Journal of the Royal Statistical Society Series A: Statistics in Society, 2026, **189**, 463–496
https://doi.org/10.1093/jrsssa/qnae157
Advance access publication 14 March 2025
Original Article

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society

Big brothering the economy: nowcasting and forecasting with port satellite images

Alessandro Spelta[1], Paolo Pagnottoni[2] and Nicolò Pecora[3]

[1]Department of Economics and Management, University of Pavia, Via San Felice, 5, Pavia 27100, Italy
[2]Department of Economics, University of Insubria, Via Monte Generoso, 71, Varese 21100, Italy
[3]Department of Economics and Social Sciences, Catholic University, Via E. Parmense 84, Piacenza 29122, Italy

Address for correspondence: Alessandro Spelta, Department of Economics and Management, University of Pavia, Via San Felice, 5, Pavia 27100, Italy. Email: alessandro.spelta@unipv.it

Abstract

The interplay between maritime commerce and macroeconomic aggregates such as gross domestic product (GDP) and international trade statistics is key for the understanding of global economic dynamics. In this paper, we construct a new indicator, the *Port Saturation Index*, based on near-to-real-time port satellite images, which are processed with deep learning techniques to extract relevant information on the size of ongoing maritime commerce. We investigate the efficacy of the index with an empirical application to nowcast economic aggregates and forecast expansions and contractions of European GDP, import, and export. Results show that the index is able to closely track real economic activity, particularly GDP, when the figure is still far from its official release, yielding a novel tool for policy makers to support decision-making.

Keywords: dynamic factor models, nowcasting, port containers, satellite images

1 Introduction

The use of alternative or unconventional data as high-frequency proxies for macroeconomic aggregates has taken hold in the literature over the recent past. In this context, high-frequency indicators representing the state of the economy can provide relevant and timely information on GDP and improve nowcasting performances (see, among others, Çakmakli & Demircan, 2023; Cimadomo et al., 2022; Ferrara & Simoni, 2023). Over the last decades, with the availability of a wide range of novel and timely indicators, research on short-term forecasting has made significant advances, developing several techniques to deal with the asynchronicity of data releases in official statistics, to accommodate for the inclusion of different frequencies, and to quantify the informational content of variables (Baffigi et al., 2004; Carriero et al., 2015; Foroni et al., 2015; Giannone et al., 2008; Hepenstrick and Marcellino, 2019; Kuzin et al., 2011; Marcellino and Schumacher, 2010; Parigi and Schlitzer, 1995; Rünstler and Sédillot, 2003), other than nowcasting (with) granular data such as regional indicators (Koop et al., 2020, 2024).

The theoretical advantage of high-frequency indicators in macroeconomic forecasting stems from several key mechanisms. Firstly, these indicators can capture economic dynamics in real-time, reducing the information lag that typically affects traditional macroeconomic data (Giannone et al., 2008). Secondly, high-frequency data can detect structural changes and economic turning points more quickly than conventional quarterly or monthly indicators (Ferrara & Simoni, 2023). Thirdly, these indicators often provide granular insights into specific sectors of the economy, allowing for a more precise decomposition of aggregate economic activity (Cimadomo et al., 2022). This theoretical framework explains why high-frequency indicators can significantly improve forecasting accuracy: they reduce the temporal aggregation bias inherent

Received: October 4, 2024. **Revised:** December 6, 2024. **Accepted:** December 18, 2024

in lower-frequency data (Marcellino and Schumacher, 2010), while their timely nature helps to capture rapid changes in economic conditions that might be smoothed out in traditional measures (Foroni et al., 2015).

Against this background, new avenues for researchers to capture real-time information and to enhance the precision of nowcasting and forecasting models have been opened. To give some prominent examples, survey data have been employed in an integrated approach with existing statistical nowcasting models to derive short-term forecasts of US real GDP density forecasts, improving their accuracy across different forecast horizons (Çakmakli & Demircan, 2023). The predictive power of Google Search Data has also been widely explored, with contributions such as that of Ferrara and Simoni (2023), which demonstrates that Google Search Data significantly enhance nowcasting accuracy, even when controlling for official variables. Also unemployment rates have been nowcasted through high-frequency Google Search Data on job searches and economic uncertainty, alongside claimant counts (Schiavoni et al., 2021), showing the relevance of such covariates in proxying the dynamics of unemployment rates. Recently, Tiozzo Pezzoli and Tosetti (2022) have utilized ground vibration time series data recorded in Beijing, China, to nowcast regional industrial production, demonstrating that vibration indicators can effectively track business cycle fluctuations, especially during economic crises. Additionally, sentiment data coming from articles published in 15 popular European newspapers have also been shown to contain signals improving real-time nowcasts of GDP growth for the Euro area (Ashwin et al., 2024). These different high-frequency indicators offer distinct advantages and limitations for economic forecasting. Google search data provide real-time insights into economic sentiment but requires sophisticated filtering to address noise (Ferrara & Simoni, 2023). Traditional regional indicators maintain consistency with official statistics but often face publication lags and revision issues (Koop et al., 2020, 2024). Survey data such as those from the Survey of Professional Forecasters (SPF) are also a valuable source of information for macroeconomic forecasting (Çakmakli & Demircan, 2023), though they present high levels of disagreement, especially during turbulent times, generating uncertainity around the estimates. In this work, we exploit satellite imagery which, despite weather-related limitations, generates near-to-real-time quantifiable indicators of physical economic activity, particularly valuable for tracking large-scale movements of goods and transportation flows across countries.

Granular satellite data on maritime commerce have been demonstrated to be a valuable source of information for monitoring supply chain disruption and nowcasting macroeconomic aggregates. As a matter of fact, maritime trade constitutes a substantial portion of global trade, with around 80% of international trade by volume being transported via sea (Canton, 2021), and containerized seaborne trade playing a primary role in global supply chain, accounting for roughly 46% of all international trade (Notteboom et al., 2022). In Bai et al. (2024), the authors construct a congestion index for a set of 50 major ports worldwide using the Automatic Identification System (AIS) data, which tracks high-frequency vessel movements, by analyzing locations, speeds, and headings of container ships. The new index is able to detect global supply chain disruptions and sheds light on the coexistence of elevated spare capacity for producers and supply scarcity in the retail market, highlighting how these factors influence output and price responses during supply chain disruptions. Using the AIS database, Mourougane et al. (2024) propose a method that identifies ports at a high granularity level to build an indicator of port congestion and trend in maritime trades, with the aim of providing relevant information to monitor trade, at the country and global levels. The AIS data have also been used to augment existing nowcasting methodologies to short-term forecast Japanese exports, and have proven to proxy the movements of exports quite well even during exogenous shocks, such as the COVID-19 pandemic, and disruption of the supply chain, such as the recent one occurred in mid-2021 (Furukawa and Hisano, 2022). Moreover, Yu et al. (2023) have recently employed satellite images to forecast global stock returns, providing evidence on the relationship between port container traffic and economic activity.

In this paper, we construct a novel index, which we call the *Port Saturation Index* (PSI), exploiting high-frequency, near-to-real-time satellite images. To build the index, we extract publicly available real-time data derived from the Sentinel-2 imagery provided by the European Space Agency (ESA) concerning the leading 16 container ports located in Europe over a 7-yr period, i.e. 2017–2023. We borrow methodologies from deep learning to perform a semantic segmentation of satellite images in order to accurately classify port areas covered by containers. Based on

this approach, we measure the time series of the proportion of port areas labelled as containers, and introduce a new weekly and monthly index, the PSI, which proxies the variation in the number of containers due to port operations and maritime traffic. This index serves as a high-frequency indicator for assessing variations in the degree of congestion and container traffic at leading ports.

In contrast to the recent related literature (Bai et al., 2024; Furukawa and Hisano, 2022; Mourougane et al., 2024), we make use of the ESA database which, unlike the AIS database, is publicly available and free to access, enhancing reproducibility and the possibilities to extend the present work. While AIS data provide detailed vessel-specific information through real-time tracking of individual ships, our approach using ESA Sentinel-2 imagery offers distinct advantages. AIS tracks vessels via transponder signals, delivering high-frequency data on positions, speeds, and headings. However, Sentinel-2 imagery provides a comprehensive view of port infrastructure, capturing actual container occupancy rates, and port congestion levels. Unlike AIS, which can be affected by compliance issues and signal interference, satellite imagery offers objective, manipulation-free measurements of port activity. A further distinctive trait of our approach is the introduction of deep learning-based semantic segmentation for precise container identification, as well as the creation of a standardized index which can be consistently applied across different ports and time periods.

We aim at validating the PSI as a high-frequency indicator by performing nowcasting and forecasting of quarterly and monthly variations in major macroeconomic aggregates, i.e. GDP, import, and export. To address the discrepancy between the frequency of the satellite-derived index and macroeconomic data, we employ a dynamic factor model (DFM) approach as in Giannone et al. (2008) and Bańbura and Modugno (2014).

First, we assess the informational content of the PSI when incorporated into a simple linear autoregressive (AR) model. We then evaluate the nowcasting performance across a spectrum of temporal horizons, with different ranges prior to the official release of the economic indicator. In order to yield a robust forecast comparison of the augmented model with the benchmark, we implement the superior predictive ability (SPA) test (Hansen, 2005). We conclude by conducting a short-term forecasting exercise to examine the power of the index in predicting expansions and contractions of macroeconomic aggregates.

Our empirical findings demonstrate that the resulting PSI is able to track major events that influenced the global and European economy, such as the trade war between U.S. and China, the Brexit period, the COVID-19 pandemic outbreak period, the phase of increasing shipping costs occurred in 2021 and the Russia–Ukraine war, as well as the recent period of trade uncertainty due to the emergence of further international conflicts in 2023. The index is particularly sensitive to large shocks to the economy, reaching its highest peaks in the proximity of Brexit and of the sharp decline in Europe seaborne trade during the COVID-19 period. Incorporating the PSI into a baseline AR model yields small but consistent improvements in predictive accuracy. This improvement is more pronounced for GDP forecasts, followed by exports and imports.

The index exhibits some appealing nowcasting and forecasting features. The application to nowcast macroeconomic aggregates of European countries with a medium-scale DFM shows the ability of our index to mostly improve nowcasting performances, particularly monthly nowcasts of GDP growth. To illustrate, including the PSI in a medium-scale DFM results in a reduction of in-sample root mean squared error (RMSE) of approximately 8% for monthly GDP with respect to the baseline model. Improvements are relatively smaller and less consistent when nowcasting higher-frequency and import/export aggregates. The predictive ability of the index is confirmed by the SPA test, which provides support on the fact that accounting for the PSI in the nowcasting model leads to outperform a competing model without the inclusion of the index, when nowcasting quarterly GDP. Moreover, in the case of monthly import/export nowcasts, the SPA test provides evidence that the inclusion of weekly PSI improves the forecasting capability of the model. While some improvement is quite generally observed across different forecast horizons, notable gains of including the index into the model arise in nowcasting when the official statistics release is still relatively far.

Finally, testing the efficacy of the index when employed as a predictor of extreme economic conditions, we find that it represents a good predictor for both quarterly GDP downturns and upturns, while it demonstrates only a modest predictive power for what concerns monthly import and export. This disparity likely stems from the different data frequencies—monthly for trade data and

quarterly for GDP—and the inherent volatility in higher-frequency trade statistics, other than the fact that the nature of the index more closely aligns with a comprehensive measure of economic expansion or contraction, rather than with a pure metric of trade volumes.

The remainder of this paper is structured as follows. Section 2 introduces the data collection and image preprocessing procedure used to construct the PSI. In Section 3, we outline the method for the PSI construction and the nowcasting methodology adopted. In Section 4, we illustrate the usefulness of the proposed index in nowcasting European macroeconomic aggregates. In Section 5, we study the extent to which the proposed index is able to predict short-term economic expansions and contractions. Section 6 presents some concluding remarks and future directions of research.

2 Data and container segmentation

In this section, we introduce the data, its collection and preprocessing, which prelude the construction of the satellite-based index. In particular, in subsection 2.1 we outline the data collection procedure and present some preliminary statistics. In subsection 2.2, we provide details on the data-preprocessing steps we follow to ensure standardization and to enhance data quality. Subsection 2.3 describes the training procedure and neural network architecture to solve the semantic segmentation task, i.e. to classify containers from image pixels.

2.1 Port satellite image acquisition

We use publicly available satellite imagery from the Sentinel-2 mission, which has been operational since 2017.[1] The Sentinel-2 satellites offer a high revisitation frequency, with a 5-day cycle at the equator and a 3-day cycle at higher latitudes. As we work with images acquired from solar synchronous satellites, each location is consistently revisited at the same local time, enabling regular time series observations. The Sentinel-2 mission provides multispectral images with a spatial resolution of 10 m, consisting of 12 spectral bands. These bands include Band B4 (red), Band B3 (green), and Band B2 (blue), which can be combined to produce an RGB image. For our analysis, we consider data at Level-1C, as these products are typically published within 100–180 min after the observation.[2]

We collect data covering the period from 1 January 2017, to 31 December 2023. We select the European ports which are listed on the top 100 ports globally as ranked by Lloyd's List's 'One Hundred Ports' index, based on their annual container throughput figures, ending up with 16 European ports.[3] These ports have been chosen for their status as major hubs in European and global maritime trade, ensuring that our analysis focuses on the most significant contributors to container traffic in the region. By selecting large, high-traffic ports, we aim to capture a representative cross-section of port activity and its economic implications. While this selection may not fully account for smaller or less active ports, it provides a basis for examining port dynamics and their relationship with broader economic trends, with our choice ensuring robustness and relevance of our findings to the primary drivers of European trade.

Identifying the container port area from satellite imagery is a challenging task, as a port typically comprises several subregions that are often geographically dispersed. This dispersion complicates the determination of the relationship between the container ports and their respective subregions. To address this issue, we delineate the boundaries of the port under study as follows.

We retrieve the port coordinates from GeoNames, a geographical database that covers all countries and contains over 11 million place names.[4] Subsequently, we define the port boundaries by exploiting the shape files provided by OpenStreetMap, which classifies land cover into different categories.[5] We extract only the zones labelled as 'industrial' within a radius of 20 km from the port coordinates centre. Then, we differentiate the container stack area of the port from other areas, and focus our analysis on the former. It is important to note that many artificial structures

[1] See, for reference, https://dataspace.copernicus.eu/.

[2] Higher-quality data products, such as Level-2A, offer improved colour accuracy due to the inclusion of an atmospheric correction algorithm and provide more detailed information through image semantic segmentation. However, these products are subject to considerable delays of 48–60 hr before being made available.

[3] See, for reference, https://lloydslist.com/one-hundred-container-ports-2022.

[4] See, for reference, https://www.geonames.org/.

[5] See, for reference, https://www.openstreetmap.org/.

Table 1. Number of port images per year

Port Name	2017	2018	2019	2020	2021	2022	2023
Algeciras	50	53	70	48	49	56	50
Antwerp	5	10	8	9	7	10	8
Barcelona	22	17	27	21	18	22	25
Bremen Bremerhaven	5	17	17	18	8	20	15
Felixstowe	7	27	18	15	14	15	13
Gdansk	13	22	15	20	17	17	13
Genoa	15	17	19	16	15	15	13
Gioia Tauro	40	32	44	44	44	52	40
Hamburg	7	19	15	18	11	17	14
Le Havre	7	27	17	22	17	17	20
London	7	27	18	15	14	15	13
Marsa	55	62	62	63	66	79	61
Piraeus	28	18	28	31	25	28	20
Rotterdam	11	24	25	21	14	20	19
Southampton	13	23	20	28	13	26	17
Valencia	24	33	37	26	21	26	32

Note. The table reports the number of port images for the 16 selected European ports over the 7-yr period 2017–2023. Each row represents a different port, while columns show the number of images for each year.

within the port, such as ship unloaders near shores, coal heaps awaiting transport, and buildings with distinctive roofs, can resemble containers in satellite images and potentially confuse deep learning models. To mitigate this issue, we manually mark the container stack area for the port under consideration.

Table 1 presents a detailed breakdown of image availability across the 16 largest European ports, including key logistics hubs such as Rotterdam, Hamburg, and Antwerp. The temporal distribution of images varies considerably both within and across ports. Notably, the dataset encompasses a total of 2,753 images, with annual totals ranging from a minimum of 309 in 2017 to a peak of 440 in 2019. This variability suggests a general trend of increasing data collection over the initial years, followed by some fluctuation in more recent periods. Among individual ports, Marsa stands out with the highest cumulative image count of 448 over the 7-yr span, sharply contrasting with Antwerp, which has the lowest total of 57 images. Such disparity is due to atmospheric conditions, and will be addressed within our index construction framework.

2.2 Data preprocessing

In this subsection, we present the data preprocessing which follows collection, where we employ established data standardization techniques to enhance data quality at different levels. Apart from the evident impact of clouds and fog on the quality of satellite images, variations in solar altitude angles across seasons also lead to different levels of sunlight and variations in shadow shapes, significantly affecting the lighting conditions of images. These factors can also pose challenges for deep learning models in accurately identifying containers from satellite images of ports. To address this issue, we employ histogram matching, a normalization technique widely utilized in satellite observations (Horn and Woodham, 1979; Rakwatin et al., 2007). This method adjusts the colour histograms of our observations by transforming each band value in the focal image, to align with the cumulative distribution function (CDF) of band values from a standard image. This correction process aims to mitigate potential measurement errors stemming from seasonal variations. We also perform contrast adjustment, by reducing atmospheric haze, relying on a dark channel prior, which is based on the observation that nonhazy images of outdoor scenes usually contain some pixels that have low signal in one or more colour channels (see He et al., 2010). Finally, since

our raw images are downloaded at different size, we resize them to 480×480 pixel, so to have a uniform input to train our deep learning architecture.

2.3 Training procedure and neural-network architecture

In this subsection, we describe the deep learning model architecture and training procedure for container segmentation within port satellite images. Identifying containers in ports can be approached as a semantic segmentation task (Yu et al., 2023), which is a growing research area in computer vision. In particular, our aim is to classify each pixel of an image as 'container', 'cloud', 'background', and 'noninterest area'. The area occupied by containers is then derived by computing the proportion of pixels classified as 'container' over the total number of pixels for each satellite image. This serves as a proxy for the number of containers present in the port. Changes in the number of containers can reflect shifts in global economic activities. Notice, however, that containers are typically stacked in multiple layers to optimize space. However, the exact number of layers cannot be accurately discerned from Sentinel-2 satellite images due to resolution limitations. As a consequence, we assume uniform layering across different container stacks.

To delineate container areas, we utilize the U-Net model (Ronneberger et al., 2015), trained with a sample of 309 images from the year 2017, which are manually labelled with the help of the segment anything model (SAM) developed by Meta.[6] SAM employs a sophisticated meta-learning strategy to achieve precise object segmentation in images. By incorporating meta-features, SAM enhances its ability to generalize across different object categories and datasets, thereby enhancing segmentation accuracy and adaptability across a wide range of visual contexts and scenarios. This approach allows us to generate daily container coverage maps for each of the 16 European port included in our study.

With the aim of automatically classify pixels in order to identify stacked containers in ports, we develop a deep learning pipeline utilizing a variant of convolutional neural network (CNN), specifically the U-Net. Generally, CNN-based models have proven effective in processing large-scale image datasets autonomously, leveraging multiple nonlinear processing layers to extract high-level features for tasks such as image classification (Krizhevsky et al., 2012) and semantic segmentation (Long et al., 2015). However, since traditional CNNs face challenges when detecting and monitoring large numbers of small, dynamically changing objects, we employ the U-Net architecture, originally designed for medical image segmentation, to perform per-pixel classification. The U-Net architecture includes an encoder–decoder structure where the encoding path extracts features through downsampling, while the decoding path upsamples the features and integrates skip connections to preserve detailed spatial information. This design allows U-Net to capture nuanced changes in container numbers within satellite images.[7]

Our U-Net architecture is extensively discussed in Appendix A. The training options of the baseline framework are set with a maximum of 500 epochs and a minibatch size of 16. The specified parameters in the training options include an initial learning rate (Learn.R.) of 0.01, ℓ_2 regularization (Regul.) of 0.01, a learning rate drop period of 10, a gradient threshold (Grad.Thr.) of 0.07, and the execution environment set to GPU, while data are shuffled every epoch. Starting from this baseline architecture, we train different U-Net models to find the one which best fits our task by tuning hyperparameters, as reported in Table 2.

In Table 2, we present the results of the segmentation network with different model parameters. For each setting of model parameters, we illustrate the following performance metrics. The Global Accuracy metric, which measures the proportion of correctly classified pixels, indicates the model's ability to distinguish containers from noncontainer elements in port areas. High Global Accuracy suggests effective congestion tracking, which is critical for an accurate port saturation indicator. Misclassification of areas adjacent to containers could inflate the port index, falsely

[6] See, for reference, https://segment-anything.com/.

[7] U-Net is effective in our study for complex port scenes and limited training data, but other approaches are also worth to be mentioned. Mask R-CNN (He et al., 2017) combines object detection and instance segmentation, ideal for densely packed containers. DeepLab (Chen et al., 2017) uses atrous convolutions for multiscale context, better handling varying container sizes. SegNet (Badrinarayanan et al., 2017) offers memory efficiency for large satellite images, while PSPNet (Zhao et al., 2017) captures global context, aiding in container distribution and port layout understanding. Traditional methods like thresholding are fast but may struggle with complex scenes, while edge detection outlines container shapes but may miss stacked ones. Region growing and the watershed algorithm (Beucher & Meyer, 1992) can help with homogeneous areas and separating touching containers, respectively.

Table 2. Model performance Results

Model setting	Global accuracy	Mean accuracy	Mean IoU	Weighted IoU	Mean BF-score
Baseline	0.800	0.748	0.432	0.734	0.570
Enc.Depth 3	0.849	0.733	0.474	0.787	0.649
Enc.Depth 5	0.825	0.559	0.412	0.758	0.619
Filt.6	0.888	0.776	0.547	0.827	0.626
Filt.10	0.806	0.717	0.435	0.739	0.627
Grad.Thr. 0.01	0.842	0.715	0.474	0.776	0.637
Grad.Thr. 1	0.809	0.705	0.426	0.743	0.653
Learn.R. 0.005	0.869	0.756	0.506	0.806	0.635
Learn.R. 0.05	0.217	0.494	0.223	0.174	0.572
Regul. 0.005	0.827	0.766	0.472	0.761	0.591
Regul. 0.05	0.787	0.691	0.402	0.717	0.606

Note. The table reports the results of the segmentation network for different model parameters. In particular, the label 'Baseline' refers to the network with parameters described in the main text. Columns report different performance metrics. The Global Accuracy is the ratio of correctly classified pixels, regardless of the class, over the total number of pixels. The Mean accuracy indicates the percentage of correctly identified pixels for each class. The Mean Intersection over Union (IoU) represents the Jaccard similarity coefficient. The Weighted Iou, averages the IoU of each class, weighted by the number of pixels in that class. The Mean Boundary F1 (BF) contour matching score indicates how well the predicted boundary of each class aligns with the true boundary.

indicating higher congestion. Mean accuracy assesses performance across individual classes-containers, background, clouds, and noninterest areas-ensuring precision in each category. Errors, such as confusing containers with infrastructure, could skew our PSI, affecting economic forecasts by misrepresenting container volumes. IoU (Intersection over Union) measures overlap accuracy between predicted and actual container areas, with Weighted IoU adjusting for class prevalence. High IoU is crucial for the index accuracy, as it limits misclassification of noncontainer areas that could distort congestion estimates, affecting economic models reliant on realistic congestion data. The BF-Score evaluates boundary accuracy, essential for clear container segmentation. High BF-Score prevents adjacent areas from being misclassified, reducing noise in PSI metrics. These metrics together ensure a robust segmentation model, where high accuracy strengthens our PSI reliability as an economic indicator.

Table 2 shows that the best performing architecture is obtained by fixing the number of first encoder filters to 6, while leaving unchanged the remaining parameter values of the baseline calibration. This results in a global accuracy of more than 88%. With an accuracy of 88%, our model effectively minimizes misclassification, ensuring high-quality port satellite imagery data that accurately reflects container movement and port congestion. This high accuracy means the model can reliably distinguish between containers and other port elements, avoiding false inflation of the PSI due to mislabelling. In other words, model outputs are robust enough to serve as a reliable nowcasting indicator, helping to predict economic activity with minimal adjustments. Accurate port imagery data also enhance the detection of real trends in port congestion, supporting timely responses to changing economic conditions and enabling more precise, data-driven economic insights.

In Figure 1, we showcase some examples of the output of our container recognition procedure. In particular, we select three ports, i.e. Bremen, Felixstowe, and Le Havre, displaying both the satellite histogram equalized images (upper row), and the probability that pixels are actually labelled as container (lower row). Brighter colours indicate a higher probability that a pixel is labelled as container. The figure shows that the outlined container recognition procedure works quite well in distinguishing portions of satellite images which are actually occupied by containers. Such preliminary assessment is an essential step before the construction of the index, as low segmentation performances could heavily affect robustness.

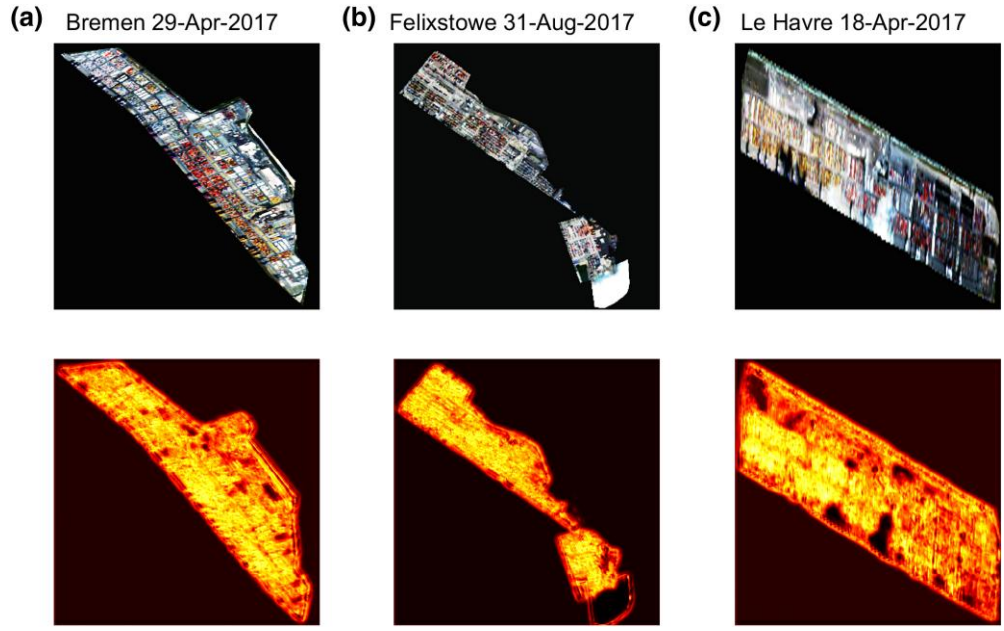

Figure 1. Container recognition results. The first row reports the sample satellite images recorded at various dates in 2017 related to three different ports, in order (left to right): Bremen (a), Felixstowe (b) and Le Havre (c). The second row contains the main output of the container segmentation model, that is a heatmap illustrating the probability that a pixel is labelled as container. Brighter colours indicate higher probabilities.

3 Methods

In this section, we outline the methods of the paper. In particular, in Subsection 3.1 we detail the construction of our proposed index, while in Subsection 3.2 we provide a description of the DFM employed for the nowcasting experiment.

3.1 PSI construction

We introduce a comprehensive index to measure port saturation, i.e. the *Port Saturation Index* (PSI). This index will serve as a pivotal tool for assessing the extent of container traffic and congestion at various ports and, thereby, as a high-frequency proxy to nowcast and forecast real economic activity. In what follows we illustrate the steps for constructing the index.

First, for each port in the sample, we compute the fraction of port area occupied by containers from the corresponding neural network output. Figure 2 illustrates the time series of the proportion of ports area labelled as containers.

From the figure, it is clear that, for most of the ports in the sample, seasonality occurs in the time series of the fraction of surface covered by containers (see, for instance, Algeciras or Piraeus, among others). A few other ports, instead, tend to behave less seasonally, such as Southampton or Le Havre. Moreover, ports like Gdansk or, partially, Piraeus and Gioia Tauro have witnessed an increase in the fraction of area occupied by containers, possibly due to an increased traffic over time, and/or to an enlargement of the area utilized for maritime traffic.

Given the purpose of nowcasting and forecasting European macroeconomic aggregates or predicting macroeconomic upturns and downturns, we summarize the dynamics of the different time series into a single index. For this reason, we estimate a one-factor DFM. In this way, we aim to explain the relationships among the proportion of surface covered by container in each port, using one single factor. This is achieved by estimating:

$$x_t = \mu_t + \Lambda F_t + e_t,$$

where x_t is a d-dimensional vector of the observed fraction of ports labelled as container, μ_t is a d-dimensional constant vector of means, Λ is a $d \times 1$ constant matrix of factor loadings, F_t is

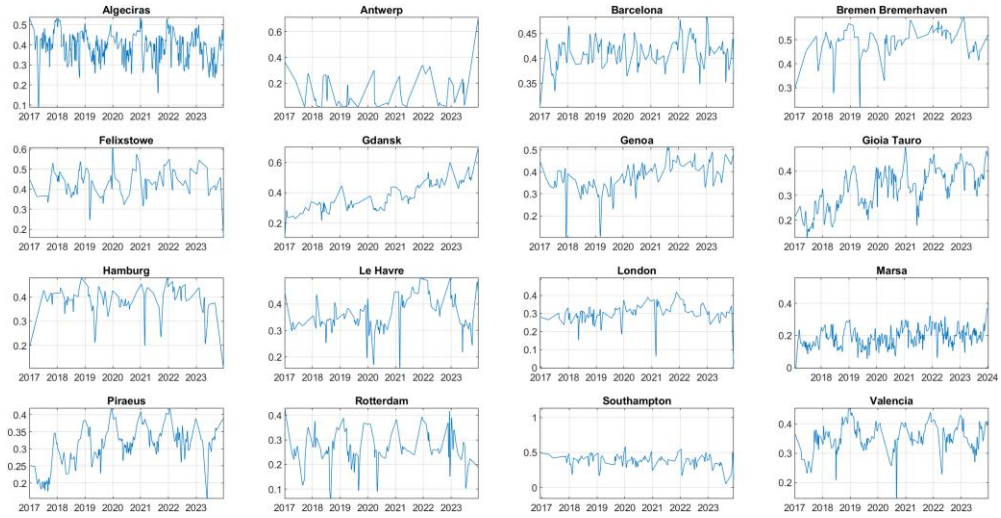

Figure 2. Time series of proportion of port areas occupied by containers. The figure reports, for each port, the time series dynamics of the estimated proportion of port surface covered by containers.

the 1-dimensional common factor, while e_t is a d-dimensional vector of independent specific factors. The common factor F_t, which encapsulates the dynamics of all series, is reported in Figure 3. From this figure, some considerations are evident. Firstly, seasonality occurs: while port saturation reaches its peak during winter, it tends to sensibly decrease and reaches its lowest levels over summer. In fact, peak shipping season runs from mid-August to the end of the year. This period sees high freight rates and tight capacity. Secondly, we also observe a growing global pattern, possibly due to various factors such as an increasing globalization trend and the rise of e-Commerce.

Based on these observations, we proceed by removing the linear trend from the data. This involves subtracting the best-fit straight line from the common factor series. After de-trending, we re-scale the time series so that the minimum value in the resulting series is zero. We call the resulting series $\hat{F}_t$. This process isolates the fluctuations from the trend, ensuring all values are non-negative and providing a normalized baseline for the subsequent analysis. Then, we define the seasonal component of the resulting time series as the 30-day moving median defined as:

$$F_t^s = \frac{1}{30}\sum_{i=0}^{29} \text{med}\ (\hat{F}_{t-i}),$$

where $\text{med}(\hat{F}_{t-i})$ represents the median value of the last 30 observations. We opt for the median instead of using a moving average given its robustness to outliers and asymmetries, which are frequently observable phenomena when analyzing real macroeconomic data. Finally, our PSI is obtained as the absolute value of the difference between the seasonal component F_t^s and the de-trended normalized series $\hat{F}_t$. In formulae:

$$\text{PSI}_t = |F_t^s - \hat{F}_t|. \tag{1}$$

We use the absolute value in defining our PSI because both extremely high and extremely low values of port saturation indicate higher congestion and, consequently, lower trading volume. An extremely positive value reflects significant congestion within the considered ports, while an extremely negative value suggests that containers may be abnormally stacked outside the European continent, both of which imply decreased trading activity. The diagram flow, which summarizes all the steps required to obtain the PSI, is illustrated in Figure 4. The figure depicts the flow of the methodology, covering all the steps from data acquisition to the implementation of nowcasting and forecasting models.

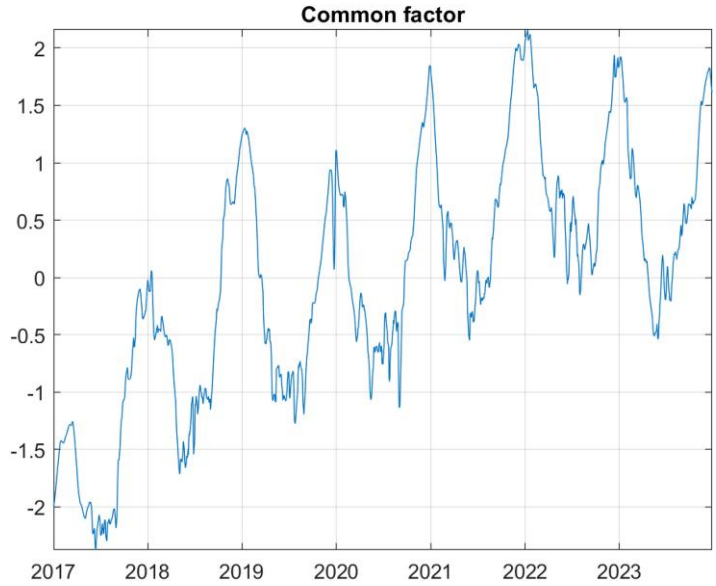

Figure 3. Estimated common factor time series. The figure shows the time series of the estimated common factor F_t related to the proportion of port surface occupied by containers.

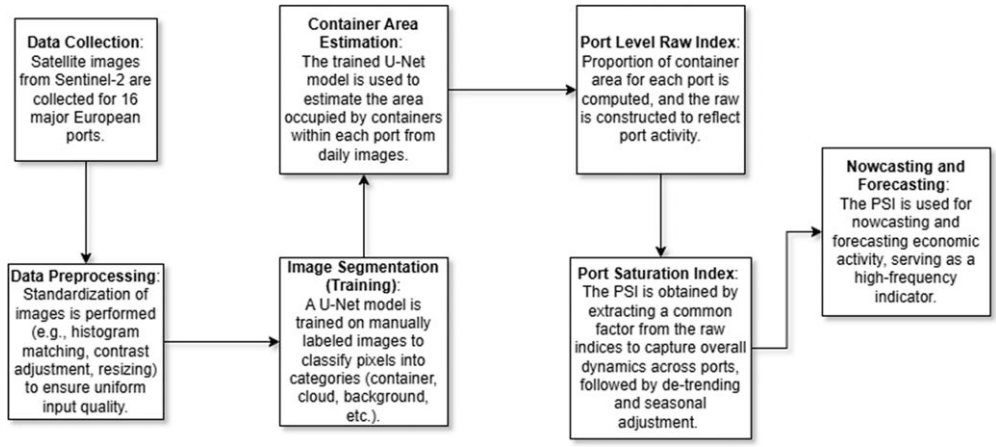

Figure 4. Diagram flow of our methodology. The figure shows the diagram flow of the methodology employed from data acquisition and Port Saturation Index construction to nowcasting and forecasting model implementation.

In Figure 5, we show the dynamics of the time series of the PSI. In the figure, we emphasize four time intervals which correspond to major events that influenced the world economy during the considered time span. In particular, we highlight the trade war between U.S. and China, as well as the Brexit period (red); the COVID-19 pandemic outbreak period (grey); the phase of increasing shipping costs occurred in 2021 (blue); the Russia–Ukraine war (green); and the recent period of trade uncertainty due to the emergence of further international conflicts (yellow). The figure shows how the time series dynamics of the PSI is intimately related to the actual state of the real economy. Indeed, the index shows large peaks in the correspondence of large shocks to the economy, such as in the vicinity of Brexit, where it reaches its maximum peak, and during the year of the COVID-19 pandemic outbreak. Other peaks are observable during periods of increased shipping costs, the Ukraine–Russia war and the recent trade uncertainty phase. During relatively tranquil periods, instead, the time series of the index assumes lower values, indicating less variation in the container coverage of ports.

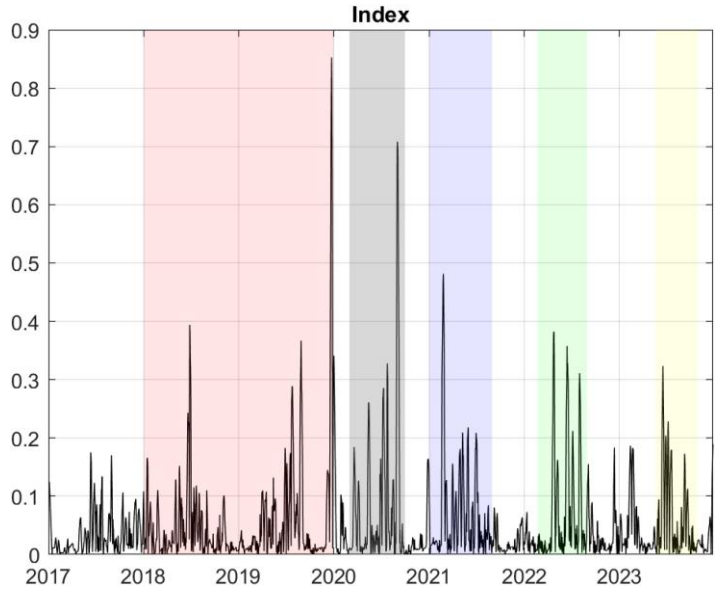

Figure 5. Port saturation index (PSI) time series. The plot represents the time series dynamics of the daily PSI. The shaded areas highlight the occurrence of major economic events. The first shaded area refers to the trade war between U.S. and China, and the Brexit period. The second area is associated to the year 2020 of COVID-19, while the blue bar highlights a phase of increased shipping costs. The third shaded area refers to the Ukraine–Russia war and the fourth area emphasizes a period of trade uncertainty due to the emergence of further conflicts.

Table 3. Augmented Dickey–Fuller (ADF) test on Port Saturation Index (PSI)

Lag	ADF	*p*-value
1	−14.73	<0.01
2	−14.61	<0.01
3	−14.91	<0.01
4	−13.12	<0.01
5	−12.14	<0.01

Note. The table reports the results of the ADF test on the PSI with lag order varying from 1 to 5. The first column refers to the lag order, the second column to the ADF test statistics, whereas the third column represents the corresponding *p*-values. The test statistics refers to an ADF test where the autoregressive model is specified with a constant and no time trend.

We test the presence of unit root by conducting an Augmented Dickey–Fuller (ADF) test on the PSI over the full sample. We let the lag order vary from 1 to 5, and implement ADF tests with a model specification with constant and no time trend. The outcomes of the ADF test are reported in Table 3. The resulting test statistics and *p*-values provide evidence against the null hypothesis of unit root, in favour of the alternative of stationarity of the time series in levels, for all the lag orders considered. The stationary behaviour of the index suggests that it can be safely used to be modelled in levels, rather than first differences, when dealing with traditional univariate or multivariate time series models.

To assess the forecasting capability of the PSI, we conduct a preliminary comparative forecasting experiment involving a simple autoregressive model of order p, namely AR(p), and an autoregressive model with exogenous covariate of the same order, ARX(p), by varying GDP, import and export growth as a dependent variable. We let the lag order of the autoregressive models vary $p = 1, .., 5$. The AR model serves as a baseline, utilizing only the historical values of the target variable to generate one-step-ahead forecasts. In contrast, the ARX model incorporates our

Table 4. RMSFE for AR(p) and ARX(p) models predicting GDP, import, and export

Lags	GDP		Import		Export	
	AR	ARX	AR	ARX	AR	ARX
1	3.3113	3.2882	3.2545	3.2341	3.9379	3.9217
2	3.2373	3.2069	3.1390	3.0684	3.7873	3.6673
3	3.2270	2.9144	3.1361	2.7203	3.7817	3.1607
4	3.2252	3.0065	3.1076	2.7547	3.7790	3.2018
5	3.2034	2.7119	3.0939	2.8797	3.7636	3.4666

Note. The table shows the root mean squared forecast error (RMSFE) of one-step-ahead forecasts for models with a lag order p varying from 1 to 5. The dependent variables considered are GDP, import, and export, whereas the models considered are AR(p) and ARX(p). The ARX(p) models incorporate our PSI as an exogenous covariate. Gross domestic product, GDP; autoregressive, AR; autoregressive with exogenous input, ARX.

proposed PSI as an additional explanatory variable, allowing us to evaluate the potential of the index to enhance predictive accuracy. This results in three distinct experiments to thoroughly examine the index forecasting power across different economic indicators for Europe.

Table 4 provides the root mean squared forecast error (RMSFE) values for AR and ARX models in predicting GDP, import, and export with varying lag orders. The ARX models incorporate the PSI as an exogenous covariate. Regarding the GDP predictions, the ARX model exhibits lower RMSFE values compared with the AR model, suggesting that the inclusion of the PSI as an exogenous variable improves the prediction accuracy. Additionally, the RMSFE values decrease as the number of lags increases, indicating that using more past observations can enhance the model predictive performance. In the case of import predictions, the ARX model outperforms the AR model, with lower RMSFE values. The RMSFE values generally decrease as the number of lags increases, but the improvement is more pronounced when moving from 2 to 3 lags. For export predictions, the ARX model demonstrates lower RMSFE values compared with the AR model, suggesting that the inclusion of the PSI enhances the prediction accuracy. Again, the RMSFE values decrease as the number of lags increases, with a more significant improvement when moving from 2 to 3 lags. Overall, the table highlights the impact of lag selection and the inclusion of the PSI as an exogenous variable on the prediction accuracy of the AR and ARX models for GDP, import, and export. The ARX model, which incorporates the additional predictor, generally outperforms the AR model, indicating the potential value of using the PSI in forecasting these economic indicators.

Figure 6 provides a visual representation of the comparative forecast performance between the AR and ARX models across the three key economic indicators, GDP, import, and export. The figure shows the RMSFE for the ARX model diverging from that of the AR model as the lag order increases. This suggests that at higher lag orders, the incremental benefit of the PSI variable increases, and the model exhibits better predictive accuracy. As far as import and export are concerned, the RMSFE of PSI-augmented ARX models is quite lower for lags greater than 2 relative to the one of the simple AR, though it gently increases for lags 4 and 5. This indicates that the PSI variable might provide a diminishing predictive power for import and export with higher lag specifications.

3.2 The econometric model

The usefulness of the proposed PSI is showcased by nowcasting the values of economic aggregates such as GDP and import/export. To this aim, we adopt a standard DFM along the lines of Bańbura and Modugno (2014); Giannone et al. (2008).[8]

Dynamic factor models (DFMs) offer a powerful and flexible framework for handling complex mixed-frequency, asynchronous time series data. Their main advantage lies in their ability to distill

[8] We also acknowledge other established techniques such as Mixed Data Sampling (Foroni et al., 2015; Hepenstrick and Marcellino, 2019; Kuzin et al., 2011; Marcellino and Schumacher, 2010), Mixed-Frequency Vector Autoregression (Carriero et al., 2015; Kuzin et al., 2011), bridge equations (Baffigi et al., 2004; Parigi and Schlitzer, 1995; Rünstler and Sédillot, 2003).

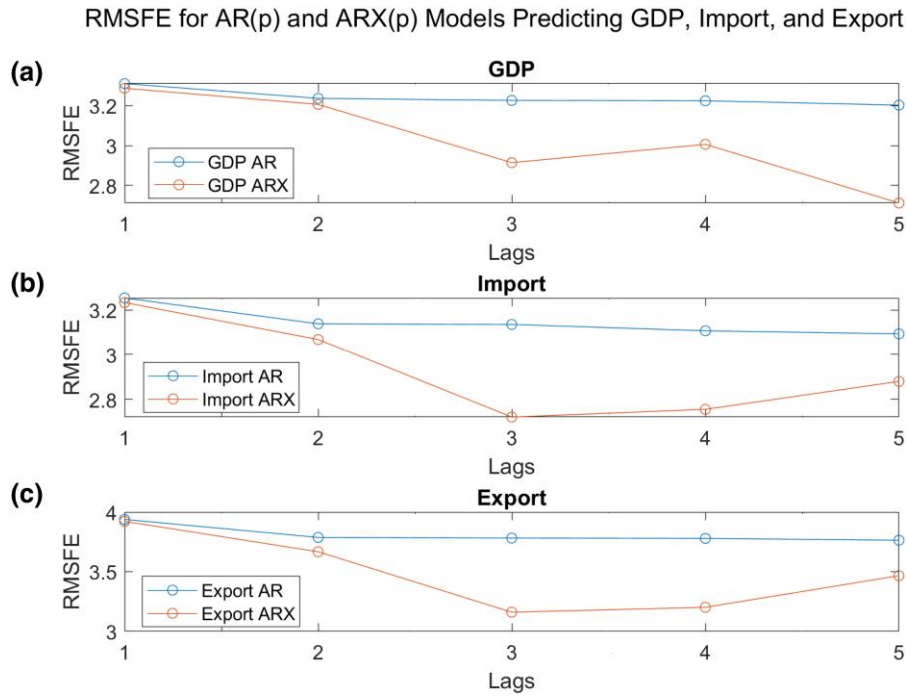

Figure 6. Comparative performance of autoregressive (AR) and autoregressive with exogenous input (ARX) models in forecasting key economic indicators. The figure visualizes the root mean squared forecast error for AR and ARX models in predicting gross domestic product (a), import (b), and export (c) across a range of lag orders ($p = 1$ to 5). The ARX models incorporate an additional exogenous covariate, namely our Port Saturation Index, while the AR models do not include this variable.

the most relevant information from numerous economic indicators into a few key underlying factors that drive economic activity. DFMs are particularly well suited for our analysis for three main reasons. First, they can efficiently process large datasets (Big Data), extracting common patterns while filtering out idiosyncratic noise. Second, their state-space representation allows for the use of Kalman filtering techniques, which is able to handle real-world data challenges such as different publication frequencies (e.g. monthly PSI versus quarterly GDP) and asynchronous data releases. Third, DFMs can incorporate new indicators like our PSI while accounting for their relationship with traditional economic measures. This approach, pioneered by Giannone et al. (2008) and Aruoba et al. (2009), has become the standard tool for economic nowcasting, widely adopted by central banks and financial institutions (Bok et al., 2018; Stock and Watson, 2017), as it provides a rigorous statistical framework for real-time economic monitoring.

We proceed by briefly revisiting the methodology of Bańbura and Modugno (2014). Let $y_t = [y_{1,t}, y_{2,t}, \ldots, y_{N,t}]'$, $t = 1, \ldots, T$, be an N-dimensional (vector) process, possibly containing missing observations. A DFM assumes that y_t admits the following factor model representation:

$$y_t = \Lambda \xi_t + \epsilon_t \quad \epsilon_t \sim \text{ i.i.d. } \mathcal{N}(0, \Sigma), \tag{2}$$

where ξ_t is a $R \times 1$ vector of (unobserved) common factors and $\epsilon_t = [\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{N,t}]'$ is the idiosyncratic component, uncorrelated with ξ_t at all leads and lags. The $N \times R$ matrix Λ contains factor loadings. It is assumed that ϵ_t is normally distributed and cross-sectionally uncorrelated, i.e. y_t follows an exact factor model. Further, the DFM assumes that the common factors ξ_t follow a stationary VAR process of order p:

$$\xi_t = A_1 \xi_{t-1} + A_2 \xi_{t-2} + \cdots + A_p \xi_{t-p} + u_t, \quad u_t \sim \text{ i.i.d. } \mathcal{N}(0, Q), \tag{3}$$

where $A_1, \ldots, A_p$ are $R \times R$ matrices of autoregressive coefficients. This econometric framework is adopted to nowcast both the GDP and the import/export time series. For the quarterly GDP, we provide monthly and weekly nowcast by aggregating our PSI at different frequencies. For the monthly import/export time series we only provide weekly nowcasts.

In line with the approach outlined by Mariano and Murasawa (2003), in the first exercise, we extend the framework to accommodate quarterly variables by creating partially observed monthly, or weekly, proxies for each of them. In the second exercise, we provide the corresponding weekly representation for each variable registered at monthly frequency. We follow the convention where the value of the quarterly variable is attributed to the third month of the respective quarter, while for the monthly variables, we assign the value of such variable to the last week of the corresponding month. Moreover, we assume that each month is composed by four weeks, as frequently assumed in the nowcasting literature (see Bańbura et al., 2013; Modugno, 2013).

Our choice of DFMs for the empirical analysis is motivated by their advantages in handling mixed-frequency indicators like the PSI. The DFM framework allows us to exploit the weekly frequency of our port saturation measure while maintaining consistency with quarterly GDP dynamics. Specifically, the factor structure captures co-movements between the PSI and traditional economic indicators, while filtering out port-specific fluctuations that may not reflect broader economic conditions. The model's state-space representation enables optimal weighting of the PSI contribution based on its historical relationship with GDP, automatically adjusting for the different publication lags of various indicators. This is particularly valuable as the PSI provides near-to-real-time signals, while traditional economic indicators often face substantial publication delays.

4 Nowcasting European macroeconomic aggregates

We now turn to the empirical results of our nowcasting experiments where we augment nowcasting models, specifically DFMs, with the PSI and evaluate their performances with or without its inclusion. We first consider GDP as the response variable for the short-term forecasts, and we subsequently model import/export, so to examine the explanatory power of the satellite-based index across different dependent variables reflecting the state of the real economy. Moreover, as far as GDP nowcasts are concerned, we employ our index both at monthly and weekly aggregation levels so to: (i) provide a more granular (weekly) forecast of the quarterly variable; (ii) study how the level of aggregation of the index impacts nowcasting performances.

The data used for the two nowcasting exercises includes key indicators of overall economic activity, encompassing industrial production, orders, retail sales, employment figures, manufacturing business survey data, inflation rates, and GDP, as in Cascaldi-Garcia et al. (2024), and our proposed PSI[9]. These variables have been broken down into more specific categories depending on the target variable we aim to nowcast. In Table 5, we provide a detailed list of the variables employed for the GDP nowcast, including their respective names, the unit of measurement, the frequency and any normalization transformations applied. Similarly, Table 6 contains covariates used in the context of the import/export nowcasting experiment. Notice that, in the latter case, along with macroeconomic variables, we also include financial predictors such as exchange rates and commodity index (see Appendix B for a detailed discussion on the selected variables employed for nowcast). After matching macroeconomic data with the data availability of port images, the final dataset we utilize covers a 7-yr period, ranging from the beginning of January 2017 to the end of December 2023.

We report, in Table 7, the in-sample RMSE when excluding (Without) or including (With) in the DFM specification the proposed PSI while nowcasting quarterly GDP growth. Results are presented for the PSI sampled at both weekly and monthly frequency. The outcomes contained in

[9] To determine the PSI, we employ a neural network model to classify each pixel of the satellite image into one of four categories: 'container', 'cloud', 'background', and 'noninterest area', as described in Section 2. The neural network is trained to identify the distinct features of containers within the port, distinguishing them from other elements such as clouds or nonrelevant areas. In particular, the last convolution layer of our U-Net maps the high-dimensional feature map to a set of class predictions for each pixel, allowing U-Net to classify every pixel in the image for semantic segmentation tasks. Once classification is completed, we compute the area occupied by containers by calculating the proportion of pixels identified as 'container' relative to the total number of pixels in the image. This proportion serves as a proxy for the number of containers present at each port, which constitutes the basis to build the PSI.

Table 5. Macroeconomic variables for gross domestic product nowcasting: units, frequency, and transformation

Series Name	Units	Freq.	Transf.
Gross Domestic Product	SWDA, Mil.Ch.2010.EUR	q	pch
Unemployment Rate	SA	m	lin
Industrial Turnover: Manufacturing	SWDA, 2015 = 100	m	pch
IP: Industry excl Construction	SA/WDA, 2015 = 100	m	pch
Industrial Production: Construction	SA/WDA, 2015 = 100	m	pch
PMI: Manufacturing Flash	SA, 50+=Expansion	m	lin
PMI: Services Business Activity Flash	SA, 50+=Expansion	m	lin
Retail Sales Volume Index	SA/WDA, 2015 = 100	m	pch
Consumer Confidence Indicator	SA	m	lin
Business Climate Indicator	SA, std-dev pts	m	lin
Exports of Goods	SA/WDA,Thous.EUR	m	pch
Imports of Goods	SA/WDA, Thous.Euros	m	pch
EU 28 excl Malta: New Passenger Car Registrations	NSA, Units	m	pca
Port saturation index	SA, Units	m/w	lin

Note. The first column identifies the names of the variables. The second column contains the unit of measurement while variable frequencies and transformations adopted are reported in the third and fourth columns, respectively. SA, seasonally adjusted; NSA, nonseasonally adjusted; WDA, working day adjusted; SWDA, seasonally working day adjusted. As concerns frequency (Freq.), m denotes a monthly variable, q identifies quarterly data while w is associated with the weekly frequency. Transformations (Transf.) are denoted by pch which refers to percentage change, lin is used when a variable is in levels (no transformation) while pca is the percentage annual change.

Table 6. Macroeconomic variables for import/export nowcasting: units, frequency, and transformation

Series Name	Units	Freq.	Transf.
Commodity Index	Units	w	pch
Exports of Goods	SA/WDA,Thous.EUR	m	pch
Imports of Goods	SA/WDA, Thous.Euros	m	pch
EU 28 excl Malta: New Passenger Car Registrations	NSA, Units	m	pca
Port Saturation Index	SA, Units	w	lin
EUR/USD	Units	w	pch
EUR/CNY	Units	w	pch
EUR/JPY	Units	w	pch
EUR/GBP	Units	w	pch
Consumer Confidence Indicator	SA	m	lin
Business Climate Indicator	SA, std-dev pts	m	lin
PMI: Manufacturing Flash	SA, 50+=Expansion	m	lin
PMI: Services Business Activity Flash	SA, 50+=Expansion	m	lin
Industrial Production: Construction	SA/WDA, 2015 = 100	m	pch
IP: Industry excl Construction	SA/WDA, 2015 = 100	m	pch
Consumer Price Index	SA/WDA, 2015 = 100	m	pch
Production Price Index	SA/WDA, 2015 = 100	m	pch

Note. The first column identifies the names of the variables. The second column contains the unit of measurement while variable frequencies and transformations adopted are reported in the third and fourth columns, respectively. SA, seasonally adjusted; NSA, nonseasonally adjusted; WDA, working day adjusted; SWDA, seasonally working day adjusted. As concerns frequency (Freq.), m denotes a monthly variable while w is associated with a weekly frequency. Transformations (Transf.) are denoted by pch, which refers to percentage change and lin is used when a variable is in levels (no transformation).

Table 7. GDP in-sample RMSE

Horizon	Without	With
Weekly	3.3869	3.3855
Monthly	2.9432	2.7079

Note. The table reports the results in terms of root mean squared error (RMSE) of the gross domestic product (GDP) in-sample forecasts excluding (Without) and including (With) the Port Saturation Index (PSI). The two rows refer to the different time aggregation of the PSI, i.e. weekly (first row) and monthly (second row).

Table 7 clearly show that including the PSI in the DFM specification improves forecast accuracy. Particularly, a reduction in RMSE is achieved both when including the weekly and the monthly PSI, with the latter being more effective in reducing the in-sample RMSE.

Figure 7 presents the difference in GDP forecast error between models that do and do not include the PSI variable. The results for the weekly nowcast are presented in the top panel while monthly nowcast results are contained in the bottom panel. The sum of forecast errors is higher in the monthly frequency case (34.5666) compared with the weekly frequency case (0.5105), indicating a better performance of the model including the PSI over the monthly frequency. Moreover, the bottom panel shows more pronounced spikes and troughs in the error difference over time, suggesting the PSI variable has a more variable impact on forecast accuracy at the monthly level compared with the smoother trend observed in the weekly frequency data (top panel). The largest positive differences in the monthly data occur in 2020, similar to the weekly data, highlighting the ability of the PSI variable to better capture the economic impacts of the COVID-19 pandemic compared with the model without it. While both figures demonstrate that the inclusion of the PSI variable generally leads to lower forecast errors, the magnitude and consistency of this improvement appears more pronounced in the monthly frequency nowcast compared with the weekly one.

The superior predictive ability (SPA) test (see Hansen, 2005) is used to evaluate whether our PSI provides additional and statistically significant information to the nowcasting model. Rejecting the null hypothesis suggests that the benchmark model that considers the PSI, is outperformed by the competing models that do not account for such a variable.

Table 8 reports the SPA test statistics for the monthly and weekly models. The SPA value for the monthly case, along with the high p-value, suggests that the nowcasting benchmark model with the inclusion of the PSI outperforms relative to the competing model. Conversely, for the weekly case, the lower p-value provides moderate evidence against the null hypothesis.

Furthermore, we evaluate the accuracy of the nowcast using a recursive out-of-sample exercise. We generate forecasts using a rolling-window framework spanning from September 2020 to December 2023. The rolling-window approach serves two purposes: it minimizes the impact of potential structural breaks and outliers, and it also circumvents challenges associated with comparing predictive performance among nested models (Medeiros et al., 2021). We produce forecasts for different time horizons. At each point within the forecast evaluation sample, we simulate real-time data availability and, similarly to Bańbura and Modugno (2014), we generate a series of projections for each reference quarter. The initial forecast is based on the data available in the first month of the preceding quarter, six months before the GDP flash release. The second forecast incorporates data available one month later, and the final forecast utilizes information from the last month of the quarter where the GDP is released. We conduct nowcast using data from the preceding and current quarters relative to the forecast reference quarter, denoted as Q(−1) and Q(0), respectively. Forecasts made in the first, second, and third month of a quarter are referred to as M1, M2, and M3, respectively. We assess the accuracy of such monthly out-of-sample forecasts via the RMSFE for different forecast horizons, namely from the first month of the preceding quarter, indicated as Q(−1)M1, to the third month of the current quarter, indicated as Q(0)M3. In Table 9, we report the RMSFE obtained when including and excluding the PSI from the model specification. The results show that, in four out of six out-of-sample monthly forecasts, the model including the PSI in its specification outperforms the one excluding it. Moreover, we notice that the monthly average out-of-sample RMSFE is lower when including

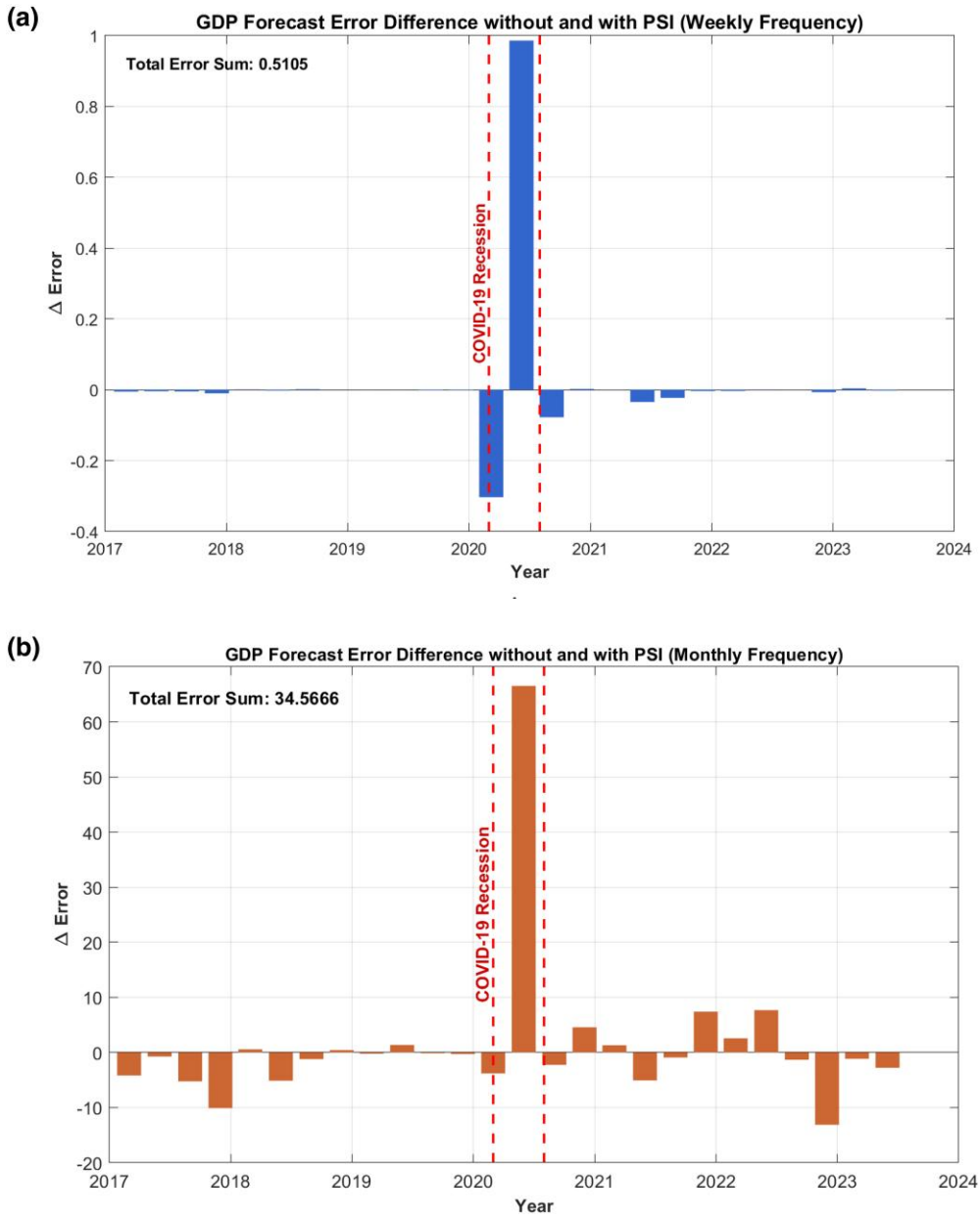

Figure 7. Gross domestic product (GDP) Forecast error difference. The figure presents the difference in GDP forecast error between models that do (a) and do not (b) include the PSI variable. The results related to weekly nowcasts are presented in the top panel. The bottom panel shows results related to the monthly nowcast.

the PSI. This confirms the ability of the monthly index to track economic activity not only when considering in-sample short-term forecasts, but also out-of-sample ones.

When employing the weekly PSI, we follow again Bańbura and Modugno (2014) to generate a series of projections for each reference month. The initial forecast is based on data available in the first week of the two months that precede the release. The second forecast incorporates data available two weeks later. According to this scheme, we conduct nowcasts using data from the current and the two preceding months, to forecast the reference month, denoted as M(−2), M(−1), and M(0), respectively. Forecasts, made every two weeks, are referred to as W1 and W3, respectively.

Table 8. Superior performance ability (SPA) test for gross domestic product nowcast

	Monthly	Weekly
SPA value	−0.48277	1.0761
p-value	(0.532)	(0.068)

Note. The table shows the results of the superior predictive ability (SPA) test across different aggregation frequencies. The benchmark model contains the Port Saturation Index (PSI), while the competing one does not include satellite information. The rejection of the null hypothesis indicates that the benchmark model with the inclusion of the PSI is outperformed by the competitor.

Table 9. GDP out-of-sample RMSFE with monthly PSI aggregation

Horizon	Without	With
Q(−1)M1	5.3511	5.0549
Q(−1)M2	4.5665	4.1273
Q(−1)M3	3.6767	3.4619
Q(0)M1	4.2149	3.6969
Q(0)M2	4.0155	4.0922
Q(0)M3	3.4921	3.5648
Aver.	4.2195	3.9997

Note. The table shows the root mean squared forecast errors (RMSFE) for varying forecast horizons, from month 1 of the preceding quarter, i.e. Q(−1)M1, to month 3 of the quarter of the gross domestic product (GDP) release, namely Q(0)M3. Aver. indicates the average error across all forecasting horizons. The two columns distinguish between the exclusion (Without) and inclusion (With) of the Port Saturation Index (PSI).

Table 10. Gross domestic product out-of-sample RMSFE with weekly PSI aggregation

Frequency	Without	With
M(−2)W1	0.9063	0.9062
M(−2)W3	0.8567	0.8566
M(−1)W1	0.8552	0.8550
M(−1)W3	0.8478	0.8464
M(0)W1	0.8478	0.8439
M(0)W3	0.8358	0.8413
Aver.	0.8582	0.8582

Note. The table shows the root mean squared forecast errors (RMSFE) for varying forecast horizons, from week 1 of the two month preceding the release, i.e. M(−2)W1, to week 3 of the month of the release, namely M(0)W3. Aver. indicates the average error across all forecasting horizons. The two columns distinguish between the exclusion (Without) and inclusion (With) of the Port Saturation Index (PSI).

In Table 10, we illustrate the out-of-sample RMSFE for varying forecast horizons, this time employing the weekly index. Results indicate that the model that includes the PSI overperforms the one without it in the majority of cases, that is five out of six. However, the average RMSFE is similar across specifications. Despite that, results overall suggest that also variations in the weekly index seem to exhibit a certain ability in tracking GDP growth.

We now investigate the capability of the weekly PSI in tracking European monthly import and export variables. To this end, Table 11 shows the monthly in-sample RMSE of import and export when excluding (Without) or including (With) the PSI in the DFM specification. The in-sample results, reported in Table 11, reveal a nuanced impact of the PSI. For export nowcasts, the inclusion of the index leads to an improvement in forecast accuracy, as highlighted by the reduction in the

Table 11. In-sample RMSE for import and export nowcast

Target Var.	Without	With
Export	4.0593	3.9497
Import	3.3903	3.4386

Note. The table reports the results in terms of root mean squared error (RMSE) of the in-sample forecasts for export (first row) and import (second row), excluding (Without) and including (With) the Port Saturation Index (PSI).

RMSE. Conversely, for import nowcasts, a slight increase in the RMSE is observed when the index is incorporated. These findings suggest that the PSI may capture more relevant information for export dynamics than for import patterns in the in-sample period.

Figure 8 presents the weekly differences in forecast errors for import (top panel) and export (bottom panel) nowcasts, comparing model performances with and without the PSI variable. In both cases, positive values reflect periods where PSI improved forecasting accuracy, while negative values indicate instances where PSI either increased error or provided no benefit. For import, the error differences are generally small and mixed until the COVID-19 recession in early 2020, where positive spikes show that PSI variables helps reduce errors during increased economic volatility. However, in the aftermath of 2020, the import error difference is mostly negative, as the total error sum (−27.6934), suggesting that PSI inclusion ultimately increases error slightly over the entire period. This indicates that while PSI is helpful during the pandemic, it is less effective during periods of economic stability. In contrast, the export model shows a different pattern. Although error differences for exports also remain modest before the pandemic, a pronounced negative spike followed by a positive one appears during the COVID-19 recession. Unlike the import model, however, the export model has a positive total error sum (73.7602), implying that PSI inclusion led to an overall improvement in forecast accuracy. This suggests that PSI variables is more consistently beneficial for export forecasts, particularly under conditions of economic stress.

To assess more rigorously the statistical significance of the improvements offered by the PSI, we also conduct the SPA test (Hansen, 2005). The corresponding results for import and export nowcasts reported in Table 12 demonstrate the effectiveness of incorporating the PSI. In fact, for export nowcasts, the negative test statistic and high p-value strongly indicate that the model including our proposal outperforms the alternative model without the satellite-based index. In the case of import nowcasts, instead, the lower p-value reveals that the superior performance of the benchmark model is less pronounced. These findings highlight the potential additional explanatory power of including the PSI in the import and export nowcasts model, particularly for export forecasts, and demonstrate that the augmented model maintains a relative superiority across both import and export nowcasts. The contrasting outcomes for exports and imports highlight potential asymmetries in how the PSI relates to different aspects of international trade. This discrepancy may be attributed to factors such as differential impacts of global supply chain disruptions, varying sensitivities to specific port activities captured by the index, or potential lag effects that differ between import and export processes.

In Tables 13 and 14, we present the RMSFE of the short-term forecasts of export and import for varying forecast horizons, when including and excluding the PSI in the model specification. We generate short-term forecasts from the first week of the two months preceding the release, indicated as M(−2)W1, until the third week of the month of the release, denoted as M(0)W3.

The out-of-sample results of Table 13 demonstrate a consistent improvement in the nowcast accuracy when incorporating the PSI. For export nowcasts, the average RMSFE decreases substantially when including the index. This improvement is observed across all forecast horizons, with a particularly notable enhancement in earlier forecast periods, i.e. relatively far from the official statistics release. Import nowcasts, as in Table 14, also show an improvement with the inclusion of the PSI, albeit to a lesser extent than exports. In fact, the average RMSFE for imports decreases across all forecasting horizons.

Overall, the out-of-sample results suggest that the PSI provides valuable real-time information for both import and export predictions, with a more pronounced benefit for export nowcasts. The

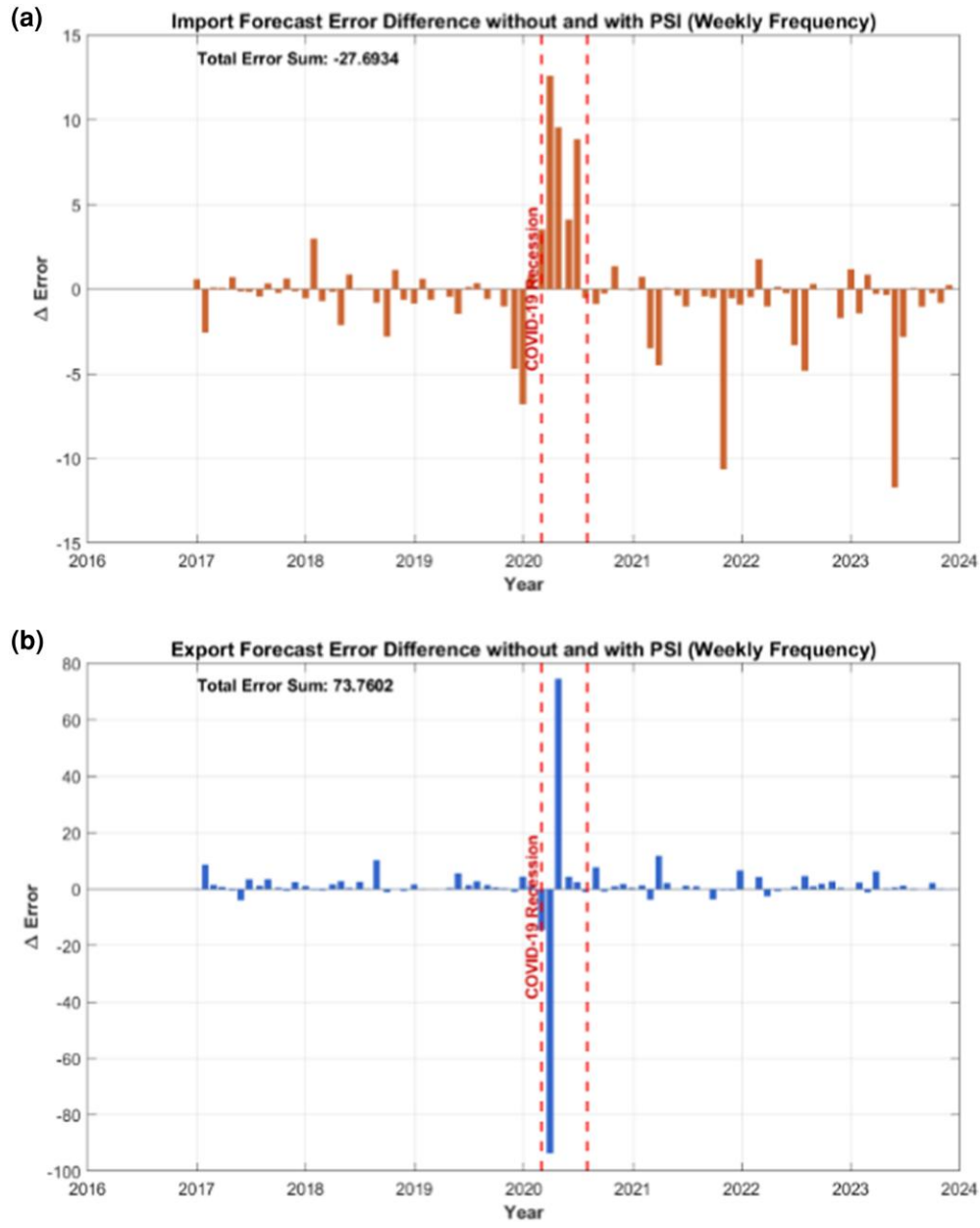

Figure 8. Forecast error differences for import and export nowcasts. The figure presents the difference in forecast error between models that do (a) and do not (b) include the Port Saturation Index variable. The results for the import variable are presented in the top panel. The bottom panel shows results for export nowcasts.

improved accuracy across various forecast horizons indicates that the index captures timely and relevant information about trade dynamics, potentially reflecting real-time changes in economic activity and global supply chain operations.

While ESA Sentinel-2 imagery provides frequent and stable observations, there are inherent limitations that must be acknowledged. One significant factor is the impact of weather conditions on data quality. Poor weather, such as clouds or fog, can obstruct satellite views, hindering the recognition of containers and port activities. To mitigate this, we specifically select Sentinel-2 RGB images with 5% or less cloud coverage from the Sentinel Hub EO Browser. Nevertheless, even

Table 12. Superior performance ability (SPA) test for import and export nowcast

	Export	Import
Value	−0.5945	0.9719
p-value	(0.4560)	(0.1744)

Note. The table shows the results of the SPA test both import and export nowcast. The benchmark model includes the PSI, whereas the augmented model does not incorporate satellite-derived data. The rejection of the null hypothesis indicates that the benchmark model is outperformed by the competing model.

Table 13. Export out-of-sample RMSFE with weekly PSI aggregation

Target Var.	Without	With
M(−2)W1	3.8398	3.0517
M(−2)W3	4.9394	2.6517
M(−1)W1	4.1776	3.1723
M(−1)W3	4.166	3.192
M(0)W1	3.7782	2.7098
M(0)W3	2.7206	2.3479
Aver.	3.9369	2.8542

Note. The table shows the root mean squared forecast error (RMSFE) of export nowcast for varying forecast horizons, from month 1 of the preceding quarter, i.e. Q(−1)M1, to month 1 of the quarter which follows, namely Q(+1)M1. Aver. indicates the average error across all forecasting horizons. The two columns distinguish between the exclusion (Without) and inclusion (With) of the Port Saturation Index (PSI).

Table 14. Import out-of-sample RMSFE with weekly PSI aggregation

Target Var.	Without	With
M(−2)W1	3.3742	2.8611
M(−2)W3	3.7190	3.2809
M(−1)W1	3.8567	3.0956
M(−1)W3	3.6273	3.3616
M(0)W1	3.5899	3.1955
M(0)W	2.9243	2.5247
Aver.	3.5152	3.0532

Note. The table shows the root mean squared forecast error (RMSFE) of Import nowcast for varying forecast horizons, from month 1 of the preceding quarter, i.e. Q(−1)M1, to month 1 of the quarter which follows, namely Q(+1)M1. Aver. indicates the average error across all forecasting horizons. The two columns distinguish between the exclusion (Without) and inclusion (With) of the Port Saturation Index (PSI).

with this criterion, there may still be instances where weather conditions affect data interpretation. Additionally, seasonal variations could influence port activities and container visibility, introducing another layer of complexity. Finally, technological disruptions, such as satellite maintenance or orbit changes, can also affect the continuity and consistency of data collection. Further, despite the 10m resolution of Sentinel-2 allowing for reliable container detection, some challenges are still present. At this resolution, each pixel represents an area roughly the size of a standard 40-foot container, meaning stacked containers are captured as a single unit, making it difficult to distinguish individual containers or accurately assess vertical stacking. This limitation can lead to underestimates of container volumes in ports with heavy vertical stacking. However, two factors help to mitigate this: first, consistent stacking practices in terminals ensure that while absolute counts may

be slightly underestimated, trends in port activity remain reliable; second, this limitation is consistent across all studied ports, maintaining a certain degree of validity in terms of comparative analysis. Higher-resolution satellite options exist (e.g. 30 cm), but their limited temporal coverage and high costs make them impractical for regular monitoring, other than the fact that they would only partially offset such limitations.

5 Predicting upturns and downturns of macroeconomic aggregates

This section examines the efficacy of our PSI as a real-time indicator for predicting expansions or contractions of real economic activity. In other words, we explore the potential of information regarding fluctuations in container proportions to serve as early warning signals for both economic downturns and upturns. To implement such analysis, we establish precise conventional definitions for economic downturns and upturns. A downturn is defined as a decrease in the (percentage) change of y_t (representing either GDP or import/export series) below the 10th percentile of its distribution. Conversely, an upturn is characterized by an increase in the (percentage) change of y_t above the 90th percentile of its distribution. To quantify downturn and upturn phenomena, we introduce two binary indices:

1. Downturn Index (Down_t):

$$\text{Down}_t = \begin{cases} 1 & \text{if } y_t < Q_{0.10}(y) \\ 0 & \text{otherwise} \end{cases}$$

   where $Q_{0.10}(y)$ denotes the 10th percentile of the y_t distribution;
2. Upturn Index (Up_t):

$$\text{Up}_t = \begin{cases} 1 & \text{if } y_t > Q_{0.90}(y) \\ 0 & \text{otherwise} \end{cases}$$

   where $Q_{0.90}(y)$ represents the 90th percentile of the y_t distribution.

These indices serve as binary indicators for economic downturns and upturns, respectively, facilitating the subsequent analysis of the relationship between port saturation and economic fluctuations. This approach is consistent with previous work in the field such as Manzan (2015), Barbaglia et al. (2023), Ferrara et al. (2022), and Tiozzo Pezzoli and Tosetti (2022). Moreover, the adopted definition of economic downturns and upturns is in agreement with established metrics of business cycle fluctuations, such as the NBER recession indicator.

To identify events in Down_t or Up_t, we employ our PSI following established crisis signaling methodologies (Spelta et al., 2020, 2021). The PSI serves as a continuous measure, indicating the likelihood of an economy being in a vulnerable or recovery state. This continuous measure is then converted into a binary prediction B using a thresholding approach. The threshold value is a critical parameter that balances the trade-off between sensitivity (correctly identifying crisis or recovery events) and specificity (avoiding false alarms). To comprehensively address this balance, we employ 1,000 evenly spaced points between the series' minimum and maximum as potential thresholds, pooling the results for analysis. To ensure consistency, we resample the PSI to match the frequency of our target variables before applying the threshold. The effectiveness of our predictions can be evaluated by comparing B to the ideal leading indicators Down_t and Up_t. This comparison is summarized in a contingency matrix, which allows us to compute the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUROC)—see Appendix C for more details on ROC and AUROC. These metrics provide a comprehensive assessment of our PSI predictive power for detecting economic crises and recoveries. Moreover, we implement the procedure in a recursive out-of-sample fashion so to assess the indicator's performance. This approach entails utilizing only information available up to the previous period to generate predictions for the next point.

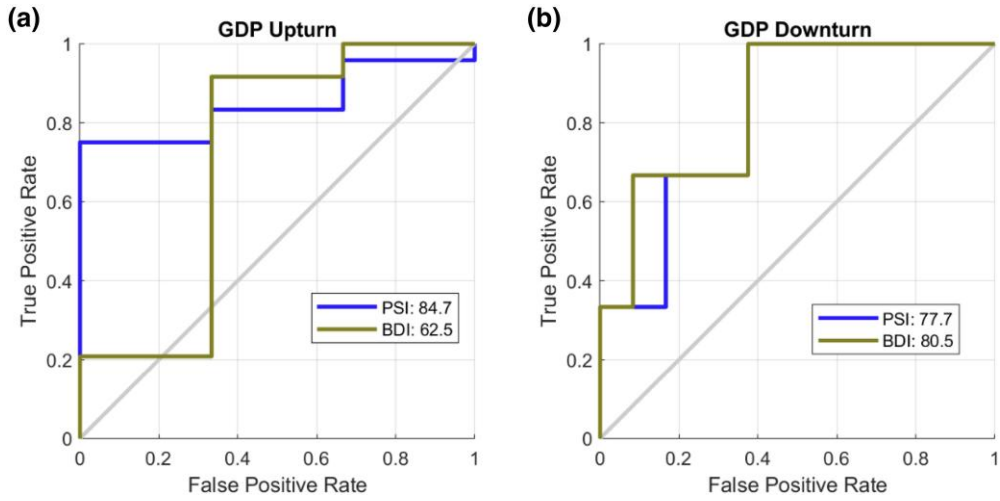

Figure 9. Receiver operating characteristic (ROC) curves of gross domestic product (GDP) upturns and downturns predictions. The figure shows the ROC curves derived from the Port Saturation Index (PSI0 employed for predicting GDP downturns (a) and upturns (b) together with the performance obtained by employing the BDI. Different colours distinguish ROC derived from the different indices and legends report the area under the ROC values associated to the curves. GDP upturns are defined as the values above the ninth decile of the GDP distribution while GDP downturns are determined as the GDP values below the first decile of the GDP distribution.

We now compare the results with those obtained through the use of an alternative indicator, the Baltic Dry Index (BDI), which is a key economic indicator that measures the cost of shipping raw materials, such as metals, coal, and grain, across the globe. Published by the London-based Baltic Exchange, the index reflects shipping rates across various vessel sizes and routes, offering insights into global trade and supply chain demand.[10] The ROC curves in Figure 9 highlight the predictive capabilities of both the PSI) and the BDI for forecasting GDP upturns and downturns. For GDP upturns, the PSI demonstrates particularly strong performance, with a steep initial ascent on the ROC curve, quickly achieving a high true positive rate at minimal false positive rates before plateauing at the maximum true positive rate. This shape indicates the PSI ability to identify GDP upturns with high accuracy and minimal false alarms. In contrast, the BDI shows moderate performance for upturns, with an AUROC of 62.5, suggesting that it is less effective than the PSI (AUROC 84.7) in identifying growth periods. For GDP downturns, however, the BDI shows an AUROC of 80.5, slightly outperforming the PSI (AUROC 77.7). This suggests that the PSI has a strong predictive capability for upturns while remaining a valid indicator also for economic slowdowns. Moreover, both indices significantly exceed the 0.5 threshold that would indicate random guessing, highlighting their effectiveness in anticipating major shifts in economic activity.

The ROC curves in Figure 10 compare the predictive performance of the PSI and the BDI in forecasting upturns and downturns for both export and import series. For export upturns, the PSI demonstrates moderate predictive power with an AUROC of 64.3, slightly outperforming the BDI, which records an AUROC of 62. Both indices show a stepped ROC curve pattern, indicating variability in predictive accuracy across different thresholds, though the PSI exhibits a marginally stronger performance. In contrast, for export downturns, the BDI outperforms the PSI. The BDI achieves an AUROC of 68, indicating a better predictive capability compared with the PSI AUROC of 58.3. Regarding import upturns, the PSI and BDI display similar predictive abilities, with the PSI still outperforming the competing indicator. The PSI achieves an AUROC of 62.5, compared with the BDI AUROC of 59.8. For import downturns, the PSI performs better than the BDI, with AUROC of 57.5 and 55.8, respectively.

Comparing the latter results to the GDP prediction performance, we observe a decline in predictive power. This discrepancy may be attributed to several factors. Firstly, the import and export series are monthly data, whereas GDP is quarterly. The higher frequency of the trade data

[10] https://www.balticexchange.com/en/index.html.

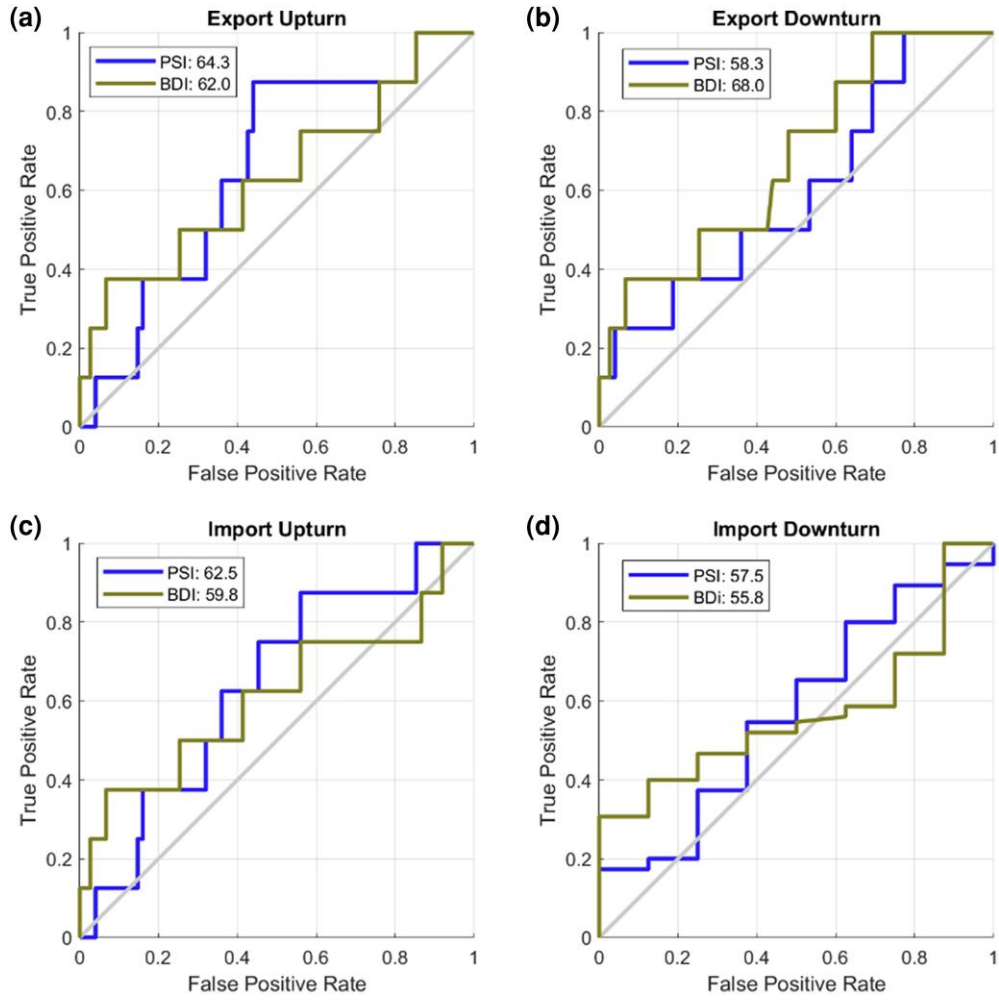

Figure 10. Receiver operating characteristic (ROC) curves of Import/Export downturns and upturns predictions. The figure shows ROC curves derived from the PSI employed for predicting downturns (a–c) and upturns (b–d) together with the performance of the BDI. Different colours distinguish ROC derived from the different indices and legends report the area under the ROC values associated to the curves. Import and export downturns are determined as the values of the series that lie below the first decile of the corresponding distribution while upturns are defined as the values above the ninth decile.

introduces more short-term volatility, potentially making it more challenging to predict accurately using the PSI. Additionally, GDP is a more comprehensive measure of economic activity, aggregating various economic factors beyond just trade. The PSI, being closely tied to trade activities, naturally shows a stronger correlation with overall economic performance (GDP) than with the more volatile and specific import and export figures. Overall, these ROC curves and their associated AUROC values provide compelling evidence that the PSI serves as a powerful predictor for both GDP downturns and upturns, while the PSI demonstrates only moderate predictive power for import and export series.

Finally, several factors could explain why import forecasts were not as significantly enhanced. Exports and imports often face distinct supply chain dynamics. Imports may be influenced by a wider range of factors, such as international production cycles, domestic demand fluctuations, and foreign trade policies, which the PSI might not fully capture. Additionally, export volumes largely respond to international demand, which can drive predictable patterns of container loading and congestion at ports. On the other hand, imports are more directly influenced by domestic

demand conditions, which can fluctuate based on local economic cycles, consumer demand, and policy changes. This makes imports more sensitive to internal economic shifts that are not always reflected in port congestion data alone. Finally, import volumes tend to be more volatile in response to changing domestic consumption trends and global supply chain disruptions and, accordingly, the PSI port-based snapshot may not fully capture these rapid shifts, leading to a weaker connection with import trends.

6 Concluding remarks

In this paper, we have proposed the use of near-to-real-time satellite image data to construct a new index, that we called PSI, which is helpful for nowcasting and forecasting economic variables. The PSI closely tracks the occurrence of major events provoking shocks to the economy, spiking at periods such as the trade war between U.S. and Brexit, the COVID-19 outbreak and the Russia–Ukraine war. The index also reacts to shipping costs and uncertainity in trade dynamics, making it a suitable indicator to mimic macroeconomic dynamics.

We have illustrated the ability of the index in tracking European macroeconomic aggregates by estimating dynamic factor models augmented by our weekly and monthly indicator to produce informed short-term forecasts. Our findings reveal that the inclusion of this novel index tend to improve nowcasting and forecasting performance, particularly for monthly GDP growth nowcasts. Further, the index exhibits some ability to predict upturns and downturns in the real economy, especially concerning comprehensive macroeconomic measures such as GDP rather than pure trade figures such as import and export.

Our main findings indicate that the use of satellite imagery to nowcast and forecast real economic activity is a promising avenue for future research, and it might be particularly effective for capturing the dynamics of international trade in real-time and at high resolution. However, there are several challenges associated with the quality of publicly available satellite data. Weather conditions, particularly cloud cover and fog, can obstruct the satellite's view and affect the accuracy of container detection. Seasonal variations may also influence the data quality, and although our filtering criterion of 5% maximum cloud coverage significantly reduces these issues, occasional weather disruptions remain a concern. In addition, alternative data products, such as Level-2A imagery with atmospheric corrections, offer enhanced accuracy in specific applications but come with the trade-off of longer processing times (48–60 hr), making them less suitable for real-time monitoring. Future research could explore combining Sentinel-2 imagery with other data sources, such as AIS or radar, to mitigate these challenges and further improve the accuracy and reliability of port activity measures, while maintaining public data accessibility and reproducibility.

The PSI faces several limitations which might foster future developments in this research direction. Firstly, regional heterogeneity in port infrastructure and operations affects comparability across single-port indices, as ports with different layouts or inland storage capabilities might show varying relationships between observed container density and actual trading activity. Secondly, the index assumes uniform container stacking patterns, potentially misrepresenting actual container volumes when ports employ varying stacking heights or nonuniform layering practices. Thirdly, while our semantic segmentation approach effectively identifies containers, it cannot distinguish between empty and full containers or capture turnover rates, potentially overestimating trade flows during congestion periods. These limitations could be, at least partially, addressed by: (i) integrating complementary data sources such as customs or port authority statistics; (ii) developing port-specific correction factors for structural differences in container handling; and (iii) incorporating alternative high-frequency indicators like truck traffic data to capture dynamics not visible from satellite imagery. Future research might explore machine learning techniques that better account for local port characteristics and stacking patterns, potentially improving the index's accuracy across different contexts.

In conclusion, while we show the usefulness of the PSI as a near-to-real-time economic tracker, there remains significant potential for future research. By incorporating advanced machine learning and artificial intelligence methods, accuracy, effectiveness and responsiveness of the PSI could be refined, by optimizing forecasts based on real-time observations, thus enabling more adaptive and precise decision-making. In the future, sharper, faster, and more frequently sampled satellite images, combined with novel techniques for image recognition and preprocessing and nowcasting,

can pave the way to a flourishing field for research and applications. The integration of PSI with other high-frequency indicators, such as those derived from the AIS database, or others pertaining to diverse domains could further enhance its predictive power. Matching satellite imagery with vessel tracking data such as the AIS database could foster the understanding, nowcasts and forecasts of global macroeconomic aggregates, as information such as vessel count, weight, speed, distance travelled, loading factor, and proxy of volume could be integrated with that of port container saturation. Also, future research could address alternative index formulations, as well as the development of nowcasting and forecasting techniques potentially able to take into account for the complex nature of high-frequency predictors of economic growth. Moreover, investigating its application in financial markets and supply chain monitoring could open new avenues for analysis and support informed decisions (see e.g. Bai et al., 2024; Yu et al., 2023).

Conflicts of interest: None declared.

Funding

This work has been partially supported by the Italian 'Ministero dell'Istruzione, dell'Università e della Ricerca' through the project 'A geo-localized data framework for managing climate risks and designing policies to support sustainable investments' (grant no. 20229CWYXC) within the PRIN 2022 program, funded by the European Union—Next Generation EU.

Data availability

Data are freely available.

Appendix A. U-Net architecture

Our U-Net architecture comprises a contracting path and an expansive path. The contracting path repeatedly applies two 3×3 convolutions (without padding), each followed by a rectified linear unit (ReLU) and a 2×2 max pooling operation with a stride of 2 for downsampling, doubling the number of feature channels at each step. The expansive path involves upsampling the feature map, followed by a 2×2 convolution (up-convolution) that halves the number of feature channels, and concatenation with the corresponding cropped feature map from the contracting path. This is followed by two 3×3 convolutions, each followed by a ReLU. Cropping is necessary due to the loss of border pixels in each convolution. The final layer uses a 1×1 convolution to map each 8-component feature vector into the desired number of classes. The depth of the encoder (Enc.Depth) is set equal to 4 indicating the number of times the input image is downsampled and upsampled during processing. The number of output channels for first encoder (Filt.) is set equal to 8. We use the weighted Cross-Entropy as loss function.

Figure A1 depicts the U-Net architecture, which is a popular convolutional neural network (CNN) model commonly used for image segmentation tasks. The U-Net architecture is characterized by its encoder–decoder structure with skip connections. The encoder part, on the left side, consists of a series of convolutional and pooling layers that progressively extract hierarchical features from the input image. This is the contracting or downsampling path, where the spatial dimensions of the feature maps are reduced while the number of feature channels increases. The decoder part, on the right side, consists of a series of deconvolutional (or upsampling) and convolutional layers that gradually restore the spatial dimensions of the feature maps. This is the expansive or upsampling path, where the feature maps are upsampled and combined with the corresponding feature maps from the encoder part through skip connections. The skip connections, represented by the dashed lines, allow the decoder to access and combine the high-resolution spatial information from the encoder, enabling the model to generate detailed segmentation maps. The 'Merge' layers in the centre of the image represent the point where the features from the encoder and decoder are combined, allowing the network to learn a more comprehensive and spatially aware representation of the input image for the port container segmentation task. The input layer on the left represents the input image, and the output layer on the right represents the segmented output, where the port containers are identified and delineated.

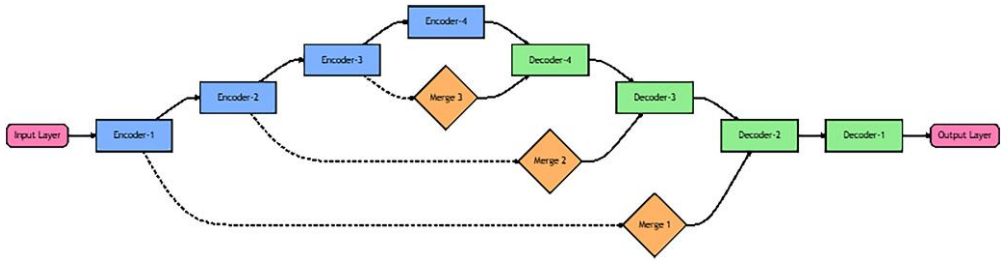

Figure A1. U-Net architecture. The U-Net architecture depicted in the image is a convolutional neural network commonly used for image segmentation tasks. It consists of an encoder–decoder structure with skip connections, allowing the model to learn a comprehensive and spatially aware representation of the input image for segmenting port containers.

Appendix B. Selection of variables for nowcasting

We here illustrate the selection of the variables used for nowcasting macroeconomic aggregates. The selection of indicators employed for nowcasting reflects a careful consideration of how different aspects of economic activity can affect GDP movements. The GDP serves as the target variable for nowcasting, representing one of the broadest measures of economic activity. The Unemployment Rate provides crucial and frequent information about labour market conditions. Changes in employment typically lag behind overall economic activity, making it valuable for confirming economic trends. When businesses face declining demand, they often adjust employment levels, making unemployment a key indicator of economic health. Industrial Turnover in Manufacturing directly reflects the monetary value of industrial activity. Because manufacturing often leads broader economic cycles and data are available monthly, changes in industrial turnover can signal wider economic shifts before they appear in GDP figures. Industrial Production (excluding Construction) captures the volume of manufacturing, mining, and utility output. As a monthly indicator, it constitutes one of the most direct measures of current economic activity and typically accounts for a significant portion of GDP movements. Similarly, Industrial Production in Construction provides insight into investment activity and construction sector health. Construction activity often reflects both business confidence (commercial construction) and household wealth effects (residential construction), making it a useful indicator of broader economic conditions. The PMI indicator (related to Manufacturing and Services) is especially valuable for nowcasting because it is among the earliest available monthly indicator and directly surveys business conditions. It can help to predict upcoming economic changes, and its quick release makes it particularly useful for real-time assessment. The Retail Sales Volume Index reflects consumer spending patterns, a major component of GDP. Monthly retail data help track household consumption, and changes in retail sales often signal shifts in consumer confidence and purchasing power. The Consumer Confidence Indicator provides insight into future consumption patterns. Since consumer spending drives a significant portion of economic activity, understanding consumer sentiment helps predicting potential changes in spending behaviour. The Business Climate Indicator aggregates business sentiment across sectors. It captures firms' assessment of current conditions and expectations, providing early signals of changes in investment, hiring, and production decisions. Trade indicators (exports and imports) reflect external sector dynamics and international demand. Monthly trade data are useful to track net exports' contribution to GDP and provides insight into global economic conditions affecting domestic activity. Finally, New Passenger Car Registrations serve as a proxy for durable goods consumption and consumer confidence, since car purchases represent significant household investments, also reflecting credit conditions.

This combination of indicators is powerful for nowcasting because it captures different aspects and timing of economic activity. The mix of monthly and weekly frequencies allows for more timely updates to GDP estimates compared with the adoption of only quarterly figures.

Finally, it is worth to mention that, as in Cascaldi-Garcia et al. (2024), we excluded daily and weekly indicators, such as financial variables, from our analysis. According to Bańbura et al. (2013), such data do not enhance the performance of nowcasting models, whether in normal periods or during recessions, due to their inherent noise and weak connection to the real economy, aside from their low-frequency nature. Additionally, we opted not to include other alternative data sources, such as web searches, electronic transactions, or textual analyses. The effectiveness of these alternative data sources in improving timely estimates of economic activity remains a topic of ongoing debate (see e.g. Larson and Sinclair, 2022).

Appendix C. ROC curves

An ROC curve is a graphical representation of the performance of a binary classification model. It visualizes the trade-off between the true positive rate (TPR) and the false positive rate (FPR) of a model at various classification thresholds. The plot is obtained by placing the TPR on the *y*-axis and the FPR on the *x*-axis, with each point on the curve corresponding to a specific threshold value. Given that the probability distributions for both true positive and false positive are known, the ROC curve is obtained as the cumulative distribution function of the detection probability in the *y*-axis versus the CDF of the false positive probability on the *x*-axis. The ideal model would have a curve that goes straight up the *y*-axis (TPR = 1) and then across the top (FPR = 0), forming a perfect right-angled triangle, indicating a perfect classifier with no false positives or false negatives. On the contrary, a model with random guessing would have a diagonal line from the bottom left (0,0) to the top right (1,1) of the plot. The diagonal divides the ROC space. Points above the diagonal represent good classification results (better than random); points below the line represent bad results (worse than random).

The AUROC curve is a widely used metric to evaluate the overall performance of a binary classification model, ranging from 0 to 1, with 1 indicating a perfect classifier and 0.5 indicating a random classifier. It represents the probability that a randomly selected positive instance will be ranked higher than a randomly selected negative instance by the classification model. In other words, the AUROC can be interpreted as the likelihood that the model will correctly identify a positive instance as having a higher probability of being positive compared with a negative instance. Figure C1 contains a graphical visualization of the ROC space for different classifiers performance.

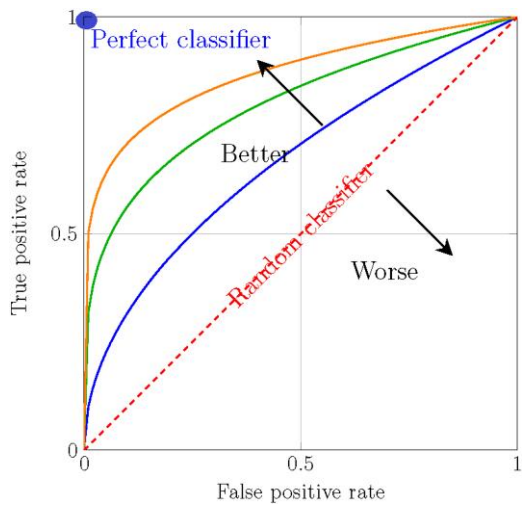

Figure C1. Receiver operating characteristic (ROC) curve. The figure illustrates the basic functioning of the ROC space comparing the performance of different classifiers.

Appendix D. Additional evidence of PSI robustness

D.1 Nonparametric analysis of unemployment rate and industrial production turnover

In this section, we conduct an additional analysis which focuses on predicting both a rise in the unemployment rate and a drawdown in the industrial production turnover. We select these variables as the unemployment rate reflects labour market health, consumer spending, and inflationary pressures, influencing economic stability and growth; on the other hand, the industrial production turnover indicates the level of economic activity, business cycle trends, and supply-side conditions, directly affecting GDP and inflation. Together, they provide a clear view of the status of both demand and supply in the economy. Figure D1 displays two ROC curves that assess the performance of a predictive model using PSI (as a predictive system indicator) to forecast unemployment upturns and industrial production turnover downturns. These variables provide critical insights into both demand (represented by unemployment) and supply (represented by industrial production) in the economy. The curves illustrate the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) across different decision thresholds. The ROC curve for unemployment upturn, with an AUROC curve of 65.2, suggests that the model demonstrates moderate predictive ability. The curve lies only slightly above the diagonal, which represents random guessing, indicating that while the model does outperform random chance, its accuracy is limited. A model with an AUROC of 65.2 indicates a fair level of discrimination, though this performance is far from ideal and suggests room for improvement, particularly in capturing shifts in labour market dynamics. In contrast, the ROC curve for industrial production downturn presents a more favourable picture, with an AUROC of 70.8. This higher value indicates that the model has a stronger capacity to distinguish between true and false positives when predicting downturns in industrial production. The curve shows a more pronounced deviation from the diagonal, implying better predictive performance relative to the unemployment upturns. From an economic perspective, the predictive model appears to perform better in capturing supply-side fluctuations, as indicated by the stronger performance in predicting industrial production downturns. On the other hand, its ability to anticipate changes in the labour market, particularly unemployment upturns, is weaker. Given the importance of both demand and supply-side factors for understanding economic cycles, improving the model's accuracy, especially concerning unemployment predictions, would be valuable. In conclusion, while the ROC curves show that the PSI-based model has some degree of predictive power for both unemployment upturns and industrial production downturns, its performance is notably better for the latter. This suggests that the model is more able to capture shifts in industrial production than in unemployment trends.

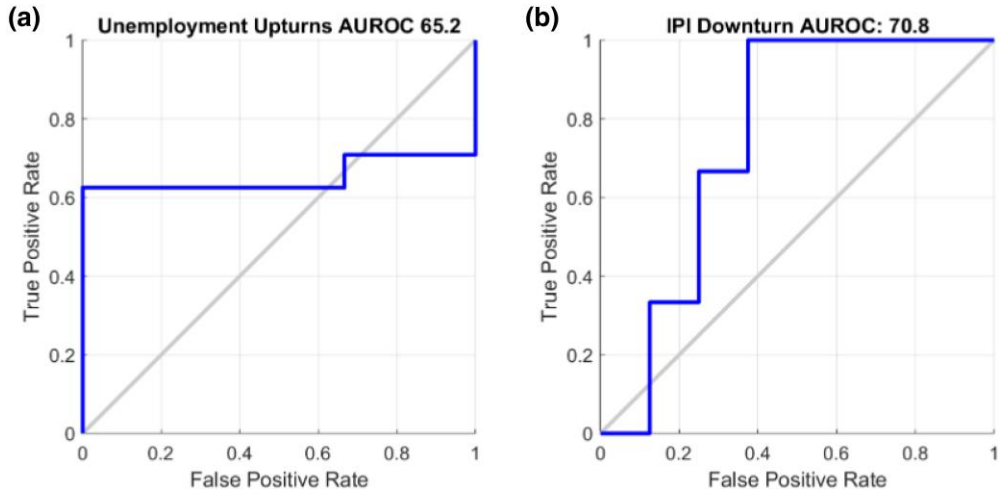

Figure D1. Receiver operating characteristic (ROC) curves of unemployment upturns and industrial production turnover downturns predictions. The figure shows the ROC curves derived from the PSI employed for predicting unemployment upturns and industrial production turnover downturns. Area under the values are reported in the figure titles.

D.2 Non parametric analysis of GDP forecast for different geographical areas

We now explore the extent to which the PSI is able to predict economic downturns and upturns at a macro-regional level, discriminating between countries belonging to Northern and Southern Europe. In particular, the northern group encompasses Belgium, Germany, Poland, Netherlands, United Kingdom. The southern group includes Spain, Greece, Italy, France, Malta. In this way

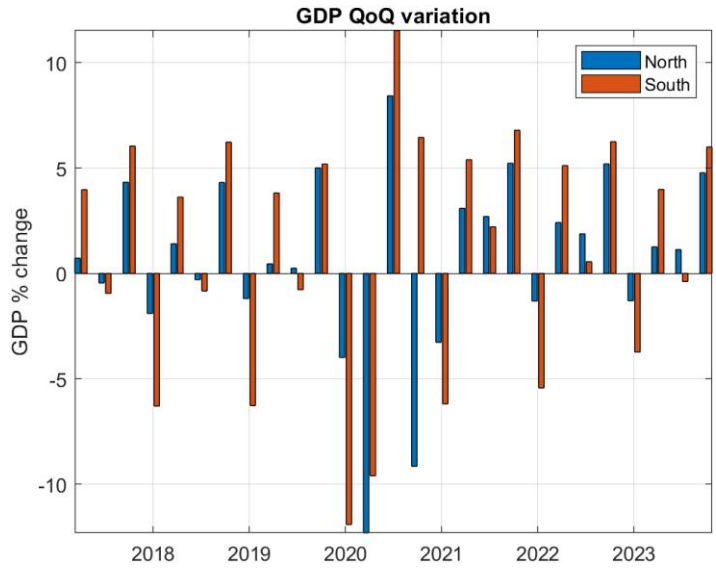

Figure D2. Gross domestic product (GDP) comparison between Northern and Southern Europe. The figure shows the quarterly percentage change (QoQ) of GDP when discriminating across Northern (Belgium, Germany, Poland, Netherlands, UK), and Southern Europe (Spain, Greece, Italy, France, Malta).

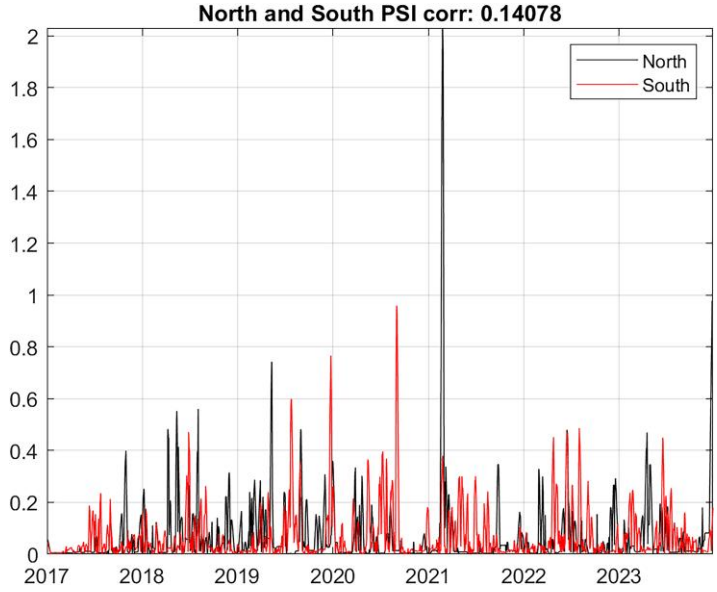

Figure D3. Port saturation index (PSI) time series for Northern and Southern Europe. The plot represents the time series dynamics of the daily PSI when discriminating between Northern (black line) and Southern (red line) European countries. The northern group includes Belgium, Germany, Poland, Netherlands, UK. The southern group includes Spain, Greece, Italy, France, Malta. The title presents the contemporaneous correlation between the two time series.

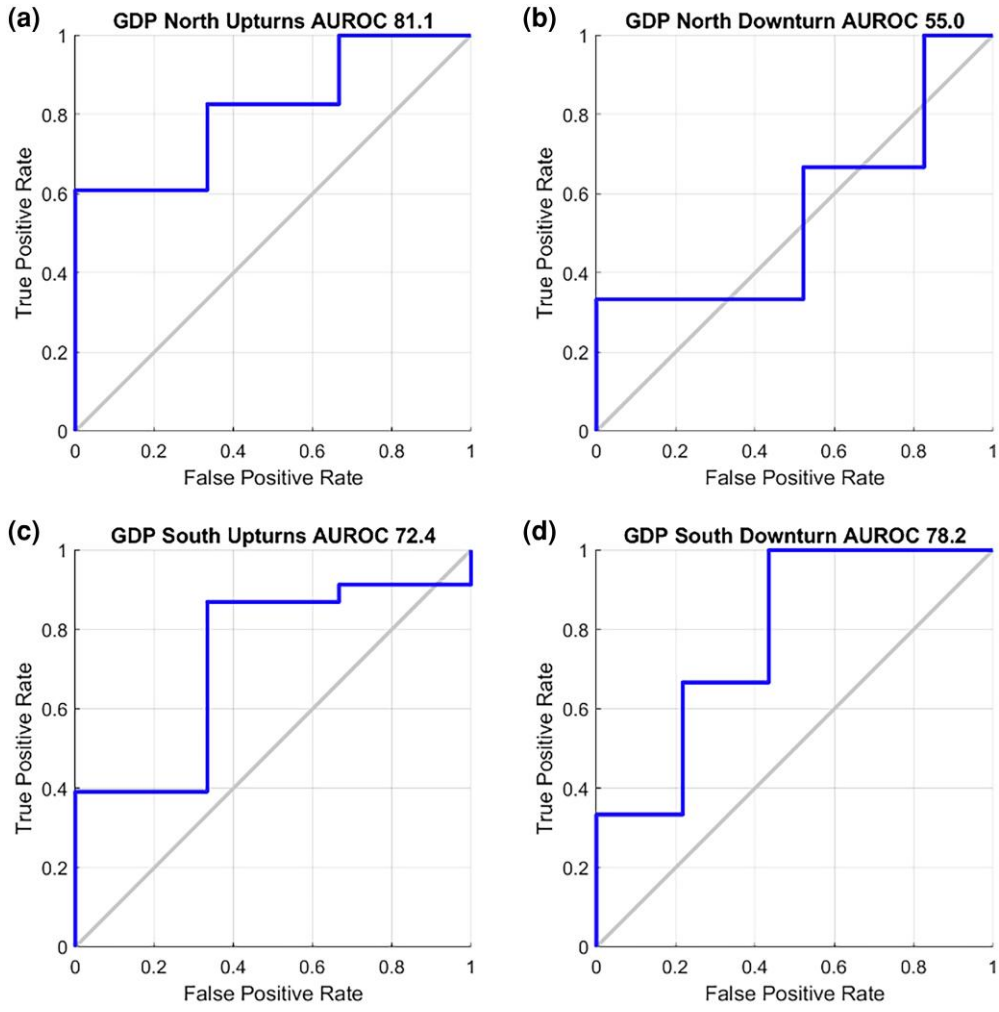

Figure D4. Receiver operating characteristic (ROC) curves of Northern and Southern Europe gross domestic product (GDP) upturns and downturns. The figure shows the ROC curves derived from the Port Saturation Index employed for predicting GDP upturns (a–c) and downturns (b–d) for Northern and Southern Europe. The northern group includes Belgium, Germany, Poland, Netherlands, UK. The southern group includes Spain, Greece, Italy, France, Malta. Area under the ROC values are reported in the figure titles.

we are able to assess the performance of our PSI when comparing macro-regions with different economic conditions. To illustrate, in Figure D2 we report the quarterly percentage change of the Northern and Southern Europe GDP. We observe a similar dynamics only during the COVID-19 period, while Southern European GDP presents a more volatile behaviour during business as usual phases. Following the same country grouping, we have constructed the PSI for both Northern and Southern Europe as shown in Figure D3. Notice that the indices are only weakly correlated.

Results on the prediction of GDP upturns and downturns are reported in Figure D4. The figure illustrates ROC curves representing the model's predictive performance for GDP upturns and downturns across Northern (Belgium, Germany, Poland, Netherlands, UK) and Southern Europe (Spain, Greece, Italy, France, Malta). Each plot's title reports the AUROC value, indicating the model's accuracy in distinguishing upturns or downturns. As far as northern countries' upturns are concerned, an AUROC of 81.1 suggests strong predictive power, as the curve rises sharply, reflecting a high true positive rate at low false positives. However, downturns of northern countries yield an AUROC of 55.0, only marginally above random, indicating weaker predictive

accuracy for this scenario. As far as southern countries are concerned, the model predicts upturns with moderate accuracy, achieving an AUROC of 72.4. Meanwhile, the model performance for southern countries' downturns is stronger, with an AUROC of 78.2, implying a reasonably effective downturn identification.

References

Aruoba S. B., Diebold F. X., & Scotti C. (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, *27*(4), 417–427. https://doi.org/10.1198/jbes.2009.07205

Ashwin J., Kalamara E., & Saiz L. (2024). Nowcasting euro area GDP with news sentiment: A tale of two crises. *Journal of Applied Econometrics*, *39*(5), 887–905. https://doi.org/10.1002/jae.v39.5

Badrinarayanan V., Kendall A., & Cipolla R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(12), 2481–2495. https://doi.org/10.1109/TPAMI.34

Baffigi A., Golinelli R., & Parigi G. (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting*, *20*(3), 447–460. https://doi.org/10.1016/S0169-2070(03)00067-0

Bai X., Fernández-Villaverde J., Li Y., & Zanetti F. (2024). *The causal effects of global supply chain disruptions on macroeconomic outcomes: Evidence and theory*. CESifo working paper.

Bańbura M., Giannone D., Modugno M., & Reichlin L. (2013). Now-casting and the real-time data flow. In G. Elliott & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 2, pp. 195–237). Elsevier.

Bańbura M., & Modugno M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, *29*(1), 133–160. https://doi.org/10.1002/jae.2306

Barbaglia L., Consoli S., & Manzan S. (2023). Forecasting with economic news. *Journal of Business & Economic Statistics*, *41*(3), 708–719. https://doi.org/10.1080/07350015.2022.2060988

Beucher S., & Meyer F. (1992). The watershed transformation applied to image segmentation. *Scanning Microscopy*, *1992*(Suppl 6), 299–314.

Bok B., Caratelli D., Giannone D., Sbordone A. M., & Tambalotti A. (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics*, *10*(1), 615–643. https://doi.org/10.1146/economics.2018.10.issue-1

Çakmakli C., & Demircan H. (2023). Using survey information for improving the density nowcasting of US GDP. *Journal of Business & Economic Statistics*, *41*(3), 667–682. https://doi.org/10.1080/07350015.2022.2058000

Canton H. (2021). United nations conference on trade and development-unctad. In Z. Bakaki, S. P. Lockhart, J. Money, K. O'Neill, P. Stoeva & J. Yao (Eds.), *The Europa directory of international organizations 2021* (pp. 172–176). Routledge.

Carriero A., Clark T. E., & Marcellino M. (2015). Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *178*(4), 837–862. https://doi.org/10.1111/rssa.12092

Cascaldi-Garcia D., Ferreira T. R., Giannone D., & Modugno M. (2024). Back to the present: Learning about the euro area through a now-casting model. *International Journal of Forecasting*, *40*(2), 661–686. https://doi.org/10.1016/j.ijforecast.2023.04.005

Chen L.-C., Papandreou G., Kokkinos I., Murphy K., & Yuille A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Cimadomo J., Giannone D., Lenza M., Monti F., & Sokol A. (2022). Nowcasting with large Bayesian vector autoregressions. *Journal of Econometrics*, *231*(2), 500–519. https://doi.org/10.1016/j.jeconom.2021.04.012

Ferrara L., Mogliani M., & Sahuc J.-G. (2022). High-frequency monitoring of growth at risk. *International Journal of Forecasting*, *38*(2), 582–595. https://doi.org/10.1016/j.ijforecast.2021.06.010

Ferrara L., & Simoni A. (2023). When are google data useful to nowcast GDP? An approach via preselection and shrinkage. *Journal of Business & Economic Statistics*, *41*(4), 1188–1202. https://doi.org/10.1080/07350015.2022.2116025

Foroni C., Marcellino M., & Schumacher C. (2015). Unrestricted mixed data sampling (midas): Midas regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *178*(1), 57–82. https://doi.org/10.1111/rssa.12043

Furukawa K., & Hisano R. (2022). A nowcasting model of exports using maritime big data (Technical Report no. 22-E-19). Bank of Japan.

Giannone D., Reichlin L., & Small D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676. https://doi.org/10.1016/j.jmoneco.2008.05.010

Hansen P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, *23*(4), 365–380. https://doi.org/10.1198/073500105000000063

He K., Gkioxari G., Dollár P., & Girshick R. (2017). Mask r-cnn. In *Proceedings of the IEEE International Conference on computer vision* (pp. 2961–2969). Computer Vision Foundation.
He K., Sun J., & Tang X. (2010). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(12), 2341–2353. https://doi.org/10.1109/TPAMI.2010.168
Hepenstrick C., & Marcellino M. (2019). Forecasting gross domestic product growth with large unbalanced data sets: The mixed frequency three-pass regression filter. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *182*(1), 69–99. https://doi.org/10.1111/rssa.12363
Horn B. K., & Woodham R. J. (1979). Destriping landsat MSS images by histogram modification. *Computer Graphics and Image Processing*, *10*(1), 69–83. https://doi.org/10.1016/0146-664X(79)90035-2
Koop G., McIntyre S., & Mitchell J. (2020). UK regional nowcasting using a mixed frequency vector auto-regressive model with entropic tilting. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *183*(1), 91–119. https://doi.org/10.1111/rssa.12491
Koop G., McIntyre S., Mitchell J., Poon A., & Wu P. (2024). Incorporating short data into large mixed-frequency vector autoregressions for regional nowcasting. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *187*(2), 477–495. https://doi.org/10.1093/jrsssa/qnad130
Krizhevsky A., Sutskever I., & Hinton G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*, 1–12.
Kuzin V., Marcellino M., & Schumacher C. (2011). Midas vs. mixed-frequency var: Nowcasting GDP in the euro area. *International Journal of Forecasting*, *27*(2), 529–542. https://doi.org/10.1016/j.ijforecast.2010.02.006
Larson W. D., & Sinclair T. M. (2022). Nowcasting unemployment insurance claims in the time of Covid-19. *International Journal of Forecasting*, *38*(2), 635–647. https://doi.org/10.1016/j.ijforecast.2021.01.001
Long J., Shelhamer E., & Darrell T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on computer vision and pattern recognition* (pp. 3431–3440). Computer Vision Foundation.
Manzan S. (2015). Forecasting the distribution of economic variables in a data-rich environment. *Journal of Business & Economic Statistics*, *33*(1), 144–164. https://doi.org/10.1080/07350015.2014.937436
Marcellino M., & Schumacher C. (2010). Factor midas for nowcasting and forecasting with ragged-edge data: A model comparison for german GDP. *Oxford Bulletin of Economics and Statistics*, *72*(4), 518–550. https://doi.org/10.1111/obes.2010.72.issue-4
Mariano R. S., & Murasawa Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, *18*(4), 427–443. https://doi.org/10.1002/jae.v18:4
Medeiros M. C., Vasconcelos G. F., Veiga Á., & Zilberman E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, *39*(1), 98–119. https://doi.org/10.1080/07350015.2019.1637745
Modugno M. (2013). Now-casting inflation using high frequency data. *International Journal of Forecasting*, *29*(4), 664–675. https://doi.org/10.1016/j.ijforecast.2012.12.003
Mourougane A., Guidetti E., & Pilgrim G. (2024). *An ocean of data: The potential of data on vessel traffic*. OECD. https://one.oecd.org/document/SDD/DOC20242/en/pdf
Notteboom T., Pallis A., & Rodrigue J.-P. (2022). *Port economics, management and policy*. Routledge.
Parigi G., & Schlitzer G. (1995). Quarterly forecasts of the Italian business cycle by means of monthly economic indicators. *Journal of Forecasting*, *14*(2), 117–141. https://doi.org/10.1002/for.v14:2
Rakwatin P., Takeuchi W., & Yasuoka Y. (2007). Stripe noise reduction in modis data by combining histogram matching with facet filter. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(6), 1844–1856. https://doi.org/10.1109/TGRS.2007.895841
Ronneberger O., Fischer P., & Brox T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18* (pp. 234–241). Springer.
Rünstler G., & Sédillot F. (2003). Short-term estimates of euro area real GDP by means of monthly data. (Technical Report). ECB working paper no. 276.
Schiavoni C., Palm F., & Smeekes S. (2021). A dynamic factor model approach to incorporate big data in state space models for official statistics. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *184*(1), 324–353. https://doi.org/10.1111/rssa.12626
Spelta A., Flori A., Pecora N., Buldyrev S., & Pammolli F. (2020). A behavioral approach to instability pathways in financial markets. *Nature Communications*, *11*(1), 1707. https://doi.org/10.1038/s41467-020-15356-z
Spelta A., Flori A., Pecora N., & Pammolli F. (2021). Financial crises: Uncovering self-organized patterns and predicting stock markets instability. *Journal of Business Research*, *129*, 736–756. https://doi.org/10.1016/j.jbusres.2019.10.043
Stock J. H., & Watson M. W. (2017). Twenty years of time series econometrics in ten pictures. *Journal of Economic Perspectives*, *31*(2), 59–86. https://doi.org/10.1257/jep.31.2.59

Tiozzo Pezzoli L., & Tosetti E. (2022). Seismonomics: Listening to the heartbeat of the economy. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(Supplement_2), S288–S309. https://doi.org/10.1111/rssa.12912

Yu H., Hao X., Wu L., Zhao Y., & Wang Y. (2023). Eye in outer space: Satellite imageries of container ports can predict world stock returns. *Humanities and Social Sciences Communications*, *10*(1), 1–16. https://doi.org/10.1057/s41599-022-01483-z

Zhao H., Shi J., Qi X., Wang X., & Jia J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE Conference on computer vision and pattern recognition* (pp. 2881–2890). Computer Vision Foundation. https://openaccess.thecvf.com/content_cvpr_2017/html/Zhao_Pyramid_Scene_Parsing_CVPR_2017_paper.html