

Massive Data Analytics for Macroeconomic Nowcasting



Peng Cheng, Laurent Ferrara, Alice Froidevaux, and Thanh-Long Huynh

Abstract Nowcasting macroeconomic aggregates have proved extremely useful for policy-makers or financial investors, in order to get real-time, reliable information to monitor a given economy or sector. Recently, we have witnessed the arrival of new large databases of alternative data, stemming from the Internet, social media, satellites, fixed sensors, or texts. By correctly accounting for those data, especially by using appropriate statistical and econometric approaches, the empirical literature has shown evidence of some gain in nowcasting ability. In this chapter, we propose to review recent advances of the literature on the topic, and we put forward innovative alternative indicators to monitor the Chinese and US economies.

1 Introduction

Real-time assessment of the economic activity in a country or a sector has proved extremely useful for policy-makers in order to implement contra-cyclical monetary or fiscal policies or for financial investors in order to rapidly shift portfolios. Indeed in advanced economies, Quarterly National Accounts are generally published on

All views expressed in this paper are those of the authors and do not represent the views of JPMorgan Chase or any of its affiliates.

P. Cheng (✉)
JPMorgan Chase, New York, NY, USA
e-mail: peng.cheng@jpmorgan.com

L. Ferrara
QuantCube Technology, Paris, France
SKEMA Business School, Lille, France
e-mail: laurent.ferrara@skema.edu

A. Froidevaux · T.-L. Huynh
QuantCube Technology, Lille, France
e-mail: af@q3-technology.com; thanh-long.huynh@q3-technology.com

a quarterly basis by Statistical Institutes, with a delay of release of about 1 or 2 months, especially the benchmark macroeconomic indicator, that is, the gross domestic product (GDP). For example, to be aware of the economic activity in the first quarter of the year (from beginning of January to end of March), we need sometimes to wait until mid-May, depending on the country considered. This is also especially true when dealing with emerging economies, where sometimes only low-frequency macroeconomic aggregates are available (e.g., annual aggregates). In this context, macroeconomic nowcasting has become extremely popular in both theoretical and empirical economic literatures. Giannone et al. [35] were the first to develop econometric models, namely, dynamic factor models, in order to propose high-frequency nowcast of US GDP growth. Currently, the Federal Reserve Bank of Atlanta (GDPNow) and the Federal Reserve Bank of New York have developed their own nowcasting tools for US GDP, available in real time. Beyond the US economy, many nowcasting tools have been proposed to monitor macroeconomic aggregates either for advanced economies (see among others [2] for the euro area and [13] for Japan) or for emerging economies (see among others [14] for Brazil or [42] for Turkey). At the global level, some papers are also trying to assess world economic conditions in real time by nowcasting the world GDP that is computed on a regular basis by the IMF when updating the *World Economic Outlook* report, four times per year (see, e.g., [28]).

Assessing economic conditions in real time is generally done by using standard official economic information such as production data, sales, opinion surveys, or high-frequency financial data. However, we recently witnessed the arrival of massive datasets that we will refer to as *alternative datasets* in opposition to *official datasets*, stemming from various sources of information. The multiplication in recent years of the number of accessible alternative data and the development of methods based on machine learning and artificial intelligence capable of handling them constitute a break in the way of following and predicting the evolution of the economy. Moreover, the power of digital data sources is the real-time access to valuable information stemming from, for example, multi-lingual social media, satellite imagery, localization data, or textual databases.

The availability of those new alternative datasets raises important questions for practitioners about their possible use. One central question is to know whether and when those alternative data can be useful in modeling and nowcasting/forecasting macroeconomic aggregates, once we control for official data. From our reading of the recent literature, it seems that the gain from using alternative data depends on the country under consideration. If the statistical system of the country is well developed, as it is generally the case in advanced economies, then alternative data are able to generate proxies that can be computed on a high-frequency basis, well in advance of the release of official data, with a high degree of reliability (see, e.g., [29], as regards the euro area). In some cases, alternative data allows a high-frequency tracking of some specific sectors (e.g., tourism or labor market). If the statistical system is weak, as it may be in some emerging or low-income economies where national accounts are only annual and where some sectors are not covered, alternative data are likely to fill, or at least narrow, some information gaps and

efficiently complement the statistical system in monitoring economic activity (see, e.g., [44]).

In this chapter, we first propose to review some empirical issues for practitioners when dealing with massive datasets for macroeconomic nowcasting. Then we give some examples of tracking for some sectors/countries, based on recent methodologies for massive datasets. The last section contains two real-time proxies for US and Chinese economic growth that have been developed by QuantCube Technology in order to track GDP growth rate on a high-frequency basis. The last section concludes by proposing some applications of macroeconomic nowcasting tools.

2 Review of the Recent Literature

This section presents a short review of the recent empirical literature on nowcasting with massive datasets of alternative data. We don't pretend to do an exhaustive review as this literature is quite large, but rather to give a flavor of recent trends. We first present the various types of alternative data that have been recently considered, then we describe econometric approaches able to deal with this kind of data.

2.1 *Various Types of Massive Data*

Macroeconomic nowcasting using alternative data involves the use of various types of massive data.

Internet data that can be obtained from web scraping techniques constitute a broad source of information, especially Google search data. Those data have been put forward by Varian [53] and Choi and Varian [19] and have been widely and successfully used in the empirical literature to forecast and nowcast various macroeconomic aggregates.¹ Forecasting prices with Google data has been also considered, for example, by Seabold and Coppola [48] who focus on a set of Latin American countries for which publication delays are quite large. Besides Google data, crowd-sourced data from online platforms, such as Yelp, provide accurate real-time geographical information. Glaeser et al. [37] present evidence that Yelp data can complement government surveys by measuring economic activity in real time at a granular level and at almost any geographic scale in the USA.

The availability of high-resolution satellite imagery has led to numerous applications in economics such as urban development, building type, roads, pollution, or agricultural productivity (for a review, see, e.g., [24]). However, as regards high-frequency nowcasting of macroeconomic aggregates, applications are more

¹Examples of applications include household consumption [19], unemployment rate [23], building permits [21], car sales [45], or GDP growth [29].

scarce. For example, Clarck et al. [20] propose to use data on satellite-recorded nighttime lights as a benchmark for comparing various published indicators of the state of the Chinese economy. Their results are consistent with the rate of Chinese growth being higher than is reported in the official statistics. Satellites can be considered as mobile sensors, but information can also be taken from fixed sensors such as weather/pollution sensors or traffic sensors/webcams. For example, Askitas and Zimmermann [5] show that toll data in Germany, which measure monthly transportation activity performed by heavy transport vehicles, are a good early indicator of German production and are thus able to predict in advance German GDP. Recently, Arslanalp et al. [4] put forward vessel traffic data from automatic identification system (AIS) as a massive data source for nowcasting trade activity in real time. They show that vessel data are good complements of existing official data sources on trade and can be used to create a real-time indicator of global trade activity.

Textual data have been also recently used for nowcasting purposes in order to compute various indexes of sentiment that are then put into standard econometric models. In general, textual analyses are useful to estimate unobserved variables that are not directly available or measured by official sources. A well-known example is economic policy uncertainty that has been estimated for various countries by Baker et al. [8] starting from a large dataset of newspapers by identifying some specific keywords. Those economic policy uncertainty (EPU) indexes have proved useful to anticipate business cycle fluctuations, as recently shown by Rogers and Xu [47], though their real-time performance has to be taken with caution. Various extensions of this approach have been proposed in the literature, such as the geopolitical risk index by Caldara and Iacovello [17] that can be used to forecast business investment. Kalamara et al. [41] recently proposed to extract sentiment from various newspapers using different machine learning methods based on dictionaries and showed that they get some improvement in terms of UK GDP forecasting accuracy. In the same vein, Fraiberger et al. [32] estimate a media sentiment index using more than 4.5 million Reuters articles published worldwide between 1991 and 2015 and show that it can be used to forecast asset prices.

Payment data by credit cards have been shown to be a valuable source of information to nowcast household consumption. These card payment data are generally free of sampling errors and are available without delays, providing thus leading and reliable information on household spending. Aastveit et al. [1] show that credit card transaction data improve both point and density forecasts for Norway and underline the usefulness of getting such information during the Covid-19 period. Other examples of application of payment data for nowcasting economic activity include among others Galbraith and Tkacz [33], who nowcast Canadian GDP and retail sales using electronic payment data, or Aprigliano et al. [3], who assess the ability of a wide range of retail payment data to accurately forecast Italian GDP and its main domestic components.

Those massive alternative data have the great advantage of being available at a very high frequency, thus leading to signals that can be delivered well ahead of official data. Also, those data are not revised, avoiding thus a major issue for

forecasters. However, there is no such thing as a free lunch. An important aspect that is not often considered in empirical works is about the cleaning of raw data. Indeed, it turns out that unstructured raw data are often polluted by outliers, seasonal patterns, or breaks, temporary or permanent. For example, when dealing with daily data, they can present two or more seasonalities (e.g., weekly and annual). In such a case, seasonal adjustment is not an easy task and should be carefully considered. An exhaustive review of various types of alternative data that can be considered for nowcasting issues is presented in [16].

2.2 *Econometric Methods to Deal with Massive Datasets*

Assume we have access to a massive dataset ready to be put into an econometric model. Generally, those datasets present two stylized facts: (1) a large number n of variables compared to the sample size T and (2) a frequency mismatch between the targeted variable (quarterly in general) and explanatory variables (monthly, weekly, or daily).

Most of the time, massive datasets have an extremely large dimension, with the number of variables much larger than the number of observations (i.e., $n \gg T$, sometimes referred to as *fat datasets*). The basic equation for nowcasting a target variable y_t using a set of variables

$$y_t = \beta_1 x_{1t} + \dots + \beta_n x_{nt} + \varepsilon_t, \quad (1)$$

where $\varepsilon_t \sim N(0, \sigma^2)$. To account for dynamics, x_{jt} can also be a lagged value of the target variable or of other explanatory variables. In such a situation, usual least-squares estimates are not necessarily a good idea as there are too many parameters to estimate, leading to a high degree of uncertainty in estimates, as well as a strong risk of over-fitting in-sample associated to poor out-of-sample performances. There are some econometric approaches to address the curse of dimensionality. Borrowing from Giannone et al. [36], we can classify those approaches in two categories: *sparse* and *dense* models. Sparse methods assume that some β_j coefficients in Eq. (1) are equal to zero. This means that only few variables have an impact on the target variable. Zeros can be imposed ex ante by the practitioners based on specific a priori information. Alternatively, zeros can be estimated using an appropriate estimation method such as the LASSO (*least absolute shrinkage and selection operator*) regularization approach [51] or some Bayesian techniques that impose some coefficients to take null values during the estimation step (see, e.g., Smith et al. [49] who develop a Bayesian approach that can shrink some coefficients to zero and allows coefficients that are shrunk to zero to vary through regimes).

In opposition, dense methods assume that all the explanatory variables have a role to play. A typical example is the dynamic factor model (DFM) that tries to estimate a common factor from all the explanatory variables in the following way:

$$x_t = \Lambda f_t + \xi_t, \quad (2)$$

where $x_t = (x_{1t}, \dots, x_{nt})'$ is a vector of n stationary time series and x_t is decomposed into a common component Λf_t where $f_t = (f_{1t}, \dots, f_{rt})'$ and Λ is the loading matrix such that $\Lambda = (\lambda_1, \dots, \lambda_n)'$ and an idiosyncratic component $\xi_t = (\xi_{1t}, \dots, \xi_{nt})'$ a vector of n mutually uncorrelated components. A VAR(p) dynamics is sometimes allowed for the vector f_t . Estimation is carried out using the diffusion index approach of Stock and Watson [50] or the generalized DFM of Forni et al. [30]. As the number r of estimated factors \hat{f}_t is generally small, they can be directly put in a second step into the regression equation to explain y_t in the following way:

$$y_t = \gamma_1 \hat{f}_{1t} + \dots + \gamma_r \hat{f}_{rt} + \varepsilon_t. \quad (3)$$

We refer, for example, to [7, 10, 9], for examples of application of this approach.

Another well-known issue when nowcasting a target macroeconomic variable with massive alternative data is the frequency mismatch as y_t is generally a low-frequency variable (e.g., quarterly), while explanatory variables x_t are generally high frequency (e.g., daily). A standard approach is to first aggregate the high-frequency variables to the low frequency by averaging and then to estimate Eq. (1) at the lowest frequency. Alternatively, mixed-data sampling (MIDAS hereafter) models have been put forward by Ghysels et al. [34] in order to avoid systematically aggregating high-frequency variables. As an example, let's consider the following MIDAS bivariate equation:

$$y_t = \beta_0 + \beta_1 \times B(\theta) x_t^{(m)} + \varepsilon_t \quad (4)$$

where $(x_t^{(m)})$ is an exogenous stationary variable sampled at a frequency higher than (y_t) such that we observe m times $(x_t^{(m)})$ over the period $[t-1, t]$. The term $B(\theta)$ controls the polynomial weights that allows the frequency mixing. Indeed, the MIDAS specification consists in smoothing the past values of $(x_t^{(m)})$ by using the polynomial $B(\theta)$ of the form:

$$B(\theta) = \sum_{k=1}^K b_k(\theta) L^{(k-1)/m} \quad (5)$$

where K is the number of data points on which the regression is based, L is the lag operator such that $L^{s/m} x_t^{(m)} = x_{t-s/m}^{(m)}$, and $b_K(\cdot)$ is the weight function that can

take various shapes. For example, as in [34], a two-parameter exponential Almon lag polynomial can be implemented such as $\theta = (\theta_1, \theta_2)$,

$$b_k(\theta) = b_k(\theta_1, \theta_2) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{k=1}^K \exp(\theta_1 k + \theta_2 k^2)} \quad (6)$$

The parameter θ is part of the estimation problem. It is only influenced by the information conveyed by the last K values of the high-frequency variable ($x_t^{(m)}$), the window size K being an exogenous specification.

A useful alternative is the unrestricted specification (U-MIDAS) put forward by Foroni et al. [31] that does not consider any specific function $b_k(\cdot)$ but assume a linear relationship of the following form:

$$y_t = \beta_0 + c_0 x_t^{(m)} + c_1 x_{t-1/m}^{(m)} + \dots + c_{mK} x_{t-(K-1)/m}^{(m)} + \varepsilon_t \quad (7)$$

The advantage of the U-MIDAS specification is that it is linear and can be easily estimated by ordinary least-squares under some reasonable assumption. However, to avoid a proliferation of parameters ($2+mK$ parameters have to be estimated), m and \tilde{K} have to be relatively small. Another possibility is to impose some parameters c_j in Eq. (7) to be equal to zero. We will use this strategy in our applications (see details in Sect. 4.1).

3 Example of Macroeconomic Applications Using Massive Alternative Data

In this section, we present three examples using the methodology that we have developed in order to nowcast growth rates of macroeconomic aggregates using the flow of information coming from alternative massive data sources. Nowcasts for current-quarter growth rates are in this way updated each time new data are published. It turns out that those macroeconomic nowcasts have the great advantage of being available well ahead of the publication of official data, sometimes several months, while being extremely reliable. In some countries, where official statistical systems are weak, such macroeconomic nowcasts can efficiently complement the standard macroeconomic indicators to monitor economic activity.

3.1 A Real-Time Proxy for Exports and Imports

3.1.1 International Trade

There are three various modes of transportation for international trade: ocean, air, and land. Each mode of transportation possesses its own advantages and drawbacks based on services, delivery schedules, costs, and inventory levels. According to *Transport and Logistics of France*, maritime market represents about 90% of the world market of imports and exports of raw materials, with a total of more than 10 million tonnes of goods traded per year, according to UNCTAD [52]. Indeed, maritime transport remains the cheapest way to carry raw materials and products. For example, raw materials in the energy sector dominate shipments by sea with 45% of total shipments. They are followed by those in the metal industry, which represents 25% in total, then by agriculture, which accounts for 13%.

Other productions such as textiles, machines, or vehicles represent only 3% of the sea transport but constitute around 50% of the value of raw materials transported because of their high value. Depending on their nature, raw materials are transported on cargo ships or tankers. Indeed, we generally refer to four main types of vessels: fishing vessels, cargo (dry cargo), tankers (liquid cargo), and offshore vessels (urgent parts and small parcels). In our study, we will only focus on the cargo ships and tankers as they represent the largest part of the volume traded by sea.

In the following of this section, we develop the methodology to analyze the ship movements and to create a proxy of imports and exports for various countries and commodities.

3.1.2 Localization Data

We get our data from the automatic identification system (AIS), the primary method of collision avoidance for water transport. AIS integrates a standardized VHF transceiver with a positioning system, such as GPS receiver, as well as other electronic navigation sensors, such as a gyro-compass. Vessels fitted with AIS transceivers can be tracked by AIS base stations located along coast lines or, when out of range of terrestrial networks, through a growing number of satellites that are fitted with special AIS receivers which are capable of de-conflicting a large number of signatures. In this respect, we are able to track more than 70,000 ships with a daily update since 2010.

3.1.3 QuantCube International Trade Index: The Case of China

The QuantCube International Trade Index that we have developed tracks the evolution of official external trade numbers in real time by analyzing shipping data from ports located all over the world and taking into account the characteristics of the ships. As an example, we will focus here on international trade exchanges

of China, but the methodology of the international trade index can be extended to various countries and adapted for specific commodities (crude oil, coal, and iron ore).

First of all, we carry out an analysis of variance of Chinese official exports and imports by products (see Trade Map, monthly data 2005–2019). It turns out that (1) “electrical machinery and equipment” and “machinery” mainly explain the variance of Chinese exports and (2) “mineral fuels, oils, and products”, “electrical machinery and equipment,” and “commodities” mainly explain the variance of Chinese imports.

As those products are transported by ships, we count the number of various ships arriving in all Chinese ports. In fact, we are interested in three various types of ships: (1) bulk cargo ships that transport commodities, (2) container cargo ships transporting electrical machinery as well as equipment and machinery, and (3) tankers transporting petroleum products. For example, the total number of container cargo ships arriving in Chinese ports for each day, from July 2012 to July 2019, is presented in Fig. 1. Similar daily series are available for bulk cargo ships and tankers.

In order to avoid too much volatility present in the daily data, we compute the rolling average over 30 days of the daily arrivals of the three selected types of ships in all the Chinese ports, such as:

$$Ship_{(i,j)}^{q3}(t) = \frac{1}{30} \sum_{m=1}^{30} X_{i,j}(t - m) \tag{8}$$

with $X_{i,j}$ the number of ship arrivals of type i [container cargo, tanker, bulk cargo] in a given Chinese port j .

Finally, we compute our final QuantCube International Trade Index for China using Eq. (8) by summing up the three types of shipping and by computing its year-over-year changes. This index is presented in Fig. 2. We get a correlation of 80%

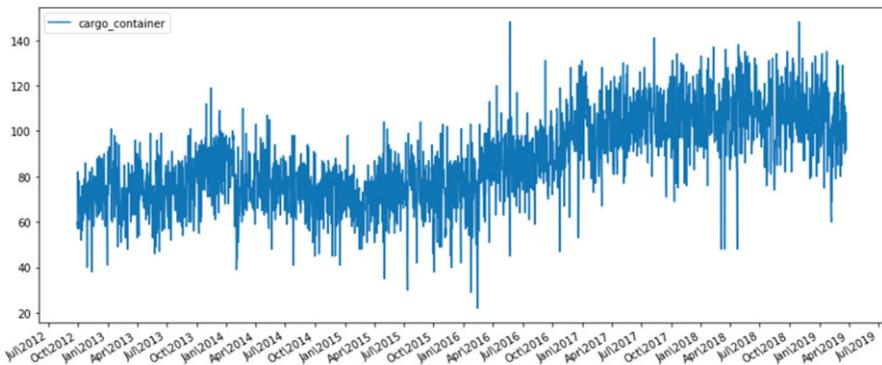


Fig. 1 Sum of cargo container arrivals in all Chinese ports



Fig. 2 China global trade index (year-over-year growth in %)

between the real-time QuantCube International Trade Index and Chinese official trade numbers (imports + exports). It is a 2-month leading index as the official numbers of imports and exports of goods are published with a delay of 2 months after the end of the reference month. We notice that our indicator clearly shows the slowing pace of total Chinese trade, mainly impacted by the increasing number of US trade sanctions since mid-2018.

For countries depending strongly on maritime exchanges, this index can reach a correlation with total country external trade numbers up to 95%. For countries relying mostly on terrestrial exchanges, it turns out that the index is still a good proxy of overseas exchanges. However, in this latter case, proxies of aerial and land exchanges can be computed to complement the information, by using cargo flights, tolls, and train schedule information.

3.2 A Real-Time Proxy for Consumption

3.2.1 Private Consumption

When tracking economic activity, private consumption is a key macroeconomic aggregate that we need to evaluate in real time. For example, in the USA, private consumption represents around 70% of the GDP growth. As official numbers of private consumption are available on a monthly basis (e.g., in the USA) or a quarterly basis (e.g., in China) and with delays of publication ranging from 1

to 3 months, alternative data sources, such as Google Trends, can convey useful information when official information is lacking.

As personal expenditures fall under the durable goods, non-durable goods, and services, we first carry out a variance analysis of the consumption for the studied countries, to highlight the key components of the consumption we have to track. For example, for the Chinese consumption, we have identified the following categories : Luxury (bags, watches, wine, jewelry), Retail sales (food, beverage, clothes, tobacco, smartphones, PC, electronics), Vehicles, Services (hotel, credit loan, transportation), and Leisure (tourism, sport, cinema, gaming). In this section, we focus on one sub-indicator of the QuantCube Chinese consumption proxy, namely, *Tourism* (Leisure category). The same methodology is developed to track the other main components of household consumption.

3.2.2 Alternative Data Sources

The touristic sub-component of the Chinese consumption is supposed to track the spending of the Chinese population for tourist trips, inside and outside the country. To create this sub-component of the consumption index, we have used touristic-related search queries retrieved by means of the Google Trends and Baidu applications. In fact, Internet search queries available through Google Trends and Baidu allow us to build a proxy of the private consumption for tourist trips per country as search queries done by tourists reflect the trends of their traveling preferences but also a prediction of their future travel destination. Google and Baidu Trends have search trends features that show how frequently a given search term is entered into Google's or Baidu's search engine relative to the site's total search volume over a given period of time. From these search queries, we built two different indexes: the tourist number per destination using the region filter "All country" and the tourist number from a specific country per destination by selecting the country in the region filter.

3.2.3 QuantCube Chinese Tourism Index

The QuantCube Chinese Tourism Index is a proxy of the tourist number from China per destination. To create this index, we first identified the 15 most visited countries by Chinese tourists that represent 60% of the total volume of Chinese tourists. We create a Chinese tourism index per country by identifying the relevant categories based on various aspects of trip planning, including transportation, touristic activities, weather, lodging, and shopping. As an example, to create our Chinese tourism index in South Korea, we identified the following relevant categories: Korea Tourism, South Korea Visa, South Korea Maps, Korea Tourism Map, South Korea Attractions, Seoul Airport, and South Korea Shopping (Fig. 3).

Finally, by summing the search query trends of those identified keywords, our Chinese Tourism Index in South Korea tracks in real time the evolution of official

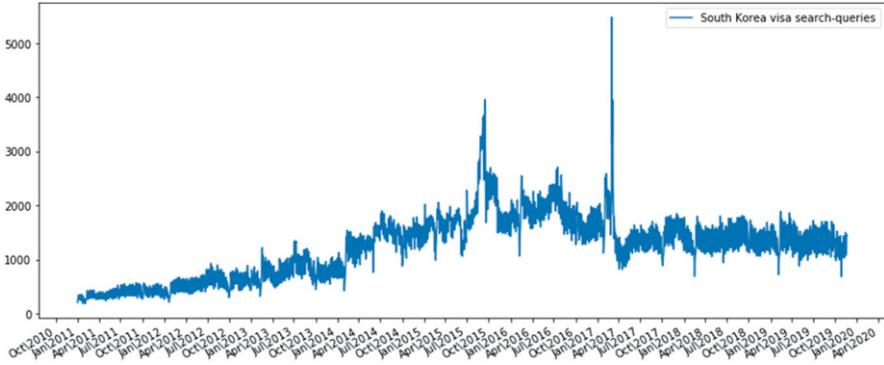


Fig. 3 Baidu: “South Korea Visa” search queries

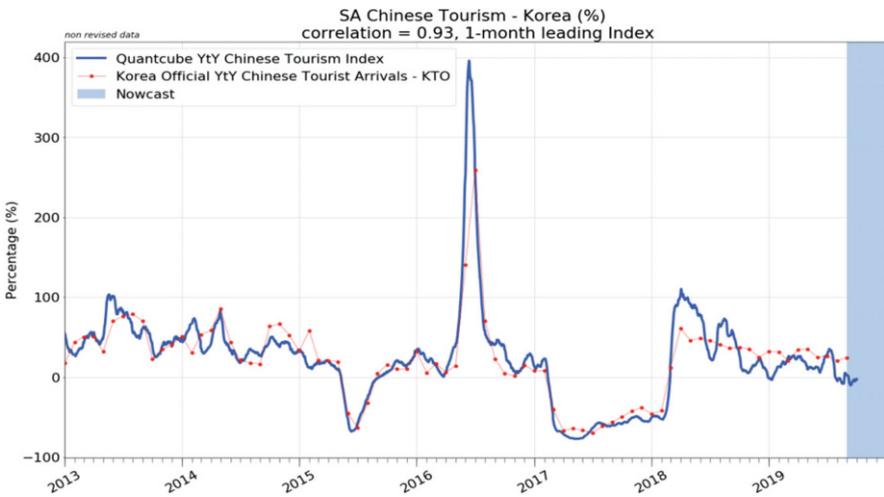


Fig. 4 South Korea Chinese tourism index short term (year-over-year in %)

Chinese tourist arrivals in Korea. We calculate the year-over-year variation of this index and validate it using official numbers of Chinese tourists in South Korea (see Fig. 4).

From Fig. 4, we observe that the QuantCube Chinese Tourism Index correctly tracks the arrivals of Chinese tourists in South Korea. As, for example, the index caught the first drop in June 2015 due to the MERS outbreak. Furthermore, in 2017, after the announcement of a future installation of the powerful Terminal High Altitude Area Defense (THAAD) system at the end of 2016, the Chinese government banned the tour groups to South Korea as economic retaliation. For 2017 as a whole, South Korea had 4.2 million Chinese visitors—down 48.3% from the previous year. The decrease in Chinese tourists leads to a 36% drop in tourist

entries. Therefore, this real-time Chinese Tourism indicator is also useful to estimate in real time the trend of the South Korea Tourism industry.

It tracks Chinese tourist arrivals with a correlation up to 95%.

Finally, we developed similar indexes to track in real time the arrivals of Chinese tourists in the most 15 visited countries (USA, Europe, etc.); we get an average correlation of 80% for the most visited countries. By aggregating those indexes we are able to construct an index for tracking the arrival of Chinese tourists over the world that provides a nice proxy of Chinese households' consumption in this specific sector.

3.3 A Real-Time Proxy for Activity Level

QuantCube Technology has developed a methodology based on the analytic of the satellite Sentinel-2 images to detect new infrastructures (commercial, logistics, industrial, or residential) and measure the evolution of the shape and the size of urban areas. But, the level of activities or exploitation of these sites is hardly determined by building inspection and could be inferred from vehicle presence from nearby streets and parking lots. For this purpose, QuantCube Technology in partnership with IRISA developed a deep learning model for vehicles counting from satellite images coming from the Pleiades sensor at 50-cm spatial resolution. In fact, we select the satellite depending on the pixel resolution needed per application.

3.3.1 Satellite Images

Satellite imagery has become more and more accessible in recent years. In particular, some public satellites provide an easy and cost-free access to their image archives, with a spatial resolution high enough for many applications concerning land characterization. For example, the ESA (European Space Agency) satellite family Sentinel-2, launched on June 23, 2015, provides 10-meter resolution multi-spectral images covering the entire world. We analyze those images for infrastructure detection. To detect and count cars we use higher-resolution VHR (very high resolution) images acquired by the Pleiades satellite (PHR-1A and PHR-1B), launched by the French Space Agency (CNES), Distribution Airbus DS. These images are pan-sharpened products obtained by the fusion of 50-cm panchromatic data (70 cm at nadir, resampled at 50 cm) and 2-m multispectral images (visible RGB (red, green, blue) and infrared bands). They cover a large region of heterogeneous environments including rural, forest, residential, as well as industrial areas, where the appearance of vehicles is influenced by shadow and occlusion effects. On the one hand, one of the advantages of satellite imaging-based applications is their natural world scalability. And on the other hand, the evolution and improvements of artificial intelligence algorithms enable us to process the huge

amount of information contained in satellite images in a straightforward way, giving a standardized and automatic solution working on real-time data.

3.3.2 Pre-processing and Modeling

Vehicle detection from satellite images is a particular case of object detection as objects are uniform and very small (around 5×8 pixels/vehicle in Pleiades images) and do not overlap. To tackle this task, we use a model called 100 layers Tiramisu (see [40]). It is a quite economical model since it has only 9 million parameters, compared to around 130 million for early deep learning network referred to as *VGG19*. The goal of the model is to exploit the feature reuse by extending DenseNet architecture while avoiding the feature explosion. To train our deep learning model, we created a training dataset using an interactive labeling tool that enables to label 60% of vehicles with one click using flood-fill methods adapted for this application. The created dataset contains 87,000 annotated vehicles from different environments depending on the level of urbanization. From the segmentation, an estimation of the number of vehicles is computed based on the size and the shape of the predictions.

3.3.3 QuantCube Activity Level Index

Finally, the model realizes a satisfying prediction for vehicle detection and counting application since the precision reached more than 85% on a validation set of 2673 vehicles belonging to urban and industrial zones. The algorithm is currently able to deal with different urban environments. As can be seen in Fig. 5, showing a view of Orly area near Paris including the prediction for the detection and counting of cars in yellow, the code is able to accurately count vehicles in some identified areas.

The example in Fig. 5 shows the number of vehicles for every identified bounding box corresponding to parking of hospitality, commercial, or logistics sites. Starting

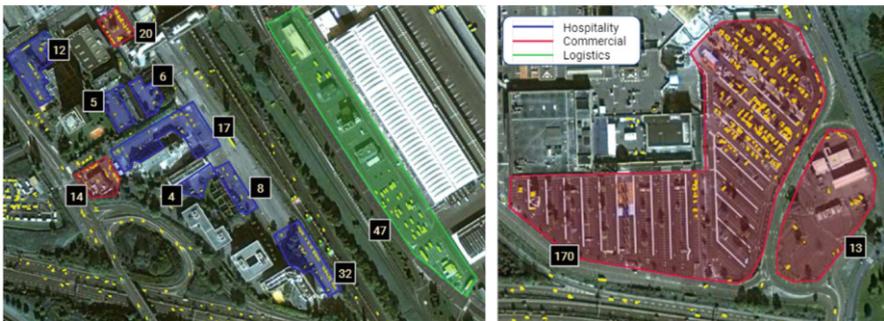


Fig. 5 Number of vehicles per zone in Orly

from this satellite-based information, we are able to compute an index to detect and count vehicles in identified sites and to track their level of activities or exploitation looking at the evolution of the index which is correlated to sales indexes. Using satellite images, it enables to create a methodology and normalized measures of the level of activities that enable financial institutions and corporate groups to anticipate new investment trends before the release of official economics numbers.

4 High-Frequency GDP Nowcasting

When dealing with the most important macroeconomic aggregate, that is, GDP, our approach relies on the expenditure approach that computes GDP by evaluating the sum of all goods and services purchased in the economy. That is, we decompose GDP into its main components, namely, consumption (C), investment (I), government spending (G), and net exports (X-M) such as:

$$GDP = C + I + G + (X - M) \quad (9)$$

Starting from this previous decomposition, our idea is to provide high-frequency nowcasts for each GDP component given in Eq. (9), using the indexes based on alternative data that we previously computed. However, depending on the country considered, we do not necessarily cover all the GDP components with our indexes. Thus, the approach that we developed within QuantCube consists in mixing in-house indexes based on alternative data and official data stemming from opinion surveys, production, or consumption. This is a way to have a high-frequency index that covers a large variety of economic activities. In this section, we present results that we get on the two largest economies in the world, that is, the USA and China.

4.1 Nowcasting US GDP

The US economy ranks as the largest economy by nominal GDP; they are the world's most technologically powerful economy and the world's largest importer and second largest exporter. In spite of some already existing nowcasting tools on the market, provided by the Atlanta Fed and the New York Fed, it seems useful to us to develop a US GDP nowcast on a daily basis.

To nowcast US GDP, we are mixing official information on household consumption (personal consumption expenditures) and on consumer sentiment (University of Michigan) with in-house indexes based on alternative data. In this respect, we use the *QuantCube International Trade Index* and *QuantCube Crude Oil Index*, developed using the methodology presented in the Sect. 3.1 of this chapter, and the *QuantCube Job Opening Index* that is a proxy of the job market and nonfarm payroll created by aggregating job offers per sector. The two official variables that

we use are published with 1-month delay and are available on a monthly frequency. However, the three QuantCube indexes are daily and are available in real time without any publication lags.

Daily US GDP nowcasts are computed using the U-MIDAS model given in Eq. (7) by imposing some constraints. Indeed we assume that only the latest values of the indexes enter the U-MIDAS equation. As those values are the averages of the last 30 days, we thus account for the recent dynamics by imposing MIDAS weights to be uniform. The US QuantCube Economic Growth Index aiming at tracking year-over-year changes in US GDP is presented in Fig. 6. We clearly see that this index is able to efficiently track US GDP growth, especially as regards peaks and troughs in the cycles. For example, focusing on the year 2016, we observe that the index anticipates the slowing pace of the US economy for this specific year which was the worst year in terms of GDP growth since 2011, at 1.6% annually. The lowest point of index was reached on October 12, 2016, giving a leading signal of a decelerating fourth quarter in 2016. As a matter of fact, the US economy lost momentum in the final 3 months of 2016.

Then, the indicator managed to catch the strong economic trend in 2017 (+2.3% annually, acceleration from the 1.6% logged in 2016). It even reflected the unexpected slowdown in the fourth quarter of 2017 two months in advance, because of surging imports, a component that is tracked in real time. Focusing on the recent Covid-19 crisis, official US GDP data shows a decline to a value slightly above zero in year-over-year growth for 2020q1, while our index reflects a large drop in subsequent months close to -6% on July 2, 2020, indicating a very negative growth in 2020q2. As regards the US economy, the Atlanta Fed and New York Fed release on a regular basis estimates of current and future quarter-over-quarter GDP

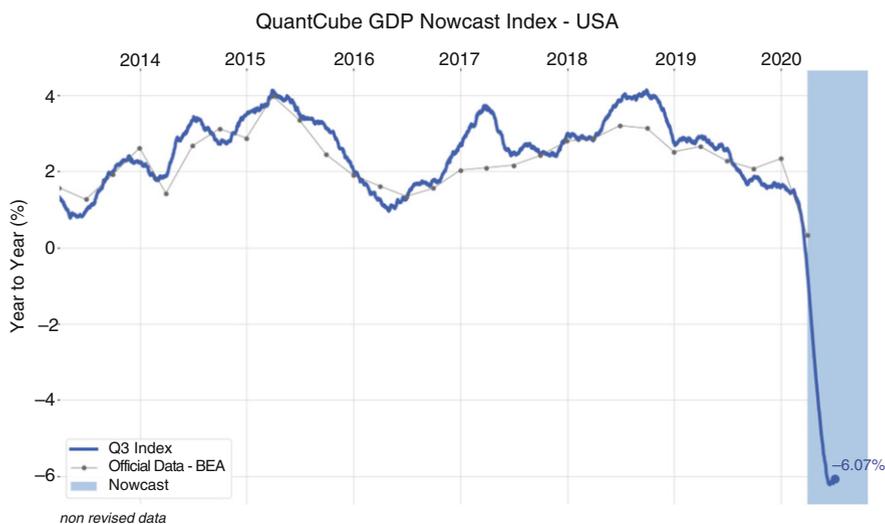


Fig. 6 US economic growth index

growth rate, expressed in annualized terms. Surprisingly, as of March 25, 2020, the nowcast of Atlanta Fed for 2020q1 was at 3.1%, and the one of New York Fed was a bit lower at 1.49% as of March 20, 2020, but still quite high. How is that? In fact, all those nowcasting tools have been extremely well built, but they only integrate official information, such as production, sales, and surveys, that is released by official sources with a lag. Some price variables, such as stock prices that are reacting more rapidly to news, are also included in the nowcasting tools, but they do not strongly contribute to the indicator. So how can we improve nowcasting tools to reflect high-frequency evolutions of economic activity, especially in times of crises? A solution is to investigate alternative data that are available on a high-frequency basis as we do with our indicator. It turns out that at the same date, our US nowcast for 2020q1 was close to zero in year-over-year terms, consistent with a quarter-over-quarter GDP growth of about -6.0% in annualized terms, perfectly in line with official figures from the BEA. This real-time economic growth indicator appears as a useful proxy to estimate in real time the state of the US economy.

4.2 Nowcasting Chinese GDP

China ranks as the second largest economy in the world by nominal GDP. It has the world's fastest growing major economy, with a growth rate of 6% in average over 30 years. It is the world's largest manufacturing economy and exporter of goods and the world's largest fastest growing consumer market and second largest importer of goods.

Yet, despite the importance for the world economy and the region, there are few studies on nowcasting Chinese economic activity (see [27]). Official GDP data are available only with a 2-month lag and are subject to several revisions.

To nowcast the Chinese GDP in real time, we use the *QuantCube International Trade Index* and the *QuantCube Commodity Trade Index* developed in the Sect. 3.1 of this chapter; the *QuantCube Job Opening Index*, a proxy of the job market created by aggregating the job offers per sector; and the *QuantCube Consumption Index* developed in Sect. 3.2. It turns out that all the variables have been developed in-house based on alternative massive datasets and are thus available on a daily frequency without any publication lags.

Daily GDP nowcasts are computed using the U-MIDAS model given in Eq. (7) by imposing the same constraints as for the USA (see previous sub-section). The China Economic Growth Index, aiming at tracking China GDP year-over-year growth, is presented in Fig. 7. First of all, we observe that our index is much more volatile than official Chinese GDP, which seems more consistent with expectations about fluctuations in GDP growth. Our measure thus reveals a bias, but it is not systematic. In fact, most of the time the *true* Chinese growth is likely to be lower than the official GDP, but for some periods of time, the estimated GDP can also be higher as, for example, in 2016–2017. The Chinese GDP index captured the deceleration of the Chinese economy from the middle of 2011. The index showed a sharp drop in Q2

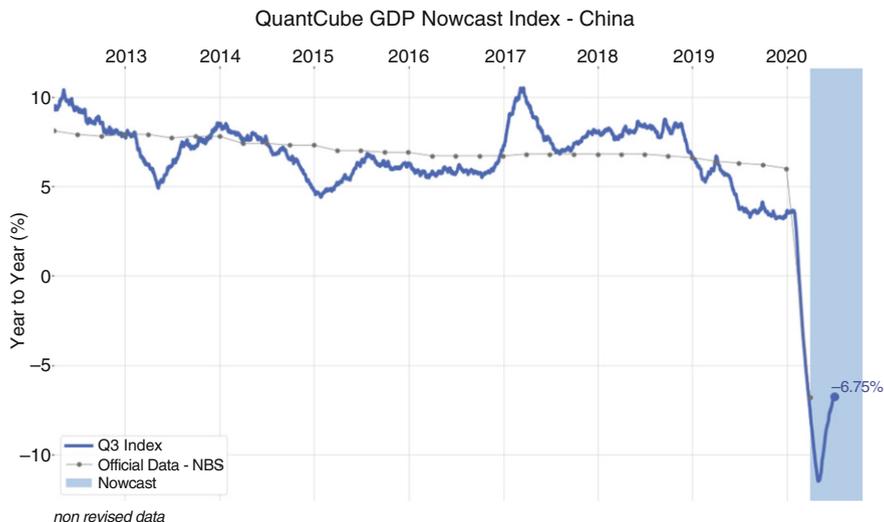


Fig. 7 China Economic Growth Index

2013, when according to several analysts, the Chinese economy actually shrank. The indicator shows the onset of the deceleration period beginning in 2014, in line with the drop in oil and commodity prices. According to our index, the Chinese economy is currently experiencing a deceleration that started beginning of 2017. This deceleration is not as smooth as in the official data disclosed by the Chinese government. In particular, a marked drop occurred in Q2 2018, amid escalating trade tensions with the USA. The year 2019 begins with a sharp drop of the index, showing that the China economy still did not reach a steady growth period. As regards the recent Covid-19 episode, QuantCube GDP Nowcast Index for China shows a sharp year-over-year decline starting at the end of January 2020 from 3.0% to a low of about -11.5% beginning of May 2020, ending at -6.7% on July 2, 2020. This drop is larger than the official data from the National Bureau of Statistics which stated a negative yearly GDP growth of -6.8% in 2020q1. Overall, this indicator is a unique valuable source of information about the state of the economy since very few economic numbers are released in China.

5 Applications in Finance

There is a long recognized intricate relationship between real macroeconomy and financial markets. Among the various academic works, Engel et al. [25] show evidence of predictive power of inflation and output gap on foreign exchange rates, while Cooper and Priestly [22] show that the output gap is a strong predictor of US government bond returns. Such studies are not limited to the fixed income market.

As early as 1967, Brown and Ball [15] show that a large portion of the variation in firm-level earnings is explained by contemporaneous macroeconomic conditions. Rangvid [46] also shows that the ratio of share price to GDP is a good predictor of stock market returns in the USA and international developed countries.

However, economic and financial market data have a substantial mismatch in the observation frequency. This presents a major challenge to analyzing the predictive power of economic data on financial asset returns, given the low signal-to-noise ratio embedded in financial assets. With the increasing accessibility of high-frequency data and computing power, real-time, high-frequency economic forecasts have become more widely available. The Federal Reserve Bank of Atlanta and New York produce nowcasting models of US GDP figures which are available at least on a weekly basis and are closely followed by the media and the financial markets. Various market participants have also developed their own economic nowcasting models. As previously pointed out in this chapter, QuantCube produces US GDP nowcasts available at daily frequency. A number of asset management firms and investment banks have also made their GDP nowcasts public. Together these publicly available and proprietary nowcast information are commonly used by discretionary portfolio managers and traders to assess investment prospects. For instance, Blackrock [39] uses their recession probability models for macroeconomic regime detection, in order to inform asset allocation decisions. Putnam Investments [6] uses global and country GDP nowcasts as key signals in their interest rates and foreign exchange strategies. While the investment industry has embraced nowcasting as an important tool in the decision-making process, evaluating the effectiveness of real-time, high-frequency economic nowcasts on financial market returns is not without its own challenges. Most of the economic nowcasts have short history and evolving methodology. Take the two publicly available US GDP nowcasts mentioned above as examples. The Atlanta Fed GDPNow was first released in 2014 and introduced a methodology change in 2017, whereas the NY Fed GDP nowcast was first released in 2016. Although longer in-sample historical time series are available, the out-of-sample historical periods would be considered relatively short by financial data standards. As a result, the literature evaluating the out-of-sample predictive power of nowcasting models is relatively sparse. Most studies have used point-in-time data to reconstruct historical economic nowcasts for backtesting purposes. We survey some of the available literature below.

Blin et al. [12] used nowcasts for timing alternative risk premia (ARP) which are investment strategies providing systematic exposures to risk factors such as value, momentum, and carry across asset classes. They showed that macroeconomic regimes based on nowcast indicators are effective in predicting ARP returns. Molodtsova and Papell [43] use real-time forecasts of Taylor rule model and show outperformance over random walk models on exchange rates during certain time periods. Carabias [18] shows that using macroeconomic nowcasts is a leading indicator of firm-level end-of-quarter realized earnings, which translates into risk-adjusted returns around earnings announcements. Beber et al. [11] developed latent factors representing economic growth and its dispersion, which together explain almost one third of the implied stock return volatility index (VIX). The results are

encouraging since modeling stock market volatility is of paramount importance for financial risk management, but historically, financial economists have struggled to identify the relationship between the macroeconomy and the stock market volatility [26]. More recently, Gu et al. [38] have shown that machine learning approaches, based on neural networks and trees, lead to a significant gain for investors, basically doubling standard approaches based on linear regressions. Obviously, more research is needed about the high-frequency relationship between macroeconomic aggregates and financial assets, but this line of research looks promising.

6 Conclusions

The methodology reported in this chapter highlights the use of large and alternative datasets to estimate the current situation in systemic countries such as China and the USA. We show that massive alternative datasets are able to account for real-time information available worldwide on a daily frequency (AIS position, flight traffic, hotel prices, satellite images, etc.). By correctly handling those data, we can create worldwide indicators calculated in a systematic way. In countries where the statistical system is weak or non-credible, we can thus rely more on alternative data sources than on official ones. In addition, the recent Covid-19 episode highlights the gain in timeliness from using alternative datasets for nowcasting macroeconomic aggregates, in comparison with standard official information. When large shifts in GDP occur, generating thus a large amount of uncertainty, it turns out that alternative data are an efficient way to assess economic conditions in real time. The challenge for practitioners is to be able to deal with massive non-structured datasets, often affected by noise, outliers, and seasonal patterns . . . and to extract the pertinent and accurate information.

References

1. Aastveit, K. A., Albuquerque, B., & Anundsen, A. K. (2020). Changing supply elasticities and regional housing booms. Bank of England Working Paper No. 844. <https://www.bankofengland.co.uk/working-paper/2020/changing-supply-elasticities-and-regional-housing-booms>.
2. Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L., & Ruenstler, G. (2011). Short-term forecasts of euro area GDP growth. *Economic Journal*, 14, C25–C44.
3. Aprigliano, V., Ardizzi, G. & Monteforte, L. (2019). Using the payment system data to forecast the economic activity. *International Journal of Central Banking* 15, 4.
4. Arslanalp, S., Marini, M., & Tumbarello, P. (2019). *Big data on vessel traffic: Nowcasting trade flows in real time*. IMF Working Paper No. 19/275.
5. Askitas, N., & Zimmermann, K. (2013). Nowcasting business cycles using toll data. *Journal of Forecasting*, 32(4), 299–306.

6. Atkin, M., Chan, A., Embre, O., Hornder, S., Solyanik, I., & Yildiz, I. (2020). *Waiting for the next wave of growth*. Macro Report, Putnam Investments. <https://www.putnam.com/institutional/content/macroReports/434-waiting-for-the-next-wave-of-growth>.
7. Baffigi, A., Golinelli, R., & Parigi, G. (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting*, 20(3), 447460.
8. Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593–1636. <https://doi.org/10.1093/qje/qjw024>.
9. Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-casting and the real-time data flow. In *Handbook of Economic Forecasting*, vol. 2. Part A (pp. 195–237). Amsterdam: Elsevier.
10. Barhoumi, K., Darne, O., & Ferrara, L. (2013). Testing the number of factors: An empirical assessment for a forecasting purpose. *Oxford Bulletin of Economics and Statistics*, 75(1), 64–79.
11. Beber, A., Brandt, M., & Luizi, M. (2015). Distilling the macroeconomic news flow. *Journal of Financial Economics*, 117, 489–507.
12. Blin, O., Ielpo, F., Lee, J., & Teiletche, J. (2020). Alternative risk premia timing: A point-in-time macro, sentiment, valuation analysis. *Forthcoming in Journal of Systematic Investing*. <http://dx.doi.org/10.2139/ssrn.3247010>.
13. Bragoli, D. (2017). Nowcasting the Japanese economy. *International Journal of Forecasting*, 33(2), 390–402.
14. Bragoli, D., Metelli, L., & Modugno, M. (2015). The importance of updating: Evidence from a Brazilian nowcasting model. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 1, 5–22.
15. Brown, P., & Ball, R. (1967). Some preliminary findings on the association between the earnings of a firm, its industry and the economy. *Journal of Accounting Research*, 5, 55–77.
16. Buono, D., Kapetanios, G., Marcellino, M., Mazzi, G. L., & Papailias, F. (2018). *Big data econometrics: Nowcasting and early estimates*. Italy: Universita Bocconi. Technical Report 82, Working Paper Series.
17. Caldara, D., & Iacovello, M. (2019). Measuring geopolitical risk. In *Federal Reserve System, International Finance Discussion Papers, number 1222*. <https://www.federalreserve.gov/econres/ifdp/files/ifdp1222.pdf>.
18. Carabias, J. M. (2018). The real-time information content of macroeconomic news: Implications for firm-level earnings expectations. *Review of Accounting Studies*, 23, 136–166.
19. Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88(s1), 2–9. <https://doi.org/10.1111/j.1475-4932.2012.00809.x>.
20. Clarck, H., Pinkovskiy, M., & Sala-i-Martin, X. (2017). China's GDP growth may be understated. In *NBER Working Paper No. 23323*.
21. Coble, D., & Pincheira, P. (2017). Nowcasting building permits with Google Trends. In *MPRA Paper 76514*. Germany: University Library of Munich.
22. Cooper, I., & Priestley, R. (2009). Time-varying risk premiums and the output gap. *Review of Financial Studies*, 22(7), 2801–2833.
23. D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting unemployment. *International Journal of Forecasting*, 33, 801–816.
24. Donaldson, D., & Storeygard, A. (2016). The view from above: Applications from satellite data in economics. *Journal of Economic Perspectives*, 30(4), 171–198.
25. Engel, C., Mark, N. C., & West, K. D. (2008). Exchange rate models are not as bad as you think. In *NBER Macroeconomics Annual 2007* (pp. 381–441). Chicago: University of Chicago.
26. Engle, R. F., & Rangel, J. G. (2008). The Spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, 21, 1187–1222.
27. Fernald, J., Hsu, E., & Spiegel, M. (2019). Is China fudging its GDP figures? In *Evidence from Trading Partner Data*. FRB San Francisco Working Paper 2019–19.
28. Ferrara, L., & Marsilli, C. (2018). Nowcasting global economic growth: A factor-augmented mixed-frequency approach. *The World Economy*, 42(3), 846–875.

29. Ferrara, L. & Simoni, A. (2019). When are Google data useful to nowcast GDP? In *An Approach via Pre-selection and Shrinkage*. Banque de France Working Paper No. 717.
30. Forni, M., Gambetti, L., Lippi, M., & Sala, L. (2017). Noisy News in business cycles. *American Economic Journal: Macroeconomics*, 9(4): 122–52. <https://doi.org/10.1257/mac.20150359>.
31. Foroni, C., Marcellino, M., & Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society A*, 178, 57–82. <https://doi.org/10.1111/rssa.12043>.
32. Fraiberger, S., Lee, D., Puy, D., & Ranciere, R. (2018). *Media sentiment and international asset prices*. IMF Working Paper No. WP/18/274.
33. Galbraith, J. W., & Tkacz, G. (2018). Nowcasting with payments system data. *International Journal of Forecasting*, 34(2), 366376.
34. Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1), 53–90. <https://doi.org/10.1080/07474930600972467>.
35. Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4), 665–676.
36. Giannone, D., Lenza, M., & Primiceri, G. E. (2017). Economic predictions with big data: The Illusion of Sparsity. In *Centre for Economic Policy Research*. Discussion Paper No. DP12256. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=12256.
37. Glaeser E., Kim, H., & Luca, M. (2017). *Nowcasting the local economy: Using Yelp data to measure economic activity*. NBER Working Paper No. 24010.
38. Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223–2273.
39. Hildebrand, P., Boivin, J., & Bartsch, E. (2020). On the risk of regime shifts. In *Macro and Market Perspectives*. New York: Blackrock Investment Institute. <https://www.blackrock.com/americas-offshore/en/insights/blackrock-investment-institute/global-macro-outlook>.
40. Jégou, S., Drozdal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017). The one hundred layers Tiramisu: Fully convolutional DenseNets for semantic segmentation. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA* (pp. 1175–1183). <https://doi.org/10.1109/CVPRW.2017.156>.
41. Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2020). Making text count: Economic forecasting using newspaper text. In *Bank of England*. Working paper No. 865. <http://dx.doi.org/10.2139/ssrn.3610770>.
42. Modugno, M., Soybilgen, B., & Yazgan, E. (2016). Nowcasting Turkish GDP and news decomposition. *International Journal of Forecasting*, 32(4), 1369–1384.
43. Molodtsova, T., & Papell, D. (2012). Taylor rule exchange rate forecasting during the financial crisis. In *National Bureau of Economic Research (NBER)*. Working paper 18330. <https://doi.org/10.3386/w18330>.
44. Narita, F., & Yin, R. (2018). In search for information: Use of Google Trends' data to narrow information gaps for low-income developing countries. Technical Report WP/18/286. IMF Working Paper.
45. Nymand-Andersen, P., & Pantelidis, E. (2018). Google econometrics: Nowcasting Euro area car sales and big data quality requirements. In *European Central Bank*. Statistics Paper Series, No. 30. <https://www.ecb.europa.eu/pub/pdf/scpsps/ecb.sps30.en.pdf?21f8c889572bc4448f92acbfe4d4d86af>.
46. Rangvid, J. (2006). Output and expected returns. *Journal of Financial Economics*, 81(3), 595–624.
47. Rogers, J. H., & Xu, J. (2019). How well does economic uncertainty forecast economic activity? *Finance and Economics Discussion Series* (pp. 2019–085). Washington: Board of Governors of the Federal Reserve System.
48. Seabold, S., & Coppola, A. (2015). Nowcasting prices using Google Trends: An application to Latin America. Policy Research Working Paper No. 7398.
49. Smith, S., Timmermann, A., & Zhu, Y. (2019). Variable selection in panel models with breaks. *Journal of Econometrics*. 212(1), 323–344.

50. Stock, J. H., & Watson, M. W. (2012). Disentangling the channels of the 2007–2009 recession. In *National Bureau of Economic Research (NBER)*. Working Paper No. w18094. <https://www.nber.org/papers/w18094>
51. Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–88. <http://www.jstor.org/stable/2346178>.
52. UNCTAD (2016). Trade and development report 2016: Structural transformation for inclusive and sustained growth. In *United Nations Conference on Trade And Development (UNCTAD)*. https://unctad.org/system/files/official-document/tdr2016_en.pdf.
53. Varian, H. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

