

The Big Read **Massachusetts Institute of Technology**

Can big data revolutionise policymaking by governments?

Mining digital information for accurate, up-to-date economic snapshots could help officials make quicker and better decisions

Robin Wigglesworth in New York JANUARY 31 2018

When Alberto Cavallo was a child growing up in Argentina in the late 1980s, the Latin American country was suffering one of its occasional crises. [Inflation was rampant](#), making even shopping trips a hectic daily dash.

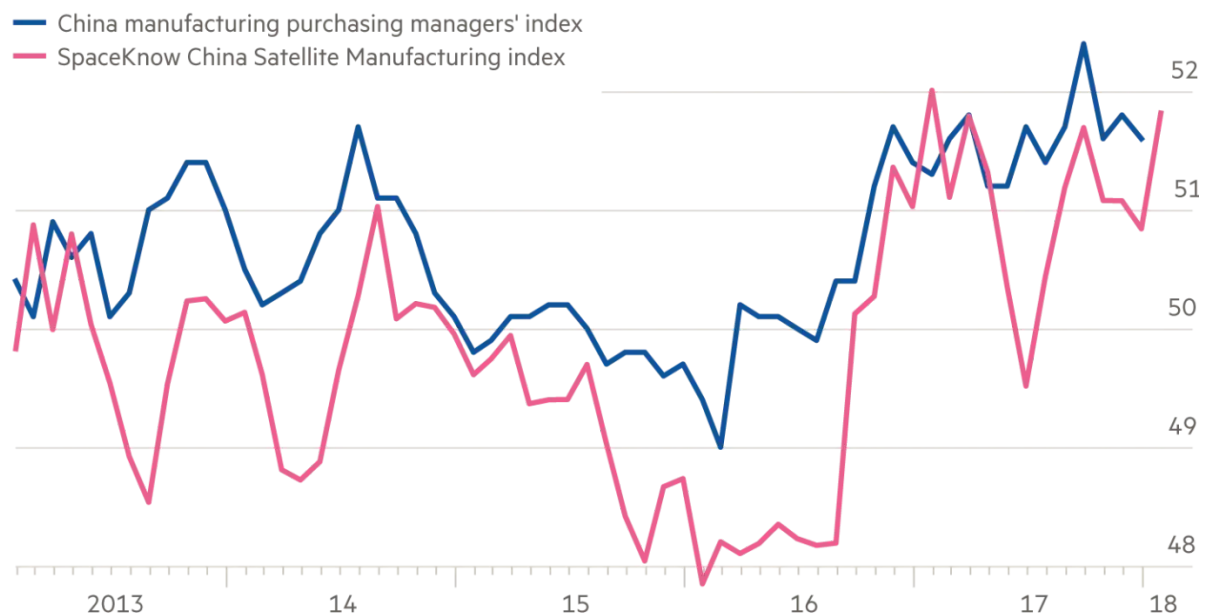
Mr Cavallo and his mother would go to the bank every day and withdraw just enough pesos for the necessary purchases, keeping the rest of their savings in dollars. They would then run to the local shop and grab what they needed as quickly as possible, hoping to get to the counter before the price list was updated again.

“If we didn’t get to the cash register in time then we had to go back to the bank and start again,” he recalls ruefully.

But the experience sowed the seeds for what has become one of the more intriguing experiments in the normally staid world of economic statistics: an attempt to harness the explosion of [“big data”](#) to enhance, complement and perhaps ultimately replace the traditional forms of data that still inform and shape the views of countless policymakers, politicians, academics and guide trillions of dollars worth of investment.

Mr Cavallo is today a professor of applied economics at MIT, where he runs the [Billion Prices Project](#) with Roberto Rigobon, another MIT professor. The project started in 2006, during a period when the then-Argentine government was accused of manipulating its inflation data. Professors Cavallo and Rigobon realised that by compiling the online prices listed by Argentine retailers they could build a more accurate and contemporaneous measure of the true inflation rate. Since a change of government in 2015-16, Argentina has published a more accurate inflation gauge.

Satellite data indicate deeper Chinese manufacturing slowdown than official statistics



Sources: China Federation of Logistics and Purchasing; SpaceKnow
© FT

The project's commercial arm, PriceStats, now collects enough data to provide daily inflation updates for 22 economies. "It was kind of an accident. But we quickly realised that it had applications elsewhere," Prof Cavallo says.

The project is just one example of a broader trend of trawling the swelling sea of big data for clues on how companies, industries [or entire economies](#) are performing. Some data are already providing useful, if imperfect, insights. But some experts forecast that the digital fingerprints of our online lives could ultimately be crunched into a real-time map of economic trends that make present-day data look as archaic as the railway freight information of the 1920s.

The trail of our digital exhaust is incomprehensibly vast. The world's annual data generation is estimated to be doubling every year, and the overall size will reach 44 zettabytes (that's trillions of gigabytes) by 2020, according to a study by [International Data Corporation](#). If all this information was placed in high-end tablet computers, the pile would reach from Earth to the moon more than six times over.



Alberto Cavallo, a professor at MIT, and colleagues have found new ways to monitor inflation using internet prices © MIT

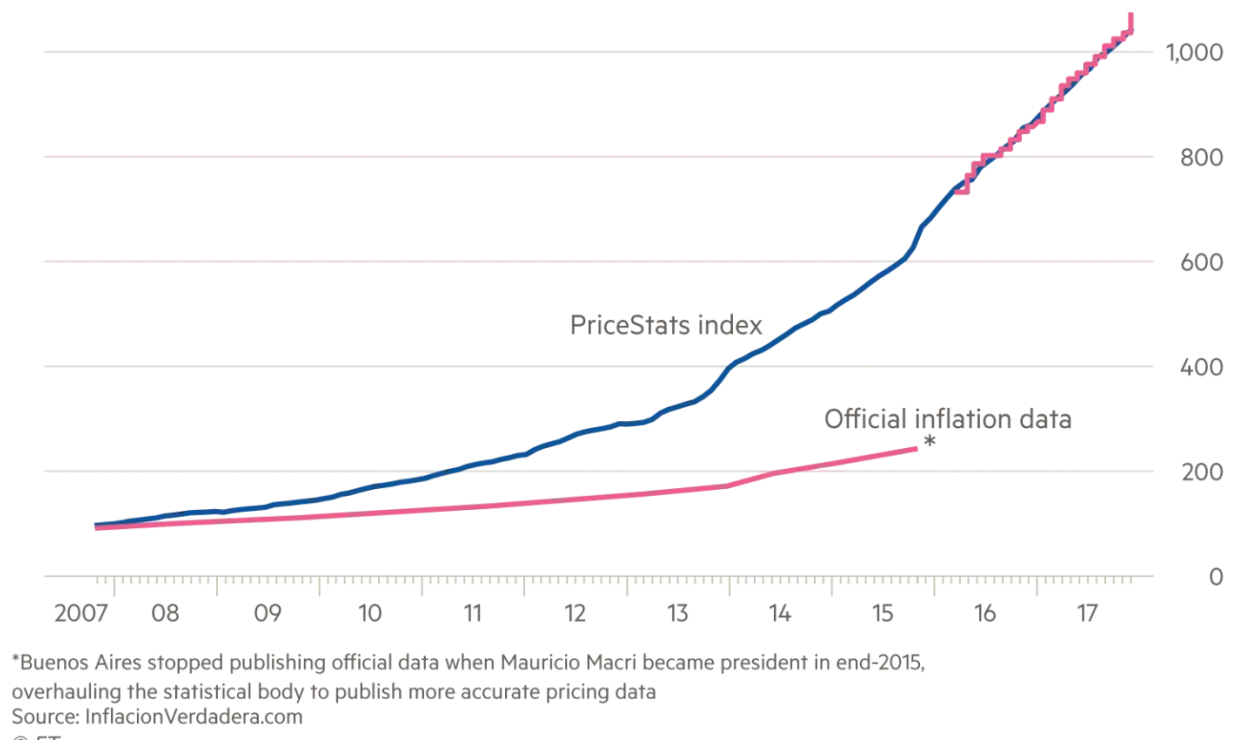
“Anything you want to know about the economy is knowable right now, if you can tap into the right data set,” says Tammer Kamel, head of Quandl, an [alternative data provider](#). “This is one of the big opportunities. These economic reports are slow but market-moving, and by lifting the right rocks you can kind of know them now.”

This may sound ambitious, given that big data can be riddled with obvious or obscure flaws and biases. But some data scientists say that as more of our lives migrate online, we might be approaching the moment when near-instantaneous economic statistics become reality.

“Marshalling all the data and putting it in the right form is not an insignificant challenge,” says Jonathan Shaw, director of a new programme at the Alan Turing Institute in London on harnessing alternative data in economic research. “(But) in 10 years’ time I imagine we will be much closer to a real-time map of the economy. If we don’t have that in a decade I’d be disappointed.”

Online prices give better measure of Argentine inflation

Indices

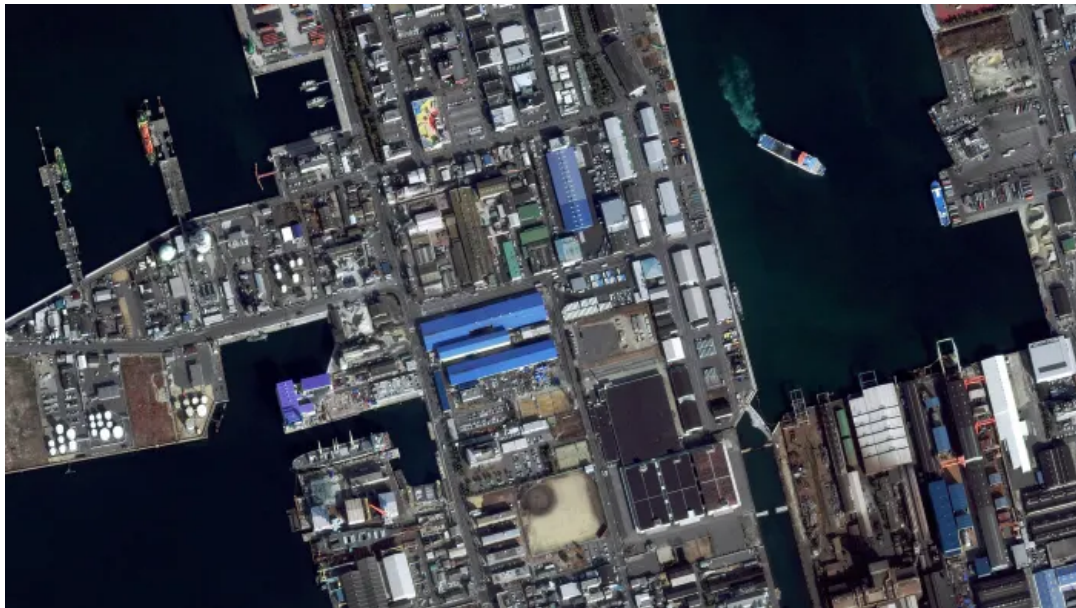


When the UK voted to leave the EU in 2016 many economists predicted a swift calamity. A survey of service sector optimism suffered its biggest drop in its 20-year history immediately after the [Brexit](#) vote, and Goldman Sachs predicted that the UK would slip into a recession. But the economy has so far proved remarkably resilient in the period before [the UK's departure](#).

Not everyone was wrongfooted. In 2015 Schroders, the UK investment group, had set up a data insights unit to help it parse reams of new digital information, including credit card data that gave it a glimpse of real-time spending patterns. Despite the pervasive sense of gloom, the data showed there had been negligible impact.

“We could tell our fund managers that things looked fine, and a few months later the official data confirmed this,” says Mark Ainsworth, head of data insights at Schroders. “All this digital data can give you more contemporaneous insight about the economy.”

The potential is dizzying. Social media feeds can be used to build real-time gauges of sentiment. Satellites in space see which ships dock where and when, whether oil tanks are full or empty, the quality of a crop or even the productivity of a blast furnace. Credit card purchases and email receipts show retail spending. Job listings from hundreds of thousands of career sites or corporate websites can reveal employment patterns. And smartphones send location data that show where we are at any given time. In time, the “[internet of things](#)” could reveal our daily eating habits through web-connected fridges.

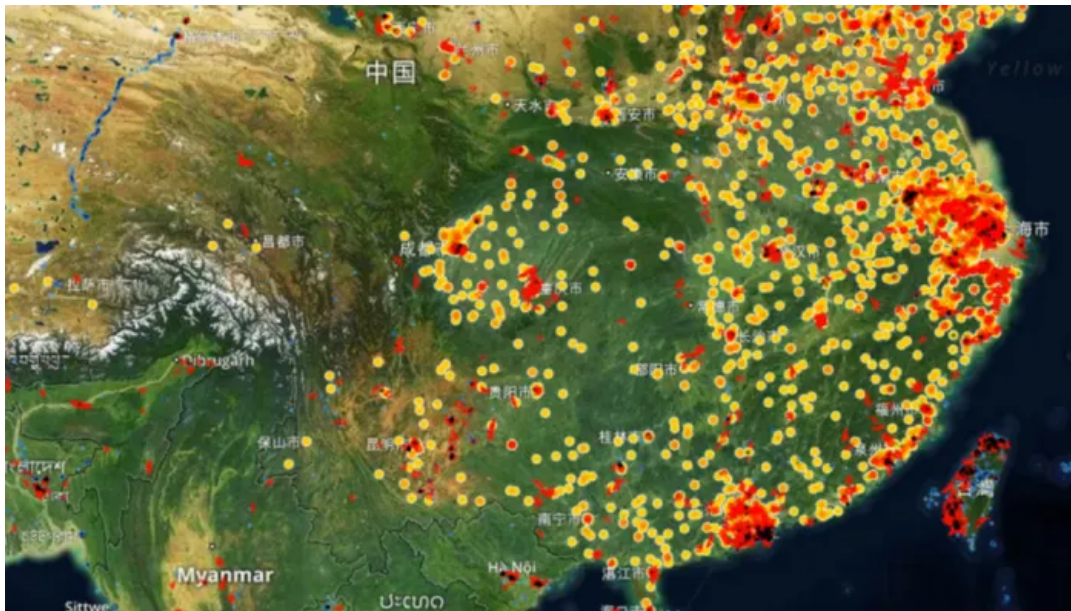


Satellite images of harbours, car parks, roads and airports can reveal much about the movement of goods and thus economic activity
© Getty

Mining these [new data sets](#) was once the preserve of sophisticated “quantitative” hedge funds. Some finance ministries, [central bankers](#) and statistics agencies are now starting to dabble in the field in order to understand the economic tides better and more swiftly — a development that could have significant public policy implications.

The financial crisis exposed major gaps in official figures. The National Bureau of Economic Research’s [business cycle dating committee](#), which is the semi-official arbiter of US economic contractions, took until December 2008 — nearly three months after Lehman Brothers went bankrupt — to declare that the US economy had actually [entered a recession](#) a year earlier. While many economists had concluded as much for some time from the rapidly souring monthly and quarterly data, the statistics did not adequately capture the pace at which the economy was tanking, recalls Diana Farrell, former deputy director of the Obama administration’s National Economic Council.

“The economy was doing a lot worse than we realised, and our policy response was predicated on a much weaker recession,” she admits.



Satellite monitoring of Chinese manufacturing facilities by SpaceKnow © SpaceKnow

How it works: measuring production in China

SpaceKnow builds the Chinese Satellite Manufacturing Index from taking millions of snapshots of more than 6,000 industrial sites across China, and uses artificial intelligence to turn activity patterns into a numeric measure of how well the country's manufacturing sector is doing.

Ms Farrell now heads the [JPMorgan Chase Institute](#), a think-tank set up by the bank to turn its own customer data into valuable economic and policy insights. Among other things, it has explored the role of the [gig economy](#), the impact of out-of-pocket [healthcare spending](#) on a family's financial wellbeing and how mortgage payment adjustments affect defaults or consumer spending. Ms Farrell says big data could have a "huge" impact on policy, especially around recessions. "There is a lot that traditional data cannot answer at extreme moments," she says. "I don't think any of this will supplant the core statistics, but it can clearly supplement it."

At the moment, the US commerce department's Bureau of Economic Analysis produces the quarterly numbers for gross domestic product, but even the "flash" reading comes with a month lag, and it is subject to frequent revisions. In the future, agencies will be able to produce much swifter data on the economy, predicts Philippe Jordan, president of CFM, a French hedge fund.

"Publishing GDP data quarterly will look old-fashioned," he says. "Giving structure to the data is immensely complex. But maybe we could start with getting monthly data on the economy rather than quarterly. That would be a good first step."



Development in North Africa (top) and Nigeria (bottom left) as captured by the SpaceKnow Africa Night Light Index © SpaceKnow

How it works: the African Light index

African statistics can be slow and misleading, so SpaceKnow measures the light intensity at night to gauge activity more quickly. Countries with low cloud density can be measured monthly, while high cloud density countries are reported quarterly.

There are still sceptics in the field. Ewan Kirk, chief investment officer of Cantab Capital, a hedge fund owned by Swiss asset manager GAM, says plenty of promising data sets that his team examines end up proving useless for investment purposes, and it is doubtful they will prove much more valuable in divining the direction of the economy.

“The economy is a really complicated thing, an order of magnitude more complicated than financial markets,” he points out. “The money right now is in being an alternative data provider, not being an alternative data user.”

Economists have become better at developing more up-to-date measures of the economy from traditional data, a practice known as “[nowcasting](#)”. Some argue that new digital data sets add next to nothing to the accuracy of a nowcasting model. For example, Canada already publishes monthly GDP data, and the UK will do so soon.

Data scientists and statisticians admit that the challenges in turning often messy data sets into something usable can be significant. Information on older citizens is often not covered by smartphone or social media data, and credit card data only capture some spending. Satellite feeds can be stymied by bad weather.

Some argue that the biggest obstacles are logistical and legal: the information is largely spread across the private sector, sitting inside banks, telecom companies, social media platforms or manufacturers. In some cases the data can be obtained — at a price — but in many cases there are legal restrictions on what companies can share, or practical limits on what they want to reveal.

Meanwhile, many government statistics agencies are often insufficiently resourced to acquire and dabble with these new data sets.

“The technical challenges are arduous, but solvable . . . People underestimate the regulatory challenges,” says Diane Coyle, a professor of economics at Manchester University, and a fellow at the UK’s [Office for National Statistics](#). She argues that statistical agencies should be given free access to important private sector data, given the public policy implications of better, faster and more granular data.

Yet there are security and privacy concerns involved in centralising enormous data sets that include often sensitive information, says Mr Ainsworth at Schroders. “The question we should ask as a society is whether we should have privacy or whether we should consolidate all this data in one place,” he says. “Because it’s digital and personal, it should be treated with respect.”



An infrared satellite image of farmland in Kansas reveals agricultural development © Getty

Is the prospect of real-time, granular and more accurate indicators derived from big data feasible or fantasy?

The sceptics say big data would not automatically equate to good data. Timeliness can come at an unacceptable cost to accuracy, and the latter should remain the priority of statistical agencies. Prof Cavallo says he sees these new digital data sources as a complement to traditional information, and doubts they will be supplanted any time soon.

“Just because we can measure everything, doesn’t mean that everything is valuable to measure,” he says.

Nonetheless, the early stages of what promises to be a digital data revolution are under way. The optimists say they can already measure economic trends in ways that would have been unthinkable just a decade ago. Existing data sets will have longer time series, which allow for more accurate modelling, and new ones will become available. That should allow participants to improve accuracy and speed up the creation of comprehensive, contemporaneous statistics on entire economies.

Prof Coyle says the field is in the “massive hype stage” of its development but predicts: “Things will progress quickly.”

Satellite data: Snapshots of China’s industrial economy

China has emerged as a fertile ground for data scientists looking to develop alternative measures of economic health, partly because of misgivings over the [quality of its official statistics](#).

While economic data in the west tend to be slow but fairly accurate, even Chinese officials have admitted that its numbers can be massaged — or “man-made” in the words of premier Li Keqiang. That has given rise to a host of alternative measures based on electricity production, loan volumes or rail cargo shipments, with one informal index even named after Mr Li.

Alternative data providers have taken this to a new level. SpaceKnow’s [China Satellite Manufacturing Index](#), which is based on 2.2bn individual snapshots of 500,000 sq km and more than 6,000 industrial sites across the country, is one of the best examples. This gauge offers investors a quicker and arguably more accurate measure of Chinese manufacturing. In 2015-16 it showed a far sharper slowdown than official surveys, most likely capturing the downturn better.

Satellite images can temporarily be foiled by simple things like bad weather, but they offer more granular and up-to-date data than traditional statistics. Orbital Insights — led by former Nasa and Google engineer James Crawford — monitors steel production in India and China from the heat emitted from blast furnaces.

Orbital Insight has also worked with the World Bank on mapping poverty rates using satellite images, and plans to launch more macroeconomic data sets. “This is the future,” says Mr Crawford. “In a few years we will have motor-level surveillance of the entire world every day.”

Letter in response to this article:

Make Big Data open, accessible and available in real time / From Gavin Hayman, Twickenham, Middx, UK

Copyright The Financial Times Limited 2020. All rights reserved.