Bank of Japan Working Paper Series

A Nowcasting Model of Exports Using Maritime Big Data

Kakuho Furukawa*
kakuho.furukawa@boj.or.jp

Ryohei Hisano**
hisanor@g.ecc.u-tokyo.ac.jp

No.22-E-19
Dec. 2022

Bank of Japan
2-1-1 Nihonbashi-Hongokucho, Chuo-ku, Tokyo 103-0021, Japan

* Research and Statistics Department
** The University of Tokyo

A Nowcasting Model of Exports Using Maritime Big Data*

Kakuho Furukawa† Ryohei Hisano‡

December 2022

Abstract

We nowcast Japan's exports using maritime big data (the Automatic Identification System data), which contains information on vessels such as their locations. The data has been only relatively rarely used for capturing economic trends. In doing so, we improve the accuracy of nowcasts by utilizing official statistics such as geographical data on ports and machine learning techniques. The analysis shows that the nowcasting model augmented with AIS data improves the performance of nowcasting relative to existing statistics (provisional reports on the *Trade Statistics of Japan*) that is available in close to real-time. In particular, the nowcasting model developed in this paper follows the movements of exports reasonably well even when they increase or decrease significantly (e.g., when the pandemic began in the spring of 2020 and when the supply chain was disrupted around mid-2021).

JEL Classification: C49, C55, E27
Keywords: Nowcasting, Alternative data, AIS data, Exports

* The authors thank Kosuke Aoki, Ichiro Fukunaga, Yoshihiko Hogen, Yosuke Kido, Yoshiyasu Koide, Teppei Nagano, Jouchi Nakajima, Tomoyuki Yagi, and staff members of the Bank of Japan for their valuable comments. All remaining errors are our own. The views expressed in this paper are those of the authors and do not necessarily reflect the official views of the Bank of Japan.
† Research and Statistics Department, Bank of Japan (kakuho.furukawa@boj.or.jp)
‡ The University of Tokyo (hisanor@g.ecc.u-tokyo.ac.jp)

1. Introduction

In recent years, there has been a growing interest in so-called "alternative data," which refers to non-traditional data based on sources different from existing statistics.[1] One of the advantages of using alternative data for economic analysis is the fact that many alternative data are available in a more timely manner (i.e., the time until the data become available is shorter) than existing statistics, which makes them useful for timely analyses of the current economic situations. In this regard, previous studies report that these alternative data contribute to improving performance of economic nowcasting (Galbraith and Tkacz, 2020; Nymand-Andersen and Pantelidis, 2018; Nakazawa, 2022; Okubo et al., 2022; Furukawa et al., 2020).[2]

In Japan, nowcasting models have been developed for GDP (Hayashi and Tachi, 2022; Maehashi and Shintani, 2020; Hara and Yamane, 2013; Chikamatsu et al., 2021; Nakazawa, 2022), private consumption (Okubo et al., 2022), and industrial production (Shintani, 2005; Furukawa et al., 2022).[3] Meanwhile, it has been relatively difficult to nowcast exports, which account for about 20 percent of Japan's GDP, with a reasonable degree of accuracy due to their large fluctuations. However, the latest studies in other countries (Arslanalp et al., 2019, 2021; Cerdeiro et al., 2020) show that maritime big data (Automatic Identification System data, see Section 3 for details) that records information on vessels including their locations, which has been used relatively rarely for economic analyses, is useful in capturing export trends in a timely manner.

Against this background, this paper utilizes the maritime data for nowcasting Japan's exports. To the best of our knowledge, this paper is the first study in Japan that utilizes the maritime data for nowcasting exports. Since seaborne trade accounts for about 70 percent of the nominal value of Japan's exports, maritime big data is likely to offer useful information on export trends (Figure 1). The unique part of this paper is the identification of vessels engaged in export activities by combining AIS data with geographic data on ports. Based on these vessels, we calculate the "export index" that serves as an explanatory variable in a nowcasting model (linear regression model). We also use a machine learning technique in calculating the export index in order to improve the

[1] This paper follows the definition of alternative data in Kameda (2022). Alternative data refers to data other than traditional economic data (e.g., monthly and quarterly macroeconomic indicators and earnings disclosure data of listed companies).

[2] Nowcasting means capturing current economic trends rather than forecasting the future. Since official statistics on economic activities are typically published with time lag, researches to capture economic trends in a more real-time fashion have been increasing in number in recent years.

[3] A notable prior study of nowcasting economic activity abroad is Giannone et al. (2008).

accuracy of the nowcasting model.

There are three main contributions of this paper. The first contribution is that the export index developed in this paper captures trends in real exports. While official statistics on exports including the Ministry of Finance's *Trade Statistics of Japan* and the Bank of Japan's *Developments in Real Exports and Real Imports* are released around the end of the following month,[4] our nowcasting model captures export trends nearly three weeks earlier than these official statistics.

Second, the export index tends to result in better outcomes than existing statistics that is also available in real-time (provisional reports of the *Trade Statistics of Japan*). In particular, we are able to follow the movements of exports reasonably well on occasions when they increase or decrease significantly such as in the spring of 2020, when the pandemic began, and around the middle of 2021, when the global supply chain was disrupted. These results suggest that AIS data, which has not been utilized widely in economic analyses in Japan, can be useful in capturing export trends in the country.

Third, we propose a method to improve the accuracy of nowcasting exports with the export index by precisely identifying vessels engaged in export activities and using a machine learning technique in calculating the export index. We consider this contribution the most important of the three. Specifically, we accurately identify which ports vessels have entered and exited by using geographic data on the boundaries of ports and calculate the export index by applying machine learning techniques such as kernel methods and deep learning in estimating the export function. We test the predictive accuracy of the nowcasting model for real exports using root mean squared error (RMSE). The results show that the model that does not utilize machine learning in calculating the export index achieve only a slight improvement compared to the model that uses only the existing statistics, while the performance improves significantly when the export index is calculated using machine learning techniques (RMSE is 5.80 for the nowcasting model with existing statistics, 5.70 for the model with export indexes calculated without machine learning, and 3.64 for the model with export indexes calculated with machine learning). This implies that there is a strong nonlinear element in the nowcasting of real exports and that machine learning techniques may be useful in capturing such relationships.

The rest of the paper is organized as follows. Section 2 reviews previous studies and

[4] For example, the *Trade Statistics of Japan* and *Developments in Real Exports and Real Imports* for June 2022 were released on July 21, 2022. For the *Trade Statistics of Japan*, provisional figures for the first ten days of each month are released at the end of the month, and provisional figures for the first twenty days of each month are released at the beginning of the following month.

summarizes the main features of this paper; Section 3 describes the data used in this analysis; Section 4 explains the nowcasting model used in this paper; Section 5 summarizes the nowcasting estimation results and their implications, and Section 6 concludes.

2. Literature Review

The analysis in this paper is primarily related to three genres of previous studies: the use of AIS data, nowcasting models using alternative data, and economic prediction using machine learning techniques. We review them in detail below in this order.

First, this paper is related to previous studies on nowcasting trade using AIS data (Adland et al., 2017; Arslanalp et al., 2019, 2021; Cerdeiro et al., 2020). Adland et al. (2017) use AIS data on tankers to nowcast crude oil trade volumes in oil-producing countries and show that AIS data is useful for capturing crude oil trade volumes, although the accuracy of prediction can deteriorate for some countries and times when trade through pipelines is large. Arslanalp et al. (2019) perform nowcasting of Malta's trade volume by aggregating the number of cargo vessels and the amount of cargo loaded using AIS data, and show that the information on vessels' locations is useful in predicting trade volume. Arslanalp et al. (2021) extend Arslanalp et al. (2019) and improve the nowcasting accuracy of trade volume in Pacific island countries by using geographic information of ports in each country and information on shipping liner schedules. Cerdeiro et al. (2020) use AIS data covering the entire world to develop indicators by country and vessel type (e.g., car carriers, containerships, etc.) to capture trends in maritime trade imports and exports in countries around the world, and find that AIS data is useful for capturing regional trade volumes (e.g., the volume of automobiles exported from Japan). These indicators are periodically updated and published in the United Nations Comtrade database.[5,6] While the number of studies on nowcasting trade volumes using AIS data has been increasing in recent years, to the best of the authors' knowledge, there have been no studies specifically focusing on Japan. In this paper, in addition to AIS data around Japan's coastal areas, we use geographical data on ports and official statistics on export activities at each port to nowcast Japan's export trends with a high degree of accuracy.

[5] https://comtrade.un.org/data/ais

[6] Note that although the database includes indicators for Japan, it is not always updated with the latest data, making it difficult to use for nowcasting. In addition, as we will see in Section 5.2., we succeed in capturing Japan's export with higher precision than Cerdeiro et al. (2020).

Second, this paper is related to nowcasting models using alternative data in that it uses non-traditional data, at least in the field of economic analyses, namely AIS data. For Japan, for example, Nakazawa (2022), Okubo et al. (2022), and Furukawa et al. (2022) construct nowcasting models using alternative data for GDP, private consumption, and industrial production, respectively. All of these studies demonstrate the usefulness of alternative data in capturing economic activities in real time. The data used in these studies (number of searches on the Internet (Nakazawa, 2022), credit card history (Okubo et al., 2022), and GPS data of mobile phones (Furukawa et al., 2022)) have also been used in many previous studies abroad. Consequently, we are rapidly accumulating knowledge about what alternative data is useful in nowcasting economic activities. For example, Galbraith and Tkacz (2015) construct a nowcasting model of GDP of Canada using transaction data of credit/debit card and checks and show that nowcasting accuracy improves compared to a model without the transaction data. Nymand-Andersen and Pantelidis (2018) also built a nowcasting model for new car sales in Europe (10 major countries) using Google search data (Google Trends) for related categories and find that the Google Trends data improves nowcasting accuracy. The analysis in this paper confirms that AIS data is also useful for nowcasting economic activities.

Third, this paper is also related to the literature that applies machine learning techniques to economic prediction. Bolhuis and Rayner (2020) point out the shortcomings of traditional Ordinary Least Square (OLS) models, such as inability to deal with multicollinearity and nonlinearity, and suggest that machine learning models can be used to overcome these shortcomings. Many previous studies in strand of alternative data also find that machine learning models outperform linear models (Furukawa et al., 2022; Anesti et al., 2021; Ashwin et al., 2021; Richardson et al., 2021). For example, Anesti et al. (2021) compare the results of GDP prediction when linear models and machine learning models are estimated with alternative data and general macro statistics, and find that the improvement in prediction accuracy with the introduction of machine learning models is particularly large when alternative data are included as explanatory variables.[7] In this paper, we also find that we can improve predictive accuracy by using machine learning techniques such as kernel methods and deep learning in the process of aggregating AIS data.

[7] In this regard, Anesti et al. (2021) point out that the use of alternative data increases the volume of information and the complexity of relationships among variables, which may increase the benefits of utilizing machine learning models that can deal with multicollinearity and nonlinearity.

3. Data

3.1. AIS data

The Automatic Identification System (AIS) is a device that allows vessels to exchange navigational information with each other. AIS transmits data including the vessel's identification codes, names, positions (latitude and longitude), draughts, which refer to the distance from the bottom of a vessel to the water surface, and destination. The data can be received by other vessels navigating in the vicinity, as well as by land-based receiving stations and satellites. Vessels meeting certain standards are required to carry AIS according to international conventions for the purpose of smooth navigation control and maritime accident prevention. In Japan, all passenger ships or vessels of 300 gross tons or more engaged in international voyages and all vessels of 500 gross tons or more are required to carry AIS. Therefore, in principle, all vessels large enough to be engaged in exports are likely to be equipped with AIS.

The radio waves emitted by AIS can be received by anyone who has a receiving medium (e.g., land-based receiving station or satellite). As a result, with the development of small satellite technology, there has been an increase in private services that provide AIS data in recent years. We use AIS data around Japan provided by VesselFinder for the period from January 1st 2017 to March 31st 2022.[8] In doing so, this paper uses AIS data on car carriers and containerships whose frequency is reduced to 6-hour units.[9] In Japan exports by car carriers and containerships account for about 60% of exports (about 85% of maritime trade) in nominal value. Moreover, given that trade by air transportation, which accounts for about 30% of exports, follows roughly the same trend as exports by containerships (discussed below), the data in this analysis can be considered to have reasonable coverage (Figure 1). In this analysis, we use AIS data for 7,003 vessels. We can see that many vessels are concentrated in the vicinity of industrial zones (Figure 2).

[8] Some providers provide the latest AIS data with a lag of a few minutes, making it one of the most real-time of the alternative data.

[9] Since there are many passenger ships and other vessels not engaged in exports sailing around Japan and AIS emits signals every few seconds, acquiring all AIS data would increase the data size drastically. We therefore filter the data on certain conditions. In addition, given that most of Japan's goods exports take place in Honshu, Kyushu, and Shikoku, we cover the coastal areas of these islands (i.e., we exclude Hokkaido and Okinawa). The data used in this analysis have 6,591,066 rows in a csv format and 1.1 GB in size.

3.2. Port data

In order to identify whether a vessel is engaged in export activities based on AIS data, it is necessary to correctly determine whether the vessel is located within a port. In this analysis, we obtain geographic data on the boundaries of ports throughout Japan as designated by the Port and Harbor Act from the Ministry of Land, Infrastructure, Transport and Tourism, and identify the area of each port. Specifically, by combining this data with Japan's coastline data, we identify port "areas" bounded by port boundaries and coastlines (Figure 3(a)). However, there are some cases where it is difficult to identify port areas from port boundary and coastline data due to straits and remote islands. Among these cases, we manually identify areas for ports with large trade volumes (Figure 3(b)), while we exclude other small ports from the analysis. As a result, we identify the areas of 302 ports, which cover 99% of Japan's exports by vessel in nominal value for the sample covering 2017-2020.

4. Construction of the "Export Index"

In this section, we first explain how to calculate the "export index" to capture export using AIS data and port data. We also discuss methods to improve the export index by using port-level official statistics.

4.1. Methodology

Step 1. Identifying vessels destined for overseas

AIS data we use in this analysis includes not only vessels sailing from Japan to overseas, but also vessels traveling between domestic ports. Therefore, in order to capture a vessel's export activities, it is important to first identify whether the vessel's next destination is overseas or not. Although AIS data contains information on the next destination, it is difficult to use for analysis because it is entered manually by the crew and the format is not standardized.[10] Therefore, we identify whether a vessel is headed overseas or to another domestic port after leaving a domestic port by continuously tracking the location of the vessel.

[10] For example, if the next destination is Tokyo, the destination item in AIS data may be entered in various formats, such as "Tokyo," "JP TOKYO," or "TYO," making it difficult to correctly match it with the actual destination.

Specifically, after a vessel leaves a port in Japan, the vessel is assumed to be headed overseas if AIS data transmitted by the vessel meets either of the following conditions: (1) the transmission data is discontinued for 48 hours or longer, or (2) the transmission data is discontinued for 24 hours or longer and the draught changes over that period. Since AIS data we use in this analysis covers the coastal area of Japan, if the vessel's AIS data is discontinued for a long period of time, we can assume that the vessel has left Japan and is headed overseas. In the first condition, (1), we set the criterion to 48 hours.[11] However, in cases where the distance between domestic and foreign ports is close, such as between Kitakyushu and South Korea, the vessel may return to Japan within 48 hours after leaving the domestic port. Therefore, under the condition (2), the vessel is considered to be headed overseas after leaving a domestic port even if the vessel's AIS data is discontinued for a short period of time, if the vessel's draught changes between before and after that period since it implies loading or unloading operation took place during this period.

Based on these criteria, a total of 109,356 vessels are identified as having an overseas next destination during the sample period.

Step 2. Estimating metric tons of cargo

Next, we estimate the metric tons of cargo loaded on a vessel by using the information on draught from AIS data. As more cargo is loaded on a vessel, the vessel sinks deeper into the water surface and the draught becomes deeper, which means that it is possible to infer the metric tons of cargo from draught. Specifically, by assuming that the draught is proportional to the weight of the vessel, we can calculate metric tons of cargo loaded on vessel i at time d ($mtc_{i,d}$) from the following equation (Figure 4).[12,13]

$$mtc_{i,d} = mtc_{i,max} \times \frac{draught_{i,d} - draught_{i,min}}{draught_{i,max} - draught_{i,min}}, \tag{1}$$

[11] Even if a vessel is heading from a domestic port to another domestic port, it may temporarily leave the coastal area depending on its route. AIS data may also be interrupted for a short period of time for some reason (e.g., AIS is switched off temporarily). Therefore, we set the time period to be long enough to prevent such vessels from being mistakenly identified as heading overseas.

[12] To be precise, the physics law states that the weight of a vessel is equal to the weight of the seawater that the vessel displaces (Archimedes' principle). Assuming that the shape of the vessel is almost rectangular and that the volume of the part of the vessel submerged under the water surface is proportional to the draught, the above relationship that the draught is proportional to the weight of the vessel is obtained.

[13] In reality, when the metric tons of cargo is small, the draught is often deepened by loading the vessel with weights such as seawater to stabilize the body. Consequently, the relationship between the metric tons of cargo and draught may not be proportional as shown in Figure 4.

where $draught_{i,d}$ is the draught of vessel i at time d, and $mtc_{i,max}$, $draught_{i,max}$ and $draught_{i,min}$ are the maximum weight of cargo that vessel i can carry, the draught at maximum loading, and the draught when empty (when no cargo is loaded), respectively. $draught_{i,d}$ is obtained from AIS data, and $mtc_{i,max}$ and $draught_{i,max}$ are obtained from the vessel database maintained by VesselFinder. For $draught_{i,min}$, the minimum draught of vessel i in the sample period ($\min_d \; draught_{i,d}$) is used as an estimate.

Step 3. Calculating the export index

Finally, we calculate the export index using the data we obtained in Steps 1 and 2. In general, vessels engaging in exports call at multiple ports for loading before heading overseas. Hence, it is necessary to track the amount of exported goods loaded at each port in order to accurately capture export activities. However, since we can only estimate the metric tons of the entire cargo at the time of arrival and departure from AIS data, it is not possible to capture the amount of exported goods loaded at each port (Figure 5). Therefore, we calculate the export index by assuming that all cargo loaded on vessels heading overseas are exports and by summing the metric tons of cargo loaded on vessels identified as heading overseas at the port of departure. As shown in the formula below, we calculate the export index in a monthly frequency for car carriers and containerships, respectively.

$$export\ \ index_{type,t} = \sum_{\substack{i,d \\ i \in S_{type} \\ destination_{i,d} = overseas}} mtc_{i,d} \tag{2}$$

where t denotes a month (e.g., January 2022), d is a time point within month t, $type$ is the type of vessels (either car carrier or containership), S_{type} is the set of vessels classified as $type$, and $destination_{i,d}$ is a variable indicating whether vessel i's next destination at time d is overseas or domestic.[14]

4.2. Result

Comparing the export index calculated above with real exports, the correlation coefficient of the year-on-year change rates for cars is 0.89, showing that the index captures the movement in real exports well (Figure 6(a)). In particular, the export index clearly captures the large decline in exports associated with the spread of the COVID-19

[14] In Equation (2), all exports loaded on vessels at domestic ports are reflected in the export index when they leave the final port of call in Japan. However, in the *Trade Statistics of Japan*, the exported goods are recorded when they pass through customs. Hence, it should be noted that the timing of recognition of exports differs between the export index in this paper and the official statistics.

in the spring of 2020, suggesting that the export index for car carriers is useful in tracking actual car export values.

On the other hand, for the export index for containerships, the correlation coefficient of the year-on-year change rate with real exports is 0.42 (Figure 6(b)). While the overall trends are similar, the index is not as accurate as that for car carriers. In particular, since the spring of 2021, its year-on-year change rate has remained consistently below that of real exports.

4.3. Improving the export index for containerships

One possible reason for the lack of accuracy of the export index for containerships compared to that for car carriers is that the types of goods exported by containerships are wide-ranging, and the value per weight can vary significantly depending on the cargo loaded. Indeed, as Equation (2) indicates, the export index is calculated based on the weight of exported goods, and thus cannot capture fluctuations in value per weight. It is possible that the increase of semiconductor-related exports after 2021 led to a higher share of goods with high values per weight (Figure 7). The fact that the year-on-year change rate of the export index during this period is lower than that of the actual exports may be due in part to such a change in the composition of exported goods.

In addition, the weight of the containers themselves, in which the exported goods are stored, could work as noise and affect the accuracy of the export index for containerships.[15] If exported goods are lightweight or containers are underutilized, the weight of the containers themselves may account for a high percentage of the metric tons of cargo, making it difficult to accurately determine trends in exports.[16]

In light of these observations, it is important to consider "value per weight of cargo" and "weight of non-exported items in cargo" in order to accurately capture exports by containerships. Although these data do not exist at the vessel level in the first place and therefore it is not possible to obtain accurate figures, they are likely to depend on the port at which the vessel calls to a certain extent. Indeed, each port has distinct characteristics in terms of "value per weight of exported goods," "weight per container," and "share of exports in shipments" (Figure 8). Therefore, even without detailed vessel-level data on

[15] In the case of the most standard size container (20-foot container), the container itself weighs about 2 tons and the maximum loading weight is about 20 to 30 tons. Hence, more than 10% of the estimated loading capacity of a containership may be the weight of the containers.

[16] It should be noted that, for logistical reasons, empty containers are often loaded onto ships. In fact, according to the *Port Statistics* (Ministry of Land, Infrastructure, Transport and Tourism), in 2020, about 40% of containers loaded at domestic ports were empty.

cargo, it is possible to infer the characteristics of the cargoes (e.g., value per unit weight of cargo) loaded on the vessels by looking at the ports to which the vessels call.[17]

Based on the above considerations, we attempt to improve the nowcasting accuracy of the export index for containerships by estimating the following "export function" for each port.

$$export_{i,p,d} = f(mtc_{i,p,d}^{ent}, mtc_{i,p,d}^{exit}, pct_{i,p,d}, \boldsymbol{\theta}_p) \tag{3}$$

where the subscripts i, p, and d indicate that vessel i left port p at time d, and $export_{i,p,d}$, $mtc_{i,p,d}^{ent}$, $mtc_{i,p,d}^{exit}$, and $pct_{i,p,d}$ denote the value of exports loaded at port p (in real term), the metric tons of cargo loaded on vessel i when it enters/exits port p, and port call time at port p, respectively. The difference in the metric tons of cargo loaded at entry and exit, which are included in the explanatory variables, is likely to be related to the amount of cargo loaded at the port. However, when vessel i loads and unloads cargo at the same time, it may not be possible to accurately estimate the amount of cargo loaded at port p from the difference in the metric tons of cargo loaded at entry and exit alone. We address this issue by additionally taking into account the port call time based on the assumption that the greater the amount of cargo loaded and unloaded, the longer the port call time. $\boldsymbol{\theta}_p$ is a set of parameters that reflect the characteristics of port p, and we estimate their values such that the estimated exported values achieve the highest accuracy. By estimating parameters for each port, it is possible, for example, to estimate the values per metric ton of cargo for vessels to be high if they leave ports where values per weight of exported goods are large. This approach is expected to improve predictive accuracy.

Once the export function in Equation (3) is estimated, the revised export index for containerships is calculated by adding up the vessel-level export values.

$$export\ \ index_{container,t}^{r} = \sum_{i,p,d} export_{i,p,d} \tag{4}$$

where d is the point in time within month t.

Since it is not possible to determine a priori the form of the export function in Equation (3), we estimate Equation (3) using kernel methods and deep learning, which can handle

[17] For example, since exports loaded at the Port of Tokyo tend to have higher values per weight than exports loaded at the Port of Kawasaki, it can be inferred that the cargoes of vessels leaving the Port of Tokyo have higher values per weight than the cargoes of vessels leaving the Port of Kawasaki.

nonlinearity. We use data from January 2017 to December 2020 for the estimation. An overview of both methods is provided below.

4.3.1. Estimating export function using kernel methods

In the estimation of Equation (3) using kernel methods, the explanatory variable $\boldsymbol{x}$, which is a vector, is first converted to a vector in higher dimensions, $\phi(\boldsymbol{x})$. The export function then is assumed to be a linear function of $\phi(\boldsymbol{x})$. In the case of this analysis, the export function is written as follows

$$export_{i,p,d} = \boldsymbol{\theta}_p^T \phi(mtc_{i,p,d}^{ent}, mtc_{i,p,d}^{exit}, pct_{i,p,d}) \tag{5}$$

Here, the superscript T means transposition. Normally, the export function is estimated to minimize the prediction error of $export_{i,p,d}$. However, in this analysis, data for $export_{i,p,d}$ do not exist in the first place. Therefore, we set up the following loss function to estimate the export function given ϕ by minimizing the prediction error of export values at the port level.[18]

$$L_p = \frac{1}{2}\sum_t \left\{ \sum_{i,d} \boldsymbol{\theta}_p^T \phi\left(\boldsymbol{x}_{i,p,d}\right) - export_{p,t} \right\}^2 + \frac{\lambda}{2}\boldsymbol{\theta}_p^T \boldsymbol{\theta}_p \tag{6}$$

where $\boldsymbol{x}_{i,p,d} = (mtc_{i,p,d}^{ent}, mtc_{i,p,d}^{exit}, pct_{i,p,d})$ and $export_{p,t}$ is the value of exports by containerships (in real terms) at port p in month t.[19] Equation (6) implies that the export function is estimated so that the error between the sum of the export value of all vessels leaving port p in month t and $export_{p,t}$ is smallest (d is a point in time within month t). The second term on the right-hand side is a regularization term that prevents the value of the parameter $\boldsymbol{\theta}_p$ from becoming too large. In this subsection, we provide an overview of kernel methods and the estimation results. For the detailed estimation method, please refer to the Appendix.

Kernel methods transform Equation (6) so that the export function is represented as a function of the inner product $\phi(\boldsymbol{x})^T\phi(\boldsymbol{x}')$ rather than that of the original vector, $\phi(\boldsymbol{x})$.

---

18 Note that although most programming languages such as Python and R provide functions for estimating Equation (5), these functions cannot be used in this analysis because, as mentioned above, data on $export_{i,p,d}$ do not exist. In the Appendix, we explain how to estimate the export function using kernel methods without using these functions.

19 The value of exports by containerships (in real terms) for port p in month t is calculated by multiplying the value of exports by containerships (in real terms, for the nation) in month t by the share of exports (in nominal terms, including transportation means other than containerships) for port p in the same year.

$\phi(\boldsymbol{x})^T\phi(\boldsymbol{x}')$ is called a kernel function and is conventionally denoted as $k(\boldsymbol{x},\boldsymbol{x}')$. Even if $\phi(\boldsymbol{x})$ is of a complex form, $k(\boldsymbol{x},\boldsymbol{x}')$ can be relatively easy to compute in certain cases. Hence, kernel methods can be used to estimate a variety of functions and make it possible to efficiently search for a form of the export function that results in high accuracy while reducing the amount of computation. Figure 9 shows the accuracies of the export function for different types of kernel functions. Based on the accuracy in the out-of-sample period, we adopt the "Gaussian kernel function."

Based on this kernel function, the two panels in Figure 10 show the relationships between the value of exports and the explanatory variables. Panel (a) shows that if the difference between the amount of cargo loaded at the time of departure and at the time of entry is negative, little exporting has taken place, and if it is positive, the value of exports increases as the difference increases. Panel (b) shows that the longer the time of port calls, the greater the value of exports is.

4.3.2. Estimating export function using deep learning

In this subsection, we provide an overview of deep learning method used in estimating the export function in Equation (3). Deep learning is a type of machine learning and expresses a function by combining neurons that resemble nerve cells (Figure 11). In general, a deep learning model is one that has a large number of hidden layers. Deep learning is a more versatile method than kernel methods since it can estimate the relationship between explanatory variables and dependent variables without making any a priori assumptions. However, at the same time, the number of parameters tends to be very large and the estimation results are not stable when the data size is small. As a result, the out-of-sample accuracy may not always be improved. In this analysis, as with kernel methods, we assume that $export_{i,p,d}$ is expressed as a function of $mtc_{i,p,d}^{ent}$, $mtc_{i,p,d}^{exit}$, and $pct_{i,p,d}$ and estimate the deep learning model such that the error between the total export value of all vessels leaving port p in month t and $export_{p,t}$ is minimized.[20,21]

The relationship between the explanatory variables and exports value based on the export function estimated using deep learning shows that same pattern as the one estimated using kernel methods. Namely, little exporting has taken place if the difference

---

[20] We first estimate models for various combinations of the number of hidden layers and the number of dimensions of the hidden layers and employed a simple deep learning model with six hidden layers and 300 dimensions of the hidden layers, which has the highest out-of-sample prediction accuracy. The model uses approximately 0.6 million parameters.

[21] A more detailed explanation on deep learning is beyond the scope of this paper and is thus omitted here. For more details, see for example Goodfellow et al. (2016).

between the amount of cargo loaded at the time of departure and at the time of entry is negative while the value of exports increases as the difference increases, and the longer the time of port calls, the greater the value of exports is (Figure 12).

4.3.3. Result

The fit of the revised export index with real exports during the out-of-sample period (January 2021 to March 2022) shows an improvement over the original version for both kernel and deep learning methods (Figure 13). It implies that estimating an export function for each port allows us to better capture the movement of real exports by taking into account changes in the composition of exported goods. Comparing kernel methods with deep learning, the former better captured real exports during the recovery phase in the first half of 2021, while the latter better captured real exports during the decline phase in the second half of 2021, making it difficult to assess which method is superior.

5. Nowcasting Model

5.1. Accuracy of nowcasting model

In this section, we build nowcasting models for the overall real exports of Japan using the export indexes by vessel type constructed in this paper. We build the models in the following two steps. In the first step, the year-on-year change rate of real exports (not seasonally adjusted) is regressed by OLS on the year-on-year change rates of the export indexes for car carriers and containerships. In the second step, the predicted year-on-year change rates of real exports (not seasonally adjusted) obtained in the first step are converted to seasonally adjusted month-on-month change rates by using the seasonal factors of real exports.[22] As a comparison, we also estimate a nowcasting model of real exports in the same step as above but using the provisional report of the *Trade Statistics of Japan*, which is an official statistics published at the end of each month, after controlling for the exchange rate fluctuation.

First, we can observe in the OLS estimation result for step 1 using all samples (January 2018 to March 2022) that the model using the export index has a higher adjusted coefficient of determination and hence has greater explanatory power for real exports than

[22] Since seasonally adjusted series are generally used to evaluate economic conditions, we focus on capturing trends in real exports on a seasonally adjusted basis. Since the time series of the alternative data in this analysis are short and therefore the export index cannot be directly seasonally adjusted, we use the two-step process in which we first estimate the model on a seasonally unadjusted basis.

the model using existing statistics (provisional report of the *Trade Statistics of Japan*) (Figure 14).[23]

Next, taking the period from January 2021 to March 2022 as the out-of-sample period, we compare the performance of the models during this period with the following four models (Models 1-4). Specifically, Model 1 uses only existing statistics (provisional report of the *Trade Statistics of Japan*), Model 2 uses the export index calculated without utilizing machine learning (calculated in Section 4.2.), and Model 3 and 4 use the export index for containerships calculated by kernel methods and deep learning, respectively. By comparing Model 1 and 2, we can verify the extent to which nowcast accuracy improves with Model 2, which includes more near-term information. In this regard, the RMSE of Model 1 is 5.80 while that of Model 2 is 5.70, suggesting that the improvement in nowcast accuracy by using more recent information is limited. On the other hand, looking at Model 3 and 4, both of which use machine learning techniques, the RMSE of Model 3 was 5.04 and that of Model 4 was 3.64, indicating that the use of deep learning significantly improves the prediction accuracy of real exports (Figure 15).[24,25] Looking at the spring of 2020, when the pandemic began, and around mid-2021, when the global supply chain was significantly disrupted, Model 4 is able to reasonably follow the large decline in exports, while model 1 is far off its prediction. The main reason for the low nowcasting accuracy of Model 1 is possibly due to the fact that the existing statistics does not include information from the middle of the month onward. Furthermore, the implication of this analysis is that even if information from after the middle of the month is available, performance would not improve unless machine learning techniques are used to account for nonlinearities.

These results indicate that the export indexes are useful in capturing real-time trends in Japan's exports. They also suggest that the nowcasting of real exports has an element of

---

[23] As part of the robustness check, we tried adding a one-period lag term for the year-on-year change rates of real exports as an explanatory variable in each formulation in Figure 14. The coefficient of the lag term turned out to be insignificant and did not affect the significance of the other variables. Therefore, we chose to use a formulation without the lag term in our main analysis.

[24] The nowcast values for the out-of-sample period are estimated on a rolling basis. That is, for example, the January 2021 nowcast is calculated using a model estimated using actual data through December 2020, and the February 2021 nowcast is calculated using a model re-estimated using actual data through January 2021. This process is repeated through March 2022 to compare the nowcast accuracy of each model.

[25] In this analysis, the out-of-sample period is set to January 2021 and beyond in order to have a sufficient number of samples when calculating export indexes using machine learning techniques. In this regard, we confirm that there is no change in the improvement in nowcast accuracy from model 1 through 4 even if we re-estimate model 1 through 4 with January 2020 and beyond set as the out-of-sample period for robustness checks,.

nonlinearity and that machine learning techniques may be useful in capturing such relationships. In this regard, the analysis can be considered consistent with Anesti et al. (2021), who argue that the advantages of using machine learning to deal with nonlinearities become large when using the alternative data as the amount of data and the complexity of the relationships among variables increase.

5.2. Comparison to the previous literature

As mentioned above, Cerdeiro et al. (2020) create indexes to capture trends in imports and exports by vessel type (car carriers, containerships, etc.) using AIS data for countries around the world, including Japan. In this section, we compare the performance of the indexes calculated by Cerdeiro et al. (2020) with that of the export index calculated in this analysis, and examine whether our method improves the nowcasting accuracy.

First, the correlation coefficients between the real exports and the export indexes calculated by Cerdeiro et al. (2020) for car carriers and containerships in Japan are 0.70 and 0.39, respectively, which are lower than those between the real exports and the export indexes calculated in this analysis (Figure 16). Moreover, in the OLS estimation where the year-on-year change rates of the real exports (seasonally unadjusted) are regressed on the export indexes, the adjusted coefficient of determination is higher for the export indexes calculated in this analysis (Figure 17). In addition, when we evaluate the nowcasting accuracy of the model by setting the period from January 2021 through March 2022 as the out-of-sample period, the model using the export indexes calculated in this analysis performs better (Figure 18). These results suggest that the export indexes calculated in this analysis are more useful for nowcasting Japan's export trends compared to those calculated by Cerdeiro et al. (2020).

These results can be attributed to the following differences in the analytical approaches of this paper and Cerdeiro et al. (2020). First, in identifying port areas, we use official geographic information data, while Cerdeiro et al. (2020) uses mechanical estimation based on vessel location data.[26] As a result, this paper is likely more accurate and has higher coverage in identifying the areas of ports in Japan than Cerdeiro et al. (2020). Second, with regard to the export index for containerships, we improve its accuracy by estimating port-specific export functions, while Cerdeiro et al. (2020) does not take such

[26] Specifically, an unsupervised learning approach is used to identify areas where many vessels are gathered as harbors based on the location information of anchored vessels.

an approach. In sum, the improvement in nowcasting accuracy can be attributed to the use of official statistics and machine learning techniques in refining the export index.

6. Conclusion

In this analysis, we calculate export indexes using AIS data, alternative data that records the information of vessels including their locations, to nowcast Japan's export. The results of the analysis show that the export indexes capture trends in real exports and compare favorably with existing statistics (provisional report of the *Trade Statistics of Japan*) that are also available in real-time. In particular, they are able to reasonably follow movements in exports when they increase or decrease significantly, such as in the spring of 2020, when the pandemic started, and around mid-2021, when the global supply chain was disrupted. In addition, in calculating the export indexes, we use geographic data on domestic ports and official statistics on export activities at each port, and apply machine learning techniques to improve the performance compared to the previous studies. These results suggest that AIS data, which has not been utilized much in economic analyses in Japan, can be useful in capturing Japan's export trends.

In this paper we assume that the relationship between vessel traffic and exports at each port is sufficiently stable. It should be noted, for example, that if the composition of items exported from a given port changes due to shifts in industrial structure and consequently the relationship between the amount of cargo carried by vessels leaving that port and the value of their exports also changes, the assumptions used for the export index calculated in this paper may no longer hold, and the accuracy of the nowcasting model may decrease. When utilizing AIS data for nowcasting trade in the future, it will be important to take this possibility into account and continue to make efforts to improve the accuracy of the analysis by updating the official statistics used to calculate export indexes as necessary.

Since our method can be applied to any types of vessels other than car carriers and containerships, building a nowcasting model of Japan's import trends using data on oil tankers and Liquefied Natural Gas carriers would be a promising topic for future analysis. Once nowcasting models of export and import trends are available, it will be also possible to conduct nowcasting of the trade balance.

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(5), pp. 657-665.

Anesti, N., E. Kalamara, and G. Kapetanios (2021), "Forecasting UK GDP Growth with Large Survey Panels," Bank of England Staff Working Paper, No. 923

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

Arslanalp, S., R. Koepke, and J. Verschuur (2021), "Tracking Trade from Space: An Application to Pacific Island Countries," IMF Working Papers, 21/225.

Ashwin, J., E. Kalamara, and L. Saiz (2021), "Nowcasting Euro Area GDP with News Sentiment: A Tale of Two Crises," ECB Working Paper Series, No. 2616.

Bolhuis, M. A. and B. Rayner (2020), "Deus ex Machina? A Framework for Macro Forecasting with Machine Learning," IMF Working Papers, Vol. 2020, Issue 45.

Cerdeiro, D. A., A. Komaromi, Y. Liu, and M. Saeed (2020), "World Seaborne Trade in Real Time: A Proof of Concept for Building AIS-based Nowcasts from Scratch," IMF Working Papers, 20/57.

Chikamatsu, K., N. Hirakata, Y. Kido, and K. Otaka (2021), "Mixed-frequency Approaches to Nowcasting GDP: An Application to Japan," *Japan and the World Economy*, vol.57: 101056.

Galbraith, J. W. and G. Tkacz (2015), "Nowcasting GDP with Electronic Payments Data," ECB Statistics Paper, No. 10.

Giannone, D., L. Reichlin, and D. Small (2008), "Nowcasting: The Real-time Informational Content of Macroeconomic Data," *Journal of Monetary Economics*, vol. 55(4), pp. 665-676.

Furukawa, K., R. Hisano, Y. Minoura and T. Yagi (2022), "A Nowcasting Model of Industrial Production Using Alternative Data and Machine

Learning Approaches," Bank of Japan Working Paper Series, No. 22-E-16.

Goodfellow, I., Y. Bengio, and A. Courville (2016), *Deep Learning*, MIT Press.Nymand-Andersen, P. and E. Pantelidis (2018), "Google Econometrics: Nowcasting Euro Area Car Sales and Big Data Quality Requirements," ECB Statistics Paper No. 30.

Hara, N., and S. Yamane (2013), "New Monthly Estimation Approach for Nowcasting GDP Growth: The Case of Japan," Bank of Japan Working Paper Series, No. 13-E-14.

Hayashi, F. and Y. Tachi (2022), "Nowcasting Japan's GDP," *Empirical Economics*, 2022.

Kameda, S. (2022), "Use of Alternative Data in the Bank of Japan's Research Activities," Bank of Japan Review, No. 22-E-1.

Maehashi, K. and M. Shintani, (2020), "Macroeconomic Forecasting Using Factor Models and Machine Learning: An Application to Japan," *Journal of the Japanese and International Economies*, vol. 58(C).

Nakazawa, T. (2022), "Constructing GDP Nowcasting Models Using Alternative Data," Bank of Japan Working Paper Series, No. 22-E-9.

Nymand-Andersen, P. and E. Pantelidis (2018), "Google Econometrics: Nowcasting Euro Area Car Sales and Big Data Quality Requirements," ECB Statistics Paper No. 30.

Okubo, T., K. Takahashi, H, Inatsugu, and M. Takahashi (2022), "Development of 'Alternative Data Consumption Index': Nowcasting Private Consumption Using Alternative Data," Bank of Japan Working Paper Series, No. 22-E-8.

Richardson, A., M. T. van Florenstein, and T. Vehbi (2021), "Nowcasting GDP Using Machine-learning Algorithms: A Real-time Assessment," *International Journal of Forecasting*, vol. 37(2), pp. 941-948.

Shintani, M. (2005), "Nonlinear Forecasting Analysis Using Diffusion Indexes: An Application to Japan," *Journal of Money, Credit and Banking* vol. 37(3) pp. 517-538.

Appendix: Estimation methodology using kernel methods

In this appendix, we outline how to estimate the export function from Equation (6) using kernel methods. First, by differentiating the loss function L_p by the parameter $\boldsymbol{\theta}_p$, we obtain the following equation

$$\frac{\partial L_p}{\partial \boldsymbol{\theta}_p} = \sum_t \left\{ \sum_{i,d} \boldsymbol{\theta}_p^T \phi(\boldsymbol{x}_{i,p,d}) - export_{p,t} \right\} \cdot \sum_{i,d} \phi(\boldsymbol{x}_{i,p,d}) + \lambda \boldsymbol{\theta}_p \tag{A1}$$

When the parameter $\boldsymbol{\theta}_p$ minimizes the loss function, the above equation is equal to zero, yielding the following equation.

$$\boldsymbol{\theta}_p = -\frac{1}{\lambda} \sum_t \{ \boldsymbol{\theta}_p^T \psi(\boldsymbol{x}_{p,t}) - export_{p,t} \} \cdot \psi(\boldsymbol{x}_{p,t}) \tag{A2}$$

where $\psi(\boldsymbol{x}_{p,t}) = \sum_{i,d} \phi(\boldsymbol{x}_{i,p,d})$ (d is a time in month t). Now, let $\boldsymbol{a}$ be a vector whose t-th element is $-\frac{1}{\lambda}\sum_t \{ \boldsymbol{\theta}_p^T \psi(\boldsymbol{x}_{p,t}) - export_{p,t} \}$ and $\boldsymbol{\Phi}$ be a matrix whose t-th row is $\psi(\boldsymbol{x}_{p,t})^T$. Then Equation (A2) can be rewritten as $\boldsymbol{\theta}_p = \boldsymbol{\Phi}^T \boldsymbol{a}$. Substituting this into the loss function L_p in Equation (6), we obtain the following equation.

$$L_p = \frac{1}{2} \boldsymbol{a}^T \boldsymbol{\Phi}\boldsymbol{\Phi}^T \boldsymbol{\Phi}\boldsymbol{\Phi}^T \boldsymbol{a} - \boldsymbol{a}^T \boldsymbol{\Phi}\boldsymbol{\Phi}^T \boldsymbol{y} + \frac{1}{2} \boldsymbol{y}^T \boldsymbol{y} + \frac{\lambda}{2} \boldsymbol{a}^T \boldsymbol{\Phi}\boldsymbol{\Phi}^T \boldsymbol{a} \tag{A3}$$

where $\boldsymbol{y}^T = (export_{p,1}, export_{p,2}, \dots)$. By denoting $\boldsymbol{\Phi}\boldsymbol{\Phi}^T$ as $\boldsymbol{K}$, Equation (A3) can be further written as follows.

$$L_p = \frac{1}{2} \boldsymbol{a}^T \boldsymbol{K}\boldsymbol{K} \boldsymbol{a} - \boldsymbol{a}^T \boldsymbol{K} \boldsymbol{y} + \frac{1}{2} \boldsymbol{y}^T \boldsymbol{y} + \frac{\lambda}{2} \boldsymbol{a}^T \boldsymbol{K} \boldsymbol{a} \tag{A4}$$

Equation (A4) expresses the loss function L_p in terms of the new parameter $\boldsymbol{a}$. Then, by differentiating the equation with respect to the parameter $\boldsymbol{a}$, we can obtain $\boldsymbol{a}$ that minimizes L_p as follows.

$$\boldsymbol{a} = (\boldsymbol{K} + \lambda \boldsymbol{I})^T \boldsymbol{y} \tag{A5}$$

where $\boldsymbol{I}$ is the identity matrix. The element of the nth row and mth column of matrix $\boldsymbol{K}$ can be written as follows

$$K_{nm} = \psi(\boldsymbol{x}_{p,n})^T \psi(\boldsymbol{x}_{p,m})$$

$$= \sum_{i,n'} \phi(\boldsymbol{x}_{i,p,n'})^T \sum_{j,m'} \phi(\boldsymbol{x}_{j,p,m'})$$

$$= \sum_{i,j,n',m'} \phi(\boldsymbol{x}_{i,p,n'})^T \phi(\boldsymbol{x}_{j,p,m'})$$

$$= \sum_{i,j,n',m'} k(\boldsymbol{x}_{i,p,n'}, \boldsymbol{x}_{j,p,m'}) \tag{A6}$$

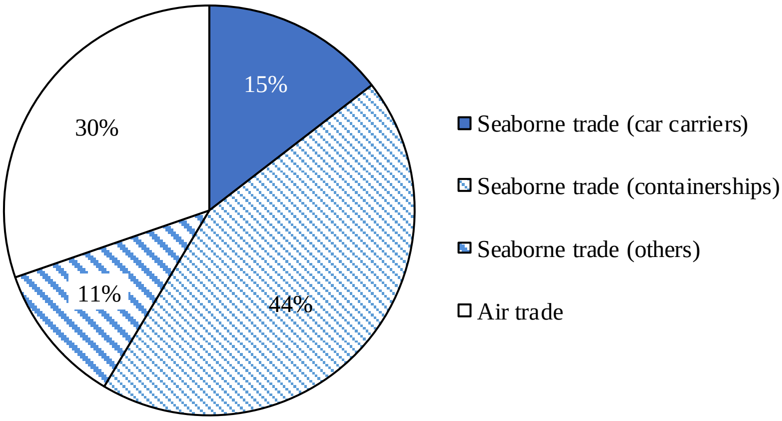
where n' and m' are the times in month n and month m, respectively. In Equation (A6), $k(\boldsymbol{x}_{i,p,n'}, \boldsymbol{x}_{j,p,m'}) = \phi(\boldsymbol{x}_{i,p,n'})^T \phi(\boldsymbol{x}_{j,p,m'})$ is called the kernel function. By combining the results of the above calculations, the following export function is obtained.

$$\begin{aligned} export_{i,p,d} &= \boldsymbol{\theta}_p^T \phi(\boldsymbol{x}_{i,p,d}) \\ &= \boldsymbol{a}^T \boldsymbol{\Phi} \phi(\boldsymbol{x}_{i,p,d}) \\ &= \boldsymbol{k}(\boldsymbol{x}_{i,p,d})^T (\boldsymbol{K} + \lambda \boldsymbol{I})^T \boldsymbol{y} \end{aligned} \tag{A7}$$

where $\boldsymbol{k}(\boldsymbol{x}_{i,p,d})$ is a vector whose n-th element is $\sum_{j,n'} k(\boldsymbol{x}_{j,p,n'}, \boldsymbol{x}_{i,p,d})$.

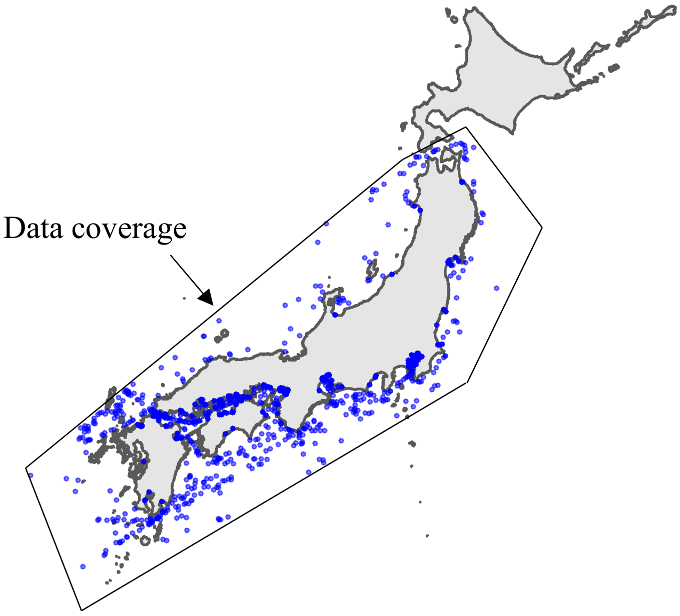
As Equation (A7) shows, in kernel methods the export function is expressed in terms of the kernel function $k(\boldsymbol{x}_{i,p,n'}, \boldsymbol{x}_{j,p,m'})$, not of the original vector $\phi(\boldsymbol{x}_{j,p,m'})$. In many cases, even if $\phi(\boldsymbol{x}_{j,p,m'})$ is of a complex form, $k(\boldsymbol{x}_{i,p,n'}, \boldsymbol{x}_{j,p,m'})$ can be relatively easy to compute. Hence, kernel methods can be used to estimate a variety of functions and make it possible to efficiently search for a form of the export function that results in high accuracy while reducing the amount of computation. In addition, the kernel function $k(\boldsymbol{x}, \boldsymbol{x}')$ has the property of taking large values when two vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ are similar. Therefore, the export function estimated using kernel methods can be interpreted as a function that predicts the amount of exports based on the amount of exports in months with similar distributions of metric tons of cargo and port call times.

Figure 1. Export shares by means of transportation



Note 1: The figures represent each item's share in nominal export values for the period between 2017 and 2021.
Note 2: All cars are assumed to be exported by car carriers.
Source: Ministry of Finance

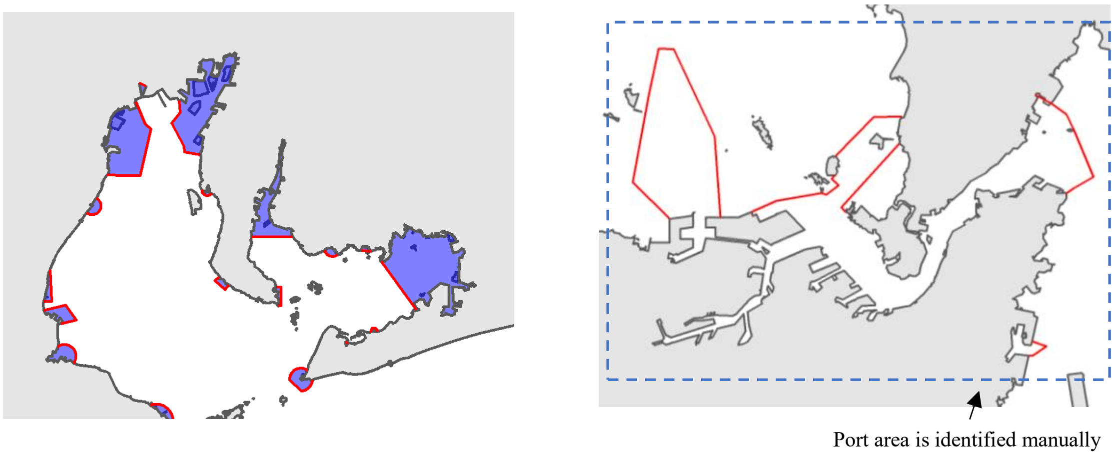
Figure 2. Distribution of vessels around Japan



Note: The dots show the locations of vessels on Jan. 1st, 2017.
Source: VesselFinder

Figure 3. Overview of port data

(a) Nagoya area (b) Kitakyushu area



Note 1: Panel (a) shows an example where port areas can be automatically identified. Panel (b) shows an example where port areas need to be identified manually due to the complicated coastal lines.
Note 2: The red lines represent boundaries of ports. The blue shadows show the identified area of each port.
Source: Ministry of Land, Infrastructure, Transport and Tourism

Figure 4. Relationship between amount of cargo and draught

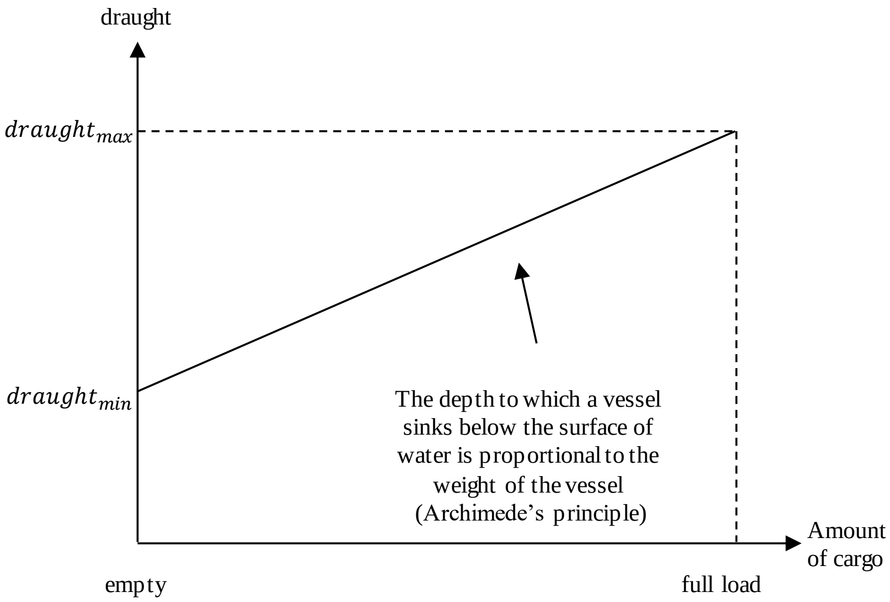
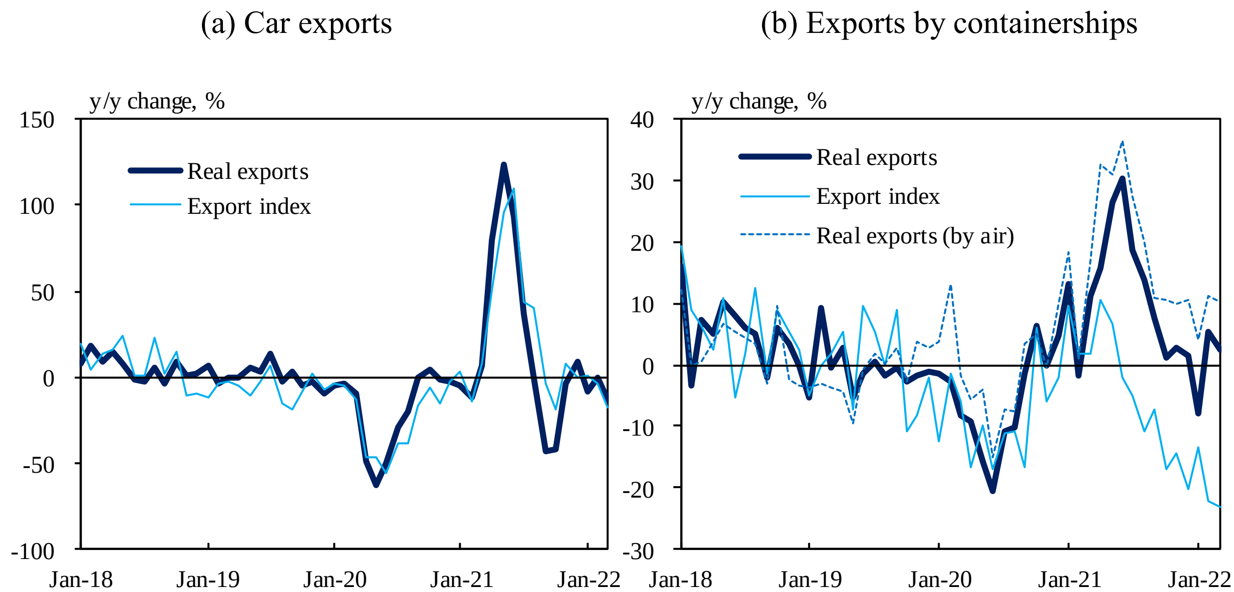


Figure 5. Example of unloading and loading by a vessel

Time of departure	Port	Ton of cargo at entry		Ton of unloaded cargo		Ton of loaded cargo	of which: exports		Ton of cargo at departure	Next destination
Apr. 1st	Kashima	100	−	20	+	30	(10)	=	110	domestic
Apr. 4th	Tokyo	110	−	30	+	20	(20)	=	100	domestic
Apr. 7th	Nagoya	100	−	10	+	50	(40)	=	140	domestic
Apr. 15th	Hakata (final port of call)	200	−	20	+	30	(30)	=	210	overseas

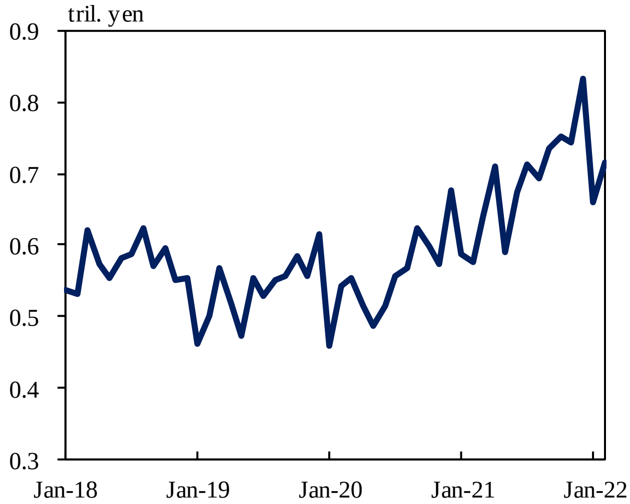
Note: Blue cells represent information that can be obtained from AIS data. If we assume that all goods loaded on vessels headed overseas are exports, the metric tons of cargo on a vessel when it leaves the final port of call (in this case 210) is equal to the sum of all exported goods loaded at domestic ports (in this case 10+20+40+…+30).

Figure 6. Export index and real exports



Note 1: The export index in Panel (a) is for car carriers, while the export index in Panel (b) is for containerships.
Note 2: The real exports by containerships and air in Panel (b) are calculated by dividing the nominal values of goods exported through the respective modes by each good's export price index.
Source: Authors' calculation based on data from Ministry of Finance

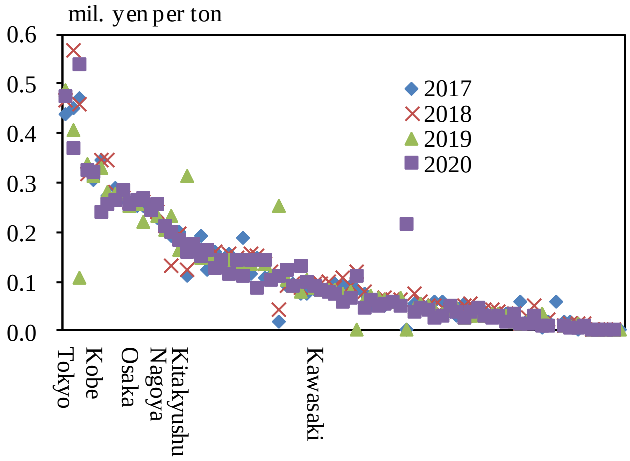
Figure 7. Exports of semiconductor-related goods



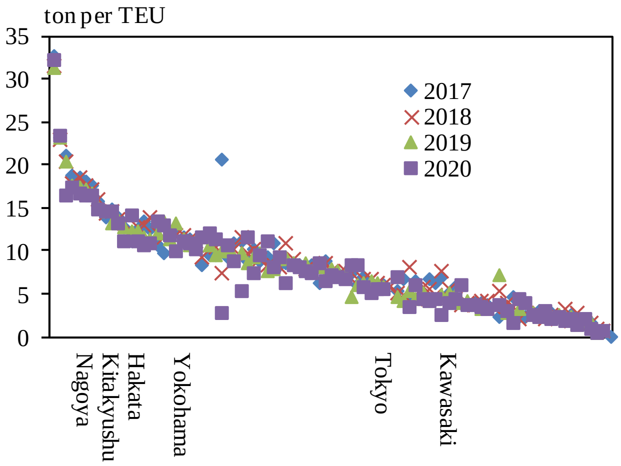
Note: The graph shows the export values of semiconductors and semiconductor production equipment.
Source: Ministry of Finance

Figure 8. Characteristics of individual ports

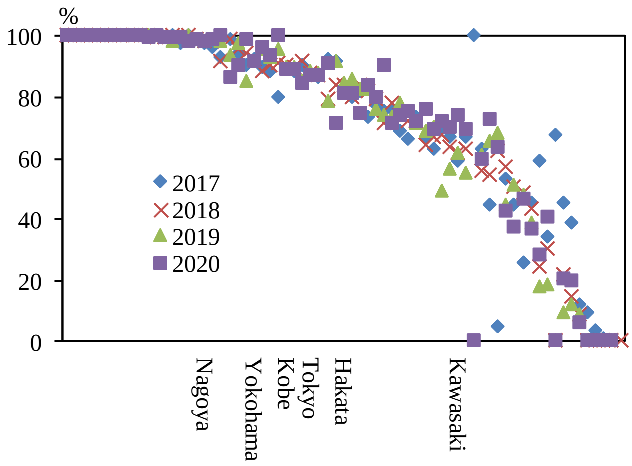
(a) Values per weight of exported goods



(b) Weight per container



(c) Share of exports in shipments



Note 1: TEU (Twenty-foot Equivalent Unit) is a measure of volume in units of twenty-foot long containers.
Note 2: Domestic ports are lined up on the horizontal axis, with labels for major ports.
Sources: Ministry of Finance, and Ministry of Land, Infrastructure, Transport and Tourism

Figure 9. Result of estimation of export functions using kernel methods

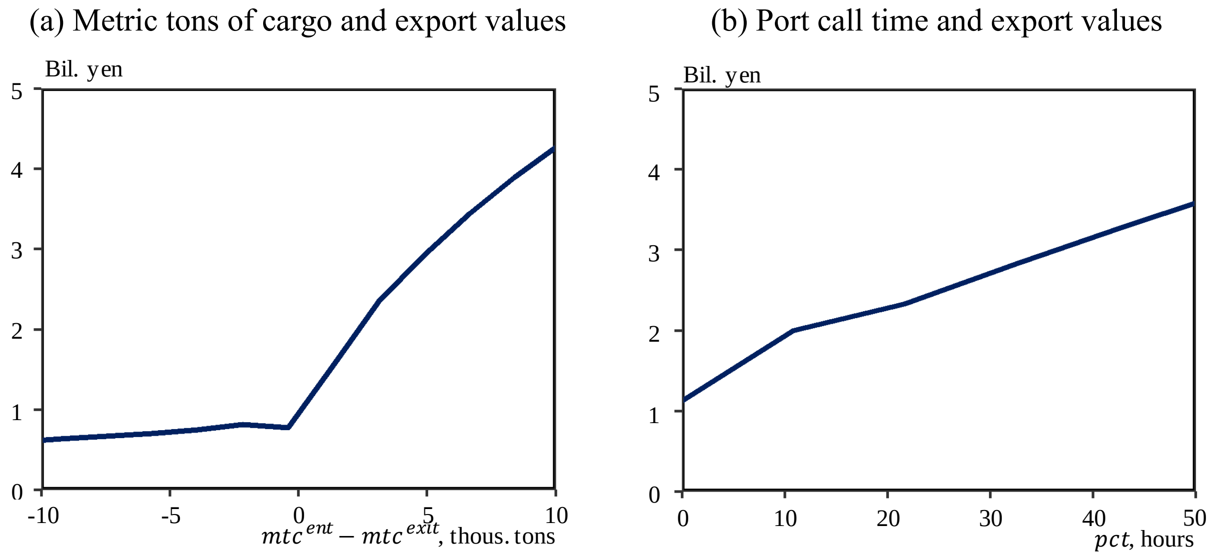
Name	$k(\boldsymbol{x}, \boldsymbol{x}')$	Description	Correlation coefficient	
			Training data	Test data
Polynomial of degree 1	$\boldsymbol{x}^T\boldsymbol{x}'$	Corresponds to $\phi(\boldsymbol{x}) = (x_1, x_2, \dots, x_n)$	0.83	0.82
Polynomial of degree 2	$(\boldsymbol{x}^T\boldsymbol{x}')^2$	Corresponds to $\phi(\boldsymbol{x}) = (x_1, x_2, \dots, x_1^2, \dots, x_1 x_2, \dots)$ (all combinations of each element of $\boldsymbol{x}$ up to 2nd degree)	0.84	0.86
Polynomial of degree 3	$(\boldsymbol{x}^T\boldsymbol{x}')^3$	Corresponds to all combinations of each element of $\boldsymbol{x}$ up to 3rd degree	0.89	0.24
Gaussian	$\exp\left(\frac{-\lVert\boldsymbol{x}-\boldsymbol{x}'\rVert^2}{2}\right)$	Corresponding ϕ is of infinite degree and cannot be written explicitly	0.99	0.91

Note 1: $\boldsymbol{x}$ is a vector whose elements are $x_1, x_2, \dots$

Note 2: The table shows the correlation coefficients between the year-on-year change rates of the export index calculated using each kernel function and of real exports by containerships.

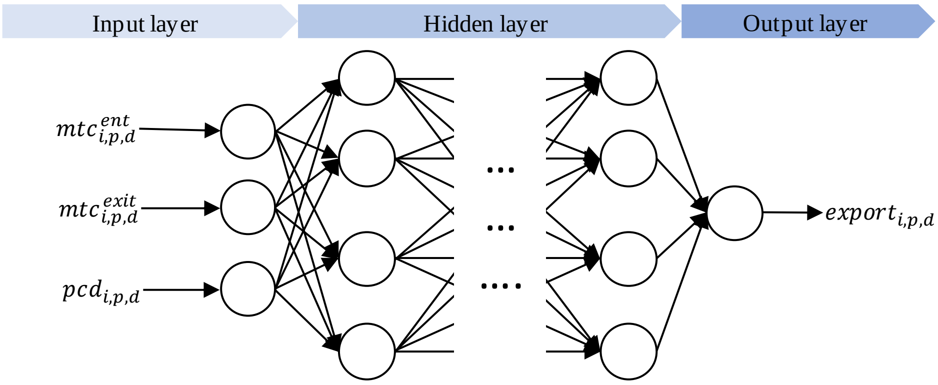
Note 3: The training data are for the period January 2017-December 2020 and the test data are for the period January 2021-March 2022. The export function is estimated based on the training data and the correlation coefficients are calculated for both data periods.

Figure 10. Estimated export function using Gaussian kernel



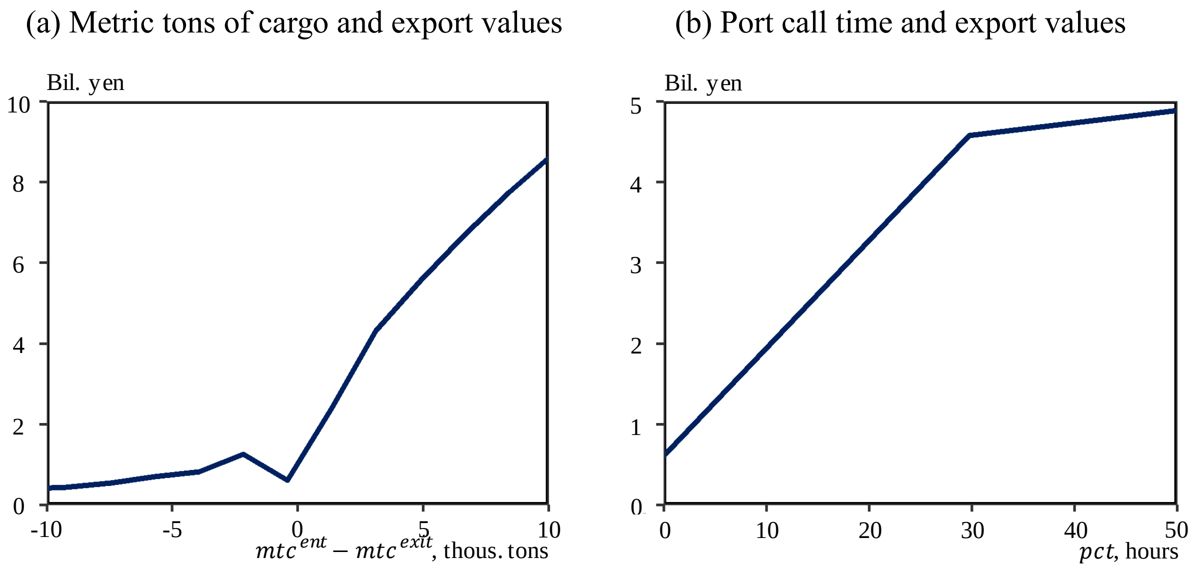
Note: The figures show the smoothed lines based on generalized additive models.

Figure 11. Overview of deep learning



Note: The circles in the figure represent neurons, and the arrows show how the output result of one neuron fed into the next neurons. See Goodfellow et al. (2016) for details.

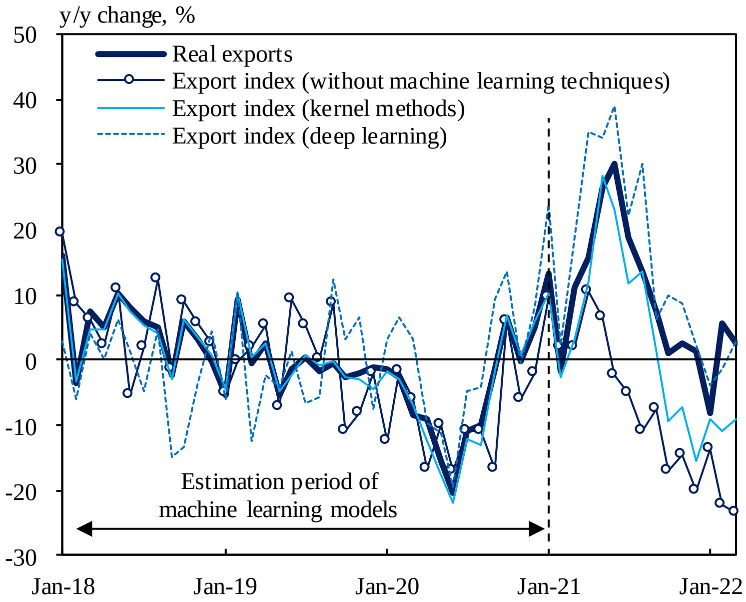
Figure 12. Estimated export function using deep learning



Note: The figures show the smoothed lines based on generalized additive models.

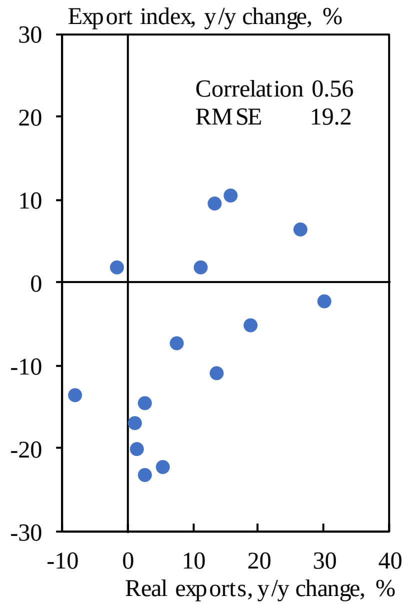
Figure 13. Export indexes for containerships

(a) Time series

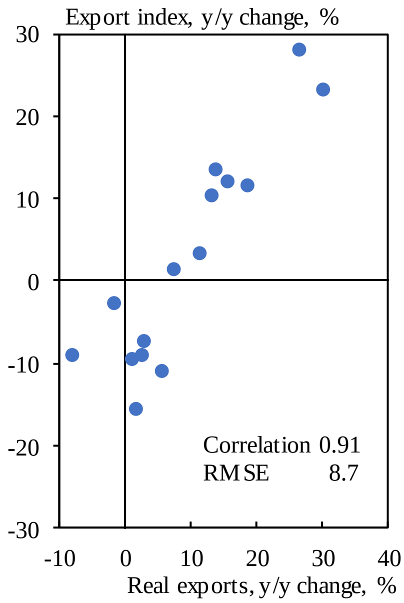


(b) Out-of-sample period

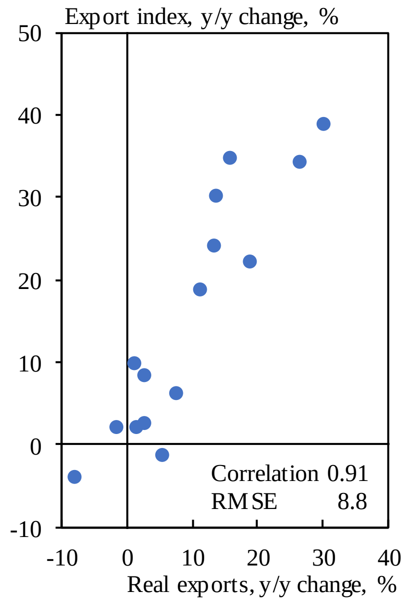
Without machine learning techniques



Kernel methods



Deep learning



Note: Real exports are real values of exported goods by containerships. Panel (b) shows the data between Jan. 2021 and Mar. 2022 (out-of-sample period).
Source: Ministry of Finance

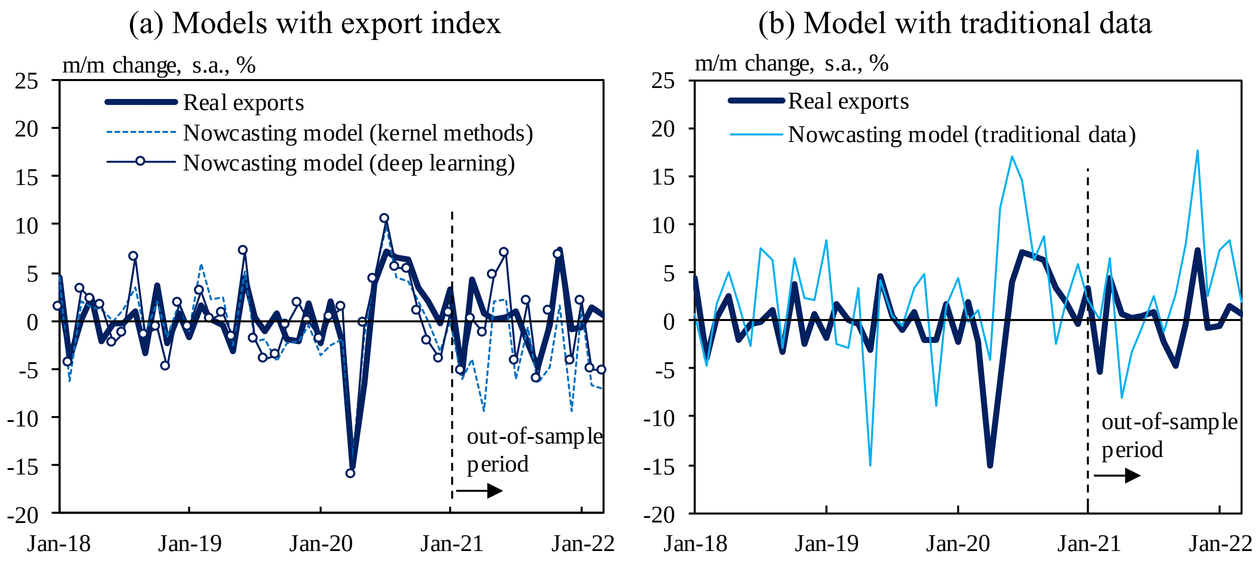
Figure 14. OLS estimation result

Dependent variable: real exports

	(1) Kernel methods	(2) Deep learning	(3)
const.	1.38 **	0.51	1.70 ***
	(0.54)	(0.50)	(0.50)
Export index (car carriers)	0.31 ***	0.28 ***	
	(0.03)	(0.02)	
Export index (containerships)	0.18 *	0.24 ***	
	(0.10)	(0.05)	
Trade statistics (provisional)			0.59 ***
			(0.03)
Adj. R2	0.88	0.91	0.66

Note 1: We take year-on-year change rate of each variable. The estimation period is between Jan. 2018 and Mar. 2022.
Note 2: The figures on the parentheses are standard errors. *, **, and *** represent statistical significance at 10%, 5%, and 1%, respectively.
Note 3: In column (1) and (2), we use the export indexes for containerships based on kernel methods and deep learning, respectively.
Note 4: The adjusted coefficient of determination of the AR(1) model for real exports for the same estimation period is 0.67.

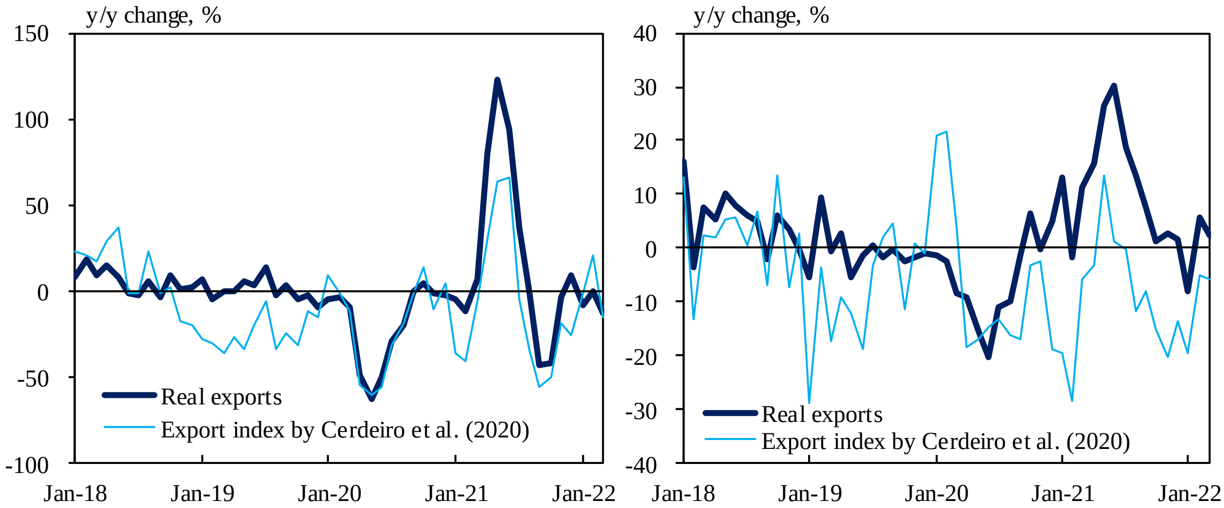
Figure 15. Estimation result of nowcasting models



Note 1: "Nowcasting model (kernel methods)" and "Nowcasting model (deep learning)" are estimated using the export indexes based on kernel methods and deep learning, respectively. "Nowcasting model (traditional data)" is estimated using the provisional report of the *Trade Statistics of Japan*.
Note 2: The RMSEs for the out-of-sample period (Jan. 2021-Mar. 2022) are 5.04 for "Nowcasting model (kernel methods)", 3.64 for "Nowcasting model (deep learning)", and 5.80 for "Nowcasting model (traditional data)." The RMSE is 5.70 when a nowcasting model is estimated with the export index for containerships without applying machine learning techniques.

Figure 16. Export index by Cerdeiro et al. (2020)

(a) Car exports (b) Exports by containerships



Sources: Cerdeiro et al. (2020) and Ministry of Finance.

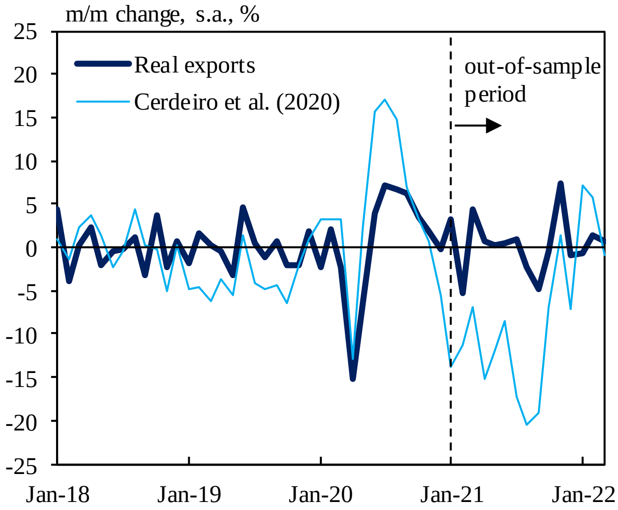
Figure 17. OLS estimation result for Cerdeiro et al. (2020)

Dependent variable: real exports

	(1) Cerdeiro et al. (2020)	(2) Our model (kernel methods)	(3) Our model (deep learning)
const.	4.57 ***	1.38 **	0.51
	(1.25)	(0.54)	(0.50)
Export index (car carriers)	0.28 ***	0.31 ***	0.28 ***
	(0.05)	(0.03)	(0.02)
Export index (containerships)	-0.01	0.18 *	0.24 ***
	(0.12)	(0.10)	(0.05)
Adj. R2	0.47	0.88	0.91

Note 1: We take year-on-year change rate of each variable. The estimation period is between Jan. 2018 and Mar. 2022.
Note 2: The figures on the parentheses are standard errors. *, **, and *** represent statistical significance at 10%, 5%, and 1%, respectively.
Note 3: In column (1) we use the export indexes for car carriers and containerships calculated by Cerdeiro et al. (2020). In column (2) and (3), we use the export index for containerships based on kernel methods and deep learning, respectively.

Figure 18. Estimation result of a nowcasting model using the export index by Cerdeiro et al. (2020)



Note: The RMSE for the out-of-sample period (Jan. 2021-Mar. 2022) is 7.44.