"High-frequency" data are especially useful for economic forecasting in periods of devastating crisis

The magnitude and suddenness of the shock caused by the Covid-19 pandemic have lessened the relevance and the predictive power of the short-term indicators commonly used to measure and forecast economic activity. Short-term economic monitoring during this time has therefore focused on using new data sources, produced at a higher frequency than monthly or quarterly. In normal times, these indicators are usually relatively ineffective for forecasting and are sometimes more volatile than the main economic aggregates – apart from new data used to monitor French activity since the Covid-19 crisis, but which are outside the scope of this study, which is concentrating on international comparisons–. However, for the four main Eurozone economies, the United States and the United Kingdom, these new data account for a large proportion of variation in the traditional production and consumption indicators. Thus while awaiting these monthly survey results, high-frequency data have proved useful for analysing and estimating activity. As a consequence, in times of crisis, as we are currently experiencing, high-frequency indicators provide additional information to that in the business tendency surveys giving a better understanding of the loss of activity in the very short term.

The predictive power of the usual indicators based on monthly business tendency surveys deteriorates as a crisis approaches and during it

In normal times, the outlook analysis and short-term forecasting carried out by INSEE are largely based on the business tendency surveys. One of the methods used to forecast economic activity –e.g. production or consumption– consists in calibrations[1] using the new information provided each month by business or household surveys. The business tendency surveys are for the most part published monthly, like the other indicators (retail sales, car registrations) used to forecast major economic aggregates, while the forecasts of economic variables are for the most part measured quarterly. In the calibration models, surveys are used up to the most recent one available, for example up to the survey for May for a Q2 forecast. Apart from times of crisis, this method provides good quality forecasts (Dubois, 2006).

However, in times of major crisis or great economic instability, these methods are less suitable. The 2008 crisis provides an example, as demonstrated in one of the focus studies in the *Point de Conjoncture* of 9 April 2020: the operational framework described here was only able to realise the magnitude of the shock very gradually. The current crisis is another example of this: the usual indicators were available only monthly and were sometimes published relatively late, given the unprecedented and very sudden nature of the shock, and the disruption in econometric relationships in these circumstances due to the scale of the crisis. This resulted in a move towards a new way of short-term monitoring, involving estimates of activity in real time using alternative data sources.

Consequently, the most recent issues of INSEE's *Points de Conjoncture* used high-frequency indicators to reflect the economic consequences of the health crisis. The main advantage of high-frequency indicators lies, by definition, in the fact that they are updated weekly or even daily, thus making it possible to monitor the situation in the economies almost instantaneously and compare them. For example, the number of Google searches for unemployment, available in *Google Trends*,[2] can be used as an indicator of job prospects, or even the number of jobseekers; *Google Trends* data on shopping centres can be a leading indicator of the number of visitors to retail outlets and hence of household consumption. Other high-frequency indicators, like electricity consumption and the concentration of nitrogen dioxide in the air can also indicate global economic activity (*Table 1*).

The purpose of this study is to assess the quality of these high-frequency indicators as advance signals of economic activity and analyse their performance compared with the traditional monthly indicators, such as the Industrial Production Index for production, retail sales for consumption. To increase the number of identification points, only high-frequency indicators available for a sufficiently long period and with at least a weekly frequency were considered. Lastly, the approach used in the relatively simple econometric models was to compare the explanatory power of high-frequency indicators rather than search for the best predictive models. The models selected do not necessarily reflect either the practices usually applied in forecasting – for example, consumer confidence is used in the models here to forecast retail sales in France for purposes of comparison, although it is rarely used in actual practice-, nor the practices currently used in France, based on bank card transaction data or scanner data from major retail outlets. Models were therefore chosen mainly for the purpose of comparing indicators in the different advanced countries. In France, bank card transaction

1. Calibrations are econometric regressions linking the economic variable that we are trying to predict, such as production for example, to monthly business tendency survey data or advanced indicators, such as retail sales or car registrations.

2. *Google Trends* are the result of searches on the Google search engine showing the popularity over time of certain search subjects or terms based on the number of searches by internet users.

data and scanner data have been extremely useful for estimating household consumption. Unfortunately, they are not made available by the national statistical institutes in the other countries at such a detailed level, which is why they have not been included in these comparisons.

In "normal" periods of the economic cycle, high-frequency data provide limited information compared with the usual indicators

In "normal" periods of the economic cycle, i.e. with limited variations in activity and thus excluding periods of crisis such as that of 2008-2009 or the current health crisis, high-frequency indicators do not significantly improve short-term forecasting of macroeconomic aggregates.

First, high-frequency indicators, such as Google searches, electricity consumption or air pollution, are very volatile (even when adjusted for climatic factors in the case of electricity consumption), much more so than macroeconomic aggregates in normal times (*Graphs 1 and 2*). Outside times of crisis, there is therefore the risk that high-frequency indicators could contain considerable statistical noise, blurring the short-term information.

Bortoli and Combes (2015) verified this using *Google Trends* data to forecast monthly household consumption. Google searches, like searches for certain products, can indeed reflect the volume of sales of these products. The authors show, however, that *Google Trends* does not make a significant improvement to the forecast of aggregate consumption by households, only to the consumption of specific items, such as clothing-footwear, for example.

More systematically, our intention was to measure the ability of high-frequency indicators to reflect the variability of macroeconomic variables (IPI, retail sales, new car registrations, etc.) compared to the usual indicators. To do this, we compared the explanatory power of two multiple linear models (via the adjusted R^2, the proportion of the variance in the endogenous variable that is predictable from the exogenous variables, adjusted to the number of variables introduced into the model), with one modelling the variable of interest using only the usual indicators (business tendency surveys) and the other adding high-frequency indicators from among those indicated in *Table 1*. By comparing these two models, the authors were able to show the contribution of information from high-frequency data orthogonal to that from surveys. To facilitate the comparison between models "with" and "without" high-frequency indicators, they were estimated over the same period, which was limited by the availability of these indicators: electricity consumption data was available from 2015, therefore the models forecasting the Industrial Production Index were estimated from 2015. For consumption and unemployment, the estimation period excluded the 2008-2009 crisis and started in 2012. To measure the average forecast benefit, the root mean square forecast error (RMSFE) was calculated for both models using a

Table 1 - Usual and high-frequency indicators used in this focus to estimate economic activity in different countries during the crisis

Macroeconomic aggregate	Usual monthly indicators	Availability of usual indicators	High-frequency indicators
Production	PMI Business tendency surveys IPI	PMI: available from the 20th of the month Business tendency surveys: available from the 25th of the month IPI: available about 40 to 50 days after the end of the month	Electricity consumption Concentration of NO2 in the air *Google Trends* "Unemployment", "Credit", "Crises" and "Consumption" Road freight indicator (Germany)
Consumption	PMI Consumer confidence Retail sales	PMI: available from the 25th of the month Confidence indicator: available from the 25th of the month	Electricity consumption; concentration of NO2 in the air *Google Trends* "Consumption", "Shopping centre", "Credit", "Unemployment", Google Trends on the topic of purchase of vehicles
Employment	Employment statistics Unemployment rate	Employment prospects indicator: available from the 25th of the month	*Google Trends* "Unemployment"

Notes:
• only the high-frequency indicators used in the prediction models presented later in this focus are listed here. As these models are standard in the different countries, some indicators available specifically in France and used in this *Point de Conjoncture* (e.g. bank card transactions) are omitted from this table. In fact, we do not have these data for the other countries monitored;
• data on the number of Google searches for "unemployment" were also used in the regressions on consumption and production for economic reasons. The number of jobseekers (potentially reflected by these searches) is strongly correlated with change in production. In addition, an increase in the number of jobseekers may have a negative effect on household consumption expenditure and encourage precautionary savings.

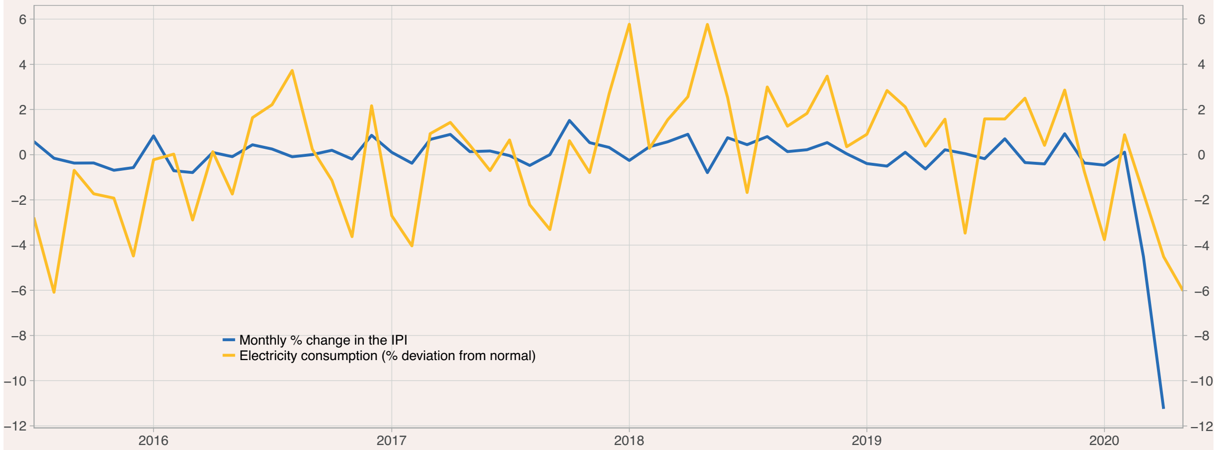
sliding window method.[3] On each date, the model was estimated up to the last available piece of data, then the forecast was calculated for the next date and compared to the indicator actually observed to obtain the "out-of-sample" forecast error.

In normal times, using high-frequency indicators improves a linear model's goodness of fit to the data only slightly. When modelling employment, for example, although *Google Trends* searches for unemployment may account for almost half of variations in the unemployment rate in the four main Eurozone countries or in monthly employment in the other advanced countries, once the Purchasing Managers' Index (PMI) and employment prospects from the Directorate-General for Economic and Financial Affairs (DG EcFIN) survey become available, high-frequency data provide only very little extra information. The increase in the adjusted R^2 following the addition of these indicators is between only 1% and 10% for an explanatory model of the French, German, Italian and Spanish unemployment rate. However, this last result can indicate the presence of an overadjustment of the model to the data. Improvement is also minimal for a model of household consumption in the Eurozone countries, measured from retail sales, and where the adjusted R^2 falls by almost 2% in Italy, and increases by only 2% and 3% in Spain and Germany respectively. High-frequency indicators seem to provide more significant information in a model of industrial production, especially in Germany, Spain and France, with an increase in the adjusted R^2 of between 8% and a little over 40%. Again, such a rise of 40% could indicate an overadjustment phenomenon. The addition of high-frequency indicators can result in a model not being sufficiently generalisable for a good forecast to be obtained with new observations. In other words, the model may wrongly pick up part of the risk of the data-generating process. For this reason, in order to

3. Models with or without high-frequency (HF) indicators were estimated over a period up to T, then used to forecast the point in T+1. The models were then estimated up to T+1 then used to forecast T+2 and so on.

1 - Outside times of crisis, high-frequency indicators (in this case electricity consumption) are more volatile than the IPI: case of the United States

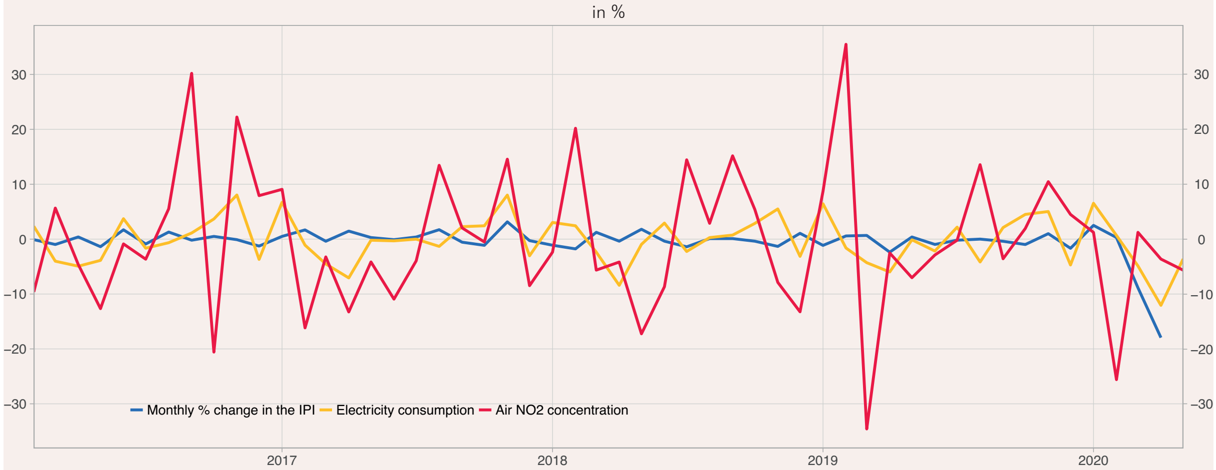
in %



Source: Federal Reserve, Energy Information Administration, INSEE calculations

2 - As in the United States, high-frequency indicators in Germany are more volatile than the IPI

in %



Source: ENTSO-E, EEA, INSEE calculations measure the quality of the information provided by the high-frequency indicators, another criterion must be used. It must be able to assess the ability of a model to correctly forecast a new observation, which is excluded from the estimate sample. This criterion is the RMSFE.

In general, the use of high-frequency indicators makes no improvement to the quality of short-term forecasts in "normal" times, i.e. outside times of crisis. Thus the forecast error of French, German and Italian industrial production increased slightly, while reductions in forecast error remained very small, like that for Spanish industrial production. Electricity consumption data, however, are much more useful at a detailed level.

In the United States, where details of monthly consumption are available 30 days after the end of the month and which is therefore forecast instead of retail sales, adding *Google Trends* "shopping centre" and "unemployment" accounts for some of the variations, but does not improve the forecast for monthly consumption in the United States. In the United Kingdom too, these high-frequency indicators do not improve the forecast for household consumption, although some can be used to improve the forecast for certain specific consumer items, such as car registrations, for example.

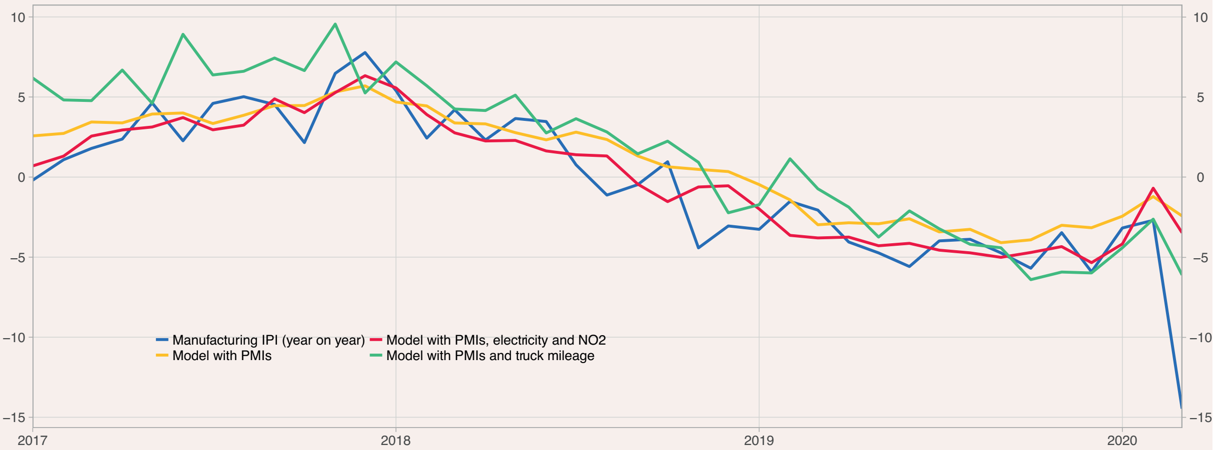
Finally, in the United States, employment statistics for a given month are published on the first Friday of the following month (except when this is a public holiday or falls on the 1st of the month). These figures are therefore available rapidly, with the result that high-frequency indicators provide much less information than in France and Germany and are therefore less useful in this case than for forecasting industrial production or household consumption.

It is during periods of crisis that certain high-frequency indicators provide a better understanding of loss of activity

In times of crisis, high-frequency indicators provide better estimates of the magnitude of the shock

3 - Truck mileage data, *Google Trends* and concentration of NO2 in the air add significantly to PMIs in the German industrial activity model

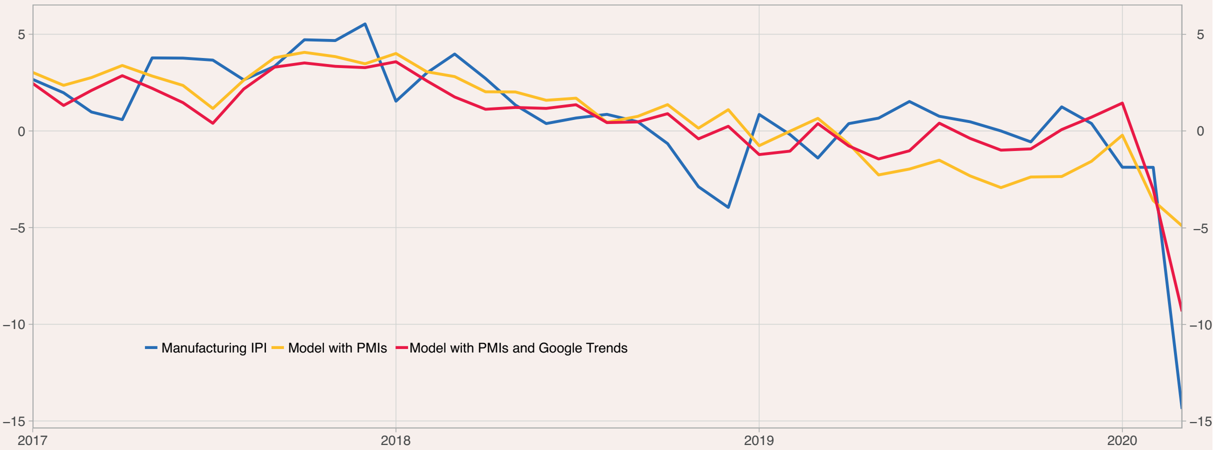
in %



Source: Destatis, EEA, Google Trends, *INSEE calculations*

4 - As in Germany, high-frequency indicators better reflect the slump in industrial production in Spain

in %



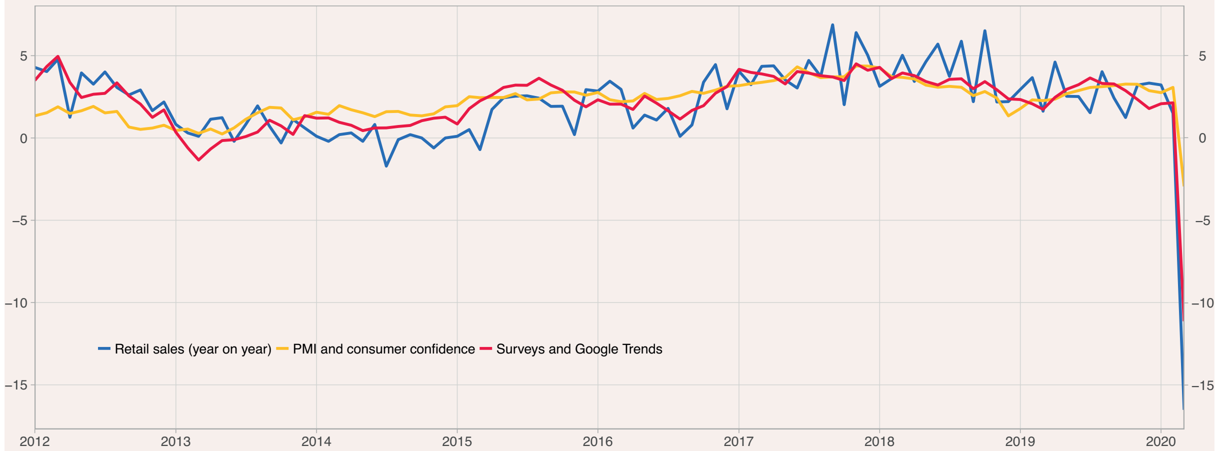
Source: Eurostat, Google Trend, *INSEE calculations* than the usual indicators. Thus, using PMIs alone, estimates of the effect of the health crisis on industrial production in March 2020 were expected to reach –2.4% and –7.5% in Germany and France respectively, against –6% and –12% when adding the high-frequency indicators, compared with an actual decline of a little over 14% and 19% respectively in Germany (*Graph 3*) and France. In Spain, high-frequency indicators also give a better appreciation of the actual decline in activity in March 2020, but less so in Italy. The drop in industrial activity observed in March in Spain was –13% and that estimated by adding high-frequency indicators (mainly *Google Trends*) was around –9.3% (against –4.9% using only PMIs, *Graph 4*). However, the difference between the estimated and the actual scale of the shock remains high, at around 4 to 7 percentage points depending on the country. Consequently, despite the use of high-frequency indicators, the econometric models have difficulty in reflecting the scale of the drop in actual activity.

Regarding the drop in consumption, high-frequency indicators provided a significant forecast benefit. For example, while retail sales fell by more than 16% in France in March, the magnitude of the shock estimated by the standard indicators was only –3%, against a decline of –11% forecast with the introduction of *Google Trends* data (*Graph 5*). However, these forecast gains are less significant for employment and consumption in Germany, Spain and Italy.

In the United Kingdom, electricity consumption did not provide any additional information. However, the addition of pollution did provide some significant information: the adjusted R^2 (in-sample) increased by 90%, to 64%. In the United States, electricity consumption and Google searches for “unemployment” provided additional information to that from the PMIs for the Industrial Production Index (*Graph 6*): the R^2 increased by 20% when these two high-frequency indicators were added

5 - *Google Trends* significantly improved the estimate of the drop in retail sales: case of France

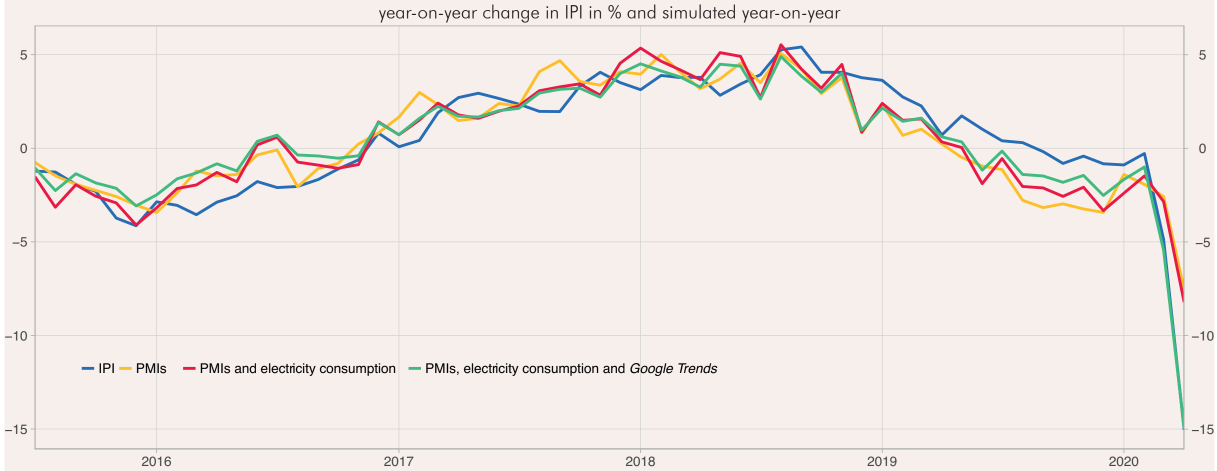
in %



Source: Google Trends, *Eurostat, INSEE calculations*

6 - As in Germany, high-frequency indicators improve the representation of industrial production in the United States

year-on-year change in IPI in % and simulated year-on-year



to the regression alongside the Institute for Supply Management's PMI. However, out-of-sample, i.e. when in order to forecast each point, the model is estimated on data available on this date (hence up to date T–1), the average forecast error decreases only very slightly. The addition of high-frequency indicators, such as electricity consumption or *Google Trends* on unemployment, greatly improves the forecast of the decline in activity in the United States (*Graph 3*). Out-of-sample, the improvement in the forecast is less but still considerable.

Ultimately, in most cases the high-frequency indicators did not provide any significant additional information to that in the business tendency surveys and brought only limited improvements to economic forecasts during "normal" times. However, in times of crisis with drastic and large-scale variations in economic activity, the usual models proved to be unsuitable for predicting economic activity. High-frequency indicators can then be used to improve forecasts a little. However, some of the high-frequency data used specifically for France but outside the scope of this study, such as scanner data or bank card transaction data, are an invaluable source of information for short-term monitoring.

The expertise and analytical skill of the economic forecaster are needed to adjust and modify the econometric models for a better understanding of the change in activity. ■

Bibliography

C. Bortoli et S. Combes, (2010) «Contribution from *Google Trends* for forecasting the short-term economic outlook in France: limited avenues», *Note de conjoncture* de March 2015

E. Dubois, (2006) « Étalonnages à l'aide d'enquêtes de conjoncture: de nouveaux résultats ». *Économie & prévision*, n° 172, janvier 2006

Insee, (2020) «Measurements and forecasting in times of crisis: a comparison with 2008-2009», *Point de conjoncture* du 9 avril 2020 ■