Trade from Space*

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

December 2018
Preliminary and Incomplete

Abstract

In this paper we set of to explore how satellite data also can be used within the field of international trade. In particular we set out to investigate how and to what extent satellite data can provide an alternative to the standard source for trade data, namely customs data, and how satellite data can enable us to improve our understanding of the drivers of international trade. The purpose of this paper is thus twofold. One, we want to construct a global comprehensive data set for trade flows exploiting detailed ship-level data on inter-country seaborne trade using satellite tracking of vessel movements and to establish that this data set can be employed as an alternative to traditional trade data. Two, we want to show how detailed satellite information on trade flows can be used to inform us about the drivers of international trade. Amongst other we use the AIS data to compute measures describing the position of ports and countries in the global trade network and investigate the impact of the location in the network on trade flows.

*We would like to thank seminar participants at the University of Oslo for valuable comments and discussion.

[†]University of Oslo inga.heiland@econ.uio.no
[‡]University of Oslo & CEPR, andreas.moxnes@econ.uio.no
[§]University of Oslo & CEPR, k.h.ulltveit-moe@econ.uio.no
[¶]University of Oslo, yuan.zi@econ.uio.no

1 Introduction

Over the last decade economists have started to exploit remote sensing data. Donaldson and Storeygard (2016) provide an overview of applications which so far has focused on environmental, development and spatial issues. In this paper we set of to explore how satellite data also can be used within the field of international trade. In particular we set out to investigate how and to what extent satellite data can provide an alternative to the standard source for trade data, namely customs data, and how satellite data can enable us to improve our understanding of the drivers of international trade.

The purpose of this paper is thus twofold. One, we want to construct a global comprehensive data set for trade flows exploiting detailed ship-level data on inter-country seaborne trade using satellite tracking of vessel movements and to establish that this data set can be employed as an alternative to traditional trade data. This has become possible due to the rapid advent over the last years of the global Automated Identification System (AIS). AIS reporting of vessel positions offers a degree of automation in data processing and aggregation that was not previously possible. Vessels send out AIS signals identifying themselves to other vessels or coastal authorities, and the International Maritime Organization requires all international voyaging vessels with above 300 Gross Tonnage and all passenger vessels to be equipped with an AIS transmitter. AIS messages include information regarding vessels identity, physical appearance, voyage-related information such as draught and destination, as well as dynamic data. AIS reports simply offer real-time information on the whereabouts of all ocean-going vessels.

The construction of an AIS based trade set is inspired by the so called light studies, see e.g. Henderson et al. (2012). In these studies night lights are used as a measure of economic growth. Compared to traditional measures of economic night lights are argued to offer a complement to GDP which typically suffers from a set of measurement problems. Moreover, while GDP numbers are not available on any consistent basis at the subnational level, much of the interesting variation in economic growth takes place within, rather than between, countries, and night lights enables the analysis at this sublevel. Similarly to night lights, AIS based trade data allow for the construction of global comprehensive trade data sets that are more detailed and accurate for example with respect to geography and time, and the same time as being more robust to misreporting which is acknowledge to be an issue with custom data.

To construct the satellite based trade data set we develop what we shall refer to as a *box accounting methodology*, where we build on the methodology that is used to track origin of value added in trade flows. This methodology allows us to move from cargo flows where the port of departure and arrival is observed but not origin and final destination, to a comprehensive data set with a full set of world wide bilateral connections.

Our second purpose is to show how detailed satellite information on trade flows can be used to inform us about the drivers of international trade. Amongst other we use the AIS data to compute measures describing the position of ports and countries in the global trade network and investigate the impact of the location in the network on trade flows.

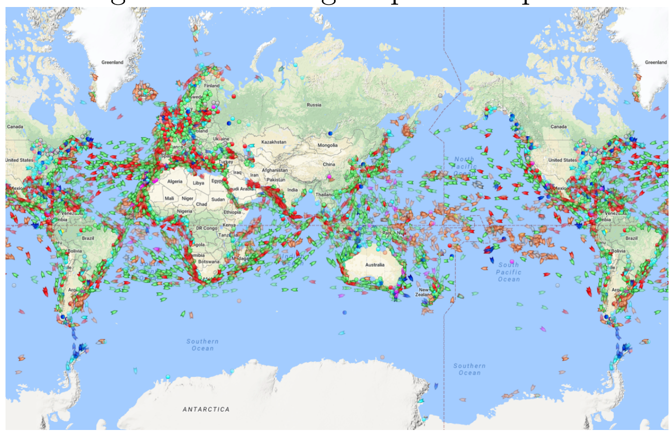
To our knowledge we are the first to try to compute a global comprehensive data based on satellite data. There are only a few recent papers that have used satellite data to explore issues related to trade. Brancaccio et al. (2017) study the role of the transportation sector in world trade focusing on search frictions and the endogeneity of trade costs. They use AIS data for a limited sample of dry bulk ships which typically carry commodities such as iron ore, coal, grain and sugar, to estimate their model. Our focus is rather on the global trade system, and our sample is the worldwide fleet of container ships which typically carry manufactured goods and is responsible for around two-thirds of world trade based on values. Unlike Brancaccio et al. (2017) we do not focus on transportation costs per se, but rather on features of the global trade system and the location of ports and countries within this system, and how these affect trade.

The rest of the paper is structured as follow. In Section 2 we outline and explain the satellite data for seaborne trade. In Section 3 we develop the box accounting methodology, while in Section 4 we take the box accounting methodology to the satellite data, compute a comprehensive global data set on trade flows and present some key descriptive statistics. In Section 5 we establish how our satellite based trade flows mirrors the trade flow based on custom data. In Section 6 we present an application of the network measures that can be calculated based on the box accounting methodology and how these can improve our understanding of the drivers of trade. Section 7 concludes.

2 AIS Data and the Construction of Draught-based Cargo

We set out to explore how satellite information on shipping activity can complement our traditional data on trade flows in order to improve our understanding of world trade and its drivers. To do so we shall combine a set of data sources. Our primary source is AIS (Automatic identification System) data from Marine Traffic that allows us to track ships around the globe through a whole year. We shall match these data with time invariant characteristics of the individual ships from the Clarkson World Fleet Register. Based on

Figure 1: Tracking Ships from Space



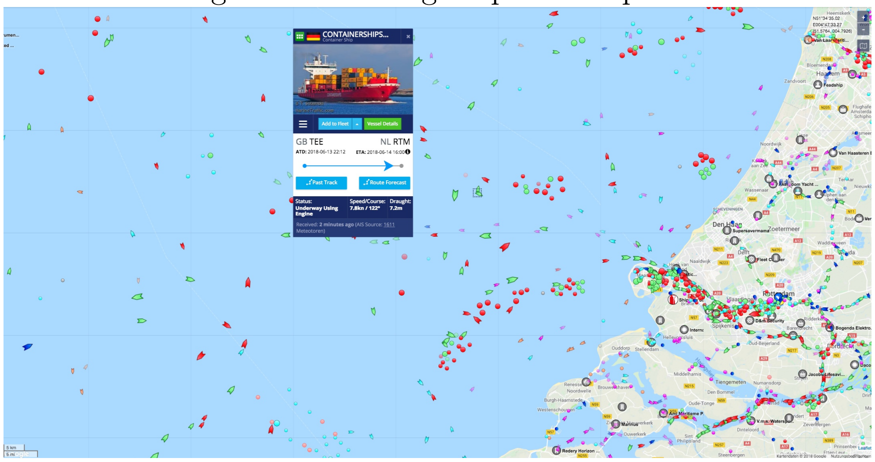
AIS and ship information we shall compute measures of the global cargo flows as well as measures of the global transport network.

2.1 AIS Data

Our data set is based on AIS (Automatic identification System) data from Marine Traffic. AIS is an automatic tracking system used on ships and by vessel traffic services (VTS). AIS is intended to assist a vessel's watchstanding officers and allow maritime authorities to track and monitor vessel movements. AIS information supplements marine radar. The International Maritime Organization's International Convention for the Safety of Life at Sea requires AIS to be fitted aboard international voyaging ships with 300 or more gross tonnage (GT), and all passenger ships regardless of size. The coverage of AIS globally has increased rapidly over the last decade, and does now allow for a global coverage of all vessels and ports of significance. AIS data are based on satellite pictures taken on a continuous basis and provides real time information on the location of a ship. Figures 1 and 2 provide examples of how the information is typically displayed. Each arrow is a ship, while the different colours are used to indicate the type of vessel (container, bulk, tanker etc).

Our aim is to match information on physical transport with data on trade flows. Every time a ship arrives or departs a port a signal is sent. This is referred to as a port call. We use data on all port calls made through the AIS system during the calendar year 2016. Every observation of a port call has a time stamp, which tells us when and on what date the call was made. The observation contains information on the name of the port, country and geographical location (latitude and longitude). In addition we get information on whether the ship is arriving or departing, whether it is in transit or not, as well as the draught of the ship at the time the port call is made. The latter is essential, since matched with other

Figure 2: Tracking Ships from Space



ship specific information, it will allow us to calculate a measure for the tonnage of the ship's cargo. Based on this measure we will construct a comprehensive data set for global cargo flows and this will form the basis for our analysis.

World seaborne trade relies on the transport by a range of different ships which operate and are loaded in distinctly different ways. In order to ensure homogeneity and thereby to be able to develop a methodology that allows us to construct a comprehensive data set for cargo, our point of departure is containerized trade, i.e.we limit the analysis of shipping activity to container ships. We note that containerized seaborne trade captures the majority of world trade. In terms of value, global seaborne container trade accounted for approximately 60 percent of all seaborne trade in 2016, while seaborne trade in turn accounted for around 80% of world merchandise trade.

2.2 Descriptives on Ships and Ports

We have detailed information on all port calls made by container ships in 2016.[1] In the Appendix Section A we describe how we clean the raw AIS data for port calls. This leaves us with a comprehensive data set for 3,146 container ships and the 498 ports they move between as described in Table 1. We see that there is a huge variation both with respect to the size of the ships (measured in deadweigt tonnes), and the numbers they pass as well the number of ships that arrive and depart from the ports. Most container ships sail fixed routes. This means that they typically visit the same port many times. Hence, we also reports summary statistics for *distinct ports*, where the mean per ships is not more than 13.

[1]Based on the ship categories used by data supplier, Marine Traffic, we use the ships categorized as "container ship" and "Cargo/containership".

Table 1: Ships and Ports

Variable:	Obs	Mean	Sd	Min	Max
Ships:					
Deadweight tonnes (dwt)	3,146	57,305	42,363	15	202,376
# ports passed	3,146	69	38	1	312
# distinct ports passed	3,146	13	7	1	46
Ports:					
# incoming ships	498	669	1,472	6	14,534
# outgoing ships	498	669	1,460	5	14,346

Note: Summary statistics are based on the port calls made by container ships in 2016. Effective dwt is calculated based on dwt and draught and is used as a proxy for cargo. Incoming and outgoing deadweight tons are in millions. Only ships with deadweight tonnes>15800 and trips with non-zero duration are used. Summary statistics include only routes taken by at least 5 ships and only routes between ports that appear both as arrival and departure ports.

2.3 Computing Draught-based Cargo

To create a comprehensive data set for global cargo flows based on the AIS data we proceed by using the information that we have on ships' draught. Draught of a ship is the distance between the surface of the water and the lowest point of the vessel. A ship's draught inform as about its load. Over the last years the use of draught-based estimates of ships' cargo has emerged in the maritime transport literature, see e.g. Adland et al. (2017). We build on this approach and limit the analysis to one type of ships, namely container ships, we are able to establish a relatively simple rule for the computation of the ships' cargo. We define a ship as empty, referred to as the ship sailing in ballast, if its draught is smaller than a given threshold, i.e. draught (H_A) $<$ ballast draught (H_B). More specifically, we define $H_B = 0.65H_s$, where H_S depicts the ship's scantling draught. *Scantling draught* is the draught the ship will have when it is fully loaded and is also referred to as *design draught*, as it is this draugt it is build for. We use 0.65 as weight based the maritime engineering literature. Given that a firm is assumed to be carrying cargo we calculate its draught-based estimate of cargo, from now on referred to as *effective dwt* (B), *as:*

$$B = dwt * (\frac{H_A - H_B}{H_S - H_B}) \tag{1}$$

As we proceed we introduce the notion *trip*. Trips are computed based on the information on time and date for a port call. A trip is the voyage between two ports, where neither the

Table 2: Ships, Trips and Port

Variable:	Obs	Mean	Sd	Min	Max
Ships:					
Share of trips in ballast (<65%)	3,146	0.12	0.15	0	1
Trips:					
Actual draught (% of scantling draught)	311,851	0.82	0.09	0.65	1
Effective dwt on loaded trips	311,851	20,989	22,508	0.09	199,744
Ports:					
Total incoming effective dwt in millions	498	13	34	0	412
Total outgoing effective dwt in millions	498	13	35	0	412

Note: Summary statistics are based on the port calls made by container ships in 2016. Effective dwt is calculated based on dwt and draught and is used as a measure for cargo. Only ships with deadweight tonnes>15800 and trips with non-zero duration are used. Summary statistics include only routes taken by at least 5 ships and only routes between ports that appear both as arrival and departure ports.

departuring port nor the arriving port is registered as a port where the ship only reports to be in transit. In Table 2 we show that on average ships do 12% of their ships in ballast, i.e. without cargo. We observe that there is substantial variation across trips with respect to draught and the effective dwt carried as well as between ports when it comes to the cargo arrving and departing.

3 Box Accounting and Box Flows

We want to use the AIS information on vessel activity combined with the draugt-based estimates of cargo to construct a comprehensive data set of global cargo flows. In this section we develop a methodology that allows us to do so. With reference to the nick name for container, "box", we shall in the following refer to cargo as *boxe*s and the methodology we develop as *box accounting*. We face one key challenge: We observe the boxes traveling between port i and j, but not the port of origin nor the final destination of the boxes, i.e. the countries of the actual trading partners.To tackle this challenge we develop a methodology that allows us to calculate number of boxes *sent* from port i to port k building on the methodology that is used to track origin of value added in trade flows.

3.1 Definitions

Let x_{ij} be the number of boxes (effective dwt) leaving port i going to port j. Then the total sum of boxes leaving port i is given by

$$Out_i = \sum_j x_{ij} \tag{2}$$

while the total sum of boxes arriving in country j is

$$In_j = \sum_i x_{ij}. \tag{3}$$

To aid the analyse we moreover define α_k as the share of incoming boxes to port k that are passed through to other ports; β_{ij} as the share of boxes to j in total boxes leaving i; and b_{ij} as the share of boxes from i in total boxes leaving j. It follows that

$$b_{ij} = \frac{\alpha_j x_{ij}}{x_j} \; for \; i \neq j, \tag{4}$$

that $b_{ii} = 0$, and that $b_{ij} = \alpha_j \frac{x_i}{x_j} \beta_{ij}$. This leaves us with a IO-matrix, B, describing the network between the ports:

$$B = \begin{bmatrix} 0 & \dots & b_{1i} & \dots & b_{1J} \\ \vdots & \ddots & & & \\ b_{i1} & & 0 & & \\ \vdots & & & \ddots & \\ b_{J1} & & & & 0 \end{bmatrix}$$

The number of boxes traveling *directly* from j to final destination k can be computed as

$$f_{jk} = (1 - \alpha_k) x_{jk}, \tag{5}$$

and the total number of boxes shipped directly from j to all final destinations $f_j = \sum_k f_{jk}$ while we note that $f_{jj} = 0$. The box accounting framework relies on a couple of key the assumptions: The passthrough coefficient α_k does not depend on origin of the boxes, port j, and the share of boxes from i in total boxes leaving j, b_{ij}, i.e. how much is absorb by the hinterland in a given location, does not depend on the final destination k.

3.2 Box Accounting

We proceed by establishing the box accounting identity, and start with a three-country example to aid the intuition for the underlying idea:

$$
\begin{aligned}
x_1 = & \underbrace{b_{12}x_2 + b_{13}x_3}_{\substack{\text{boxes from 1 shipped} \\ \text{to 2,3 and passed on}}} \quad + \quad \underbrace{f_{12} + f_{13}}_{\substack{\text{boxes from 1 shipped} \\ \text{to 2,3 that stay}}} \\
= & \frac{\alpha_2 x_{12}}{x_2} x_2 + \frac{\alpha_3 x_{13}}{x_3} x_3 + (1-\alpha_2)x_{12} + (1-\alpha_3)x_{13} \\
= & \underbrace{\sum_k x_{1k}}_{\substack{\text{total boxes shipped} \\ \text{out of 1}}}
\end{aligned}
$$

On this backgroun we establish the N-port accounting identity

$$X = BX + F, \tag{6}$$

which implies that

$$X = LF \qquad \text{where } L = (I-B)^{-1} = \sum_{m=0}^{\infty} B^m. \tag{7}$$

We can show the N-port accounting identity in detail as

$$
\begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_J \end{bmatrix} = \begin{bmatrix} 0 & \dots & b_{1i} & \dots & b_{1J} \\ \vdots & \ddots & & & \\ b_{i1} & & 0 & & \\ \vdots & & & \ddots & \\ b_{J1} & & & & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_j \\ \vdots \\ x_J \end{bmatrix} + \begin{bmatrix} f_1 \\ \vdots \\ f_i \\ \vdots \\ f_J \end{bmatrix} = L \begin{bmatrix} f_1 \\ \vdots \\ f_i \\ \vdots \\ f_J \end{bmatrix}
$$

3.3 Box Accounting by Destination Port, Departure Port, and Route

We can now compute number of boxes travelling from i to k along any possible route. By splitting up $F = \sum_k F^k$ into destination ports k we obtain total boxes absorbed in k by departure port (in rows): $X^k = LF^k$

$$\begin{bmatrix} x_1^k \\ \vdots \\ x_i^k \\ \vdots \\ x_J^k \end{bmatrix} = L \begin{bmatrix} f_{1k} \\ \vdots \\ f_{ik} \\ \vdots \\ f_{Jk} \end{bmatrix}$$

where x_i^k is the number of boxes traveling through port i that are ultimately absorbed in k but can take any route in between. However these boxes are not equal to, nor do they include, the boxes added in i. We proceed by computing the boxes travelling from i to k along a m-hop route. Making use of

$$L = (I - B)^{-1} = \sum_{m=0}^{\infty} B^m \tag{8}$$

we can split up X^k according to m-hop travel routes. Boxes absorbed in k arriving on direct route (0 hops), by departure port is given by

$$X^{k0} = B^0 F^k = F^k \tag{9}$$

where $x_i^{k0} = f_{ik}$ is the ith element of X^{k0} and we note that B^0 means B to the power of 0, i.e. B^0 is the first element of $\sum_{m=0}^{\infty} B^m$ while the other superscripts $k, k0$ are not exponents but indices, indicating destination country k, respectively, k reached through a 0-hop route. Based on this we define

- Boxes absorbed in k arriving on **one-hop routes**, by departure port:

$$X^{k1} = BF^k \tag{10}$$

- Boxes absorbed in k arriving on **two-hop routes**, by departure port:

$$X^{k2} = B^2 F^k \tag{11}$$

- Boxes absorbed in k arriving on **m-hop route,** by departure port:

$$X^{km} = B^m F^k \tag{12}$$

- Total boxes absorbed in k, via any route by departure ports:

$$X^k = \sum_{m=0}^{\infty} X^{km} = \sum_{m=0}^{\infty} B^m F^k = LF^k \tag{13}$$

3.4 Box Accounting by Destination Port, Departure Port, Number of Hops, and last Hop Port

We can get more information about the routes by turning the vector F^k into a diagonal matrix. In contrast to the above, we do not sum over the second to last port, i.e., the port from which boxes reach the final destination. Hence, we can compute boxes absorbed in k arriving on **one-hop routes**, by start port and hop port:

$$\mathbf{X}^{1k} = \begin{bmatrix} x_{11k} & \dots & x_{1jk} & \dots & x_{1Jk} \\ \vdots & \ddots & & & \\ x_{i1k} & & & & \\ \vdots & & & \ddots & \\ x_{J1k} & & & & x_{JJk} \end{bmatrix} = B \begin{bmatrix} f_{1k} & \dots & 0 & \dots & 0 \\ \vdots & \ddots & & & \\ 0 & & f_{kk} & & \\ \vdots & & & \ddots & \\ 0 & & & & f_{Jk} \end{bmatrix}$$

with typical element

$$x^{k1}_{ijk} = x_{ijk} = b_{ij} f_{jk} = b_{ij} x_{jk}(1-\alpha_k) = \alpha_j \frac{x_{ij}}{x_j} x_{jk}(1-\alpha_k). \tag{14}$$

of which $\nu_i x^{k_1}_{ijk}$ boxes are from country i's hinterland. We can moreover compute boxes absorbed in k arriving on **two-hop route**s, by departure port and route as

$$\mathbf{X}^{\mathbf{k2}} = B^2 diag(F^k) \tag{15}$$

with typical element

$$x^{k2}_{ij_2k} = \sum_{j_1} x_{ij_1j_2k} \quad \text{where} \quad x_{ij_1j_2k} = b_{ij_1} b_{j_1j_2} x_{j_2k}(1-\alpha_k) \tag{16}$$

3.5 Box Accounting Identity

Alike the computation of bilateral trade in value added, we can now compute the number of boxes from port i's hinterland that is absorbed by port k's hinterland. We have that the share of boxes added in i in total boxes leaving i is

$$\nu_i = 1 - \sum_{h \neq i} b_{hi} \tag{17}$$

It follows that we have boxes from i absorbed by k traveling directly: $\nu_i x^{k0}_i = \nu_i f_{ik} = \nu_i(1-\alpha_k)x_{ik}$, via all m-hop routes: $\nu_i x^{km}_i$, and via any route: $\nu_i x^k_i$, where x^{km}_i is the ith element of X^{km}. This allows us to calculate the following identity for box flows:

$$\begin{aligned}\underbrace{x_j}_{Out_j} &= \underbrace{\alpha_j \sum_i x_{ij}}_{Passthrough_j} + \underbrace{\nu_j x_j}_{Send_j} \\ &= \alpha_j \underbrace{\sum_i x_{ij}}_{In_j} + \underbrace{\nu_j x_j}_{Send_j}\end{aligned} \tag{18}$$

where Out_j depicts the flow of boxes leaving port j, In_j depicts the flow of boxes arriving in port j, $Send_j$ depicts the flow of boxes sent from port j and originating form its hinterland. We note that

$$\alpha_j = \frac{Out_j - Send_j}{In_j} \tag{19}$$

and

$$\nu_i = \frac{Send_i}{Out_i}. \tag{20}$$

Finally, we have that

$$1 - \alpha_k = \frac{Receive_k}{In_k} \tag{21}$$

where $Receive_k = \sum_i x_i^k$ is the total number of boxes that *stay* in a location and

$$In_k - Receive_k = Out_k - Send_k. \tag{22}$$

Note that the aggregate accounting identity $X = LF$ itself does not rest on proportionality assumptions on the key coefficients. We need the proportionality assumptions outlined above for two purposes. To compute the unobserved b_{ij} we assume that the passthrough coefficient α_j is the same for every origin i. This assumption implies that the unobserved share of boxes that stay also does not depend on i and hence we can compute the boxes that stay as $f_{ik} = (1 - \alpha_k)x_{ik}$. When we break up the aggregate accounting identity into destinations k we assume implicitly that the travel history of boxes traveling from j to k is the same as that of boxes traveling from j to ℓ. In other words, $X = LF$ does not in general imply $X^k = LF^k \; \forall k$. The general formula is $X^k = L^k F^k$ where $L^k = (I - B^k)^{-1}$ and a typical element of B^k, b_{ij}^k, describes the share of boxes from i contained in boxes passing j with final destination k. $X = LF$ implies $X^k = LF^k \; \forall k$ only if $L^k = L \; \forall k$ which implies $b_{ij}^k = b_{ij} \; \forall ijk$. I.e., the second proportionality assumption requires that the foreign shares of of boxes passing through j do not vary across destinations k. This implies that the share of boxes added, $\nu_i = 1 - \sum_{h \neq i} b_{hi}$, also does not depend on the destination of a shipment.

3.6 Computing Box Flows Combining AIS and Box Accounting

We want to compute global bilateral cargo flows, *box flows*, between ports, that in turn can be aggregated up to the country level. To do so we now combine the box accounting methodology with AIS data. Based on the identity established in the box accounting (see (22)), we only need to calculate one parameter for each port, (α_k) , the share of incoming boxes to port j that passes through relative to total number of incoming boxes, in order to be able to construct a comprehensive data set of global box flows for 2016. It follows from (21) that to do so we need information on In_k and $Receive_k$. The former we observe and can be calculated based on the port-to-port information on shipping activity and our draught based measure of cargo. The latter we need to compute subject to an assumption on ships onloading and offloading practice.

Recall that $Receive_k = \sum_i x_i^k$ is the total number of boxes that arrives and *stay* in location k. We compute $Receive_k$ using changes in ship's draught at port k. Our point of departure are all ships s leaving port k for which departure draught $<$ arrival draught, i.e. ships that by definition must have offloaded more than they onloaded. Based on this universe we let $Receive_{sk}$ be calculated based on the diffence in draught between arrival and departure:

$$Receive_{sk} = (draught_{s,k,arrival} - draught_{s,k,departure}) * DWT_s \tag{23}$$

This measure represents a lower bound for cargo received in a location and will thus provide us with a lower bound for $(1 - \alpha_k)$. It will underestimate the arrived cargo if a firm has departure draught $<$ arrival draught but its gross offload is bigger than its net offload. It will also underestimate the arrived cargo for ships with $Receive_{sk} > 0$ but departure draught $>$ arrival draught which will be the case if the onloaded cargo exceeds the offloaded cargo. We choose this as our preferred measure since unlike all other alternatives the direction of the mismeasurement is unambiguous. In the Appendix Section B for a thorough discussion of the alternatives.

4 Descriptives on Global Box Flows

Our detailed AIS data on port-to-port flows provides extensive information on cargo flows. Using the draught based cargo measure and the box accounting methodology we are now able to transform these port-to-port data into a comprehensive global data set with bilateral

Table 3: Top 10 Ports: In and Out going Boxes

Port Name	Country Name	Out	In	Out+In	Out-In	Rank
Port of Singapore*	Singapore	412.07	411.61	823.68	0.46	1
Port of Shekou*	China	259.84	269.68	529.52	-9.84	2
Port of Shanghai	China	281.68	216.04	497.72	65.64	3
Port of Hong Kong	Hong Kong	206.83	223.14	429.97	-16.31	4
Port of Port Klang	Malaysia	191.93	197.77	389.70	-5.85	5
Busan New Port*	Korea	182.15	187.49	369.64	-5.33	6
Port of Yantian	China	152.25	124.64	276.90	27.61	7
Port of Ningbo	China	153.60	121.84	275.44	31.75	8
Port of Kaohsiung	Taiwan	135.00	128.07	263.06	6.93	9
Rotterdam Maasvlakte*	Netherlands	116.36	123.68	240.04	-7.32	10

Note: This table reports the incoming (In), outgoing (Out), Out+In, Out-In effective deadweight tons in millions of the 10 ports with the highest In+Out effective deadweight tons. Ports are ranked by their In+Out numbers. Summary statistics are based on port calls made by container ships. Effective dwt is calculated based on dwt and draught and is used as a proxy for cargo. Incoming and outgoing deadweight tons are in millions. Only ships with deadweight tonnes>15800 and trips with non-zero duration are used. Summary statistics include only routes taken by at least 5 ships and in only routes between ports that appear both as arrival and departure ports. * indicates a cluster of ports located that are located < 30 km from each other.

cargo flows between ports, and which in turn can be aggregated up to bilateral country data for the calendar year 2016. In this section we explore features of this data set.

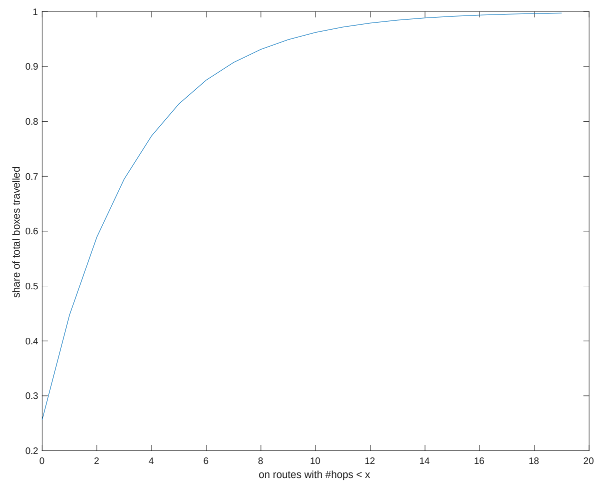
In the following we shall consistently refer to cargo as boxes, but note that our measure of cargo is still effective deadweight (dwt) tonnes, i.e. our draught based cargo estimate. In Table 3 we rank the ten most busy ports based on the sum of boxes coming in and going out. We observe that there is a lot of variation across ports and that with the exception of Rotterdam in the Netherlands, the busiest ones are all located in Asia.

In Figure 3 we show the cumulative share of boxes traveling according to the number of hops on their travelled routes. We see that around fifty percent of the cargo travelling has routes with two hops or less.

In Figure 4 we illustrate the relationship between the amount of boxes that leave each port (Out_j) and the amount of boxes that are sent ($Send_j$), i.e. exported, from the port. It follows from the box accounting methodology that $Send_j$ must be smaller or equal to Out_j, and the plot of all global ports confirms this, at the same time as it illustrate that there is considerable variation regarding the amount of cargo leaving a port that is passing through and has another port of origin.

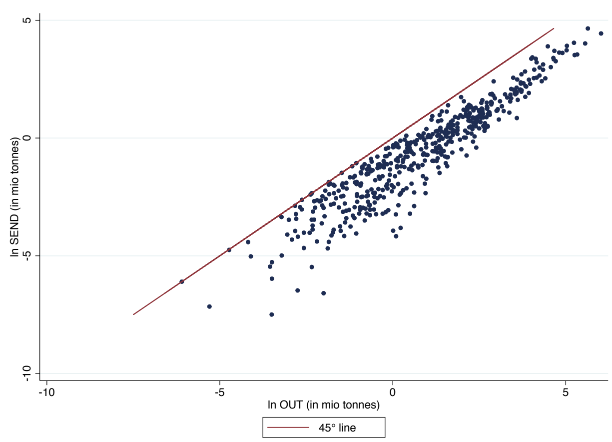
I Figure 5 we illustrate the flipside, namely the relationship between the amount of boxes

Figure 3: Boxes, Trips and Hops



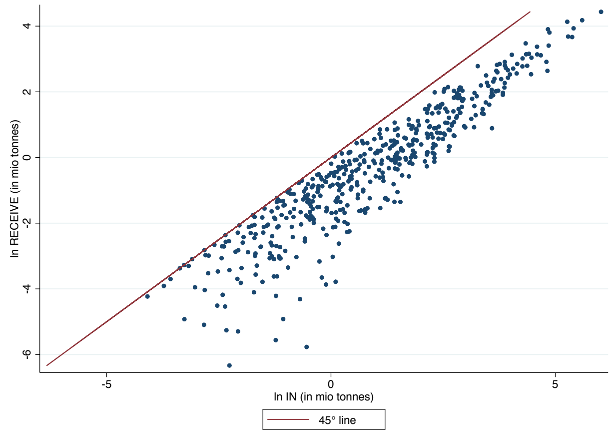
Note: The figure shows the cumulative share of boxes traveling on multi-hop routes.

Figure 4: $Send_j$ and Out_j



Note: The figure shows the relationship between $Send_j$ and Out_j boxes by port.

Figure 5: $Receive_j$ versus In_j



Note: The figure shows the relationship between $Receive_j$ and In_j boxes by port.

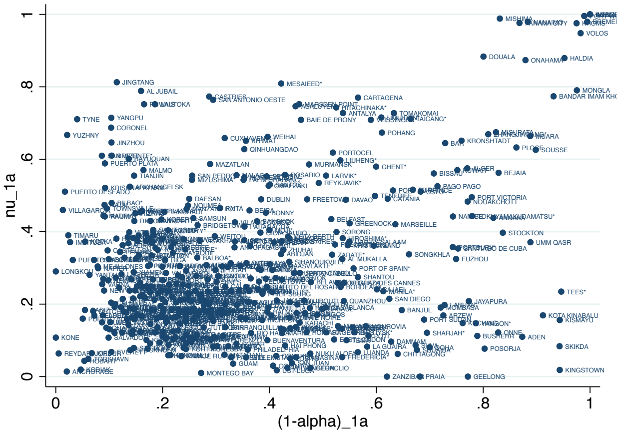
that arrive in each port (In_j) and the amount of boxes that are absorbed ($Receive_j$), i.e. imported and not passed through. It follows from the box accounting methodology that $Receive_j$ must be smaller or equal to In, and the plot of all global ports confirms this, at the same time as it illustrate that there is considerable variation regarding the amount of cargo arriving a port that is passing through and has another port of origin.

Finally, we characterize the ports based on the port specific coefficients computed by the boxing accounting methodology, recalling that $\nu_j = Send_j/Out_j$ and $(1-\alpha_j) = Receive_j/In_j$. Both these coefficients provide information on where the port is in the global cargo chain. If the port typically the end destination for cargo flows, then $1 - \alpha_k$ will be high. Similarly, if ν_j is high, then the port is typically at the beginning of the global cargo chain. We observe that there is substantial variation across ports with respect to where they are in the global cargo chain.

5 Satelitte versus Reported Trade Data

Our first aim is to establish whether inter-country seaborne trade flows based on the tracking of vessel movements over time can provide an alternative to reported customs data. Hence, inspired by the light studies, where night lights are used as a measure of economic growth (see e.g. Henderson et al. (2012)), we set out to examine whether the box flows mirrors the bilateral reported trade flow. Hence we want to compare our box flow measures $Send_{ik}$ and Out_{ik} to reported export flows from country i to country k. To do so we first construct

Figure 6: Port Heterogeneity: ν_j vs $(1 - \alpha_j)$



Note: The figure shows the computed ν_i on the vertical axis and $(1 - \alpha_i)$ on the horizontal axis by port.

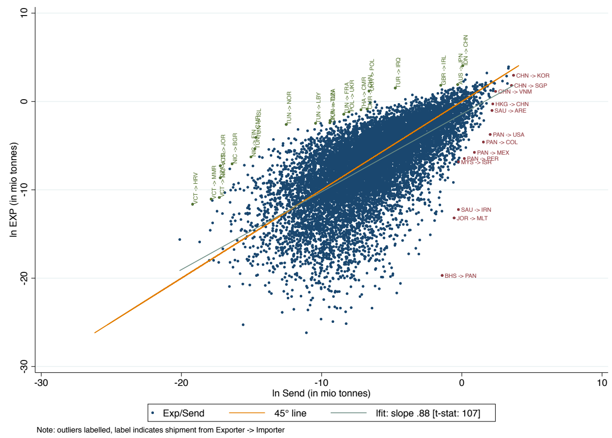
country level measures of of box flow measures of trade, and proceed by aggregating the port-to-port level data to the country-to-country level using weighted averages, i.e. we use the tonnes sent by a given port in total tonnes sent as weights.

Next, we use the comprehensive trade data set available from Comtrade and BACI as well as US Customs data that allows us to identify containerized products.[2] We calculate three measures of bilateral exports based on these sources: total export value, total exports quantity in tonnes, and containerized export quantity. The latter use information on US trade for to identify what products can be catergorised as containerized product. This is the measure of exports which we expect to be most closedly related to the box flow measures. Appendix Section C explains in details how this variable is constructed.

In Table 4 we provide a summary of the bilateral flows between country pairs based on the box flow data set as well as the trade data. We observe that the mean for $Send_{ik}$ and containerized exports quantity are identical while the maximum values are in roughly in the same ball park, while as expected the figures for the Out flows are as expected considerably higher as it can be thought of as a type of upper bound for the outward flows of goods from the country of departure.

[2]See the Appendix Section C for details on the trade and customs data.

Figure 7: Bilateral Trade: Exports versus Send



Note: The figure compare goods flows based on trade statistics and Ais data. Exports refer to containerized exports in mio tonnes. Numbers in log. Outliers are labeled with labels indicating shipment from exporter to importer country. 2016 data, observations where exports or $Send_j$ are missing are dropped.

Table 4: Bilateral Trade: Summary statistics

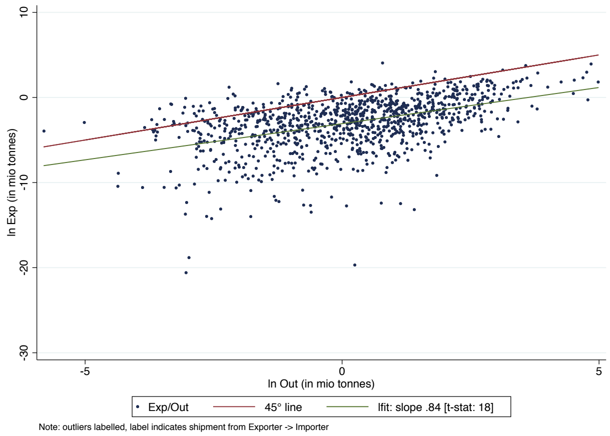
	Obs	Mean	St	Min	Max
Send (in mio computed tonnes)	13,469	0.10	0.79	0	39.92
Out (in mio computed tonnes)	13,469	0.33	3.29	0	146.10
Containerized exports quantity (in tonnes)	13,469	0.10	0.95	0	57.60
Total exports value (in mio USD)	13,469	0.95	7.75	0	428.17
Total exports quantity (in mio tonnes)	13,469	0.93	11.21	0	808.43

Note: Bilateral relationships for which *Send* or export is missing are dropped.

We explore the correlation between containerized exports quantity and the box flow measures for export. In Figure 7 we plot exports versus *Send* box flows for each bilateral country relationship. We observe that there is a high correlation between the two with a correlation close to .90. We also examine the relationship between *Out* box flows and exports, see Figure 8. Also here we observe a high correlation, but again as expected, not as high as for *Send*. Finally, we also aggregate up to the country level and compare exports and *Send*, see Figure 9, and again the correlation is above .80.

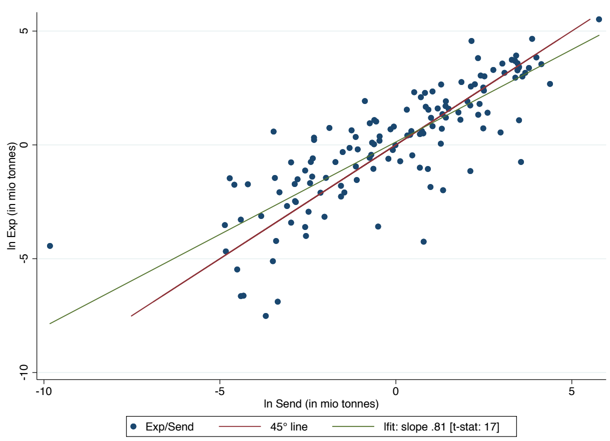
On this background we now perform a standard gravity analysis to explore the drivers of reported trade flows as compared to the box flow measures of trade. As we only have

Figure 8: Bilateral Trade: Exports versus Out



Note: The figure compare goods flows based on trade statistics and Ais data. Exports refer to containerized exports in mio tonnes. Numbers in log. Outliers are labeled with labels indicating shipment from exporter to importer country. 2016 data, observations where exports or $Send_j$ are missing are dropped.

Figure 9: Country-level trade: Exports versus Send



Note: The figure compare goods flows based on trade statistics and Ais data. Exports refer to containerized exports in mio tonnes. Numbers in log. 2016 data, observations where exports or $Send_j$ are missing are dropped.

Table 5: Gravity analysis with Satelitte Data

Variable:	ln Exp Value (1)	ln Exp Quant (2)	ln Container Exp (3)	ln Send (4)	ln Out (5)
ln GDP origin country	1.191^a	1.308^a	1.230^a	0.896^a	0.319^a
	(0.032)	(0.045)	(0.042)	(0.055)	(0.043)
ln GDP destination country	0.908^a	0.972^a	0.880^a	0.599^a	0.194^a
	(0.027)	(0.044)	(0.038)	(0.046)	(0.031)
ln Distance	-1.209^a	-1.683^a	-1.196^a	-1.131^a	-0.066
	(0.084)	(0.108)	(0.113)	(0.117)	(0.100)
Contiguity	0.355^c	0.495^c	0.316	0.573^b	0.448^b
	(0.195)	(0.257)	(0.238)	(0.240)	(0.173)
Common language	0.728^a	0.647^a	0.600^a	0.812^a	0.339^b
	(0.127)	(0.163)	(0.172)	(0.202)	(0.142)
FTA	0.461^a	0.253	0.340^b	-0.153	0.050
	(0.127)	(0.168)	(0.162)	(0.195)	(0.183)
Observations	11,963	11,938	11,866	11,963	1020
Adjusted R^2	0.71	0.61	0.58	0.486	0.23

Note: All dependent variables in logs for the year 2016. *ln Exp value* refers to total export value (containerized and non-containerized exports) in mio USD, *ln Exp Quant* refers to export quantity (containerized+non-containerized) in mio tonnes, *ln Container Exp* refers to containerized export quantity in mio tonnes, *ln Send* refers to computed export based on AIS data, *ln Out* refers to computed outflow of tonnes based on AIS data. *FTA* denotes joint membership in a free trade agreement (Source: WTO). Standard errors clustered by origin and destination country in brackets. ^{a} p< 0.01, ^{b} p< 0.05, ^{c} p< 0.1.

box flow data for one year, our analysis will be cross sectional. The results are reported in Table 5. For the two measures we expect to be closest to each other, containerized exports and *Send* boxes, we observe coefficients and degree of significance that are very close to each other for all independent variables apart from the dummy variable that indicates whether there is a free trade agreement between the two countries (FTA). While this has significant positive impact on reported trade, it does not have a significant impact on box flows. Our results suggest that our box flow data set can complement the reported customs data for trade and thus allow for improved understanding of global trade flows and their drivers.

6 Application I: Network Analysis

We proceed by exploring possible applications for the use of the AIS data. Combining the AIS data with our Box Accounting methodology we are able to shed light on the the position of ports/countries within the global container shipping network. This allows us to explore the connectedness and remoteness of ports and countries and the travel time betwen ports. Hence, the data at hand potentially allows for an improved understanding of the global trade network and the drivers of trade cost. Employing the box accounting methodology we proceed by constructing a comprehensive global data set for a set of network features of the box flow data.

6.1 Constructing Network Centrality Measures

We start by providing an overview of the netwok measures we will focus on and how they are defined and constructed. We limit the analysis to the set of measures that we believe are important drivers of trade costs.

Number of Hops As illustrated in Figure 3 most cargo flows do not travel directly from the port/country of origin to the port/country of destination. 75 percent of the trips have one or more hobs. We constructed three measures based on the routes' hobs: (i) *Minimum number of hobs* is shortest connection between two ports in terms of hops. Using network terminology, this measure represents the degrees of separation in unweighted directed network. We calculate this based on the raw data for port-to-port using the shortest path tool of MATLAB's graph package as well as based on our box accounting framework. (ii) *Modal number of hobs* is the route length in terms of hops along which the largest number of boxes flows between two ports according to the box accounting framework. (iii) *Weighted average of hobs* is the weighted average of number of hops between two ports across all possible routes of varying lengths and is calculated based on the box accounting framework.

Travel Time Based on the time stamps provided in our port-to-port data we observe *direct travel time* between two ports, but this is obviously only available for a subset of port-to-port connections. Namely, for the connections where we observe ships sailing directly from the departure to the arrival port.[3] The direct travel time between two ports is computed as the median over all ships' trip durations. Note that these travel times are *directed*, i.e. the travel time from D to A need not equal the travel time from A to D.

We also use the MATLAB graph tools and the raw port-to-port data to calculate the *indirect minimum travel time*, defined as the shortest travel time between two ports which are

[3]Note, that as we calculate travel time, we exclude trips which involve crossings of anchorages where the ship does not indicate that it is *in transit.* Moreover, we exclude port-to-port connections where less than 5 ships were observed over the whole year. Note also that the travel time reflects the time it takes to get from port D's geofence to port A's geofence. It does not include time spent traveling, waiting or lading/unlading, within the port area.

not directly connected. We compute this measure as shortest path in weighted directed network where edges reflect direct connections and weights are the direct travel times described just above.

Based on the box accounting framework we eventually calculate the *weighted average travel time* as the weighted average time over all routes along which boxes flow according to the box accounting. Weights are given by relative number of boxes on a route. More specifically, we definet_{ik} as travel time on direct route between i and k that is observed in the data whenever a direct route exists. Letting x_{ik}^k be boxes leaving i on any route with final destination k, we have that based on the same proportionality assumptions as before, the weighted average travel time across all routes between i and k is

$$\bar{t}_{ik} = \sum_j \frac{b_{ij} x_{jk}^k}{x_{ik}^k} (t_{ij} + \bar{t}_{jk}) + \frac{f_{ik}}{x_{ik}^k} t_{ik}. \tag{24}$$

Hence,$\bar{t}_{ik}$ is weighted average of direct and indirect travel time, $\frac{f_{ik}}{x_{ik}^k}$ is the share of total boxes leaving i with final destination k that travel directly, while $\frac{b_{ij} x_{jk}^k}{x_{ik}^k}$ is the share that travels to j first and from there takes any possible route to final destination k. Indirect travel time consists of the time it takes to get to j directly plus the weighted average time it takes to get to k from there. We can illustrate the calculation with a three-country example in matrix form:

$$\begin{aligned}
\begin{bmatrix} \bar{t}_{13} \\ \bar{t}_{23} \\ \bar{t}_{33} \end{bmatrix} &= \begin{bmatrix} 0 & \frac{b_{12} x_{23}^3}{x_{13}^k} & \frac{b_{13} x_{33}^3}{x_{13}^k} \\ \frac{b_{21} x_{13}^3}{x_{23}^k} & 0 & \frac{b_{23} x_{33}^3}{x_{23}^3} \\ \frac{b_{31} x_{13}^3}{x_{33}^3} & \frac{b_{32} x_{23}^3}{x_{33}^k} & 0 \end{bmatrix} \begin{bmatrix} \bar{t}_{13} \\ \bar{t}_{23} \\ \bar{t}_{33} \end{bmatrix} \\
&+ diag \left\{ \begin{bmatrix} 0 & \frac{b_{12} x_{23}^3}{x_{13}^k} & \frac{b_{13} x_{33}^3}{x_{13}^k} \\ \frac{b_{21} x_{13}^3}{x_{23}^k} & 0 & \frac{b_{23} x_{33}^3}{x_{23}^3} \\ \frac{b_{31} x_{13}^3}{x_{33}^3} & \frac{b_{32} x_{23}^3}{x_{33}^k} & 0 \end{bmatrix} \begin{bmatrix} 0 & t_{21} & t_{31} \\ t_{12} & 0 & t_{32} \\ t_{13} & t_{23} & 0 \end{bmatrix} \right\} \\
&+ \begin{bmatrix} \frac{f_{13}}{x_{13}^3} & 0 & 0 \\ 0 & \frac{f_{23}}{x_{23}^3} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} t_{13} \\ t_{23} \\ 0 \end{bmatrix}
\end{aligned}$$

while

$$\bar{t}_k = \mathcal{B}^k \bar{t}_k + \mathcal{B}^k T' + \mathcal{F}^k t_k \tag{25}$$

gives the general expression where $\mathcal{B}^k$ collects the weights $\frac{b_{ij} x_{jk}^k}{x_{ik}^k}$ of all indirect routes from i to k, $\mathcal{F}^k$ collects the shares of direct routes $\frac{f_{ik}}{x_{ik}^k}$ for all i,$t_k = [t_{1k}, ...t_{jk}, ...t_{Jk}]$ ' collects all ports' direct shipping time to destination k and $T = [t_1, ...t_k, ...t_K]$. We can solve the above equation for $\bar{t}_k$ as follows:

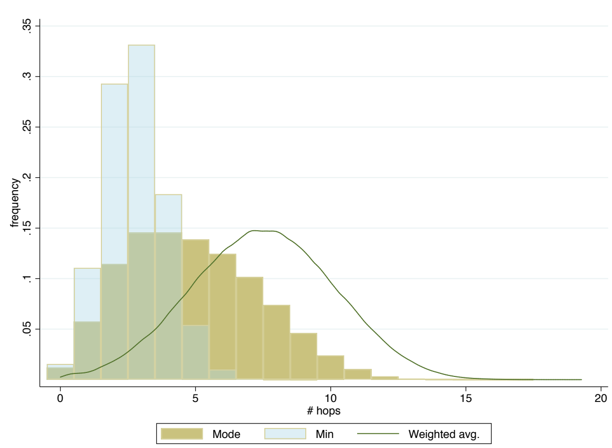
$$\bar{t}_k = (I - \mathcal{B}^k)^{-1}[diag(\mathcal{B}^k T') + \mathcal{F}^k t_k] \tag{26}$$

As discussed above, the box accounting framework implies that some share of a shipment returns to the departure port. Likewise, the time accounting will generate weighted average travel times of ports with themselves (i.e. $\bar{t}_{33} \geq 0$ in the above 3-country example).

6.2 Descriptives on Network Centrality Measures

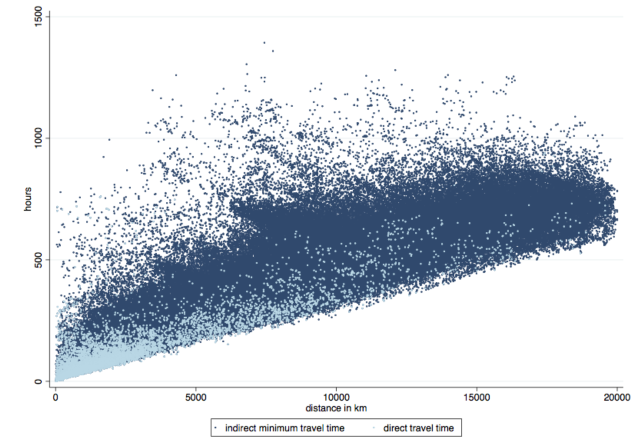
Figure 10 shows the frequency of minimum hops, modal hops, and the distribution of the weighted average hops among all port pairs. Figure 11 plots direct and indirect minimum travel times against geodetic distance for port pairs. We observe that there is a clear but not complete correlation between the time and distance measures. In Figure 12 we compare the distributions for direct travel time, minimum indirect travel time and weighted average time, and observe and the distributions for minimum and weighted average travel time are far from coinciding and the median weighted average time is far higher than the median for the minimum travel time. Figure 13compares the minimum travel times based on the shortest-path algorithm to the weighted average travel times from the box accounting. Each dot is a port-to-port connection. By definition, all dots lie above the 45-degree line.

Figure 10: Minimum, Modal, and Weighted Average # Hops - Distribution Across Port Pairs



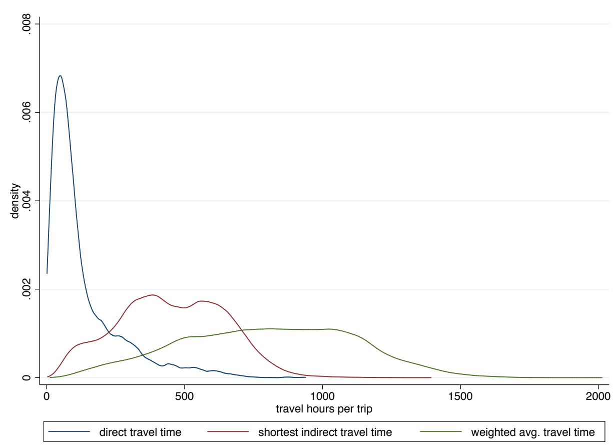
Note: Figure shows the distribution of minimum, modal, and weighted average # hops across port pairs. Computations are based on the box acounting methodology and port-to-port shipments from the AIS data.

Figure 11: Direct and Indirect Shipping Time - Distance



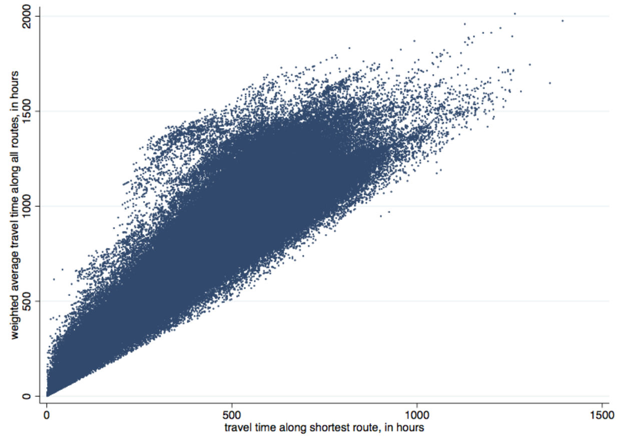
Note: Figure shows travel times between two ports. All computions are based on observed travel times between all regular (non-anchorage) ports in the AIS. Routes with less than 5 ships are dropped. Indirect travel time computed as shortest path in port network where edges, reflecting direct connections, are weighted by direct travel times.

Figure 12: Shipping Times - Distribution Across Trips



Note: Distribution of travel times between regular (non-anchorage) ports in the AIS data (median across ships on the same routes). Routes with less than 5 ships are dropped. Largest 1% of observations of indirect travel time winsorized for visibility.

Figure 13: Weighted Average vs Minimum Shipping Time

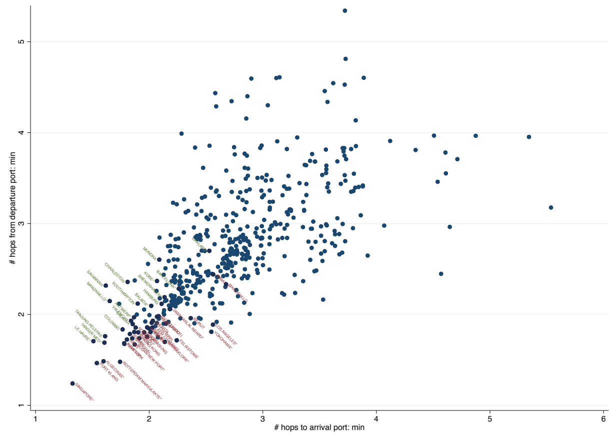


Note: Figure shows travel times between two ports. All computions are based on observed travel times between regular (non-anchorage) ports in the AIS. Routes with less than 5 ships are dropped. Indirect travel time computed as shortest path in port network where edges, reflecting direct connections, are weighted by direct travel times. Weighted average travel time computed using box accounting framework.

Using minimum number of hops computed based on the box accounting framework, in Figure 14 we can show how well connected – how central – ports are as origin and destination port respectively. We see that connectivity dependence on the direction of trade, while for many ports the two measures are clearly correlated. We also observe that ports differ substantially in the degree of connectivity.

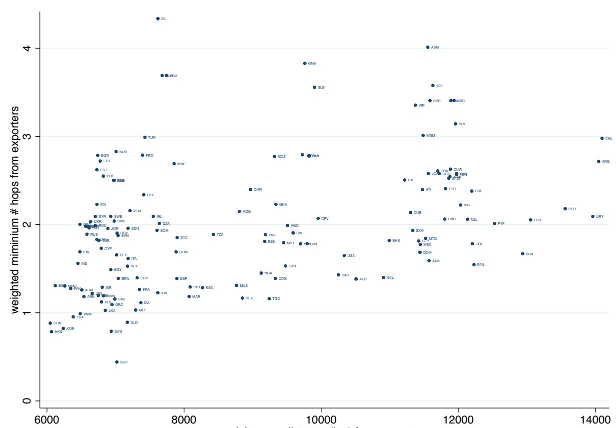
Country Level Analysis Aggregating over a country's ports, we obtain country level measures of centrality as origin or destination. Figures 15, 16 and 17 show how, the country's centrality as a destination market in terms of minimum number of hops, modal number of hops, and travel time compares to the countries centrality in terms of geogrphical distance. All measure in these figures, including distance, are computed as weighted averages across departure ports with weights reflecting the departure ports' share in global tonnes sent. Irrespective of centrality measure we see that some countries are well connected both when it comes to hops and distance, while others might by geographically rather central, but more remote when it comes to the number of hops or the time needed in order for goods from this country to reach their export market.

Figure 14: Port-Level Connectivity - Minimum # hops to Destination/Origin



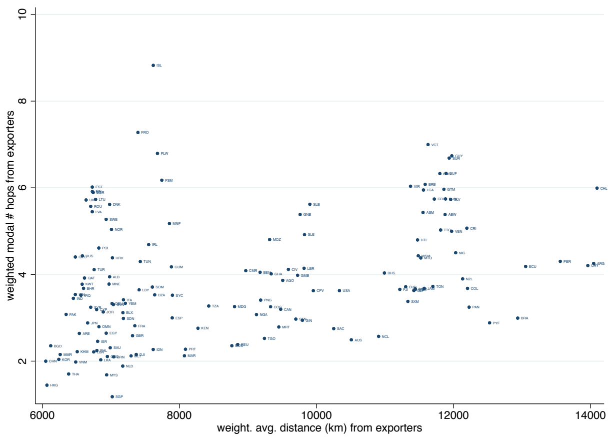
Note: Figure shows for all ports the average over the minimum # hops to all destination (from all origin) ports on the vertical (horizontal) axis. Largest 10% of ports are labelled.

Figure 15: Country-Level Connectivity - Minimum # hops



Note: Figure shows by country the weight. avg. travel time and distance from all exporting ports. Weights equal relative port size, i.e. a port's total shipments sent in global shipments sent.

Figure 16: Country-Level Connectivity - Modal # hops

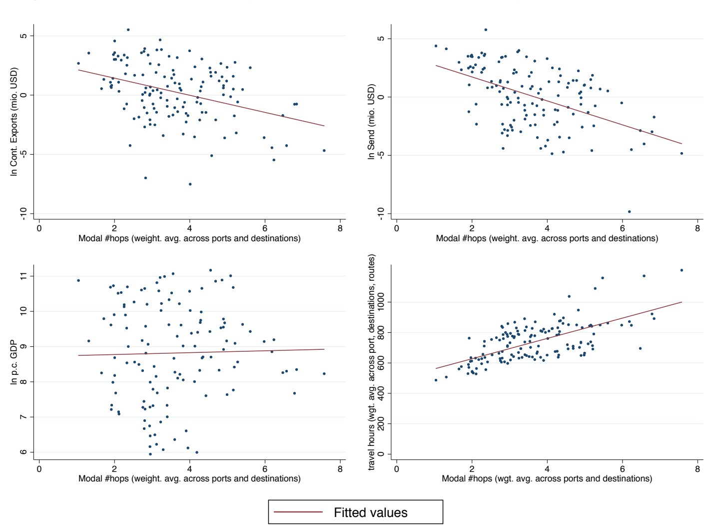


Note: Figure shows by country the weight. avg. modal # hops and distance from all exporting ports. Weights equal relative port size, i.e. a port's total shipments sent in global shipments sent.

Figure 17: Country-Level Connectivity - Average Travel Time

Note: Figure shows by country the weight. avg. travel time and distance from all exporting ports. Weights equal relative port size, i.e. a port's total shipments sent in global shipments sent.

Figure 18: Country-Level Connectivity, Trade, GDP



Note: Modal # hops to all export markets weighted with the exporters size, i.e. total shipments sent.

6.3 The Role of Centrality for Trade

The standard gravity model includes a time invariant measure of geographical distance. However, this is a rather crude proxy for trade cost. To improve the understanding of the role of location for trade, we want to examine the role of centrality of a location for trade more thoroughly using the meaures of centrality calculated above. We first offer a graphical presentation by some key relationships, see Figure 18, while in Table 6 we report the results from a cross sectional gravity analysis using a set of standard right hand side variable and adding to these measures for weighted average travel time and the modal number of hops. As left hand side variables we use reported total export value for all containerized as well as non-containerized exports (*lnExpVal*), containerized export quantity in millions tonnes (*lnContExp*) and our box flow measure for trade flows (*ln*Send). We see that conditional upon geographical distance, travel time has a significant neagtive effect on bilateral trade. The same applies to the the number of hobs it takes to get from the origin to the destination country. Accounting for travel time and number of hobs on the export route reduces the magnitude of the effect of distance.

Table 6: Gravity analysis with satelitte data

Variable:	ln ExpVal (1)	lnContExp (2)	lnSend (3)	lnExpVal (4)	lnContExp (5)	lnSend (6)	lnExpVal (7)	lnContExp (8)	lnSend (9)
lnDistance	-1.526^a	-1.576^a	-1.602^a	-1.019^a	-0.858^a	0.050	-1.025^a	-0.869^a	0.027
	(0.079)	(0.104)	(0.045)	(0.109)	(0.147)	(0.044)	(0.108)	(0.145)	(0.033)
lnTravelHours				-0.747^a	-1.055^a	-2.432^a	-0.578^a	-0.758^a	-1.806^a
				(0.094)	(0.130)	(0.067)	(0.119)	(0.157)	(0.063)
Modal#hops							-0.065^b	-0.115^a	-0.242^a
							(0.028)	(0.036)	(0.014)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13,143	13,029	13,143	13,143	13,029	13,143	13,143	13,029	13,143
Adjusted R^2	0.80	0.70	0.94	0.80	0.71	0.99	0.80	0.71	0.99

Note: All dependent variables in logs. *lnExpVal* refers to total export value (containerized and non-containerized exports in mio USD, *lnContExp* refers to containerized export quantity in mio tonnes, *lnSend* refers to computed export based on AIS data. Controls include dummy variables for for contiguity, common language, and joint membership in a free trade agreement (FTA). Standard errors clustered by origin and destination country in brackets. ^{a}p<0.01,^{b} p<0.05,^{c} p< 0.1.

7 Concluding remarks

References

Adland, R., H. Jia, and S. P. Strandenes (2017). Are ais-based trade volume estimates reliable? the case of crude oil exports. *Maritime Policy and Management 44*(5), 657–665.

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

Donaldson, D. and A. Storeygard (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives 30*(4), 171–198.

Henderson, V., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *American Economic Review 102*(2), 994–1028.

Appendix

A Constructing the port-to-port data set

Step 1. Cleaning of the AIS data Our point of departure are the AIS data containing all port calls made in 2016 that has been provided to us by Marine Traffic. As explained above, to ensure homeogenity with respect to the ship type and the calculatio of cargo we limit the analysis to container ships.[4] Marine Traffic gives each ship a unique identifier (Ship ID). We start out with close to 5300 ships based on this identifier. We use this to identify each ship's travel history. Based on international ship register a ship also has an IMO number and an MMSI number as well as a Ship Name. Ideally there should be a perfect match between IMO, MMSI, and Ship ID. However, for around 5% of the ships this is not the case. For some we are able to establish a match based on the set of characteristics at hand, while some we have to drop. This leaves us with a data set with 5165 ships.

Step 2. Merge ship specific variables with Clarkson database We merge the cleaned AIS data with information on scantling draught (m) from Clarkson World Fleet register database.

Step 3. Clean ships' travelling patterns In some cases the the arrival port doesn't equal next departure port. These trips are dropped.

Step 4. We observe a ship arriving or departing from a port, we know whether it is In transit, or whether the port is an anchorage. We drop all port observations that when a ship is "In transit", or is passing an anchorage.

B Computing arriving Cargo (*Receive*)

$(1-\alpha_i)$, by definition, observes

$$1-\alpha_i = \frac{Receive_i}{In_i} \tag{27}$$

We observe In_ibut not $Receive_i$ and need to calculate the latter to compute $(1-\alpha_i)$. Here we present alternative estimates of $Receive_i$ and thus $(1-\alpha_i)$, while in the analysis we employ alternative $(1a)$.

[4] Based on the ship categories used by data supplier, Marine Traffic, we use the ships categorised as "container ship" and "Cargo/containership".

1. base estimates of the numerator on ships where $Receive_i > 0$ must be true, i.e. that are leaving the port with departure draught < arrival draught
   (a) lower bound: $(1 - \alpha_i)$ = (effective dwt based on the difference btw arr and dep draught) /total cargo of all incoming ships)
      i. underestimate if: ships with $Receive_i > 0$ but departure draught > arrival draught
         A. underestimate if cargo = gross off load − net off load > 0 for ships with departure draught < arrival draught
      ii. $(1-\alpha_i)$ = (effective dwt based on the arr draught) /total cargo of all incoming ships)
         A. underestimate if: ships with $Receive_i > 0$ but departure draught > arrival draught
         B. overestimate: if gross off load < arrival draught/cargo (i.e. if ships with departure draught < arrival draught do not off load everything at arrival for ships)
2. base estimates of the numerator on ships where that (I) are leaving the port with departure draught < arrival draught without having been in anchorage in between arrival and departure of the port, and (II) all ships that are leaving the port and have been in anchorage after arrival at the port
   (a) $(1 - \alpha_i)$ = ((effective dwt based on the difference btw arr and dep draught of group I) + (effective dwt based on the difference in draught when they arrive at the port and leave the port for anchorage for group II)) /(total cargo of all incoming ships)
3. upper bound: $(1 - \alpha_i)$ = (effective dwt based on the arr draught) /total cargo of all incoming ships) =1
4. use all firms, but let numerator ($Receive_i$) estimates differ depending on whether the ship is leaving the port with departure draught$\gtrless$ arrival draught
   (a) if departure draught < arrival draught:
      i. ν_i= (effective dwt based on the difference btw dep and arr draught) /total cargo of all outgoing ships)
      ii. if departure draught > arrival draught

A. ν_i= (effective dwt based on the arr draught) /total cargo of all outgoing ships)

iii. joining (a) and (b) gives

A. ν_i= ((effective dwt based on the difference btw dep and arr draught for all ships with d<a)+(effective dwt based on the arr draught for all ships with d>a)) /total cargo of all outgoing ships)

C Trade Data and Construction of Trade Variables

C.0.5 Boxes traded (tonnes shipped in containers) from COMTRADE/BACI data

Tonnes shipped in containers are computed based on bilateral product-level trade flows as reported in the *COMTRADE/BACI* database and estimated product-country-pair-specific container shares based on U.S. customs data.

BACI data. We use the product-level trade flows from the BACI database for 2016 which records trade flows for 219 countries and 5199 products classified at the six-digit level of the HS2012 classification. The database records 7,596,147 positive trade values at the product-country-pair level. For 7,512,408 (98.9%) of these flows the weight traded in tonnes is also known. The total trade value corresponding to the observations with missing weights amounts to 2.6 % of global trade value. Among the 219 countries are two country groups: Belgium and Luxembourg are aggregated and so are the member countries of the South African Customs Union (South Africa, Swaziland, Namibia, Lesotho, and Botswana).

COMTRADE data. The COMTRADE database for 2016[5] contains imports and exports reported by 148 countries with respect to a maximum number of 233 countries. We make two sample adjustments: We drop trade flows which are not reported in HS2012 (corresponding to .8% of global import value) and we merge Belgium and Luxembourg as well as the countries of the South African Customs Union to achieve comparability with the BACI database. The resulting dataset has 11,749,437 observations. For about 59% (57%) of all trade flows in terms of value (weight) we observe the "mirrowed" flow, i.e. the same trade flow reported by both the importing and exporting party. Table 7 summarizes the datasets.

U.S. customs data. Container shares by product and country pair are approximated using U.S. customs data provided by the *Foreign Trade Division* of the *US Census Bureau*

[5] downloaded from https://comtrade.un.org/api/get/bulk/C/A/2016/ALL/HS on June 6th 2018

Table 7: Comtrade and BACI data overview

	all	HS2012 all	adjusted countries all	imp val	exp val	imp qty	exp qty	mirrow flows val	qty
COMTRADE									
# obs (in mio.)	12.25	11.95	11.75	7.60	7.19	7.60	6.93	4.47	4.31
# reporters	148	134	131	131	131	131	131	131	131
# partners	233	233	228	228	228	228	228	131	131
# 6-digit products	5774	5204	5204	5204	5204	5204	5204	5201	5195
BACI									
# obs (in mio.)			7.60	7.60		7.52			
# importers			219	219		219			
# exporters			219	219		219			
# 6-digit products			5199	5199		5198			

Column header "HS2012" denotes the subset of observations reported in HS2012, column header "adjusted countries" means that Belgium and Luxembourg and the members of the South African Customs Unions have been aggregated, "mirrow flows" refer to observations of trade flows which are reported both by the importing and the exporting party.

and made available through Peter Schott's trade data website[6]. This dataset contains the universe of annual U.S. import and export values and quantities (units and kg) differentiated by partner country, product, transport mode (air, vessel, containterized vessel, other), and district of entry (roughly 50 U.S. districts). Exports are further differentiated by foreign vs domestic origin, imports are differentiated further by district of unlading (which may differ from district of entry), by type of use (U.S. consumption, total), and valuation (fob, cif). Exported and imported quantities in units by product and partner are reported for totals but not by transport mode. For shipments by air, vessel, and container vessel, kilograms are reported. The product classifications used for imports and exports feature 10 digits with 2-,4-, and 6-digit codes corresponding to the *Harmonized System (HS)* classification. Table 8 describes the scope of the dataset.

Table 8: U.S. Customs Data 2012-2015

	Obs$^{a)}$	Mean	Std. Dev.	Min	Max
Exports					
# countries	4	225	.5	224	225
# products	4	5160	.5	5159	5160
# products p. country	899	1194	1196	1	4913
Imports					
# countries	4	223	.8	222	224
# products	4	5150	5	5146	5157
# products p. country	892	595	969	1	4538

a) For # countries, products (# products p. country), Obs equals # years (# years × # countries)

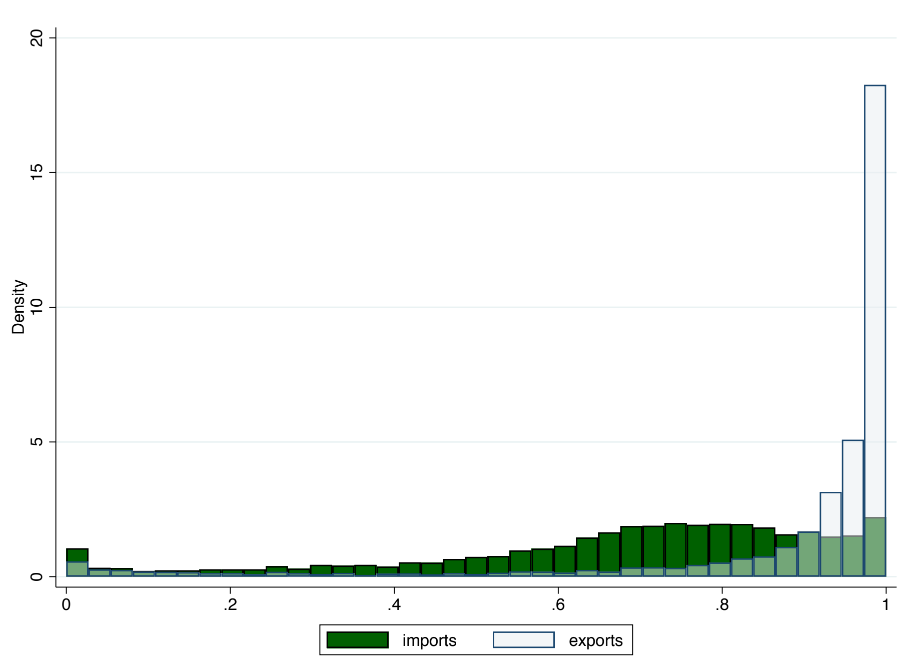
[6]http://faculty.som.yale.edu/peterschott/sub_international.htm.

Approximating container shares. The U.S. data is used to calculate approximate shares of containerized trade between any two countries in the following way: First, we compute by product and partner country the *share of containerized trade in total trade by vessel,* separately for import and export flows

$$s^C_{exp,p,i} = \frac{C_{exp,p,i,USA}}{V_{exp,p,i,USA}} \qquad \text{and} \qquad s^C_{imp,p,i} = \frac{C_{imp,p,i,USA}}{V_{imp,p,i,USA}}$$

where $C_{exp,p,i,USA}$ ($V_{exp,p,i,USA}$) denotes the total weight of product p exported from country i to the U.S. in a container vessel (any type of vessel) as observed in the U.S. customs data for a given period of time. Analogously for imports. Presuming that the U.S. has fully implemented container technology on the import and export side, these shares are informative about the partner country's degree of container utilization for trade of a particular product.

Figure 19: Shares of container trade in U.S. imports and exports: Distribution across products



Note: The figure shows the distribution of the share of containerized trade (in kg) in total trade by vessel across products for U.S. exports and imports. Shares are weighted averages across trade partners and years 2012-2015. Source: U.S. Census Bureau, FTD

Figure 19 shows the distribution of $s^C_{exp,p,i}$ and $s^C_{imp,p,i}$ (weighted averages over all partners) across products. Note that these shares compared containerized vessel trade to overall vessel trade. Products which are never shipped by vessel are thus excluded (but enter the analysis again below). There is a stark difference between exports and imports. On the export side, the vast majority of products has container shares close to one. Whereas on the import side, shares are more evenly distributed between zero and one. For both trade directions there is mass point at zero indicating that there is a group of products which are never transported in a container vessel. Table 10 lists all products for which the container shares $s^C_{exp,p,i}$ and $s^C_{imp,p,i}$ (weighted averages over all partners) are both below 1%. Most prominently the list, contains live animals and chemicals. Table 9 shows container shares and the corresponding vessel trade and total trade for all HS 2-digit categories. Besides the aformentioned product categories, we find small container shares for commodities (grains, cereals, seeds, ores, mineral oil, wood, and iron and steel), as well as for large and bulky items such as ships and other vehicles. Most other product categories feature container shares above or close to 90%.

Second, we compute by product and geographical characteristics of partner countries (overseas, land-connected to the U.S) the *share of trade by vessel in total trade* according to

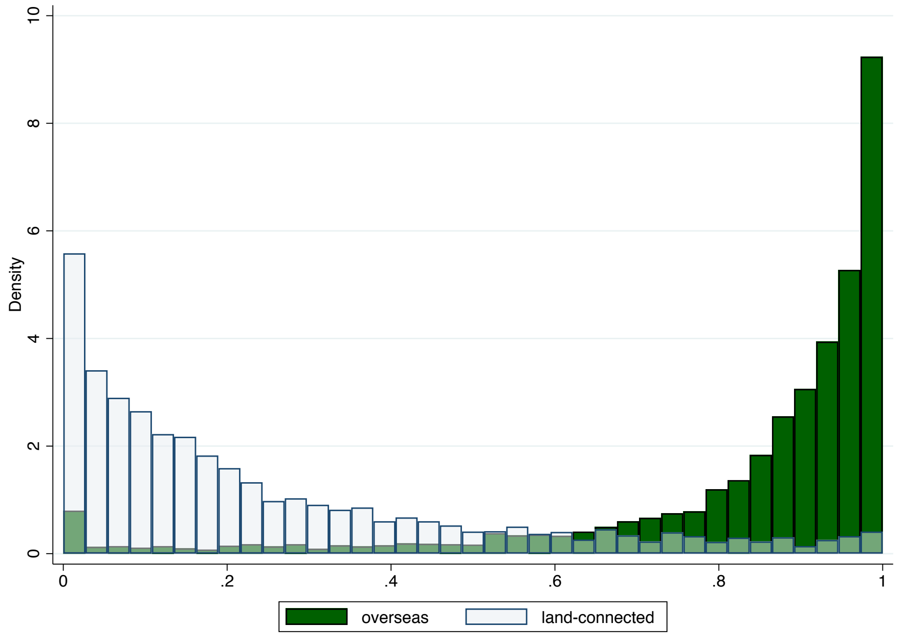
$$
\begin{aligned}
s^V_{p,Overseas} &= \frac{\sum_{j \in J^O}(V_{exp,j,USA} + V_{imp,j,USA})}{\sum_{j \in J^O}(T_{exp,j,USA} + T_{imp,j,USA})} \quad \text{and} \\
s^V_{p,LandConnected} &= \frac{\sum_{j \in J^L}(V_{exp,j,USA} + V_{imp,j,USA})}{\sum_{j \in J^L}(T_{exp,j,USA} + T_{imp,j,USA})}
\end{aligned}
$$

where $T_{exp,j,USA}$ ($T_{imp,j,USA}$) denotes the total weight of product p exported from (imported by) country j to (from) the U.S..[7] $J^L(J^O)$ denotes the set of countries which are (are not) connected to the U.S. via land. These differentiated shares reflect the fact that shipping by vessel is a less likely transport mode if a product is more suitable for transport by air, or if the partner country is on the same continent thus enabling transport by truck or freight train. Figure 20 shows the distribution of $s^V_{p,Overseas}$ and $s^V_{p,LandConnected}$ across products, conveying a strong difference between trade with overseas countries, which is dominated by vessel shipping, and land-connected countries, for which vessel shares are more often small or zero.

Table 11 lists all HS 6-digit products which are never traded by vessel according to the U.S. data. The list includes live animals, rare-earth-metals, chemicals, diamonds, and various kinds of ships. Table 9 shows that in terms of total trade volume (import + exports) in USD, mineral fuels and oil dominate vessel trade with the U.S., followed by nuclear machinery, vehicles, and electrical machinery.

[7] In contrast to trade values, trade flows in terms of weight are only available for air and vessel shipments, but not for other transport modes. Assuming that f.o.b. prices do not vary across transport modes, we compute $T_{exp,j,USA}$ and $T_{imp,j,USA}$ by scaling the sum of kgs shipped by air and vessel with the inverse of the share of trade value shipped air and vessel in total trade value.

Figure 20: Shares of vessel trade in U.S. total trade: Distribution across products



Note: The figure shows the distribution of the share of vessel trade (in kg) in total trade by vessel across products for U.S. trade with overseas and land-connected countries. Shares are weighted averages across trade partners and years 2012-2015. Source: U.S. Census Bureau, FTD

Finally, we use the importer and exporter-specific container shares and the vessel shares by geographic characteristic to approximate container trade between any two countries that do not have a land connection according to

$$C_{exp,p,i,j} = \min[s^C_{exp,p,i}, s^C_{imp,p,j}] * s^V_{p,overseas} * T_{exp,i,j},$$

where $T_{exp,i,j}$ is the export flow (in terms of weight) from i to j as observed in the COMTRADE/BACI data. For country pairs i and j which are connected by land, we use $s^V_{p,LandConnected}$ in place of $s^V_{p,Overseas}$.[8] As regards the degree of containerization, we have a measure for both the importer and the exporter. We presume that the extent to which trade between them is containerized is dominated by the partner with the lower containerization and hence use the minimum of the two shares in a relationship.

[8]More precisely, we define countries as connected by land if they belong to the same continent and neither of them is an island.

Table 9: U.S. container shares and trade volumns by HS2012 2-digit product group

HS2012 chapter	HS2012 2-digit	Description	Container share	Vessel trade mio. USD	Total trade mio. USD
I	1	LIVE ANIMALS	0.19	171	3961
I	2	MEAT AND EDIBLE MEAT OFFAL	0.75	15037	23206
I	3	FISH AND CRUSTACEANS, MOLLUSCS AND OTHER AQUATIC I	0.67	12732	19036
I	4	DAIRY PRODUCE; BIRDS' EGGS; NATURAL HONEY; EDIBLE	0.81	5261	7502
I	5	PRODUCTS OF ANIMAL ORIGIN, NOT ELSEWHERE SPECIFIED	0.92	1307	2031
II	6	LIVE TREES AND OTHER PLANTS; BULBS, ROOTS AND THE	0.94	327	2079
II	7	EDIBLE VEGETABLES AND CERTAIN ROOTS AND TUBERS	0.83	2560	11746
II	8	EDIBLE FRUIT AND NUTS; PEEL OF CITRUS FRUIT OR MEL	0.73	16583	24911
II	9	COFFEE, TEA, MATE AND SPICES	0.99	7186	8657
II	10	CEREALS	0.05	19464	23867
II	11	PRODUCTS OF THE MILLING INDUSTRY; MALT; STARCHES;	0.86	981	2360
II	12	OIL SEEDS AND OLEAGINOUS FRUITS; MISCELLANEOUS GRA	0.12	25707	29785
II	13	LAC; GUMS, RESINS AND OTHER VEGETABLE SAPS AND EXT	0.97	2667	3296
II	14	VEGETABLE PLAITING MATERIALS; VEGETABLE PRODUCTS N	0.97	81	128
III	15	ANIMAL OR VEGETABLE FATS AND OILS AND THEIR CLEAVA	0.28	6105	9482
IV	16	PREPARATIONS OF MEAT, OF FISH OR OF CRUSTACEANS, M	0.95	4846	6924
IV	17	SUGARS AND SUGAR CONFECTIONERY	0.32	3215	6289
IV	18	COCOA AND COCOA PREPARATIONS	0.84	3495	6154
IV	19	PREPARATIONS OF CEREALS, FLOUR, STARCH OR MILK; PA	0.91	3343	9690
IV	20	PREPARATIONS OF VEGETABLES, FRUIT, NUTS OR OTHER P	0.86	7593	11833
IV	21	MISCELLANEOUS EDIBLE PREPARATIONS	0.88	6147	11949
IV	22	BEVERAGES, SPIRITS AND VINEGAR	0.77	20305	27130
IV	23	RESIDUES AND WASTE FROM THE FOOD INDUSTRIES; PREPA	0.36	9165	13975
IV	24	TOBACCO AND MANUFACTURED TOBACCO SUBSTITUTES	0.69	3088	3593
V	25	SALT; SULPHUR; EARTHS AND STONE; PLASTERING MATERI	0.13	3959	5583
V	26	ORES, SLAG AND ASH	0.04	7310	10654
V	27	MINERAL FUELS, MINERAL OILS AND PRODUCTS OF THEIR	0.02	367351	470725
VI	28	INORGANIC CHEMICALS; ORGANIC OR INORGANIC COMPOUND	0.22	17478	26496
VI	29	ORGANIC CHEMICALS	0.28	49983	94452
VI	30	PHARMACEUTICAL PRODUCTS	0.88	22158	111663
VI	31	FERTILISERS	0.09	7439	12264
VI	32	TANNING OR DYEING EXTRACTS; TANNINS AND THEIR DERI	0.90	5752	11414
VI	33	ESSENTIAL OILS AND RESINOIDS; PERFUMERY, COSMETIC	0.85	12307	21164
VI	34	SOAP, ORGANIC SURFACE-ACTIVE AGENTS, WASHING PREPA	0.87	5569	9952
VI	35	ALBUMINOIDAL SUBSTANCES; MODIFIED STARCHES; GLUES;	0.86	3366	5752
VI	36	EXPLOSIVES; PYROTECHNIC PRODUCTS; MATCHES; PYROPHO	0.98	591	1497
VI	37	PHOTOGRAPHIC OR CINEMATOGRAPHIC GOODS	0.88	2328	3889
VI	38	MISCELLANEOUS CHEMICAL PRODUCTS	0.61	18045	37339
VII	39	PLASTICS AND ARTICLES THEREOF	0.91	52168	103583
VII	40	RUBBER AND ARTICLES THEREOF	0.94	26256	40805
VIII	41	RAW HIDES AND SKINS (OTHER THAN FURSKINS) AND LEAT	0.97	3201	4145
VIII	42	ARTICLES OF LEATHER; SADDLERY AND HARNESS; TRAVEL	0.97	9656	13893
VIII	43	FURSKINS AND ARTIFICIAL FUR; MANUFACTURES THEREOF	0.92	79	1028
IX	44	WOOD AND ARTICLES OF WOOD; WOOD CHARCOAL	0.35	12874	23995
IX	45	CORK AND ARTICLES OF CORK	0.98	231	280
IX	46	MANUFACTURES OF STRAW, OF ESPARTO OR OF OTHER PLAI	0.95	444	524
X	47	PULP OF WOOD OR OF OTHER FIBROUS CELLULOSIC MATERI	0.52	8991	12170
X	48	PAPER AND PAPERBOARD; ARTICLES OF PAPER PULP, OF P	0.75	14173	31204

continues on next page

HS2012 chapter	HS2012 2-digit	Description	Container share	Vessel trade mio. USD	Total trade mio. USD
continuing from previous page					
X	49	PRINTED BOOKS, NEWSPAPERS, PICTURES AND OTHER PROD	0.94	3366	9152
XI	50	SILK	0.77	33	148
XI	51	WOOL, FINE OR COARSE ANIMAL HAIR; HORSEHAIR YARN A	0.90	163	378
XI	52	COTTON	0.90	6966	8107
XI	53	OTHER VEGETABLE TEXTILE FIBRES; PAPER YARN AND WOV	0.98	144	232
XI	54	MAN-MADE FILAMENTS; STRIP AND THE LIKE OF MAN-MADE	0.94	2365	4003
XI	55	MAN-MADE STAPLE FIBRES	0.89	3442	4389
XI	56	WADDING, FELT AND NONWOVENS; SPECIAL YARNS; TWINE,	0.93	2556	4195
XI	57	CARPETS AND OTHER TEXTILE FLOOR COVERINGS	0.95	2167	3338
XI	58	SPECIAL WOVEN FABRICS; TUFTED TEXTILE FABRICS; LAC	0.81	621	1196
XI	59	IMPREGNATED, COATED, COVERED OR LAMINATED TEXTILE	0.93	1826	4278
XI	60	KNITTED OR CROCHETED FABRICS	0.82	1445	2021
XI	61	ARTICLES OF APPAREL AND CLOTHING ACCESSORIES, KNIT	0.98	37546	45453
XI	62	ARTICLES OF APPAREL AND CLOTHING ACCESSORIES, NOT	0.97	28392	39397
XI	63	OTHER MADE-UP TEXTILE ARTICLES; SETS; WORN CLOTHIN	0.92	12199	14796
XII	64	FOOTWEAR, GAITERS AND THE LIKE; PARTS OF SUCH ARTI	0.99	22394	26411
XII	65	HEADGEAR AND PARTS THEREOF	0.96	1790	2297
XII	66	UMBRELLAS, SUN UMBRELLAS, WALKING STICKS, SEAT-STI	0.98	521	561
XII	67	PREPARED FEATHERS AND DOWN AND ARTICLES MADE OF FE	0.96	1574	1939
XIII	68	ARTICLES OF STONE, PLASTER, CEMENT, ASBESTOS, MICA	0.94	5700	9869
XIII	69	CERAMIC PRODUCTS	0.93	4579	7150
XIII	70	GLASS AND GLASSWARE	0.96	5785	11934
XIV	71	NATURAL OR CULTURED PEARLS, PRECIOUS OR SEMI-PRECI	0.91	3997	105800
XV	72	IRON AND STEEL	0.18	30845	46898
XV	73	ARTICLES OF IRON OR STEEL	0.55	32143	55757
XV	74	COPPER AND ARTICLES THEREOF	0.72	10125	17967
XV	75	NICKEL AND ARTICLES THEREOF	0.91	2786	5096
XV	76	ALUMINIUM AND ARTICLES THEREOF	0.77	13394	28134
XV	78	LEAD AND ARTICLES THEREOF	0.63	454	1206
XV	79	ZINC AND ARTICLES THEREOF	0.82	587	2039
XV	80	TIN AND ARTICLES THEREOF	0.94	818	966
XV	81	OTHER BASE METALS; CERMETS; ARTICLES THEREOF	0.89	2816	5165
XV	82	TOOLS, IMPLEMENTS, CUTLERY, SPOONS AND FORKS, OF B	0.94	6315	13528
XV	83	MISCELLANEOUS ARTICLES OF BASE METAL	0.94	7295	14326
XVI	84	NUCLEAR REACTORS, BOILERS, MACHINERY AND MECHANICA	0.82	192628	484838
XVI	85	ELECTRICAL MACHINERY AND EQUIPMENT AND PARTS THERE	0.92	107450	414530
XVII	86	RAILWAY OR TRAMWAY LOCOMOTIVES, ROLLING STOCK AND	0.75	2792	5691
XVII	87	VEHICLES OTHER THAN RAILWAY OR TRAMWAY ROLLING STO	0.43	187055	379715
XVII	88	AIRCRAFT, SPACECRAFT, AND PARTS THEREOF	0.80	9643	140089
XVII	89	SHIPS, BOATS AND FLOATING STRUCTURES	0.23	3586	4848
XVIII	90	OPTICAL, PHOTOGRAPHIC, CINEMATOGRAPHIC, MEASURING,	0.87	23442	142397
XVIII	91	CLOCKS AND WATCHES AND PARTS THEREOF	0.96	773	5639
XVIII	92	MUSICAL INSTRUMENTS; PARTS AND ACCESSORIES OF SUCH	0.96	1243	1957
XIX	93	ARMS AND AMMUNITION; PARTS AND ACCESSORIES THEREOF	0.93	3627	7702
XX	94	FURNITURE; BEDDING, MATTRESSES, MATTRESS SUPPORTS,	0.98	35747	59438
XX	95	TOYS, GAMES AND SPORTS REQUISITES; PARTS AND ACCES	0.98	23769	32037
XX	96	MISCELLANEOUS MANUFACTURED ARTICLES	0.95	4118	7412
XXI	97	WORKS OF ART, COLLECTORS' PIECES AND ANTIQUES	0.94	676	13768

The table lists shares of containerized trade in total vessel trade by HS2012 2-digit product groups (weighted averages across products, countries, imports and exports, and years 2012-2015) as well as volumns traded (imports+exports) by vessel in total (in mio. USD, averages across years 2012-2015)

Table 10: HS2012 6-digit products with zero container shares but positive vessel trade shares

HS2012	Description
10410	Live sheep
10512	Live domestic turkeys, weighing <= 185 g
10613	Live camels and other camelids [Camelidae]
10614	Live rabbits and hares
20751	Fresh or chilled domestic geese, not cut in pieces
30273	Fresh or chilled carp "Cyprinus carpio, Carassius carassius, Ctenopharyngodon idellus, Hypophthalmic
261220	Thorium ores and concentrates
290382	Aldrin (ISO), chlordane (ISO) and heptachlor (ISO)
290892	4,6-Dinitro-o-cresol (DNOC (ISO)) and its salts
291040	Dieldrin "ISO" "INN"
293341	Levorphanol "INN" and its salts
293750	Prostaglandins, thromboxanes and leukotrienes, their derivatives and structural analogues, used prim
293951	Fenetylline "INN" and its salts
293963	Lysergic acid and its salts
382473	Mixtures containing hydrobromofluorocarbons "HBFCs"
411330	Leather further prepared after tanning or crusting "incl. parchment-dressed leather", of reptiles„
880240	Aeroplanes and other powered aircraft of an of an unladen weight > 15.000 kg (excl. helicopters and
911110	Cases for wrist-watches, pocket-watches and other watches of heading 9101 or 9102, of precious metal

The table lists HS2012 6-digit products for which the share of containerized vessel trade in total vessel trade (weighted average across imports, exports, years 2012-2015) is below 1%

Table 11: HS2012 6-digit products with zero vessel shares

HS2012	Description
10391	Live pure-bred swine, weighing < 50 kg (excl. pure-bred for breeding)
10420	Live goats
10514	Live domestic geese, weighing <= 185 g
10515	Live domestic guinea fowls, weighing <= 185 g
10611	Live primates
10612	Live whales, dolphins and porpoises (mammals of the order Cetacea); manatees and dugongs (mammals of
10620	Live reptiles "e.g. snakes, turtles, alligators, caymans, iguanas, gavials and lizards"
10631	Live birds of prey
10632	Live psittaciformes "incl. parrots, parrakeets, macaws and cockatoos"
10633	Live ostriches, and emus [Dromaius novaehollandiae]
10639	Live birds (excl. birds of prey, psittaciformes, parrots, parrakeets, macaws, cockatoos, ostriches a
30191	Live trout "Salmo trutta, Oncorhynchus mykiss, Oncorhynchus clarki, Oncorhynchus aguabonita, Oncorhy
30195	Live southern bluefin tunas "Thunnus maccoyii"
30222	Fresh or chilled plaice "Pleuronectes platessa"
30256	Fresh or chilled blue whiting "Micromesistius poutassou, Micromesistius australis"
121140	Poppy straw, fresh or dried, whether or not cut, crushed or powdered
252410	Crocidolite asbestos (excl. products made from crocidolite)
292146	Amfetamine "INN", benzfetamine "INN", dexamfetamine "INN", etilamfetamine "INN", fencamfamine "INN",
292424	Ethinamate "INN"
293050	Captafol "ISO" and methamidophos "ISO"
293355	Loprazolam "INN", mecloqualone "INN", methaqualone "INN" and zipeprol "INN", and salts thereof
293943	Cathine "INN" and its salts
293962	Ergotamine "INN" and its salts
293991	Cocaine, ecgonine, levometamfetamine, metamfetamine "INN", metamfetamine racemate, and salts, esters
430211	Tanned or dressed furskins of mink, whole, with or without heads, tails or paws, not assembled
710210	Diamonds, unsorted
710221	Industrial diamonds unworked or simply sawn, cleaved or bruted
710231	Non-industrial diamonds unworked or simply sawn, cleaved or bruted (excl. industrial diamonds)
710239	Diamonds, worked, but not mounted or set (excl. industrial diamonds)
711031	Rhodium, unwrought or in powder form
711041	Iridium, osmium and ruthenium, unwrought or in powder form
811251	Unwrought thallium; thallium powders
890110	Cruise ships, excursion boats and similar vessels principally designed for the transport of persons;
890120	Tankers
890130	Refrigerated vessels (excl. tankers)

The table lists HS2012 6-digit products for which the share of vessel trade in total trade (weighted average across imports, exports, years 2012-2015) is zero for both countries with and without a land-connection to the U.S..

Implementation. For $T_{exp,i,j}$ we use BACI/COMTRADE data for 2016. To compute $s^C_{exp,p,i}$, $s^C_{imp,p,i}$, $s^V_{p,Overseas}$, and $s^V_{p,LandConnected}$, we aggregate the use US customs data across U.S. districts, sum up to the six-digit level, and then compute weighted averages across the years 2012 and 2015.[9] Since the U.S. does not trade all 6-digit products with all countries, we compute three more aggregate versions of $s^C_{exp,p,i}$, $s^C_{imp,p,i}$; one at the 4-digit level and weighted averages across partners at the 6-digit and 4-digit level. These are used to substitute missing values at the most detailed level if necessary. Likewise, we compute $s^V_{p,Overseas}$ and $s^V_{p,LandConnected}$ at the four-digit level to replace missings. Table 12 shows exemplary for the BACI data the percentages of trade flows covered by the subsequent replacements with shares computed at less disaggregated levels.

Table 12: Merge BACI with vessel and container shares

	$s^V_{LandConnected}$	$s^V_{Overseas}$	s^C_{exp}	s^C_{imp}
HS 6-digit	97	96.9	82.5	78.1
HS 6-digit weight. avg.			13.6	18.1
HS 4-digit	3	3.1	2.7	2.9
HS 4-digit weight. avg.			1	.8
Zero vessel trade			.2	.2

The table shows percentages of total trade flows covered by most (HS 6-digit by partner) to least (HS 4-digit weight. avg across partners) disaggregated container and vessel shares derived from U.S. customs data.

Table 13 summarizes the final BACI-data based dataset on global vessel and container vessel trade by HS chapter. The last row shows the totals across chapters, revealing that 71 of global trade in 2016 (in terms of weight) takes place through vessel shipments, and 11% of global trade used container vessels. Obviously food, metal, minerals and metals are mainly carried by respectively bulk and tanker ships. We also know that transport vehicles, in particular cars are typically carried on vessels designed particularly for this purpose. However, for other manufactured goods we see that approximately 50% of global trade relies on container vessels.

[9] We use multiple years to cover as many trade relationships and products as possible. 2012 is the first year the HS2012 classification was applied, 2015 is the most recent year for which the U.S. customs data is available to us.

Table 13: Global vessel and container trade by HS chapter

HS 2012	Description	Imp mio USD	Imp qty mio tonnes	Share in product total Vessel in %	Share in product total Cont. in %	Share in global trade Vessel in %	Share in global trade Cont. in %	Share in global trade Imp qty in %
I	LIVE ANIMALS; ANIMAL PRODUCTS	315671	122	52	35	3	1	1
II	VEGETABLE PRODUCTS	452433	853	75	17	10	6	6
III	ANIMAL OR VEGETABLE FATS AND OILS AND THEIR CLEAVAGE PRODUCTS; PREPARED EDIBLE FATS; ANIMAL OR VEGETABLE WAXES	88110	101	77	17	1	1	1
IV	PREPARED FOODSTUFFS; BEVERAGES, SPIRITS AND VINEGAR; TOBACCO AND MANUFACTURED TOBACCO SUBSTITUTES	534283	1182	31	15	13	4	9
V	MINERAL PRODUCTS	1615411	8370	83	3	20	70	61
VI	PRODUCTS OF THE CHEMICAL OR ALLIED INDUSTRIES	1447101	745	67	18	9	5	5
VII	PLASTICS AND ARTICLES THEREOF; RUBBER AND ARTICLES THEREOF	675227	282	55	46	9	2	2
VIII	RAW HIDES AND SKINS, LEATHER, FURSKINS AND ARTICLES THEREOF; SADDLERY AND HARNESS; TRAVEL GOODS, HANDBAGS AND SIMILAR CONTAINERS; ARTICLES OF ANIMAL GUT	94810	9	57	49	0	0	0
IX	WOOD AND ARTICLES OF WOOD; WOOD CHARCOAL; CORK AND ARTICLES OF CORK; MANUFACTURES OF STRAW, OF ESPARTO OR OF OTHER PLAITING MATERIALS; BASKETWARE AND WICKERWORK	127038	355	50	22	5	2	3
X	PULP OF WOOD OR OF OTHER FIBROUS CELLULOSIC MATERIAL; RECOVERED (WASTE AND SCRAP) PAPER OR PAPERBOARD; PAPER AND PAPERBOARD AND ARTICLES THEREOF	229350	268	56	28	5	2	2
XI	TEXTILES AND TEXTILE ARTICLES	695071	86	67	54	3	1	1
XII	FOOTWEAR, HEADGEAR, UMBRELLAS, SUN UMBRELLAS, WALKING STICKS, SEAT-STICKS, WHIPS, RIDING-CROPS AND PARTS THEREOF; PREPARED FEATHERS AND ARTICLES MADE THEREWITH; ARTIFICIAL FLOWERS; ARTICLES OF HUMAN HAIR	147521	8	65	51	0	0	0
XIII	ARTICLES OF STONE, PLASTER, CEMENT, ASBESTOS, MICA OR SIMILAR MATERIALS; CERAMIC PRODUCTS; GLASS AND GLASSWARE	153960	182	48	38	5	1	1
XIV	NATURAL OR CULTURED PEARLS, PRECIOUS OR SEMI-PRECIOUS STONES, PRECIOUS METALS, METALS CLAD WITH PRECIOUS METAL, AND ARTICLES THEREOF; IMITATION JEWELLERY; COIN	626958	1	52	41	0	0	0
XV	BASE METALS AND ARTICLES OF BASE METAL	983979	761	56	15	8	4	6
XVI	MACHINERY AND MECHANICAL APPLIANCES; ELECTRICAL EQUIPMENT; PARTS THEREOF; SOUND RECORDERS AND REPRODUCERS, TELEVISION IMAGE AND SOUND RECORDERS AND REPRODUCERS, AND PARTS AND ACCESSORIES OF SUCH ARTICLES	3825776	192	48	32	4	1	1
XVII	VEHICLES, AIRCRAFT, VESSELS AND ASSOCIATED TRANSPORT EQUIPMENT	1753846	179	49	15	2	1	1
XVIII	OPTICAL, PHOTOGRAPHIC, CINEMATOGRAPHIC, MEASURING, CHECKING, PRECISION, MEDICAL OR SURGICAL INSTRUMENTS AND APPARATUS; CLOCKS AND WATCHES; MUSICAL INSTRUMENTS; PARTS AND ACCESSORIES THEREOF	539404	8	50	36	0	0	0
XIX	ARMS AND AMMUNITION; PARTS AND ACCESSORIES THEREOF	12541	0	61	46	0	0	0
XX	MISCELLANEOUS MANUFACTURED ARTICLES	357571	60	49	41	2	0	0
XXI	WORKS OF ART, COLLECTORS' PIECES AND ANTIQUES	21354	0	66	51	0	0	0
I-XXI	**TOTAL**	1.47e+07	13765	71	11	100	100	100