WP/20/57

IMF Working Paper

World Seaborne Trade in Real Time:
A Proof of Concept for Building AIS-based Nowcasts from Scratch

by Diego A. Cerdeiro, Andras Komaromi, Yang Liu and Mamoon Saeed

INTERNATIONAL MONETARY FUND

 WP/20/57

IMF Working Paper

Institute for Capacity Development
Strategy, Policy, and Review Department

World Seaborne Trade in Real Time: A Proof of Concept for Building AIS-based Nowcasts from Scratch♠

Prepared by Diego A. Cerdeiro, Andras Komaromi, Yang Liu and Mamoon Saeed♣

Authorized for distribution by Valerie Cerra and Martin D. Kaufman

May 2020

Abstract

Maritime data from the Automatic Identification System (AIS) have emerged as a potential source for real time information on trade activity. However, no globally applicable end-to-end solution has been published to transform raw AIS messages into economically meaningful, policy-relevant indicators of international trade. Our paper proposes and tests a set of algorithms to fill this gap. We build indicators of world seaborne trade using raw data from the radio signals that the global vessel fleet emits for navigational safety purposes. We leverage different machine-learning techniques to identify port boundaries, construct port-to-port voyages, and estimate trade volumes at the world, bilateral and within-country levels. Our methodology achieves a good fit with official trade statistics for many countries and for the world in aggregate. We also show the usefulness of our approach for sectoral analyses of crude oil trade, and for event studies such as Hurricane Maria and the effect of measures taken to contain the spread of the novel coronavirus. Going forward, ongoing refinements of our algorithms, additional data on vessel characteristics, and country-specific knowledge should help improve the performance of our general approach for several country cases.

JEL Classification Numbers: F14, F47, E66.
Keywords: Nowcasting, trade, economic activity, applied machine learning, AIS, big data.
Author's E-Mail Address: dcerdeiro@imf.org, akomaromi@imf.org, yliu10@imf.org, msaeed@imf.org

♠ We are very grateful to Martin Kaufman, Tam Bayoumi, Jörg Decressin, Brad McDonald and Gian Maria Milesi-Ferretti for early discussions that helped shape this project. Special thanks go to Marco Marini for his contributions in the early stages of this project. We also thank Master Mariner Iancu Alter for extremely helpful discussions, Bram Hendricks of the Netherlands Bureau for Economic Policy Analysis for kindly sharing the country-level data used to assess the performance of our indicators, and Raju Huidrom, Diane Kostroch, Ayoub Mharzi, Srobona Mitra, Cyril Rebillard, Daniel Rodriguez, Alberto Sanchez, Jacinta Shirakawa, Alessandra Sozzi, Andrew Tiffin and Philippe Wingender for very helpful comments.
♣ Diego A. Cerdeiro: IMF, Strategy, Policy, and Review Department; Andras Komaromi: IMF, Institute for Capacity Development; Yang Liu: IMF, Information Technology Department; Mamoon Saeed: IMF, Information Technology Department.

Table of Contents

I. INTRODUCTION

World trade is intertwined with overall economic activity. Fluctuations in trade can reflect changes in global growth, or they themselves can spur or hamper growth. Businesses and policy makers therefore devote great efforts to monitoring high frequency trade data from various sources to inform their near-term outlooks and make informed investment decisions or craft policy recommendations. However, most proxies of global trade are available with a one- to three-month lag and some high frequency indicators miss large swaths of world trade. In this paper, we build real-time indicators of world seaborne trade using raw data emitted by the global vessel fleet through their transponders to provide a more immediate picture of global trade flows than is currently available.

Maritime transport is at the core of world trade and it can be tracked in real time. Over 80 percent of global merchandise trade by volume and more than 70 percent of its value is carried by the international shipping industry (UNCTAD, 2017). Cargo ships are equipped with a device that periodically emits a signal (Automatic Identification System message, or AIS) which contains information on the vessel's location, speed, draught, etc. Various private providers use terrestrial receivers and satellites to catch these AIS messages and transform them into structured data feeds. We use different machine-learning techniques to transform these data into estimates of trade volume at the world, bilateral and within-country levels.

Our methodology has good predictive power for many countries. Moreover, in some cases we usually have the highest correlation with a one-month lead. This implies that in some cases our results can be useful not just to nowcast trade but to forecast it as well. It can be instructive to lay out this finding with an example. In countries with high data quality, a publication lag of around two months is normal. This means that at the end of February or early March official monthly customs data report the evolution of trade until the end of the previous year. At that point in time, our methodology allows us to calculate the 3m/3m growth rate from AIS data and – in some cases – this will be a good predictor for the quarter-on-quarter growth rate of actual imports for the first quarter of the year. We infer that this predictive power stems from AIS messages sometimes frontrunning the recording of trade by national customs offices.

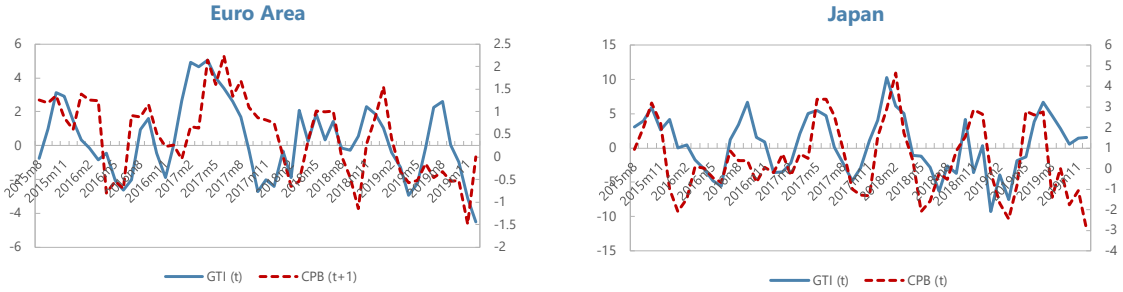
Figure 1 illustrates some of our results for the cases of the Japan and the Euro Area. There, and henceforth, we label our index the Global Trade Intelligence (GTI) index. The correlation between the monthly GTI and official (CPB) *growth rates* is as high as 0.49 for Japan and, and of 0.47 for the Euro Area when official data are measured with a one-month lead.[1] The general approach does not work as well for all countries, either because shares of

[1] Official trade volume data were collected and shared by the Netherlands Bureau for Economic Policy Analysis (CPB). CPB uses the X12 seasonal adjustment method that we also applied to our AIS-based GTI index. Hence, the reported correlations are not coming from predictable seasonal fluctuations. See Section V below for further discussion. Both our estimates and CPB's include intra-Euro Area trade.

trade by sea are low or because the geography or infrastructure of countries' ports is substantially different from the official U.S. data used to train the algorithm. For the world as a whole, our results have a monthly pairwise correlation with official statistics of nearly 0.9 in levels, and around 0.4 in quarter-on-quarter growth rates.

Figure 1. GTI Index and Official Data

3m/3m growth rates, seasonally adjusted



Sources: Authors' calculations, and Netherlands Bureau of Economic Analysis (CPB).

We also show the usefulness of our methodology for sectoral analyses by benchmarking our estimates for the case of oil with available data for crude oil trade. The correlation in levels for global crude oil imports is 0.73 in the raw data, and 0.81 on a 3-month moving average basis. In growth rates, the correlation is as high as 0.47. We close our results section by showing how our methodology can be used in event studies, illustrating this with two prominent cases: the effects of Hurricane Maria in Puerto Rico, and the impact of the novel coronavirus outbreak in China.

We think that our results should be valuable to policy makers, especially in periods where time is of the essence. To the best of our knowledge, the estimates by the CPB are the 'gold standard' monthly trade volume indicators that are representative of world trade. On or around the 23rd day of a given month *t*, CPB will publish its first estimates of trade volumes corresponding to month *t-2*. As far as a policy maker is concerned, this means that the CPB publishes its first estimates for month *t-2* with a 7 to 11-week lag, depending on whether the reference point is the beginning or the end of month *t-2*. Because of the challenges involved in producing deflators, the first estimate for month *t-2* usually undergoes revisions, so that the effective lag is often of 11 to 15 weeks. Our current implementation of the methods presented in this paper produce import estimates with a 5-day lag, and export estimates with a 10-day lag.[2] In cases where our index tracks official data best with a lead, the timeliness advantage is of course further enhanced.

[2] As described in Section V, our volume indicators rely on information on how deep the ship is into the water (the draught). This information is manually entered by crews, in some cases with a lag. Therefore, if we want to estimate the volume loaded/offloaded at a given port, we may need a few more days of data to observe any draught updates. Hence, to be cautious, we censor the last 5 days of our estimates for imports. In the case of exports, we censor the last 10 days because outgoing cargo is only classified as export once it reaches a

In keeping with the spirit of the standard disclaimer for IMF working papers – that they describe research in progress and are published to elicit comments and to encourage debate – we are still working on methodological refinements. We are making our results to date available for two reasons. First, because we think the results presented here are an encouraging first step in using AIS data to monitor world trade activity. And second, because of the enormous need for real-time monitoring tools during the current COVID-19 pandemic. At the end of the paper, we describe specific areas of ongoing improvements and avenues for future research that build on this proof of concept.

To the best of our knowledge, our paper is the first to propose the construction of trade volume indicators relying solely on AIS data and publicly available sources of information. In other words, we back out trade patterns from scratch using (almost) just ships' radio messages. A limited set of papers has sought, with mixed results, a similar objective but using proprietary port-call data provided by private vendors. Since those port call data are proprietary, it is generally unknown how they are constructed and they are not benchmarked against port calls recorded by customs agencies. In contrast, we provide a transparent algorithm to identify port calls using as benchmark four years of daily official vessel entrances recorded by U.S. customs.

First, we build a dataset of potential port calls using historical AIS messages and a spatial clustering algorithm. Since the algorithm uses AIS data as input but is not provided any training sample on where ports are actually located in the world (i.e. it has no 'teacher'), it belongs to the class of *unsupervised* machine learning techniques. Second, we combine this set of potential port calls with official, vessel-level U.S. entrance-record data. Since in this case we do have official data to target, we can resort to *supervised* machine learning techniques. Specifically, we use these official data to train a classifier that returns a prediction of whether any given potential port call is a port call or not. In doing so, our paper is also the first to report the accuracy of our (AIS-derived) port call data with respect to official statistics at the vessel level. Third, we estimate the volume of trade between any two ports and identify import, export, and internal trade volumes. Finally, we benchmark of our results against official trade volume data.

Our paper contributes to an incipient and growing literature that uses AIS data to monitor trade. Perhaps the two most closely related papers are by Adland, Jia and Strandenes (2017), and Arslanalp, Marini and Tumbarello (2019).[3] In one of the earliest contributions, Adland et

different country. Further work may allow us to reduce this lag for the case of exports. *Stricto sensu*, our estimates are therefore in fact *near* real time.

[3] For a general discussion of potential applications of big data to help improve official statistics, see Hammer, Kostroch and Quirós (2017). AIS data in particular have also been used to calibrate theoretical models of trade, providing the ability to perform experiments such as the effect of closing or opening certain routes (Brancaccio et al., forthcoming; Heiland et al., 2019).

al. (2017) combine AIS data with proprietary port-call data with the aim of nowcasting crude oil exports. Arslanalp et al. (2019) also combine AIS with proprietary port call data for Malta and develop indicators to gauge trade activity.

Our paper differs from these contributions in terms of coverage and methodology. While we can break down our indicators at the level of vessel type (crude oil tanker, dry bulk, etc.), our main objective in this proof of concept is to gauge overall trade activity. Similarly, while we illustrate the power of AIS data with some country case studies, our exercise is not tailored to any specific country but aims to cover the entire world. From a methodological point of view, the main difference is that our approach does not rely on any proprietary port-call dataset but builds this information from the AIS data themselves. Furthermore, in the process of training our algorithm we are also the first to assess the accuracy of AIS-derived port-call data at the vessel level.[4] The more general approach and the fact that we only use raw AIS data do not come at the expense of performance when compared to the performance in those narrower studies. For example, Arslanalp et al. (2019) find a correlation of 0.65 when comparing their trade gauge *in levels* with quarterly official statistics for Malta. With our general method that is not tailored to any specific country, we find a monthly correlation in seasonally-adjusted levels of 0.88 for the world as a whole. This is not to say that expert knowledge of specific cases cannot improve our general, global approach to the problem. We believe that this is in fact very much the case for some countries.

The rest of the paper is organized as follows. Section II gives a very brief description of AIS data. Section III outlines our approach to identifying ports from AIS data, and mapping these ports to countries. The results of this section yield a set of polygons; any ship sending an AIS message from within these polygons is considered a potential port call. Section IV lays out our approach for teasing out actual port calls from this set of potential port calls. Section V presents our methodology for estimating the volume of trade between any two ports and how we identify imports and exports. Section VI compares our estimated gauges of trade with official data at the aggregate and sectoral levels, and illustrates the usefulness of our methodology for event studies. Section VII outlines already identified steps to refine the methodology and potential avenues for future research. Section VIII presents concluding remarks.

[4] More specifically, we check whether we can find in our AIS-derived port calls the specific vessels that entered ports at specific dates in official statistics. This is different from the approach performed in the aforementioned papers, where the only measure of fit is the comparison with aggregate time series data on the number of incoming vessels.

II. AIS DATA[5]

AIS stands for Automatic Identification System. International Maritime Organization (IMO) regulations require, since end-2004, AIS transponders to be fitted aboard "[…] all ships of 300 gross tonnage and upwards engaged on international voyages, cargo ships of 500 gross tonnage and upwards not engaged on international voyages and all passenger ships irrespective of size." AIS transponders send radio messages so as to provide a periodic real-time feed to other ships and coastal authorities and thus enhance maritime security. These radio messages can be picked up either by terrestrial stations (if the ship is near a shore) or by satellites.

Besides unique ship identifiers, AIS messages include some data that are automatically generated, and some that are manually entered by ships' crews. Automatically generated data include, for example, the ship's position (latitude and longitude) and its speed. Among the manually entered fields are, for example, the ship's draught (i.e. the vertical distance between the ship's keel and the waterline), the destination and the navigational status (e.g. moored, anchored, under way using engine, etc).

The AIS data we use in this paper were provided by MarineTraffic, and cover the period from January 1, 2015 to April 18, 2020. The dataset includes over one billion messages from over 50,000 distinct ships. While most ships send AIS messages with a frequency of 2-10 seconds, the data we purchased are down-sampled to the hourly frequency.[6]

III. ELICITING PORTS FROM RAW AIS DATA – UNSUPERVISED LEARNING

The first, and arguably main challenge when making sense of AIS data for trade-nowcasting purposes is to identify when a ship's message comes from within the boundaries of a port. The scale of the challenge is given by both the sheer number of ports and the fact that port boundaries might change or new ports might be built over time. The National Geospatial Intelligence Agency's World Port Index (WPI), which compiles characteristics of major ports, identifies 3,669 ports. Even if one were to use domain knowledge to draw port boundaries for all these ports, this would still not address the problem that boundaries might become obsolete over time.

Our approach in this paper is to infer port boundaries, the first building block of our nowcasting indicators, from the AIS data themselves. We use unsupervised machine learning

[5] Our description in this section is deliberately brief. Almost every paper using AIS data has a good description of what AIS messages contain and how they are collected. Because this data source is not yet so widely known, we have included this section so that the paper is self-contained for all readers.

[6] Due technological reasons such as message collision and lower satellite coverage in the deep oceans, the actual coverage is slightly lower (see e.g. Natale et al., 2015). However, we still have about one message per ship every two hours on average.

techniques to identify geographical clusters of vessels that, based on certain AIS features, are presumed to be at a port. The method we use is described in subsection A below, and its implementation is presented in subsection B. Because we are ultimately interested in building trade gauges between ports and countries, we lastly map our ML-identified port boundaries to ports, and hence countries, in the WPI database. This is presented in subsection C.

A. Conceptual approach: Density-based clustering

To identify port boundaries using AIS data, we rely on the fact that loading and offloading requires vessels to stand still and on the assumption that crews will, in general, correctly update the vessels' navigational status when at port. For that reason, our starting point is the set of messages with speed less than 0.5 knots per hour and reporting a navigational status of *moored*. For some vessel types above a certain size that may load and offload without mooring, we also consider messages with speed less than 0.5 knots per hours and reporting a navigational status of *anchored*.[7] Let D denote this set of messages. The question is how to identify port boundaries equipped with the latitude and longitude of messages in D.

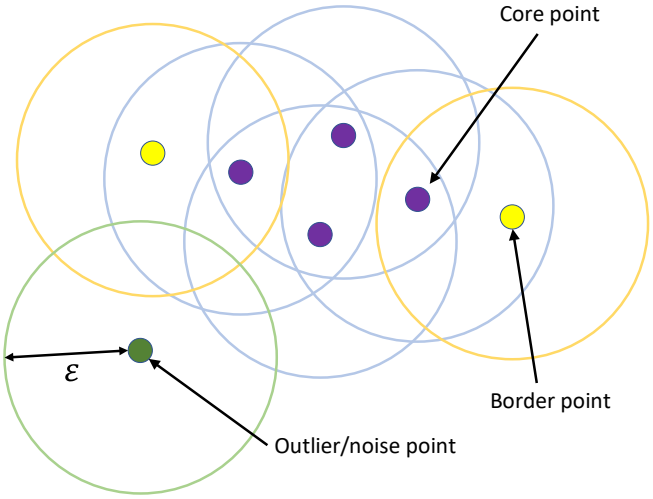
We use the density-based algorithm for discovering clusters in large spatial databases with noise (DBSCAN henceforth; Ester et al., 1996). DBSCAN groups observations that are close to each other into clusters, while points in low-density regions are considered outliers. To do this, it requires two exogenous inputs: a radius ε to define neighborhoods, and a minimum number of points N to define what (given the radius) are areas that are sufficiently dense so as to be considered a cluster. Specifically, for any point p, the ε-neighborhood (ε being the radius) is defined as:

$$N_\varepsilon(p) = \{q \in D \mid dist(p,q) \leq \varepsilon\},$$

where $dist(.,.)$ denotes the geodesic distance. Given a choice of ε, a point p is defined as a core point if at least N points are in its ε-neighborhood. A point q is *directly reachable* from p if p is within q's ε-neighborhood. A point q is *reachable* from p if there is a path $p_1, \ldots, p_n$ $p_1 = p$ and $p_n = q$ such that each p_{i+1} is directly reachable from p_i. All points not reachable from any other points are outliers or noise points. A *cluster* is defined as the set of points reachable from a core point. Figure 2 illustrates what is meant by core, border, and outlier points.

[7] These are bulk carriers width deadweight tonnage (DWT) over 75,000 metric tons, oil/chemical tankers with DWT over 50,000 metric tons, and crude oil tankers with DWT over 100,000 metric tons. In both the mooring and anchored cases, the 0.5-knot threshold is used solely so as to filter out clearly erroneous navigational status reports. Spiliopoulos et al. (2018), for example, use this same threshold.

Figure 2. Core, border and outlier points in DBSCAN



Notes: In this figure, if DBSCAN is set with distance ε and minimum number of points $N = 4$, then it will identify a cluster including all purple (core) dots and yellow (border) dots. The green dot will be identified as noise.

The application of DBSCAN yields a set of *clusters*, but what we need in order to identify port calls are geographical boundaries. For each cluster, we define its associated *polygon* as the convex hull of the cluster.

Before turning to how we operationalize these ideas in our data, it is useful to discuss the advantages and drawbacks of the DBSCAN algorithm. DBSCAN has clear advantages compared to several alternatives. For example, the WPI database that we use to map polygons to ports characterizes each port's location with a single latitude-longitude pair. In principle, one could take port boundaries to be circles of a given radius centered around the latitude and longitude measures in the WPI database. This would imply a one-size-fits-all approach to port boundaries, where all ports have the same shape and size. Evidently, in the real world port boundaries are dramatically different both in shape and in size depending on the geography around them.

DBSCAN is not the only machine-learning method that can be used to identify clusters. K-means clustering is another popular and very efficient method (Lloyd, 1982; Forgy, 1965). Compared to DBSCAN, however, K-means is heavily influenced by outliers, as it forces every point into a cluster. Even more importantly, K-means requires a priori knowledge of the number of clusters K, a very serious limitation in our application. Spatial data analysis methods such as hotspot analysis (Getis, 1992) attempt to generalize Poisson processes to applications in the space domain. The basic idea consists in comparing the concentration of points in space to the expected number given a random distribution of events. The limitation, however, is that these methods would require prior knowledge of the shape of port boundaries. In our application, such boundaries are not readily available.

The ability of DBSCAN to elicit clusters of arbitrary shapes is well suited to our application, insofar as the geography of ports is widely heterogeneous. Furthermore, by not requiring prior knowledge of the number of clusters and being able to tease out outliers, DBSCAN has also clear advantages over methods such as K-means. Yet, DBSCAN does have certain limitations. Chief among them is that it requires two exogenous parameters: the distance ε and the minimum number of points N. In our application, in order not to miss trade at ports that are less-frequently visited, we set these parameters such as to err on the side of having too many polygons. Recent extensions of the algorithm to endogenize these parameters constitute a promising avenue for future refinements of our results.[8]

B. Implementation: Weighted DBSCAN

Not all AIS messages are relevant for the construction of mooring and anchorage clusters. First, we only keep messages where the navigational status is moored and the speed is below 0.5 knots per hour. To identify port calls of certain ship types that may engage in trade while anchored (see footnote 6 above) we also use messages where the navigational status is anchored and the speed is below 0.5 knots per hour.[9] This leaves us with 189.0 million observations, out of the 854.7 million observations of the original 4-year dataset. Second, to reduce computational time, we round up the latitude and longitude fields. For latitudes between 0 and 45 degrees, we round to the nearest fourth decimal place. This corresponds to around 11 meters at the equator and around 8 meters at 45-degree latitude. For latitudes greater than 45 degrees, we round to the nearest one-fifth of the fourth decimal place (i.e. in 0.0002 increments). Finally, we collapse the dataset by rounded-up latitude and longitude, creating a new variable (*position_weight*) that counts the number of AIS messages corresponding to each rounded-up position.

Our weighted DBSCAN algorithm requires two parameters: a radius ε, and a minimum number of points N. We set $\varepsilon = 2{,}000$ meters and $N = 1{,}000$.[10] Our prior is that within this distance ε we should be able to observe sufficiently many messages to justify the point as a core point. In our application, the minimum number of neighbors to define a core point depends on the timespan of the dataset and the sampling frequency. Loosely speaking, we set

[8] The original authors of DBSCAN have proposed a new method, called OPTICS, which is less sensitive to input parameters. Given the size of our dataset, we adopted a weighted version of DBSCAN where each point on the grid in weighted by the number of messages in that point. A similar extension of OPTICS could be used to refine our results.

[9] To avoid overlap between our resulting mooring and anchorage polygons, we subtract from the area of anchorage polygons the area of mooring polygons. This implies that while our mooring polygons are convex sets by construction, the anchorage polygons need not be.

[10] Across the paper, distances on the map (including between message locations, and also between centroids and ports) were computed using the haversine formula.

the minimum number of observations N as equivalent to a single-berth terminal having to be occupied at least about five percent of the time to be identified as a cluster by the algorithm. With 12 daily messages over four years, a berth that is constantly occupied would have 12x365x4=17,520 messages. We therefore set DBSCAN's parameter to 1,000. Figure 3 below shows the sensitivity, in terms of number of core and noise points, under various choices of the parameter N. By and large, the number of noise points does not change dramatically under different choices of ε .

Figure 3. Number of core and noise points under different calibrations

different choices of ε; number of points in millions

	Core	Noise
500m	186.2	2.8
1000m	187.2	1.8
2000m	187.8	1.3
5000m	188.2	0.8

C. Assigning polygons to ports and countries

For each AIS-derived polygon, we first calculate the latitude and longitude of its centroid and try to identify the sovereign country where the centroid lies. The assignment of polygons to port groups is based on the cluster's centroid's position and works as follows.

Assigning centroids to countries. In order to map centroids' positions to countries, we use a database of polygons identifying territorial waters of all countries in the world (version 10 of the Maritime Boundaries Geodatabase; Flanders Marine Institute, 2018). For some polygons, however, their centroids happen to be located over land, so that no jurisdiction is found when searching across sea polygons. We therefore also identify the country of the centroid by using polygons over land areas, using the World Borders Dataset.[11] Whenever our search over water returns no sovereign country, we use the results of a search over land.

Of the around 3,500 cluster centroids, all but around 1,300 are mapped to a country based on maritime boundaries. Of those 1,300, all but around 500 are mapped to a country based on land boundaries. That is, we end up with about 3,000 centroids mapped to a country and about 500 centroids that are not assigned to any country.

Assigning centroids to head ports. This is done in two steps. First, if the polygon's centroid has a country assigned, we look for the set of head ports within a 30 km radius of the centroid that belongs to the same country. If this set is not empty, then we assign the polygon to the head port in this set that is closest to the centroid. Otherwise, we assign the polygon to the

[11] Available at https://thematicmapping.org/downloads/world_borders.php

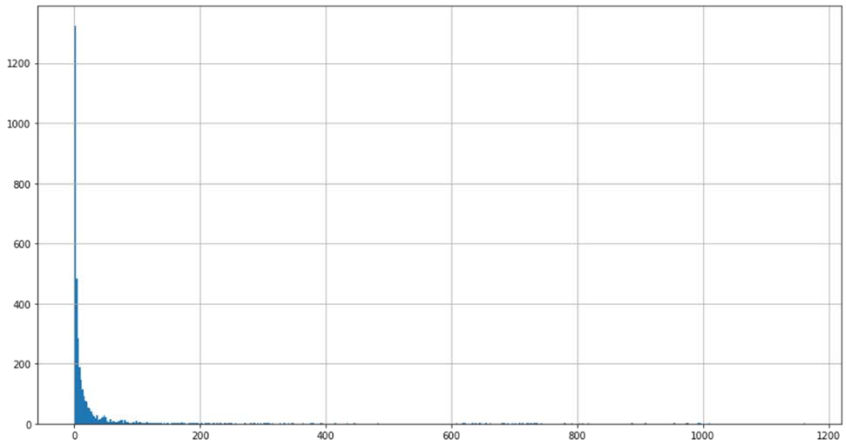
WPI port within the same country that is closest to the centroid. Secondly, if the polygon's centroid does not have a country assigned, then we look for the set of head ports within a 30 km radius of the centroid. If this set is not empty, then we assign the cluster to the head port in this set that is closest to the centroid. Otherwise, we assign the cluster to the WPI port that is closest to the centroid.

The two-step, hierarchical procedure that first searches for head ports helps avoid situations where a cluster is assigned to a non-head port i that belongs to head port j when in fact head port k is at a smaller distance from the cluster's centroid. Further note that the calibration based on a 30-km distance is consistent with the one used to construct port groups.

Removing assignment if nearest port is too far away. Figure 4 below shows the resulting histogram of centroids' distance to the nearest port. When this distance is greater than 75 km, we remove the assignment. Clusters that, because of this step, end up not being assigned to any port constitute what we denote 'synthetic ports'. Synthetic ports that lie on the waters or land of a country are synthetic ports of that country. We have a total of 16 synthetic ports, which we assign to the closest countries. In order to minimize the possibility of misclassification, we manually validate each of these assignments.[12]

Figure 4. Histogram of clusters' distance to closest WPI port

distance in km



IV. IDENTIFYING PORT CALLS – SUPERVISED LEARNING

Once we have identified port areas from the raw AIS data, the question is how to infer whether a message received from within a port polygon is a true port call or not. Ships may traverse port polygons without stopping at port facilities that allow them to load or offload cargo. Such examples include ships going up the river in Rotterdam or Buenos Aires,

[12] We should note that, in Figure 5, some polygons appear to be far from the closest WPI ports but are not necessarily far from the shores of the closest country.

crossing the San Francisco Bay to reach Oakland, or stepping on a polygon on their way through the Singapore Strait.

We make use of official statistics on vessel entrances and clearances at U.S. ports to identify true port calls from the set of polygon visits. More specifically, the U.S. Army Corps of Engineers' Navigation Data Center (NDC henceforth), in partnership with the U.S. Department of Homeland Security's Customs and Border Protection agency, compiles what is known as the national waterway data. NDC data record most port calls on U.S. ports, including 23 different fields of potential interest. Importantly, the dataset includes the vessels' IMOs and date of entry.[13] As of March 2020, NDC data are available through end 2018.[14]

A. Mapping AIS-derived port visits and official port calls

We start the analysis by looking for matches between our polygon visits and the official port-call data produced by NDC. We do this by mapping every GTI and NDC port call to one of the following categories:

Matched port calls. For each NDC port call we search for the same IMO in our port visit data over a window of +/- 2 days around the date the port call was recorded by NDC. The rationale for looking up to two days *back* in time is that by law vessels have up to two days to file their entrance report .[15] At the other end, there are two reasons to look for port calls up to two days *after* the NDC records. First, NDC records the day in local U.S. time, but the timestamp in AIS data is in UTC time. UTC timestamps can be 5-9 hours ahead of U.S. time, depending on the U.S. time zone and whether daylight saving time is in effect. That is, there can be vessels that arrive at Los Angeles at 6 pm, but it is already the next day in UTC. We limit, but do not eliminate, this source of discrepancy by shifting AIS timestamps by six hours across the board. Second, a vessel may report entry ahead of entering the port, for example while waiting at anchorage. Sometimes a single vessel may have multiple port calls within this window either in our database or in NDC database or in both. In those cases, we match NDC records to the ones with the closest dates in our dataset. If this does not break the tie, we give preference to the earliest record.

False negatives. These are IMOs recording port calls in the NDC database for which we do not find a port call in our data over the +/- 2-day window.

---

[13] The U.S. Code for Federal Regulations stipulates that all vessels, with the exception of U.S.-flagged ships coming directly from another U.S. port and without foreign goods onboard, must file an entrance statement (Code for Federal Regulations, Title 19, §4.3). U.S.-flagged ships make up only 0.5 percent of our dataset.
[14] We exclude Puerto Rico and U.S. Virgin Islands from all the analyses in this section.
[15] Code for Federal Regulations, Title 19, §4.3.

False positives. These are IMOs recording port calls in our database that remain unmatched to NDC port calls over the +/- 2-day window. Since some vessels are not required to file an entrance record, a false positive does not necessarily imply that a vessel did not enter a port.

As a matter of fact, one should not treat the official data from NDC as capturing the universe of calls at U.S. ports. In particular, NDC is unlikely to record port calls by U.S.-flagged vessels coming from another U.S. port (see footnote 10 above). We will therefore restrict attention in what follows to non-U.S.-flagged vessels that in our raw database arrive from non-U.S. ports.

B. Checking for blind spots

For the period 2015-2018, we record more than 500,000 distinct ship-date pairs with messages coming from within polygons associated with U.S. ports, whereas NDC shows nearly 290,000 entrance records. Around 215,000 NDC port calls can be mapped to an GTI polygon visit within a +/- 2-day window. As shown in Table 1, nearly all matched port calls are found on the same day in both datasets, or with a one-day lag/lead in our database. In this initial matching, about 71,000 observations are classified as false negatives (no GTI record), and over 300,000 as false positives (GTI record unmatched to an NDC record). The next subsection discusses how to correctly measure false positives and lays out our approach to tackling them. Here we focus on analyzing false negatives, and show that the actual number is much lower than suggested by the initial matching.

Table 1. Searching official port calls in GTI polygons
including all vessels

	Frequency	Percent	Cumulative
Found by GTI on day t =			
-2	804	0.1	0.1
-1	9,458	1.6	1.7
0	189,538	31.4	33.1
+1	10,427	1.7	34.8
+2	4,460	0.7	35.5
In ML-AIS but not in NDC (false positives)	318,168	52.6	88.1
In NDC but not in ML-AIS (false negatives)	71,674	11.9	100.0
Total	**604,529**	**100.0**	

The underlying AIS data that we use contain IMO-registered vessels that participate in global trade and comprise the core commercial markets of the world fleet (dry bulk, containers, wet bulk, gas carriers-LNG/LPG, etc.). To test the adequacy of our algorithms and refine our calibrations, an analysis of false negatives must first tease out how many of the unmatched NDC port calls correspond to IMOs that never show up in our underlying data. These are mostly small ships (e.g. tugs, barges, service vessels) that do not participate in international trade.

Of the around 71,000 false negative port calls, about 58,000 correspond to IMOs that are not in our data. Excluding these IMOs from the analysis reveals that our algorithm can match 94.3 percent of NDC's port calls (Table 2).

Table 2. Searching official port calls in GTI polygons
excluding vessels not appearing in underlying AIS database

	Frequency	Percent	Cumulative
Found by GTI on day t =			
-2	804	0.4	0.4
-1	9,458	4.2	4.5
0	189,538	83.3	87.8
+1	10,427	4.6	92.4
+2	4,460	2.0	94.3
In NDC but not in ML-AIS (false negatives)	12,934	5.7	100.0
Total	**227,621**	**100.0**	

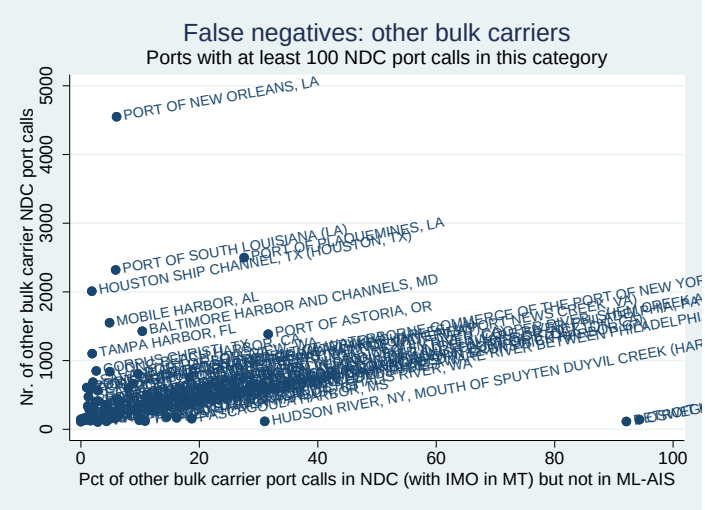
At 94.3 percent, we are able to capture the overwhelming majority official U.S. port calls. The natural question is whether this overall high figure masks weaknesses in detecting certain types of vessels. Table 3 shows total official port calls broken down by false negative and true positives for the top ten vessel types visiting U.S. ports. For example, of the 71,598 official port calls by container ships, 69,599 are detected by our algorithm, implying a false negative rate of 2.79 percent. The false negative rate is below 10 percent in all cases presented in Table 3, except for self-discharging bulk carriers. Figure A1 in Appendix A shows that, within this category, we tend to miss smaller vessels but capture the larger ones.

Table 3. Searching official port calls: Vessel-type breakdown
excluding vessels not appearing in underlying AIS database

	In GTI polygon?		Total
	No	Yes	
CONTAINER SHIP	1,999	69,599	71,598
	2.79	97.21	100
BULK CARRIER	3,216	31,666	34,882
	9.22	90.78	100
OIL/CHEMICAL TANKER	1,949	30,375	32,324
	6.03	93.97	100
GENERAL CARGO	1,258	19,232	20,490
	6.14	93.86	100
VEHICLES CARRIER	618	18,347	18,965
	3.26	96.74	100
CRUDE OIL TANKER	830	17,352	18,182
	4.56	95.44	100
SELF DISCHARGING BULK CARRIER	1,186	5,422	6,608
	17.95	82.05	100
LPG TANKER	310	5,683	5,993
	5.17	94.83	100
OIL PRODUCTS TANKER	308	3,877	4,185
	7.36	92.64	100
RO-RO CARGO	100	3,248	3,348
	2.99	97.01	100

Different port geographies may, in principle, lead to different false negative rates. For example, we may be missing port calls in ports with certain geographic characteristics that our spatial clustering algorithm couldn't handle well. Detecting any such problem would be important, especially as we will use the U.S. data to train a random forest classifier, which then will be applied to other countries. For example, for a country with its main port in a river delta (as is Rotterdam for the Netherlands), it can be especially useful to know whether GTI works well in New Orleans, where ships go up the Mississippi. We therefore analyzed the false negative rate by U.S. port and vessel type. [16] Figure 5 illustrates the case of bulk carriers, showing that false negative rates are relatively low for the most relevant ports.

Figure 5. False negatives by U.S. port: The case of bulk carriers



C. Teasing out false positives via supervised learning

While Table 1 shows that our port polygons capture nearly all official port calls, it also reveals a more pervasive problem – an abundance of false positives. That is, vessels that step on polygons but do not have entry records in the NDC data. As noted above, NDC need not record all port calls by all vessels: U.S.-flagged vessels coming from other U.S. ports and loaded only with U.S. goods are not required *de jure* to complete a port entrance form. In practice, it seems that these vessels indeed often don't file for entry. For example, of the GTI port visits for which we can find a corresponding NDC port call, about half have a previous port visit in the U.S. This fraction jumps to nearly 70 percent for GTI port visits for which we *cannot* find a corresponding NDC port call. Similarly, in the GTI port visits for which we can find a corresponding NDC port call, about 3 percent of the vessels are U.S.-flagged. This fraction jumps to 17 percent for GTI port visits for which we *cannot* find a corresponding NDC port call. For this reason, and to ensure an unbiased implementation of our

[16] It is in fact this analysis that led us to use messages with navigational status *anchored* in the construction of polygons for three specific vessel types (see footnote 4).

classification method, in what follows we drop all observations where the previous GTI port of call is in the U.S. and the vessel is U.S. flagged.

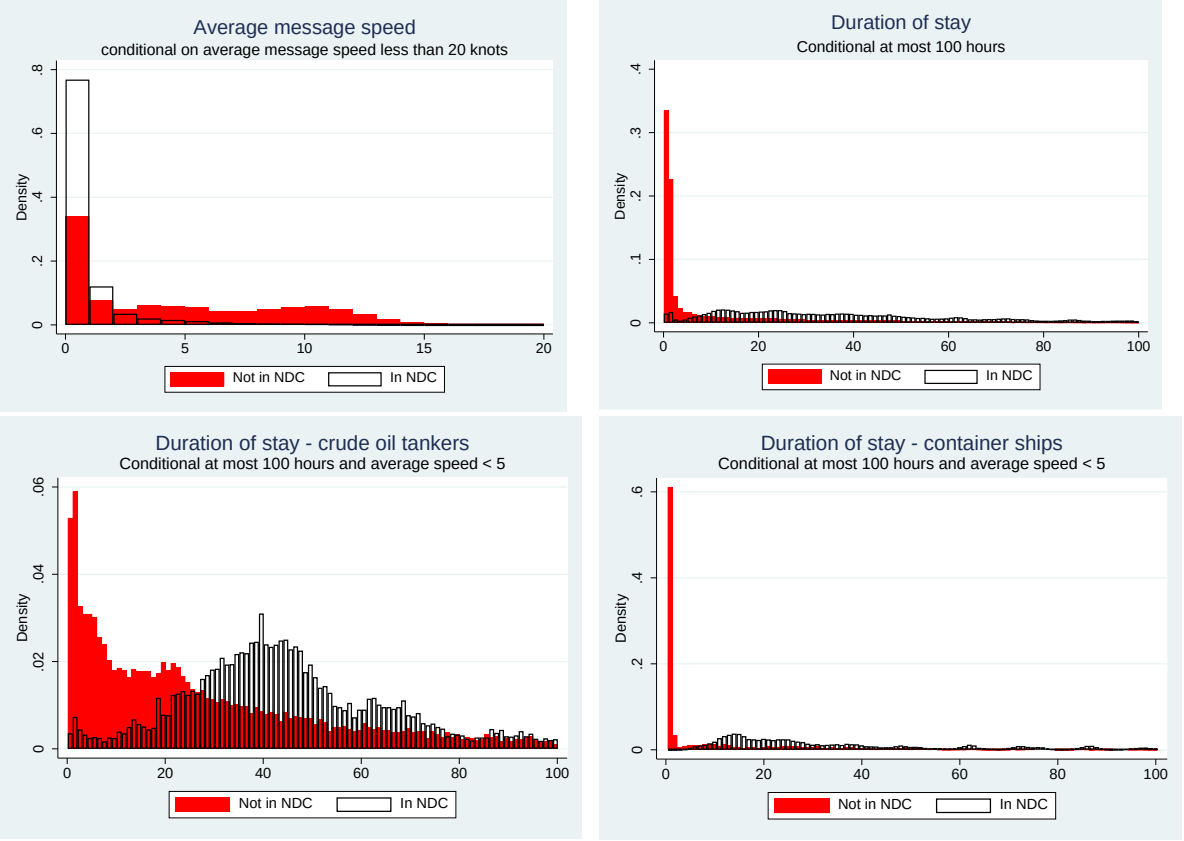
This still leaves us with plenty of false positives, that is AIS-derived polygon visits that we cannot match to official entry records.[17] Fortunately, we can rely on several features of these visits to infer whether the messages coming from within a polygon represent a port call. For example, a vessel speeding through the polygon at an average speed of 10 knots per hour and/or spending less than one hour within the polygon is unlikely to be able to load or offload any cargo during its polygon visit.

In principle, teasing out these false positives on the basis of a set of covariates is a classification problem akin to the challenge in international macroeconomics of predicting crises (see e.g. Frankel and Rose, 1996; Kaminsky, Lizondo and Reinhart, 1998; Catao and Milesi-Ferretti, 2014). Following in the footsteps of this literature, we could for example estimate a probit model of the binary variable of each visit being an NDC-recorded port call (labeled as a 1) or not (labeled as 0).

The problem with estimating models with an underlying linearity assumption (such as probit or logistic regression) is that many of the covariates that we use have non-linear and non-monotonic effects on the probability of a given observation being a port call. Consider, for example, how average speed and duration of stay within the polygon are related to the likelihood of finding a matching NDC port call for a particular polygon visit. In the case of average speed, the relation is nonlinear in the sense that, for average speeds of less than around 3 knots per hour, the distribution of average speed is similar for true and false positives (Figure 6, top left panel). For average speeds above 5, however, false positives are far more frequent than true port calls. Similarly, low and very large values of duration of stay are associated with a higher frequency of false positives. In between, however, one finds that the distribution of false positives is not very different from that of true port calls (Figure 6, top right panel). These distributions, in turn, are different at the vessel-type level, as the technology of loading and offloading varies depending on the type of cargo carried by the ship, as well as possibly the ship's size (Figure 6, bottom panels).

[17] There are around 336,000 observations of U.S.-flagged vessels where the previous GTI port of call is in the U.S., which implies that we also drop some true positives from the analysis. False positives drop the most, as expected, from around 320,000 to around 90,000.

Figure 6. Teasing out false positives using AIS data



These nonlinearities and nonmonotonicities call for a flexible, nonparametric approach to distinguish false positives from true port calls. We use a random forest classifier to predict whether a vessel's visit to a polygon is related to a port call. A random forest classifier is a so-called *ensemble* learning method that creates multiple decision trees in randomly selected subspaces of the space spanned by our covariates (known as 'features' in the machine learning literature) and yields a classification of zeros and ones by majority vote of all individual trees. Each decision tree is constructed with randomly-sampled covariates and subsample of the data (also known as training sample) to minimize a cost function (Ho, 1995; Ho, 1998; Breiman, 2001). The cost function used is Gini's diversity index:

$$C = 1 - \sum_{i=0}^{1} p_i^2$$

where p_i is the deadweith-tonnage-weighted frequency of observations labeled as class i under a node of the tree. Since in our application misclassifications are more costly the larger the ship, we penalize them in proportion to vessels' gross tonnages.

The tree is grown every time by splitting at a feature to minimize the Gini's diversity index of the parent node until it is impossible to reduce the Gini's diversity index or we reach a

constraint that the number of points under a node must be larger than a given threshold. This threshold is a hyper parameter and is used to limit the complexity of the model and thus improve out of sample prediction power. Random forests can help reduce overfitting of individual decision trees by simply introducing randomness and using majority vote of different trees (Friedman et al., 2001).

We construct a total of 16 covariates or features of polygon visits that we feed into the classifier (Table 4). The information contained in these variable ranges from characteristics of the ship (e.g. gross tonnage) or the polygon (whether it is an anchorage polygon), to measurement of events taking place within the polygon (e.g. number of messages with navigational status anchored or moored), to dynamic features that account for events in multiple polygons (e.g. number of hours from previous mooring-polygon visit). Based on preliminary bivariate analysis and intuition, we conjecture that these features will help separate polygon visits that are not true port calls.

The inclusion of variables related to ships' draught warrants some discussion. Draught corresponds to the vertical distance between the bottom of the ships' keel to the waterline. It can therefore be a useful gauge of how loaded vessels are. What is worth flagging is that, contrary to, for example, vessels' speed and position, draught is manually fed by vessels' crews to AIS transponders. As a result, measurement errors in draught readings can be pervasive. This includes not just potential errors in typing in the information, but also possibly lags in updating it when large changes in draught take place (marginal changes in draught take place constantly, including because of the burning of fuel).[18]

Table 4. Features used to tease out false positives

Variable	Definition
Anchored message ratio	Fraction of messages with status anchored (over anchored+moored)
Total no. anchored/moored messages	Count of anchored or moored messages
Is anchorage polygon	1 if it is an anchorage (i.e. not mooring) polygon
Hours from draught change	Hours between last draught change and polygon entry
Dummy previous mooring call	1 if other mooring call in past 48 hours
Hours from previous mooring call	Hours from end of previous mooring call
No. draught decrease	Count of draught decreases
No. draught increase	Count of draught increases
Share of low-speed messages	Share of messages with speed<0.5 knots
No. messages	Count of messages
Average message speed	Average speed of messages
Average speed	Distance over time from [first message out of] to [first message in] polygon
Duration of stay	Time from [first message out of] to [first message in] polygon
Summer deadweight tonnage	Summer deadweight tonnage
Gross tonnage	Gross tonnage
Ship type	Ship type

[18] We will come back below to issues related to the draught variable when discussing our construction of a volume indicator.

Our approach to training the random forest classifier works as follows. We have a total of 472,085 labeled port visits. We randomly pick 10 percent of the data points as the test set and the rest is used to train and optimize hyper parameters using five-fold cross validation (Kohavi, 1995).[19] Hyper parameter tuning is done using Bayesian optimization (Snoek at al., 2012).[20] We tune two hyper parameters: the minimum number of points at each leaf node, and the number of trees constructed. The exercise yields a minimum number of points at each leaf node of 1, and 232 trees.[21]

Table 5. Model Comparison on Testing Sample

	RANDOM FORESTS	SVM	LOGISTIC REGRESSION
ACCURACY $\frac{TP+TN}{Total\ Sample}$	91.3%	84.5%	84.7%
PRECISION $\frac{TP}{TP+FP}$	89.4%	80.9%	81.2%
RECALL $\frac{TP}{TP+FN}$	91.1%	85.0%	84.8%

Notes: TP = # of true positives, TN = # of true negatives, FP = # of false positives, FN = # of false negatives

Table 5 shows different measures of how well the random forest classifier works in the testing sample, i.e. the subsample that we left aside when training the model. We compare these results with the ones stemming from two other popular classification methods in the machine learning literature: support vector machines (SVM; Cortes and Vapnik, 1995), and logistic regression.

The results in Table 5 show an overall good performance of the random forest classifier. Moreover, the classifier outperforms the two alternatives in the three different metrics shown. The metrics shown on Table 5 are derived from each model's confusion matrix. The confusion matrix maps each observation in our AIS-derived polygon visits into one of four categories: (i) those that the model classifies as port call and show up as port calls in official statistics (true positives), (ii) those that the model classifies as port call but do not show up in

[19] That is, we randomly partition the training dataset into 5 equal-sized subsamples. Of the 5 subsamples, a single subsample is retained as the validation data for testing the model (e.g. calculating the error rate, etc.), and the remaining 4 subsamples are used to train the model. The cross-validation process is then repeated 5 times, with each of the 5 subsamples used exactly once as the validation data. The 5 results can then be averaged to produce a single estimation.

[20] Hyper parameters tuning consists in maximizing out-of-sample performance over hyper parameters. It is in general difficult to calculate the derivatives of this objective function. The Bayesian approach constructs a probabilistic model (usually Gaussian processes) for the objective function and then exploits this model to make decisions about where to next evaluate the function.

[21] Our random forest model uses all the features to grow each tree. Randomness is introduced by bootstrapping only a fraction 0.6 of the training samples for each tree.

official statistics (false positives), (iii) those that the model does not classify as port call and do not show up in official statistics (true negatives), and (iv) those that the model does not classify as port call but show up in official statistics (false negatives). As is clear from this description, false positives correspond to type I errors, whereas false negatives correspond to type II errors.

The first metric in Table 5, denoted *accuracy*, corresponds to the number of correctly-classified port calls (the sum of true positives and true negatives) in ratio to the total sample of polygon visits. Since it conflates type I and type II errors, a high level of accuracy as the one observed in Table 5 for our classifier may mask error-type-specific problems. As far as nowcasting trade goes, both types of errors would introduce equal amounts of noise to our trade indicators. It is nonetheless important to know whether one type of error is more pervasive, as false positives would imply over-estimation of trade flows whereas false negatives would be associated with under-estimation.

The second and third measures presented in Table 5 shed light on these two types of errors. *Precision* corresponds to the number of cases that both the model and the official data identify as port calls (true positives) in ratio to the sum of true positives and cases the model identifies as port calls but the official data do not (false positives). At nearly 90 percent, the precision of the random forest classifier is substantially higher than for the two alternative classifiers. *Recall* measures the number of true positives in ratio to the total number of official port calls (i.e. true positives plus false negatives). Importantly, compared to the alternative methods, the random forest classifier achieves substantially higher precision while also improving on recall. In other words, its predictions of certain polygon visits not being true port calls does not come at the expense of missing actual port calls. The random forest model outperforms SVM and logistic regression by being able to account better for the non-linearities present in our application.

The results shown in Table 5 correspond to certain exogenous thresholds in each model. For example, and most familiar to economists, the results for the logistic regression corresponds to a rule whereby an estimated probability of being a port call greater than or equal to (lower than) 0.5 yields a predicted label of port call (not a port call). Potentially, one can choose a different rule as a tradeoff between true positive rate and false positive rate. For example, one could label as predicted port call those observations with a logistic-estimated probability greater than 0.6 instead of 0.5. This will naturally lead to a lower incidence of false positives and a higher incidence of false negatives. The Receiver Operating Characteristics (ROC) curve, shown in Figure 7 for the three different classification methods, illustrates precisely this tradeoff. The figure shows clearly that random forest classifier is unambiguously the best classification method among the three methods we considered.

Figure 7. ROC Curve of Different Models on Test Set

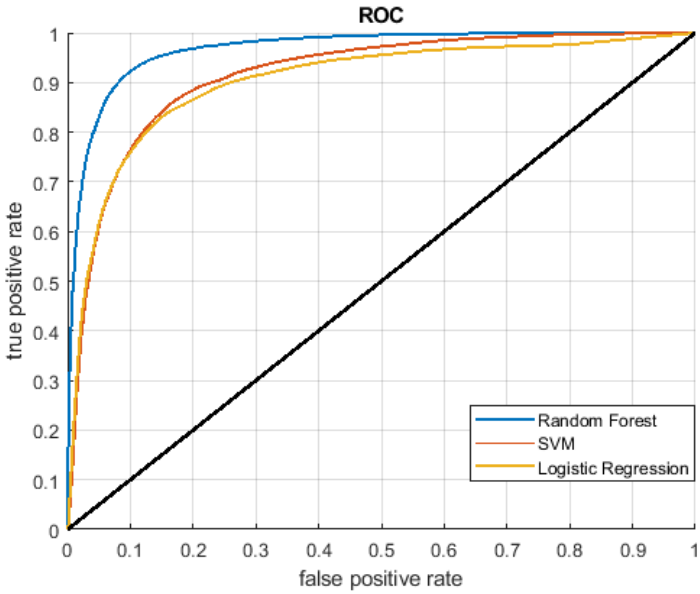
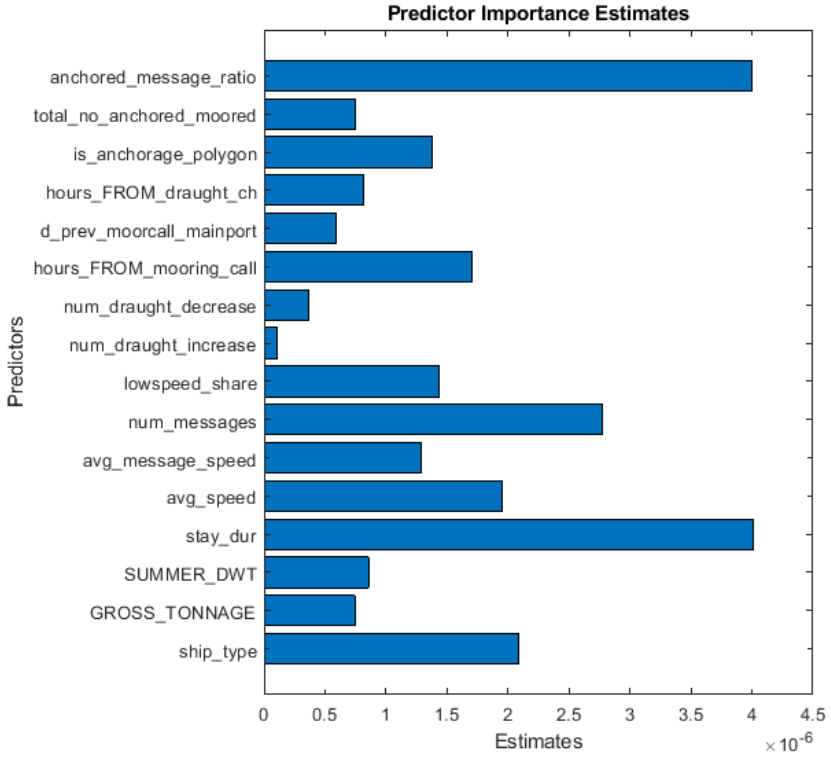


Figure 8 shows the importance of different covariates of our trained random forest model. The importance of a covariate or feature is estimated by summing changes in the Gini's diversity index due to splits on every predictor and dividing the sum by the number of branch nodes. As shown in the plot, the two most important features are the anchored message ratio and the duration stay within the polygon. The first of these is a feature of the polygons themselves, not the polygon visits. Its importance reflects the fact that most of our false positives come from anchorage polygons, where by construction the fraction of anchored messages will be higher. The importance of duration of stay, in turn, is intuitive since vessels need a certain amount of time to load/unload cargo and if the stay is too short or too long, it becomes less likely that the polygon visit is trade related. Coupled with other variables such as number of messages from polygon, ship type, and time to next mooring polygon, these are strong predictors of whether a polygon visit is related to a port call or not.

Figure 8. Feature Importance in the Random Forest Model



V. THE FINAL STEP: VOYAGES AND THE VOLUME OF TRADE

Once the random forest classifier has been trained using the official port-call data for the U.S., we apply it to our worldwide AIS-derived polygon-visit dataset. That is, we feed the same 23 features described above so that the classifier returns a predicted label of either 0 (not a port call) or 1 (port call) for each polygon visit. This constitutes our estimated global port call database.

It is straightforward to get voyages between these estimated port calls. The only thing to note is that when constructing these voyages, we 'see through' certain ports. This is done so as to avoid incorrectly classifying the origin/destination at the bilateral level, and is done at the expense of losing trade gauges for these jurisdictions. In particular, we discard polygon visits where the jurisdiction is Panama or Gibraltar, and two ports along the Suez Canal (WPI names: *As Suways*, and *Bur Said (Port Said)*). These are well-known bottlenecks in the maritime transportation network.

Equipped with these voyages, the last remaining step consists in assigning to each trip a gauge of the volume of trade involved. With our existing dataset, the best possible gauge of the metric tons of cargo transported by a ship i at time t is a function of the ship's deadweight tonnage and draught data. Specifically, let mtc_{it}^{IN} denote the inbound metric tons

of cargo transported by ship i at time t into a given port. Further let dwt_i denote the deadweight tonnage of i, $d_{i,max}$ denote maximum (or design) draught, and $d_{i,ballast}$ denote its draught when not transporting any goods (or ballast draught), and d_{it} denote the observed draught. Then, we can estimate the incoming cargo weight as:

$$mtc_{it}^{IN} = dwt_i \times \frac{d_{it} - d_{i,ballast}}{d_{i,max} - d_{i,ballast}},$$

where we adjust the ship's total capacity with the current utilization rate.

We have data on ships' deadweight tonnage dwt_i, and design draught $d_{i,max}$. As mentioned above, draught is manually entered by the crew and thus prone to measurement error. We take d_{it} to be the last observed draught before entering a polygon, as we presume that crews will more carefully report their draught shortly before entering a port so as to inform port authorities that the vessel will be able to enter without problems.

We do not have in our data a measure of ballast draught, and therefore impute it as follows. For each vessel, we calculate the ratio of the first percentile of draught as observed in the four years of AIS messages to the vessel's design draught. Then, by ship type (crude oil tanker, container ship, etc.) and by deadweight tonnage tertile (within ship type), we get the median of this ratio. Let mr_{rs} denote this median ratio for ship type r and size tertile s. For each ship i of type r and tertile size s, $i(r,s)$, we impute the ballast draught as

$$d_{i(r,s),ballast} = mr_{rs} \times d_{i,max} .$$

The estimated metric tons of cargo mtc_{it} gives us a gauge of the volume of trade in each voyage. Voyages, of course, can take place between ports of different countries or within a country. Moreover, a vessel coming from overseas may offload cargo (imports) in multiple domestic ports, or load cargo (for exports) in multiple domestic ports. We therefore need a mapping from metric tons of cargo in port-to-port voyages to volumes of imports, exports, and internal trade.

Our procedure to assign metric tons of cargo to imports, exports, and internal trade works as follows:

(1) Recall that mtc_{it}^{IN} denotes the metric tons of cargo transported by a ship i at time t into a given port. We estimate the outgoing cargo weight as $mtc_{it}^{OUT} = mtc_{it+1}^{IN}$, that is the outgoing cargo weight is equal to the incoming cargo weight of the next port. As stated above, this is motivated by the idea that incoming draught may be more precise than leaving draught. Next, from the incoming and outgoing cargo weight, we define the net cargo offloaded at the port as $mtc_{it}^{OL} = mtc_{it}^{IN} - mtc_{it}^{OUT}$. Note that this number can be negative or positive.

(2) If a ship enters a port in a new country (that is, if the previous port call was in a different country), we check the sign of mtc_{it}^{OL}. If it is positive, we say that it is an import. If it is negative, we don't count any imports on this trip, even if the ship later goes to another port in the same country and offloads some cargo. The reason is that we cannot determine if it is offloading foreign goods or goods that it picked up within the country (internal trade).

(3) If at the first port of call in the country there was an import event as defined above, we keep counting as imports any subsequent drops in metric tons of cargo (i.e. so long as $mtc_{it+j}^{OL} \geq 0$ for *every* $j = 0, \dots, J$ until $mtc_{it+(J+1)}^{OL} < 0$, i.e. up to the first instance in which the ship starts 'gaining' metric tons of cargo).

This procedure yields a well-defined measure of imports, but there is an obvious timing issue: what should be the date of importation if the ship makes multiple importing port calls? We assign the date of each port call to the relevant portion of offloaded cargo.

In order to define exported cargo we follow the same process as outlined above but starting backwards from the last port of the ships' voyage through the country's ports. That is, we first find the last port call within the country. If mtc_{it}^{OL} is negative (that is, the ship 'gained' metric tons of cargo), we say it is an export event. Then we go one port call back in the same country (if there is any). If the ship also gained cargo, we also say it is exports. And we count every previous loadings as exports, so long as the ship keeps (continuously) collecting cargo.

After this process, we are able to group every port call into one of three categories: (i) importing; (ii) exporting, and (iii) internal trade. We also have cargo weight estimates for these events. We conclude by aggregating across the desired country groupings, vessel types and time periods. For example, we can create monthly country-level import and export volume indicators by adding up all the imported and exported cargo weight for each country and each month. If we were interested in weekly bilateral crude oil trade, we could focus only on crude oil tankers and aggregate by week and country pairs.

VI. RESULTS

The results from our volume gauges for port-to-port voyages have applications at the aggregate macro-analyses level, can be used for sectoral analysis, and for event studies. We explore each of these applications in the three subsections that follow.

A. Macroeconomic nowcasting

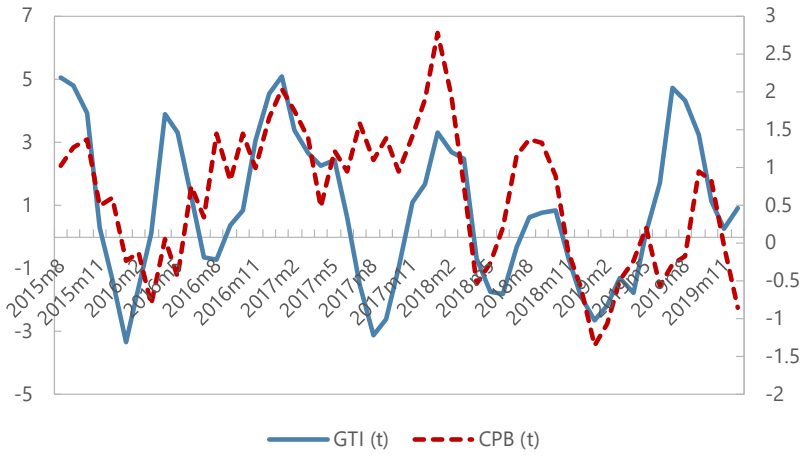
Table 6 presents the comparison of our economy-level GTI indices for exports, import and trade with official statistics as reported by CPB. The comparison is also done under different

transformations: the seasonally-adjusted data in levels, 3-month moving averages, and 3-month/3-month growth rates. The results are ordered from worst to best performance on the basis of tracking the growth rate of import volumes (i.e. the third-to-last column). It is worth noting that, unfortunately, CPB was unable to share with us non-seasonally adjusted series. In all our comparisons, therefore, we aim to use the same seasonal adjustment process as CPB. These comparisons might look better if we could either compare non-seasonally adjusted series, or perform ourselves the seasonal adjustment of customs series as well as our GTI series.[22]

At the global level, our world seasonally adjusted monthly import volume gauge has a correlation of 0.88 with official statistics. In 3m/3m growth rate terms, the correlation is of 0.4, with the two series shown in Figure 9. The performance is fairly similar for the case of world exports, with only the growth rate of exports showing a slightly poorer tracking. For the 3-month moving averages in levels, the median correlation with the set of economies in the CPB sample is of 0.68 for imports, and of 0.46 for exports. In growth rates, the median correlations are of 0.27 for imports, and 0.18 for exports.[23]

Figure 9. World: GTI Index and Official Data

imports, 3m/3m growth rates, seasonally adjusted



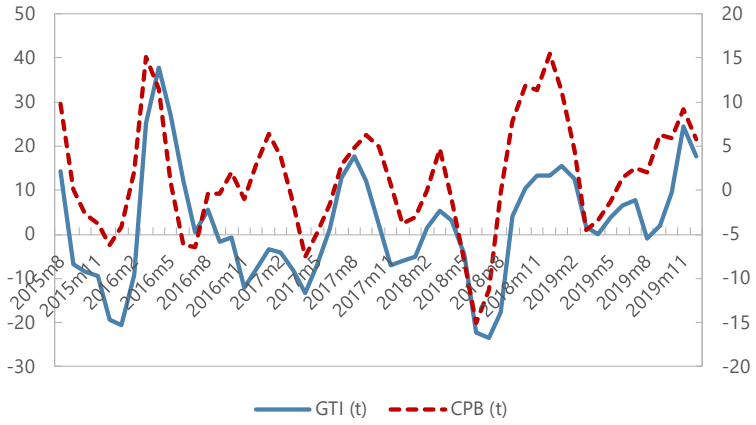
[22] Relatedly, in Table 6 all correlations are using our index and CPB data contemporaneously, but – as in the case of the Euro Area – the lead that yields the best correlation with customs data differs by country. This should be taken into account when using the data for nowcasting.

[23] It is worth noting that the correlation for the world in aggregate is higher than for the individual economies presented in Table 6. There are at least two reasons that would explain this. First, because in many cases the port of importation/exportation might not belong to the country importing/exporting. This is also the likely explanation for why our index tracks much better the Euro Area as a whole than some of its individual members. Second, our index might feature a relatively good fit with countries for which we have been unable to get official monthly trade volume data.

We note that, in constructing our current indices, we add up the estimated metric tons of cargo across all types of ships. More generally, this might explain why the index we use may work particularly well with certain countries with less-diversified imports or exports. For example, the correlation with Argentina's exports is the highest among all economies, as shown in Table 6 and illustrated in Figure 10. Appendix B explores the possibility of using nominal trade weights to construct our indices.

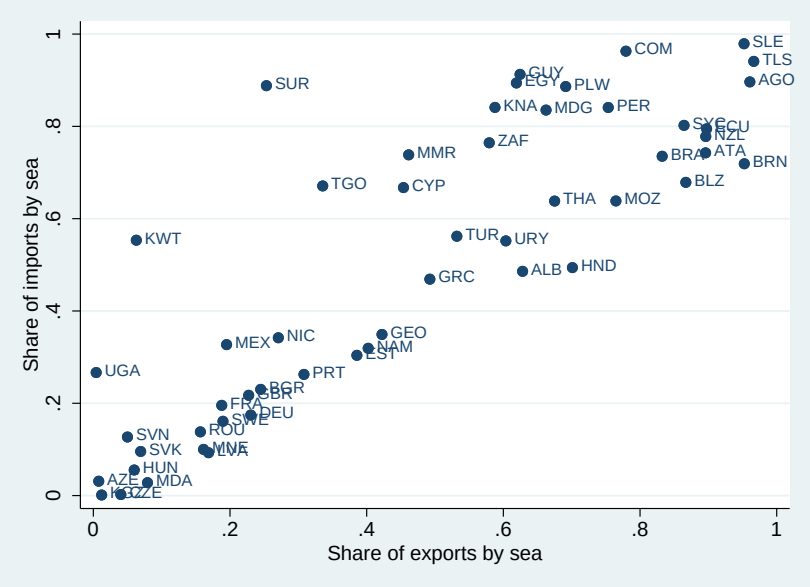
Figure 10. Argentina: GTI Index and Official Data

exports, 3m/3m growth rates, seasonally adjusted



Finally, it is worth noting that, while we have produced results for every country in the world that has a coastline, we have only been able to benchmark our trade volume estimates for 43 economies (including the Euro Area) and the world. This is the result of the current dearth of publicly available monthly trade volume data. While the set of individual economies we did benchmark covers most of global trade, our methodology might potentially also be very useful for a vastly larger number of cases. In particular, our benchmarking sample is heavily biased toward advanced economies, but several emerging market and developing economies feature a very high share of trade that takes place by sea. This is illustrated in Figure 11, where we plot every economy that, as of April 2020, has reported the mode of transportation in its annual COMTRADE data. Given this landscape, expanding the extent of benchmarking of our results seems important.

Figure 11. Share of trade by sea in 2017



Source: UN COMTRADE.

Table 6. Benchmarking GTI indices at the economy level

	Transformation and Trade Flow								
	Raw (level)			3m mov. av. (level)			3m/3m growth		
	Import	Export	Total	Import	Export	Total	Import	Export	Total
Malta	0.09	-0.23	-0.14	-0.21	-0.16	-0.27	-0.13	-0.19	-0.29
India	0.78	0.32	0.69	0.84	0.30	0.70	-0.13	0.17	-0.05
Estonia	0.05	0.28	0.13	0.07	0.43	0.20	-0.03	0.09	0.02
Rep. of Korea	0.41	0.37	0.42	0.46	0.59	0.54	-0.02	-0.11	-0.22
Taiwan Province of China	-0.13	0.41	0.22	-0.21	0.51	0.23	0.02	0.22	0.19
Romania	0.61	0.16	0.55	0.74	0.05	0.70	0.03	0.15	0.16
Netherlands	0.05	0.62	0.55	0.18	0.72	0.71	0.04	0.21	0.24
Germany	0.44	-0.08	0.36	0.63	-0.04	0.57	0.05	-0.23	-0.13
Lithuania	0.62	0.63	0.73	0.83	0.83	0.92	0.07	0.48	0.16
New Zealand	0.63	0.32	0.62	0.84	0.68	0.84	0.08	0.32	0.32
Cyprus	0.14	-0.08	-0.19	-0.01	-0.26	-0.25	0.08	-0.01	-0.11
Brazil	0.32	0.43	0.35	0.54	0.56	0.55	0.11	0.62	0.42
USA	0.58	0.87	0.85	0.66	0.92	0.89	0.12	0.38	0.32
Russian Federation	0.03	0.87	0.59	0.05	0.89	0.67	0.19	0.19	0.13
United Kingdom	-0.09	0.21	0.21	-0.26	0.37	-0.04	0.19	0.38	0.42
Canada	0.71	0.62	0.73	0.84	0.75	0.86	0.19	0.49	0.08
Finland	0.54	0.57	0.70	0.71	0.77	0.83	0.23	-0.23	0.05
Argentina	0.58	0.69	0.45	0.77	0.68	0.48	0.23	0.72	0.49
Indonesia	0.68	0.12	0.53	0.78	0.23	0.67	0.23	0.03	0.06
Slovenia	0.39	-0.25	0.26	0.64	-0.39	0.51	0.24	-0.06	0.20
Mexico	0.76	0.57	0.80	0.81	0.68	0.87	0.26	-0.12	0.08
Singapore	0.43	-0.29	0.18	0.56	-0.56	0.15	0.26	0.29	0.16
Greece	0.70	0.80	0.82	0.86	0.86	0.89	0.27	0.44	0.38
Poland	0.89	-0.30	0.85	0.95	-0.43	0.91	0.28	0.04	0.12
Belgium	0.41	0.24	0.40	0.69	0.48	0.66	0.28	-0.24	-0.01
Latvia	0.65	-0.37	0.50	0.82	-0.44	0.67	0.30	0.15	0.10
Ireland	0.57	-0.28	0.29	0.82	-0.52	0.45	0.31	-0.05	0.11
Portugal	0.20	-0.62	-0.49	0.28	-0.76	-0.63	0.31	0.42	0.24
Spain	0.67	0.73	0.74	0.85	0.82	0.86	0.31	-0.04	0.02
Denmark	0.44	0.31	0.55	0.75	0.56	0.78	0.33	0.15	0.40
Australia	0.48	0.24	0.53	0.69	0.15	0.63	0.33	0.46	0.19
Norway	0.36	0.27	0.33	0.59	0.40	0.60	0.33	0.22	0.09
Euro Area	0.76	0.76	0.83	0.89	0.89	0.93	0.34	0.11	0.04
France	0.45	0.10	0.46	0.67	0.03	0.54	0.34	0.37	0.31
China	0.56	0.69	0.71	0.56	0.76	0.75	0.39	0.49	0.46
Iceland	0.56	0.30	0.59	0.80	0.50	0.85	0.39	0.44	0.42
WORLD	**0.85**	**0.85**	**0.86**	**0.88**	**0.87**	**0.88**	**0.40**	**0.32**	**0.40**
Bulgaria	0.30	0.44	0.43	0.45	0.59	0.67	0.40	0.33	0.33
Hong Kong SAR, China	0.26	-0.37	-0.20	0.44	-0.48	-0.22	0.42	-0.18	0.00
Croatia	0.44	-0.10	0.38	0.58	-0.15	0.46	0.43	-0.06	0.29
Italy	0.27	-0.05	0.16	0.31	-0.11	0.12	0.47	-0.05	0.24
Japan	0.30	0.13	0.29	0.33	0.09	0.25	0.49	0.27	0.49
Sweden	0.66	0.03	0.62	0.76	-0.04	0.76	0.53	-0.12	0.30
Turkey	0.68	0.75	0.75	0.77	0.86	0.88	0.68	0.23	0.38

B. Sectoral analysis: Crude oil trade

In most cases, AIS data on their own do not allow for analyses at the commodity level. Vast product varieties can be transported by two specific types of vessels: those designed to transport containers, and those designed to transport dry bulk commodities. The technology required to transport a narrower set of commodities, however, implies that only a few types of vessels can carry them. As noted by Adland et al. (2017), the most salient case is the that of crude oil. This provides an ideal testing ground to analyze the performance of our general methodology at a more granular level.[24]

In this section we compare our estimated metric tons of cargo for crude oil tankers with data from the Joint Organization Data Initiative (JODI). JODI compiles and makes publicly available oil and gas international trade data for a wide range of economies. We use JODI's monthly data on crude oil trade, measured in thousands of metric tons of oil. In addition to compiling volume-of-trade data, JODI also performs data quality assessments and encodes the results of this exercise in their databases. In order to understand where lack of fit to data is more likely a result of the need for refinements in our methodology, and where it might reflect data-quality issues, we present all results alongside JODI's assessment code for the specific country and data flow shown. JODI's quality assessment code can take four values: 1 (*Results of the assessment show reasonable levels of comparability*), 2 (*Consult metadata/Use with caution*), 3 (*Data has not been assessed*), and 4 (*Data under verification*).

Tables 7 and 8 below show the correlations of our metric tons of cargo estimates with JODI's crude oil data for imports and exports, respectively. In each table, economies are sorted first by JODI's assessment code, and second by the volume of trade. Unlike the case of Table 6 above, in these tables trade data are not seasonally adjusted.[25]

In their comparisons using annual data for 2013-2015, Adland et al. (2017) noted that the match of seaborne trade estimates to JODI statistics is affected by the use of pipelines and transshipments. Despite these inherent problems, the results presented in the tables below show a reasonably good fit for global crude oil imports. The correlation in levels for global imports is of 0.73 in the raw data, and of 0.81 on a 3-month moving average basis. In growth rates, the correlation is as high as 0.47. This fit between the GTI and JODI growth series is illustrated in Figure 12.

[24] Finished vehicles are also mostly transported by one type of vessel, known as roll-on/roll-off vessels. Compared to crude oil, however, they represent a wider set of product varieties, requiring e.g. the construction of deflators to perform volume comparisons.

[25] This responds to two reasons. First, in the analysis of Table 6 we had no alternative but to seasonally adjust our series in order to make them comparable to the seasonally adjusted CPB series. Secondly, it is less clear to us that oil trade should exhibit seasonal patterns and we therefore refrained from manipulating the original data thereby possibly introducing noise.

At the country level, the import gauge does not always track JODI data as well. In some important cases, as in Germany and the Netherlands, this is most likely due to the port of arrival differing from the actual destination. Possibly as a result of this, the GTI index does not track well every Euro Area country, but does a good job for the Euro Area as a whole, with correlations in levels and growth rates well above 0.5.

The case of exports is quite different, and this might reflect data quality issues. As can be seen in Tables 7 and 8, JODI's global imports far exceed global exports (by over 15 percent). Possibly even more important, imports with a JODI assessment code of 1 represent 83 percent of global imports, but the same figure for global exports is of a mere 39 percent. In this context, it is not surprising that our tracking of JODI's *export* data is poor at the global level, as well as for many large exporters. Our index, which is free from weaknesses in statistical reporting, may in some cases be useful as a source of independent information on oil exports.

Figure 12. World: GTI and JODI crude oil import growth

imports, 3m/3m growth rates, seasonally adjusted

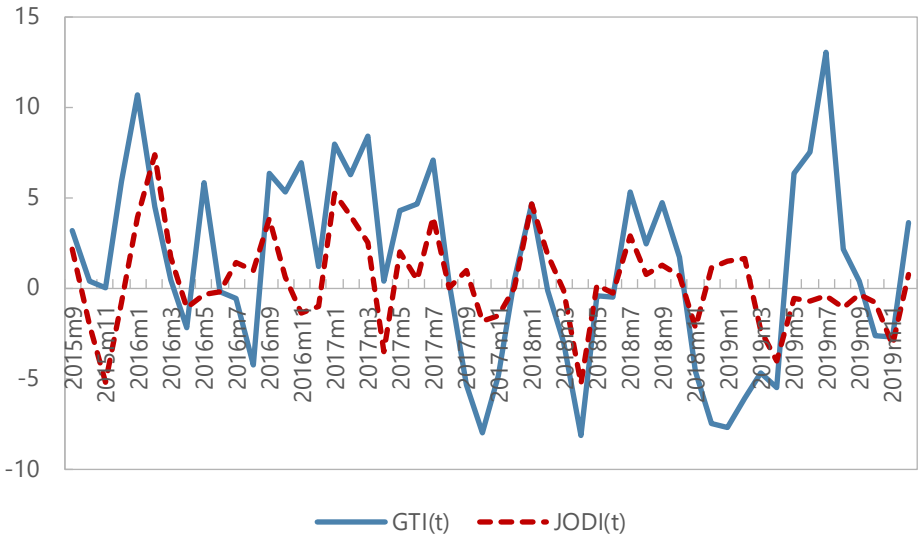


Table 7. Benchmarking crude oil imports

	JODI data quality code	Thousands of metric tons	Transformation		
			Raw (level)	3m mov.av. (level)	3m/3m (growth)
China	1	35,510	0.28	0.19	0.20
USA	1	31,082	0.52	0.71	0.46
Japan	1	12,799	0.52	0.46	0.30
Rep. of Korea	1	12,203	0.05	0.04	-0.32
Germany	1	7,403	-0.12	-0.09	0.08
Spain	1	5,492	0.50	0.63	0.74
Italy	1	5,269	0.40	0.48	0.34
Netherlands	1	4,542	0.17	0.33	0.27
France	1	4,451	0.42	0.55	0.65
Thailand	1	3,772	0.21	0.32	0.10
Taiwan Province of China	1	3,616	0.03	-0.17	0.10
United Kingdom	1	3,605	0.44	0.62	0.47
Canada	1	3,182	0.02	-0.11	0.40
Belgium	1	2,766	0.19	0.26	0.21
Poland	1	2,164	0.37	0.46	0.38
Turkey	1	2,153	0.26	0.36	0.35
Greece	1	1,938	0.30	0.39	0.26
Sweden	1	1,606	0.28	0.08	0.29
Australia	1	1,372	0.22	0.44	0.16
Portugal	1	1,096	0.43	0.30	0.27
Finland	1	934	0.21	0.41	0.29
Brazil	1	782	0.22	0.44	-0.04
Chile	1	723	0.17	0.26	0.13
New Zealand	1	430	0.30	0.11	0.25
Denmark	1	380	0.21	0.15	0.15
Peru	1	330	0.11	0.14	0.44
Ireland	1	257	0.21	-0.21	0.12
Uruguay	1	190	0.45	0.35	0.31
Trinidad and Tobago	1	170	0.39	0.41	0.09
Norway	1	161	0.53	0.72	0.01
Dominican Rep.	1	98	0.65	0.60	0.22
Jamaica	1	87	0.45	0.47	0.54
Nicaragua	1	56	-0.19	-0.47	0.24
Colombia	1	55	0.62	0.74	0.78
Russian Federation	1	30	-0.05	0.01	0.44
Argentina	1	14	-0.10	-0.36	-0.59
Mexico	1	3	0.03	-0.02	0.80
India	3	18,069	0.39	0.55	-0.12
Singapore	3	3,705	0.10	0.42	0.02
South Africa	3	1,039	0.02	-0.07	-0.06
Indonesia	3	1,021	0.16	0.24	0.48
Bahrain	3	908	0.06	0.11	0.12
Malaysia	3	899	0.22	0.48	-0.02
Lithuania	3	789	-0.17	-0.01	0.10
Philippines	3	784	-0.08	-0.09	0.08
Romania	3	648	0.23	0.53	0.49
Bulgaria	3	539	0.26	0.20	0.15
Egypt	3	444	-0.06	-0.16	-0.22
Croatia	3	214	0.21	0.40	0.02
United Arab Emirates	3	141	0.24	0.36	0.31
Papua New Guinea	3	101	0.46	0.52	0.21
Ukraine	3	58	0.27	0.55	0.05
Ecuador	3	56	0.10	0.25	0.18
Tunisia	3	46	0.23	0.47	0.01
Morocco	3	28	0.01	0.03	-0.24
Venezuela	3	27	0.08	-0.03	0.08
Algeria	3	15	-0.07	-0.09	-0.06
Oman	3	1	-0.09	-0.28	
Georgia	3	1	-0.06	0.02	0.87
Albania	3	0	-0.02	-0.06	
WORLD		180,254	0.73	0.81	0.47
Euro Area		34,936	0.58	0.68	0.60

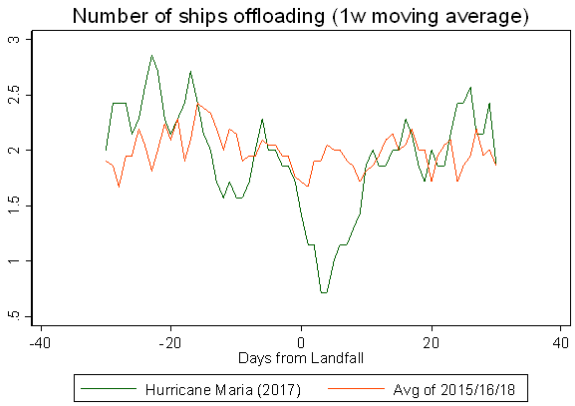
Table 8. Benchmarking crude oil exports

	JODI data quality code	Thousands of metric tons	Transformation		
			Raw (level)	3m mov.av. (level)	3m/3m (growth)
Russian Federation	1	21,159	0.13	0.23	-0.34
Canada	1	12,018	0.49	0.65	0.21
USA	1	6,030	0.92	0.96	0.54
Norway	1	5,360	0.52	0.60	0.57
Mexico	1	5,160	-0.07	-0.25	-0.20
Brazil	1	4,498	0.57	0.79	0.40
United Kingdom	1	2,976	0.18	-0.06	0.26
Colombia	1	939	0.50	0.73	-0.15
Australia	1	878	0.51	0.69	0.43
Brunei Darussalam	1	422	0.10	0.18	-0.88
Denmark	1	292	0.05	-0.21	0.02
China	1	228	-0.08	-0.27	-0.06
Argentina	1	184	0.37	0.40	0.25
New Zealand	1	113	0.17	0.28	0.23
Thailand	1	98	-0.14	-0.16	0.25
Trinidad and Tobago	1	50	-0.20	-0.34	0.41
Italy	1	44	0.07	0.23	0.04
Netherlands	1	30	0.00	-0.02	-0.06
Guatemala	1	30	0.36	0.36	0.09
Poland	1	20	-0.20	-0.39	-0.25
Sweden	1	19	0.34	0.43	0.89
Greece	1	14	0.01	0.03	-0.12
Peru	1	7	-0.13	-0.24	0.01
Germany	1	7	0.03	-0.05	0.18
Ireland	1	6	0.00	0.19	-0.40
Saudi Arabia	3	30,228	-0.20	-0.18	-0.34
Iraq	3	15,011	0.46	0.57	0.02
Kuwait	3	8,385	-0.24	-0.28	-0.26
United Arab Emirates	3	7,658	0.39	0.46	0.16
Nigeria	3	7,629	0.03	0.19	0.25
Angola	3	5,921	0.47	0.70	0.15
Iran	3	5,174	0.40	0.57	0.01
Venezuela	3	2,285	0.39	0.46	0.07
Algeria	3	1,896	0.00	-0.23	-0.09
Ecuador	3	1,697	0.13	0.21	-0.21
Qatar	3	1,570	-0.02	-0.05	0.17
Malaysia	3	1,323	-0.07	-0.26	0.00
Indonesia	3	834	0.12	0.32	0.12
Gabon	3	683	0.06	0.17	-0.18
Oman	3	677	-0.38	-0.57	0.44
Egypt	3	551	0.09	0.11	0.06
Bahrain	3	435	0.38	0.64	-0.11
Tunisia	3	130	-0.20	0.04	-0.02
Equatorial Guinea	3	94	-0.09	-0.13	0.94
Albania	3	56	-0.09	-0.19	-0.94
Philippines	3	46	-0.01	-0.11	-0.12
Singapore	3	16	0.08	-0.02	0.04
Lithuania	3	4	-0.10	-0.15	-0.09
Romania	3	3	0.11	0.18	-0.06
Croatia	3	2	-0.12	-0.16	-0.69
Ukraine	3	2	0.47	0.36	
Georgia	3	0	-0.06	0.03	-0.46
WORLD		155,446	0.07	0.01	0.11
Euro Area		105	0.08	0.24	-0.12

C. Event studies

Since our AIS-derived trade gauges are built at the port level, our results can also be used for micro-level analyses around specific events. We illustrate this by looking at the effects of Hurricane Maria, which made landfall in Puerto Rico on September 20, 2017. The catastrophic Category 5 hurricane is estimated to have caused as many as 4,600 deaths (Kishore et al., 2018), and its total USD damage makes it one of the top five costliest disasters for the U.S. (Smith, 2020). Figure 13 illustrates the disruption caused in the immediate aftermath of the hurricane. Accounting for seasonal patterns, the trade gauge shown in Figure 13 shows that offloading vessel traffic declined by as much as 75 percent and took around 10 days to recover, in coincidence with U.S. President Trump temporarily waiving the 1920 Jones Act[26] that puts restrictions on the ships and ships' crews that are allowed to transport goods between U.S. ports.

Figure 13. Hurricane Maria in Puerto Rico



Since our results are at a daily frequency, we can also use them to bring into focus country-level patterns that may not always show up as crisply in monthly data. Furthermore, since the results can be produced in real time, they can also help guide policy makers during crises as they unfold. The current COVID-19 pandemic has in fact dramatically increased the demand for tools to monitor economic activity in real time.

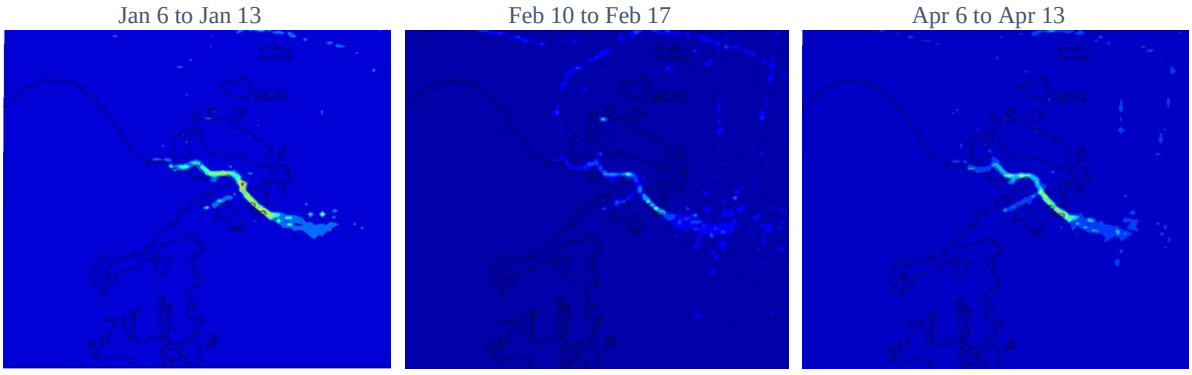
To give a visual sense of the effects of what has been dubbed the Great Lockdown (Gopinath, 2020), Figure 14 shows the frequency of ship messages around the Port of Ningbo, one of the busiest ports in China, over three different weeks in 2020. The panel on the left corresponds to the first week of 2020, where the port is seemingly teeming with activity. The panel in the middle shows a week in the midst of the strictest pandemic containment measures. By the account of this picture, things went quite dark. The right panel

[26] See e.g. "Trump Waives Jones Act for Puerto Rico, Easing Hurricane Aid Shipments," The New York Times, September 28, 2017.

shows the second week of April – lights are up again, but on close inspection maybe not as bright as pre-crisis.

Figure 14. Port of Ningbo, China

frequency of AIS messages with speed above 0.5 knots, 2020 data



Visual aids can be very powerful, but more structure is needed to create economically-meaningful, policy-relevant indicators to help answer simple questions such as: How big is the drop in trade activity? Should it be attributed mostly to exports or to imports?

Figure 15 shows our daily estimated metric tons of exports for China in the first three and a half months of 2020. The series is normalized by the estimated volume of exports for the previous three years. To purge the results from any calendar effects, these base years are adjusted for the Lunar New Year. The results show the dramatic fall in Chinese exports in the wake of lockdown measures to contain the spread of the novel coronavirus. Exports resumed in early to mid-March, though by mid-April the recovery remained incomplete.

Figure 15. China: Estimated metric tons of exports

relative to 2017-2019 avg, 30-day moving average, LNY-adjusted

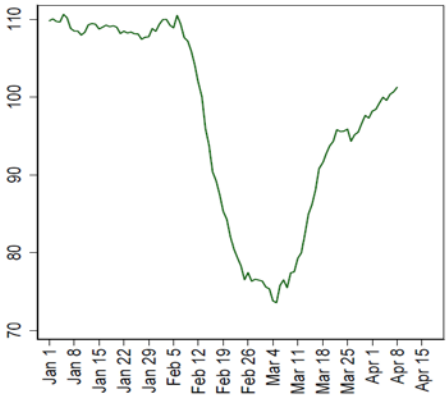
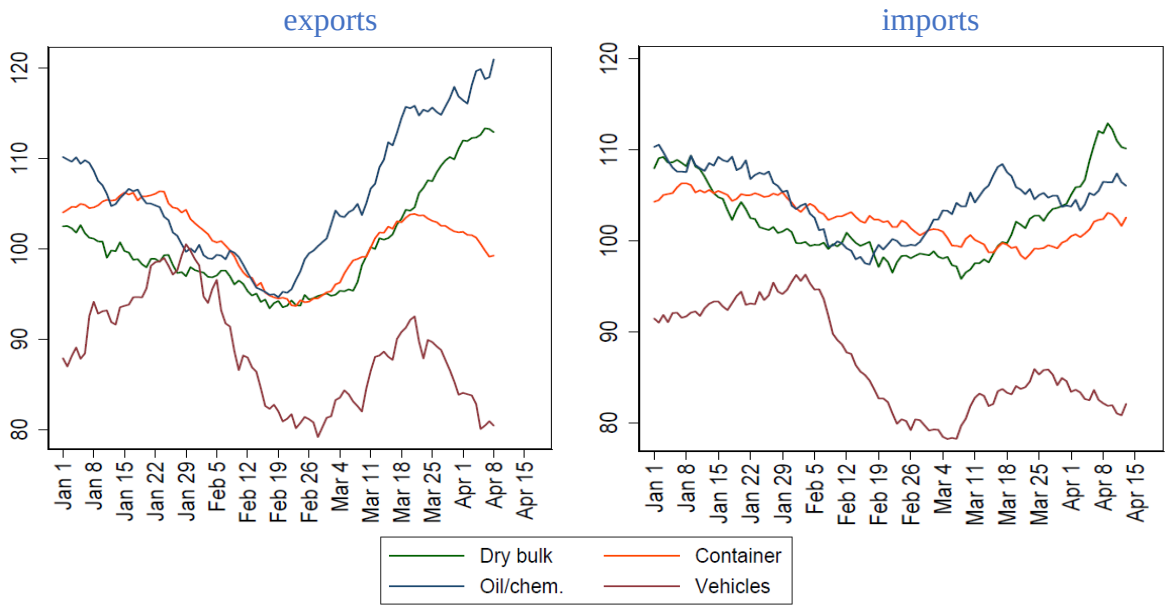


Figure 16 shows the real-time estimates of world trade, for exports and imports. To shed light on the potential heterogeneous effects at the sectoral level, the results are broken down by broad vessel categories. As China started reopening its economy, world exports initially

recovered across the board. In the specific case of oil and related products, the recent export performance is especially strong but is not fully matched by an increase in world imports – in line with reports that crude oil is being stored at sea. More recently, however, exports of less commoditized goods (those transported in containers, and finished vehicles) appear on course for a second dip. The situation is perhaps best reflected in the very weak readings for vehicle exports and imports as companies halt production and households postpone purchases of durable goods.

Figure 16. Real-time estimates of world trade
relative to 2017-2019 average, 30-day moving averages



VII. ONGOING DEVELOPMENTS AND VENUES FOR FUTURE IMPROVEMENTS

Our proof of concept clearly demonstrates that it is possible to build a real-time global trade nowcasting system relying only on raw AIS data and off-the-shelf machine learning algorithms. However, there are several areas where ongoing efforts we are currently undertaking will likely enhance the performance and granularity of our nowcasts. The planned improvements on our agenda fall under three categories: (i) refinements in the algorithms that do not require new data or any external information; (ii) changes in trade volume estimation that require information not included in the raw AIS data; and finally (iii) modifications that require deep domain knowledge about some segments of the shipping industry or particular ports. The next paragraphs provide specific examples in all these categories, setting up a roadmap for the next phase of the development of our Global Trade Intelligence index.

First, there is room to upgrade the construction of port polygons using more sophisticated spatial clustering algorithms. DBSCAN is very efficient in identifying clusters with arbitrary size and shape, and it handles outliers (e.g. erroneous GPS positions) very well. However, it struggles to find clusters with significantly differing densities. In our application, some ports are substantially busier than others either because of their importance in the trade network or because of the vessel types they specialize in (e.g. containers vs tankers). Since DBSCAN uses a constant density-threshold to find locally dense areas, it may miss some low-traffic ports or lump them together with nearby high-traffic ports. Ideally, instead of looking for areas above a set density threshold, we would like the find contours on the map where ship density suddenly drops. We are therefore in the process of implementing an alternative algorithm to produce our port polygons. The algorithm, called OPTICS (Ankerst et al., 1999), is well-suited for exactly this purpose, and it is likely to improve the automatic identification of port boundaries.

Second, we can significantly improve the forecasting properties of our trade index by predicting ships destination while they are still sailing the oceans. In our current methodology, we only estimate trade flows when ships arrive at a port or leave a port. However, manually entered (unstructured) textual information in the AIS messages about the vessel's destination, the stability of main shipping routes and the history of the vessel's previous voyages enable us to infer the likely destination port. Since voyages between regions of the world often take many weeks, this enhancement of the methodology could increase the forecasting capability from one month to 2-3 months. Again, machine learning algorithms can be very useful to predict ship trajectories from loosely structured data.

Third, more detailed vessel information could increase the precision of our cargo weight estimation. Our current methodology only requires AIS messages and publicly available information, which we see as an advantage, because we do not need to rely on proprietary, expensive and hard-to-verify external information such port calls. Nonetheless, vessels' physical characteristics from ship registers would allow us to use more precise formulas for cargo weight estimation. Our current method assumes that cargo weight changes linearly with the ship's draught, which is equivalent to assuming that the ship's hull is a rectangular cuboid. This is obviously not the case, and in the engineering literature various form coefficients (for example, the block coefficient) are used to describe the shape of the hull (MAN, mimeo). These coefficients would make it possible to adjust our metric tons calculations (Jia, Prakash, and Smith, 2019). Similarly, we are forced to estimate vessels' ballast draught from the observed AIS data, but in principal we could get more precise information from a ship register.

Finally, we must admit that no machine learning algorithm can handle all possible real-life situations. Sometimes ports feature unusual geographies. Oil platforms and floating production, storage and offloading units (FPSOs) in the deep seas, or riverside ports accommodating large ocean-going vessels can present a challenge. Similarly, certain vessel

or cargo types can have unique characteristics. For example, the practice of lightering of oil tankers means that anchoring large oil tankers offload cargo to smaller tankers or barges, which then carry the oil to shore. In these cases, we may double count imported cargo if we register both events as port calls. To handle all these special cases, deep domain knowledge of the particular ports or the specific segment of the maritime shipping industry is needed. Incorporating this kind of expert knowledge should improve our results significantly. However, having an automated system that produces trade nowcasts with minimal human intervention, such as our proposed methodology, is an indispensable prerequisite to any such manual fine-tuning.

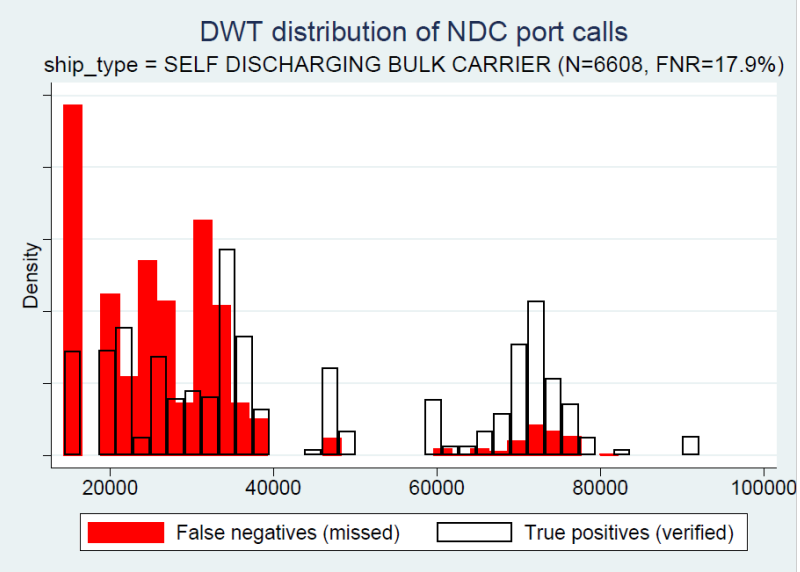
VIII. CONCLUDING REMARKS

In this paper we have proposed an end-to-end solution for the construction of trade volume indicators relying solely on AIS data and publicly available sources of information. Our methodology achieves a good fit with official trade statistics for many countries and for the world in aggregate. We have also shown the usefulness of our approach for sectoral analyses of crude oil trade, and for different event studies where granular data are essential. In all, we think our results are an encouraging first step in using AIS data to monitor world trade activity. Going forward, ongoing refinements of our algorithms, additional data on vessel characteristics, and country-specific knowledge should help further improve the performance of our general approach for several country cases. Finally, monthly official trade volume statistics for more countries than the ones benchmarked here will also be essential to understand the potential our results might have for a very large set of less developed countries that rely heavily on seaborne trade and that might benefit substantially from reliable nowcasting tools.

APPENDIX

A. Further analysis of false negatives

Figure A1. Missed port calls of self-discharging bulk carriers



B. Trade-weighted indices

For some countries, it might be possible to improve upon our baseline approach by using trade weights, e.g. by assigning to the estimated metric tons of cargo of dry bulk carriers a weight equal to the non-oil commodity trade of that specific country, etc. To test this alternative approach, we first constructed a coarse vessel classification. We then mapped this classification to HS 4-digit codes (see Table A1).[27] Then, using COMTRADE data for 2012-2014, we created weights for each coarse vessel type. We then applied these weights when constructing the trade indices.

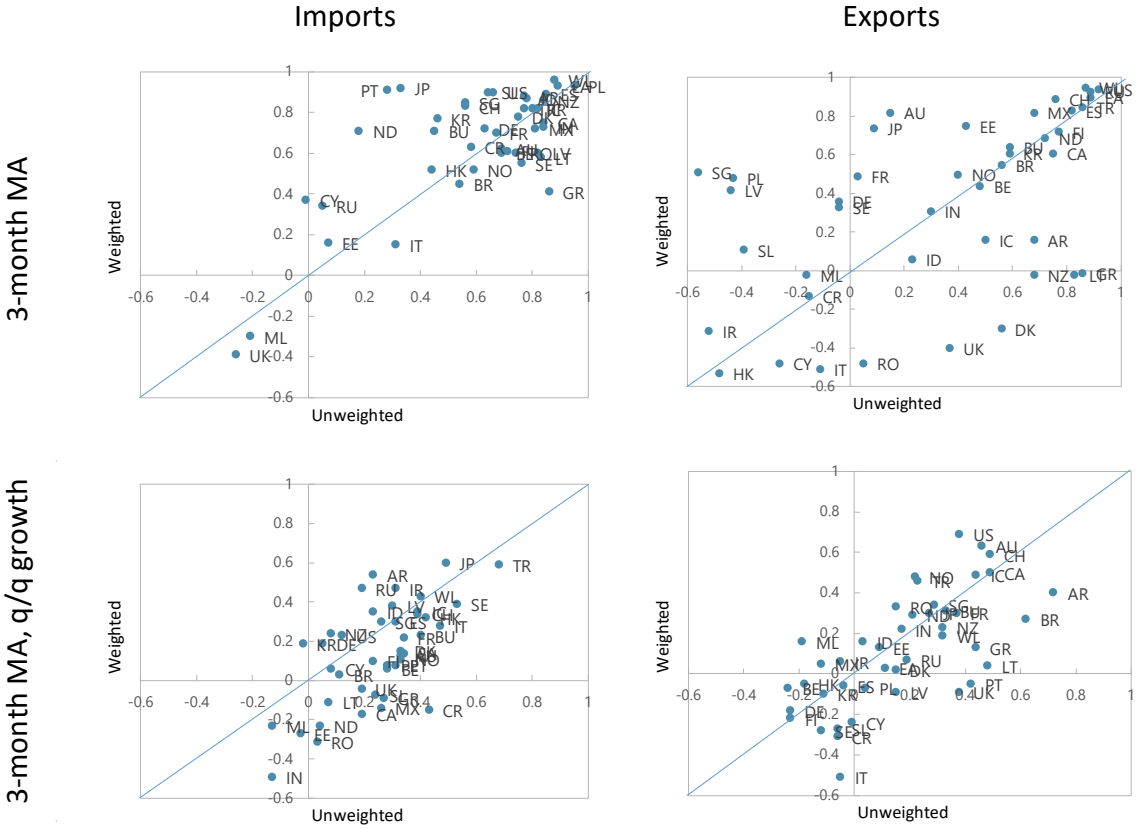
[27] For an alternative mapping from vessel types to HS codes, see Liu et al. (2019).

Table A1. Mapping vessel types to HS codes

Coarse vessel classification	HS codes
Dry bulk	09-14, 17, 25-26, 27 (except 2709-2711), 68, 72-81
Container/General cargo	06, 18-24, 30-67, 69-71, 82-86, 87 (except 8701-8705), 90-97
Oil & chemicals	2709-2710, 28-29
Roll on/roll off	8701-8705
LPG & LNG tankers	2711
Foodstuff (other than dry bulk)	01-05, 07-08, 15-16
Memo:	
Unassigned HS codes	88-89

Figure 11 illustrates how this alternative approach compares to our baseline approach of simply adding up all estimated metric tons of cargo. Our fit improves for some economies, but certainly not for all. Notably, in the case of imports the improvements are less ambiguous when the series are measured in levels. In growth rates, on the other hand, the performance is overall worse. At least when considered in isolation, this alternative approach does not appear to yield unambiguous improvements in our performance.

Figure A2. GTI-CPB correlations: Unweighted v. Weighted GTI Indices



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 & Management, Vol. 44, No. 5, pp. 657-665.

Ankerts, M., M.M. Breunig, H.-P. Kriegel and J. Sander. 1999. "OPTICS: Ordering Points To Identify the Clustering Structure," Proceedings ACM SIGMOD, International Conference on Management of Data.

Arslanalp, S., M. Marini and P. Tumbarello. 2019. "Big Data on Vessel Traffic: Nowcasting Trade Flows in Real Time," IMF Working Paper 19/275.

Brancaccio, G., M. Kalouptsidi and T. Papageorgiou. Forthcoming. "Geography, Transportation and Endogenous Trade Costs," Econometrica.

Breiman, L. 2001. "Random forests," Machine learning 5-32.

Catao, L. and G.M. Milesi-Ferretti. 2014. "External liabilities and crises," Journal of International Economics, Vol. 94 (1).

Cortes, C. and V. Vapnik. 1995. "Support-Vector Networks," Machine Learning, 20, 273-297.

Ester, M., H.-P. Kriegel, J. Sander, X. Xu. 1996. "A density-based algorithm for discovering clusters in large spatial databases with noise," Proceedings of the Second International Conference on Knowledge Discovery and Data Mining.

Flanders Marine Institute (2018). Maritime Boundaries Geodatabase: Maritime Boundaries and Exclusive Economic Zones (200NM), version 10.

Frankel, J.A. and A.K. Rose. 1996. "Currency Crashes in Emerging Markets: An Empirical Treatment," International Finance Discussion Papers, Federal Reserve Board of Governors.

Friedman, J., T. Hastie, and R. Tibshirani. 2001. The elements of statistical learning. New York: Springer series in statistics.

Forgy, E.W. 1965. "Cluster analysis of multivariate data: efficiency versus interpretability of classifications," Biometrics. 21 (3): 768–769.

Getis, A. 1992. “The Analysis of Spatial Association by Use of Distance Statistics,” Geographical Analysis, Vol. 24, No. 3.

Gopinath, G. 2020. “The Great Lockdown: Worst Economic Downturn Since the Great Depression,” IMF Blog, April 14, 2020.

Hammer, C.L., D.C. Kostroch and G. Quirós. 2017. “Big Data: Potential, Challenges, and Statistical Implications,” IMF Staff Discussion Note 17/06.

Heiland, I., A. Moxnes, K.H. Ulltveit-Moe and Y. Zi. 2019. “Trade From Space: Shipping Networks and The Global Implications of Local Shocks,” mimeo.

Ho, T.K. 1995. “Random Decision Forests,” the 3rd International Conference on Document Analysis and Recognition. Montreal, QC: IEEE. 278–282.

Ho, T.K. 1998. “The random subspace method for constructing decision forests,” IEEE transactions on pattern analysis and machine intelligence 832-844.

Kaminsky, G., S. Lizondo and C.M. Reinhart. 1998. “Leading Indicators of Currency Crises,” IMF Staff Papers, Vol. 45 (1).

Kishore, N., D. Marques, A. Mahmud, M.V. Kiang, Ii Rodriguez, A. Fuller, P. Ebner, C. Sorensen, F. Racy, J. Lemery, L. Maas, J. Leaning, et al. 2018. “Mortality in Puerto Rico after Hurricane Maria,” New England Journal of Medicine, 379, 2.

Jia, H., V. Prakash, T. Smith. 2019. “Estimating vessel payloads in bulk shipping using AIS data,” Int. J. Shipping and Transport Logistics, Vol. 11, No. 1, 2019 25

Kohavi, Ron. 1995. “A study of cross-validation and bootstrap for accuracy estimation and model selection,” the International Joint Conference on Artificial Intelligence.

Liu, H., Z. Meng, Z. Lv et al. 2019. “Emissions and health impacts from global shipping embodied in US–China bilateral trade,” Nat Sustain 2, 1027–1033.

Lloyd, S. 1982. “Least squares quantization in PCM,” IEEE Transactions on Information Theory, 28 (2): 129–137.

MAN. Mimeo. “Basic Principles of Ship Propulsion,” available at https://spain.mandieselturbo.com/docs/librariesprovider10/sistemas-propulsivos-marinos/basic-principles-of-ship-propulsion.pdf

Natale, F., M. Gibin, A. Alessandrini, M. Vespe and A. Paulrud. 2015. “Mapping fishing effort through AIS data,” PloS one, 10(6).

National Geospatial-Intelligence Agency (NGA). 2017. World Port Index 2017.

Smith, A.B. 2020. “2010-2019: A landmark decade of U.S. billion-dollar weather and climate disasters,” National Oceanic and Atmospheric Administration.

Snoek, J., H. Larochelle, R. P. Adams. 2012. “Practical Bayesian optimization of machine learning algorithms,” Advances in neural information processing systems 2951-2959.

Spiliopoulos G., D. Zissis and K. Chatzikokolakis. 2018. “A Big Data Driven Approach to Extracting Global Trade Patterns,” in: Doulkeridis C., Vouros G., Qu Q., Wang S. (eds) Mobility Analytics for Spatio-Temporal and Social Data. MATES 2017. Lecture Notes in Computer Science, vol 10731.

United Nations Conference on Trade and Development (UNCTAD). 2017. Review of Maritime Transport.