ORIGINAL ARTICLE

Can maritime big data be applied to shipping industry analysis? Focussing on commodities and vessel sizes of dry bulk carriers

Kei Kanamoto[1] · Liwen Murong[1] · Minato Nakashima[1] · Ryuichi Shibasaki[2]

Abstract

Enriched navigational information provided by an automatic identification system (AIS) could improve the estimation accuracy of trade patterns analysis by using different data sources. This paper estimates the global trade flow pattern of dry bulk cargo by commodity, namely iron ore, coal, grains, fertilisers, and iron and steel. We use AIS data and the information on commodities handled in ports, estimated by using a two-tiered Geohash geocoding. Estimation results are accurate at country level except for iron and steel. The results are used to quantify the impact of the previously identified variables on vessel size selection by regression analysis and a multinomial logit model. Finally, our model is used to forecast the future shipping demand by vessel type and commodity.

Keywords AIS · Dry bulk · Port-based global cargo flow · Iron ore · Coal · Grains · Vessel size · AXS dry · Vessel movement

1 Introduction

Resource (commodities) trade is critical to world economic growth because it supports the economic activities that consume energy. Given that international bulk shipping is highly volatile due to its dependence on energy supply and demand, it

✉ Ryuichi Shibasaki
  shibasaki@tmi.t.u-tokyo.ac.jp

[1] Department of Systems Innovation, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-8656, Japan

[2] Resilience Engineering Research Center, Department of Technology Management for Innovation, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-8656, Japan

is crucial for carriers to comprehensively understand global trade flows. Although statistics on international trade are admittedly widely available, such as the UN Comtrade database, they are generally limited to country level.

The developments in the automatic identification system (AIS) and satellite communication capabilities have allowed historical data to build up on ports of call, as well as sailing information on vessel movements. Although AIS data have been used in a variety of research applications, research on logistics to estimate global trade flows and transport patterns for each type of cargo remains scarce because the type of cargo and its handling at ports cannot be ascertained directly from the data. In this paper, we focus on dry bulk shipping, which accounts for 44% of seaborne trade volume. Dry bulk carriers carry three major bulk cargoes—iron ore, coal and grains—and minor bulks (e.g., wood, fertilisers, and iron and steel). There is a significant number of dry bulk carriers, calling at many ports. Moreover, a vessel often transports several commodities, therefore estimating the commodities carried by dry bulk carriers with AIS data is difficult.

In the first part of this study, we estimate the cargo flows for dry bulk carriers by commodity on a port-to-port basis using AIS data and berth information in ports. The second part of this research explores the potential factors related to the choice of vessel size, by using the already estimated cargo flows. Multiple regression and a multi-nominal-logit model are applied. Finally, future shipping demand by vessel size and by commodity is estimated by the developed model.

2 Literature review

AIS has been used in various maritime analyses, as summarised in Tu et al. (2017) and Yang et al. (2019). Lechtenberg et al. (2019) analysed trading vessels using AIS information such as estimated time of arrival, next destination and anchoring time. Adland et al. (2018) analysed the effect of various factors related to a voyage on vessel speed by using multiple regression analysis in the case of dry bulk shipping, with vessel speeds extracted from AIS data.

However, few studies have accurately analysed transport volume and shipping cargo using AIS data. Adland et al. (2017) estimated the export quantity of crude oil by country based on AIS data, and compared the results with existing statistics. Shibasaki et al. (2020) estimated global cargo flows on a port basis for liquefied natural gas (LNG) carriers, achieving high estimation accuracy. Arifin et al. (2018) forecasted port-to-port global cargo flows, but the accuracy of their estimation was confirmed only for iron ore and coal shipping between Japan and Australia. Arslanalp et al. (2019) attempted to estimate the amount of trade for each vessel type in real time by using AIS data but did not estimate the contents (cargo transported) in dry bulk carriers. Thus, although some studies have estimated global cargo flows, no studies have estimated with high accuracy the multi-commodity global cargo flows transported by dry bulk carriers.

In regard to the analysis of vessel size, Wada et al. (2018) estimated the demand for newly built bulk carriers focussing on iron ore, coal and grains, and on a flow prediction model, order prediction model, construction model and vessel allocation model.

Borthen et al. (2018) developed a genetic-algorithm-based method to address the vessel assignment problem for platform supply vessels in Norway, with the aim of finding the optimal size. Santos and Guedes Soares (2017) established a vessel size selection model to optimise freight and shipping costs for roll-on/roll-off (RO-RO) vessels, sailing from the Portuguese Port of Leixões to the Port of Rotterdam in the Netherlands. The authors considered a series of factors, including fuel costs, emission control areas and limitation of vessel drafts and lengths. Ko (2011) proposed a dynamic factor model related to bulk shipping indices and used maximum likelihood estimation. The author revealed the synchronicity and idiosyncrasy of dry bulk sub-markets to unobserved dynamic common factors, i.e. the extent to which dry bulk sub-markets conform with changes in common factors. His findings revealed the susceptibility of vessel type to the global market situation.

Moreover, there are few in-depth studies related to the analysis of vessel size focussing on the dry bulk shipping market. Goulielmos (2013) compared panamax and capesize vessels to test whether the larger vessel entailed a higher risk. Based on time series analysis of time contracts, he found that the risk was lower for larger-sized vessels. Alizadeh et al. (2016) investigated scrapping probability in the dry bulk market and established logit models based on vessel type, confirming the influence of market variables including expected recovery time, bunker price, freight market volatility and interest rate on determining the probability of scrapping. Jia et al. (2019) proposed a method to estimate vessel payload of bulk carriers, by using drafts information from AIS. The effectiveness of the proposed method was verified by multiple regression analysis. Moreover, the relationship of deadweight tonnage (DWT) of bulk carriers and cargo payload was revealed. According to our review of the extant literature, no study so far has been carried out to estimate vessel size for each commodity in dry bulk shipping, the reason being the unavailability of data. Based on a database of commodities information, we explore potential factors and develop a commodity-based vessel size selection model.

3 Data overview

3.1 AIS data

We use 2016 AIS data of the Seasearcher database of Lloyd's List Intelligence. AIS data generally include static information (e.g. IMO number, and vessel length and width), dynamic information (e.g. latitude and longitude position, course over ground and speed over ground) and voyage-related information (e.g. draft and destination). Dynamic data is recorded every 2 s to 10 s, depending on the speed of the sailing vessel, and every 3 min at anchorage. Because the voyage-related information is input manually, erroneous input or instances of missing data occur.

Table 1 Range of DWT and the number of dry bulk carriers for each vessel type. Source: authors based on AIS data

Vessel type	Vessel capacity (10^3 DWT)	Number of vessels (2016)
Very large ore carrier (VLOC)	~220	223
Capesize	~120	1424
Neo-panamax	~85	556
Panamax	~65	2165
Handymax	~40	3383
Handysize	~20	2663
Minibulk	~20	1149

3.2 Berth information

AXS Dry data, provided by AXS Marine, is a database of integrated data on dry bulk carriers. Based on vessels, ports and contract information, AXS Dry data records the pair of the arrival- and departure port, for each voyage, with commodity information and draft changes. We use AXS Dry data to develop the list of commodities handled at each berth of a port.

3.3 Vessel size classification

The size of dry bulk carriers is classified into seven categories (Table 1). The *very large ore carrier* (VLOC) is the biggest vessel size among dry bulk carriers. Vessels too large to transit the Suez Canal (SC) must sail around the Cape of Good Hope—these are classified as *capesize* vessels. The original criterion for a *panamax* vessel was the width of the lock chambers of the Panama Canal (PC)—their deadweight ranged between 65,000 tons to 85,000 tons. The enlargement of the PC in 2016 made the transit of larger vessels possible. These are now classified as *neo-panamax* vessels. The significance of draft and other limitations of smaller size vessels, such as the *handymax* and the *handysize* vessels, is less relevant.

4 Estimation method of global cargo flow by commodity

4.1 Overview

The basic purpose of estimation is to develop a commodity list of what is handled in each port, as in Shibasaki et al. (2020). The list is developed based on AXS Dry data, and the commodity is confirmed by matching the list to the export and import ports. Note that several commodities are often handled at the same port, or even at the same berth. Therefore, it is necessary to identify a berth for each commodity. We

do this by using AIS data, and we develop a commodity list for each berth wherever the port handles multi-commodities. Further, in cases where a berth handles multi-commodities, additional treatment is necessary which is explained below.

4.2 Determining the loading and discharging ports

Other than for loading and discharging cargo, vessels stop for refuelling, repairs and offshore waiting. In this study, by using the draft information of AIS data, we assume that if the rate of draft increase was 20% or more than the maximum change of draft at each stopping point, the cargo was loaded, and if the rate of draft decrease was 20% or more, the cargo was discharged. We assume these conditions for determining the loading and discharging occurrences because draft can slightly change even when refuelling or taking ballast water.

In bulk shipping, the cargo is often loaded or discharged sequentially in several ports because of the capacity constraints of ports. However, as the accuracy of draft information (which is manually input) is relatively low, we do not consider the loading or discharging of cargo at multiple ports. Alternatively, we assume that all cargo is loaded or discharged in the largest port in terms of the magnitude of draft change.

4.3 Extracting record of ports of call from AIS data[1]

To estimate cargo flows, the loading and discharging ports should be extracted from AIS data, including the following information:

- Port name.
- Arrival time, departure time and time at berth.
- Location of handling berth.
- Pre-arrival draft, post-departure draft and the magnitude of draft change.

We use Geohash for extraction. This is a geocoding technique that manages all positional coordinates on the Earth (i.e. latitude and longitude) by allocating the character codes of plural digits. One of the features of Geohash is its flexibility; for example, if the number of digits of Geohash increases, division becomes smaller, and the analytical precision increases. In this study, a four-digit Geohash (Geohash 4) with 10 km order is used when identifying the port, and a seven-digit Geohash (Geohash 7) with 100 m order is used when identifying the berth. The first reason why this two-tiered identification is applied is to prevent incremental change in information of those berthing or mooring places, such as when vessels handle the cargo in a ship-to-ship (STS) transfer (Fig. 1). Another reason is to measure draft changes accurately. In this paper, draft variation is extracted by comparing the arrival draft at each stopping point. However, the time of changing the draft recorded in AIS data is often behind the actual time of draft change because the draft data is manually input.

---

[1] The non-technical reader can skip this section without loss of argument continuity.
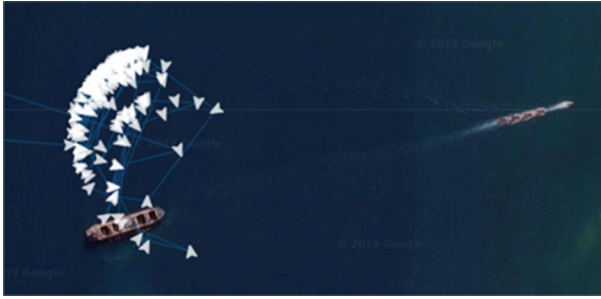
Fig. 1 AIS data in the case of ship-to-ship transfer. Source: Authors

If the staying point is too subdivided by using Geohash 7, the draft change is not associated with the staying point where cargoes are loaded or discharged.

Therefore, the staying points are extracted from AIS data by using the following procedure:

Step 1 Interpolate the missing values on vessel speed and draft with the value immediately before.

Step 2 Generate Geohash 4 and Geohash 7 from the position coordinates (latitude and longitude information).

Step 3 Define the case whereby a vessel is continuously in the same Geohash 4 for 5 hours or more as 'staying' and extract the staying start time and end time.

Step 4 Assume the most-counted draft during the staying period in Geohash 4 (acquired in step 3) as before-arrival draft and associate Geohash 7, whose staying period is the longest with the berthing or mooring point. Furthermore, assume the draft of the next staying point (in Geohash 4) as post-starting draft.

Step 5 Determine the export and import berths based on the difference in the pre-arrival and post-starting draft.

4.4 Developing the commodity list in each port or berth

The commodity list in each port or berth is developed based on the cargo information handled in each staying point, which is obtained by associating the staying points (in Geohash 7) extracted from AIS data with ports of call acquired from AXS Dry data, which contains commodity information. Among the 9856 vessels that are common in both AIS and AXS Dry data, 64,602 shipping records including the pairs of export and import ports included in AXS Dry data, are associated with a staying point (in Geohash 7), on the condition that the staying period (in Geohash 4) acquired from AIS data overlaps with AXS Dry data.

As a result, the export ports list consists of 830 single-commodity ports (ports where handling one commodity is identifiable without reference to berths) and 544 multi-commodity ports. Berths in multi-commodity ports are subdivided

Table 2 Number of export and import ports and berths handling a single commodity. Source: Authors

Commodity	Single-commodity ports		Single-commodity berths in multi-commodity ports	
	Export	Import	Export berths	Import berths
Agri-products	40	21	233	266
Breakbulk	172	83	649	707
Bulk	256	156	1076	1347
Coal	76	112	2225	1505
Fertilisers	52	37	363	572
Grains	95	64	1128	1277
Iron ore	48	19	520	502
Iron and steel	91	45	661	918
Sum	830	537	6855	7094

Table 3 Number of combination of commodities for multi-commodity exporting berths

First commodity	Second commodity								Sum
	Agri-products	Break-bulk	Bulk	Coal	Fertilisers	Grains	Iron ore	Iron and steel	
Agri-products	0	5	3	2	0	35	0	1	46
Breakbulk	3	0	14	3	6	10		8	44
Bulk	6	20	0	25	30	21	21	20	143
Coal	0	9	20	0	1	5	6	7	48
Fertilisers	1	5	22	0	0	7	1	8	44
Grains	68	7	31	6	13	0	3	11	139
Iron ore	1	3	23	7	2	9	0	10	55
Iron and steel	2	5	24	10	8	20	2	0	71
Sum	81	54	137	53	60	107	33	65	590

*Including 61 berths where more than three types of commodities are exported

into 6855 single-commodity berths (in terms of Geohash 7) and 590 multi-commodity berths. The import ports list comprises 537 single-commodity ports and 941 multi-commodity ports; berths in multi-commodity ports are subdivided into 7094 single-commodity berths and 2524 multi-commodity berths. The breakdown by commodity of the single-commodity ports and single-commodity berths are presented in Table 2. Additionally, the combinations of the commodities of the multi-commodity exporting and importing berths are presented in Tables 3 and 4.

Table 4 Number of combination of commodities for multi-commodity importing berths. Source: Authors

First commodity	Second commodity								Sum
	Agri-products	Break-bulk	Bulk	Coal	Fertilisers	Grains	Iron ore	Iron and steel	
Agri-products	0	9	21	7	10	66	3	18	134
Breakbulk	16	0	70	28	27	30	4	25	200
Bulk	28	76	0	118	73	81	66	74	516
Coal	15	38	187	0	22	26	95	53	436
Fertilisers	19	17	66	15	0	43	3	28	191
Grains	163	36	104	31	77	0	12	79	502
Iron ore	4	8	72	183	11	7	0	8	293
Iron and steel	17	36	64	27	45	42	18	0	249
Sum	262	220	584	409	265	295	201	285	2521

Including 812 berths where more than three types of commodities are exported

Table 5 Commodity estimate reference of export and import berth combinations. Source: Authors

Import Export	Berth D (Single-commodity)	Berth E (Multi-commodity)	Berth F (Unknown)
Berth A (single-commodity)	Berth A's commodity	Berth A's commodity	Berth A's commodity
Berth B (multi-commodity)	Berth D's commodity	*Common cargo between berth B and E* *or* *Berth B's commodity*	Berth B's commodity
Berth C (unknown)	Berth D's commodity	Berth E's commodity	Unknown

4.5 Estimating commodity for laden shipping

The commodity estimate for laden shipping is done by simultaneously matching the export and import ports based on the commodity list. If either port is a single-commodity one, *that* commodity is then assigned. If both are multi-commodity ports, then the commodity(-ies) handled at a certain berth is (are) confirmed. Table 5 presents a matrix for estimating the commodity of each laden shipping, based on the commodity information of the export and import berths. Even at berth level, if a single-commodity berth and a multi-commodity berth are mixed, the commodity connected to the single-commodity berth is assigned. In rare cases, although both the export and import port are single-commodity ones (or single-commodity berths), the commodity is inconsistent. In that case, the commodity of the export port is used (italicized in Table 5), because this is more likely to be certain due to the limited

Table 6 Estimated vessel capacity in laden shipping by commodity and comparison with the observed trade amount. Source: Authors

	AIS (mil. ton)	AXS (mil. ton)	AIS/AXS (%)	Clarkson (mil. ton)	AIS/Clarkson (%)	IHS (mil. ton)	AIS/IHS (%)
Iron ore	1651	1482	111	1418	116	1511	109
Coal	1257	1082	116	1141	110	1165	108
Grains	517	451	115	450	115	416	124
Fertilisers	149	149	100	164	91	135	110
Iron and steel	185	222	84	408	45	315	59

Table 7 Estimated average load factor of each commodity. Source: Authors

Commodity	Average load factor (%)
Iron ore	86
Coal	91
Grains	87
Fertilisers	90

number of export ports. In the case where both export and import berths are multi-commodity ones, if the commodity cannot be determined, then the most intensively exported commodity at the export side is selected.

5 Estimation results of global cargo flows

5.1 Estimation of load factors

Port-to-port dry bulk cargo flows are estimated for iron ore, coal and grains, as well as for two minor bulks, i.e. fertilisers, and iron and steel. In general, global trade statistics are available only at national level, therefore the reasonableness of our estimation results is examined on a country basis by using existing statistics, including the Dry Bulk Trade Outlook, provided by Clarkson Research, and the GTA Forecasting, provided by IHS Markit.

Table 6 presents the estimated total amount of vessel capacity (DWT) in laden shipping for each commodity. The table reveals that, except for iron and steel, the estimated value is normally higher than the total global trade amount, calculated by AXS, Clarkson and IHS. The reason for the low coverage for iron and steel is that its definition may differ between AIS data and statistical sources.

Based on the difference between total vessel capacity and trade amount provided by Clarkson or IHS, the average load factor per commodity is calculated (except for iron and steel) in Table 7. In a different approach, one might separately estimate load
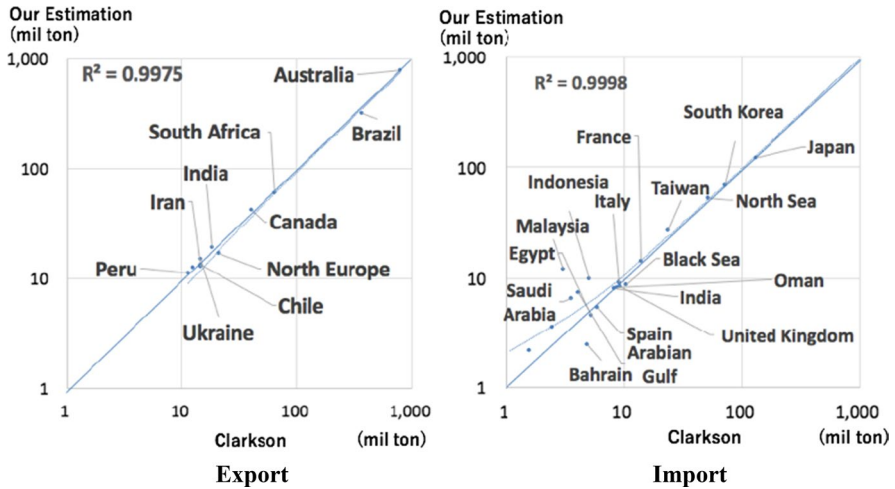
Fig. 2 Comparison of observed and estimated amounts of iron ore by country. Source: Authors

factors from draft changes in AIS data, but this would require a much more elaborate examination of the reliability of manually input draft data. Therefore, we have used the average load factors as a *prototype* of estimation and then estimated the amount of laden shipping for each commodity by multiplying the vessel capacity of each laden shipping by the average load factors.

5.2 Comparison of estimated and observed export and import amounts by country

5.2.1 Iron ore

Figure 2 presents the comparisons of the observed (provided by Clarkson) and estimated exported and imported amounts of iron ore by country. The estimated amount agrees with the observed amount in every country for both exports and imports. The coefficient of determination R^2 between the estimated and observed amount is 0.9975 for exporting countries and 0.9998 for importing countries.

5.2.2 Coal

Figure 3 presents the comparisons of the observed (provided by Clarkson) and the estimated exported and imported amounts of coal. The estimated amount also agrees with the observed amount in every country for both exports and imports. The coefficient of determination R^2 is 0.9829 for exporting countries and 0.9915 for importing countries. The imported amounts in some Southeast Asian countries, including the Philippines, Thailand and Vietnam, are underestimated.
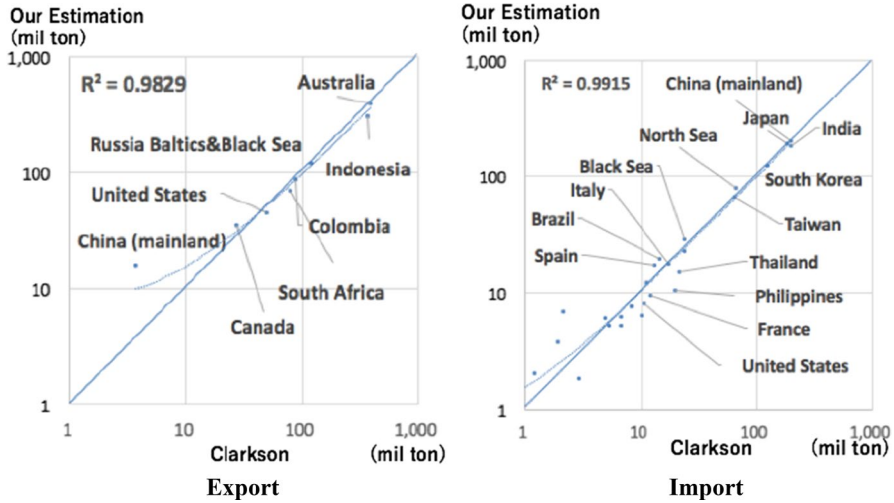
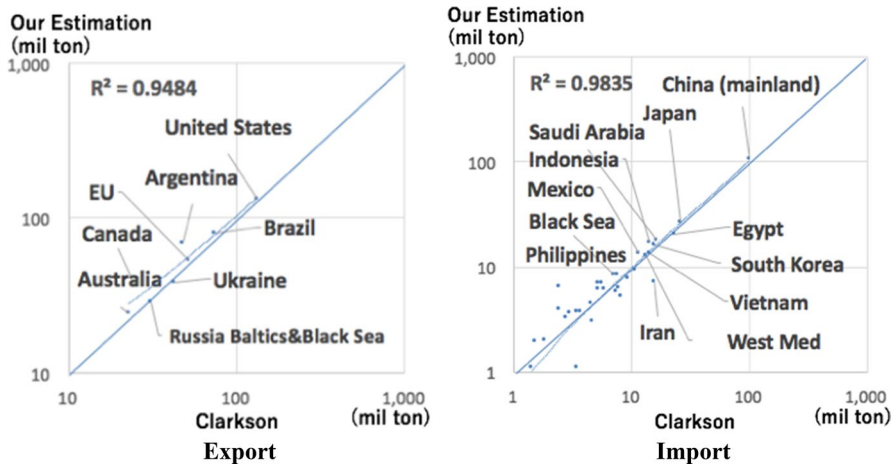**Fig. 3** Comparison of observed and estimated amounts of coal by country. Source: Authors

Fig. 4 Comparison of observed and estimated amounts of grains by country. Source: Authors

5.2.3 Grains

Figure 4 presents the comparisons of the observed (provided by Clarkson) and the estimated exported and imported amounts of grains. The estimated amount agrees with the observed amount in every country for both exports and imports. The coefficient of determination R^2 is 0.9484 for exporting countries and 0.9834 for importing countries. The imported amount in Iran is significantly underestimated.
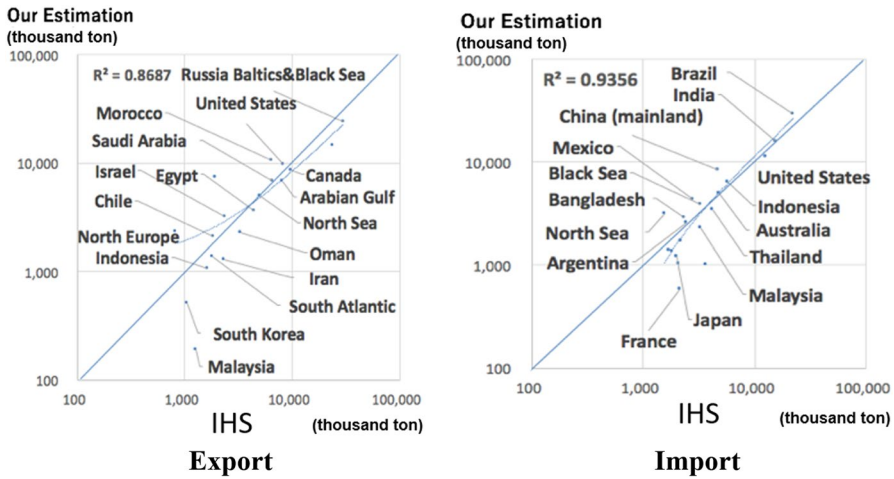
Fig. 5 Comparison of observed and estimated amounts of fertilisers by country. Source: Authors

5.2.4 Fertilisers

Figure 5 presents the comparisons of the observed (provided by IHS) and estimated exported and imported amounts of fertilisers. The imported amount is accurately estimated: the coefficient of determination R^2 is 0.8687 for exporting countries and 0.9356 for importing countries. The exported amount is also accurately estimated, except for countries with a small amount of trade.

5.2.5 Iron and steel

We compare the observed (provided by IHS) and estimated exported and imported amounts for iron and steel. However, the coefficient of determination R^2 is 0.8697 for exporting countries and 0.2448 for importing countries. The reason for the difference in estimation accuracy is that the exported amount from China, the largest exporting country of iron and steel, is accurately estimated, however the overall estimation accuracy is poor. Therefore, this category is eliminated from all following analyses.

5.3 Comparison of monthly estimated and observed amounts

Monthly trade statistics are available for some countries but not all. Figure 6 compares the estimated and observed amounts (provided by the Japan Maritime Center, 2017) of imports of iron ore and coal in Japan by month in 2016 (the ratio of the estimated amount to the observed amount). The difference from the observed amount in January for iron ore can be explained by the absence of export information at the end of 2015 from AIS data. Furthermore, similar to LNG imports in Shibasaki et al. (2020), observed and estimated amounts tend to deviate slightly in July and August,
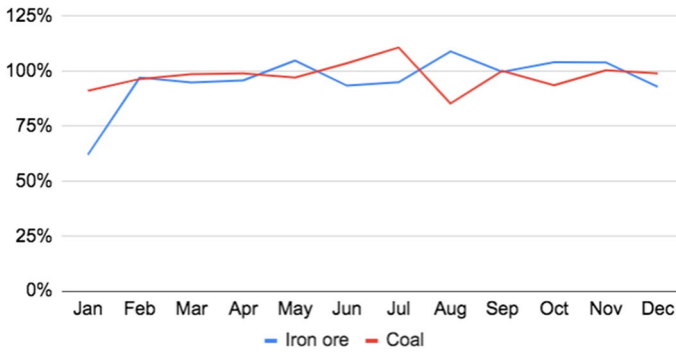
Fig. 6 Comparison of the observed and estimated monthly iron ore and coal imports in Japanese ports (2016). Source: Authors

Table 8 Comparison of observed and estimated coal imports in selected Japanese ports. Source: Authors

Port	Observed amount (A) thousand ton	Estimated amount (B) thousand ton	Estimation accuracy (B/A) (%)
Kinuura	9784	10,676	109
Fukuyama	9068	8094	89
Kita-Kyushu	8790	6058	69
Tokuyama-Kudamatsu	7960	7333	92
Oita	7132	11,699	164
Mizushima	7074	6969	99
Tachibana	6647	6241	94
Matsuura	6456	6903	107
Kawasaki	6408	1258	20
Onahama	6362	6530	103
Japan total	181,236	194,076	107

compared with other months. This suggests the presence of seasonal characteristics in terms of the differences in port arrival times and customs procedures, but our estimates generally agree with the observed amounts, even on a monthly basis.

5.4 Comparison of estimated and observed amounts by port

Trade statistics for each port are available for some countries, but not all. Table 8 compares the estimated and observed import amounts (obtained from Japanese port statistics produced by the Ministry of Land, Infrastructure, Transport and Tourism, Japan) of coal in the top ten Japanese ports. Although the import amount in almost all ports is estimated with high accuracy, for Oita Port it is overestimated, and for Kitakyushu and Kawasaki Ports it is underestimated. For example, in our estimation, imports at Kawasaki Port are often recorded in Yokosuka Port located in the mouth of Tokyo Bay, but

the observed volume in Yokosuka Port is zero. The source of these errors is considered to be the different timing of draft changes and a shorter staying period than the threshold (i.e. 5 h). Furthermore, sums of the estimated amounts in Oita Port and Kitakyushu Port roughly agree with the observed amounts (the difference ratio is 115%). Thus, a problem exists in estimating destinations whenever cargo is imported into multiple ports (discussed in the next subsection).

5.5 Challenges of the proposed method

We reviewed AIS data for countries whose estimated amount differed significantly from the observed amounts, finding several points for improvement, including the following:

5.5.1 Successive loading and discharging at multiple ports

As discussed in the previous subsection, there are cases of loading or discharging at more than two ports in succession. In these cases, the estimated and observed amounts may differ because this study assumes all cargo is loaded or discharged at the berth where draft changes are the highest. For example, for the import of iron ore in Bahrain, vessels often enter the port after discharging part of their cargo in the Qatar area, which causes the underestimation of Bahrain's import amounts.

5.5.2 Wrong communication of AIS data

Due to reasons such as congestion, weather and communications equipment, the possibility exists that AIS data cannot be properly received. For example, in the sea area around the Philippines and Vietnam, sometimes AIS data cannot be properly collected because of data confusion due to the large number of vessels. As a result, there may be wrong estimations at import ports in Indonesia for cargo that is actually discharged in the Philippines or Vietnam.

5.5.3 Turning off AIS

Although AIS should be installed on vessels according to regulations, the crew could purposely interrupt the transmission. For example, in an area off the Somali coast, AIS devices were purposely turned off to avoid ships being discovered by Somali pirates. Furthermore, AIS is sometimes manually turned off when vessels enter sanctioned countries such as North Korea and Iran. In this case, the staying points cannot be extracted and estimation accuracy deteriorates.

6 Modelling vessel size selection based on the estimated cargo flow

In this section, the analysis concerning vessel size and commodities in dry bulk cargo shipping is conducted using the estimated global cargo flows. First, variables influencing the choice of vessel size and their correlations are explored. The contribution of each variable is then quantitatively examined through multiple regression and a multinomial logit model. Notably, iron and steel is excluded from the analysis because the estimation results do not describe the actual trade well.

6.1 Variables of vessel size selection

Based on earlier works such as that by Stopford (2008) and the results of our survey, we have considered five types of variable in the following analysis for vessel size choice, consisting of trade volume (which is the main target in analyses such as Arifin et al. 2018), voyage distance (Adland et al. 2018), port entry restrictions (Jia et al. 2019), limitations of shipping routes (Dinwoodie et al. 2014) and the dry bulk shipping freight index (Wu et al. 2018).

Trade volume is considered for each commodity. The use of large-size vessels between countries with large trade volumes can reduce shipping costs per ton by reducing the frequency of voyages. Port entry restrictions are considered in both loading and discharging ports. Large-size vessels must raise draft by reducing the load factor to be able to enter in ports with draft limitations. Among several characteristics of vessel size such as length, width and water displacement, we use draft as a representative index of port entry restrictions due to lack of data. Regarding limitations of shipping routes, we consider only the Suez Canal and the Panama Canal because these canals cannot accommodate the largest vessels. However, shipping companies have an incentive to build larger vessels to save on shipping costs because the Suez Canal toll per ton decreases as vessel size increases. In this study, the shares of SC and PC transits are considered as variables, estimated for each pair of loading and discharging ports by dividing the annual number of vessels to pass through canals by the total annual number of vessel transits in that pair. In regard to the dry bulk shipping freight index, we employ the rates of the four Baltic Exchange sub-indices, i.e. Baltic Capesize Index (BCI), Baltic Panamax Index (BPI), Baltic Supramax Index (BSI) and Baltic Handysize Index (BHsI) of the Baltic Exchange Dry Index (BDI), to avoid multicollinearity.

Figure 7 shows the correlation matrix of the above variables, including vessel size (DWT). Based on the correlation coefficients between the explanatory variables, the BSI is not used in the following analysis, being collinear with BHsI.

	DWT	Voyage distance	Loading max draft	Discharging max draft	Panama share	Suez share	BCI/ BDI	BPI/ BDI	BSI/ BDI	BHsI/ BDI
DWT	1.00	0.14	0.54	0.50	-0.06	-0.03	-0.01	-0.02	0.01	0.01
Voyage distance		1.00	0.17	0.06	0.04	0.08	0.01	-0.01	-0.01	0.00
Loading max draft			1.00	0.24	-0.02	-0.03	-0.02	0.01	0.01	0.01
Discharging max draft				1.00	-0.07	-0.06	0.00	0.00	0.00	0.00
Panama share					1.00	-0.01	-0.01	0.02	0.01	0.00
Suez share						1.00	-0.01	0.00	0.02	0.01
BCI/BDI							1.00	-0.42	-0.83	-0.88
BPI/BDI								1.00	0.23	0.11
BSI/BDI									1.00	0.85
BHsI/BDI										1.00
Trade volume — Iron ore	0.28	-0.21	-0.07	0.41	-0.04	-0.13	0.01	0.00	0.00	-0.01
Trade volume — Coal	0.16	-0.05	0.19	0.00	-0.11	-0.08	0.02	-0.01	-0.01	-0.01
Trade volume — Grains	0.42	0.18	0.05	0.37	0.18	-0.14	0.01	-0.04	0.02	-0.03
Trade volume — Fertilizers	0.23	0.03	0.06	0.18	-0.09	-0.10	0.02	-0.01	0.00	-0.02

Fig. 7 Correlation matrix of variables in vessel size selection. Source: Authors

6.2 Multiple regression model

The factors mentioned above are used as explanatory variables in the multiple regression model set by commodity, whereas DWT capacity is used as the dependent variable.

$$DWT_{crsn} = \alpha_{c0} + \alpha_{c1} \cdot V_{cij} + \alpha_{c2} \cdot D_{rs} + \alpha_{c3} \cdot o_r + \alpha_{c4} \cdot u_s + \alpha_{c5} \cdot S_{rs} + \alpha_{c6} \cdot P_{rs}$$
$$+ \alpha_{c7} \cdot \left(BCI_n/BDI_n \right) + \alpha_{c8} \cdot \left(BPI_n/BDI_n \right) + \alpha_{c9} \cdot \left(BHsI_n/BDI_n \right),$$

(1)

where DWT_{crsn} is the DWT capacity of vessel n employed for transporting commodity c from loading port r to discharging port s; V_{cij} is the annual trade volume from country i to country j for commodity c; D_{rs} is the voyage distance from loading port r to discharging port s; o_r and u_s are the draft limits for loading port r and discharging port s, respectively; S_{rs} and P_{rs} are observed shares in vessel transits of the SC and the PC for each port pair, respectively; BCI_n, BPI_n, $BHsI_n$, BDI_n are the daily values of the four indices, when vessel n departed from loading port; and α_{cm} are the coefficients of each explanatory variable m ($m = 0, 1, \ldots 9$) for commodity c.

Figure 8 presents the estimated results and the t-values of each explanatory variable, the values of R^2 and the effect size (Cohen's f^2) by commodity. Cohen's f^2 is defined as follows:

$$f^2 = \frac{R^2}{1 - R^2}.$$

(2)

Dependent variable: DWT	Iron ore	Coal	Grains	Fertilizers
α_{c0}: constant	-146,000 (-7.36)	-25.900 (-2.05)	35,300 (4.97)	10,600 (1.11)
α_{c1}: Trade volume (kton)	0.0480 (23.76)	0.137 (16.22)	0.607 (35.99)	3.51 (12.06)
α_{c2}: Voyage distance (km)	0.956 (11.58)	0.678 (14.49)	0.441 (19.81)	0.283 (9.78)
α_{c3}: Loading max draft (m)	10,400 (57.79)	2,120 (22.32)	759 (16.08)	1,060 (15.93)
α_{c4}: Discharging max draft (m)	5,570 (33.23)	3,900 (41.53)	899 (17.65)	528 (8.22)
α_{c5}: Panama share	-56,400 (-3.96)	-1,620 (-0.50)	-3,630 (-4.79)	1,300 (1.10)
α_{c6}: Suez share	9,980 (2.35)	18,900 (4.12)	-3,690 (-4.87)	9,010 (7.21)
α_{c7}: BCI/BDI	-3,140 (-1.09)	-2,420 (-1.30)	-2,420 (-2.31)	503 (0.35)
α_{c8}: BPI/BDI	-16,500 (-2.92)	-1,060 (-0.29)	-4,910 (-2.39)	-1,570 (-0.56)
α_{c9}: BHsI/BDI	-15,000 (-0.74)	-2,510 (-0.19)	-13,700 (-1.87)	3.980 (0.40)
Sample size	9,085	13,473	9,498	3,237
R^2	0.47	0.19	0.29	0.19
Cohen's f^2	0.90	0.23	0.41	0.23

t-value in parentheses

Fig. 8 Estimated results of multiple regression model for vessel size selection. Source: Authors

As shown in Fig. 8, the value of R^2 for each commodity is small, whereas the t-values of some estimated variables are significantly large. Furthermore, the effect sizes for iron ore and grains are large, whereas those for coal and fertilisers are small. Therefore, the proposed multiple regression model is capable of explaining the influence of variables on each commodity but it is not appropriate for forecasting. The detailed commodity-based analysis is presented below.

6.2.1 Iron ore

Figure 8 shows that draft limit in ports, trade volumes and voyage distance have a significant effect on vessel size. The t-values of draft limits in both loading and discharging ports are comparatively large. Larger vessels—VLOC and capesize—are generally employed in iron ore shipping, whereas it is possible to use a panamax or neo-panamax vessel if draft limits exist. Moreover, based on the t-values in Fig. 8, the influence of trade volumes is greater than that of voyage distances. Due to the stability of iron ore shipping among countries with large trade volumes, more long-term contracts for large-size vessels could emerge.

6.2.2 Coal

Draft limit in ports, trade volumes and voyage distance are of great importance in coal shipping. Among them, draft limits in discharging ports is significantly influential. In particular, this variable is significant for countries with large import demand

but insufficient port facilities, such as Japan. Moreover, the size of vessels passing through the SC is significantly large because of the toll system, as mentioned above. However, vessels transiting the PC do not contribute much to the vessel size selection because the tolls of the PC depend on net tonnage with a fixed rate.

6.2.3 Grains

The most important factor in grains shipping is trade volumes, followed by voyage distance and draft limits. Grains shipping consumes comparatively less fuel because the stowage factor for grains is about three times that of iron ore (Stopford 2008). Due to relatively smaller lot sizes (demand) even for long-distance voyages, vessels smaller than the handymax size are often used. This phenomenon makes trade volumes more important than voyage distance when selecting vessel size.

The impact of dry bulk shipping indices on grains transport is more significant than in the case of other commodities. One possible reason is strong seasonality, namely the balance of supply and demand in grains transport tends to fluctuate within certain time periods (as well as with dry bulk shipping indices) than other commodities.

6.2.4 Fertilisers

The influence of draft limits in loading ports and trade volumes is comparatively large, followed by voyage distance, draft limits in discharging ports and the SC transit. Handymax and handysize vessels are mostly used because fertilisers are frequently shipped between very restricted ports, such as those in the Middle East, one of the most important export regions of fertilisers.

6.3 Multinomial logit model

6.3.1 Overview of the model

In the dry bulk shipping industry, it is more important in a practical sense to choose vessel type (categorised in Table 1), than vessel capacity. We apply a multinomial logit model, which is often employed for the discrete choice problem among three or more alternatives. In the following model, the freight rate indices are excluded due to their regression results. Equation (3) describes the probability $Pr_{crs}^{(l)}$ of selecting vessel type l ($l = 1, 2, \ldots 7$; from VLOC to minibulk) for transporting commodity c between loading port r to discharging port s.

$$\Pr_{crs}^{(l)} = \frac{\exp\left(\beta_{c0}^{(l)} + \beta_{c1}^{(l)}V_{cij} + \beta_{c2}^{(l)}D_{rs} + \beta_{c3}^{(l)}o_r + \beta_{c4}^{(l)}u_s + \beta_{c5}^{(l)}S_{rs} + \beta_{c6}^{(l)}P_{rs}\right)}{\sum_{k=1}^{7} \exp\left(\beta_{c0}^{(k)} + \beta_{c1}^{(k)}V_{cij} + \beta_{c2}^{(k)}D_{rs} + \beta_{c3}^{(k)}o_r + \beta_{c4}^{(k)}u_s + \beta_{c5}^{(k)}S_{rs} + \beta_{c6}^{(k)}P_{rs}\right)},$$

(3)

where $\beta_{cm}^{(l)}$ are the coefficients of each explanatory variable m ($m = 0, 1, \ldots 6$) for commodity c.

Table 9 Accuracy of the multinomial logit model. Source: Authors

Commodity	Number of choices	First hit rate (%)	Second hit rate (%)
Iron ore	6	65.0	84.8
Coal	4	50.6	79.8
Grains	5	55.4	81.2
Fertilisers	4	60.1	90.8

Vessel type [Number of records]	VLOC [1040]	Capesize [5269]	Neo-Panamax [685]	Panamax [898]	Handymax [805]	Handysize [388]
Trade volume (kton)	3.62E-08* (0.01)	2.91E-07* (0.00)	5.69E-09 (0.54)	-7.30E-08* (0.00)	-1.30E-07* (0.00)	-1.30E-07* (0.00)
Voyage distance (km)	-1.55E-06* (0.08)	1.33E-05* (0.00)	-3.33E-06* (0.00)	-3.74E-06* (0.00)	-1.49E-06* (0.02)	-3.18E-06* (0.00)
Loading max draft (-m)	3.50E-02* (0.00)	1.23E-02* (0.00)	-5.00E-03* (0.00)	-1.77E-02* (0.00)	-1.69E-02* (0.00)	-7.70E-03* (0.00)
Discharging max draft (-m)	2.21E-02* (0.00)	6.60E-03* (0.00)	-7.70E-03* (0.00)	-7.00E-03* (0.00)	-8.00E-03* (0.00)	-6.10E-03* (0.00)
PC share	-2.76E+01 (1.00)	3.31E+01 (1.00)	-1.28E+01 (1.00)	3.76E+00 (1.00)	2.68E+00 (1.00)	7.70E-01 (1.00)
SC share	-2.87E-01* (0.06)	3.82E-01* (0.01)	-3.89E-02 (0.20)	-9.60E-02* (0.00)	5.19E-02* (0.00)	-1.21E-02 (0.41)

p-value in parentheses with marginal effect above
* 10% level of significance (p ≤ 0.10)

Fig. 9 Estimation results of the multinomial logit model for iron ore shipping. Source: Authors

We employ Newton's method (Kelley 2003) to estimate the marginal effect of each explanatory variable. As the number of selected samples is extremely small, calculating the Hessian matrix and obtaining resolution becomes more challenging, therefore the minibulk vessel is integrated into the handysize category in the iron ore model. Moreover, the neo-panamax vessel is integrated into the panamax category, and again the minibulk category is integrated with the handysize one in the coal model because of the low accuracy of the original model.

6.3.2 Estimation results

The accuracy of the multinomial logit model for vessel type selection is defined as the *hit rate* of the estimated and observed vessel type. The first hit rate is defined based on the vessel type with the highest probability, whereas the second hit rate is defined based on the vessel types with the highest and second highest probabilities. Table 9 presents the accuracy of each model, showing more than 50% for the first hit rate and approximately 80–90% for the second hit rate.

6.3.2.1 Iron ore Figure 9 shows the estimation results for iron ore shipping. Most of the estimated marginal effects are statistically significant, except for the share of the PC transit. As the draft limit in port becomes deeper, the probability that the VLOC is selected most increases, followed by capesize vessels. Moreover, the marginal effects of the draft limit in the loading port on these vessels are larger than in the discharging

Vessel type [Number of records]	Capesize [1873]	Panamax [6511]	Handymax [3316]	Handysize [1773]
Trade volume (kton)	3.15E-07*	1.98E-06*	1.17E-06*	-3.46E-06*
	(0.00)	(0.00)	(0.00)	(0.00)
Voyage distance (km)	5.81E-06*	1.71E-05*	3.53E-06*	-2.65E-05*
	(0.00)	(0.00)	(0.00)	(0.00)
Loading max draft (-m)	1.07E-02*	6.10E-03*	-5.80E-03*	-1.10E-02*
	(0.00)	(0.00)	(0.00)	(0.00)
Discharging max draft (-m)	2.36E-02*	4.50E-03*	-2.18E-02*	-6.30E-03*
	(0.00)	(0.00)	(0.00)	(0.00)
PC share	-2.10E-01*	1.68E-01*	1.63E-01*	-1.21E-01*
	(0.00)	(0.00)	(0.00)	(0.00)
SC share	4.16E-02	-4.42E-02	9.20E-03	-6.60E-03
	(0.18)	(0.45)	(0.86)	(0.89)

p-value in parentheses with marginal effect above
* 10% level of significance ($p \leq 0.10$)

Fig. 10 Estimation results of the multinomial logit model for coal shipping. Source: Authors

Vessel type [Number of records]	Neo-Panamax [208]	Panamax [2841]	Handymax [2454]	Handysize [3580]	Minibulk [415]
Trade volume (kton)	1.43E-07	2.00E-05*	3.12E-06*	-4.49E-06*	-2.00E-05*
	(0.47)	(0.00)	(0.00)	(0.00)	(0.00)
Voyage distance (km)	1.36E-06*	1.25E-05*	1.23E-05*	-1.16E-05*	-1.46E-05*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Loading max draft (-m)	1.40E-03*	1.04E-02*	2.90E-03*	-1.18E-02*	-3.00E-03*
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
Discharging max draught (-m)	1.90E-03*	1.04E-02*	2.60E-03*	-1.46E-02*	-3.00E-04
	(0.00)	(0.00)	(0.02)	(0.00)	(0.63)
PC share	-7.98E-01	-4.18E-02	4.88E-01	3.91E-01	-3.87E-02
	(0.91)	(0.99)	(0.81)	(0.86)	(0.83)
SC share	4.30E-03	1.21E-01*	1.49E-01*	-2.16E-01*	-5.72E-02*
	(0.34)	(0.00)	(0.00)	(0.00)	(0.00)

p-value in parentheses with marginal effect above
* 10% level of significance ($p \leq 0.10$)

Fig. 11 Estimation results of the multinomial logit model for grains shipping. Source: Authors

port. The increase in the trade volume or voyage distance increases the probability of selecting capesize vessels. The SC share increases the probability of capesize vessels because some of them can just about transit the SC, but decreases that of the VLOC because these vessels cannot transit the canal without reducing the cargo payload.

6.3.2.2 Coal Figure 10 presents the estimation results for coal. All the estimated marginal effects are statistically significant, except for the SC transit share. Similarly to iron ore, as the draft limit in the ports becomes deeper, the probability of the largest vessel (capesize) increases the most, followed by the second-largest vessel (panamax), whereas trade volume and voyage distance mostly raise the probability of the second-largest vessel. The PC transit share increases the probability of the panamax and handymax vessels, whereas it decreases the probability of the capesize vessels because it is difficult for them to transit the PC.

6.3.2.3 Grains Figure 11 presents the estimation results for grains. Most of the estimated marginal effects are statistically significant, except for the PC transit share.

Vessel type [Number of records]	Panamax [185]	Handymax [1064]	Handysize [1759]	Minibulk [229]
Trade volume (kton)	2.00E-05*	7.00E-05*	7.00E-05*	-1.50E-04*
	(0.00)	(0.00)	(0.00)	(0.00)
Voyage distance (km)	1.81E-06*	1.56E-05*	6.92E-06*	-2.43E-05*
	(0.00)	(0.00)	(0.00)	(0.00)
Loading max draft (-m)	8.10E-03*	1.63E-02*	-2.38E-02*	-6.00E-04
	(0.00)	(0.00)	(0.00)	(0.61)
Discharging max draft (-m)	4.00E-03*	6.70E-03*	-6.80E-03*	-3.90E-03*
	(0.00)	(0.00)	(0.00)	(0.00)
PC share	-1.96E-02	-7.24E-02*	4.82E-02	4.37E-02*
	(0.35)	(0.06)	(0.23)	(0.02)
SC share	3.38E-02*	2.47E-01*	-2.32E-01*	-4.90E-02
	(0.05)	(0.00)	(0.00)	(0.11)

p-value in parentheses with marginal effect above
* 10% level of significance ($p \leq 0.10$)

Fig. 12 Estimation results of the multinomial logit model for fertilisers shipping. Source: Authors

Excluding the neo-panamax vessels, for which the number of records is the smallest, as the trade volumes increase, the change in probability for selecting each vessel type is proportional to the size of vessel type. Namely, as the trade volumes increase, the probability of selecting the panamax vessel increases the most, followed by the handymax vessel, whereas the probability of selecting the minibulk vessel decreases most, followed by the handysize vessel. A similar trend is observed in the draft limits, except for the largest (neo-panamax) and smallest (minibulk) vessels. By contrast, the estimated marginal effect of the panamax and handymax vessels in voyage distance is almost the same, demonstrating that the handymax vessel is normally used for long-distance grains shipping, along with the panamax vessel because of good fuel performance.

6.3.2.4 Fertilisers Figure 12 presents the estimation results for fertilisers shipping. The marginal effects are statistically significant for most variables. As with grains shipping, by excluding the panamax vessels, of which the number of records is the smallest, the probability of the handymax vessel increases the most, as the trade volume, voyage distance, draft limit or the SC transit share increases. By contrast, the PC transit share mostly increases the probability of the minibulk vessels. One explanation for this finding could be that vessels shipping from the east coast of the United States to the west coast of Colombia are supposed to pass through the PC and most of them are minibulk vessels.

6.3.3 Summary of estimation results

In most cases, an increase in trade volume, voyage distance or maximum draft in ports contributes to increasing the probability that larger vessels are selected. More specifically, mitigating draft limits significantly affects the use of the largest vessels, whereas trade volume and voyage distance generally affect even the smaller vessels. For the draft limit, and focussing on the VLOC and capesize vessels, a similar trend to the multiple regression model can be observed, as the limit in the loading port is
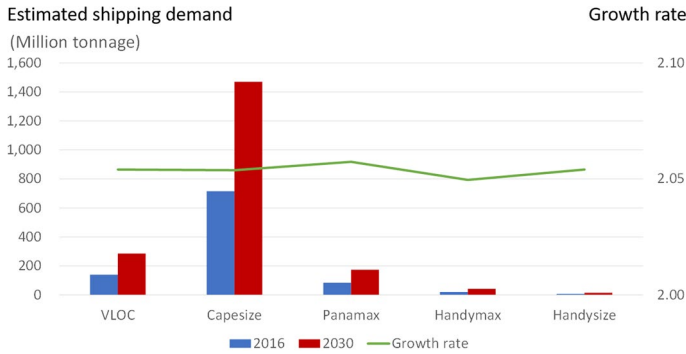
Estimated shipping demand   Growth rate

Fig. 13 Estimated shipping demand for iron ore by vessel type in 2030. Source: Authors

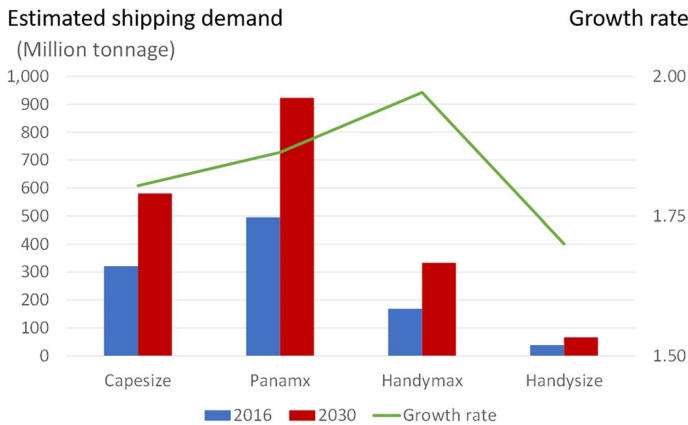Estimated shipping demand   Growth rate

Fig. 14 Estimated shipping demand for coal by vessel type in 2030. Source: Authors

dominant in iron ore shipping, whereas in the discharging port it is dominant in coal shipping.

The effects of transiting the SC and PC are more complicated because this may also reflect the geographical characteristics of trade for each commodity. In cases where the SC or PC share is statistically significant, the use of vessels that cannot transit the canal fully laden is discouraged, whereas the use of larger vessels beneath the limit is encouraged, especially in the SC, to save on canal tolls.

6.4 Estimation of future shipping demand by vessel type

Using the multinomial logit model of the previous subsection, we estimate the future trend of shipping demand by vessel type. The future shipping demand of each commodity is input based on the forecasts of IHS (as of 2030), whereas the other exploratory variables are unchanged.

Figure 13 shows the estimated future shipping demand for iron ore for each vessel type in 2030, together with the shipping demand in 2016. The trade volume of all vessel types will increase in 2030, and the growth rate of the panamax vessels will be slightly higher than that of other types. Figure 14 shows the estimated future shipping demand for coal for each vessel type in 2030, together with the shipping demand in 2016. While the demand for all vessel types will increase in 2030, similar to iron ore shipping, the growth rate of each vessel type is different; namely, that of handymax vessels is expected to be the highest.

7 Conclusions

This study first estimated the global port-to-port cargo flows of dry bulk shipping by commodity using AIS data and the information on commodities handled in ports and berths acquired from AXS Dry data and the two-tiered Geohash geocoding. Estimated trade volumes of exports and imports by country for three major dry bulk cargoes (iron ore, coal and grains) and two minor dry bulk cargoes (fertilisers and iron and steel) were compared with observed volumes. As a result, the trade volumes of iron ore, coal, grains and fertilisers were estimated with high accuracy compared with the observed export and import volume at country level, whereas the estimation accuracy for iron and steel deteriorated because of the different definitions of iron and steel in each set of statistics. The volume of imports by month and by port in Japan was generally estimated accurately, however this estimate should be improved in some ports where large over- or under-estimation occurs.

Next, we quantified the relationship between commodities and vessel size. After examining the factors that contributed to selecting a certain vessel size, multiple regression and a multinomial logit model were developed. Our findings were as follows: For iron ore, the draft limit of the loading port played a critical role in selecting vessel size, especially larger vessels such as the VLOC and capesize vessels. Additionally, considering the stable relationship between countries with large trade volumes, the influence of trade volume on vessel size was demonstrated to be more important than the voyage distance. For coal, the draft limit at the discharging port was more important in selecting vessel size and it was more critical at a discharging port which could not accommodate larger vessels. For grains, supply and demand areas were dispersed and fuel consumption was relatively low. Therefore, there was a high probability in selecting a handymax vessel, even for long-distance voyages, demonstrating the small influence that voyage distance had on the choice of vessel size. Moreover, the dry bulk shipping freight indices had some influence on vessel size selection because of the seasonality associated with the supply and demand of grains shipping. For fertilisers, many variables had a positive influence on selecting the handymax vessel. The transit of the PC indirectly made the handysize or minibulk vessel a choice for shipping between the USA and South America. Furthermore, in the transport of any commodity, if the vessel was transiting the SC or PC, the use of vessels that cannot transit the canal without reducing the cargo payload was discouraged. By contrast, the toll system of the SC strengthened the incentive

to use larger vessels, because the larger the vessel size, the lower the toll per unit of cargo.

Finally, future shipping demand by vessel type (as of 2030) was estimated in iron ore and coal shipping as an example application of the developed model. The results for coal shipping revealed that the growing demand for importing into developing countries will stimulate the demand for relatively small-size vessels, such as the handymax vessels. In addition to the forecast of the future demand by vessel type for each commodity, we considered many applications of port-to-port global cargo flows of dry bulk goods, for practical or industrial purposes. For example, our estimated results could allow more detailed analyses of the shipping industry, such as port and berth usage rates, or the operational efficiency of tramp shipping. This could be achieved by observing the characteristics of tramp shipping and its trading patterns. Moreover, a more precise forecast of short-term or real-time cargo demand is possible, which could be used to optimise vessel choice or even orderbook.

The accuracy of our estimates requires further improvement through a closer investigation of the data and algorithms, including those of iron and steel shipping, and expanding the range of estimation to other minor bulk cargo or different years. For example, a more precise estimation of loading factors by voyage, which could be achieved by improving the reliability of draft change information along with other specifications of the vessel, such as in Jia et al. (2019), would help increase estimation accuracy in cases of successive loading or discharging at multiple ports. Another possibility for improvement is to interpolate by other vessel movement data if AIS data are missing or wrong.

The vessel size selection model could also be improved. For example, we did not apply the ordered logit model in this study because some variables, such as the SC transit share, non-linearly affected vessel size. However, the results of our model should be compared with other alternative models. The integrated model in all commodities should also be examined. Furthermore, confirmation of the robustness of the developed models by using other years' data is necessary and would also enable a time series analysis of vessel types. Regarding our forecasts of the future demand of vessel types, an in-depth study that considers forecasts of port-based cargo flows and the future changes in draft limits would improve the precision of our predictions.

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* 44 (5): 657–665.

Adland, R., H. Jia, and S.P. Strandenes. 2018. The determinants of vessel capacity utilization: The case of Brazilian iron ore exports. *Transportation Research* 110A: 191–201.

Adland, R., and H. Jia. 2018. Dynamic speed choice in bulk shipping. *Maritime Economics & Logistics* 20 (2): 253–266.

Alizadeh, A.H., S.P. Strandenes, and H. Thanopoulou. 2016. Capacity retirement in the dry bulk market: A vessel based logit model. *Transportation Research* 92E: 28–42.

Arifin, M.D., K. Hamada, N. Hirata, K. Ihara, and Y. Koide. 2018. Development of ship allocation models using marine logistics data and its application to bulk carrier demand forecasting and basic planning support. *Journal of the Japan Society of Naval Architects and Ocean Engineers* 27: 139–148.

Arslanalp, S., Marini, M., and Tumbarello, P., 2019, Big data on vessel traffic: Nowcasting trade flows in real time, IMF Working Paper, https://www.imf.org/en/Publications/WP/Issues/2019/12/13/Big-Data-on-Vessel-Traffic-Nowcasting-Trade-Flows-in-Real-Time-48837

AXS Marine, AXS dry, Date of access: 1/2020. https://public.axsmarine.com/axsdry

Borthen, T., H. Loennechen, X. Wang, K. Fagerholt, and T. Vidal. 2018. A genetic search-based heuristic for a fleet size and periodic routing problem with application to offshore supply planning. *EURO Journal on Transportation and Logistics* 7 (2): 121–150.

Clarkson Research, 2019. Dry Bulk Trade Outlook.

Dinwoodie, J., M. Landamore, and P. Rigot-Muller. 2014. Dry bulk shipping flows to 2050: Delphi perceptions of early career specialists. *Technological Forecasting and Social Change* 88: 64–75.

Goulielmos, A. 2013. An econometric analysis of the impact of vessel size on weekly time charters: A study in volatility (Panamax and Cape). *International Journal of Transport Economics* 40 (1): 31–48.

IHS Markit, GTA Forecasting, Date of access:1/2020. https://ihsmarkit.com/products/gta-forecasting.html

Japan Maritime Center, 2017. Trends in maritime dry bulk cargo shipping, Date of access: 3/2020. https://www.jpmac.or.jp/relation/drybulk_cargo_pdf/d2017-09.pdf. (in Japanese)

Jia, H., V. Prakash, and T. Smith. 2019. Estimating vessel payloads in bulk shipping using AIS data. *International Journal of Shipping and Transport Logistics* 11 (1): 25–40.

Kelley, C.T. 2003. Solving nonlinear equations with Newton's method. *SIAM Publications. doi* 10 (1137/1): 9780898718898.

Kline, Investor Relations (Shipping Market Information), Date of access: 1/2020. https://www.kline.co.jp/ja/ir/finance/shipping.html.

Ko, B.-W. 2011. An Application of Dynamic Factor Model to Dry Bulk Market - Focusing on the Analysis of Synchronicity and Idiosyncrasy in the Sub-Markets with Different Ship Size. *KMI International Journal of Maritime Affairs and Fisheries* 3 (1): 69–82.

Lechtenberg, S., D. de Siqueira Braga, and B. Hellingrath. 2019. Automatic identification system (AIS) data based ship-supply forecasting. In Digital Transformation in Maritime and City Logistics: Smart Solutions for Logistics. *Proceedings of the Hamburg International Conference of Logistics (HICL)* 28: 3–24.

Ministry of Land, Infrastructure, Transport and Tourism, Japan, 2016, Trends in imported cargo by 5 major products, Date of access: 9/2020. https://www.mlit.go.jp/k-toukei/h28genkyo.html (in Japanese)

Santos, T.A., and C. Guedes Soares. 2017. Methodology for ro-ro ship and fleet sizing with application to short sea shipping. *Maritime Policy & Management* 44 (7): 859–881.

Shibasaki, R., K. Kanamoto, and T. Suzuki. 2020. Estimating global pattern of LNG supply chain: A port-based approach by vessel movement database. *Maritime Policy & Management* 47 (2): 143–171.

Stopford, M., 2008, Maritime Economics 3rd edition, Routledge.

Tu, E., G. Zhang, L. Rachmawati, E. Rajabally, and G.B. Huang. 2017. Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology. *IEEE Transactions on Intelligent Transportation Systems* 99: 1–24.

Wada, Y., K. Hamada, N. Hirata, K. Seki, and S. Yamada. 2018. A system dynamics model for shipbuilding demand forecasting. *Journal of Marine Science and Technology* 23 (2): 236–252.

Wu, Y., J. Yin, and P. Sheng. 2018. The dynamics of dry bulk shipping market under the shipping cycle perspective: Market relationships and Volatility. *Transportation Research Record* 2672 (11): 1–9.

Yang, D., L. Wu, S. Wang, H. Jia, and K.X. Li. 2019. How big data enriches maritime research—A critical review of Automatic Identification System (AIS) data applications. *Transport Reviews* 39 (6): 755–773.