

# Geography, Search Frictions and Endogenous Trade Costs

Giulia Brancaccio\*, Myrto Kalouptsi†, Theodore Papageorgiou‡§

*Princeton University, Harvard University, CEPR and NBER, McGill University*

June 2017

We leverage detailed data on vessel movements and shipping contracts to shed new light on world trade costs and trade flows. The data reveal new facts about shipping patterns, and motivate us to build a framework modeling the behavior of exporters and ships. Our framework has two novel features: (i) trade costs are endogenous and determined jointly with trade flows; as a result, they depend on the entire network of countries; (ii) search frictions between exporters and ships can limit trade. We estimate the model and recover flexibly the matching process between ships and exporters. Endogenous trade costs provide a novel link to understand trade patterns and we showcase this by considering the impact of (i) an improvement in shipping efficiency; (ii) a slow-down in China; (iii) the opening of the Northwest Passage; (iv) search frictions.

**Keywords:** geography, trade costs, matching function estimation, trade flows, transportation, shipping, search frictions, trade imbalances

---

\*Princeton University, Department of Economics, Fisher Hall, Princeton, NJ 08544, USA, giulia@princeton.edu

†Harvard University, Department of Economics, Littauer Center 124, Cambridge MA 02138, USA, myrto@fas.harvard.edu

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

§We are thankful to Pol Antras, Costas Arkolakis, Aris Christou, Manolis Galenianos, Jakub Kastl, Fabian Lange, Robin Lee, Marc Melitz, Guido Menzio and Ariel Pakes for their many helpful comments. We also gratefully acknowledge financial assistance from the Griswold Center for Economic Policy Studies at Princeton University.

# 1 Introduction

More than 80% of international trade is carried by the global shipping industry.<sup>1</sup> To export, an exporter has to find an available vessel and contract for a voyage and a price.<sup>2</sup> In turn, the ship is optimally choosing its travels in search of cargo, thinking about its future options. This spatial equilibrium between exporters and ships determines the trade costs (i.e. transportation prices) that different countries face, as well as the trade flows between different regions of the world. What is the role of geography (i.e. country locations, natural inheritance of goods) in determining trade costs and flows? Is the matching process between exporters and ships efficient? How does ship behavior affect the behavior of exporters and the resulting trade flows?

In this paper, we use detailed micro-data on vessel movements, as well as shipping contracts between exporters and ships to shed new light on world trade costs and trade flows. The data both reveal novel facts about shipping patterns, and motivate us to build a framework modeling the behavior of exporters and transportation agents (ships). Our framework has two novel features. First, trade costs are endogenous and determined jointly with trade flows. As such, trade costs depend on the entire network of countries, rather than just the bilateral (distance between) trading partners. Endogenous trade costs provide a novel link to understand trade patterns. Second, search frictions between exporters and ships can limit trade. We use the data to estimate our model and recover flexibly the matching process between ships and exporters. Finally, we use our framework to tackle a number of questions of interest: How does an improvement in shipping efficiency affect world trade flows? How do shocks propagate through the world; for instance, how would a Chinese slow-down trickle through the network of countries, or how would the opening of the Northwest Passage affect trade costs and trade flows? What is the loss due to search frictions between exporters and ships?

We focus on dry bulk ships, which carry mostly commodities (grain, ore, coal, etc.) and whereby an exporter hires the entire vessel for a specific voyage. We begin by using our data to uncover some novel facts about (i) world trade flows; (ii) trade costs; (iii) the matching process between ships and exporters. First, satellite data of ships' movements reveal that most countries are either net importers or net exporters and that, related to this, at any point in time a staggering 45% of ships are traveling without cargo (ballast). This natural trade imbalance, often overlooked in the trade literature, is a key

---

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

<sup>2</sup>Different segments of shipping function differently. In this paper we focus on bulk shipping, where exporters of bulk commodities fill up an entire vessel and hire it for a single trip in a spot market (much like a taxi or a rental car); see Section 2.2.

driver of trade costs. Second, shipping prices are largely asymmetric and depend on the destination’s trade imbalance: all else equal, the prospect of having to ballast after the destination port 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. Third, we uncover evidence suggestive of search frictions between ships and exporters: 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. Moreover, the law of one price does not hold: shipping prices exhibit substantial dispersion within a time-origin-destination triplet.

We build a dynamic spatial search model of the global shipping industry, in the spirit of the search and matching literature, that captures the observed empirical patterns and explores the importance of endogenous trade costs. The model features three key ingredients: (i) geography; (ii) search frictions; (iii) forward-looking ships and exporters that optimally choose their travels and exporting destinations respectively. Geography enters the model through different trip durations across different locations. In addition to natural geography, locations differ in their economic geography, namely their natural inheritance in commodities of different value. Search frictions between exporters and ships prevent the matching of all possible pairs. Ships are homogeneous and forward looking: when matched with an exporter, they negotiate the price taking into account matching opportunities and ballast options at the destination. When unmatched, ships decide whether to wait at their location or ballast someplace else, taking into account their expected discounted stream of profits at each location.

We estimate our model using the collected data. There are two sets of core model primitives: (i) the matching function between ships and exporters, as well as the global distribution of searching exporters; and (ii) ship sailing and port waiting costs, as well as exporter valuations and costs. Using data on the number of ships and matches, as well as the weather, we obtain the former; using data on shipping prices, ship ballast choices and exporter destination choices, we estimate the latter.

In particular, we adopt a novel approach to flexibly recover both the matching function (which gives the number of matches as a function of the number of agents searching on each side of the market), as well as the searching exporters. A sizable literature has estimated matching functions in different contexts (e.g. labor markets, taxicabs).<sup>3</sup> Here, we observe ships and matches, but not searching exporters. Our approach draws from the literature on nonparametric identification (Matzkin (2003)) and, to our knowledge, we are

---

<sup>3</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, Frechette et al. (2016) and Buchholz (2016) have used a parametric assumption on the matching function, to recover the passengers.

the first to apply it to matching function estimation. It relies on the joint density of matches and ships, as well as sea weather as an instrument that exogenously changes arriving ships. We make two contributions. First, unlike the existing literature, 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 implications. To provide some intuition, consider the following test for search frictions: weather shocks exogenously shift ship arrivals at port; in regions with more ships than exporters, this should not affect matches unless there are search frictions. We show that here, matches are indeed affected by weather shocks, which both suggests that search frictions are present and allows us to recover the curvature of the matching function.

The remaining primitives are obtained from ship and exporter choices, and prices. In particular, we first recover ship sailing and port costs via Maximum Likelihood, formed by the optimal ballast choice probabilities. As ships are forward looking this is a dynamic discrete choice problem (similar to Rust (1987)). Then, we obtain exporter valuations directly from prices: once ship primitives are known, we employ the surplus sharing condition derived from Nash bargaining to back out each valuation corresponding to each individual contract price. Finally, we use trade flows (loaded trips) to recover exporter costs by destination.

Our estimated model provides a unique framework to study how trade costs and flows are jointly determined. The novel feature here is that trade costs are endogenous and we perform a number of counterfactual exercises to illustrate why this matters. More specifically, our counterfactuals showcase three mechanisms present in our economy: First, changes in the model’s primitives also affect ships’ outside opportunities and this has an indirect effect on prices and exports. Second, the change in a country’s trade costs depends heavily on its network of trading partners and its geographical proximity to large exporters and importers. Third, reductions in impediments to trade, such as improvements in transportation efficiency or a reduction in search frictions, benefit disproportionately exporters who are (i) large; and (ii) export high value commodities, as ships are more likely to reallocate there. As a result this leads to “polarization”, whereby differences in export volumes of different countries widen.

We first consider a decline in the cost of sailing (improvement in shipping efficiency). This naturally increases the value of a match between a ship and an exporter and thus pushes down the price while increasing exports. However, the decline in the sailing cost improves a ship’s bargaining position, as it

makes ballasting less costly. This latter effect dampens the original increase in exporting, and it pushes up the exports of net exporters disproportionately more than the exports of net importers.

We illustrate the importance of trade networks and market conditions in neighboring countries, by considering a slow-down in China. In such an event, the reallocation of ships over space can amplify the effect of the slow-down in neighboring regions. More specifically, besides the direct effect to countries whose exports relied heavily on the Chinese economy, our model points out that there is a secondary effect driven by the reduced supply of ships in that region of the world: the many ships that ended up in China are no longer around. This impacts negatively both China’s own exports (import-export complementarity), but also neighboring countries’ toward which these ships would ballast; in contrast, distant countries may benefit from ships reallocating there.

We also consider the opening of the Northwest Passage. Melting the arctic ice would reduce the travel costs between Northeast America and the Far East; although the former experiences an increase in exporting, the latter suffers a decline because of the higher outside option of ships ending a trip there. Moreover, although the shock is local, it has global effects: other countries’ exports are also lower due to ships’ higher outside options. Exporters close to Northeast America (e.g. Brazil) are disproportionately hurt, as ships that used to ballast there now ballast to Northeast America; in contrast, other big exporters, such as Australia, are shielded by their closeness to the biggest importers (China, India).

Finally, we quantify the trade lost due to search frictions. We demonstrate that exporting universally goes up considerably. In addition, trade shifts towards countries with bigger exporters, as differences in frictions across regions are no longer relevant and exporter size becomes a more important determinant of trade.

**Related Literature** We relate equally 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 this literature, trade costs are treated as “residuals” that explain the gap between actual and predicted bilateral trade flows conditional on variables such as size, distance, common border/language, etc. Here, we consider what happens to trade flows when 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. In addition, we document important

features of trade costs, often not taken into account; for instance, trade costs are asymmetric and depend on the trip’s origin and destination, as well as the entire country network. 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; Asturias (2016) explores the impact of the number of shipping firms on transport prices and trade; Wong (2017) 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)).

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 exploring how the location and neighborhood of each country interact with the functioning of transportation agents and thus shapes up trade costs and flows. We illustrate that both a country’s location (i.e. distances from all other countries) and its natural inheritance are key features of the equilibrium. This natural asymmetry contributes to trade imbalances which are often overlooked in the literature (one exception is Reyes-Heroles (2016)); as we argue however, these imbalances are crucial in determining a country’s exports and trade costs. One feature of the trade literature that our paper is missing is that we do not determine product and input prices in equilibrium. The latter may be reasonable as we focus only on commodities and so wages and capital prices may be taken as exogenous. The former would require additional data and is an interesting avenue for future research.

Second, our paper relates to the search and matching literature (see Rogerson et al. (2005) for a survey). On one hand, our model is essentially a search model in the spirit 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. Lagos (2000), Lagos (2003) and Buchholz (2016) have also adopted similar spatial search models for taxi cabs; an important difference here is that prices are set in equilibrium,

while in the taxi market prices are exogenously set by regulation. In our setup this is important, since by endogenizing the trade costs we can consider how they change trade flows in the different counterfactuals. Moreover, as mentioned 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), Aguirregabiria and Mira (2007), Bajari et al. (2007), Pakes et al. (2007)) in matching conditional choice probabilities that involve value functions (applications include Ryan (2012) and Collard-Wexler (2013)). Buchholz (2016) and Frechette et al. (2016) also explore dynamic decisions in the context of taxi cabs' search and shift choices respectively. Finally, Kalouptsi (2014) has also looked at the shipping industry, albeit at the entry decisions of shipowners and the resulting investment cycles in new ships, while Kalouptsi (2017) 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 a number of facts on trade flows, transportation prices and search frictions. Section 4 describes the model. Section 5 lays out our empirical strategy, while Section 6 presents the estimation results. Section 7 discusses the counterfactuals, while Section 8 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 Trade in Dry Bulk Commodities

Dry bulk shipping involves vessels designed to carry a homogeneous unpacked dry cargo, for individual shippers on non-scheduled routes. The entire cargo carried belongs to one owner (the exporter). Bulk carriers operate like taxi cabs: a specific cargo is transported individually in 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>4,5</sup>

There are four different 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 (see Kalouptside (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.

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>6</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 2001 to 2016, collected by Clarksons Research, a major global shipbroking firm. An observation is a transaction between a shipowner and a charterer, for transportation of a specific cargo, on specific dates, from a specific origin to a specific destination. We observe the name of the vessel, the identity of the charterer who hires the ship, the contract signing date, the agreed loading and unloading dates, the agreed upon trip 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 2009 and 2016. AIS transceivers automatically broadcast information, such as the ships’ positions (longitude and latitude), speed, and level of draft (i.e. the vertical distance between the waterline and the bottom of the ship’s hull), at regular intervals of at most six minutes. The level of draft allows us to determine whether a ship is loaded or not at any point in time.

---

<sup>4</sup>It is worth noting that bulk ships are very different from containerships, which operate like buses: containerships carry cargos (mostly manufactured goods) from many different cargo owners in container boxes, along fixed itineraries/routes according to a timetable. It is technologically impossible to substitute bulk with container shipping.

<sup>5</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 world trade value (WTO (2015)).

<sup>6</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.

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  $0.75^\circ$  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 numerous sources (e.g. UNCTAD, FAO, IEA, Comtrade).

	Mean	St. Dev	Median	Min	Max
Contract price per day ( $10^5$ US dollars)	0.105	0.054	0.095	0.01	0.7
Contract trip price ( $10^5$ US dollars)	1.417	9	1.17	0.07	8.315
Contracts per ship	2.108	1.445	2	1	12
Loaded trip duration (weeks)	2.50	2.06	2.02	0.14	10.41
Empty trip (ballast) duration (weeks)	1.74	1.63	1.29	0.09	8.98
Days between contract signing and loading date	6.11	6.686	4	0	28
Prob of ship staying at port conditional on being unmatched	0.77	0.12	0.76	0.59	0.95

Table 1: Summary statistics.

**Summary statistics** Our final dataset involves 5,410 ships between 2012 and 2016.<sup>7</sup> We end up with 7,652 shipping contracts, for which we know the price, as well as the exact origin and destination. As shown in Table 1, the average price is 10,000 dollars per day (or 140,000 dollars if we take the trip duration into account), with substantial variation (the standard deviation is 5,000 dollars per day). We have 233,580 ship-week observations at which the ship decides to either ballast someplace or stay at its current location. Loaded trips last on average 2.5 weeks, with some variation within an origin-destination pair. Ballast trips last less, 1.74 weeks on average. Contracts are signed on average six days prior to the loading date.<sup>8</sup> Upon signing a contract, about 42% of ships are already in the loading port. Ships that do not find a cargo, remain in their current location with probability 77%. Clarksons reports the product carried in a small subsample of the contract dataset (about 20%). The main commodity categories are grain (23%), iron ore (20%), coal (20%), alumina ore (6%), chemicals/fertilizer (6%) and minor bulks like wood chips and sands. Finally, it is worth noting that our sample period is one of low shipping demand and severe ship oversupply due to high ship investment between 2005 and 2008 (see Kalouptside (2014,

<sup>7</sup>We only use contracts during the sample period of the satellite data. Moreover, we drop the first two years (2009-2011) in the matching function estimation, as satellites are still launched at that time and the geographic coverage is more limited.

<sup>8</sup>As practitioners say, “a ship is not a train”; it is not possible for a ship to promise too far in advance arrival to load at a specific port, due to the uncertainties of prior travels (weather, port/canal congestion, port strikes, etc.).

2017)).

### 3 Facts

In this section, we present a number of novel facts: we document geographic patterns of trade through vessel movements, we explore the nature of trade costs (i.e. shipping prices) and how they correlate with trade imbalances and product values, and we discuss some descriptive evidence suggestive of search frictions. Our findings guide the model formulation in Section 4.

#### 3.1 Trade Flows, Geography and Ballasting

Figure 1 plots vessel locations during a 10 day period. It reveals that some of the most frequent voyages are between Australia and China, Brazil and China, as well as Northwest America and China. This graph does not distinguish between loaded and ballast voyages; Figure 14 in the Appendix presents a chart of the loaded trips between world regions. The most popular loaded trip is from Australia to China (around 5% of loaded trips in our dataset). The most popular ballast trip is the reverse, from China to Australia (5.7% of ballast trips). It is no accident that China dominates the observed flows: in recent years, Chinese growth has led to massive imports of raw materials for industrial expansion and infrastructure building. In turn, Australia, Brazil and Northwest America are large exporters of minerals, grain, coal, etc.

China’s example suggests that global trade in commodities features a substantial imbalance, partly owing to the natural inheritance of different countries in raw materials: we next illustrate that this feature is quite general and that most countries are either net importers or net exporters of commodities. Figure 2 plots the difference between the number of ships departing loaded and the number of ships arriving loaded, over the sum of the two. A positive ratio indicates that a country is a net exporter, while a negative ratio suggests that the country is a net importer; a ratio close to zero implies balanced trade. As shown, trade flows in most countries in the world are considerably imbalanced. Australia, Brazil and Northwest America are big exporters, whereas China and India are big importers. This feature of trade is not unique to raw materials; container shipping exhibits similar asymmetries, suggesting that trade imbalances may affect trade costs through a mechanism similar to the one discussed in the present paper; the direction of the imbalance, however, may be different (e.g. China is a big exporter in containers rather than a big importer).

A consequence of the imbalanced nature of international trade, is that ships spend much of their time

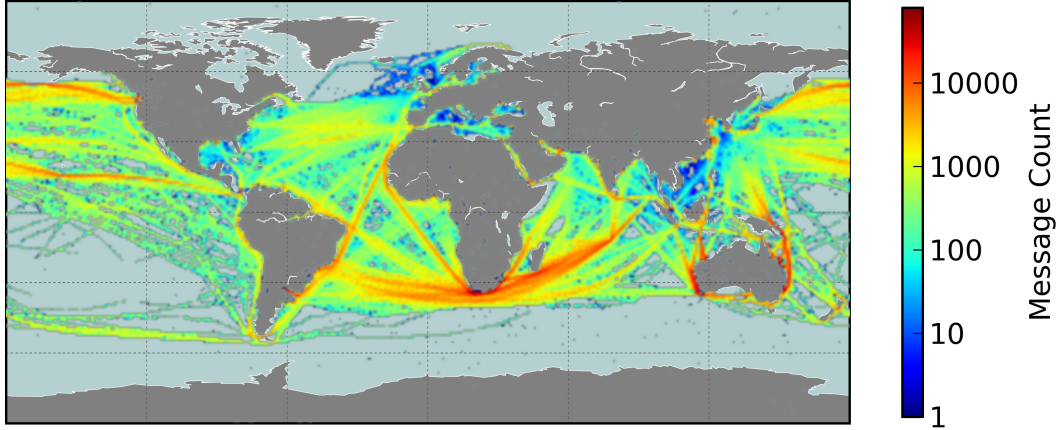


Figure 1: Number of AIS signals (messages) received in a ten day period.

traveling ballast, i.e. without cargo. We find that the fraction of the miles a ship travels ballast over the total miles traveled is about 45%.

The imbalanced nature of trade, although an important empirical feature, is often overlooked in the trade literature. In this paper, we do not assume balanced trade and in fact this asymmetry is a key driver of our model and empirical findings.

### 3.2 Trade Costs (Shipping Prices)

We next turn to the nature of trade costs that exporters face. 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 day, while a trip from Australia to China costs substantially more, at 10,000 dollars per day on average, net of fuel costs.<sup>9</sup> In fact, most trips exhibit substantial asymmetry: the average ratio of the price from origin  $i$  to destination  $j$  to the price from  $j$  to  $i$  (highest over lowest), is equal to 1.6 and can be as high as 4.1.<sup>10</sup> This empirical pattern suggests that bilateral distance is not the only determinant of trade costs.

To delve deeper into the determinants of trade costs, we present some price regressions in Table 2. The first column presents the results of a log-price regression on ship types and country of origin fixed effects which already account for 66% of the price variation, suggesting that geography is important in explaining trade costs. The second column adds destination fixed effects and, interestingly, the fraction of

<sup>9</sup>This price asymmetry has been documented also in container shipping; see e.g. Wong (2017) and references therein.

<sup>10</sup>This is calculated using the 15 geographical regions employed in our empirical exercise below (see Section 6), to guarantee sufficient data per origin/destination.

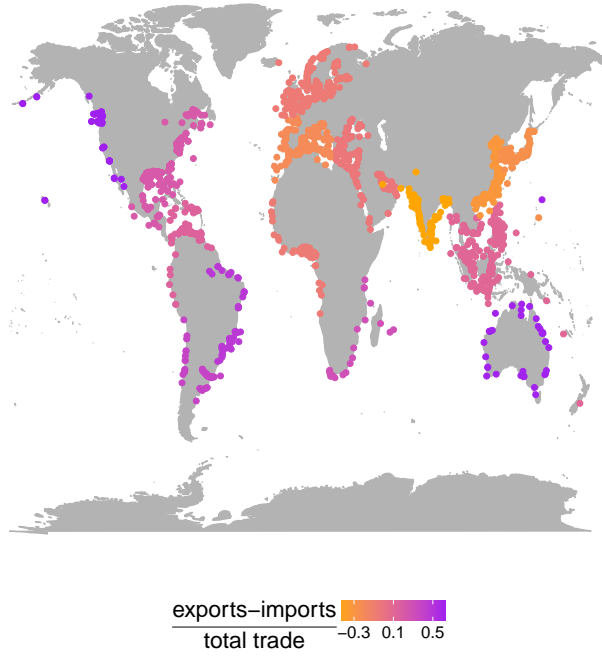


Figure 2: Trade Imbalance. Difference between exports (ships leaving loaded) and imports (ships arriving loaded) over total trade (all ships).

price variation explained increases. This suggests that ships may demand a premium to travel towards a destination with low exports, to compensate for the difficulty of finding a new cargo originating from that destination. To control more directly for this effect, in the third column of Table 2 we consider (i) the probability that ships leave ballast from the destination; and (ii) conditional on leaving ballast, the miles a ship travels ballast from the destination on average. As expected, we find that both variables are positive and significant and lead to substantially higher prices. Indeed, a 1% increase in the average distance traveled ballast after the destination, is associated with a 0.17% increase in prices. Similarly, exporting to a destination where the probability of a ballast trip afterwards is ten percentage points higher, costs on average 2.3% more.<sup>11</sup>

Finally, it is worth emphasizing that ship heterogeneity is not a first order issue. First, even in our small dataset, the majority of ships (80%) are seen carrying at least 2 of the 3 main products (coal, ore and grain), which suggests that ships do not specialize on certain products. Similarly, we observe that most ships travel to most regions, suggesting that they do not specialize geographically either. Ship fixed effects have no explanatory power in either price regressions or ballast probability regressions. Finally, this is consistent with much of the evidence provided in Kalouptsi (2014); for instance hedonic regressions of ship resale prices suggest that unobserved heterogeneity is not an important consideration.

<sup>11</sup>To confirm that this result is not driven by the different composition of products exported towards different destinations, the last column of Table 2 also controls for the product carried for the subsample of contracts reporting this information. We discuss how trade costs depend on the value of the product in Section 3.3.

	I	II	III	IV
	log(price per day)			
Handymax	-0.148** (0.014)	-0.136** (0.014)	-0.123** (0.014)	0.027 (0.120)
Handysize	-0.397** (0.017)	-0.330** (0.018)	-0.343** (0.017)	-0.209** (0.124)
Panamax	-0.223** (0.013)	-0.214** (0.013)	-0.212** (0.013)	-0.117 (0.119)
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)
Probability of ballast			0.234** (0.030)	0.556** (0.081)
Average duration of ballast trip (log)			0.166** (0.014)	0.065** (0.032)
Constant	10.304** (0.068)	10.284** (0.103)	9.127** (0.099)	8.915** (0.408)
Destination FE	No	Yes	No	No
Origin FE	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes
Quarter FE	Yes	Yes	Yes	Yes
Obs	11,014	11,014	11,011	1,662
R <sup>2</sup>	0.663	0.694	0.674	0.664

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

Table 2: Shipping price regressions.

### 3.3 Search Frictions

In this section we present some descriptive evidence suggesting that search frictions inhibit the matching of all available ships and exporters. Overall, it is not straightforward to know a priori whether a market (here, the market for sea transport) suffers from search frictions. In labor markets, where search frictions are generally thought to be present, two main empirical regularities are often offered as evidence: (i) the coexistence of unemployed workers and vacant firms; and (ii) wage inequality among observationally identical workers. In this section we show that a similar set of empirical conditions hold in shipping, suggesting that search frictions may be present here as well.

We first consider whether there is evidence of unrealized matches, along the lines of (i). In particular, we cannot replicate the argument done for labor markets, since our data reports only ships and matches, not searching exporters (similar to vacancies in labor markets); we can, however, consider a different moment that has a similar flavor.

Figure 3 displays the weekly number of ships that arrive empty and load, as well as the number of ships that leave empty, in Norway and in Chile. Both countries are net exporters. In Norway, several ships arrive empty and load, but almost no ship departs empty. In Chile, however, the picture is quite different: we simultaneously see ships that arrive empty to load, and ships that depart empty. In other words, 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: why does the ship that depart empty, not pick up the cargo, instead of having another ship arrive from elsewhere to pick it up?

We next perform this exercise for all net exporting countries, by computing the bi-weekly ratio of the incoming empty and loading ships over the outgoing empty ships for a given country. In the absence of frictions, one may expect this ratio to be close to zero. However, as shown in Figure 4 which depicts the histogram of these ratios, most countries are more similar to Chile, than Norway. Indeed, the average ratio is well above zero and for some countries it is even above 0.5. In addition, this pattern is quite robust in a number of dimensions.<sup>12</sup> While in labor markets, as some researchers have argued, observed or unobserved heterogeneity may partly explain the co-existence of unemployment and vacancies (a vacancy

---

<sup>12</sup>This figure is robust to alternative market definitions, time periods and ship types. Figure 15 in the Appendix presents this histogram by ship type: 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 the same if done by port rather than country. As mentioned above, ships tend to carry all products; thus we do not believe this pattern is explained by product switching. Labor contracts are usually about 5-8 months long and the crew flies between their home and the relevant port. 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 toward 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%).

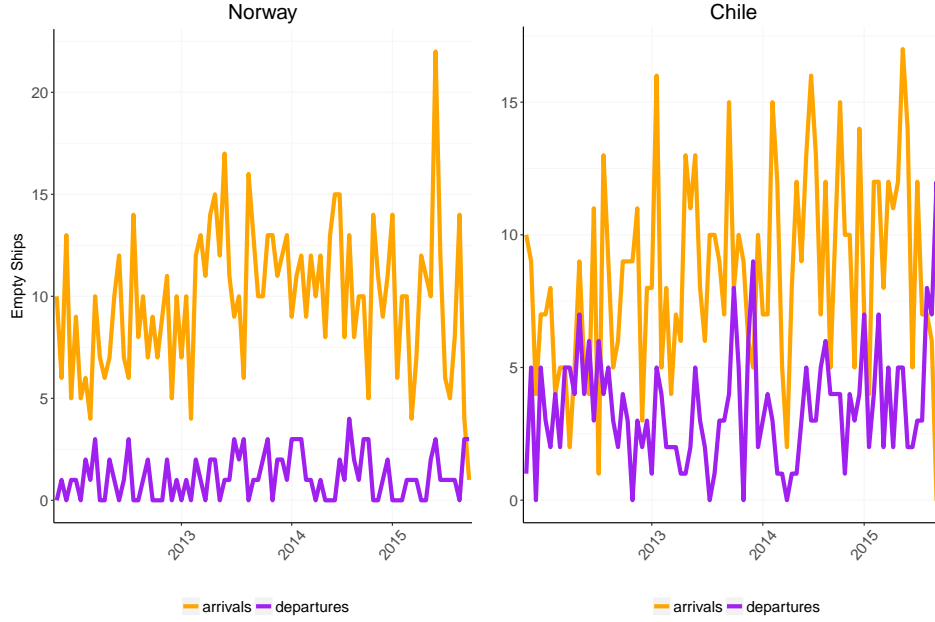


Figure 3: Flow of ships arriving empty and loading, and ships leaving empty in 2 week intervals.

for a chemical engineer may not be of interest to a high school dropout), in this market the importance of heterogeneity is much more limited: as discussed above ships are widely considered to offer homogeneous services and do not specialize geographically or in terms of products.

Again inspired by the labor literature, we investigate a second aspect of this market that is suggestive of search frictions: 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. This observation has generated a substantial and widely influential literature on frictional wage inequality, i.e. wage inequality that is driven by search frictions.<sup>13</sup>

In the shipping market a similar empirical regularity is present. 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).<sup>14</sup> For instance, in the most popular trip from Australia to China the coefficient of variation is on average 37% and ranges from 21% to 55% across quarters.

<sup>13</sup>See for instance Burdett and Mortensen (1998), Postel-Vinay and Robin (2002), Mortensen (2003) and references therein.

<sup>14</sup>Again here we use the 15 geographical regions employed in our empirical exercise below (see Section 6), to guarantee sufficient data per origin/destination.

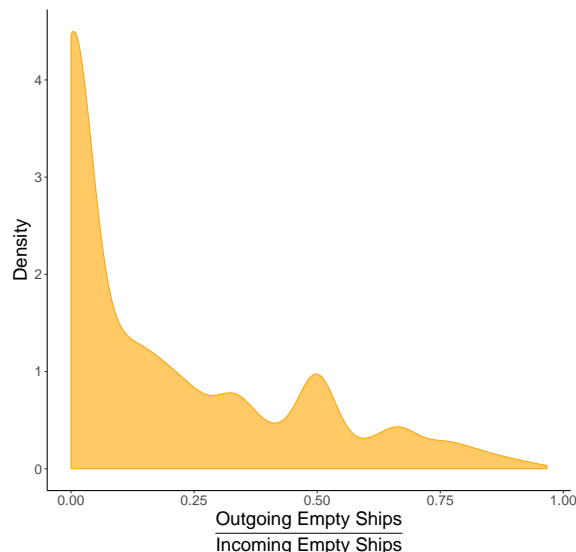


Figure 4: Histogram of the ratio of outgoing empty over incoming empty and loading ships in net exporting countries.

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. In the absence of frictions, if there are more ships than exporters, as is arguably the case during our sample period, we would expect prices to be bid down to the ships’ opportunity cost.<sup>15</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. In the context of the shipping market, our model predicts that exporters with higher valuations pay more when there are search frictions, consistent with the evidence in Table 2.

Finally, we perform a simple “dispatcher” simulation, which assigns every observed match to the ship that is closest to it. We find that the fraction of distance a ship ballasts in the simulation is 38% on average, significantly less than in the data (45%). In this simulation, ships do not optimally choose where to search; we return to this in the counterfactuals of Section 7.

We revisit search frictions in Section 5.1, where we both provide a test, as well as estimate a non-parametric matching function, thus flexibly measuring the extent of search frictions. We close this section by briefly discussing what the nature of search frictions may be. The mere existence of brokers suggests that search frictions are an issue in this market. Information frictions may still prevail though; when a ship is unmatched in a certain geographical region (e.g. East Americas) her broker may not “meet” the broker of a specific exporter in one of the ports. Interviews with industry participants indeed suggest that information

<sup>15</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.

can get lost between the different brokers of different shipowners and exporters; interestingly, oftentimes more than two brokers may mediate the deal. Unfortunately direct data on brokerage arrangements are not available. In a market with a small number of large exporters, it might be easier for them to be matched to the existing ships. Consistent with this, we find that price dispersion is negatively correlated with the Herfindahl Index of the observed ship charterers.

## 4 Model

We next introduce a dynamic spatial search model of the global shipping industry. Geography enters the model through different trip durations across different locations. There are two types of agents: exporters and ships. Exporters choose whether and where to export, while unmatched ships choose where to ballast. Following the search and matching literature we model new matches every period using a matching function, which captures the implications of frictional trading, in a parsimonious fashion (Pissarides (2000)). In other words, we do not explicitly model the meeting technology between exporters and ships, but given the evidence presented in Section 3.3 we allow for the possibility of search frictions.

### 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  $\beta$ .

**Exporters** At each location  $i$  and period  $t$ , there are  $f_{it}$  freights that need to be delivered to another location. We use the words freight, exporter and cargo interchangeably. Exporters have heterogeneous valuations,  $v$ , from exporting to their destination. The valuation of a freight going from  $i$  to  $j$  is drawn from the distribution  $F_{ij}^v$  with mean  $\mu_{ij}$ .<sup>16</sup> Unmatched freights survive with probability  $\delta > 0$ . 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 valuation  $v$ .

---

<sup>16</sup>In this paper, we do not consider the determination of commodity prices; in other words, we take exporter valuations to be exogenous.  $v$  is meant to capture the exporter's revenue. Determining this object in equilibrium requires additional data on exporters and is an interesting avenue for future research.

**Ships** There are  $S$  homogeneous ships in the world.<sup>17,18</sup> In every period, a ship is either traveling loaded or ballast, from some location  $i$  to some location  $j$ , or it is at port in some region  $i$ . A ship at port in location  $i$  incurs a per period cost  $c_i^u$ , while a ship sailing from  $i$  to  $j$  incurs a per period cost  $c_{ij}^s$ . The duration of a trip between region  $i$  and region  $j$  is stochastic: a traveling ship arrives at  $j$  this period with probability  $\xi_{ij}$ , so that the average duration of the trip is  $1/\xi_{ij}$ .<sup>19</sup>

**Matching** Freights can only be delivered to their destination by ships and each ship can carry (at most) one freight. As discussed earlier, we capture the process through which exporters are paired with available ships in a market by a matching function, whereby the number of matches at time  $t$  are

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

where  $s_{it}$  is the number of unmatched ships in  $i$ . The matching function  $m_i(s_{it}, f_{it})$  is increasing in both arguments, and whenever  $m_{it} = \min(s_{it}, f_{it})$  the matching process is frictionless. Let  $\lambda_{it}$  denote the probability with which an unmatched ship in location  $i$  meets a freight;  $\lambda_{it} = m_{it}/s_{it}$ . Similarly, let  $\lambda_{it}^f$  denote the probability with which an unmatched freight meets a ship;  $\lambda_{it}^f = m_{it}/f_{it}$ .

When a ship and a freight meet, they can either agree on a price to be paid by the freight to the ship or they both revert to their outside options. Note that the outside option of a freight 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 between the freight and the ship via the price-setting mechanism. We assume that the price,  $\tau_{ijv}$ , paid to the ship delivering a freight of valuation  $v$  from region  $i$  to destination  $j$ , is determined by generalized Nash bargaining, with  $\gamma \in (0, 1)$  denoting the exporter's bargaining power. The price is paid upfront and the ship commits to begin its voyage immediately to region  $j$ .

In what follows, 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.<sup>20</sup> The agreed upon price guarantees that

---

<sup>17</sup>We follow Kalouptsi (2014) and assume constant returns to scale so that a shipowner is a ship. Similarly, a freight owner is a freight, so that he does not choose the export tonnage. We also ignore the different ship sizes in the model; estimation results are robust when we consider types separately.

<sup>18</sup>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, 2017).

<sup>19</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>20</sup>An exporter  $(v, j)$  either forms a match with any ship it meets as long as his value,  $v$ , is high enough to generate a positive surplus given the destination,  $j$ , or cannot agree to a mutually acceptable price with any ship. Given that in this context valuations are an order of magnitude greater than the shipping price, this assumption is fairly innocuous (see Section 5.3).

the exporter prefers not to wait for the following period (recall that ships are homogeneous) and that the ship does not gain from either waiting or ballasting elsewhere.

Finally, it is worth noting that matching occurs only if ships and freights are in the same region. This assumption is consistent with our data, where most ships are already in the region of the freight when signing a contract; contracts are signed only six days before loading; and there is substantial ship excess supply in all regions. In our estimation, each “region” is fairly broad usually comprising of several countries.

**Timing** The timing of each period is as follows:

1. In each region  $i$ , existing ships and exporters match.
2. In each region  $i$ , ships at port pay the port costs, draw additive iid preference shocks  $\epsilon = [\epsilon_1, \dots, \epsilon_I] \in \mathbb{R}^I$  from distribution  $F^\epsilon$  and decide whether to (i) stay in their current region and wait for freight; or (ii) ballast toward some destination  $j$ .<sup>21</sup>
3. Unmatched ships that decided to ballast away begin traveling to their chosen destination. All ships already traveling from  $i$  to  $j$  arrive at  $j$  with probability  $\xi_{ij}$ . Existing exporters disappear with probability  $1 - \delta$ .
4. In each region  $i$ ,  $\mathcal{E}_i$  potential exporters decide whether and to which destination to export to. The exporters that do enter the market draw their valuations from  $F_{ij}^v$ , pay cost  $\kappa_{ij}$  and join the pool of unmatched exporters the following period.

**States and Transitions** The state variable of a ship in region  $i$  includes its current location  $i$ , as well as the vector  $(s_t, f_t, s_t^w)$  where  $s_t = [s_{1t}, \dots, s_{It}]$ ,  $f_t = [f_{1t}, \dots, f_{It}]$  and the  $I^2 - I$  dimensional vector  $s_t^w$ , with entries  $s_{ijt}^w$ , denotes the number of ships traveling from  $i$  to  $j$  in period  $t$ . The state variable of an existing exporter in  $i$  includes his location  $i$ , valuation  $v$  and destination  $j$ , as well as the vector  $(s_t, f_t, s_t^w)$ . Exporters in region  $i$  at time  $t$  transition as follows:

$$f_{it+1} = \delta(f_{it} - m_i(s_{it}, f_{it})) + d_i \quad (1)$$

---

<sup>21</sup>Recall that the average trip duration is 2-3 weeks, while contracts are signed on average six days in advance, so that ships often sail towards a destination without having signed a contract already. This is fairly intuitive given the large trade imbalances and high port costs.

with  $d_i$  the (endogenous) flow of new freights. Ships at location  $i$  transition as follows:

$$s_{it+1} = (s_{it} - m_i(s_{it}, f_{it})) P_{ii} + \sum_{j \neq i} \xi_{ji} s_{jit}^w \quad (2)$$

where  $P_{ij}$  is the probability of an unmatched ship ballasting from  $i$  to  $j$  (determined endogenously, see below). 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  $\xi_{ji}$  arrive. Finally, ships that are traveling from  $i$  to  $j$ ,  $s_{ijt}^w$  evolve as follows:

$$s_{ijt+1}^w = (1 - \xi_{ij}) s_{ijt}^w + P_{ij} (s_{it} - m_i(s_{it}, f_{it})) + G_{ij} m_i(s_{it}, f_{it}) \quad (3)$$

where  $G_{ij}$  is the probability of going loaded from  $i$  to  $j$  (determined endogenously, see below). In words, fraction  $\xi_{ij}$  of the traveling ships arrive, fraction  $P_{ij}$  of ships that remained unmatched in location  $i$  chose to ballast to  $j$  and finally,  $G_{ij}$  of ships matched in  $i$  depart loaded to  $j$ .

## 4.2 Equilibrium

We derive the optimal behavior of exporters and ships, as well as the equilibrium prices and trade flows. In this paper, we consider the steady state of this model.<sup>22</sup> This assumption is not unreasonable for the data at hand, which covers a period (2012-2016) that is uniformly characterized by ship oversupply and relatively low demand for shipping services. More specifically, we assume that agents view the spatial distribution of ships and freights,  $(s_t, f_t, s_t^w)$ , as fixed and make decisions based on their steady-state values: given the short-lived nature of their decisions (where to ballast) it does not feel unreasonable that they ignore aggregate long-run shocks when making these weekly choices.

**Ships** Let  $W_{ij}$  denote the value of a ship traveling from  $i$  to  $j$  (empty or loaded). Then:

$$W_{ij} = -c_{ij}^s + \xi_{ij} \beta U_j + (1 - \xi_{ij}) \beta W_{ij} \quad (4)$$

In words, the ship that is traveling from  $i$  to  $j$ , pays per period cost of transit  $c_{ij}^s$ ; with probability  $\xi_{ij}$  it arrives at its destination  $j$ , where it will begin unmatched with value  $U_j$  defined below; finally, with the

---

<sup>22</sup>Given that a ship can travel to most ports in the world in under a month, any transition dynamics to a new steady state will be short. This is convenient in our counterfactual analysis of Section 7, where we are able to compare steady states directly.

remaining probability  $1 - \xi_{ij}$  the ship does not arrive and keeps traveling. The ship arrives at  $j$  after  $1/\xi_{ij}$  periods on average.

Consider now a ship in region  $i$ . This ship obtains:

$$U_i = -c_i^u + \lambda_i E_{j,v} V_{ijv} + (1 - \lambda_i) J_i \quad (5)$$

In words, the ship is matched with probability  $\lambda_i$ , in which case it obtains the value of a matched ship  $V_{ijv}$  defined below. The ship takes expectation over the type of freight it meets, i.e. its value and destination. With the remaining probability,  $1 - \lambda_i$ , the ship does not find a freight and it obtains the value  $J_i$  also defined below. Finally, the ship pays the per period port cost  $c_i^u$ .

If matched with an exporter with value  $v$  and destination  $j$ , the ship receives the agreed upon price,  $\tau_{ijv}$ , and begins traveling, so that:

$$V_{ijv} = \tau_{ijv} + W_{ij} \quad (6)$$

If the ship remains unmatched, it faces the choice of either staying at  $i$  and matching there the following period with probability  $\lambda_i$ , or ballasting away from  $i$  in search of better opportunities. In the latter case, the ship can choose among all possible destinations. In particular, if unmatched, the ship draws preference shocks  $\epsilon \in \mathbb{R}^I$ , from a double exponential distribution  $F^\epsilon$ , with standard deviation  $\sigma$ . The unmatched ship's value function is :

$$J_i(\epsilon) = \max \left\{ \beta U_i + \sigma \epsilon_i, \max_{j \neq i} W_{ij} + \sigma \epsilon_j \right\} \quad (7)$$

and let:

$$J_i \equiv E_\epsilon J_i(\epsilon) = \sigma \log \left( \exp \frac{\beta U_i}{\sigma} + \sum_{j \neq i} \exp \frac{W_{ij}}{\sigma} \right) + \sigma \gamma^{euler}$$

where  $\gamma^{euler}$  is the Euler constant.<sup>23</sup> In words, if the ship stays in its current region  $i$ , it obtains value  $U_i$ ; 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, while  $P_{ij}$  denote the probability that a ship at  $i$  chooses to ballast to  $j$ . We have:

$$P_{ii} = \frac{\exp(\beta U_i / \sigma)}{\exp(\beta U_i / \sigma) + \sum_{l \neq i} \exp(W_{il} / \sigma)} \quad (8)$$

---

<sup>23</sup>The formula for the ex ante value function  $J_i = E_\epsilon J_i(\epsilon)$  is the closed form expression for the expectation of the maximum over multiple choices, and is obtained by integrating  $J_i(\epsilon)$  over the double exponential distribution of  $\epsilon$ .

and

$$P_{ij} = \frac{\exp(W_{ij}/\sigma)}{\exp(\beta U_i/\sigma) + \sum_{l \neq i} \exp(W_{il}/\sigma)}. \quad (9)$$

**Exporters** We start with existing exporters and then consider exporter entry. An exporter that is matched in location  $i$  receives value:

$$V_{ijv}^f = v - \tau_{ijv}, \quad (10)$$

in words, he obtains his delivery value,  $v$  and pays the agreed price,  $\tau_{ijv}$ , which is derived below. An exporter that does not get matched receives no payoff in the period and survives with probability  $\delta$ ; if so, the following period with probability  $\lambda_i^f$  he gets matched and receives  $V_{ijv}^f$ , while with the remaining probability  $1 - \lambda_i^f$  he remains unmatched again:

$$J_{ijv}^f = \beta \delta \left( \lambda_i^f V_{ijv}^f + (1 - \lambda_i^f) J_{ijv}^f \right). \quad (11)$$

There are  $\mathcal{E}_i$  ex ante homogeneous potential exporters in location  $i$  in every period. Each potential entrant in region  $i$ , makes a discrete choice between not exporting, as well as which destination  $j$  to export to, subject to production and exporting costs  $\kappa_{ij}$ , as well as random preference shocks  $\epsilon^f \in \mathbb{R}^I$ , distributed according to a double exponential distribution. Upon deciding to become an existing exporter in  $i$  with destination  $j$ , the entrant draws valuation  $v$  from  $F_{ij}^v$ . Therefore, a potential entrant solves:

$$J_i^{ef} = \max \left\{ \epsilon_0^f, \max_{j \neq i} \left\{ E_v J_{ijv}^f - \kappa_{ij} + \epsilon_j^f \right\} \right\}$$

where we denote by 0 the (outside) option of not exporting and normalize the payoff in that case to zero.<sup>24</sup>

Potential exporters' behavior is given by the choice probabilities:

$$\tilde{G}_{ij} \equiv \frac{\exp(J_{ij}^f - \kappa_{ij})}{1 + \sum_{l \neq i} \exp(J_{il}^f - \kappa_{il})} \quad (12)$$

and

$$\tilde{G}_{i0} \equiv \frac{1}{1 + \sum_{l \neq i} \exp(J_{il}^f - \kappa_{il})} \quad (13)$$

where  $J_{ij}^f \equiv E_v J_{ijv}^f$ .

---

<sup>24</sup>It is possible to allow the potential exporters to know their valuations (across destinations) before making their exporting choice, but this would make the estimation computationally more demanding.

Therefore the number of entrant exporters in  $i$  equals:

$$d_i = \mathcal{E}_i (1 - \tilde{G}_{i0}) \quad (14)$$

It is worth noting that the distribution of export destinations conditional on exporting is given by

$$G_{ij} \equiv \frac{\tilde{G}_{ij}}{1 - \tilde{G}_{i0}} \quad (15)$$

This is the distribution that ships employ when forming expectations over the potential matches in different regions in equation (5).

**Prices** As discussed above, the rents generated by a match between a freight and a ship, are split via Nash bargaining. This implies the usual surplus sharing condition:

$$\gamma (V_{ijv} - J_i) = (1 - \gamma) (V_{ijv}^f - J_{ijv}^f) \quad (16)$$

We use the above condition to solve out for the equilibrium price  $\tau_{ijv}$ , in the following lemma:

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

$$\tau_{ijv} = \frac{\gamma (1 - \beta\delta (1 - \lambda_i^f))}{1 - \beta\delta (1 - \gamma\lambda_i^f)} (J_i - W_{ij}) + \frac{(1 - \gamma) (1 - \beta\delta)}{1 - \beta\delta (1 - \gamma\lambda_i^f)} v \quad (17)$$

*Proof.* Substitute  $V_{ijv}$ ,  $V_{ijv}^f$ ,  $J_{ijv}^f$  and  $U_i$  in (16). □

In other words, the price is a linear combination of the exporter's value,  $v$ , and the difference between the ship's value of transporting the freight,  $W_{ij}$ , and its outside option,  $J_i$ .

It is worth noting that the price is decreasing in the value of a ship traveling from  $i$  to  $j$ ,  $W_{ij}$ . Recall that  $W_{ij}$  depends on both the conditions at the destination through  $U_j$ , and distance, captured by  $\xi_{ij}$ , since  $W_{ij} = -c_{ij}^s / (1 - \beta(1 - \xi_{ij})) + \xi_{ij}\beta U_j / (1 - \beta(1 - \xi_{ij}))$ . In other words, destinations that are unappealing to ships because there are few freights there and the probability of ballasting afterwards is high, would command higher prices. This is consistent with the evidence presented in Table 2. The same holds for destinations that are further away (low  $\xi_{ij}$ ), have low value freights or severe search frictions. Moreover,  $U_j$  also controls for conditions at all possible ballast destinations from  $j$ , as well as for conditions at all

possible export destinations from  $j$ . Finally, note that the price between  $i$  and  $j$  depends on all countries rather than just  $i$  and  $j$ , both through the outside option of the ship,  $J_i$ , as well as through the conditions in the export/ballast destinations from  $j$ , captured in  $U_j$ , as discussed above. We return to the importance of the network when we perform model counterfactuals in Section 7.

In addition, exporters that have a higher value,  $v$ , pay higher prices, again consistent with evidence in Table 2. As discussed in Section 3.3, this is true because the law of one price no longer holds when there are search frictions. In a world without search frictions and more ships than freights, the shipping price is given by  $\tau_{ij} = J_i - W_{ij}$ ; therefore the previous properties of trade costs (dependence on distance, destination and entire network of countries) are independent of the presence of search frictions and *still hold*.<sup>25</sup>

**Steady State Equilibrium** We next define the steady state equilibrium for this model and prove that it exists.

**Definition.** A steady state equilibrium,  $(s^*, f^*, s^{w*})$ , is a distribution of ships and exporters over locations, that satisfies the following conditions:

- (i) Ships' optimal behavior,  $P_{ij}(s^*, f^*, s^{w*})$  follows (8) and (9) and expectations employ (15).
- (ii) Potential exporters' behavior,  $\tilde{G}_{ij}(s^*, f^*, s^{w*})$ , follows (12) and (13) and entrants are determined from (14).
- (iii) Prices are determined by Nash bargaining, according to (17).
- (iv) Ships and freights satisfy the following steady state equations (established in Proposition 1 below):

$$s_i^* = \sum_j P_{ji}(s^*, f^*, s^{w*}) (s_j^* - m_j(s_j^*, f_j^*)) + \sum_{j \neq i} G_{ji}(s^*, f^*, s^{w*}) m_j(s_j^*, f_j^*) \quad (18)$$

$$f_i^* = \delta (f_i^* - m_i(s_i^*, f_i^*)) + E_i \left( 1 - \tilde{G}_{i0}(s^*, f^*, s^{w*}) \right) \quad (19)$$

$$s_{ij}^{w*} = \frac{1}{\xi_{ij}} (P_{ij}(s^*, f^*, s^{w*}) (s_i^* - m_i(s_i^*, f_i^*)) + G_{ij}(s^*, f^*, s^{w*}) m_i(s_i^*, f_i^*))$$

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

*Proof.* See the Appendix. □

---

<sup>25</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  $V_{ijv} = J_i$ . Substituting in for  $V_{ijv}$  from equation (6) leads to the above price equation.

**Trade Flows** Finally, we characterize the steady state trade flow between regions  $i$  and  $j$ , which is equal to:

$$\mathcal{E}_i \tilde{G}_{ij} = \mathcal{E}_i \frac{\exp(J_{ij}^f - \kappa_{ij})}{1 + \sum_{l \neq i} \exp(J_{il}^f - \kappa_{il})} = \mathcal{E}_i \frac{\exp(\alpha_i (\mu_{ij} - \tau_{ij}) - \kappa_{ij})}{1 + \sum_{l \neq i} \exp(\alpha_i (\mu_{il} - \tau_{il}) - \kappa_{il})}$$

where  $\alpha_i = \beta \delta \lambda_i^f / (1 - \beta \delta (1 - \lambda_i^f))$ , since in the steady state:

$$J_{ij}^f \equiv E_v J_{ijv}^f = E_v \frac{\beta \delta \lambda_i^f (v - \tau_{ijv})}{1 - \beta \delta (1 - \lambda_i^f)} = \alpha_i (\mu_{ij} - \tau_{ij}) \quad (20)$$

where  $\tau_{ij} \equiv E_v \tau_{ijv}$ . This equation is reminiscent of 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  $\{\lambda_i^f, \mu_{ij}, \kappa_{ij}, \mathcal{E}_i\}$  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, the endogenous trade costs,  $\tau_{ij}$ , for all  $j$ . The important addition here is that the trade flow depends on all countries through the endogenous trade cost  $\tau_{ij}$ ; indeed, recall that  $\tau_{ij}$  depends on all locations both through the outside option of the ship that can ballast anywhere, but also because the ship cares about the conditions in the ballast and export destinations from location  $j$ . Therefore, any change in the primitives affects trade flows both directly, but also indirectly through its impact on trade costs. We illustrate this when we perform model counterfactuals in Section 7.

## 5 Empirical Strategy

In this section we lay out the empirical strategy followed to estimate the model of Section 4. The main model primitives we wish to recover are: the matching function and searching exporters, the ship travel and port costs, as well as the distribution of exporter valuations and their costs. We describe the empirical strategy here, while in Section 6 we present the results.

### 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 (2016), and Frechette

et al. (2016) have used such data, coupled with a “parametric” assumption on the matching function to recover the hailing passengers.<sup>26</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. Our approach makes two contributions to the literature. First, we do not take a stance on the presence and magnitude of search frictions in the industry. Why is this important? Consider the case of no search frictions, so that the matching function is

$$m_{it} = \min(s_{it}, f_{it}) \quad (21)$$

In words, all potential matches are realized. In contrast, if there are search frictions, we have:

$$m_{it} = m_i(s_{it}, f_{it}) \leq \min(s_{it}, f_{it}) \quad (22)$$

so that some potential matches are not realized. If one side of the market is unobserved (here  $f_{it}$ ) or mismeasured (arguably, in labor markets searching workers are imperfectly measured; Lange and Papa-georgiou (2017)) it is not straightforward to differentiate (21) from (22). Indeed, when a ship/taxi is traveling empty is it because no exporter/passenger was searching or because an exporter/passenger was there but did not get to meet the ship/taxi due to frictions? Our approach, allows us to disentangle the two.

Our second contribution is to avoid imposing parametric restrictions on the matching function. The literature has imposed functional forms such as the Cobb-Douglas. The desire to be non-parametric is not just “stylistic” when it comes to matching functions: parametric restrictions are directly linked to welfare implications. 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.

Our approach borrows from the literature on nonparametric identification (Matzkin (2003)). Roughly, the method leverages (i) an invertibility assumption between matches and freights, (ii) the observed relationship between ships and matches, (iii) an instrument that shifts the number of ships, and (iv) a

---

<sup>26</sup>Buchholz (2016) assumes an “urn-ball” matching function. Frechette et al. (2016) 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.

restriction on the matching function that allows us to disentangle monotonic transformations. To provide some intuition, we outline a simple version of the methodology. We then formalize the argument. We refer the interested reader to Matzkin (2003) for further details.

Suppose that (i) the matching function  $m_i(s_i, f_i)$  is continuous and strictly increasing in  $f_i$  and, (ii) that  $s_i$  is independent of  $f_i$ . The first assumption is natural in our context: more freights should lead to more matches, all else equal. The second assumption will prove useful in presenting the estimation methodology, but is likely not valid in our case, as the spatial distribution of ships and freights is determined jointly in equilibrium; we relax this assumption below.

Let  $F_{m|s}$  denote the distribution of matches conditional on ships, and  $F_f$  the distribution of freights,  $f$ . Then at a given point  $(s_{it}, f_{it}, m_{it})$  we have:

$$\begin{aligned}
F_{m|s}(m_{it}|s_{it}) &= \Pr(m_i(s, f) \leq m_{it}|s_{it}) \\
\text{monotonicity} &= \Pr(f \leq m_i^{-1}(s, m_{it})|s_{it}) \\
\text{independence} &= \Pr(f \leq m_i^{-1}(s_{it}, m_{it})) \\
&= F_f(f_{it})
\end{aligned} \tag{23}$$

In words, the conditional distribution of matches (outcome) on ships (observed covariate) at a point  $(m_{it}, s_{it})$  is equal to the distribution of freights at the corresponding (unobserved) point  $f_{it}$ . Equation (23) is our main relationship for the identification and estimation of both freights and the matching function. However, (23) alone is not sufficient: it is not possible to distinguish monotonic transformations of  $f$  and  $m(\cdot)$ . To do so, a restriction on either the distribution  $F_f$  or the matching function is required. In this paper we assume that the matching function is homogeneous of degree one, so that:  $m_i(\alpha s_i, \alpha f_i) = \alpha m_i(s_i, f_i)$ , all  $\alpha > 0$ .<sup>27</sup> The intuition behind the identification argument is as follows: the correlation between  $s_i$  and  $m_i$  informs us on  $\partial m_i(s_i, f_i)/\partial s_i$ , since the sensitivity of matches to changes in ships is observed and  $s_i$  is independent of  $f_i$  by assumption; then, due to homogeneity, this derivative also delivers the derivative  $\partial m_i(s_i, f_i)/\partial f_i$ ; and once these derivatives are known, so is the matching function, which can now be inverted to provide the freights as well.<sup>28</sup>

---

<sup>27</sup>In the labor literature, most matching function estimates find support for constant returns to scale (see Petrongolo and Pissarides (2001)). Given that the nature of search frictions is not necessarily that different (in both cases it is a shortcut for information frictions about which ships/freights may be available), we consider this a natural starting point.

<sup>28</sup>We could alternatively impose an assumption on the distribution  $F_f$ . For example, if we assume that  $F_f$  is uniform on  $[0, 1]$ , we can use (23) to recover  $f_{it}$  pointwise from the conditional distribution of  $m$  on  $s$ ; once freights are recovered, we also instantly know the (inverse) matching function. Bajari and Benkard (2005) employ this methodology to nonparametrically

Finally, as mentioned above, independence of ships and freights is not a natural assumption in our setting. We do, however, have a plausibly valid instrument: sea weather shifts the arrival of ships in a port without affecting the number of freights. We therefore assume that ships  $s$  are a function of the instrument  $z$  and a shock  $\eta$ , such that ships and freights are independent conditional on  $\eta$ . This allows us to modify (23) by conditioning on  $\hat{\eta}$  as well as on  $s$  to obtain the distribution of freights.

Proposition 2 formalizes these arguments:

**Proposition 2.** (i) Suppose that  $m(s, f)$  is continuous, (positively) homogeneous of degree 1 and strictly increasing in  $f$ . Suppose further that  $s$  and  $f$  are independent. Finally, suppose that  $m$  is known for a specific pair  $(s^*, f^*)$  so that  $m^* = m(s^*, f^*)$ , with  $m^* \neq 0$ . Then, the function  $m(\cdot)$  is identified.  
(ii) Suppose there exists an instrument  $z$  such that

$$s = h(z, \eta) \quad (24)$$

with  $z$  independent from  $(f, \eta)$ . Assume that proper conditions hold so that  $\eta$  can be uniquely recovered from (24). Then,  $s$  and  $f$  are conditionally independent given  $\eta$ . The distribution  $F_f$  and the matching function can be recovered from:

$$\begin{aligned} F_{f|\eta}(\phi) &= F_{m|s=\frac{\phi}{f^*}s^*, \eta} \left( \frac{\phi}{f^*} m^* \right) \\ F_f(\phi) &= \int F_{f|\eta}(\phi) f_\eta(\eta) d\eta \\ m(s, \phi) &= \int F_{m|s, \eta}^{-1}(F_{f|\eta}(\phi)) f_\eta(\eta) d\eta \end{aligned}$$

The matching function is estimated separately for each region  $i$ ; Section 6 presents the results.<sup>29,30</sup>

## 5.2 Ship Costs

We next turn to the ships' sailing costs,  $c_{ij}^s$ , and port costs,  $c_i^u$ , as well as the standard deviation of the shocks,  $\sigma$ , which is identified because the observed prices pin down the scale of payoffs (in dollars). We obtain the parameters of interest,  $\theta = \{c_{ij}^s, c_j^u, \sigma\}$ , from the ships' optimal ballast choice probabilities, (8) and (9), which are a function of the ships' value functions  $(U_i, W_{ij})$ , which in turn depend on the parameters

---

estimate hedonic price equations and unobserved product quality in the case of personal computers.

<sup>29</sup>Following Proposition 2, we need a known point  $(m^*, s^*, f^*)$ , such that  $m(\alpha s^*, \alpha f^*) = \alpha m^*$ . We choose  $1 = m(s^*, 1)$ , so that one exporter is always matched when there are  $s^*$  ships. We set  $s^*$  such that in all locations  $m_i \leq f_i$  and we iterate over  $s^*$ . Note that this approach delivers a conservative bound on search frictions, since in principle we could allow for higher levels of exporters.

<sup>30</sup>We interpret the observed time-series variation as driven by short-run deviations from the steady state values.

of interest,  $\theta$ . We estimate  $\theta$  via Maximum Likelihood. We use a nested fixed point algorithm to solve for the ship value functions at every guess of the parameter values (Rust (1987)), compute the predicted choice probabilities and then calculate the likelihood. We calibrate the discount factor to  $\beta = 0.995$ .

Since our model features a number of inter-related value functions ( $W, U, V$ ), 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  $\theta$ , the map  $T_\theta : \mathbb{R}^I \rightarrow \mathbb{R}^I$ ,  $U \rightarrow T_\theta(U)$  with,*

$$T_\theta(U) = -c_i^u + \lambda_i \sum_{j \neq i} G_{ij} \tau_{ij} + \lambda_i \sum_{j \neq i} G_{ij} \left[ -\frac{c_{ij}^s}{1 - \beta(1 - \xi_{ij})} + \beta \xi_{ij} \frac{U_j}{1 - \beta(1 - \xi_{ij})} \right] + (1 - \lambda_i) J_i(\theta, U)$$

where  $\tau_{ij} = E_v \tau_{ijv}$  is the mean price from  $i$  to  $j$ , is a contraction and  $U(\theta)$  is the unique fixed point.

*Proof.* See the Appendix. □

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

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

- (a) Solve for  $W^m$  from:

$$W_{ij}^m = \frac{-c_{ij}^s + \xi_{ij} \beta U_j^m}{1 - \beta(1 - \xi_{ij})}$$

- (b) Update  $J^m$  from:

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

- (c) Update  $U^{m+1}$  from:

$$U_i^{m+1} = -c_i^u + \lambda_i E_{j,v} \tau_{ijv} + \lambda_i \sum_{j \neq i} G_{ij} W_{ij}^m + (1 - \lambda_i) J_i^m$$

where we use the actual average prices from  $i$  to  $j$ , i.e.,  $E_{j,v} \tau_{ijv} = \sum_{j \neq i} G_{ij} \tau_{ij}$ . Note that  $\lambda_i$  is known (it is simply the average ratio  $\frac{1}{T} \sum m_{it}/s_{it}$ ). Similarly,  $G_{ij}$ , the probability that an exporter ships from  $i$  to  $j$  (conditional on exporting), is obtained directly from the observed trade flows (see Section 5.3).

3. Form the likelihood using the choice probabilities:

$$\mathcal{L} = \sum_i \sum_j \sum_l \sum_t y_{ijlt} \log P_{ij}(\theta) = \sum_i \sum_j \log P_{ij}(\theta)^{n_{ij}} \quad (25)$$

where  $y_{ijlm}$  is an indicator equal to 1 if ship  $l$  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  $P_{ij}(\theta)$  are given by (8) and (9).<sup>31</sup>

**Identification** As is always the case in dynamic discrete choice models, 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 the Appendix. The additional restrictions amount to using the observed fuel price to determine  $c_{ij}^s$  for some  $i, j$ ; see Section 6.2. We estimate the port costs  $c_1^u, \dots, c_I^u$ , which may be capturing heterogeneous costs difficult to measure (actual port costs, ability to wait outside of port, etc.), as well as  $\sigma$ . We present our results in Section 6.

### 5.3 Exporter Valuations and Costs

Using the observed shipping prices, we back out exporter valuations in a straightforward manner. Indeed, consider the equilibrium price (17) solved with respect to the exporter's valuation:

$$v = \frac{1 - \beta\delta \left(1 - \gamma\lambda_i^f\right)}{(1 - \gamma)(1 - \beta\delta)} \tau_{ijv} - \frac{\gamma \left(1 - \beta\delta \left(1 - \lambda_i^f\right)\right)}{(1 - \gamma)(1 - \beta\delta)} (J_i - W_{ij}) \quad (26)$$

Note that once  $\theta = \{c_{ij}^s, c_i^u, \sigma\}$  is known, so is  $J_i$  and  $W_{ij}$ . Moreover,  $\tau_{ijv}$  is observed, while  $\lambda_i^f$  is recovered from the matching function ( $\lambda_i^f$  is the average ratio,  $\frac{1}{T} \sum_t m_{it}/f_{it}$ , where  $f_{it}$  is estimated). We calibrate the freight survival probability to  $\delta = 0.99$ . Now, equation (26) has two unknowns: the valuation  $v$  and the bargaining coefficient  $\gamma$ . First, we pin down  $\gamma$  from external information on the average value of international trade in commodities and obtain  $\gamma = 0.3$ .<sup>32</sup> Now, given this estimate for  $\gamma$ , we re-

<sup>31</sup>We assume that our data comes from one steady state, so that  $(s_t, f_t, s_t^w)$  is fixed at  $(s^*, f^*, s^{w*})$ ; hence  $P_{ij}(\theta)$  does not depend on  $t$  (it also does not depend on the ship  $l$  since ships are homogeneous). During our sample period of 2012-2016 the industry did not experience any major shocks. We also estimated the model separately by season to allow for seasonal time variation and find our results to be similar across the four seasons.

<sup>32</sup>To obtain the average valuation worldwide, we first collect the average price of the five most common commodities (iron ore, coal, grain, steel and urea) from Index Mundi, and multiply it by the average tonnage carried by a bulk carrier (this is equal to the average bulker size times its utilization rate; see Footnote 34). We then set  $\bar{\mu}$  as their weighted average based on each commodity's frequency in shipping contracts; we find  $\bar{\mu}$  to equal 7 million US dollars. Finally, solving (26) with respect

cover exporter valuations point-wise from (26) and can obtain their distribution,  $F_{ij}^v$ , nonparametrically.<sup>33</sup> Note that valuations are drawn from an origin-destination specific distribution, which allows for arbitrary correlation between a cargo's valuation and destination.

The exporter costs  $\kappa_{ij}$  capture both the cost of production, as well as any export costs beyond shipping prices and are estimated from the exporters' chosen destinations. Indeed, given the choice probabilities  $\tilde{G}_{ij}$  defined in (12) we can recover  $\kappa_{ij}$  as follows (Berry (1994)):

$$\kappa_{ij} = J_{ij}^f - \left( \ln \tilde{G}_{ij} - \ln \tilde{G}_{i0} \right) \quad (27)$$

where  $J_{ij}^f$  is now known; indeed, recall from (20) that  $J_{ij}^f = \alpha_i(\mu_{ij} - \tau_{ij})$  and we can now calculate  $\alpha_i = \lambda_i^f / (1 - \beta\delta(1 - \lambda_i^f))$ , as well as the mean valuations  $\mu_{ij}$ . Finally, the satellite data provides direct information on the frequencies  $G_{ij}$ , the proportion of loaded trips from  $i$  to  $j$  (see Figure 14 of the Appendix). We do, however, need to determine the share of the outside option or equivalently, the number of entrants  $d_i$  and potential entrants  $\mathcal{E}_i$ , in order to compute  $\tilde{G}_{ij} = G_{ij}(1 - \tilde{G}_{i0})$ . We obtain the number of entrants by solving for  $d_i$  from the freight transition (1) and taking the average. The number of potential entrants  $\mathcal{E}_i$  is set equal to the total production of the relevant commodities for each region  $i$ .<sup>34</sup>

## 6 Results

In this section we present the results from our empirical analysis. Throughout the estimation, we consider 15 geographical regions, depicted in Figure 16 in the Appendix.<sup>35</sup> To determine the regions, we employ a clustering algorithm that minimizes the within-region distance between ports.

---

to  $\gamma$  and averaging over  $i, j, v$  yields:

$$\gamma = \frac{(1 - \beta\delta)(\bar{\mu} - \bar{\tau})}{\beta\delta E_{ijv} \lambda_i^f \tau_{ijv} + (1 - \beta\delta)\bar{\mu} - E_{ij} \left( 1 - \beta\delta(1 - \lambda_i^f) \right) (J_i - W_{ij})}$$

where  $\bar{\tau}$  is the average observed price.

<sup>33</sup>We have assumed that the exporter obtains zero payoff when he does not find a match. It is not possible to separately identify valuations from such inventory costs or scrap values.

<sup>34</sup>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. A ship operates on average 340 days per year (due to maintenance, repairs, etc.) and has a deadweight utilization of about 65%. A region's total production serves as an upper bound to the region's exports.

<sup>35</sup>The trade-off here is that we need a large number of observations per region, while allowing for sufficient geographical detail. 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 inter-regional trips.

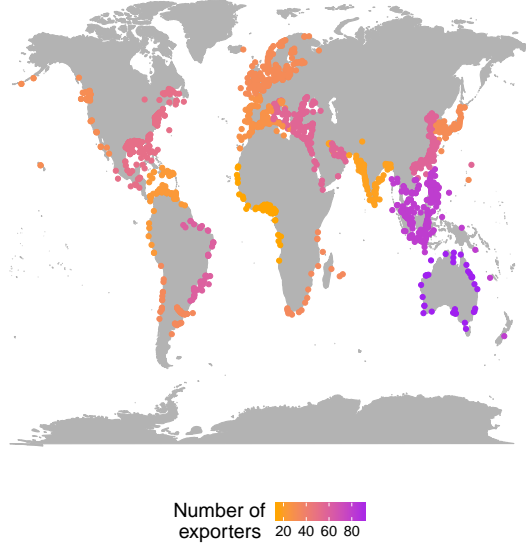


Figure 5: Average weekly number of exporters.

## 6.1 Matching Function

**Search Frictions Test** Before presenting the main results, we provide a simple test for search frictions, inspired by the model and the empirical methodology outlined in Section 5.1. Suppose that for some region  $i$ , it is known that there are more ships than exporters, i.e.  $\min(s_{it}, f_{it}) = f_{it}$ . If there are no search frictions, so that  $m_{it} = \min(s_{it}, f_{it}) = f_{it}$ , exogenously changing the number of ships does not affect the number of matches. In contrast, if there are search frictions, any exogenous change in the number of ships changes the number of matches. We can thus test for search frictions by using the exogenous changes in ocean weather conditions to explore whether changing the number of ships in regions with a lot more ships than exporters affects the realized number of matches. To proxy for weather conditions, we employ the unpredictable component of wind at sea.<sup>36</sup> Since we do not observe freights directly, for each region we consider periods when there are substantially more ships than matches (though recall that overall during our sample period there is substantial ship excess supply). Table 4 in the Appendix presents the results across regions during weeks with at least twice as many ships as matches. We find that indeed matches are affected by weather conditions, consistent with the presence of search frictions.

<sup>36</sup>In particular, we divide the sea surrounding each region into 8 different 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 on their lag component and season fixed effects. We experiment with the lag structure and the results are overall robust. Finally, the results are robust to running the VAR jointly for neighboring zones.

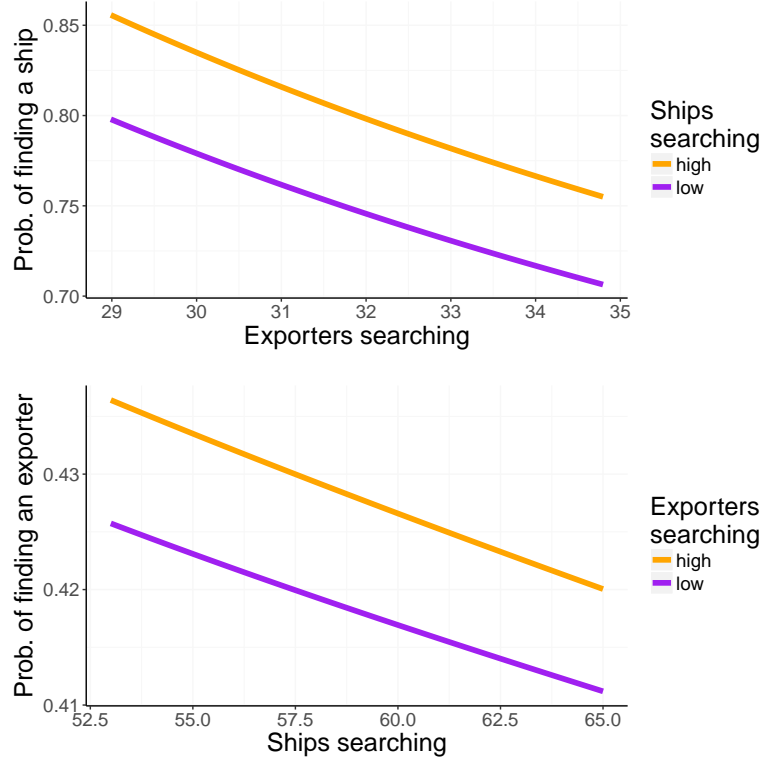


Figure 6: Estimated match finding probabilities for ships and exporters,  $\lambda$  and  $\lambda^f$  in the West Coast of South America. “High” corresponds to the 40<sup>th</sup> percentile, while “low” corresponds to the 60<sup>th</sup> percentile.

**Matching Function Estimates** We now turn to the results from our main methodology of Section 5.1. Table 5 in the Appendix presents the results from the first stage regression of the number of ships on the weather for all regions and reveals that ocean wind has significant impact. Figure 5 presents the weekly average number of exporters in the world. Exporters are concentrated in Australia, the East Coast of South America and Southeast Asia. India, Africa and Central America have the fewest freights.

To visualize the matching function, Figure 6 plots matching rates for ships and exporters, for the West Coast of South America as one example. The top panel plots the matching rate for exporters,  $\lambda^f$ , as a function of the number of exporters searching and for different levels of ships. Note that, as expected,  $\lambda^f$  declines as the market gets crowded with exporters. Similarly, the bottom panel plots the probability that a ship finds a match,  $\lambda$ , as a function of the number of ships and for different levels of exporters. Again, this probability declines in the number of searching ships. It is also worth noting, that exporters have substantially higher chances of finding a match than ships, consistent with our sample period of high ship supply and low demand, as well as our conservative scale restriction on the exporters (see footnote 29). This is true in all regions.

To measure the extent of search frictions in different regions, we compute the average percentage

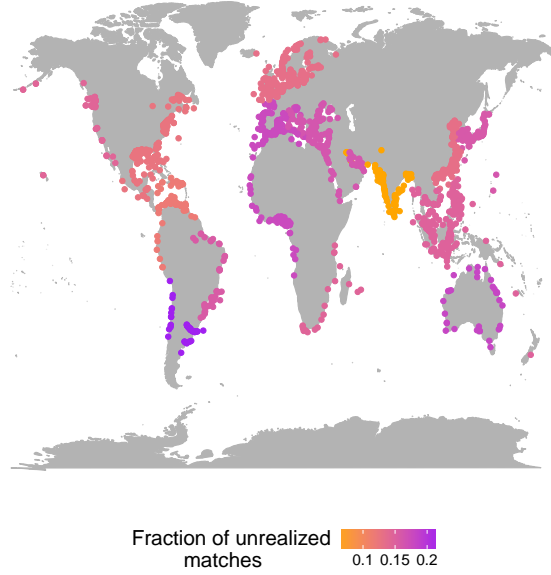


Figure 7: Average weekly share of “unrealized” matches because of search frictions.

of weekly “unrealized” matches; i.e.  $(\min\{s_i, f_i\} - m_i) / \min\{s_i, f_i\}$ . The results are plotted in Figure 7 and reveal that search frictions are heterogeneous over space and may be sizable, with up to 20% of potential matches “unrealized” weekly in regions like West Africa and parts of South America and Europe. On average, 17.2% of potential matches are “unrealized”.<sup>37,38</sup> Figure 17 in the Appendix, depicts these ratios for all regions, as well as the minimum between ships and freights, along with confidence intervals constructed from 500 bootstrap samples. We can reject that the matching function equals the minimum for all regions.

The ratio of “unrealized” matches correlates well with the observed within-region price dispersion, an indicator of search frictions. It also correlates with the ratio of incoming and outgoing ships, discussed in Section 3.3, as for instance there are more “unrealized” matches in Chile than Norway. 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.

<sup>37</sup>Results are overall robust if instead of imposing that the matching function is homogeneous of degree one, we fix the distribution of  $f$ ; see Footnote 28. In our case, a  $[0, 1]$  uniform distribution for freights does not sound plausible since we need to also satisfy  $m_{it} \leq f_{it}$  for all  $i, t$ . Therefore, we instead experimented with more flexible distributions (normal, log-normal) and calibrated their parameters so that this inequality is always just satisfied- this again yields the most conservatively estimated level of search frictions.

<sup>38</sup>It is worth noting that this does not imply that in the absence of search frictions we would have 17.2% more matches; this is simply a measure of the severity of search frictions in different regions.

## 6.2 Ship Costs

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 all port wait costs  $c_i^u$ , for all  $i$ .<sup>39</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. We thus set this cost for one of the groups (for trips originating from the East Coast of North and South America) equal to the average fuel price.<sup>40</sup> Note also that 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 3 report the results.<sup>41</sup> Not surprisingly, sailing costs are fairly homogeneous. Port wait costs are more heterogeneous and large, ranging between 130,000 and 260,000 US dollars per week. Consistent with industry narratives, it is costly to let a ship waiting at port, both due to direct port and security fees, as well as the rapid depreciation of the ship's machinery and electronics. Ports in the Americas are the most expensive, while ports in China, India, Southeast Asia and the Middle East are the cheapest. The standard deviation of the preference shocks,  $\sigma$ , is estimated at about 11,000 US dollars, roughly 10% of price, which implies that the preference shocks do not account for a disproportionately large part of utility or ballast decisions. As shown in Figure 18 of the Appendix, the fit is very good, as our predicted choice probabilities are very close to the observed ones.<sup>42</sup>

## 6.3 Exporter Valuations and Costs

In Figure 8 we plot the average exporter valuations across origins, while the third column of Table 3 reports the estimates. There is substantial heterogeneity in valuations across space. South and North America have the highest valuations, while Europe and Southeast Asia have the lowest. This ranking is reasonable, as for instance, Brazil exports grain which is expensive, while Southeast Asia exports mostly coal, which is one of the cheapest commodities. We generalize this example by focusing on grain, the most

---

<sup>39</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>40</sup>The average weekly price of fuel is 69,100\$. We have experimented heavily with different types of identification restrictions and the results are robust. In particular, we have considered (i)  $c_{ij}^s = c^s$  for all  $i$  and  $j$ ; (ii) coarser and finer groups; (iii)  $c_{ij}^s$  clustered by the distance between  $i$  and  $j$ , as well as clustered by the weather between  $i$  and  $j$  (both capture nonlinear effects of distance on the sailing cost). The ship also incurs operating costs (crew, maintenance, etc.). However, these are fixed costs of operation; as such they do not affect the ships' decisions and can be ignored.

<sup>41</sup>The standard errors are computed from 500 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>42</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 shipowners do not realize that under a mean-reverting demand for seaborne trade prices will go up eventually (see Kalouptsi (2014)). If we compute the equilibrium under higher exporter valuations that lead to prices closer to the ones observed before 2010, the ship value function indeed becomes positive.

expensive frequently shipped commodity. In particular, using data from Comtrade, we explore whether countries that have a high share of grain exports tend to have higher estimated valuations. The results, shown in Figure 9, reveal that indeed there is a positive correlation between the two, suggesting that exporters with higher valuations may be producers of more expensive products.<sup>43</sup> Of course, there may be other factors determining the valuation of an exporter such as inventory control, just in time production, etc. On average, the average price  $\tau_{ij}$  is equal to about 5% of the mean valuation  $\mu_{ij}$ , consistent with other estimates in the literature (e.g. UNCTAD (2015), Hummels et al. (2009)).

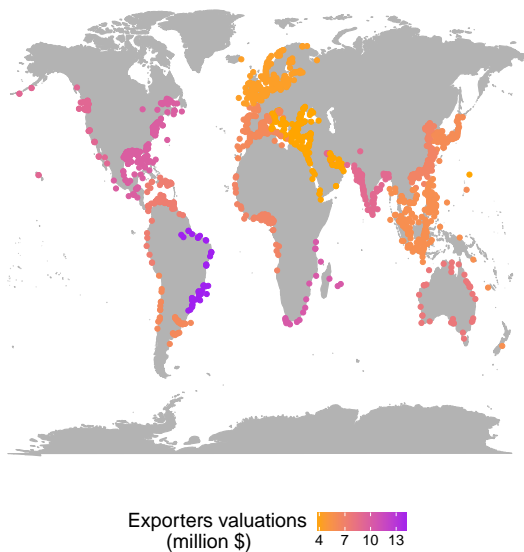


Figure 8: Average exporter valuations.

---

<sup>43</sup>As discussed in Section 3.3, the dependence of prices on the type of good is suggestive of search frictions in the market. In our estimation we are able to back out the valuations given our estimates for search frictions. The external validation discussed in this paragraph also supports our conclusions that search frictions are present in these markets.

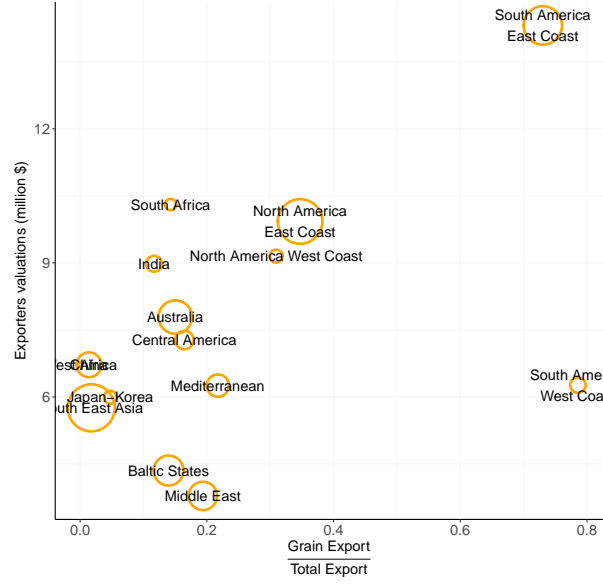


Figure 9: Average exporter valuations and share of exports in grain (source, Comtrade). The size of the circle proxies the number of observations.

Finally, we turn to exporter costs. The share of the exporters' outside option, computed from the total commodity production in the region, is on average 62%. China and India (South America, Australia and Southeast Asia) feature the highest (lowest) outside share, consistent with their high (low) imports. The estimated exporter costs 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  $\mu_{ij}$ . 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 production costs, as well as other exporting costs.

Our estimation results are robust when we estimate the costs separately by season. Moreover, we have performed our estimation and counterfactuals for Handymax and Handysize vessels alone (the segments for which we have sufficient data at the baseline region-week level) and found that the resulting counterfactuals are very similar.

	Port Costs $c_u$	Sailing Costs $c_s$	Exporters Valuations $\mu_v$	Preference Shock $\sigma$
North America West Coast	2.458 (0.07)	0.693 (0.002)	79.605 (2.038)	
North America East Coast	2.271 (0.021)	- -	103.145 (2.229)	
Central America	1.846 (0.022)	0.693 (0.002)	73.161 (3.007)	
South America West Coast	1.996 (0.026)	0.693 (0.002)	59.063 (1.679)	
South America East Coast	2.563 (0.027)	- -	125.877 (3.001)	
West Africa	1.421 (0.015)	0.64 (0.002)	57.838 (2.658)	
Mediterranean	1.637 (0.018)	0.568 (0.003)	59.87 (2.475)	
North Europe	1.399 (0.009)	0.568 (0.003)	49.199 (1.959)	
South Africa	2.478 (0.035)	0.64 (0.002)	99.074 (2.907)	
Middle East	1.273 (0.007)	0.568 (0.003)	44.203 (2.355)	
India	1.48 (0.014)	0.624 (0.003)	84.722 (4.2)	
South East Asia	1.67 (0.008)	0.56 (0.002)	72.282 (3.324)	
China	1.438 (0.01)	0.558 (0.002)	66.382 (3.61)	
Australia	2.635 (0.025)	0.56 (0.002)	70.507 (2.543)	
Japan-Korea	1.53 (0.022)	0.558 (0.002)	55.589 (2.514)	
				0.117 (0.0008)

Note: all the estimates are in 100,000 USD. Standard errors computed from 500 bootstrap samples.

Table 3: Ship costs and exporter valuation estimates. The sailing cost for the East Coast of North and South America is set equal to 0.69 (the fuel cost).

## 7 Counterfactuals

In this section, we use our estimated model to explore a number of questions of interest through the lens of endogenous trade costs. We consider: the change in exports following an improvement in shipping efficiency, a Chinese slow-down, the opening of the Northwest Passage and the trade reduction due to search frictions.

These counterfactuals illustrate three novel mechanisms of our setup. First, as model primitives change and directly affect the value of a match, the ship’s outside option,  $J_i$ , is also affected and provides a new channel that impacts trade costs and exporting. Second, the economic and geographic network of countries matters. For instance, when a shock makes a trading partner  $j$  more attractive for ships, say because it offers more options to ships that end their trip there, then  $W_{ij}$  increases and from equation (17), prices to  $j$  go down and exports increase. Similarly, exporting countries that are close to large net importers, benefit from the “glut” of ships that end their trip there and wish to ballast elsewhere. Third, how each country is affected in every one of the counterfactuals below depends crucially on (i) the size of its trade imbalance and (ii) its mean value,  $v$ , especially when it is a net exporter. For instance, as we will see below, under improvements in shipping technology or matching efficiency, ship reallocation leads to “polarization”, whereby more ships are now more likely to reallocate to large, high value exporting countries

To perform the counterfactuals, we compute the steady state spatial equilibrium distribution of ships and exporters. In the Appendix, we provide the computational algorithm employed.<sup>44</sup>

### 7.1 Change in Trade due to a Change in Shipping Efficiency

We consider how a 10% decrease in the cost of shipping,  $c^s$ , affects shipping prices,  $\tau$ , and trade flows. This change in transportation costs has two effects:

First, there is a direct increase in the surplus of all matches, since now a match between a ship and a freight is more valuable.<sup>45</sup> All else equal, this reduces export prices,  $\tau$ , which in turn increases the value of an unmatched exporter,  $J^f$ , and thus induces more entry into the export market.

Second, reducing  $c^s$  implies that ballasting is cheaper and ships can reallocate across space more freely. Therefore, their “outside option”,  $J$ , is now higher. Since Nash bargaining requires that both parties

<sup>44</sup>Our algorithm always converged to the same solution, even from very different starting values.

<sup>45</sup>Formally, using the ship and freight value functions, the match surplus is given by

$$S_{ijv} = v - \frac{c_{ij}^s}{1 - \beta(1 - \xi_{i,j})} + \frac{\xi_{ij}\beta}{1 - \beta(1 - \xi_{ij})}U_j - J_i - J_{ijv}^f.$$

A decline in  $c_{ij}^s$ , holding everything else constant, directly increases  $S_{ijv}$ .

receive their outside option plus a share of the surplus, an increase in the ship’s outside option leads, all else equal, to an increase in prices,  $\tau$  (see the price equation (17)). Put differently, reduced transportation costs imply that ships are less “tied” to their current region and as freights’ monopsony power is reduced, ships receive higher prices. This effect tends to mitigate the increase in freight entry driven by the direct effect on the surplus.

Figure 10 presents the results and showcases that there is considerable heterogeneity in different regions’ reaction to a change in transportation costs. The impact of the second effect on prices is easily seen, as prices decline by at most 3%, instead of 10% (the decline in  $c^s$ ). North America, the East Coast of South America, Australia and Southeast Asia see the highest increase in exporting (5-7%). India, China, Japan and Southern Europe are among the regions where exporting increases the least (1-3%). We find that the second effect, capturing the change in ships’ outside option, correlates both with a country’s net exporting status, as well as the average valuation of its exporters. Indeed, net exporters experience a higher increase in their exports, than net importers all else equal. Ships ending a trip at net importing countries now face lower ballast costs and are thus less likely to wait there for a cargo; their outside option is higher and they can command higher prices. For instance, we find that in India, China and Japan the second effect mutes the direct effect substantially: if ships could not reallocate, the increase in exports in these countries would have been three times higher. In contrast, high value net exporters (e.g. Northeast America, as well as Brazil) benefit from the increased willingness of ships to ballast, and they do see an increase in the number of ships ballasting there: as transport costs, and thus distance, now matter less, mean freight valuations,  $\mu$ , become a relatively more important determinant of ships’ decisions.

## 7.2 Chinese Slow-down

We next explore how shocks propagate in a world where trade costs are endogenous by considering a Chinese slow-down (i.e. a reduction in the mean valuation of freights going to China,  $\mu_{i,china}$ , by 10%). Figure 11 presents the resulting trade costs and trade flows across the world.

We begin discussing the results by looking at China itself: shipping prices from China increase by 2%, while exporting declines by 6%. This illustrates the 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 led to reduced trade costs for Chinese exporters. Therefore, when imports decline, fewer ships end up in China and Chinese exporters are hurt.

Next, note that China’s trading partners, such as Australia, Southeast Asia and Brazil, naturally

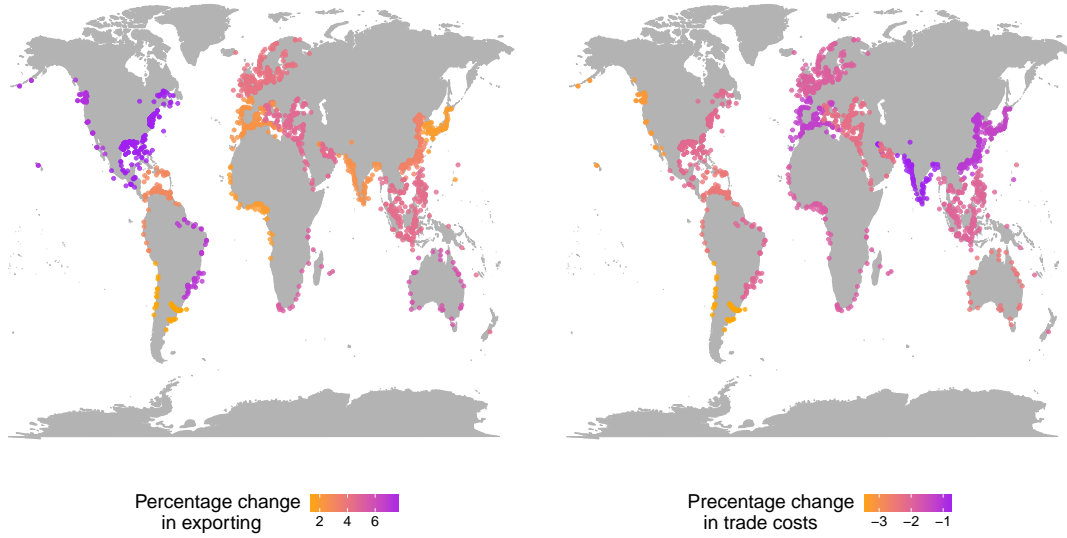


Figure 10: Change in exports and trade costs under a 10% decline in the sailing cost  $c^s$ .

experience a substantial decline in their exports (16-40%). In addition, much like in the previous counterfactual, the decline in exporting is dampened by the reduction in ships' outside option as ships are overall worse off. Prices also decline, both because China is a relatively expensive destination ( $\mu_{i,China}$  is high), but also because ships' outside option is lower.

Finally, it is instructive to trace out the geographically heterogeneous response of exports to this shock, in order to investigate the role of the network of countries. The dampening accruing from ships' lower outside options is much lower for the big exporters close to China, such as Australia and Southeast Asia, and much larger for further away exporters, such as Brazil. This underscores the importance of being close to a large net importer like China: exporting countries in that "pocket" of the world, gain not just by directly exporting to China, but also indirectly from the increased supply of ships in that region. Indeed, the Southeast Pacific region consists of some close-by large exporters (Australia, Southeast Asia) and importers (China, Japan, India) and thus benefits from a "cheap" supply of ships that remains in the area ballasting and trading between these countries. When Chinese demand falls these ships reallocate to other parts of the world and tend to push up exports there, dampening the overall decline. To see this, we compute that if ships were not able to reallocate and ships' outside options had not changed, Brazilian exports would have fallen by 21% rather than 15%, while in Northeast America exports would have declined by 15% rather than 9%. In contrast, Chinese exports would have barely been affected, while the decline in Southeast Asian, Japanese and Australian exports would have been roughly the same.

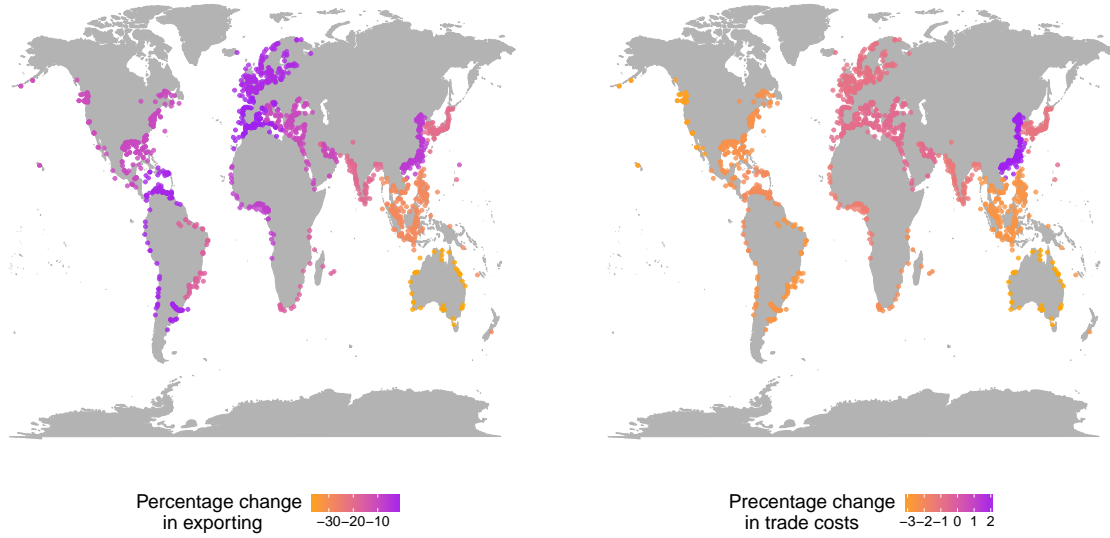


Figure 11: Change in exports and trade costs under a Chinese slow-down.

### 7.3 Opening the Northwest Passage

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 Arctic sea ice; with global warming and ice thinning, there is public discussion about opening the passage to be exploited for shipping. The Northwest Passage would reduce the travel costs between Northeast America and the Far East.

To simulate the opening of the Northwest Passage, we reduce transport costs between the East Coast of North America and China/Japan/S.Korea by 30%.<sup>46</sup> Figure 12 presents the resulting change in exports by region. Naturally, Northeast America sees its exporting rise, by 8.6%. Interestingly, exports from China and Japan/S.Korea fall, as ballasting is now less costly for ships: when in China or Japan they can now ballast to Northeast America more cheaply. Indeed, we find that the probability of an unmatched ship staying in China (Japan/S. Korea),  $P_{ii}$ , is now 25% (28%) lower. The ships' higher outside option, tends to increase prices and decrease exporting.

Finally, Figure 12 reveals that other countries, not directly affected by the opening of the Northwest Passage experience changes in their trade. Overall, exports decline by up to about 3%. This decline is due to the ships' higher outside option. The response of different countries is quite heterogeneous and illustrates the neighbor and network effects. For instance, we see that Brazil, as well as Northwest America are hurt the most. Indeed, as these countries are close to Northeast America ships that used to ballast

<sup>46</sup>We calculate the change in sailing costs of traveling via the Northwest Passage from Ostreng et al. (2013).

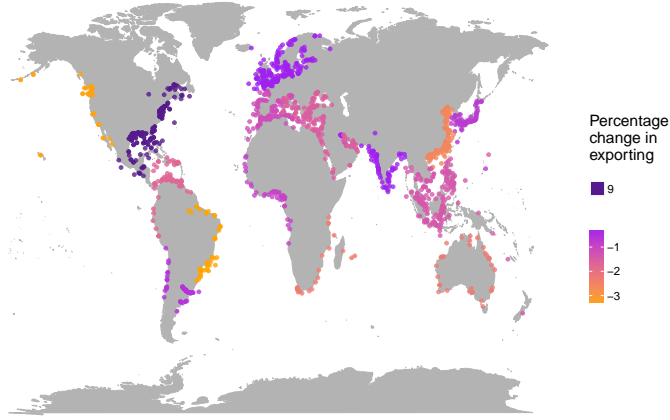


Figure 12: Change in exports if the Northwest Passage becomes a viable commercial route.

there to load now prefer to ballast to Northeast America instead. We find that the number of ships that ballast to Brazil (Northwest America) is lower by about 2% (3%), while ballasting to Northeast America is 13% higher. In contrast, Australian and Southeast Asian exports do not decline as much, as these countries are shielded by their closeness to China and India. These effects are present because of the endogenous trade cost and demonstrate the spillover of such policy changes through the network of countries.

#### 7.4 The Trade Lost because of Search Frictions

Finally, we quantify the trade lost because of search frictions. To do so, we shut down search frictions by setting  $m(s, f) = \min\{s, f\}$ .<sup>47</sup> Figure 13 presents the resulting change in exports by region, as well as the change in ballasting. We find that exporting would be 6-45% higher across different regions in the world with an average increase of 23%.<sup>48</sup> It is of course not surprising that we see higher exports globally in the absence of search frictions. Interestingly, the change in trade is disperse geographically. While countries that experience more severe frictions, as captured in Figure 7, roughly experience somewhat larger increases in exports, this is not always the case. Indeed, a country's net exporting status is a more important determinant of the extent to which the country can attract economic activity. Large net exporters like Brazil and the Northeast America experience disproportionally large increases in exports, as differences in frictions across regions are no longer relevant and exporting size becomes a more important determinant of trade. The correlation between the change in exports and a country's trade imbalance is

<sup>47</sup>In practice we set  $m(s, f) = \alpha \min\{s, f\}$  with  $\alpha = 0.99$ , in order to maintain the Nash bargaining setup and the model comparable.

<sup>48</sup>It is worth noting that the impact of search frictions is quantitatively important also because a ship completes several trips during the course of a year and therefore finds itself "unemployed" extremely often; in contrast, in labor markets, the average worker experiences unemployment once every few years.

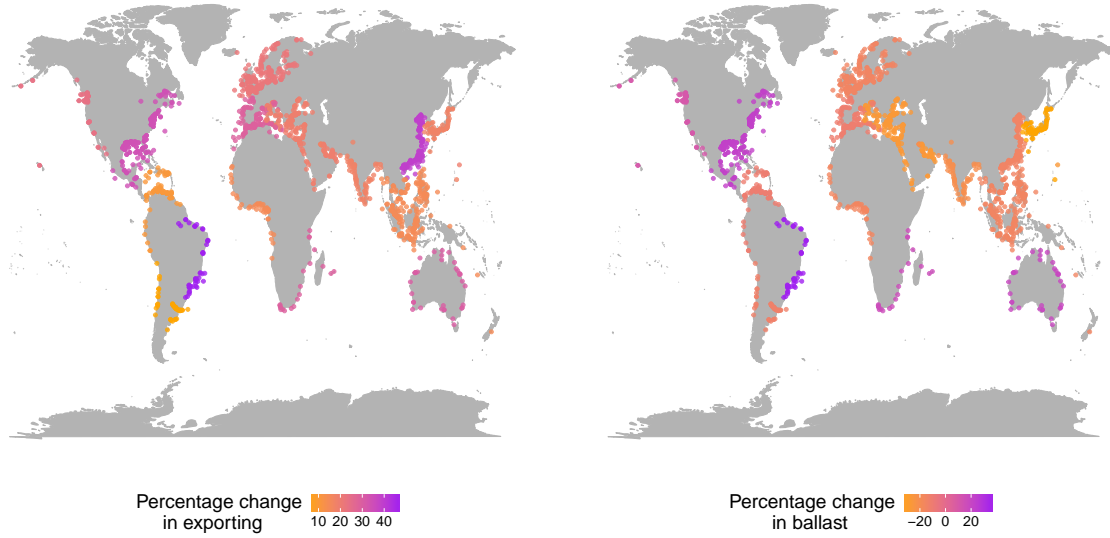


Figure 13: Change in exporting in a world without search frictions.

0.9. Since search frictions have been removed, ships ballast more towards large net exporters. Indeed, in a world with no search frictions, as shown in in Figure 13, large net exporters such as Brazil, Australia and North America experience large increases in the number of ships ballasting there. As previously discussed, when impediments to trade are reduced, country differences in their export status and size become relatively more important determinants of ships’ ballast decisions and the resulting trade flows.

## 8 Conclusion

In this paper, we build a dynamic spatial search model for world ships and exporters. Using unique data on shipping contracts and ship movements we recover the main primitives of interest: the matching function between ships and freights, the distribution of searching exporters, ship costs and exporter valuations. Our methodology allows us to obtain the matching process flexibly, without relying on assumptions regarding the extent of search frictions or the parametric form of the matching function. We demonstrate that accounting for the endogeneity of trade costs is important in both descriptive analysis (e.g. elasticities, shock propagation), as well as policy analysis (e.g. transportation infrastructure planning). Finally, we find that search frictions substantially reduce world trade.

## References

AGUIRREGABIRIA, V. AND P. MIRA (2007): “Sequential Estimation of Dynamic Discrete Games,” *Econometrica*, 75, 1–53.

- 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.
- ASTURIAS, J. (2016): “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.
- BUCHHOLZ, N. (2016): “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.
- COLLARD-WEXLER, A. (2013): “Demand Fluctuations in the Ready-Mix Concrete Industry,” *Econometrica*, 81, 1003–1037.
- 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 (2016): “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.
- 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.
- 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.
- (2017): “Detection and Impact of Industrial Subsidies, the Case of Chinese Shipbuilding,” *forthcoming*, *Review of Economic Studies*.
- KALOUPTSIDI, M., P. T. SCOTT, AND E. SOUZA-RODRIGUES (2016): “Identification of Counterfactuals in Dynamic Discrete Choice Models,” *NBER working paper*, 21527.

- 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.
- LANGE, F. AND T. PAPAGEORGIOU (2017): “Estimating Matching Functions when Search Intensity is Unobserved,” *mimeo, McGill University*.
- 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.
- 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.
- PISSARIDES, C. A. (2000): *Equilibrium Unemployment Theory*, Cambridge, MA: MIT Press, 2nd ed.
- POSTEL-VINAY, F. AND J.-M. ROBIN (2002): “Equilibrium Wage Dispersion with Worker and Employer Heterogeneity,” *Econometrica*, 70, 2295–2350.
- REYES-HEROLES, R. (2016): “The Role of Trade Costs in the Surge of Trade Imbalances,” *mimeo, Federal Reserve Board*.
- 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.
- 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.
- WONG, W. F. (2017): “The Round Trip Effect: Endogenous Transport Costs and International Trade,” *mimeo, University of Wisconsin-Madison*.
- WTO (2015): “International Trade Statistics,” Tech. rep., World Trade Organization, Geneva, Switzerland.

## A Construction of Ship Travel Histories

Here, we describe the construction of ships’ travel histories. The first task is to identify stops that ships make, using the EE data which provides the exact location of ships every six minutes. 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 our information for 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-SKorea-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 provided by <https://sea-distances.org/>, and find that we match trip durations well.

## B Construction of Searching Ships

Here we describe the construction of the vector 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  $t$ .

## C Additional Figures and Tables

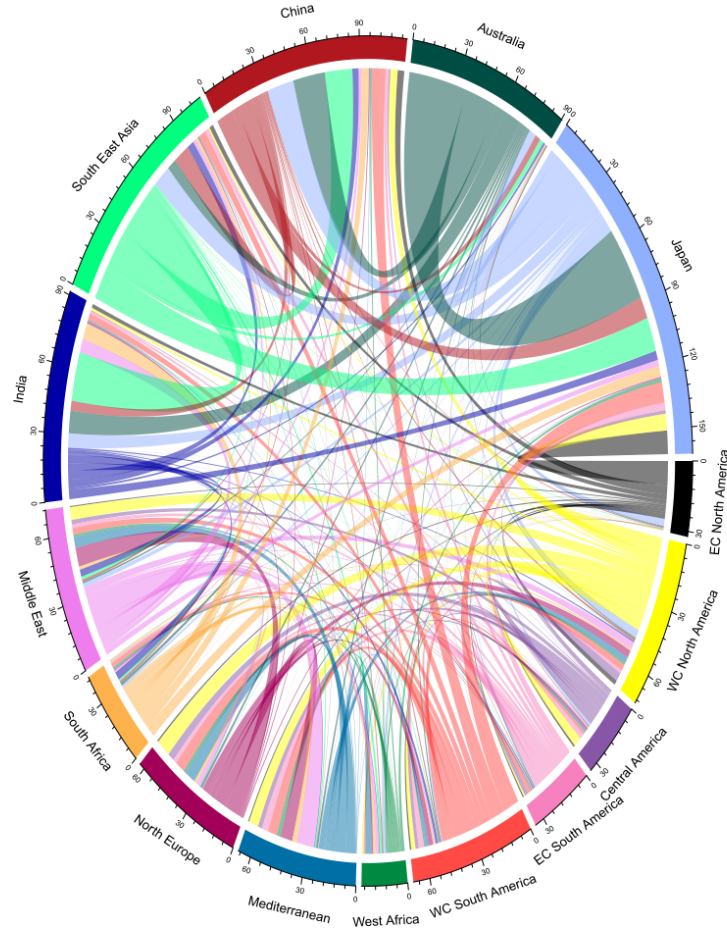


Figure 14: Flows of loaded trips. The colored bar along the perimeter of the circle is proportional to the total number of incoming and outgoing loaded ships for each region. The number of outgoing loaded ships from a region are represented as rays of the same color as the color bar, and are directed towards the destination region. The width of each outgoing ray is proportional to the number of loaded ships headed from the region of origin to the region of destination.

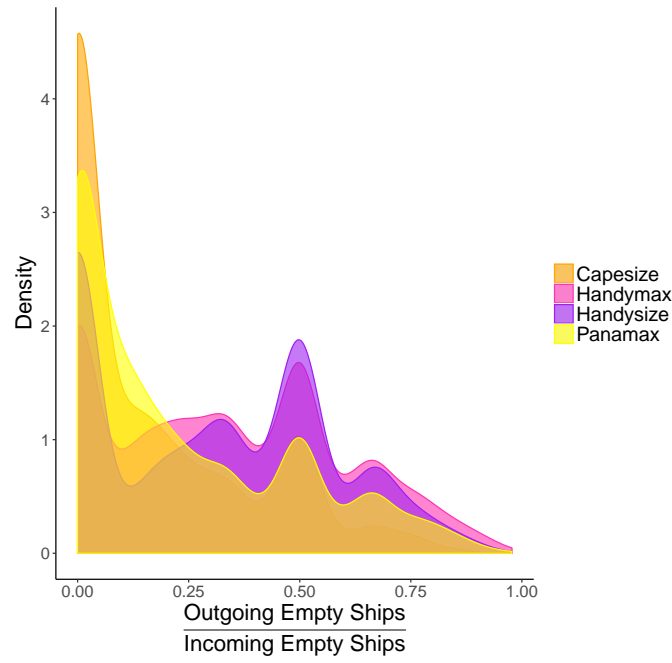


Figure 15: Histogram of the ratio of outgoing empty, over incoming empty and loading ships in net exporting countries, by ship type.

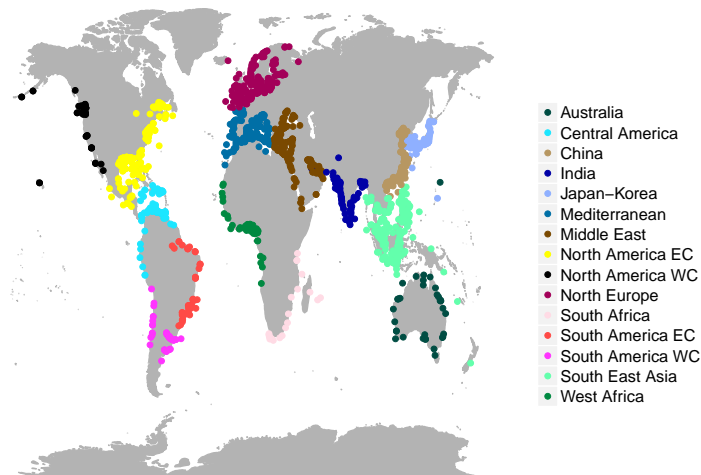


Figure 16: Definition of regions. Each color depicts one of the 15 geographical regions.

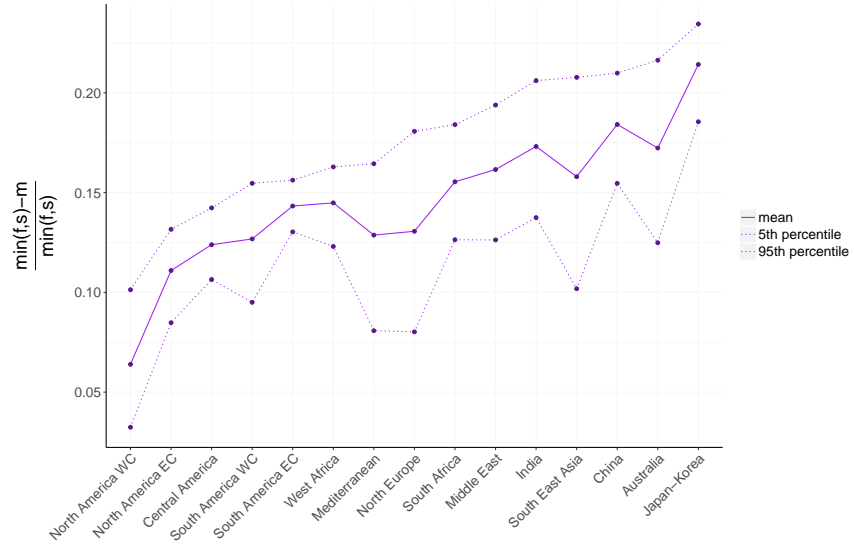


Figure 17: Average weekly share of unrealized matches due to search frictions, with confidence intervals from 500 bootstrap samples.

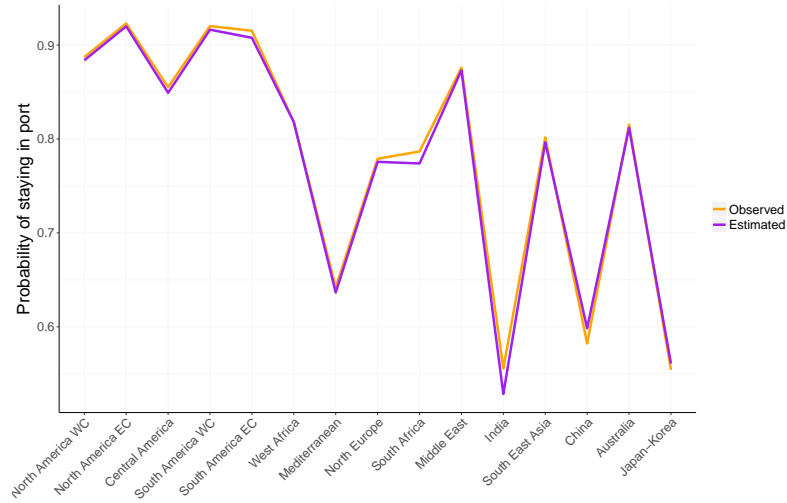


Figure 18: Observed and estimated probability of waiting in port  $P_{ii}$ .

	N	$R^2$	Joint Significance	$\frac{s}{m}$
North America West Coast	193	0.146	0	2.706
North America East Coast	200	0.17	0.013	3.172
Central America	199	0.272	0	3.451
South America West Coast	198	0.246	0	2.913
South America East Coast	200	0.269	0	4.083
West Africa	200	0.261	0	5.862
Mediterranean	200	0.358	0	4.244
North Europe	200	0.23	0	3.577
South Africa	200	0.083	0.01	2.862
Middle East	200	0.147	0.001	3.86
India	200	0.12	0.018	8.58
South East Asia	200	0.18	0.005	3.334
China	200	0.177	0	6.194
Australia	187	0.17	0.008	2.457
Japan-Korea	200	0.16	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 the  $R^2$ ; the third column joint significance; and the last column the average ratio between matches and ships in each region. To proxy for the unpredictable component of weather, we divide the sea surrounding each region into 8 different zones (North East, South East, South West and North West both close to the coast and in open sea), 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. Finally, 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.

	N	$R^2$	Joint Significance
North America West Coast	200	0.101	0.004
North America East Coast	200	0.106	0.0002
Central America	200	0.175	0.0007
South America West Coast	200	0.418	0
South America East Coast	200	0.178	0
West Africa	200	0.138	0.0001
Mediterranean	200	0.181	0
North Europe	200	0.138	0.0003
South Africa	200	0.066	0.064
Middle East	200	0.162	0.0012
India	200	0.157	0.0001
South East Asia	200	0.081	0.0008
China	200	0.176	0
Australia	200	0.049	0.02
Japan-Korea	200	0.036	0.12

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 number of observations; the second column the  $R^2$ ; the third column joint significance. To proxy for the unpredictable component of weather, we divide the sea surrounding each region into 8 different zones (North East, South East, South West and North West both close to the coast and in open sea), 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. Finally, 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.

## D Proof of Proposition 1

We first derive (18) and (19). Suppose  $f_{it}, s_{it}, s_{ijt}^w$  approach  $f_i, s_i, s_{ij}^w$  as  $t \rightarrow \infty$ . Then (2) becomes:

$$s_i = (s_i - m_i(s_i, f_i)) P_{ii} + \sum_{j \neq i} \xi_{ji} s_{ji}^w \quad (28)$$

while for a ship traveling from  $j$  to  $i$ , (3) becomes:

$$s_{ji}^w = (1 - \xi_{ji}) s_{ji}^w + P_{ji} s_j + (G_{ji} - P_{ji}) m_j(s_j, f_j) \quad (29)$$

or

$$\xi_{ji} s_{ji}^w = P_{ji} s_j + (G_{ji} - P_{ji}) m_j = P_{ji} (s_j - m_j) + G_{ji} m_j$$

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

$$\sum_{j \neq i} \xi_{ji} s_{ji}^w = \sum_{j \neq i} P_{ji} (s_j - m_j) + \sum_{j \neq i} G_{ji} m_j$$

and replacing in (28) we get (18).

Equation (19) is a direct consequence of (1) and (14).

The steady state equations (18) and (19) 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  $h : \mathbb{R}^{2I} \rightarrow \mathbb{R}^{2I}$ ,  $h = (h^s, h^f)$  with:

$$h_i^s(s, f) = \sum_{j=1}^I P_{ji}(s, f) (s_j - m_j(s_j, f_j)) + \sum_{j \neq i} G_{ji} m_j(s, f)$$

$$h_i^f(s, f) = \delta (f_i - m_i(s_i, f_i)) + \mathcal{E}_i \sum_{j \neq 0, i} \tilde{G}_{ij}(s, f)$$

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  $\lambda, \lambda^f$  are zero at the origin and continuous. It remains to show that  $F(X) \subseteq X$ . Let  $(s, f) \in X$ . Then,

$f_i \leq \mathcal{E}_i/(1 - \delta)$  and  $\sum_{i=1}^I s_i \leq S$ . Now,

$$h_i^s(s, f) = \sum_{j=1}^I P_{ji}(s, f) (s_j - \lambda_j(s_j, f_j) s_j) + \sum_{j \neq i} G_{ji} \lambda_j(s, f) s_j$$

or

$$h_i^s(s, f) = \sum_{j=1}^I s_j [P_{ji}(s, f) (1 - \lambda_j(s_j, f_j)) + G_{ji} \lambda_j(s, f)]$$

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

$$\sum_{i=1}^I h_i^s(s, f) = \sum_{j=1}^I s_j \left[ \sum_{i=1}^I P_{ji}(s, f) (1 - \lambda_j(s_j, f_j)) + \sum_{i=1}^I G_{ji} \lambda_j(s, f) \right]$$

or

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

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

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

$$h_i^f \leq \delta f_i + \mathcal{E}_i \sum_{j \neq 0, i} \tilde{G}_{ij}(s, f) \leq \delta f_i + \mathcal{E}_i \leq \delta \frac{\mathcal{E}_i}{1 - \delta} + E_i = \frac{\mathcal{E}_i}{1 - \delta}$$

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

## E Proof of Proposition 2

(i) Following Matzkin (2003), two matching functions  $m(\cdot)$  and  $\tilde{m}(\cdot)$  are observationally equivalent if there exists a strictly increasing and differentiable function  $g(\cdot)$  such that:

$$\tilde{m}(s, f) = m(s, g(f))$$

Let  $\lambda > 0$  and fix  $\bar{s}, \bar{f}$ . Then

$$\tilde{m}(\lambda \bar{s}, \lambda \bar{f}) = \lambda \tilde{m}(\bar{s}, \bar{f}) = \lambda \bar{m}$$

Furthermore,

$$\tilde{m}(\lambda \bar{s}, \lambda \bar{f}) = m(\lambda \bar{s}, g(\lambda \bar{f})) = \lambda m(\bar{s}, \frac{1}{\lambda} g(\lambda \bar{f}))$$

Therefore,

$$\bar{m} = \tilde{m}(\bar{s}, \bar{f}) = m(\bar{s}, \frac{1}{\lambda}g(\lambda\bar{f}))$$

Invertibility implies that  $\bar{f} = \tilde{m}^{-1}(\bar{s}, \bar{m})$  and  $\frac{1}{\lambda}g(\lambda\bar{f}) = m^{-1}(\bar{s}, \bar{m})$ , or

$$g(\lambda\bar{f}) = \lambda m^{-1}(\bar{s}, \bar{m})$$

Differentiate with respect to  $\lambda$  to obtain  $\bar{f}g'(\lambda\bar{f}) = m^{-1}(\bar{s}, \bar{m})$ , which for  $\lambda = 1$  becomes  $g'(\bar{f})\bar{f} = m^{-1}(\bar{s}, \bar{m}) = g(\bar{f})$ . Therefore, the Euler condition is satisfied and  $g(\cdot)$  is homogeneous of degree 1. Since  $g(\cdot)$  is a function of a real variable, the only possibility is  $g(f) = cf$  with  $c > 0$ , a constant. Finally, we use the a priori knowledge of the point  $m^* = m(s^*, f^*)$  to establish that  $c = 1$ . Indeed, by definition,  $m(s^*, f^*) = \tilde{m}(s^*, f^*) = m^*$ . But also,  $m(s^*, cf^*) = \tilde{m}(s^*, f^*)$ . Therefore,  $cf^* = f^*$  and since  $f^* \neq 0$ ,  $c = 1$ .

(ii) Conditional on  $\eta$ ,  $s$  is a function of  $z$  which in turn is by assumption independent from  $f$ . It follows that  $s$  and  $f$  are conditionally independent given  $\eta$ . At a point  $\phi$  we have that:

$$F_{f|\eta}(\phi) = \Pr(f \leq \phi|\eta) = \Pr(f \leq \phi|\eta, s) = \Pr(m(s, f) \leq m(s, \phi)|\eta, s) = F_{m|s, \eta}(m(s, \phi))$$

Hence,

$$m(s, \phi) = F_{m|s, \eta}^{-1}(F_{f|\eta}(\phi))$$

which we integrate over  $\eta$  to obtain the result. Let  $\phi = \frac{\phi}{f^*}f^*$ . Then,

$$F_{f|\eta}(\phi) = F_{m|s=\frac{\phi}{f^*}s^*, \eta}\left(m\left(\frac{\phi}{f^*}s^*, \frac{\phi}{f^*}f^*\right)\right) = F_{m|s=\frac{\phi}{f^*}s^*, \eta}\left(\frac{\phi}{f^*}m^*\right)$$

## F Proof of Lemma 2

Fix  $\theta$ . Let  $\phi_{ij} = \frac{1}{1-\beta(1-\xi_{ij})}$ . The map  $T_\theta(U)$  is differentiable with respect to  $U$  with Jacobian:

$$\frac{\partial T_\theta(U)}{\partial U} = \beta(DG + (I - D)P) \odot Z \quad (30)$$

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  $\phi_{ij}\xi_{ij}$  and  $\odot$  denotes the pointwise

product. Indeed, the  $(i, j)$  entry of  $\frac{\partial T}{\partial U}$  is

$$\left(\frac{\partial T}{\partial U}\right)_{ij} = 1 \{i = j\} - \beta \lambda_i G_{ij} \xi_{ij} \phi_{ij} - (1 - \lambda_i) \frac{\partial J_i}{\partial U_j}$$

Now,

$$\frac{\partial J_i}{\partial U_j} = \frac{1}{e^{\frac{\beta U_i}{\sigma}} + \sum_k e^{\frac{W_{ik}}{\sigma}}} e^{\frac{W_{ij}}{\sigma}} \frac{\partial W_{ij}}{\partial U_j} = \beta P_{ij} \xi_{ij} \phi_{ij}$$

and thus

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

which in matrix form becomes (30) (as a convention set  $\xi_{ii} = 1$ ). 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  $0 < \xi_{ij} \phi_{ij} \leq 1$ . Thus,

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

and therefore  $\|H\| \leq 1$ . We deduce that  $\|\frac{\partial T_\theta(U)}{\partial U}\| \leq \beta < 1$ . The mean value theorem then implies

$$\|T_\theta(U) - T_\theta(U')\| \leq \beta \|U - U'\|$$

## G Identification of Ship Port and Sailing Costs

**Proposition 3.** *Given the choice probabilities  $P_{ij}(\theta)$ , the parameters  $\{\frac{c_{ij}^s}{\sigma}, \frac{c_i^u}{\sigma}, \frac{1}{\sigma}\}$  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  $\phi_{ij} = \frac{1}{1 - \beta(1 - \xi_{ij})}$ . The Hotz and Miller (1993) inversion states:

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

Substituting from (4)-(5) we obtain:

$$\sigma \log \frac{P_{ij}}{P_{ii}} = -\phi_{ij} c_{ij}^s + \beta \xi_{ij} \phi_{ij} U_j(\theta) - \beta U_i(\theta) \quad (31)$$

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

$$\log P_{ij} = \frac{W_{ij}}{\sigma} - \frac{J_i}{\sigma} + \gamma^{euler}$$

or:

$$\sigma \log P_{ij} = -\phi_{ij} c_{ij}^s + \beta \xi_{ij} \phi_{ij} U_j(\theta) - J_i + \sigma \gamma^{euler} \quad (32)$$

and

$$\sigma \log P_{ii} = \beta U_i(\theta) - J_i + \sigma \gamma^{euler} \quad (33)$$

Now, replace  $W_{ij}$  from (32) into the definition of  $U$ , (5) to get:

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

where  $\tau_i \equiv E_{v,j} \tau_{ijv} = \sum_{j \neq i} G_{ij} \tau_{ij}$ . Substitute  $J_i$  from (33):

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

so that given the CCP's,  $U_i$  is an affine function of  $c^u$  and  $\sigma$ . Next, we replace this into the Hotz and Miller (1993) inversion (31) to obtain:

$$\begin{aligned} c_{ij}^s &= \frac{\beta}{\phi_{ij}(1-\beta)} c_i^u - \frac{\beta}{1-\beta} \xi_{ij} c_j^u + \\ &+ \sigma \left( \frac{\beta}{1-\beta} \left( \xi_{ij} \left[ (1-\lambda_j) \gamma^{euler} + \lambda_j \sum_{l \neq j} G_{jl} \log P_{jl} - \log P_{jj} \right] - \frac{1}{\phi_{ij}} \left[ (1-\lambda_i) \gamma^{euler} + \lambda_i \sum_{l \neq i} G_{il} \log P_{il} - \log P_{ii} \right] \right) \right) \\ &- \frac{\sigma}{\phi_{ij}} \log \frac{P_{ij}}{P_{ii}} + \frac{\beta}{1-\beta} \xi_{ij} \lambda_j \tau_j - \frac{\beta}{(1-\beta)\phi_{ij}} \lambda_i \tau_i \end{aligned}$$

Note that

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

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

We divide by  $\sigma$ :

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

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

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

## H 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 Section 7.

1. Make an initial guess for  $\{s^0, f^0, U^0\}$ .
2. At each iteration  $m$ , inherit  $\{s^m, f^m, U^m\}$

(a) Update the ship's and exporter's optimal policies by repeating the following steps  $K$  times.<sup>49</sup>

- i. Solve for  $W^{m+1}$  from:

$$W_{ij}^{m+1} = \frac{-c_{ij}^s + \xi_{ij} \beta U_j^m}{1 - \beta(1 - \xi_{ij})}$$

- ii. Update  $J^{m+1}$  from:

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

- iii. Compute the equilibrium prices using

$$\tau_{ijv}^m = \frac{\gamma \left( 1 - \beta \delta \left( 1 - \lambda_i^{f,m} \right) \right)}{1 - \beta \delta \left( 1 - \gamma \lambda_i^{f,m} \right)} \left( J_i^{m+1} - W_{ij}^{m+1} \right) + \frac{(1-\gamma)(1-\beta\delta)}{1 - \beta \delta \left( 1 - \gamma \lambda_i^f \right)} \mu_{ij}$$

- iv. Update  $\tilde{G}$ :

$$\tilde{G}_{ij}^{m+1} \equiv \frac{\exp \left( \frac{\lambda_i^f (\mu_{ij} - \tau_{ij}^m)}{1 - \beta \delta (1 - \lambda_i^f)} - \kappa_{ij} \right)}{1 + \sum_{l \neq i} \exp \left( \frac{\lambda_i^f (\mu_{il} - \tau_{il}^m)}{1 - \beta \delta (1 - \lambda_i^f)} - \kappa_{il} \right)}$$

<sup>49</sup>K is chosen to accelerate the convergence in the spirit of standard modified policy iteration methods.

v. Update  $U^m$ :

$$U_i^{m+1} = -c_i^u + \lambda_i E_{v,j} \tau_{ijv} + \lambda_i \sum_{j \neq i} \left( \frac{\tilde{G}_{ij}^{m+1}}{1 - \tilde{G}_{i0}^{m+1}} \right) W_{ij}^{m+1} + (1 - \lambda_i) J_i^{m+1}$$

vi. Obtain the ships ballast choices  $(P_{ij}^{m+1})_{i=1:I, j=1:I}$

3. Update to  $\{\tilde{s}^{m+1}, \tilde{f}^{m+1}\}$  from:

$$\tilde{f}_i^{m+1} = \delta_i (f_i^m - m_i^m) + \mathcal{E}_i (1 - \tilde{G}_{i0}^{m+1})$$

and

$$\tilde{s}_i^{m+1} = \sum_j P_{ji}^{m+1} (s_j^m - m_j^m) + \sum_j \frac{\tilde{G}_{ij}^{m+1}}{1 - \tilde{G}_{i0}^{m+1}} G_{ji}^{m+1} m_j^m$$

4. If  $\|\tilde{s}^{m+1} - s^m\| < \epsilon$ ,  $\|\tilde{f}^{m+1} - f^m\| < \epsilon$  and  $\|U^{m+1} - U^m\| < \epsilon$ , stop, otherwise update freights and ships as follows:

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

where  $\alpha$  is a smoothing parameter.