ARTICLE

Check for updates

Eye in outer space: satellite imageries of container ports can predict world stock returns

Honghai Yu[1], Xianfeng Hao[1], Liangyu Wu[1], Yuqi Zhao[2] & Yudong Wang [3 ✉]

Forecasting stock returns is challenging. Traditional economic data that are available to all investors are published with lags and suffer from the problem of frequent revisions. Consequently, they often fail to forecast stock returns. For this reason, investors are increasingly interested in seeking alternative data. This paper forecasts stock returns using satellite-based information on shipping containers, which can capture economic activity in real-time. The container coverage area in each port is identified from 83,672 satellite images via the U-Net method and used as a proxy for the number of containers. Forecast combination over univariate predictive regression is used to generate return forecasts. The results indicate that the number of containers in ports can significantly predict stock index returns in 27 out of 33 countries at a daily frequency for the 2019–2021 period. An investor making use of satellite data on marine ports can, on average, receive an annualized return of 16.38%. The predictability can be explained by the predictive relationship between port container numbers and economic activity. In future studies, satellite data can be applied to monitor and forecast other economic indicators.

[1] School of Management and Engineering, Nanjing University, Nanjing, China. [2] School of Economics and Management, Tongji University, Shanghai, China. [3] School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China. ✉email: wangyudongnj@126.com

Introduction

Economic theory holds that asset returns are functions of the state variables of the real economy and that the real economy displays business-cycle fluctuations. If the quantity and price of aggregate risk are linked to economic fluctuations, one can expect return predictability to exist (Campbell and Shiller, 1988; Fama and French, 1989; Campbell and Cochrane, 1999). However, in practice, forecasting stock returns remains notoriously difficult, although many economic variables have been developed for this purpose in the literature. The predictors examined include dividend–price ratio, dividend yield, earnings–price ratio, dividend–payout ratio (Campbell and Shiller, 1988), stock volatility (Guo, 2006), book-to-market ratio (Kothari and Shanken, 1997; Pontiff and Schall, 1998), T-Bill rate (Campbell, 1987), term spread (Fama and French, 1989), and inflation rate (Fama and Schwert, 1977). Welch and Goyal (2008) systematically examine the forecasting performance of popular economic indicators in published papers and find no evidence that any of them can significantly beat the no-predictability model (historical average). Goyal et al. (2021) further confirm this finding based on the reexamination of the forecasting performance using new predictors developed in recent literature.

Apart from challenging modeling issues, there are three reasons for the failure of macroeconomic data to forecast stock returns. First, macroeconomic indicators such as the consumer price index (CPI) and gross domestic production (GDP) are always published with delays, even if the government data are credible. Therefore, one has to execute real-time predictions of future returns using lagged macroeconomic data instead of current-period data. The use of lagged data undermines forecasting performance. Second, many macroeconomic indicators undergo revisions after initial publication. Using unrevised data in real-time biases the forecasting results. Third, traditional economic data are publicly available, and there is virtually no cost apart from basic data processing equipment to acquire the information. When this new information enters the market, it is instantaneously integrated into the price if the number of investors trading on it is sufficient (Jensen, 1978). As a result, the predictive value of announcement-based economic data is extraordinarily short-lived due to its public accessibility. In addition, macroeconomic data are published at monthly or quarterly frequencies, making it difficult to forecast stock returns at daily or weekly frequencies. Therefore, frequent data that contain economic information that is not easily available through simple searches in real-time can be expected to improve the forecasting of stock returns.

Recent technological advances in artificial intelligence and big data have revolutionized information collection. Hedge funds have turned to satellite technology to gain a real-time information advantage in understanding economic activity. They use satellite imagery to obtain information on mines, ports, plantations, or farmland before making investments. Several recently established companies provide satellite-based forecasts of economic indicators. For example, Orbital Insight, Planet Labs, Spire Global, and Space Know track industrial facilities, real estate properties, foot traffic activity, oil refineries, petrochemical plants, and auto manufacturing centers to generate information about commercial properties. In 2016, 70 of the 74 clients of Orbital Insight, the biggest geospatial analytics company, were hedge funds (see https://orbitalinsight.com). Institutional investors increasingly attempt to glean investment insights from such imagery.

We use information from satellite images of container ports to predict global stock market returns. Economic globalization depends on the rapid and efficient movement of goods via containerization. Globally, 90% of non-bulk dry cargo is now shipped by container. The number of containers at a port can be regarded as an indicator of macroeconomic information. The rationale is that an increase in the number of containers stacked in the port implies decreased demand for shipping service, and thus predicts lower economic activity. Stock prices are expected to fall accordingly. We extract real-time information from the Sentinel-2 images for the top 48 container ports (as ranked by throughput) from the European Space Agency (ESA).

Our database contains 83,672 temporally and spatially matched multispectral daytime 10 m/pixel images from Sentinel-2 satellites. The database covers the period from January 1, 2017, to November 1, 2021. To segment the container areas, we train the U-Net model (Ronneberger et al., 2015) with 3711 hand-labeled images in 2017. In this way, we can obtain a series of container coverage areas at a daily frequency for each of the 48 ports under consideration. The container area series are used to predict daily stock returns on 33 stock indices in major countries. Our results indicate that the combined container area information reveals significant return predictability for most of the 33 markets for the period from 2019 to 2021. Investment strategies based on the container information generate an economically considerable profit, with an annualized return of 16.38% and a Sharpe ratio of 1.19. Extended analysis shows the close links between the predictive power of satellite-based data and the macroeconomy.

The ability of our satellite-based container information to predict stock returns can be explained by its ability to anticipate economic activity. Global marine trade links production activity and the consumption of goods and is thus highly informative about economic activity. We investigate the predictive ability of the number of containers in relation to the growth of industrial production. Our results suggest a negative predictive relationship in 27 out of 28 countries at the horizon of four months, and this predictive ability is statistically significant in 15 cases. The significantly negative predictive relationship is also found when regressing the world average growth rate of industrial production on the past change in container numbers. The number of containers has a greater ability to predict industrial production during the COVID-19 period, echoing the stronger return predictability after the pandemic.

We compare the information content of our satellite-based container number data with popular shipping data, such as the freight rates indicator (Kilian, 2009) and container throughput (Döhrn and Maatsch, 2012; Döhrn, 2019; Kilian et al., 2021). We find that the global number of containers significantly predicts these two shipping indicators. Our data lead the traditional indicators for the horizons of 2 months. The strong relationship between our container indicator and existing indicators is not affected by the COVID-19 pandemic.

The advantages of satellite imagery as a data source for economic studies have been documented in the literature. Several studies use satellite image data to measure economic variables including GDP growth (Henderson et al., 2012), economic inequity (Chen and Nordhaus, 2011), income distribution (Mirza et al., 2021), sustainable development (Burke et al., 2021), and rural household poverty (Jean et al., 2016; Watmough et al., 2019). Most of these studies use satellite data on a night light, which makes it possible to compare economic activity across different areas. For a single area, night light data show minor variance over time, and thus they are not appropriate for time series prediction analysis. Recent studies reveal that the satellite imagery from Orbital Insight of parking lots can anticipate retailer sales performance that is not yet announced and mute price reactions to earnings announcements (Katona et al., 2018; Zhu, 2019). Unlike these studies, which rely on ready-made forecasts from commercial satellite companies, we build our own database from public satellite data. This is an important
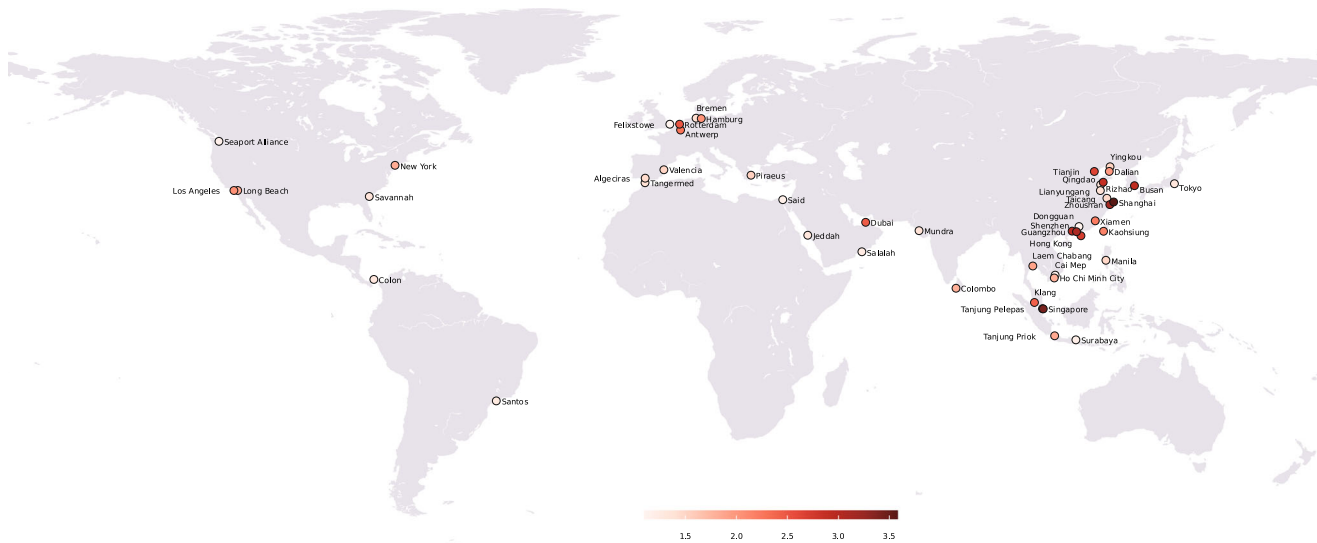
Fig. 1 Global distribution of container ports. Location of top 48 container ports ranked by throughput (TEU) data in 2020 obtained from the Institute for Shipping Economics and Logistics (ISL). TEU stands for the 20-foot equivalent container. The darkness of each point represents its throughput (units: 10 million TEUs), where the port with the higher throughput is darker.

difference because the high cost of commercial satellite data makes them inaccessible to many retail investors. The return predictability based on public satellite data is useful to many more market participants and has stronger economic implications.

The application of satellite data is also found in a few finance literature. Katona et al. (2018) use parking lot traffic signals extracted from satellite imagery and find that unequal access to satellite data increases information asymmetry among market participants. Zhu (2019) finds that satellite-based estimates of normalized car counts in parking lots of retailers predict earnings that are not yet announced. Mukherjee et al. (2021) show that after the introduction of satellite-based imagery data, oil price responses to government announcements of oil inventory are smoother. As a contribution, we directly show that the satellite-based estimates of the number of containers predict world stock returns, enriching the literature on the application of satellite data in empirical asset pricing studies.

The remainder of this paper is organized as follows: Section "Data" shows the details of satellite data processing and the stock returns data. Section "Forecasting results" reports the forecasting results. The section "Understanding the source of return predictability" gives some explanations about the source of return predictability. Section "Discussion" performs discussions on the application of satellite data. The last section concludes the paper.

Data
The identification of containers in ports. We collect publicly available and freely distributable satellite imagery from the Sentinel-2 mission. The dataset consists of 83,672 RGB images of 48 major ports from January 1, 2017, to November 1, 2021. Figure 1 provides the global distribution of those container ports. The details about satellite imagery processing and model training are given in Appendix.

The identification of containers in the ports can be treated as a semantic binary segmentation task, which is an increasingly popular domain in computer vision. A semantic binary segmentation task takes an image as input and outputs binary classification results for each pixel in the image. In this task, the two classes are "container" and "non-container." Note that containers are usually stacked in several layers to save floor space. However, the number of layers cannot be recognized precisely from the Sentinel-2 satellite images due to the limitation of

resolution. We arbitrarily assume that different container stacks have the same number of layers. In this way, we count the number of pixels in each satellite image that are classified as "container" and take this as the proxy of the number of containers in the port. Changes in the number of containers can reflect the dynamics of global economic activities.

Specifically, we use U-Net (Ronneberger et al., 2015), a conventional deep-learning model for semantic segmentation tasks, to identify containers from satellite images. As a variant of convolutional neural networks (CNN), U-Net uses a unique U-shaped architecture and skip connections to capture multi-scale contour information. U-Net has been shown to be efficient in cell segmentation tasks (Ronneberger et al., 2015). After multiple iterations and improvements (Dolz et al., 2018; Zhou et al., 2020), it shows excellent performance in medical imagery semantic segmentation tasks like CT pancreas segmentation (Oktay et al., 2018) and cancer detection (Huang et al., 2021). In recent years, U-Net has become a major image segmentation method in various research areas, especially satellite image segmentation. For example, researchers have used U-Net to locate photovoltaic solar energy-generating units from space (Kruitwagen et al., 2021) and to forecast seasonal arctic sea ice (Andersson et al., 2021).

We construct a unique training set for our U-Net model. Traditional satellite or aerial data set image semantic segmentation tasks rely on precise standard datasets. For example, the Massachusetts roads data set provides images covering more than 2600 square kilometers and precisely labels the shape of every road in each image. However, because of the lack of research into container identification in satellite images, a standard dataset for container recognition is not yet available. Therefore, it is necessary for us to reconstruct a dataset oriented toward our task. Specifically, we pick out all Sentinel-2 satellite images in 2017, the earliest year in our dataset. After abandoning images with 5% or greater cloud coverage, we label the remaining 3711 images by hand. We identify "container" or "non-container" areas of each image, and use this as our dataset for U-Net model training. As we use data from 2017 to train the model, and then identify images and predict stock returns after that period, this procedure avoids forward-looking bias due to the application of future information.

Table 1 Experiment results of the U-Net model.

Input size	Number of convolution layers	Accuracy	Recall	*F*-score
120*120	8	0.717	0.809	0.760
	13	0.713	0.828	0.766
	18	0.714	0.821	0.763
	23	0.718	0.871	0.787
240*240	8	0.785	0.847	0.814
	13	0.791	0.879	0.832
	18	0.787	0.860	0.821
	23	0.780	0.840	0.808
480*480	8	0.930	0.958	0.944
	13	0.952	0.956	0.954
	18	**0.959**	0.948	0.953
	23	0.955	**0.961**	**0.958**
600*600	8	0.848	0.863	0.855
	13	0.841	0.841	0.840
	18	0.839	0.859	0.848
	23	0.831	0.883	0.855

This table compares the model performance under different groups of hyperparameters, including the input size of images and the number of convolution layers. The accuracy, recall, and *F*-score are evaluated through 10-fold cross-validation.
Bold values represents the maximum value achieved under the corresponding criterion.

We train different U-Net models to find the one that best fits our task by selecting two hyperparameters: the size of input images and the depth of the network. First, given the convolution kernel size (normally 3 × 3), the input size of images influences the perception ability of the convolution kernel. Second, the depth of the network, which is usually referred to as the number of convolution layers in U-Net, determines the level of contour information that can be used in the model. A deeper layer generally corresponds to a higher level of information. Table 1 shows the analysis results of those two hyperparameters. We find that the highest identifying accuracy is achieved when inputting medium-size images (480 × 480 pixels) to a deeper network structure (23 convolution layers). The preference for deep layers echoes the simplicity of the container stack shape because a deeper network leads to a better capacity for abstraction, which works best when the object contour is simple. Finally, our selected model for identification achieves 93.20% accuracy, 92.45% recall, and 92.81% *F*-score in the testing set, demonstrating good performance.

We then measure the container coverage areas, which are our proxy for container numbers, from each image based on our training model. It is essential to evaluate if our model performs consistently for each image. Thus, we visualize and analyze the global spatial distribution change in the number of containers over time for a better understanding of model performance in practice. A simple test is to see whether the identification results for the spatial distribution of containers in 2017 look similar to the true spatial distribution in that year. If so, we can conclude that the model performs consistently well. We do another test using the exogenous shock caused by COVID-19. Ports in the United States faced severe congestion due to COVID-19 (Meeks et al., 2021), as a lack of truck drivers and other laborers caused a huge number of containers to pile up. If our model performs well, it will capture a significant growth in the number of containers in 2021.

Figure 2 shows the identification results of our U-Net model for the Port of Los Angeles, the largest port in the United States. The first column introduces the Sentinel-2 satellite image, and the yellow areas indicate seven subregions of the container port. The subgraphs in the second column show the container distribution in the training set in 2017, in which the color of each pixel indicates the percentage of pixels classified as "container"; a brighter color represents a higher percentage. The third to last columns show the recognition results for our model from 2017 to 2021. The results of the two tests can be found in Fig. 2. First, the similarity of the second and third columns reveals the consistency of the training labels and recognition results. Second, the abnormal growth in the number of containers in 2021, as shown in the last column, is consistent with the port congestion in 2021 (Meeks et al., 2021). Both pieces of evidence demonstrate the stability of our model.

To predict the stock market return, we calculate the daily average change in the number of containers in the ports (GNC), which is defined as $\text{GNC}_{i,t} = \frac{\log(\text{NC}_{i,t}) - \log(\text{NC}_{i,s})}{t-s}$, where $\text{NC}_{i,t}$ represents the number of containers in port i at time t, and s represents the most recent date before time t for which cloud-free (clear) satellite imagery is available. Dividing the log difference by $t-s$ is used to standardize the daily change in the number of containers, which eliminates the influence of the uneven distribution of observations over time. Intuitively, an increase in GNC indicates higher port congestion and lower trading volume, which heralds future economic downturn and predicts lower stock returns.

Stock returns data. Our dataset of international stock returns is related to 33 market-level indices from 28 countries, including 18 developed markets and 12 emerging markets. We collect the daily stock returns from January 2017 through November 2021 from the *Wind Database*. Table 2 presents the details of the stock index under consideration.

We recursively generate daily stock returns from January 2019 through November 2021, and therefore the data from January 2017 through December 2018 are used as the initial estimation sample. The selection of this initial estimation sample involves a trade-off. On the one hand, we require more initial data to get a more reliable regression estimate in the process of computing the first return forecast. On the other hand, we also require a longer out-of-sample evaluation period to obtain a more accurate evaluation result. As a compromise, we select 40% of the data as the initial estimation sample. In this way, our out-of-sample period covers January 2019 through November 2021, which spans periods both before and after the COVID-19 pandemic. To check the robustness of our results to the sample selection, we re-examine the forecasting performance using an alternative initial estimation sample from January 2017 to December 2017. The results are shown in the online appendix.

Stock indices trade in different time zones. Due to such consideration, for each stock index, we convert the Sentinel-2 image UTC time to the local time to match the data. We use the satellite imagery available within 24 h before closing time to predict the close-to-close return on the next business day, which avoids the introduction of forward-looking information.

Competing shipping indicators. We investigate the linkages between our container number index and existing shipping indicators including the container throughput index and baltic dry index (BDI). Although these indexes are published with lags, they can well capture global economic activity (Kilian, 2009). Figure 3 plots the global number of containers (GNC) index, BDI, and the container throughput index (RWI_ISL). We can observe that the GNC index changes inversely with competing shipping indicators.

Forecasting results
Forecasting procedure. Following the majority of papers on return predictability, we assume a linear relationship between the
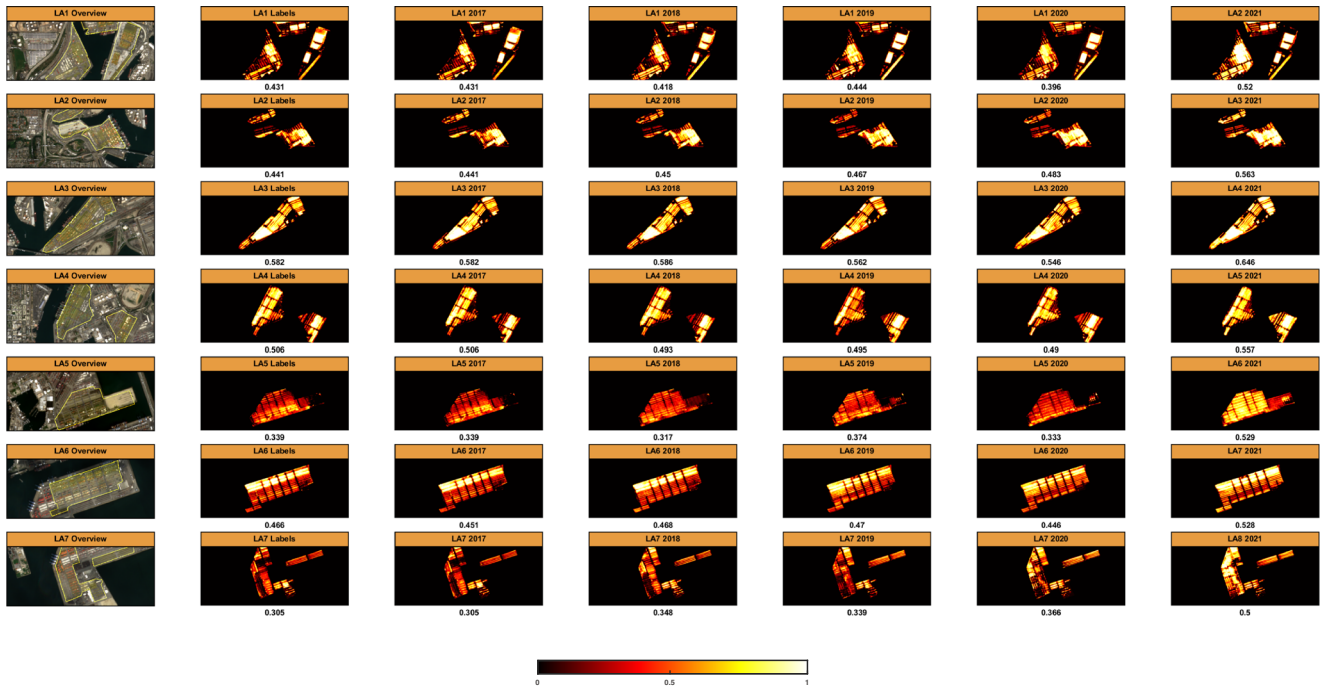
Fig. 2 Container recognition results for Port of Los Angeles. The first column shows satellite images at 7 different parts of Port of Los Angeles (LA) and the yellow areas indicate the container yards. The second column is the ground truth of containers' spatial distributions in 2017, labeled by hand. The brighter color of pixels means a higher container frequency in one specific year, and the annual average frequency is marked at the bottom of each diagram. Container recognition results by U-Net are shown from the third column to the last column, representing results from 2017 to 2021, respectively.

GNC and the stock index return (Goyal and Welch, 2003; Welch and Goyal, 2008). The specification of predictive regressions for stock returns on the lagged predictor variable of interest can be written as follows:

$$r_{t+h} = X_t \cdot \beta + \varepsilon_{t+h}, \ t = 1, ..., T - h \tag{1}$$

where r_{t+h} represents the average daily continuously compounded stock returns in excess of the risk-free rate from time t to $t+h$; X_t represents a vector that consists of a return predictor and an intercept; and ε_{t+1} is the error term, $\varepsilon_t \sim$ i.i.d. $(0, \sigma_\varepsilon^2)$. The parameter estimates of the predictive regression can be simply obtained via ordinary least squares (OLS). As investors do in practice, we execute one-step-ahead forecasting to generate out-of-sample return predictions. Specifically, the return forecasts from univariate predictive regression can be written as follows:

$$\widehat{r}_{t+h} = X_t \widehat{\beta}_t, t = M, ..., T - h \tag{2}$$

where $\hat{\beta}_t$ is the parameter estimates of β using the information until time t. At time t, we regress $\{r_\tau\}_{h+1}^t$ on the $\{x_\tau\}_1^{t-h}$ to obtain parameter $\hat{\beta}_t$ via the OLS method. The parameter estimates are updated at each point in time for $t \geq M$ using extending windows, where M denotes the initial sample length to execute parameter estimation.

Next, we integrate multiple informations from the global ports by pooling the individual return forecasts. It has been shown that combining forecasts is an effective method for extracting information from high-dimensional predictors in economic forecasting (Timmermann, 2006). Forecast combinations use the weighted average of forecasts from individual models, as given by

$$\widehat{r}_{t+h,\text{comb}} = \sum_{i=1}^N \widehat{\pi}_{t+h,i} \widehat{r}_{t+h,i}, \quad t = M, ..., T - h \tag{3}$$

where $\hat{r}_{t+h,i}$ denotes the forecasts from model i, $\hat{\pi}_{t,i}$ is the ex-ante

weight assigned to model i formed at time k, and N is the number of predictive models. In this paper, we consider an equal-weighted mean combination, which uses equal weight $\hat{\pi}_{t+h,i} = 1/N$. Although this weighting scheme is simple, recent empirical and simulation studies have shown that it is not necessarily outperformed by more sophisticated combinations (Smith and Wallis, 2009; Claeskens et al., 2016). Note that this paper focuses on plain OLS forecasting techniques and a naïve combination strategy, in the interest of straightforwardly testing the predictive power of the new GNC indicator.

Statistical predictability. We examine whether satellite-based container data are helpful in predicting global stock market returns out-of-sample. The out-of-sample forecast at time $t + 1$ is made using the data available up to time t. The out-of-sample R^2 (R^2_{OoS}) is used to evaluate the forecasting performance, defined as follows:

$$R^2_{\text{OoS}} = 1 - \frac{\sum_{t=M=h}^T (r_t - \hat{r}_t)^2}{\sum_{t=M+h}^T (r_t - \bar{r}_t)^2} \tag{4}$$

where $\hat{r}_t$ is the return forecast, $\bar{r}_t$ is the prevailing mean forecast, and r_t is the realized return at time t. Therefore, R^2_{OoS} quantifies the reduction of the forecasting loss of the given model relative to the forecasting loss of the benchmark model. A positive R^2_{OoS} implies that the given model outperforms the benchmark model. We use the common benchmark of the historical average, which is typically hard to beat (Welch and Goyal, 2008). The statistic developed by Clark and West is applied to test the statistical significance of return predictability (Clark and West, 2007).

Figure 4 plots R^2_{OoS} values of the equal-weighted forecast combinations using univariate predictive regression with the change in container numbers in each of the 48 ports. The combination model that aggregates information from global ports dominates the no-predictability benchmark of the historical

Table 2 International stock index.

Abbreviation	Stock index	Country	Time zone
AEX	Amsterdam Exchanges Index	Netherlands	Europe/Amsterdam
ATG	Athens General Composite Index	Greece	Europe/Athens
ATX	Austrian Traded Index	Austria	Europe/Vienna
BFX	BEL 20 Index	Belgium	Europe/Brussels
FCHI	CAC 40 Index	France	Europe/Paris
FTMIB	FTSE MIB Index	Italy	Europe/Rome
FTSE	FTSE 100 Index	United Kingdom	Europe/London
GDAXI	DAX Performance-Index	Germany	Europe/Berlin
IBEX	IBEX 35 Index	Spain	Europe/Madrid
OMXSPI	OMX Stockholm PI	Sweden	Europe/Stockholm
OSEAX	Oslo SE All-share Index	Norway	Europe/Oslo
PSI20	Euronext Lisbon PSI Index	Portugal	Europe/Lisbon
RTS	Russian Trading System Index	Russian Federation	Europe/Moscow
SSMI	Swiss Market Index	Switzerland	Europe/Zurich
XU100	Borsa Istanbul 100 Index	Turkey	Europe/Istanbul
WIG30	Warsaw SE WIG 30 Index	Poland	Europe/Warsaw
SSEC	Shanghai SE Composite Index	China	Asia/Shanghai
BSESN	S&P BSE Sensex Index	India	Asia/Calcutta
CSI300	Shanghai Shenzhen CSI 300 Index	China	Asia/Shanghai
HSI	Hang Seng Index	China	Asia/Shanghai
KLSE	FTSE Bursa Malaysia KLCI Index	Malaysia	Asia/Kuala Lumpur
KS11	KOSPI Composite Index	South Korea	Asia/Seoul
N225	Nikkei 225 Index	Japan	Asia/Tokyo
STI	Straits Times Index	Singapore	Asia/Singapore
TASI	Tadawul All Share Index	Saudi Arabia	Asia/Riyadh
COLCAP	MSCI COLCAP Index	Colombia	America/Bogota
DJI	Dow Jones Industrial Average Index	United States	America/New York
GSPC	Standard & Poor's 500 Index	United States	America/New York
GSPTSE	S&P/TSX Composite Index	Canada	America/Toronto
IXIC	NASDAQ Composite Index	United States	America/New York
MXX	IPC MEXICO Index	Mexico	America/Mexico City
JTOPI	South Africa Top 40 Index	South Africa	Africa/Johannesburg
AORD	All Ordinaries Index	Australia	Australia/Sydney

This table reports the abbreviation and basic information of international stock indices used in this paper. Our dataset contains 33 market-level indices from 28 countries, including 18 developed markets (Australia, Austria, Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, the United Kingdom, and the United States) and 12 emerging markets (China, Colombia, Greece, India, Malaysia, Mexico, Poland, Russia, Saudi Arabia, South Africa, South Korea, and Turkey).

average at horizons of up to 5 days. Specifically, at the horizon of 1 day, we observe positive R^2_{OoS} in all markets, and 27 of these values are statistically significant at the 10% level. The average daily R^2_{OoS} of 33 markets can reach 0.0529%. This powerful predictive ability can also be seen for longer forecasting horizons. The R^2_{OoS} values are positive in all cases, with average magnitudes of about 0.05%. The long-horizon predictability suggests that the information from satellite imagery cannot be absorbed into the price immediately but is digested slowly.

We carry out statistical inference by testing whether the given model forecasts yield significant improvements over the benchmark forecasts. The asymptotic statistics suffer from the problem of small-sample bias. In addition, they may have the wrong size even when small-sample bias is considered. Due to such considerations, we follow Mark (1995) in using a bootstrap-based Diebold and Mariano (1995) statistic to examine the significance of the forecasting improvement. Specifically, we execute statistical inference using a stationary bootstrap procedure from Politis and Romano (1994) under the null hypothesis that the equity premium is unpredictable. The Diebold and Mariano (1995) (DM) statistic for testing equal predictive ability between the given model and benchmark model is given by

$$DM = \frac{\overline{f_t}}{\sqrt{\frac{\hat{S}_{f_t}}{N_{f_t}}}} \quad (5)$$

where $\overline{f_t} = \frac{1}{N_{f_t}} \sum \hat{f_t}$, $\hat{S}_{f_t} = \frac{1}{N_{f_t}} \sum \left(\hat{f_t} - \overline{f_t} \right)^2$ and $\hat{f_t} = \left(\overline{r_t} - r_t \right)^2 - \left(\hat{r_t} - r_t \right)^2$. The number of resamples is set as 2000 when bootstrapping, and the block length is optimally estimated from the data using the selection procedure of Patton et al. (2009).

As shown in Table 3, we find that the forecast improvement of the GNC model is statistically significant for most cases after accounting for size distortion. The bootstrap-based DM tests show a significantly positive R^2_{OoS} at the 10% level for 29 of 33 markets at the horizon of one day. This finding holds for longer forecasting horizons. Overall, the forecast improvement of the GNC model is consistent across different time periods and is insensitive to sample size distortion.

We further use an alternative evaluation, success ratio (SR), which measures how often the model generates forecasts with the correct sign. The criterion is given by

$$SR = \frac{1}{T - M - h + 1} \sum_{t=M+h}^{T} l_t, l_t = I\left(\left(r_t - \hat{r}_{t,model} \right)^2 < r_t^2 \right) \quad (6)$$

The SR criterion is less sensitive to the R^2_{OoS} metric. The PT test (Pesaran and Timmermann, 2009) is used to examine whether the success ratio of each model is significantly >50%. Figure 5 exhibits the corresponding results. We find that the combined information from individual ports correctly predicts the sign of
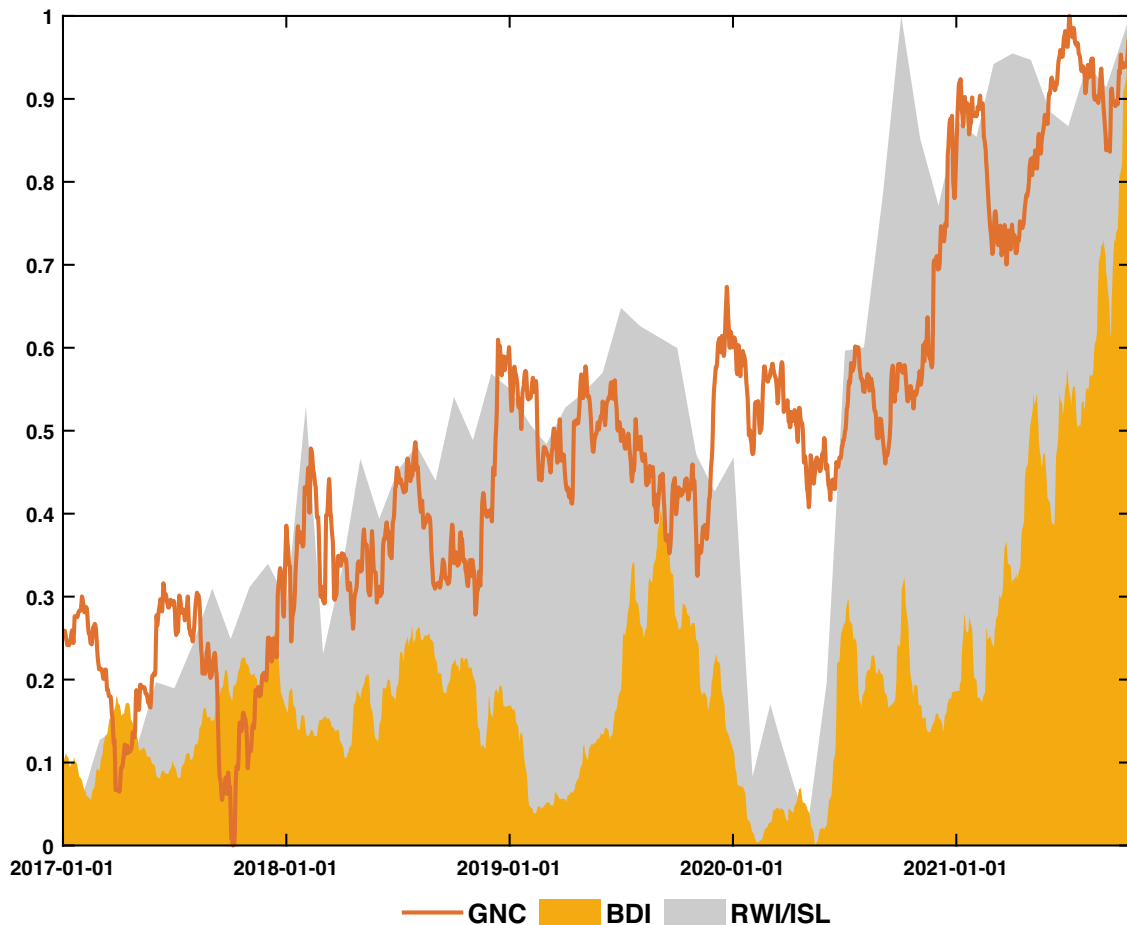
Fig. 3 The global number of containers index. GNC represents the global number of containers calculated by summing the numbers for all 48 ports. For comparison, we also plot the baltic dry index (BDI) and container throughput index collected by the Leibniz Institute for Economic Research (RWI) and the Institute for Shipping Economics and Logistics (ISL). The data are min–max normalized between 0 and 1.

the change in the market index more frequently than tossing a coin in most cases. The success ratios are higher than 0.5 in the case of 27 markets for the forecasting horizon of one day, and 23 of these values are significant. The models are also more successful when predicting the direction of change than a random-walk benchmark at longer horizons, although the directional predictability is slightly weakened.

Forecasting performance over time. We check the predictive power of container data over time. A concern is that the model cannot consistently beat the benchmark but only show predictive ability during short periods. To address this issue, we calculate the cumulative sum of squared prediction error difference (CSSED) proposed by Welch and Goyal (2008), defined as follows:

$$\text{CSSED}_{\text{model},t} = \sum_{\tau=M+h}^{t} \left(e_{\text{bench},\tau}^2 - e_{\text{model},\tau}^2 \right), \quad t = M + h, \ldots, T \tag{7}$$

where $e_{\text{bench},\tau}$ and $e_{\text{model},\tau}$ denote time τ forecast errors associated with the historical average benchmark model and the given model, respectively. The CSSED measure has become a standard indicator in the finance literature for evaluating out-of-sample return predictability (Goyal et al., 2021). The CSSED curve can illustrate whether a model of interest produces more accurate return forecasts than the benchmark model for any given evaluation period by redrawing the horizontal zero line to

the beginning of the out-of-sample period. Intuitively, when the forecast generated from the given model outperforms the forecast generated from the benchmark model at time $t + 1$, the CSSED increases from time t to $t + 1$.

Figure 6 plots the CSSEDs of the combined GNC models relative to the historical average benchmark. At the horizon of 1 day, all the curves have slopes that are predominantly positive, suggesting that the forecasts conditional on the number of containers consistently outperform the prevailing mean benchmark over time. At longer horizons, the CSSED curves are also positively sloped and have less frequent falloffs. The smoother line illustrates the higher robustness of the predictive power of the container information in forecasting long-term returns. More importantly, we find an interesting pattern in all markets, in that the CSSEDs jumped upward during the global depression caused by COVID-19, especially in the American and European markets. The evidence indicates much stronger return predictability during the COVID-19 recession.

The better performance of the container indicators during the COVID-19 period can be linked to the supply disruptions caused by friction in the container shipping market. The COVID-19 pandemic produced a historic global crisis. To prevent an outbreak, many countries went into lockdown and imposed unprecedented containment measures. These led to unexpected frictions in container shipping markets, such as port congestion, shipping delays, or shortages along the supply chain, which seriously reduced economic activity. These frictions increased suddenly in early 2020, as the pandemic spread across China and
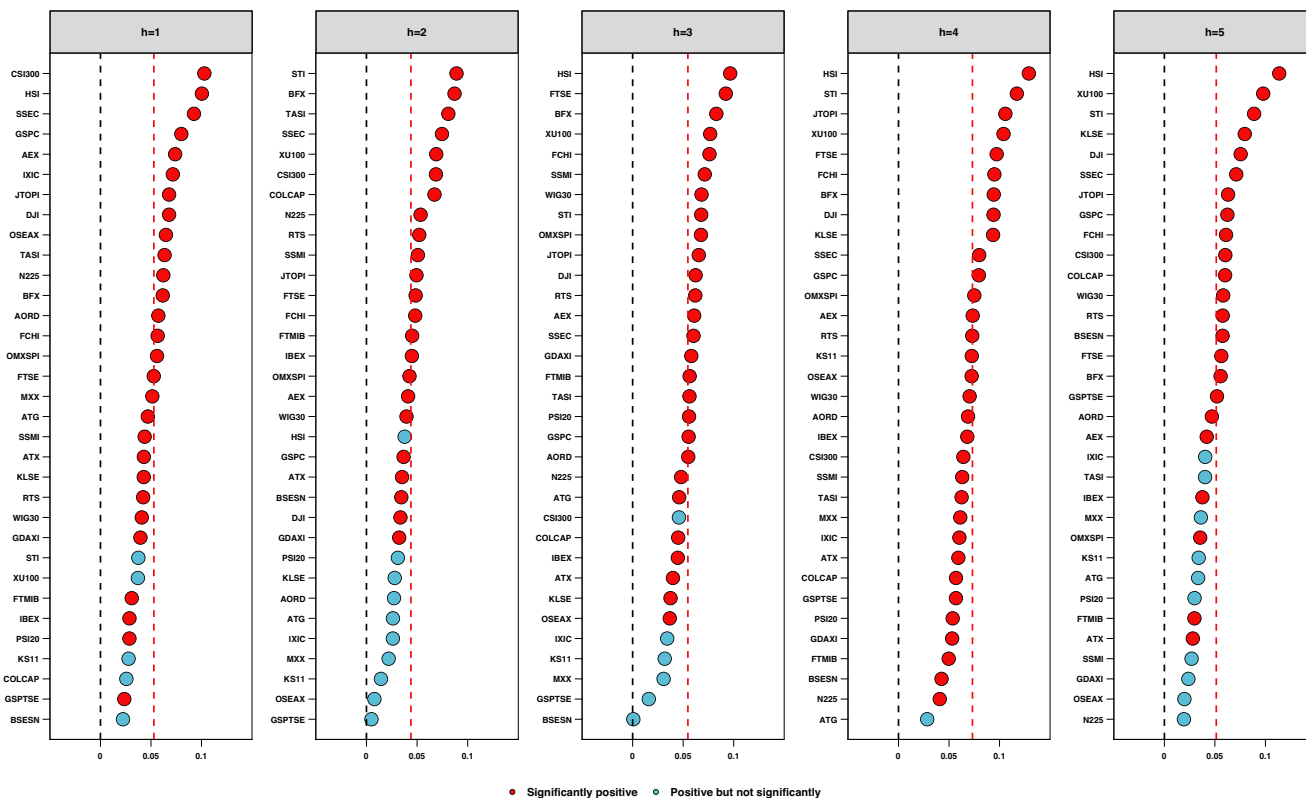
Fig. 4 Forecasting performance evaluated by out-of-sample R-square. This figure plots the scatter of the out-of-sample R-square for different stock market indexes. The forecast quality is evaluated by the out-of-sample R^2 (R^2_{OoS}) defined by the percent reduction of mean squared prediction error of the given model relative to the benchmark model of historical mean, $R^2_{OoS} = 1 - \sum_{t=M+h}^{T} (r_t - \hat{r}_t)^2 / \sum_{t=M+h}^{T} (r_t - \bar{r}_t)^2$, where $\hat{r}_t$, $\bar{r}_t$ and r_t are the return forecasts from the model of interest, prevailing mean forecasts and the realized returns, respectively. The full sample period is from 2017-01-01 through 2021-11-01 and the out-of-sample period starts on 2019-01-01 (i.e., $M = 500$). We multiply the R^2_{OoS} by 100 to denote percentage values. We measure the statistical significance relative to the prevailing mean model using the Clark and West (2007) test statistic. The cases significant at 10% level are highlighted in red.

then the rest of the world; lockdowns disrupted global supply chains and reduced the volume of container shipping.

Figure 3 exhibits a sharply increasing number of containers stacked in port, implying blocked shipping capacity. Unexpected shocks along supply chains prevent demand from being realized, particularly for durable goods (Notteboom et al., 2021). For example, the U.S. domestic demand component of overall goods consumption dropped approximately 20% in early 2020, and durable goods consumption dropped more than 30% (Kilian et al., 2021). Kilian et al. (2021) show that the primary determinant of the economic contraction in developed countries in early 2020 was the sharp drop in domestic demand. In contrast to the Great Recession, the change in consumption played a more important role in determining real economic activity during the COVID-19 crisis period. Using data on container trade volume in North America, Kilian et al. (2021) provide quantitative evidence of the impact of friction in the container shipping market on the real economy during the COVID-19 crisis. They find that these frictions led to an abrupt drop in container trade, triggering an economic recession in the U.S. in March 2020. Furthermore, they find that the recovery of North American container trade in late 2020 was substantially driven by lower frictions in the container shipping market rather than by a recovery of U.S. import and export demand. Figure 6 shows a pattern in which the forecasting performance of the satellite-based container indicator improves during the COVID-19 pandemic. The CSSED curves jump for most stock market indices in March 2020, echoing the effects of shocks to container market frictions on economic activities.

We find that the CSSED jump is more prominent in the U.S. and European markets. This finding can be explained by the heterogeneous effects of lockdown policies on different countries. Starting in mid-January 2020, China implemented an unprecedented series of national lockdown policies to contain the spread of the virus. Because of the interconnected global shipping network, the local shock in China propagated to other regions and became a global shock. Bai et al. (2022) use data derived from the Automatic Identification System to quantify the impact of pandemic lockdown policies on global port calls. They find that in February 2020, container shipping at ports with high levels of connectivity to Chinese ports was significantly affected by Chinese lockdown policies, with a time lag of two to three weeks, depending on the voyage duration of a container ship. The effects of China's lockdown on close neighbors were less serious than the effects on highly connected ports in distant regions. As a result, in mid-March 2020, significant supply chain breaks occurred in European and American ports but not in Asian ports. The unexpected frictions in container shipping markets caused a sharp drop in domestic demand in these countries, which explains the jump in return predictability revealed by the satellite-based container indicator.

Economic significance of return predictability. While the container number data show statistically significant return predictability, investors are more concerned about whether they can generate economic gains. Following the literature (Welch and

	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
Table 3 Forecasting improvement significance evaluated by bootstrap-based Diebold–Mariano (1995) statistic.					
AEX	0.003***	0.028**	0.006***	0.001***	0.030**
ATG	0.035**	0.190	0.090*	0.184	0.154
ATX	0.035**	0.075*	0.051*	0.008***	0.072*
BFX	0.004***	0.002***	0.001***	0.000***	0.009***
FCHI	0.001***	0.021**	0.002***	0.000***	0.009***
FTMIB	0.062*	0.038**	0.014**	0.014**	0.067*
FTSE	0.020**	0.057*	0.003***	0.001***	0.008***
GDAXI	0.044**	0.073*	0.006***	0.008***	0.129
IBEX	0.084*	0.061*	0.040**	0.009***	0.060*
OMXSPI	0.012**	0.052*	0.006***	0.003***	0.035**
OSEAX	0.009***	0.368	0.043**	0.003***	0.178
PSI20	0.099*	0.121	0.022**	0.022**	0.101
RTS	0.033**	0.023**	0.009***	0.006***	0.024**
SSMI	0.053*	0.024**	0.010***	0.021**	0.157
XU100	0.191	0.033**	0.008***	0.002***	0.003***
WIG30	0.066*	0.103	0.005***	0.004***	0.007***
SSEC	0.014**	0.037**	0.067*	0.049**	0.041**
BSESN	0.142	0.049**	0.443	0.066*	0.015**
CSI300	0.005***	0.047**	0.114	0.068*	0.045**
HSI	0.001***	0.144	0.014**	0.002***	0.003***
KLSE	0.051*	0.122	0.120	0.007***	0.007***
KS11	0.136	0.252	0.147	0.006***	0.103
N225	0.041**	0.066*	0.073*	0.088*	0.257
STI	0.117	0.004***	0.010**	0.001***	0.000***
TASI	0.041**	0.023**	0.038**	0.058*	0.139
COLCAP	0.096*	0.001***	0.013**	0.001***	0.004***
DJI	0.000***	0.099*	0.015**	0.001***	0.001***
GSPC	0.000***	0.086*	0.037**	0.007***	0.021**
GSPTSE	0.056*	0.355	0.234	0.002***	0.006***
IXIC	0.003***	0.183	0.172	0.037**	0.095*
MXX	0.044**	0.235	0.139	0.025**	0.108
JTOPI	0.005***	0.071*	0.002***	0.000***	0.005***
AORD	0.017**	0.214	0.041**	0.004***	0.020**
Amount of significant DM statistics	29	22	26	32	24

This table reports the *p*-value of the statistical significance of the satellite-based container model relative to the prevailing mean model using a bootstrap-based Diebold and Mariano (1995) statistic. We draw inferences by using a stationary bootstrap produce of Politis and Romano (1994) under the null hypothesis that the equity premium is unpredictable. The number of resamples is set as 2000 when bootstrapping, and the block length will be optimally estimated from the data use the selection procedure of Patton et al. (2009). The sample period covers from 2019:01 to 2021:11. The asterisks ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The last line reports the amount of cases significant at the 10% level.

Goyal, 2008; Rapach et al., 2010), we consider a risk-neutral investor who executes zero-investment strategies that invest in a risky asset (stock market index) financed with bills or in a risk-free asset (bills) financed by shorting the stock index. A popular benchmark is the unconditional investment strategy, namely buy-and-hold, in which the investor holds the stock index all of the time and only earns the equity premium itself. Although the buy-and-hold approach is naïve and no-timing, it is rarely dominated by sophisticated timing investment strategies conditional on various indicators, as identified in previous papers (Goyal et al., 2021).

In contrast, if the investor has information regarding the number of containers, he/she can seek time investments. In order to achieve a more realistic performance of our portfolio strategy, we start trading with a day lag, when the satellite information is available to investors. Here, we test the economic predictability of the number of containers, and whether it can generate higher returns than the naïve buy-and-hold strategy. Specifically, we consider two timing strategies, tilted and untilted ones, according to the return forecasts. The untilted investment strategy invests in the market index using money financed by bills when the market is expected to be bullish, but shorts the market index and saves money in bills when it is expected to be bearish. This strategy judges market trends based on whether the return forecasts are above or below the historical average. Unlike the untilted strategy, the tilted strategy switches from long stocks to short only if the

signal is very bearish, at the 25th percentile rather than the mean. The weight assigned to the risky asset is scaled according to the Z-score specification for both timing strategies. The forecast at a given time subtracts the prevailing mean (untilted) or first quartile (tilted) of the forecasts and then divides by the prevailing standard deviation. These two strategies earn the same as the unconditional strategy when the forecast is bullish, but the opposite when the forecast is bearish.

Table 4 reports the mean return and Sharpe ratio for the three investment strategies: buy-and-hold, tilted, and untilted. We find that both of the timing strategies that are conditional on return forecasts based on the number of containers can beat the benchmark of the buy-and-hold strategy. The untilted timing strategy outperforms the naïve strategy in 26 of 33 markets, with a mean return of about 14.85% per year and annual SR of 1.16, almost double the mean and SR of the benchmark strategy. The titled timing strategy shows even better performance and outperforms the naïve strategy in 30 of 33 markets, with an annualized return of 16.38% and SR of 1.19. Overall, the return predictability revealed by the number of containers is economically meaningful and profitable.

Understanding the source of return predictability
A comparison with existing shipping data. The satellite-based container number data can predict returns because they are
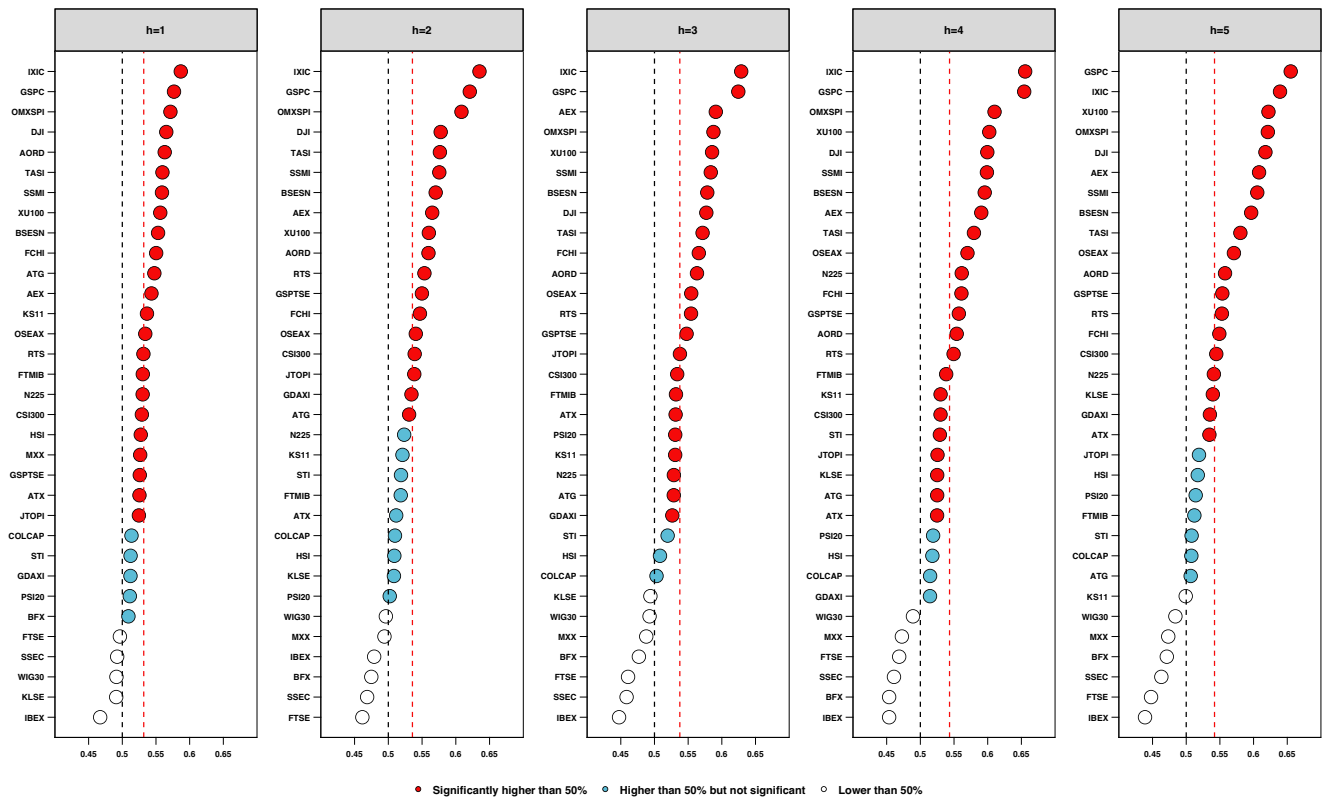
Fig. 5 Forecasting performance evaluated by directional accuracy. The forecast quality is evaluated by success ratio (SR) defined by how often the model generates forecasts with the correct direction, $SR = 1/(T - M - h + 1)\sum_{t=M+h}^{T} l_t$, $l_t = I\left(\left(r_t - \hat{r}_{t,\text{model}} \right)^2 < r_t^2 \right)$, where $\hat{r}_{t,\text{model}}$ and r_t are the return forecasts from the model of interest and the realized returns, respectively. The full sample period is from 2017-01-01 through 2021-11-01 and the out-of-sample period starts on 2019-01-01 (i.e., $M = 500$). We measure the statistical significance using the Pesaran and Timmermann (2009) test statistic. The cases significant at the 10% level are highlighted in red.

forward-looking with regard to economic activity. A natural question is whether these data are more accurate than other types of shipping data used to measure economic activity, such as the freight rates indicator (Kilian, 2009) and container throughput (Döhrn and Maatsch, 2012; Döhrn, 2019; Kilian et al., 2021). If the answer is no, our container data are meaningless.

Intuitively, on the premise that the total amount of containers in economic production is time-invariant in the short term, a larger number of idle containers stacked in the yard means fewer containers in the maritime logistics supply chain. Thus, an increase in the number of containers in the port implies reduced demand for shipping services. We first test this hypothesis by checking the predictive relationship between the number of containers in the port and the container throughput. The specification of the model is given by

$$\Delta\text{RWI\_ISL}_{t+h} = \alpha + \beta_1 \Delta\text{GNC}_t + \beta_2 \Delta\text{RWI\_ISL}_t + \varepsilon_{t+h} \quad (8)$$

The dependent variable RWI_ISL is the h-month change in container throughput, measured by the container throughput index. GNC_t is the aggregate number of containers over all 48 ports under consideration. It is a global trade indicator collected by the Leibniz Institute for Economic Research (RWI) and the Institute for Shipping Economics and Logistics (ISL) and includes information from the 82 biggest ports in the world that handle around 60% of global container throughput. The RWI_ISL index is highly correlated with global trade measures in processed goods and is a leading indicator of industrial production (Döhrn and Maatsch, 2012; Döhrn, 2019). The explanatory variables include

the lagged growth rate of the number of containers and the growth rate of container throughput.

We test the link between the number of containers and freight rates using the following equation:

$$\Delta\text{BDI}_{t+h} = \alpha + \beta_1 \Delta\text{GNC}_t + \beta_2 \Delta\text{BDI}_t + \varepsilon_{t+h} \quad (9)$$

The dependent variable is the baltic dry index (BDI), which is a good proxy of global real economic activity in industrial commodity markets. The increase in economic activity boosts demand for shipping services, causing freight rates to grow (Kilian, 2009). Similarly, the explanatory variables are the lagged growth rate in container numbers and the lagged dependent variable.

Table 5 reports the regression results at horizons of 1–4 months. The significance of the coefficients is shown by the t-statistics based on heteroscedasticity-robust standard errors. The explanatory variable ΔGNC_t is standardized. The slope coefficient of ΔGNC_t is significantly negative for both predictive regressions at the horizon of 2 months. A one standard deviation increases in ΔGNC_t is associated with a 27.2% decrease in container throughput changes and a 26.1% decrease in baltic dry index changes. This indicates that satellite-based information leads the traditional shipping indicators.

Predictive ability for real economic activity. The motivation for using port imagery to predict stock market returns comes from the close link between container shipping trade and the real economy. Maritime transport is the backbone of the global economy (McConville, 1999; Kilian, 2009; Branch and Stopford, 2013). Because 90% of the world's non-bulk dry cargo is shipped in
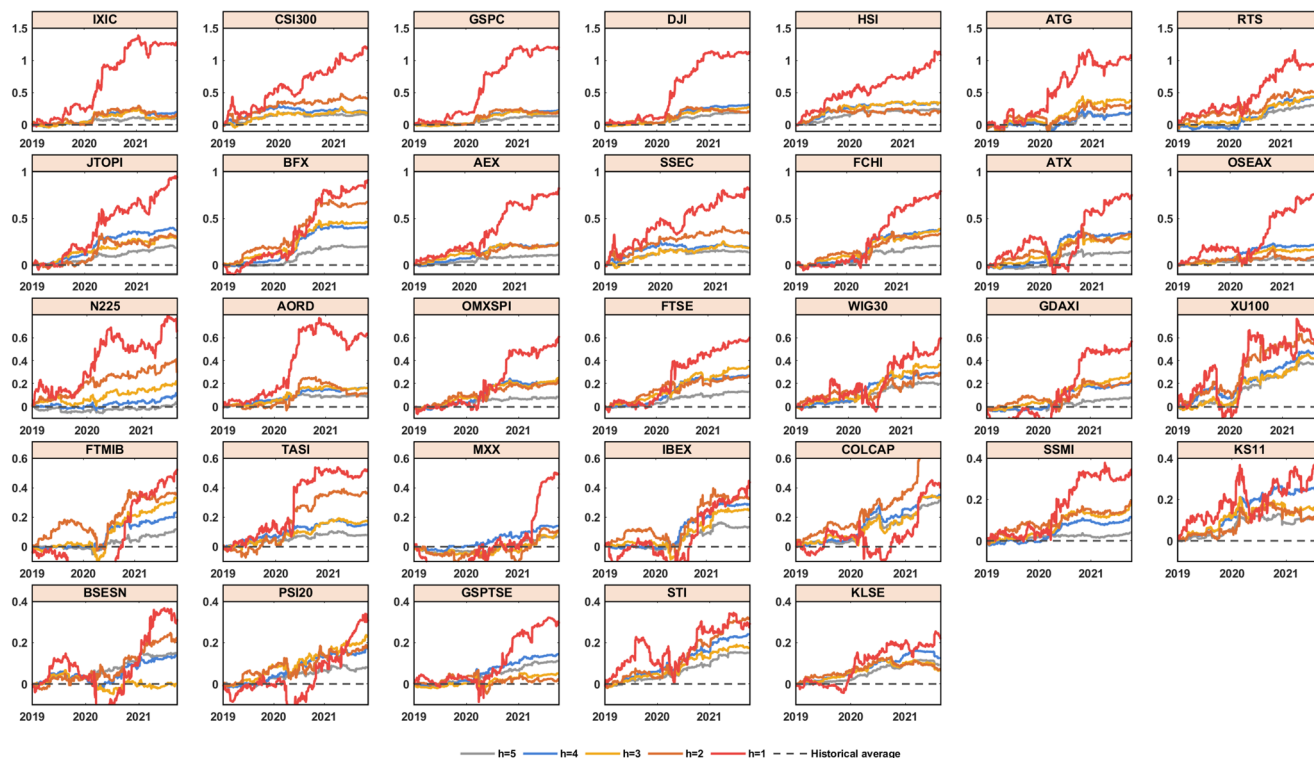
Fig. 6 Cumulative sum of squared prediction error differences relative to the historical average forecasts. This figure plots the cumulative sum of squared prediction error difference (CSSED) over time of the model of interest relative to the benchmark of the historical average model. The CSSED is computed as $\text{CSSED}_{\text{model},t} = \sum_{\tau=M+1}^{t} \left(e_{\text{bench},\tau}^2 - e_{\text{model},\tau}^2 \right)$, where $e_{\text{bench},\tau}$ and $e_{\text{model},\tau}$ denote time τ forecast errors associated with the historical average benchmark model and the given model, respectively.

containers, container trade volume is considered an appropriate measure of economic activity. Several studies have shown that container trade volume is a leading indicator of global or regional industrial production (Döhrn and Maatsch, 2012; Döhrn, 2019; Michail, 2020; Kilian et al., 2021). According to the dividend discount model, asset returns are driven by time-varying discount rates, which are functions of the state variables of the real economy (Rapach and Zhou, 2013). Therefore, economic variables that measure or predict the state of the economy are useful for predicting returns (Fama and French, 1989; Campbell and Cochrane, 1999; Cochrane, 2007, 2011). Thus, we expect that container shipping trade volume can forecast stock returns if it measures economic activity well. However, shipping trade data are not available in real-time and are always released with a one- or two-month delay (Döhrn and Maatsch, 2012). Fortunately, the satellite-based technique provides an innovative information acquisition channel, through which the information can be incorporated into prices before the government's announcement (Mukherjee et al., 2021).

We construct a real-time container shipping trade indicator based on port satellite imagery. In general, an increase in the number of containers at a port is a sign of port inefficiency, which is typically caused by congestion in transportation over urban roads, railways, and waterways (Talley, 2006). Such frictions in the container shipping markets restrict the container shipping trade. Although the literature reveals a relationship between container shipping volume and domestic demand, few studies directly link the number of containers in the yard to consumption. In contrast, the number of containers measures idle transport capacity, which is negatively related to container shipping volume. For this reason, we test whether the number of containers can anticipate real output growth. The motivation

for our test is that the global marine trade links the production activity and consumption of goods and is a key node for monitoring economic activity. Therefore, the predictive ability of the number of containers may come from its prognostication of the real economy.

We create our macroeconomic growth forecasts using the following autoregressive distributed lag model:

$$\Delta\text{RIP}_{t+h} = \alpha + \beta_1\Delta\text{GNC}_t + \beta_2\Delta\text{RIP}_t + \varepsilon_{t+h} \qquad (10)$$

where RIP_{t+h} is the growth rate of industrial production during the period $t+h$. The set of right-hand-side predictors includes the growth rate in the number of containers in the ports (ΔGNC_t) and the one-month lagged dependent variable (ΔRIP_{t+h-1}) used to accommodate autocorrelation.

Table 6 reports the regression results for 28 countries with horizons from 1 to 6 months. The last row of the table shows the results for the average growth rate of industrial production across these countries. It is evident that the number of containers can negatively predict the growth of industrial production in 27 out of 28 countries at the horizon of 4 months. In particular, 15 of the correlations show statistical significance at the 10% level. Consistent results are obtained when regressing the average growth rate of industrial production on the change in container numbers. Overall, we conclude that the link between the number of containers in ports and global real economic activity has a major role in the predictive ability of container numbers.

To shed light on the underlying economic mechanism, we now focus on the predictive ability of the satellite-based container indicator around the outbreak of COVID-19. In Table 5, we find that our satellite-based container indicator can negatively predict

shipping indicators such as the freight rate indicator of Kilian (2009) and the container throughput indicator of Döhrn and Maatsch (2012). We investigate whether the negative predictive relationship demonstrates structural breaks around the COVID-

19 crisis. Specifically, we use a predictive regression with an interaction item,

$$y_{t+h} = \alpha + \beta_1 \Delta GNC_t + \beta_2 \Delta GNC_t \times COVID_t + \beta_3 COVID_t + \beta_4 y_t + \varepsilon_{t+h}$$

(11)

where y_{t+h} represents one of the two competing shipping indicators under consideration, the freight rate and container throughput. ΔGNC_t is the change in the aggregate numbers of containers over all 48 ports. $COVID_t$ is a dummy variable that takes the value of one after the outbreak of COVID-19, defined as the period after December 2019. The coefficients that we are interested in are β_1 and β_2. As shown in Table 7, the estimates of β_1 are significantly negative at the horizon of two months, indicating that GNC predicts the traditional shipping indicators. More importantly, the estimate of β_2 is not significantly different from zero. This finding suggests that the predictive relationship is not affected by the shocks of COVID-19.

Next, we discuss the relationship between the satellite-based container indicator and real economic activity over the shocks of the COVID-19 period. Specifically, we use the following regression:

$$\Delta IP_{t+h} = \alpha + \beta_1 \Delta GNC_t + \beta_2 \Delta GNC_t \times COVID_t + \beta_3 COVID_t + \beta_4 \Delta IP_t + \varepsilon_{t+h}$$

(12)

Panel C of Table 7 reports the corresponding estimation results. We find that the β_1 estimate is significantly negative at the horizon of four months, implying that GNC leads the real economic activities at four months. The estimate of β_2 is also significantly negative, consistent with the enhanced predictive power of our indicator during the COVID-19 pandemic period. In summary, GNC is a leading measure of shipping trade and its measuring ability is not affected by the shock of the COVID-19 crisis. The simultaneous enhancement of the container information's ability to predict the stock market and global industrial output in the COVID-19 epidemic further suggests that its link to the real economy is an important source of its return predictability.

Discussion
In this paper, we investigate the ability of satellite imagery data of container ports to predict global stock market returns. The number of containers, extracted using the deep learning method, reveals significant return predictability and delivers sizeable investment gains.

Market efficiency under costly information acquisition. The efficient market hypothesis states that stock prices reflect all

Table 4 Portfolio performance of the number of containers.

	Buy-and-hold		Untilted strategy		Tilted strategy	
	Mean	SR	Mean	SR	Mean	SR
AEX	14.20	0.74	**17.27**	**1.50**	**20.17**	**1.55**
ATG	12.70	0.48	**13.59**	**0.77**	**15.38**	**0.81**
ATX	9.00	0.38	**17.19**	**1.05**	**18.41**	**1.06**
BFX	7.40	0.33	**27.31**	**1.89**	**28.53**	**1.84**
FCHI	10.30	0.48	**19.30**	**1.50**	**21.00**	**1.49**
FTMIB	11.20	0.48	**17.52**	**1.28**	**19.24**	**1.30**
FTSE	1.60	0.09	**13.42**	**1.03**	**13.69**	**1.00**
GDAXI	12.00	0.56	**16.27**	**1.23**	**18.47**	**1.27**
IBEX	0.60	0.02	**13.63**	**0.92**	**13.69**	**0.87**
OMXSPI	18.90	1.02	18.48	**1.64**	**21.23**	**1.75**
OSEAX	11.10	0.62	6.39	0.51	8.43	**0.63**
PSI20	4.90	0.25	**10.05**	**0.83**	**10.90**	**0.82**
RTS	15.70	0.66	**25.35**	**1.52**	**28.42**	**1.54**
SSMI	10.90	0.68	**14.39**	**1.56**	**16.11**	**1.59**
XU100	8.80	0.64	**20.34**	**1.40**	**22.24**	**1.44**
WIG30	2.40	0.13	**12.29**	**0.79**	**12.65**	**0.77**
SSEC	9.20	0.63	**15.62**	**1.45**	**17.16**	**1.51**
BSESN	15.30	0.81	**17.90**	**1.30**	**20.49**	**1.36**
CSI300	12.90	0.76	**17.08**	**1.38**	**19.70**	**1.48**
HSI	−4.50	−0.20	**5.14**	**0.41**	**4.62**	**0.34**
KLSE	−4.10	−0.22	**2.85**	**0.44**	**2.37**	**0.33**
KS11	10.10	0.57	8.89	**0.65**	**10.70**	**0.74**
N225	9.90	0.56	**14.18**	**1.14**	**15.67**	**1.16**
STI	−0.30	0.01	**14.36**	**1.45**	**14.39**	**1.35**
TASI	12.70	0.83	**19.10**	**1.64**	**21.78**	**1.71**
COLCAP	−1.90	−0.04	**21.00**	**1.57**	**20.90**	**1.46**
DJI	13.10	0.59	**14.91**	**1.29**	**16.52**	**1.35**
GSPC	18.40	0.87	17.87	**1.57**	**20.24**	**1.67**
GSPTSE	11.10	0.58	6.97	**0.77**	8.76	**0.90**
IXIC	26.00	1.12	18.85	**1.42**	22.43	**1.57**
MXX	4.90	0.41	**8.37**	**0.68**	**9.50**	**0.71**
JTOPI	6.00	0.38	**15.15**	**1.05**	**16.12**	**1.05**
AORD	9.70	0.53	8.98	**0.68**	**10.53**	**0.76**
Average	9.10	0.50	**14.85**	**1.16**	**16.38**	**1.19**

This table reports the annualized mean return and sharp ratio (SR) of non-timed strategies and timed strategies conditional on return forecasts. The non-timed strategy (buy-and-hold) is always bullish and earns the equity premium. The conditional timing strategy invests in the equity premium financed by the T-bill if the predictive return is above historical average (Untilted), or above its 25th percentile (Tilted); and the opposite otherwise. The weight assigned to the risky asset is scaled according to Z-score, which is equal to the return forecast minus the prevailing mean, divided by the prevailing standard deviation. The outperformed cases of timed strategies relative all the non-timed strategy are highlighted in bold.

Table 5 Forward-looking content of satellite imagery data.

	h = 1	h = 2	h = 3	h = 4
Panel A: Predictive regression for ΔRWI_ISL_{t+h}				
ΔGNC_t	0.037 (0.27)	−0.285** (−2.128)	0.132 (0.99)	−0.052 (−0.376)
ΔRWI_ISL_t	−0.167 (−1.222)	0.075 (0.558)	0.299** (2.249)	−0.243* (−1.756)
Adjusted R^2	2.76%	8.12%	11.85%	6.55%
Panel B: Predictive regression for ΔBDI_{t+h}				
ΔGNC_t	−0.188 (−1.405)	−0.271** (−2.095)	0.163 (1.228)	0.097 (0.716)
ΔBDI_t	0.092 (0.691)	−0.039 (−0.302)	0.157 (1.183)	−0.151 (−1.128)
Adjusted R^2	4.27%	8.05%	5.69%	3.29%

This table provides evidence that growth in the number of containers in the global ports (GNC) predicts a decrease in global container throughput (RWI) and the decrease of freight rate (BDI). The sample period covers from 2017:01 to 2021:10. We report t-statistics based on heteroscedasticity robust standard errors in parentheses. The asterisks ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6 Forecasting real industry production growth with the number of containers.						
	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
Netherlands	0.154	−0.09	0.07	−0.225	−0.012	0.036
Greece	0.089	−0.07	0.114	−0.078	−0.172	0.048
Austria	−0.016	0.071	−0.001	−0.356**	0.116	0.119
Belgium	0.033	0.019	−0.077	−0.094	0.189	0.046
France	−0.062	0.042	−0.154	−0.278*	0.159	0.134
Italy	0.008	0.037	−0.246*	−0.23	0.273*	0.052
UK	−0.112	0.021	0.041	−0.326**	0.113	0.061
Germany	−0.069	0.058	−0.044	−0.408***	0.167	0.12
Spain	−0.076	−0.02	−0.071	−0.269*	0.232	0.077
Sweden	0.059	−0.101	0.03	−0.325**	0.095	0.086
Norway	0.086	0.074	−0.05	−0.089	−0.052	0.018
Portugal	0.065	0.027	0.003	−0.354***	−0.032	0.036
Russia	−0.037	−0.009	0.124	−0.068	0.062	−0.103
Switzerland	0.139	−0.011	−0.025	−0.129	−0.123	0.203
Turkey	−0.085	0.079	0.002	−0.363**	0.126	0.015
Poland	−0.057	0.063	−0.037	−0.335**	0.163	0.11
China	−0.119	−0.168	0.282**	−0.037	0.063	0.062
India	0	0.075	0.05	−0.512***	0.206	0.014
Malaysia	−0.113	−0.173	0.146	−0.312**	0.254*	0.085
South Korea	0.061	0.024	0.126	−0.193	−0.196	0.213
Japan	0.02	0.041	0.05	−0.166	−0.219	0.159
Singapore	0.347***	−0.164	0.001	0.065	−0.16	0.116
Colombia	0.114	0.022	−0.14	−0.192	−0.032	0.229
America	−0.109	0.022	0.018	−0.327**	0.096	0.127
Canada	−0.067	0.072	−0.041	−0.281*	0.131	0.072
Mexico	−0.029	0.04	0.044	−0.349**	−0.044	0.194
South Africa	−0.125	−0.005	0.095	−0.358**	0.198	0.008
Australia	0.009	−0.236*	−0.297**	−0.309**	0.079	0.152
Average	0.004	0.014	−0.005	−0.42***	0.116	0.153

This table provides evidence that growth in the number of containers in the global ports (GNC) predicts the real industry production growth for different countries. The sample covers the period from 2017:01 to 2021:10. We report t-statistics based on heteroscedasticity robust standard errors in parentheses. The asterisks ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7 Forward-looking content of satellite imagery data during COVID-19 pandemic.				
	$h=1$	$h=2$	$h=3$	$h=4$
Panel A: Predictive regression for ΔRWI_ISL_{t+h}				
ΔGNC_t	0.005 (0.031)	−0.311* (−1.911)	0.105 (0.652)	−0.119 (−0.727)
$\Delta GNC_t \times COVID_t$	−0.029 (−0.099)	0.163 (0.579)	0.003 (0.012)	0.072 (0.250)
$COVID_t$	0.099 (0.354)	0.382 (1.389)	0.306 (1.113)	0.456 (1.613)
ΔRWI_ISL_t	−0.163 (−1.181)	0.057 (0.422)	0.297** (2.225)	−0.248* (−1.821)
Adjusted R^2	2.90%	10.39%	13.55%	11.79%
Panel B: Predictive regression for ΔBDI_{t+h}				
ΔGNC_t	−0.220 (−1.373)	−0.319** (−1.991)	0.192 (1.189)	0.104 (0.621)
$\Delta GNC_t \times COVID_t$	0.182 (0.646)	−0.095 (−0.337)	0.025 (0.087)	0.009 (0.030)
$COVID_t$	0.391 (1.461)	0.552** (2.042)	0.471* (1.703)	0.513* (1.787)
ΔBDI_t	0.093 (0.703)	−0.053 (−0.399)	0.148 (1.091)	−0.150 (−1.074)
Adjusted R^2	12.24%	13.35%	12.47%	9.14%
Panel C: Predictive regression for ΔAIP_{t+h}				
ΔGNC_t	0.062 (0.362)	−0.125 (−0.747)	−0.036 (−0.209)	−0.244*** (−2.975)
$\Delta GNC_t \times COVID_t$	−0.012 (−0.040)	0.396 (1.341)	0.091 (0.301)	−0.510* (−1.689)
$COVID_t$	−0.112 (−0.388)	−0.082 (−0.285)	0.008 (0.028)	0.181 (0.579)
ΔIP_t	0.037 (0.260)	−0.171 (−1.230)	−0.178 (−1.251)	−0.066 (−0.531)
Adjusted R^2	0.78%	5.68%	3.20%	9.79%

This table investigates the predictive ability of global ports (GNC) to global container throughput (RWI), the decrease in freight rate (BDI), and the growth rate of averaged real industry production (AIP). To examine the change of predictive relationship around COVID-19, we add a dummy variable that takes the value of one on the COVID-19 period, defined as the time after 2020. The sample period covers from 2017:01 to 2021:11. We report t-statistics based on heteroscedasticity robust standard errors in parentheses. The asterisks ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

available information (Fama, 1970). In an environment with perfect information, the market value of information approaches zero (Copeland and Friedman, 1992; Sunder, 1992) and there is no motivation for information-gathering. However, this hypothesis relies on the assumption that information acquisition is costless. Our finding of the return predictability of satellite information is consistent with the argument for market efficiency under costly information acquisition (Grossman and Stiglitz, 1980; Verrecchia, 1982). According to this argument, an investor who spends resources on collecting information expects to receive

compensation in a market where information acquisition is costly. The information acquisition leads to informed trades and improves the information content in prices, further promoting price efficiency. When informed investors observe positive information about security, they bid the current price up, and conversely, the future return will be low. In this way, the private information is quickly incorporated into stock prices. This process provides informed traders with profits and accordingly generates return predictability.

The application of satellite data increases price informativeness. In comparison with traditional data sources such as government or corporate announcements, satellite data have the advantages of real-time availability, frequency, and quality. As a result, this alternative data source provides forward-looking information about future earnings that can be incorporated into stock prices (Katona et al., 2018; Zhu, 2019; Mukherjee et al., 2021). The information about future earnings contained in satellite imagery effectively reduces insider trading. As a result, the availability of such alternative data decreases information asymmetry between firm insiders and outside investors. Managers have less opportunity to trade profitably on their private information about future earnings because a fraction of such information is available to outside investors through an alternative source. Therefore, we believe that the use of satellite data in investment decisions will improve market price efficiency.

The horizon of return predictability. We find return predictability at horizons of up to 5 days using satellite imagery data. One may wonder why satellite information is not instantly incorporated into the price—for example, on the same day that the information is released. Instead, the existence of return predictability implies a lead–lag effect.

We provide two explanations. First, the pre-processing and delivery of satellite data take time. Satellite data are usually sold as different levels of products instead of the raw data. Because the raw data only contain the original sensor signals, they cannot provide any useful insights to investors. They must be subjected to a series of pre-processing steps. For example, the transformation of the Sentinel-2 raw data into human-understandable images includes decompression, radiometric corrections, geometric viewing model refinement, resampling, and conversion to reflectance. Normally, it takes nearly 3 h to publish an early version of satellite data products and 48–60 h for carefully calibrated versions. This delay is inevitable for any satellite platform. As most raw satellite images are captured from 10:30 a.m. to 1:30 p.m. local time, the extracted satellite information usually arrives in the stock market with a one-day or longer delay.

Second, processing satellite images requires techniques that unsophisticated individual investors lack. The high threshold for processing satellite data naturally prevents most investors from using them, making their price effect gradual instead of instant. Jensen (1978) argues that when new information enters the market it is instantly integrated into stock prices if enough investors take it into account, leading to extremely short-term return predictability. However, only a portion of investors takes satellite data into account. Most satellite products and customized services are so expensive that only institutional traders can afford them. Although there are some publicly available satellite data resources accessible to individual investors, most do not have enough image-processing techniques to interpret the information in a timely fashion. Thus, the prices partially reflect the information of informed individuals (arbitrageurs) (Grossman and Stiglitz, 1980). This private information diffuses through the market as it is acquired by investors. The average private signal

can be revealed by a sequence of prices, and uninformed investors condition their trades on public observables such as prices and order flows. When less-informed investors observe private signals, they may rationally behave like price chasers (Grundy and McNichols, 1989; Wang, 1993). As a result, the private satellite information is integrated into the price. It takes longer for satellite information to be incorporated into market price, based on Jensen's (1978) argument.

Therefore, it is reasonable that our satellite-based container number index shows market return predictability at horizons of up to 5 days. The predictability across several days in fact reflects the further revelation of existing private information as time goes by.

Conclusions

Due to the development of big data technology, satellite imagery data are increasingly appreciated by sophisticated investors. In comparison with traditional economic data such as industrial production and consumer price indices, satellite data have greater frequency and are available in real-time. This paper provides a fresh example of the application of satellite data in financial investment decision-making.

As changes in the number of containers reflect real economic activities, we draw on container information based on satellite imagery of major global ports using a deep learning method. Our results suggest that the combined container information predicts stock returns in global mainstream markets. The return predictability is both statistically and economically significant. The predictive power of container information became stronger after the COVID-19 pandemic, especially in the U.S. and European markets. We also show that satellite-based container information leads to the traditional shipping indicators of freight rates and container throughput. The container information is also found to be closely linked to global economic activity, serving as a potential explanation for the return predictability.

Container number data is not available in real-time. Container coverage drawn from satellite images provides a good proxy and can reflect economic changes, thus helping investors in the stock market to obtain excess returns. We execute an investment experiment and find that an investor making use of satellite imagery of seaports will on average receive annualized returns of 16%. In practice, many reasons may reduce the investment gains using satellite information. For example, it typically takes several hours or even longer time to transform the original sensor signals into images and process satellite images. We have accounted for such lags in portfolio analysis. Nevertheless, during the period of data processing, container information may flow into the stock market via other channels such as online text, although the availability of these alternative data remains another problem. We believe that with the development of big data technology, satellite images provide an important information source to help improve investment decisions.

References

Andersson TR, Hosking JS, Perez-Ortiz M, Paige B, Elliott A, Russell C, Law S, Jones DC, Wilkinson J, Phillips T, Byrne J, Tietsche S, Sarojini BB,

Blanchard-Wrigglesworth E, Aksenov Y, Downie R, Shuckburgh E (2021) Seasonal Arctic sea ice forecasting with probabilistic deep learning. Nat Commun 12(1):1–12. https://doi.org/10.1038/s41467-021-25257-4

Bai X, Xu M, Han T, Yang D (2022) Quantifying the impact of pandemic lockdown policies on global port calls. Transp Res Part A: Policy Pract 164:224–241. https://doi.org/10.1016/j.tra.2022.08.002

Branch A, Stopford M (2013) Maritime economics. Routledge

Burke M, Driscoll A, Lobell DB, Ermon S (2021) Using satellite imagery to understand and promote sustainable development. Science 371(6535):eabe8628. https://doi.org/10.1126/science.abe8628

Campbell JY (1987) Stock returns and the term structure. J Finan Econ 18(2):373–399. https://doi.org/10.1016/0304-405X(87)90045-6

Campbell JY, Cochrane JH (1999) By force of habit: a consumption-based explanation of aggregate stock market behavior. J Political Econ 107(2):205–251. https://doi.org/10.1086/250059

Campbell JY, Shiller RJ (1988) The dividend-price ratio and expectations of future dividends and discount factors. Rev Finan Stud 1(3):195–228. https://doi.org/10.1093/rfs/1.3.195

Chen X, Nordhaus WD (2011) Using luminosity data as a proxy for economic statistics. Proc Natl Acad Sci USA 108(21):8589–8594. https://doi.org/10.1073/pnas.1017031108

Claeskens G, Magnus JR, Vasnev AL, Wang W (2016) The forecast combination puzzle: a simple theoretical explanation. Int J Forecast 32(3):754–762. https://doi.org/10.1016/j.ijforecast.2015.12.005

Clark TE, West KD (2007) Approximately normal tests for equal predictive accuracy in nested models. J Econometr 138(1):291–311. https://doi.org/10.1016/j.jeconom.2006.05.023

Cochrane JH (2007) Financial markets and the real economy. In: Mehra R (Ed.) Handbook of the equity premium. Elsevier, Amsterdam

Cochrane JH (2011) Presidential address: discount rates. J Finance 66(4):1047–1108. https://doi.org/10.1111/j.1540-6261.2011.01671.x

Copeland TE, Friedman D (1992) The market value of information: some experimental results. J Bus 241:266. https://www.jstor.org/stable/2353164

Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13(3):253–263. https://doi.org/10.1080/07350015.1995.10524599

Döhrn R (2019) Sieben Jahre RWI/ISL-Containerumschlag-Index–ein Erfahrungsbericht. Wirtschaftsdienst 99(3):224–226. https://doi.org/10.1007/s10273-019-2422-z

Döhrn R, Maatsch S (2012) Der RWI/ISL-Containerumschlag-Index: Ein neuer Frühindikator für den Welthandel. Wirtschaftsdienst 92(5):352–354. https://doi.org/10.1007/s10273-012-1387-y

Dolz J, Ben Ayed I, Desrosiers C (2018) Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) International MICCAI brain lesion workshop. Springer, pp. 271–282

Fama EF (1970) Efficient capital markets: a review of theory and empirical work. J Financ 25(2):383–417. https://doi.org/10.2307/2325486

Fama EF, French KR (1989) Business conditions and expected returns on stocks and bonds. J Finan Econ 25(1):23–49. https://doi.org/10.1016/0304-405X(89)90095-0

Fama EF, Schwert GW (1977) Asset returns and inflation. J Finan Econ 5(2):115–146. https://doi.org/10.1016/0304-405X(77)90014-9

Goyal A, Welch I (2003) Predicting the equity premium with dividend ratios. Manag Sci 49(5):639–654. https://doi.org/10.1287/mnsc.49.5.639.15149

Goyal A, Welch I, Zafirov A (2021) A comprehensive look at the empirical performance of equity premium prediction II. https://doi.org/10.2139/ssrn.3929119

Grossman SJ, Stiglitz JE (1980) On the impossibility of informationally efficient markets. Am Econ Rev 70(3):393–408. http://www.jstor.org/stable/1805228

Grundy BD, McNichols M (1989) Trade and the revelation of information through prices and direct disclosure. Rev Finan Stud 2(4):495–526. https://doi.org/10.1093/rfs/2.4.495

Guo H (2006) On the out-of-sample predictability of stock market returns. J Bus 79(2):645–670

Henderson JV, Storeygard A, Weil DN (2012) Measuring economic growth from outer space. Am Econ Rev 102(2):994–1028. https://doi.org/10.1257/aer.102.2.994

Huang Y-J, Dou Q, Wang Z-X, Liu L-Z, Jin Y, Li C-F, Wang L, Chen H, Xu R-H (2021) 3-D RoI-aware U-net for accurate and efficient colorectal tumor segmentation. IEEE Trans Cybern 51(11):5397–5408. https://doi.org/10.1109/TCYB.2020.2980145

Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S (2016) Combining satellite imagery and machine learning to predict poverty. Science 353(6301):790–794. https://doi.org/10.1126/science.aaf7894

Jensen MC (1978) Some anomalous evidence regarding market efficiency. J Finan Econ 6(2/3):95–101. https://doi.org/10.1016/0304-405X(78)90025-9

Katona Z, Painter M, Patatoukas PN, Zeng J (2018) On the capital market consequences of alternative data: evidence from outer space. In: 9th Miami behavioral finance conference. https://doi.org/10.2139/ssrn.3222741

Kilian L (2009) Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. Am Econ Rev 99(3):1053–69. https://doi.org/10.1257/aer.99.3.1053

Kilian L, Nomikos N K, Zhou X (2021) Container trade and the US recovery. Center for Financial Studies Working Paper (659). https://doi.org/10.2139/ssrn.3963305

Kothari SP, Shanken J (1997) Book-to-market, dividend yield, and expected market returns: a time-series analysis. J Finan Econ 44(2):169–203. https://doi.org/10.1016/S0304-405X(97)00002-0

Kruitwagen L, Story KT, Friedrich J, Byers L, Skillman S, Hepburn C (2021) A global inventory of photovoltaic solar energy generating units. Nature 598(7882):604–610. https://doi.org/10.1038/s41586-021-03957-7

Mark NC (1995) Exchange rates and fundamentals: evidence on long-horizon predictability. Am Econ Rev 85(1):201–218. https://www.jstor.org/stable/2118004

McConville J (1999) Economics of maritime transport: theory and practice. Witherby

Meeks A, Isidore C, Yurkevich V (2021) North America's biggest container port faces record backlog. https://edition.cnn.com/2021/10/18/business/container-port-record-backlog/index.html

Michail NA (2020) World economic growth and seaborne trade volume: quantifying the relationship. Transp Res Interdiscip Perspects 4:100108. https://doi.org/10.1016/j.trip.2020.100108

Mirza MU, Xu C, van Bavel B, van Nes EH, Scheffer M (2021) Global inequality remotely sensed. Proc Natl Acad Sci USA 118(18). https://doi.org/10.1073/pnas.1919913118

Mukherjee A, Panayotov G, Shon J (2021) Eye in the sky: private satellites and government macro data. J Finan Econ 141(1):234–254. https://doi.org/10.1016/j.jfineco.2021.03.002

Notteboom T, Pallis T, Rodrigue J-P (2021) Disruptions and resilience in global container shipping and ports: the COVID-19 pandemic versus the 2008–2009 financial crisis. Marit Econ Logist 23(2):179–210. https://doi.org/10.1057/s41278-020-00180-5

Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B (2018) Attention u-net: learning where to look for the pancreas. arXiv preprint https://doi.org/10.48550/arXiv.1804.03999

Patton A, Politis DN, White H (2009) Correction to "Automatic block-length selection for the dependent bootstrap" by D. Politis and H. White. Econom Rev 28(4):372–375. https://doi.org/10.1080/07474930802459016

Pesaran MH, Timmermann A (2009) Testing dependence among serially correlated multicategory variables. J Am Stat Assoc 104(485):325–337. https://doi.org/10.1198/jasa.2009.0113

Politis DN, Romano JP (1994) The stationary bootstrap. J Am Stat Assoc 89(428):1303–1313. https://doi.org/10.1080/01621459.1994.10476870

Pontiff J, Schall LD (1998) Book-to-market ratios as predictors of market returns. J Finan Econ 49(2):141–160. https://doi.org/10.1016/S0304-405X(98)00020-8

Rapach D, Zhou G (2013) Forecasting stock returns. In: Elliott G, Timmermann A (eds) Handbook of economic forecasting. Elsevier, pp. 328–383

Rapach DE, Strauss JK, Zhou G (2010) Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Rev Finan Stud 23(2):821–862. https://doi.org/10.1093/rfs/hhp063

Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds) International conference on medical image computing and computer-assisted intervention. Springer, pp. 234–241

Smith J, Wallis KF (2009) A simple explanation of the forecast combination puzzle. Oxf Bull Econ Stat 71(3):331–355. https://doi.org/10.1111/j.1468-0084.2008.00541.x10.1111/j.1468-0084.2008.00541.x

Sunder S (1992) Market for information: experimental evidence. Econometrica 667–695. https://doi.org/10.2307/2951588

Talley WK (2006) An economic theory of the port. Res Transp Econ 16:43–65. https://doi.org/10.1016/S0739-8859(06)16003-5

Timmermann A (2006) Forecast combinations. In: Elliott G, Granger CWJ, Timmermann A (eds) Handbook of Economic Forecasting, Ch. 4, vol 1. pp. 135–196

Verrecchia RE (1982) Information acquisition in a noisy rational expectations economy. Econometrica 50:1415–1430. https://doi.org/10.2307/1913389

Wang J (1993) A model of intertemporal asset prices under asymmetric information. Rev Econ Stud 60(2):249–282. https://doi.org/10.2307/2298057

Watmough GR, Marcinko CL, Sullivan C, Tschirhart K, Mutuo PK, Palm CA, Svenning J-C (2019) Socioecologically informed use of remote sensing data to predict rural household poverty. Proc Natl Acad Sci USA 116(4):1213–1218. https://doi.org/10.1073/pnas.1812969116

Welch I, Goyal A (2008) A comprehensive look at the empirical performance of equity premium prediction. Rev Finan Stud 21(4):1455–1508. https://doi.org/10.1093/rfs/hhm014

Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2020) Unet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging 39(6):1856–1867. https://doi.org/10.1109/TMI.2019.2959609

Zhu C (2019) Big data as a governance mechanism. Rev Finan Stud 32(5):2021–2061. https://doi.org/10.1093/rfs/hhy081

Author contributions
HY, YW, XH, and LW conceived the study. XH, LW, and YZ collected and analyzed satellite imagery. XH and LW designed data preprocessing pipelines trained deep learning models, and YZ analyzed their output. XH and YW finished the forecasting process. YW, XH, and LW wrote the paper. HY and YW managed the project and provided funding support. These authors contributed equally to this work.

Competing interests
The authors declare no competing interests.

Ethical approval
All experimental data are obtained from publicly available data sources and do not include any personal information.

Informed consent
This article does not contain any studies with human participants performed by any of the authors.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-023-01891-9.

Correspondence and requests for materials should be addressed to Yudong Wang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.