OECD Economics Department Working Papers No. 1634

Tracking activity in real time with Google Trends

Nicolas Woloszko

https://dx.doi.org/10.1787/6b9c7518-en

OECD

Organisation for Economic Co-operation and Development

ECO/WKP(2020)42

Unclassified English - Or. English

27 November 2020

ECONOMICS DEPARTMENT

TRACKING ACTIVITY IN REAL TIME WITH GOOGLE TRENDS

ECONOMICS DEPARTMENT WORKING PAPERS No. 1634

By Nicolas Woloszko

Authorised for publication by Luiz de Mello, Director, Policy Studies Branch, Economics Department

All Economics Department Working Papers are available at www.oecd.org/eco/workingpapers.

JT03469118

ABSTRACT/RESUMÉ

Tracking activity in real time with Google Trends

This paper introduces the OECD Weekly Tracker of economic activity for 46 OECD and G20 countries using Google Trends search data. The Tracker performs well in pseudo-real time simulations including around the COVID-19 crisis. The underlying model adds to the previous Google Trends literature in two respects: (1) the data are adjusted for common long-term bias and (2) the data include variables based on both Google Search categories *and* topics (the latter being a collection of related keywords), thus further exploiting the potential of Google Trends. The paper highlights the predictive power of specific topics, including “bankruptcies”, “economic crisis”, “investment”, “luggage” and “mortgage”. Calibration is performed using a neural network that captures non-linear patterns, which are shown to be consistent with economic intuition using machine learning interpretability tools (“Shapley values”). The tracker sheds light on the recent downturn and the dynamics of the rebound, and provides evidence about lasting shifts in consumption patterns.

Keywords: nowcasting, Google Trends, high-frequency, machine learning, neural network, interpretability, COVID-19

JEL: C45, C53, C55, E37

Suivre l’activité économique en temps réel avec Google Trends

Ce document présente le Tracker Hebdomadaire d'activité économique de l'OCDE pour 46 pays de l'OCDE et du G20 en utilisant les données de recherche de Google Trends. Le Tracker fonctionne bien dans les simulations en pseudo-temps réel, y compris autour de la crise liées au COVID-19. Le modèle sous-jacent innove par rapport à la literature existante autour de Google Trends à deux égards : (1) les données sont ajustées pour tenir compte des biais de long terme et (2) les données comprennent des variables basées à la fois sur les catégories et les sujets de recherche Google (ces derniers étant une collection de mots-clés apparentés), ce qui permet d'exploiter davantage le potentiel de Google Trends. Le document souligne le pouvoir prédictif de sujets spécifiques, notamment les "faillites", la "crise économique", les "investissements", les "bagages" et les "hypothèques". Le calibrage est effectué à l'aide d'un réseau de neurones qui capture des relations non linéaires, dont il est démontré qu'elles sont conformes à l'intuition économique grâce à des outils d'interprétabilité en apprentissage automatique ("valeurs de Shapley"). Le Tracker éclaire le récent ralentissement et la dynamique de la reprise, et documente des changements dans les habitudes de consommation.

Mots clefs: nowcasting, Google Trends, haute fréquence, apprentissage statistique, réseau de neurones, interprétabilité, COVID-19

JEL: C45, C53, C55, E37

Table of contents

Tables

Figures

Boxes

Tracking activity in real time with Google Trends

Nicolas Woloszko[1]

1. Introduction and Summary

1. A pre-requisite for good macroeconomic policymaking is timely information on the current state of the economy, particularly when economic activity is changing rapidly. Given that GDP is usually only available on a quarterly basis (with first estimates typically published only 4 weeks or more after the end of the quarter), policymakers and forecasters have long made use of more timely higher frequency data, such as survey-based indicators like Purchasing Managers' Indices (PMIs). However, both the current crisis and the earlier ones have shown that the underlying relationship with survey-based indicators can become unreliable when changes in economic activity are abrupt and massive (Vermeulen, 2012[1]).This problem has prompted a search for alternative high-frequency indicators of economic activity. This paper discusses one such indicator based on Google Trends, which are used to construct a Weekly Tracker that provides real-time estimates of GDP growth in 46 G20, OECD and partner countries.

2. The OECD Weekly Tracker of GDP growth attempts to fill the gap in real-time high-frequency indicators of activity with a large country coverage. To the author's knowledge, the Tracker is the first weekly GDP proxy that can be compared across a large array of OECD and G20 countries. The Tracker provides estimates of year-on-year growth rate in weekly GDP with a 5-day delay. It applies a single machine learning algorithm on a panel of Google Trends data for 46 countries. The algorithm flexibly captures cross-country heterogeneity and provides comparable estimates across countries. It exploits the full potential of Google Trends data by aggregating together information about search behaviour related to consumption, labour markets, housing, industrial activity and uncertainty. The Tracker provides high-frequency and real-time information about the COVID-19 crisis and subsequent rebound, as well as the impact of confinement measures.

3. The Tracker uses a two-step model to nowcast weekly GDP growth based on Google Trends. First, a *quarterly model* of GDP growth is estimated based on Google Trends search intensities at a quarterly frequency. Second, the relationship between Google Trends and activity, using the same elasticities estimated from the quarterly model, is applied to the *weekly* Google Trends series to yield a weekly tracker. The relationship between Google Trends variables and GDP growth is fitted using a machine learning algorithm ("neural network"). The algorithm captures non-linearities that are likely to be

[1] The author is a member of the OECD Economics Department Macroeconomic Analysis Division and NAEC Innovation LAB, and thanks David Turner, Sebastian Barnes, Annabelle Mourougane, Boris Cournède, Laurence Boone, Luiz de Mello, Alain de Serres, Nigel Pain, Sylvain Benoit, Frédéric Gonzales, Daniela Glocker, Tim Bulman, Hermès Morgavi, Sebastien Turban, Ana Chico Sanchez, and Véronica Humi, as well as participants of the Bank of England and Banca d'Italia 2020 conferences on economics and machine learning. The author also thanks Hal Varian for granting the NAEC Innovation Lab access to the Google Trends API.

key in extreme situations, but which are difficult to estimate with more conventional econometric approaches.

4. Using modern machine learning interpretability tools, this paper exploits the neural network to derive insights about non-linear patterns captured by the model that are consistent with economic intuition. For instance, searches for unemployment benefits are stronger predictors of activity around times when lay-offs increase and thus become dominant with regards to hiring in explaining changes in employment. Model interpretability tools also highlight the most important variables and the macroeconomic predictive power of a number of topics including “bankruptcies”, “economic crisis”, “investment”, “luggage” and “mortgage”.

5. The model of GDP growth based on Google Trends proves to perform well in out-of-sample nowcast simulations. On average across OECD and G20 countries, the quarterly model based on Google Trends has a Root Mean Squared Error (RMSE) lower by 17% than an auto-regressive model that just uses lags of year-on-year GDP growth. It captures a sizeable share of business cycle variations, including around the Global Financial Crisis (when the available data for training the algorithm was much smaller) and the euro area sovereign debt crisis. The timing of the downturn and subsequent rebound is well captured by the model, although the full magnitude of the negative shock in the second quarter of 2020 is typically under-estimated, given its unprecedented scale. The tracker thus provides a useful tool for real-time narrative analysis on a weekly basis, although it does not on average outperform models based on more standard variables, once these are eventually released. It also provides evidence of lasting shifts in consumption patterns away from services implying social interactions.

6. The paper is organised as follows. The second section describes the Google Trends data, data issues and data pre-processing. The third section introduces the non-linear modelling approach. The fourth section displays results of pseudo-real time simulations. The fifth section provides insights into the inner workings of the model using interpretability tools. And the sixth one shows the Weekly Tracker and provides insights on the 2020 recession.

2. The COVID-19 crisis called for the use of high-frequency indicators

7. The 2020 crisis is unique in its magnitude and speed, and highlighted the caveats of standard indicators. Leading indicators most commonly used by policymakers fall in two categories: “hard” and “soft” (Table 1). Hard indicators are collected by national administrations or statistical agencies and suffer from publication delays ranging from one to three months, which is a major constraint for policymakers facing rapid fluctuations in activity. Soft indicators are timelier, but can become less informative about GDP during recessions. PMIs and confidence surveys are often based on averages of qualitative answers based on the net balance of respondents’ optimism or pessimism, which limits their ability to quantify the magnitude of an ongoing crisis.

Table 1. Standard indicators were outpaced by the crisis

Indicator	Type	Frequency	Release	Relationship to GDP
GDP	Hard	Quarterly (monthly for GBR, CAN and SWE)	Usually 1-2 months after the end of the quarter	
Industrial production	Hard	Monthly	Around 30-55 days after the end of the month	Linear
Retail sales	Hard	Monthly	Around 8-10 weeks after the end of the month	Linear
PMIs	Soft	Monthly	Around start of the next month	Linear in normal times, non-linear around crises
Consumer confidence	Soft	Monthly	Around start of the next month	Linear in normal times, non-linear around crises
Google Mobility	High-frequency	Daily	With a 7-day delay	Difficult to calibrate as historical data start mid-February 2020.
Google Trends	High-frequency	Daily, Weekly or Monthly	With a 5-day delay	Model-based relationship

Source: OECD.

8. As a specific example, the information provided by standard indicators to French policymakers when they implemented the lockdown in mid-March illustrates the limitations of these traditional gauges at a time of crisis. After the lockdown was implemented on 17 March, the first releases were the flash PMIs on 24 March, which sent mixed signals reflecting the uneven nature of the shock as the manufacturing PMI fell moderately (to 42.9) while the services PMI fell to an all-time low (29.0). On 27 March, consumer confidence readings for February edged down marginally (to 103 from 104), well above market expectations (of 92), consistent with the unexpectedly high business confidence released one day before. Flash GDP releases for the first quarter came out on 30 April at -5.8% quarter-on-quarter, which did not provide specific information about activity in March as the GDP figure is a quarterly average. The first traditional hard indicators to provide information about activity in March were household consumption (-17.9% month-on-month) and industrial production (-16.2% month-on-month), but these were only published on 30 April and 7 May, respectively, over six weeks after the start of the lockdown.

9. The need to quickly understand the impact of the COVID-19 pandemic to calibrate policy advice has made high frequency data not only more relevant, but often in the first instance the only way of measuring the impact of the crisis in real time. The swift economic policy responses was in part made possible by the existence of programmes to facilitate work-time sharing among employees and which in many countries were already in place and thus ready to be activated. Other programmes with features that could be made more contingent on the state of the economy include for instance unemployment benefits, or support to businesses in financial distress that would neither overload them with debt nor distort competition and which could help to limit the unnecessary liquidation of otherwise solvent and viable. The need to calibrate such policy actions on real-time assessments of economic activity increases reliance on high-frequency indicators.

10. The past few years have seen the emergence of new types of high-frequency indicators. These include flight departures, restaurant bookings, mobility reports based on anonymised personal data from Google and Apple, air quality indices, news-based indicators such as the Economic Policy Uncertainty Index (Baker, Bloom and Davis, 2013[2]), electricity consumption, and credit card transactions. These new indicators are often available on a daily or real-time basis and for a range of countries. Policy institutions and National Statistical Agencies (NSOs) across the world have turned to such alternative data, including the ECB (Benatti et al., 2020[3]), the Bank of England (Bank of England, 2020[4]), INSEE (INSEE, 2020[5]), the Federal Reserve banks of Saint Louis (Kliesen, 2020[6]) and Cleveland (Knotek and Zaman, 2020[7]),

and the IMF (Chen et al., 2020[8]). Relatedly, the Harvard-based project "Opportunity Insights" gathered a large number of high-frequency data on the US economy from private companies. The OECD has leveraged a number of high-frequency indicators (OECD, 2020[9]), including Google Mobility reports (based on the locations of Google Maps users). This paper focuses on Google Trends data, which provides aggregate information from Google Search.

3. Exploiting the full potential of Google Trends

11. In the past decade, a growing literature has provided evidence of the usefulness of Google Trends data for 'nowcasting' the current state of the economy (Varian and Choi, 2009[10]; Carrière-Swallow and Labbé, 2010[11]; Chen et al., 2015[12]; Narita and Yin, 2018[13]; Ferrara and Simoni, 2019[14]). Papers have studied the link between Google Trends data and employment or unemployment (Baker and Fradkin, 2017[15]; Fondeur and Karamé, 2013[16]; D'Amuri et al., 2012[17]), as well as consumption (Morgavi, 2020[18]), trade (Gonzales, Jaax and Mourougane, 2020[19]), digitization (Pisu, Costa and Hwang, 2020[20]), or housing prices (Askitas and Zimmermann, 2009[21]; Wu and Brynjolfsson, 2015[22]) and construction (Cournède, Ziemann and De Pace, 2020[23]). More recently, Google Trends data have also been used to assess the impact of the COVID-19 crisis (Abay, Tafere and Woldemichael, 2020[24]; Doerr and Gambacorta, 2020[25]).

12. Google Trends provides Search Volume Indices, which measure search intensity (number of searches for a given keyword divided by total searches) by location and period. Queries can be made by keyword, category of keywords or topic. Queries based on keywords are language-specific and subject to ambiguity. Google Trends series for the keyword "apple" mixes up searches for the fruit and the company. Both categories and topics are harmonised across languages, and queries based on categories or topics yield series comparable across countries. Google Trends thus provide a dataset of monthly panel data. Focusing on observations from 2004 to 2020 for 46 countries, the monthly data give a total of 8370 observations. Using topics and categories allows for more general models as topics and categories abstract away from keywords and provide a representation of search interest for things rather than specific terms.

13. Google has classified searches into 1200 categories[2]. These allocate individual searches to (multiple) categories using a probabilistic algorithm (Varian and Choi, 2009[10]). Categories are structured as a 5-level hierarchical classification. For instance, the category "Autos and Vehicles" aggregate together all searches related to cars, whereas an equivalent query based on keywords would have to explicitly combine each possible car name and brand.

14. Topics address the ambiguity problem of keywords. The topic "Apple (company)" allows to single out searches related to the company not the fruit, and combines searches for keywords such as "apple watch", "ipad", and "macbook". Google has created topics that aggregate together multiple requests made on Google Search based on their purpose and meaning, by taking into account where users click. There is no fixed list of topics and topics selection implied exploration from the Google Trends website.

15. Categories and topics also have drawbacks. The exact content of topics and categories is opaque and the algorithm that allocates keywords can make arbitrary choices. For instance, the topic "unemployment benefits" encompasses mostly French keywords in Canada, and is thus informative about labour markets in French-speaking Québec rather than the whole country. This further warrants the use of machine learning algorithm capable of flexibly capturing cross-country causal heterogeneity.

[2] The list of all categories can be found in this repo: https://github.com/pat310/google-trends-api/wiki/Google-Trends-Categories

16. The present exercise exploits the full potential of the Google Trends data by using both category-based and topics-based searches. Two hundred fifteen categories were selected from 1 200 on a judgmental basis, as selecting data based on judgment may provide better results than data-driven selection (Combes, Bortoli and Clément, 2016[26]). Thirty-three topics were retained on a judgmental basis as well (see details in Annex B). A number of papers have exploited search categories for nowcasting (Scott and Varian, 2014[27]; Suhoy, 2009[28]; Wu and Brynjolfsson, 2015[22]; Ferrara and Simoni, 2019[14]; Combes, Bortoli and Clément, 2016[26]). Fewer use search topics (Narita and Yin, 2018[13]; Fetzer et al., 2020[29]). To the author's knowledge, this paper is the first one to combine search categories and topics for economic nowcasting, along with (Gonzales, Jaax and Mourougane, 2020[19]).

17. Google Trends categories and topics cover a large number of economic sectors, with a strong though not exclusive focus on consumption. The Tracker built in this paper exploits Google Trends variables related to consumption goods (*e.g.*, food and drinks, vehicle brands, home appliances) or services (*e.g.*, performing arts, travel, sports, restaurants, arts and entertainment), which represent a large share of GDP. It also includes search intensities informative about labour markets (*e.g.*, unemployment benefits, jobs), housing and construction[3] (*e.g.*, real estate agencies, credit and lending, forbearance), a large array of business services (*e.g.*, venture capital, commercial vehicles), bankruptcies (*e.g.*, bankruptcy), which can be tightly related to the business cycle. Searches performed as part of some industrial activities are also included (*e.g.*, maritime transport, agricultural equipment), which can provide information on the supply side. Lastly, it includes searches whose intensity suggests economic anxiety (*e.g.*, economic crisis, economic news) in order to better capture crises (Fetzer et al., 2020[29]) as well as poverty (*e.g.,* food bank). Signals about multiple facets of the economy can be aggregated to infer a timely picture of the macro economy. Using many variables also reduces the risk related to structural breaks in specific series, which was highlighted by the failure of the "Google Flu" experiment[4].

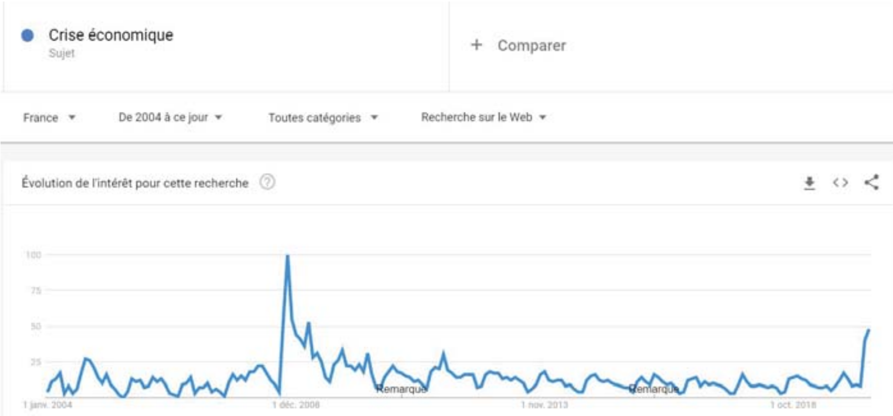
18. High-frequency and big data are often subject to limitations as the original purpose of their collection is usually not scientific analysis. These caveats call for specific attention and statistical pre-processing. Google Trends variables are transformed to year-over-year growth rate in order to remove seasonality. Breaks occurring in January 2011 and January 2016 caused by changes in the data collection process are addressed by smoothing the year-on-year growth rates. Finally, as the Google Search user base has increased dramatically since 2004, the *relative* search intensities of most search categories decrease over time. The methodology used to filter out the long-term bias as well as the detailed pre-processing treatments are described in detail in Annex A.

[3] Google Trends data were used in (Cournède, Ziemann and De Pace, 2020[23]) in order to provide real-time estimates of construction PMIs for a large country coverage

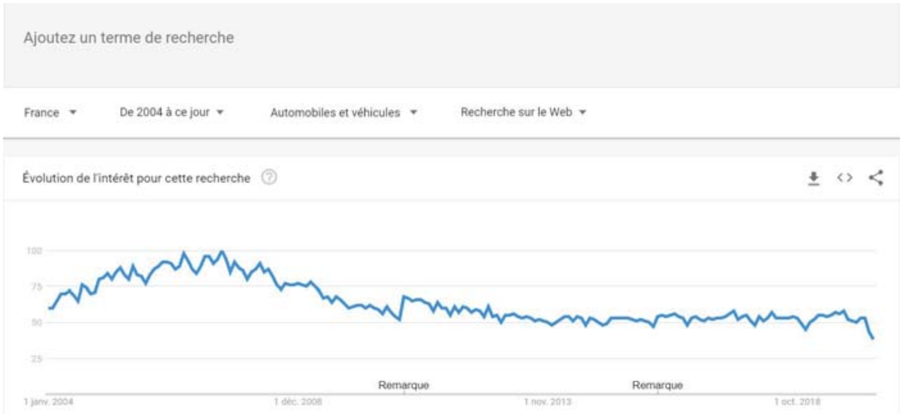
[4] In 2009, Google started tracking influenza epidemics based on searches for "influenza" or related symptoms (Ginsberg et al., 2009[55]). In 2013, the experiment was shown to be limited by media coverage of influenza epidemics during major outbreaks that were causing surges in Google searches unrelated to the virus propagation (Butler, 2013[56]).

Figure 1. Queries in Google Trends: beyond keyword searches

A. Topic-based query (economic crisis)



B. Category-based query (automobile and vehicles)



Note: The topic-based search aggregates multiple keywords including “subprimes”, “crise des subprimes” and so on. It is valid across languages. For instance if the country filter is set to Spain, it will aggregate Spanish keywords relating to crises including “crisis economica” and “crisis españa. The category-based query aggregates all searches falling into the “Automobile and vehicles” category. It is also harmonised across languages.
Source: Google Trends.

4. A neural panel model of GDP growth

19. This paper constructs a model of GDP growth from Google Trends search volume indices. The model aims at nowcasting GDP growth at a weekly frequency. The model is fitted using quarterly Google Trends series and applied to weekly Google Trends series in order to provide a weekly tracker. It can capture multiple non-linearities, as no assumption is made as to the shape of the relationship between Google Trends search intensities and economic activity. Lastly, it exploits a distinctive feature of Google Trends data, i.e. variables comparable across 46 OECD and G20 countries, by pooling countries together.

4.1. From quarterly GDP growth to a weekly tracker: a bridge model of GDP growth

20. The Weekly Tracker uses a two-step model to nowcast weekly GDP growth[5] based on Google Trends. First, a quarterly model of GDP growth is estimated based on Google Trends search intensities at a quarterly frequency using a panel model of 46 G20, OECD and partner countries:

$$y_{iq} = f\left(d\, svi_{c,q}\, , cfe_i \right) + \sigma_i \tag{4}$$

Where the growth rate of GDP on the same quarter of the previous year (y_{iq})[6] is modelled as a non-linear function f of the year-over-year log-difference of quarterly averages of search volume indices ($d\, svi_{c,q}$) for categories (indexed by c) and country dummies (cfe_i), plus some white noise σ_i. Second, the function $\hat{f}$ estimated from the quarterly model is applied to the weekly Google Trends series, assuming that this relationship is frequency-neutral, in order to yield a weekly tracker:

$$\widehat{y_{iw}} = \hat{f}\left(d\, svi_{c,w}\, , cfe_i \right) \tag{5}$$

The OECD Weekly Tracker can thus be interpreted as an estimate of the year-over-year growth rate of "weekly GDP" (same week compared to previous year). Two countries have monthly GDP series (the United Kingdom and Canada), which are used in to train the model along with quarterly GDP series for other countries. Monthly GDP series are regressed on of the year-over-year log-difference of monthly averages of search volume indices.

21. The model covers all OECD countries as well as G20 members (excluding the European Union) and partner countries. China and Saudi Arabia are excluded from the sample as well, as the relationship between activity and searches on Google seem more heterogeneous in these two countries[7]. The resulting model includes 263 variables and is trained on 2 806 observations, corresponding to 46 countries observed along 61 quarters since 2005.

22. The model of quarterly GDP yearly growth rate is built with the objective to infer weekly estimates. Consequently, lower-frequency variables cannot enter the model, thus excluding a lagged dependent term. Quarterly variables widely used for short-term forecasting may explain significant shares of the variations but often come with lags and perform poorly around turning points. Moreover, lower frequency variables cannot improve the granularity of weekly estimates. This constraint represents a major challenge with regards to existing literature that often relies on autoregressive models (Varian and Choi, 2009[10]) or adding Google Trends data to standard quarterly variables (Ferrara and Simoni, 2019[14]).

4.2. A non-linear algorithm

23. The relationship between Google Trends variables and GDP growth is fitted using a machine learning algorithm ("neural network", see (Csáji, 2001[30])). Google Trends "big" data make it possible to use such algorithms that are powerful but require large samples. The algorithm captures non-linearities that are likely to be key when there are extreme movements in GDP, but which are difficult to estimate with more conventional econometric approaches. Cross-country differences related to Google Search's market penetration or institutional settings are flexibly captured as the neural network allows for all possible interactions between Google Trends variables and country dummies.

---

[5] The dependent variable is GDP (expenditure approach) in growth rate compared to the same quarter of the previous year, seasonally adjusted, at a quarterly frequency, taken from the OECD Quarterly National Accounts (QNA).

[6] For the United Kingdom and Canada, monthly GDP series are available and were used along with monthly log-differences of Google Trends series.

[7] Google curtailed its activities in mainland China in 2010 following censorship-related rows with the authorities.

24. The neural network can be thought of as an alternative to using dynamic factors or principal components as it reduces the dimensionality to a number of intermediate components in the middle layer before making a prediction. The multi-layer structure helps avoid overfitting. As opposed to PCA, it allows for capturing non-linear relationships. Variables are pre-processed using normalisation. The main caveat of neural network is their black-box nature, which is addressed using machine learning interpretability techniques in section 6.

25. While a vast research has focused on the inclusion of the GT indicator as an explanatory variable in conventional autoregressive models, papers have used factor models of multiple GT categories (Vosen and Schmidt, 2011[31]; Balakrishnan and Dixit, 2013[32]). Other papers used linear shrinkage methods such as Ridge (Ferrara and Simoni, 2019[14]) or Spike-and-Slab (Scott and Varian, 2014[33]). Fewer papers have used non-linear methods: Burdeau and Kinzler (2017[34]) experimented with Support Vector Machines (SVMs) and boosting and reported better results from non-linear approaches.

26. Neural networks have had attracted little attention from macroeconomists compared to tree-based methods such as Random Forests, mostly because of the small size of macroeconomic data. By providing variables comparable across countries for a large number of countries and at a high frequency, Google Trends creates opportunities for using a wider array of algorithms and econometric methods, be it for prediction or policy analysis.

Box 1. Training the neural network

Additional details on the training on the neural network algorithm.

Architecture and technical details. The neural network algorithm used in this paper is a standard multi-layer perceptron implemented with most of the default parameters in Python statistical software *scikit-learn.* It includes two hidden layers of respectively 100 and 20 neurons. Each neuron uses a "*relu*" activation function. The activation function takes a weighted sum of input signals (the variables values) and yields the linear combination of inputs provided it is higher than a given threshold. The weights and threshold are optimised using stochastic gradient descent.

Standard Scaler. It has become an industry standard to scale the variable values prior to fitting the algorithm. Early experiments proved that Quantile Scalers performed badly especially when it comes to predicting extreme values in times of crises. Standard Scalers do not treat extreme values as outliers and thus allow better performance around downturns.

Ensemble. Neural networks are notoriously sensitive to the initial random parameters. The choice of random parameters proved to have a strong effect on the results. In order to curb the effect of that randomness, the tracker uses an ensemble of five neural networks initialised with random parameters, whose predictions are averaged over.

Hyper-parameter optimisation. The use of gridsearch for hyperparameter optimisation was purposely avoided. Even when perfomed on a training set prior to the forecast simulations, gridsearch can lead to "overfitting the validation set": users may experiment with many parameter grids and simulation settings leading to biased simulation results. In order to prevent that issue, parameter optimisation was performed through a simple trial-and-error process aiming at finding a good level of fit with reasonable performance rather than at maximising goodness-of-fit. An additional guarantee against overfitting the validation set is provided by the generality of the model over a large number of countries, which reduces the likeliness of bias in performance measurement caused by *ad hoc* hyper-parameter selection.

4.3. To pool or not to pool: a panel nowcasting model for 46 countries

27. The panel nature of the data raises the question of whether to pool countries together or run country-wise models. Country-specific models seem more intuitive as various levels of internet penetration, habits, culture, demography and institutions could explain possibly large differences in country-specific elasticities, but would involve many more variables (248) than observations (61). Conversely, pooling countries together increases the sample size and thus estimation accuracy[8].

28. This paper uses a neural panel model, which exploits a large sample of observations from 46 countries while capturing cross-country heterogeneity. Neural networks are able to handle heterogeneity in the data as long as country dummies are included. A neural network whose architecture incudes an intermediate layer with enough neurons (in our case, 100) can flexibly model each possible interaction between Google Trends variables and country dummies. Each neuron takes as input signals from Google Trends variables and country dummies, and returns a non-linear function of the weighted sum of these inputs. As a result, the model can capture country-specific elasticities.[9]

29. The next section provides an out-of-sample assessment of the quarterly model. Predictions from quarterly SVIs annual growth rates are compared against official GDP year-on-year growth rates at each quarter. This exercise will provide a notion of the model performance, while the weekly predictions cannot be evaluated as there is no official record of weekly GDP.

Box 2. Sparse or dense?

A useful consideration prior to choosing a model type is whether the data generating process is sparse or dense. Sparse processes involves only a few variables relevant for predicting the outcome while most others are noise. Dense processes relates to cases where a large number of variables will have small but significant impact on the y variable. Sparse models will thus try and select the most important variables, as would Random Forests or Lasso. Dense models such as Dynamic Factor Models or Ridge will include a larger number of variables while constraining the total number of parameters in order to prevent overfitting.

The dense-or-sparse question is pivotal in economic predictive modelling. In a recent paper, Giannone, Lenza and Primiceri (2018[35]) shed light on the "illusion of sparsity" in macro, whereby a small number of variables are key while a large number of variables may remain significant. The absence of a clear-cut answer would explain the long-standing debate between proponents of DFMs against Vector Auto-Regressive models (VARs).

Experiments with Elastic Net (Zou and Hastie, 2005[36]) in a spirit close to (Giannone, Lenza and Primiceri, 2018[35]) showed dense patterns in the Google Trends data. This finding suggest that neural networks are better suited. In addition, Random Forests and XGBoost regressors (which are both sparse algorithms), both with standard scikit-learn parameters, underperformed the neural network used in this paper.

---

[8] The alternative between the two options can thus be thought of as bias-variance trade-off: introducing some bias by using average elasticities rather than country-specific ones allows to substantially reduce the variance of the estimator and increases overall predictive accuracy.

[9] As a consequence, neural networks with few but large layers may be preferred.

5. How well can the OECD Weekly Tracker nowcast the economy?

30. The predictive performance of the model underlying the Weekly Tracker is assessed using pseudo-real time simulations on the quarterly series. Pseudo-real time simulations emulate the conditions a forecaster would have faced at each time period, by looping on time and using only past or present data (except for the fact that revised series are used, not vintages). The simulations suggest that the Google Trends Tracker provides relevant leading information on GDP growth, economic crises and business cycles in almost all 46 countries in the sample. With one exception, it accurately signals the COVID-19 downturn in all countries.

31. The dependent variable to be explained is GDP growth at M-1, which is GDP growth one month before its official release. Simulations are an out-of-sample exercise and emulate a forecast made at the end of the last month of the current quarter. A Q2 GDP growth forecast will thus use Google Trends data up until June and the algorithm will be trained on Google Trends and GDP data up to Q1. At each iteration, the algorithm training parameters are optimised using early stopping and a 10% hold-out sample from the training set.

32. The quarterly model of annual GDP growth based on Google Trends performs well in out-of-sample nowcast simulations (Table 2). On average across 46 countries, the quarterly model based on Google Trends has a Root Mean Squared Error (RMSE) that is 17% lower than an autoregressive model that just uses lags of year-on-year GDP growth.[10] It captures a sizeable share of business cycle variations, including around the Global Financial Crisis (when the available data for training the algorithm was much smaller) and the euro area sovereign debt crisis (Figure 2, additional results in Annex C). Its RMSE is on average 8% lower than an autoregressive model in 2008-10 and 41% lower in 2020. The timing of the downturn and subsequent rebound is well captured by the model, although the full magnitude of the negative shock in 2020 Q2 is typically under-estimated, given its unprecedented scale. The mean average error in predicting year-on-year GDP growth in the first (second) quarter was 2.42 (3.86) percentage points, compared with actual falls in GDP for the median country of 0.12% (10.4%). The Tracker thus provides a useful tool for real-time narrative analysis on a weekly basis, although it does not on average outperform models based on more standard variables, once these are eventually released.

Figure 2. Nowcasting GDP growth with Google trends (M-1 forecast)

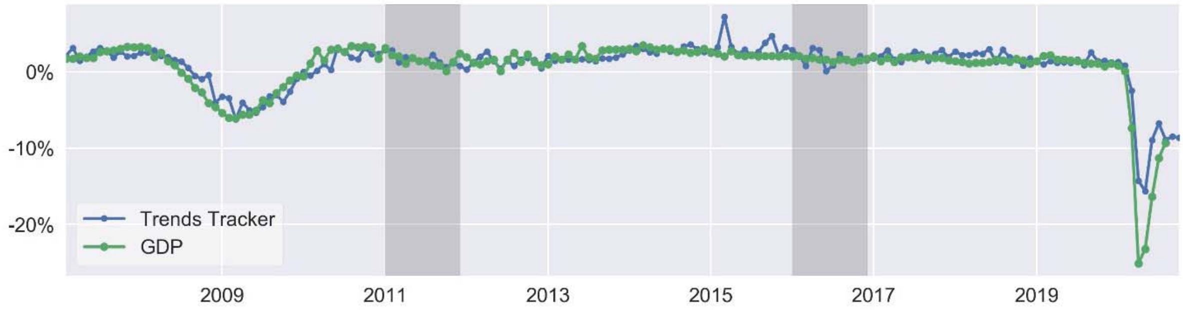
Nowcasting GDP in growth rate compared to the same quarter of the previous year, seasonally adjusted, from 2006 Q3 to 2020 Q2.

Panel A. United States

[10] For the G7 countries, the improvement in the RMSE relative to the use of an autoregressive model is even larger, at 26%.

Figure 2. Nowcasting GDP growth with Google trends (M-1 forecast) (*contd.*)

Panel B. United Kingdom



Panel C. Spain

Panel D. Italy

Panel E. Germany

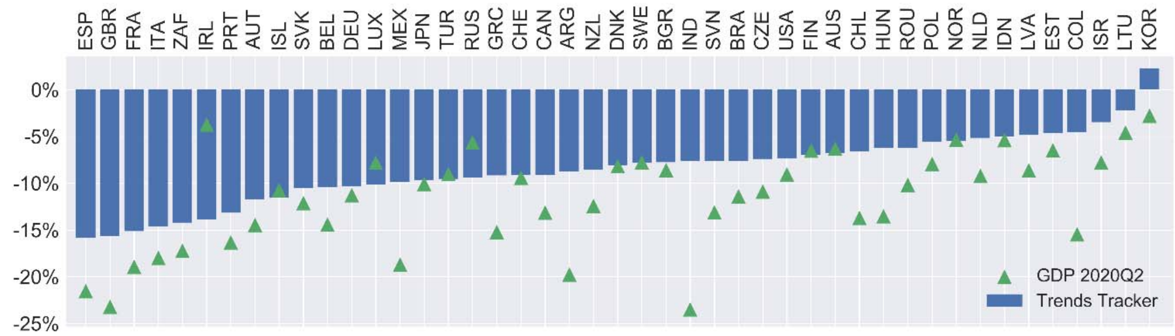
Figure 2. Nowcasting GDP growth with Google trends (M-1 forecast) (*contd.*)

Panel F. France

Note: The quarterly model is applied to 3-month moving averages of Google Trends series and yields monthly estimates that can be compared to quarterly GDP growth for February (Q1), May (Q2), August (Q3) and November (Q4). Shaded areas in 2011 and 2016 are years when the tracker is unavailable due to structural breaks in Google Trends data preventing the calculation of year-on-year growth rates in search intensities. Simulations are based on the latest GDP data, not the real-time vintages. For each quarter, the forecast is made 5 days after the end of the month, so 3-7 weeks before the GDP is published.
Source: Google Trends; OECD Quarterly National Accounts; and OECD calculations.

33. Figure 3 provides a closer focus on the COVID-19 crisis. The Weekly Tracker captures on average 60% of the fall observed in the second quarter. The tracker captures the sign of GDP growth accurately for all countries but Korea. The country ranking is relatively accurate. The extreme nature of the COVID-19 shock makes it challenging to predict the resulting growth dynamics based on history. This good overall performance reflects the ability of Google Trends to capture in a qualitative way the rapid change of economic situation during the second quarter and to translate this into a meaningful number.

Figure 3. Tracker's predictions for Q2 2020



Source: Google Trends, OECD Quarterly National Accounts and OECD calculations.

Table 2. Forecast performance

	2006-2020			2008-2010			2020		
	Unadjusted	Standardised	Relative to AR(4)	Unadjusted	Standardised	Relative to AR(4)	Unadjusted	Standardised	Relative to AR(4)
Poland	2.16	0.92	0.39	3.33	1.43	1.27	4.23	1.81	0.57
US	1.37	0.62	0.49	2.06	0.93	0.87	3.05	1.38	0.44
Japan	1.97	0.70	0.56	3.30	1.17	0.61	1.24	0.44	0.22
Belgium	1.58	0.59	0.57	2.42	0.91	1.92	4.92	1.85	0.64
South Africa	1.76	0.93	0.61	1.92	1.02	1.01	3.26	1.72	9.39
Slovak Rep.	3.05	0.70	0.61	3.65	0.84	0.40	4.73	1.09	0.63
Hungary	2.26	0.60	0.66	2.73	0.73	0.78	6.32	1.68	0.62
Netherlands	1.35	0.56	0.68	1.48	0.61	0.45	4.19	1.74	0.68
Lithuania	3.36	0.60	0.69	5.28	0.94	0.61	3.18	0.57	0.77
Chile	2.05	0.60	0.70	2.47	0.73	0.94	6.06	1.78	0.54
Mexico	2.26	0.60	0.71	2.79	0.73	1.17	7.13	1.88	0.59
Bulgaria	1.80	0.64	0.73	2.53	0.90	0.54	2.35	0.83	5.28
Germany	1.77	0.58	0.73	2.94	0.96	0.75	2.62	0.85	0.37
Norway	1.20	0.67	0.74	1.29	0.73	0.92	1.29	0.72	0.30
Canada	1.52	0.57	0.75	1.50	0.57	0.93	4.35	1.65	0.51
Denmark	1.78	0.69	0.76	2.69	1.04	0.88	1.49	0.58	0.26
Spain	2.19	0.55	0.79	1.46	0.36	0.53	8.41	2.10	0.78
Austria	1.67	0.59	0.79	3.09	1.10	1.27	3.48	1.24	0.46
Brazil	2.14	0.62	0.81	2.59	0.75	0.53	0.22	0.06	0.11
Slovenia	2.42	0.62	0.81	3.31	0.85	0.88	5.15	1.32	1.04
UK	2.31	0.63	0.82	1.86	0.51	1.54	9.78	2.68	0.74
Portugal	2.17	0.65	0.83	2.62	0.78	1.89	4.91	1.47	0.51
Finland	2.26	0.67	0.87	4.60	1.36	0.99	1.46	0.43	0.49
Italy	1.98	0.61	0.88	2.83	0.87	0.85	5.22	1.61	0.69
Indonesia	1.17	0.72	0.89	1.45	0.89	1.61	2.19	1.35	0.38
France	1.72	0.54	0.89	2.08	0.66	1.38	6.50	2.05	0.78
Sweden	1.88	0.62	0.90	2.86	0.95	0.89	2.47	0.82	0.37
Turkey	3.62	0.78	0.90	4.91	1.05	0.59	1.57	0.34	0.44
Estonia	4.76	0.80	0.91	9.28	1.57	0.81	0.17	0.03	0.04
India	2.17	0.92	0.92	2.68	1.13	0.56	4.44	1.88	2.37
Czech Rep.	1.93	0.58	0.94	1.90	0.57	0.60	4.80	1.44	0.88
Latvia	3.90	0.60	0.96	6.82	1.05	0.81	4.29	0.66	0.78
Israel	1.90	0.84	1.03	2.10	0.93	0.94	2.98	1.32	0.49
Iceland	3.76	0.86	1.08	5.79	1.33	1.03	0.75	0.17	0.16
Luxembourg	3.10	0.89	1.08	5.15	1.48	1.92	2.45	0.70	0.69
New Zealand	1.13	0.70	1.12	1.30	0.80	0.98	0.83	0.52	0.42
Argentina	4.14	0.78	1.14	5.70	1.07	0.97	3.57	0.67	0.62
Ireland	6.46	0.84	1.15	4.54	0.59	0.54	6.93	0.90	12.69
Switzerland	1.44	0.88	1.18	2.21	1.35	0.99	1.39	0.85	0.41
Australia	1.27	1.52	1.20	2.07	2.47	2.47	1.73	2.07	1.50
Korea	2.34	1.13	1.21	2.66	1.29	0.72	2.18	1.06	0.57
Greece	3.02	0.73	1.22	3.68	0.88	1.07	3.77	0.91	1.77
Russia	2.67	0.66	1.34	5.15	1.27	1.28	1.72	0.42	6.73
OECD - G20	2.14	0.66	0.83	2.69	0.93	0.92	3.26	1.09	0.59

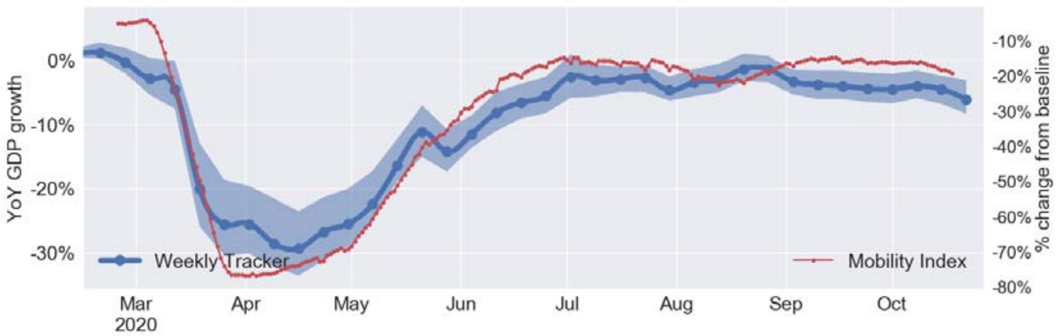
Notes: The first column (*Unadjusted*) indicates root mean square error in forecast simulations. RMSEs are standardised by the GDP growth series standard deviation in the second column (*Standardised*), and divided by the RMSE of an AR(4) model in the third one [*Relative to AR(4)*].The sample includes OECD and G20 countries but the EU, China and Saudi Arabia. The "OECD – G20" row provides the median of performance indicators over the 46 countries. Simulations for the full sample are performed over a segment starting in September 2006 and ending either in Q1 2020 or Q2 2020 depending on official GDP figures availability. Simulations around the time of the GFC are performed over 01-2008 to 12-2009. Simulations in 2020 cover Q1 and Q2.
Source: OECD calculations.

34. The Weekly Tracker model is further validated by a close correlation with weekly movements in mobility. Valid pseudo-real time simulations run at the quarterly frequency do not imply that the Weekly Tracker accurately captures short-term activity variations within quarters. The frequency-neutrality

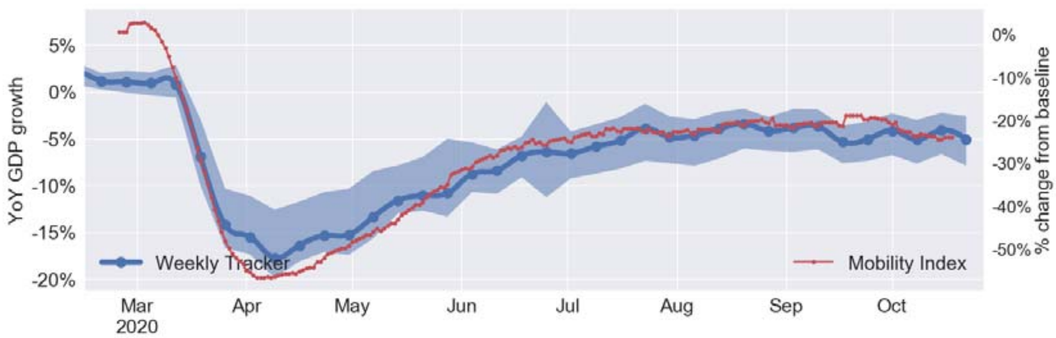
assumption is thus further assessed in Figure 4, which compares the Weekly Tracker to the Google Mobility Index derived from Google Maps data which was proved to track activity well (Fernández-Villaverde and Jones, 2020[37]; IMF, 2020[38]; OECD, 2020[9]). The relative magnitudes of variations in GDP growth seem well captured by the Weekly Tracker, as well as the timing of the fall in activity in March. In France, Canada, New Zealand, and the United Kingdom, the drop in the Weekly Tracker and in the Mobility Index are simultaneous, while in Japan the Weekly Tracker leads the Mobility Index. The timing and relative magnitude of the evolutions of the Weekly Tracker and the Mobility Index around the rebound in May-July are also very close.

Figure 4. The OECD Weekly Tracker and Google Mobility

A. France



B. Canada



C. New Zealand

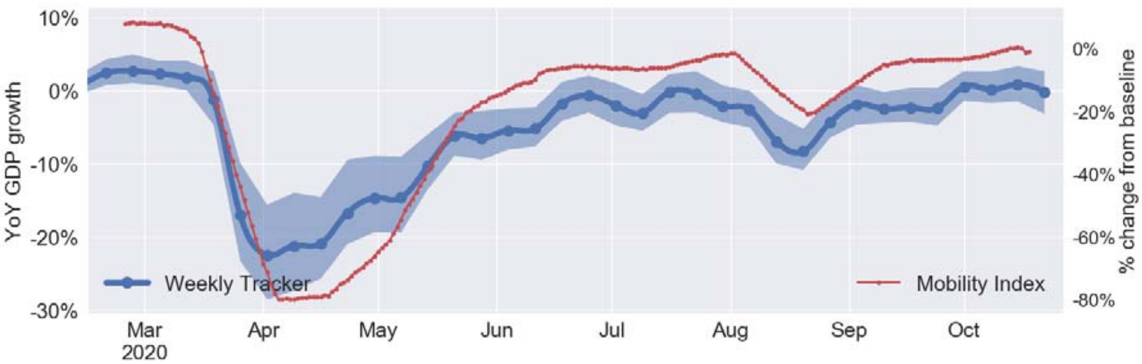
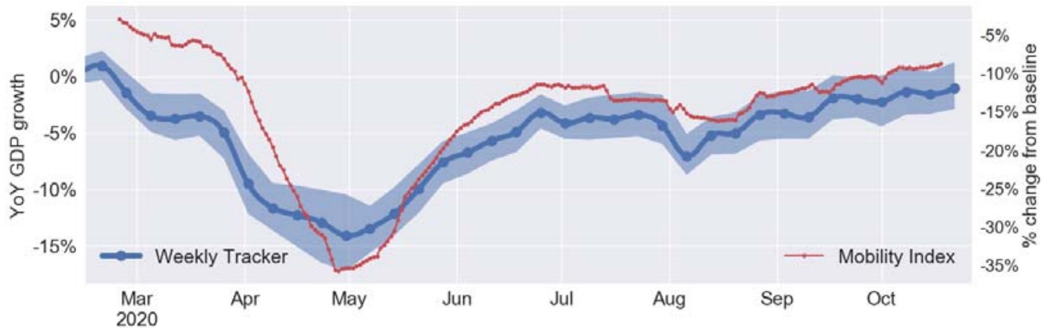
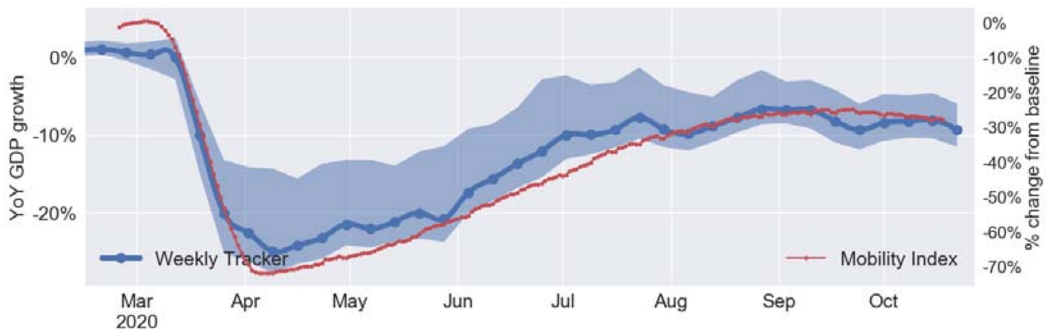


Figure 4. The OECD Weekly Tracker and Google Mobility (*contd.*)

D. Japan



E. United Kingdom



Note: The Mobility Index (red line, right axis) is the average of the Google Mobility indices for work and leisure. The OECD Weekly Tracker (blue line, left axis) is predicted by a model trained on quarterly observations (and monthly observations for the UK).
Source: OECD Weekly Tracker; Google Mobility reports.

6. Model insights: a dive into the black box

35. Whereas machine-learning algorithm are sometimes seen as "black boxes", recent research has strived to provide interpretability techniques to "open the black box" (Renard et al., 2019[39]; Joseph, 2019[40]; Ribeiro, Singh and Guestrin, 2016[41]; Zhao and Hastie, 2019[42]; Craven and Shavlik, 1996[43]; Laugel et al., 2019[44]). Interpretability techniques can either aim at explaining given predictions (local interpretability) or the general functioning of a model (global interpretability). Understanding the drivers of the predictions made by the neural network behind the Weekly Tracker is key to ensure that the model is consistent with economic intuition and does not rely on spurious patterns. This section also sheds light on insights derived from interesting patterns learnt by the algorithm.

6.1. Shapley values: explaining machine learning with game theory

36. A recent tool ('SHAP') (Lundberg and Lee, 2017[45]) has already become an industry standard (Tiffin, 2019[46]; Joseph, 2019[40]) by providing both local and global interpretability. This method decomposes the predictions made by any algorithm into variable contributions (their "Shapley values"). It uses Shapley values, a method from coalitional game theory designed to fairly distribute a 'pay-out' from

a multi-player game[11]. In this case, the pay-out is the prediction minus its average $\hat{y} - E(\hat{y})$, and the players are the variable values. The Shapley value is the average marginal contribution of a variable value to the prediction over all possible "coalitions". A coalition is defined as a number of variables taking the value that is observed rather than their average or any arbitrary value. Shapley values sum to the model prediction. With a linear model, the Shapley value for observation i and variable x_j is simply equal to $\beta_j x_{ij} - \beta_j E(x_j)$. In the context of a model that captures multiple interactions, SHAP provides variable contributions all else equal, by averaging over the various possible variable combinations.

37. Shapley values are a powerful interpretability tool. The Shapley value is the only attribution method that combines the following properties: efficiency (Shapley values sum to the prediction minus its average), symmetry (two variable values have the same Shapley value should they contribute equally to all coalition), dummy (a variable value with no impact on the prediction whatever the coalition has Shapley value equal to zero) and additivity. The Shapley values are based on a mathematical theory and distribute the variable contributions 'fairly'. Decomposing a given prediction into Shapley values provides local interpretability. Conversely, Shapley values for a given variable can be plotted against that variable (which gives a partial dependence plot) to provide global insights on the model.

6.2. A dive into the model inner workings

38. Figure 5[12] shows that variables with the largest contributions, that is, most important variables in the model, are searches corresponding to "unemployment", "investment" and "student loan". The contribution of the topic "Unemployment" is highly negative when the search intensity is high, and around zero for lower search intensities. Search topics are high in the ranking (unemployment, investment, crisis, recession, economic crisis, exportation, bankruptcy, mortgage, interest, luggage, recruitment), although there are initially only 33 topic-based variables against 215 category-based variables, thus providing additional evidence to the relevance of using topics.

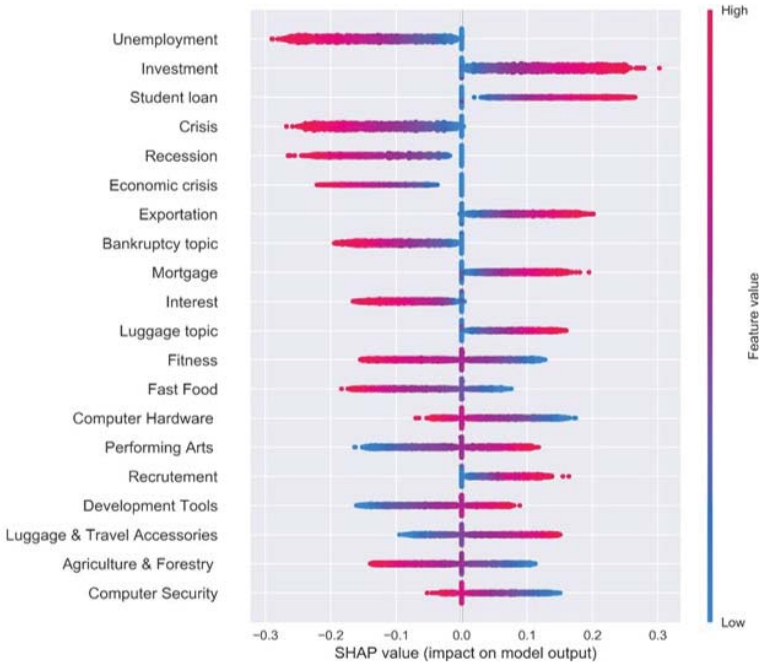
39. Results are consistent with intuition. For instance, higher searches around the topic "investment" may indicate wealth effects and signal higher GDP growth. Topics indicating economic anxiety ("crisis", "recession", "economics crisis") are strong predictors of lower growth. They may also reflect media coverage[13]. Searches around the bankruptcy topic may be related to information-seeking by individuals or firms going bankrupt, as well as attention towards large bankruptcies (such as Lehman Brothers'). Consumption items are also important predictors, as they reflect consumption behaviours ('luggage', 'fitness', 'performing arts'). Lastly, a few variables may provide information on firm behaviour ('recruitment', 'computer hardware', 'development tools').

[11] An intuitive explanation of Shapley values is provided in (Molnar, 2020[54]).

[12] Shapley Values are computed using the Python library "SHAP", built by Lundberg and Lee.

[13] From this point of view, Google Trends appears as an alternative to news-based quantitative analysis as internet searches follow (or drive) media attention.

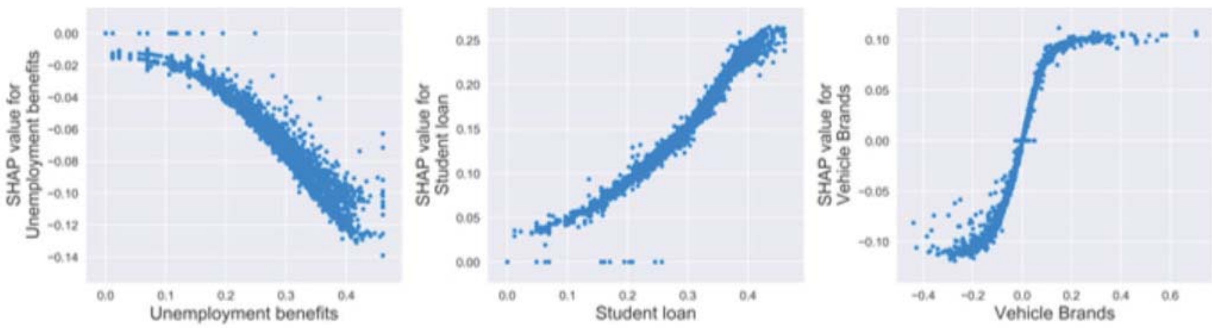
Figure 5. Most important variables and their contributions to predictions



Note: Shapley values are the contributions of a variable to the GDP growth estimate predicted by the model. Variables are ranked by importance, and for each variable. Each point correspond to an observation (that is a given month * a given country) and its colour depends on the value of the variable.
Source: OECD calculations

40. Figure 6 sheds light on selected non-linearities using partial dependency plots based on Shapley values. In a linear context the plot would show a straight line with slope β. In the present case, the neural network captures non-linearities, hence the non-linear shape of the partial dependence curves. Interestingly, the elasticity of the GDP growth to searches for unemployment benefits captured by the model is lower (the slope is flatter) on the left of the panel, and higher on the right when the search intensity is higher. This pattern suggests that searches for unemployment benefits are stronger predictors of activity around times when lay-offs increase and thus become dominant with regards to hiring in explaining changes in employment.

Figure 6. Partial dependence plots



Note: Partial dependence plots show Shapley values of a given variable against that variable value. Each dot corresponds to an observation in the data. The X-axis shows the log (first two panels) or log differenced (third panel) search index for a topic, and the Y-axis the contribution of that topic to the prediction ($\hat{y}$) made by the algorithm. In a linear model, the curve would be straight.
Source: OECD calculations.

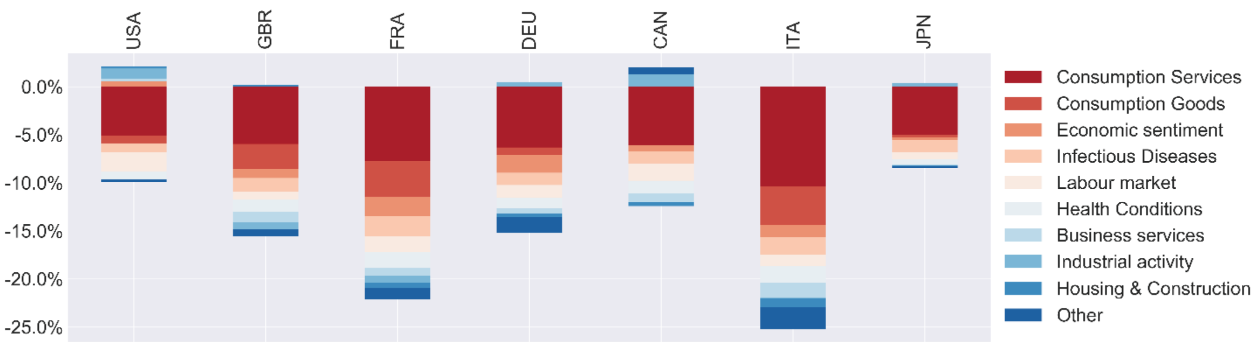
6.3. From Shapley values to sectoral insights

41. Figure 7 provides insights on the predictions made for March 2020. It shows the contributions of the variables most important for the predictions across countries. Google Trends variables have been aggregated into a small number of economically significant groups for the sake of readability. Interestingly, topics related to various consumption items (e.g., food and drinks, vehicle brands) are the main drivers of the low estimates. Topics indicating economic anxiety ('crisis', 'economic crisis', 'recession', 'economic news') have the second largest contributions across countries. They capture the intensity of the economic fear, probably capturing views shared in the public sphere. Variables associated with unemployment benefits or unemployment also have a major contribution. This may reflect individuals losing their jobs who look up the internet for how to receive benefits.

42. Variables related to business services (e.g., data management, accounting, consulting) and industrial activity (e.g., agriculture and forestry, manufacturing, mail and package delivery) have smaller contributions. This probably reflects the fact that the supply side may not captured by Google Trends as well as consumption, and should not be taken at face value. As the algorithm tries and maximises the accuracy of the prediction, the weights it gives to industrial activity results from a trade-off between the actual weight of industrial activity in explaining GDP and the relative quality of the signal it gets for industrial activity, which is probably noisier than information about consumption behaviours.

Figure 7. A focus on March 2020

Contributions of the main common variables to the prediction for 2020 Q2, G7 countries



Note: Bars show Shapley values for the prediction made for 2020 Q2. Google Trends variables are aggregated together into significant groups detailed in Annex B.
Source: Google Trends and OECD calculations.

7. The OECD Weekly Tracker

43. The model of GDP growth described in Section 4. was trained and assessed using data at a quarterly frequency, and is being used in this section to provide a weekly index of economic activity. The Weekly Tracker fully exploits the timeliness and granularity of the Google Trends data, and provides weekly information on business cycles in real time.[14]

44. The model is estimated using all known GDP growth rates by the time of the writing. The weekly series are subject to the same pre-processing that is described in Section 2, except there is no need to filter out the common long-term trend (as only the past five years are available). The weekly series are

[14] The Google Trends data come with a 5-day delay.

calibrated on the monthly series and the yearly log differences are computed as described above. Confidence bands are derived using bootstrap. The model is trained on 1300 samples drawn with replacement from the data. The resulting 1300 predictions are used to compute the 90% confidence intervals (CI).

45. The OECD Weekly Tracker provides early and timely insights on economic activity during the COVID-19 crisis and subsequent recovery. The magnitude of the shock to economic activity in March was extreme, as confirmed by GDP figures for Q2. The Trackers suggest that in a number of countries there was a rebound in April and May, with impetus slowing from June.

7.1. The COVID-19 crisis: a week-by-week analysis

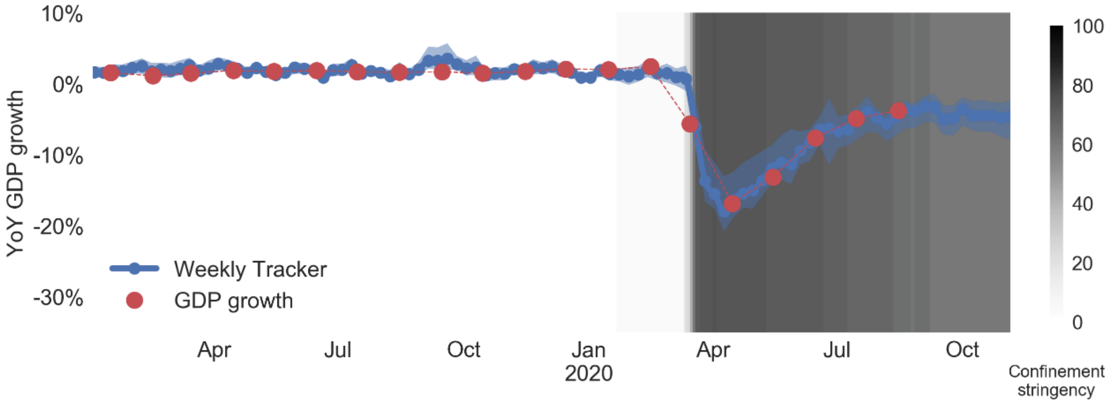
46. The OECD Weekly Tracker suggests that this crisis caused major fluctuations in economic activity which were too abrupt to be captured by monthly indicators. Over the preceding years, 2017-19, a high-frequency proxy of GDP growth would not have added much useful information content (Figure 8). However, in 2020, changes in economic activity were more rapid and pronounced, providing a clear advantage from having a weekly proxy for GDP. During the course of March 2020, the Weekly Tracker suggests that for the United States, year-on-year GDP growth fell from +2.4% during the first week to -10.2% in the last week, before reaching -14.7% in mid-April. In India, it fell from +1.6% in the second week to -15.3% in the last week of March, declines of a magnitude later corroborated by actual industrial production figures (-16.3% year-on-year in April). The shock was also particularly sudden in many major European economies: for example, in the United Kingdom the Weekly Tracker suggests annual GDP growth fell from +0.37% to -20% in the course of March, reaching -24% in mid-April. In contrast, in addition to being subject to longer publication delays, lower-frequency indicators provide a more distorted picture of both the pattern of the downturn and the recovery dynamics, when activity is changing rapidly.

47. The OECD Weekly Tracker suggests that the immediate impact on GDP of the global pandemic was particularly heterogeneous across advanced economies (AEs) (Figure 8). In France and Italy, where particularly stringent lockdowns were implemented, activity is estimated to have fallen suddenly by around 29% below its 2019 level by early April (which is broadly consistent with GDP outturns for the second quarter). In countries where the lockdowns were less stringent, activity is estimated to have fallen slightly less abruptly: by 25% in the United Kingdom and by around 13-17% in Germany, Japan, Canada and Australia (again broadly consistent with GDP outturns for the second quarter). Korea, where epidemic control relied more on track-and-test than lockdown policies, had the lowest short-term drop, with the proxy measure of weekly GDP only falling by 4% below a year earlier in the worst week of April. While there is a clear impact from exiting lockdowns, the Weekly Tracker suggests the recovery in economic activity was much more gradual than following the initial impositions.

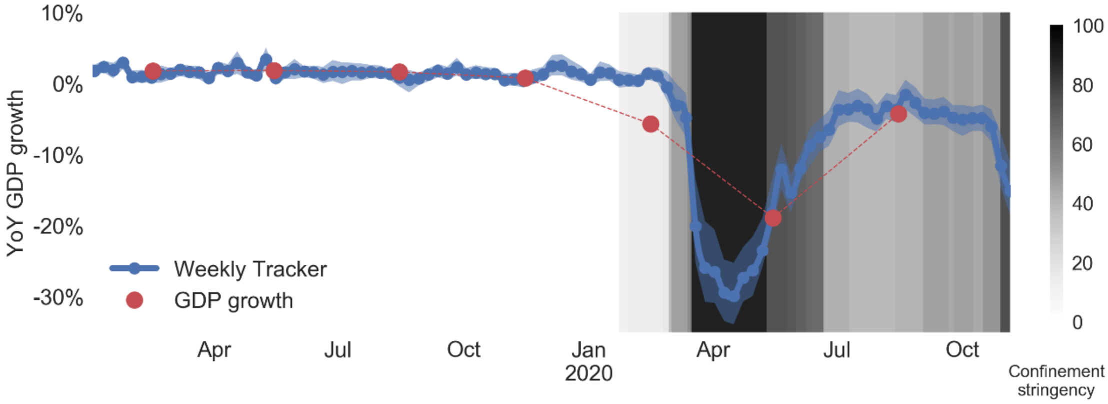
Figure 8. Weekly Tracker: advanced economies

Model estimates of "weekly GDP" growth with regard to same week of previous year

A. Canada



B. France



C. Germany

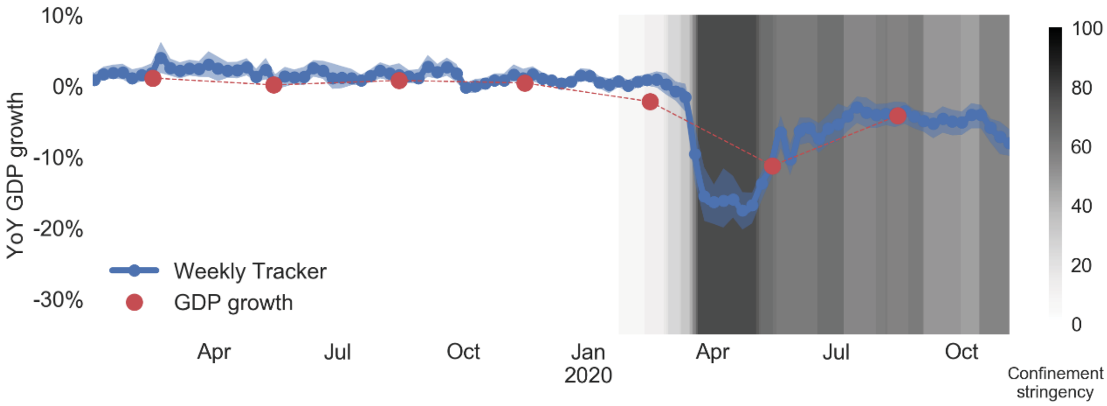
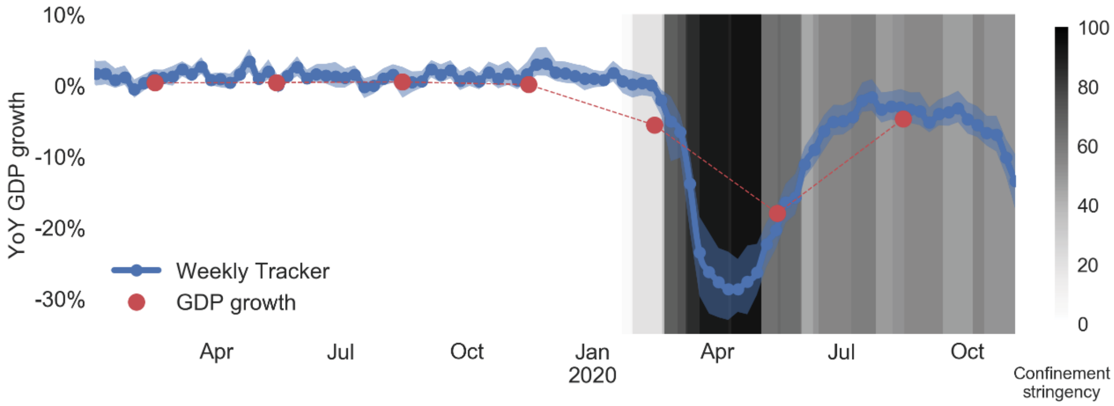
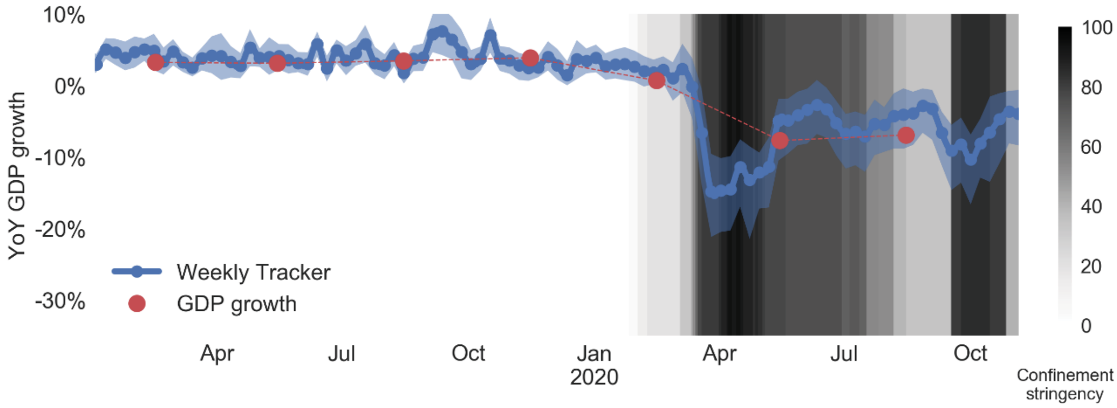


Figure 9. Weekly Tracker: advanced economies (*contd.*)

D. Italy



E. Israel



F. Japan

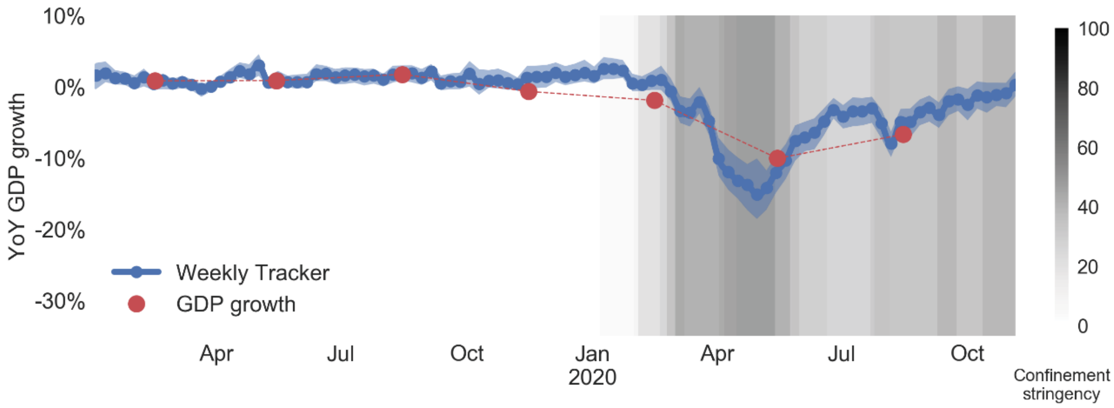
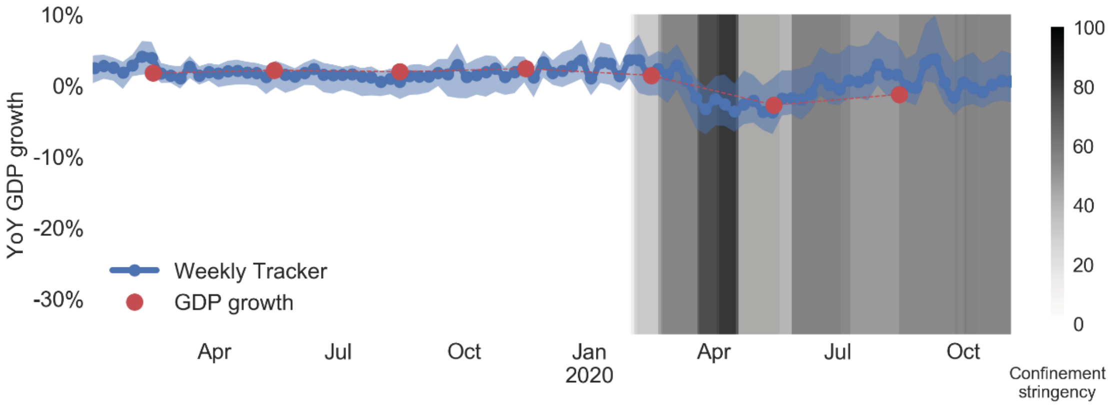
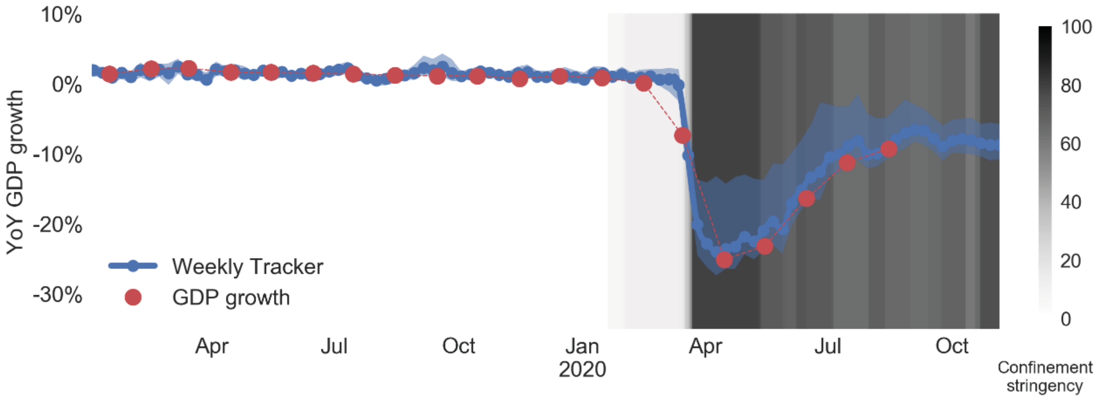


Figure 9. Weekly Tracker: advanced economies *(contd.)*

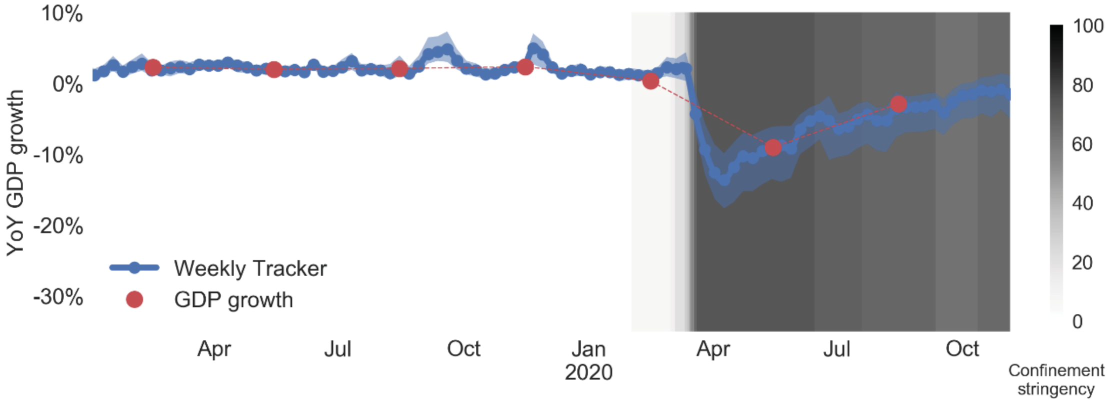
G. Korea



H. United Kingdom



I. United States



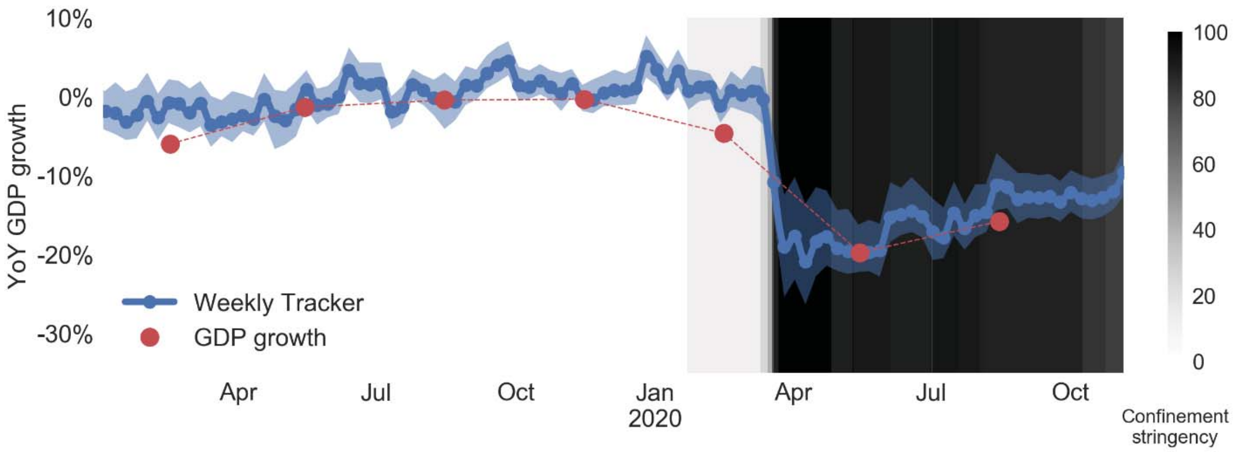
Note: The blue confidence band shows 95% confidence intervals. Red dots representing GDP growth are official outturns except 2020 Q3 for Israel, which is the *Economic Outlook* projection. Monthly GDP growth series are used when available (for the United Kingdom and Canada). The darkness of the grey background reflects the strictness of the confinement measures based on the OECD COVID-19 tracker (darker = stricter).

Source: *OECD Economic Outlook* 108 database; OECD Weekly Tracker; UK Office for National Statistics; StatCan; and Oxford COVID-19 Government Response Tracker.

48. Many emerging-market economies (EMEs) exhibit a similar sudden fall in activity based on the Weekly Tracker, although the rebound differs widely across countries (Figure 9). The initial shock to activity is estimated to be particularly strong in India (-20%), Mexico (-19%), South Africa (-19%), Argentina (-18%), Turkey (-15%) and Brazil (-13%) with regards to the same weeks of 2019. Russia and Indonesia were hit less hard, as the Weekly Tracker suggests that activity at the trough was around 11% lower than in 2019. The fall in activity was particularly swift in Argentina and India, which implemented very stringent confinement policies.

Figure 9. Weekly Tracker: selected emerging economies

A. Argentina



B. South Africa

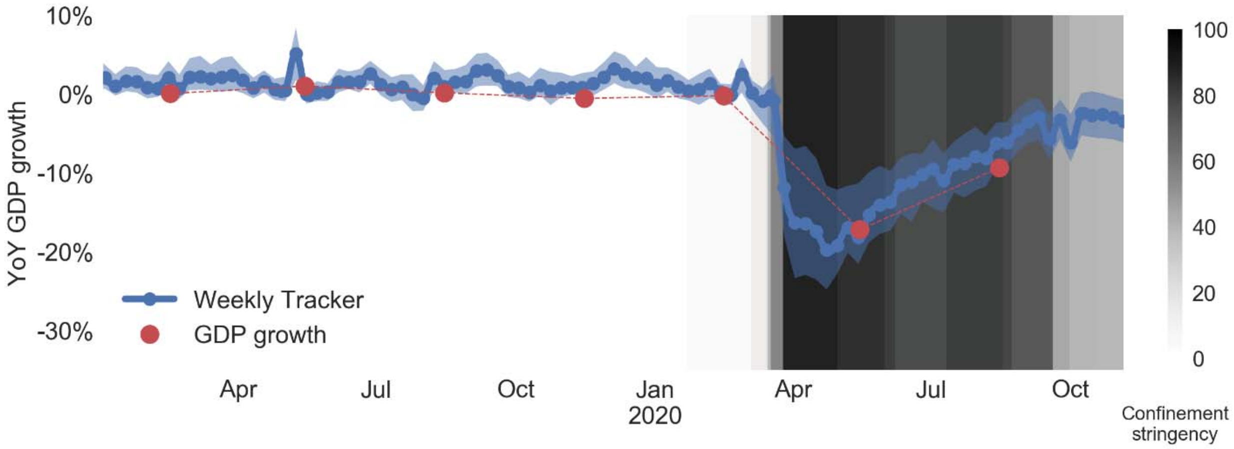
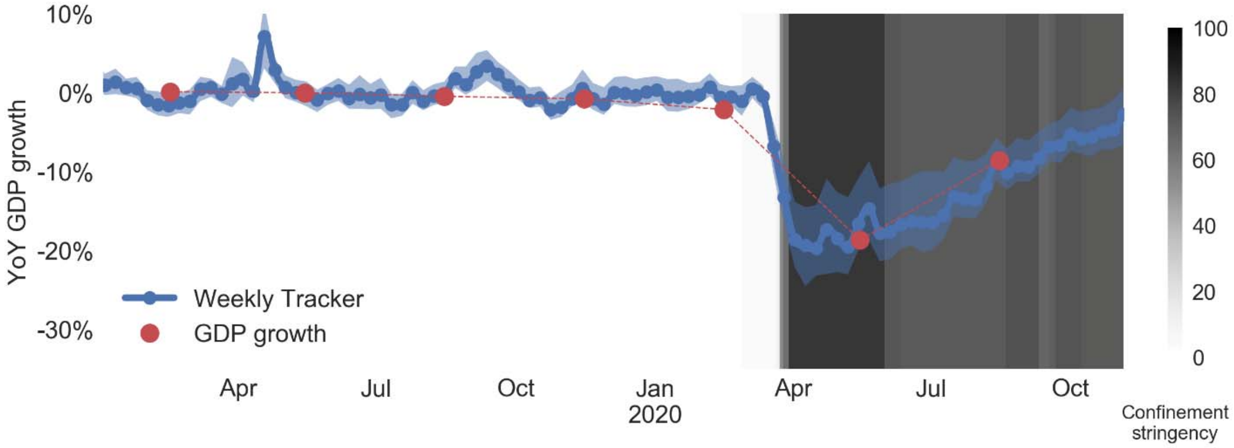
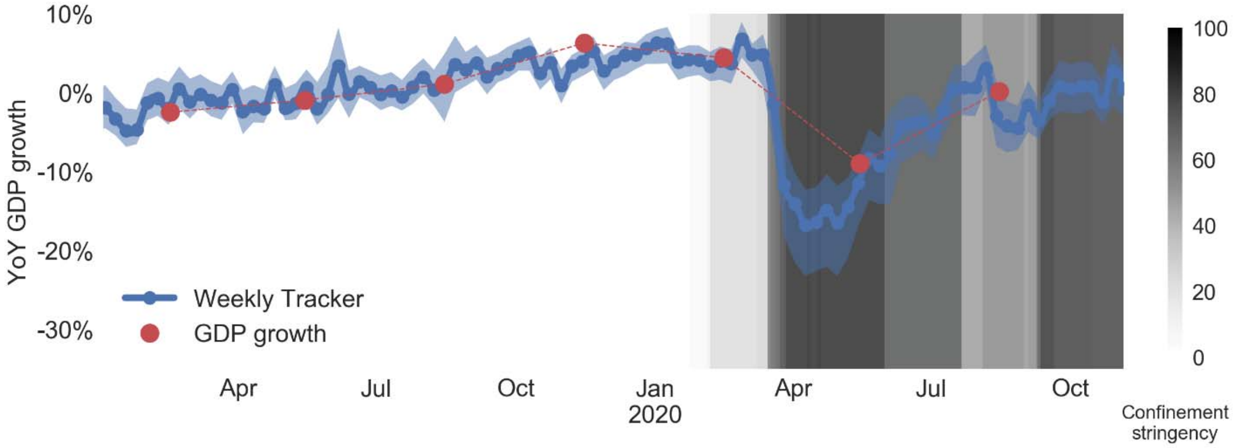


Figure 9. Weekly Tracker: selected emerging economies (*contd.*)

C. Mexico



D. Turkey



E. Brazil

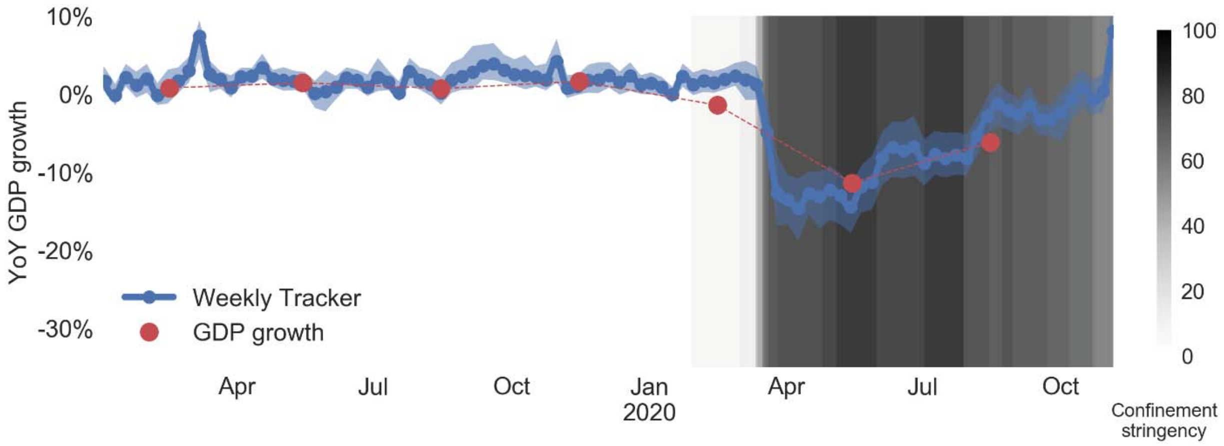
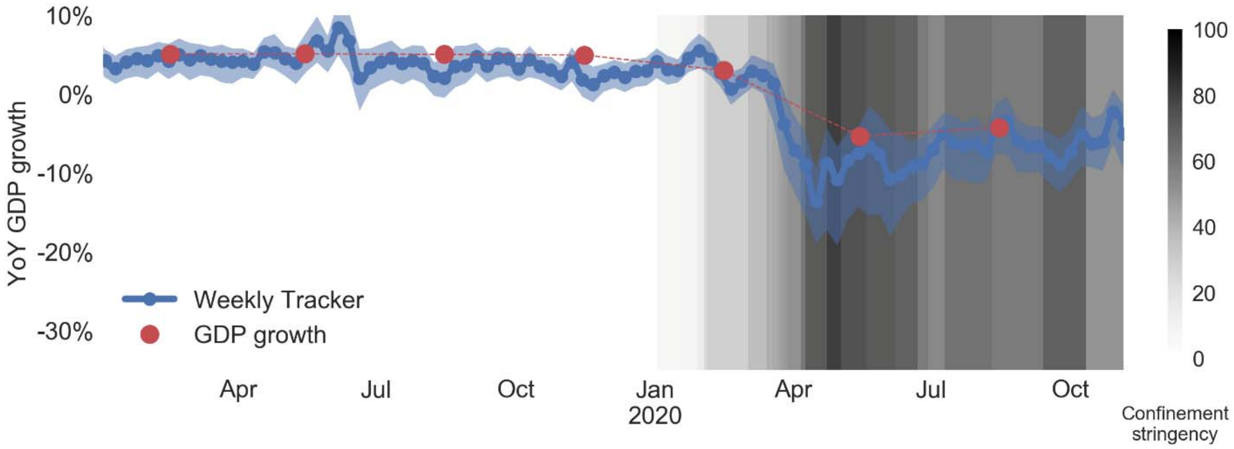
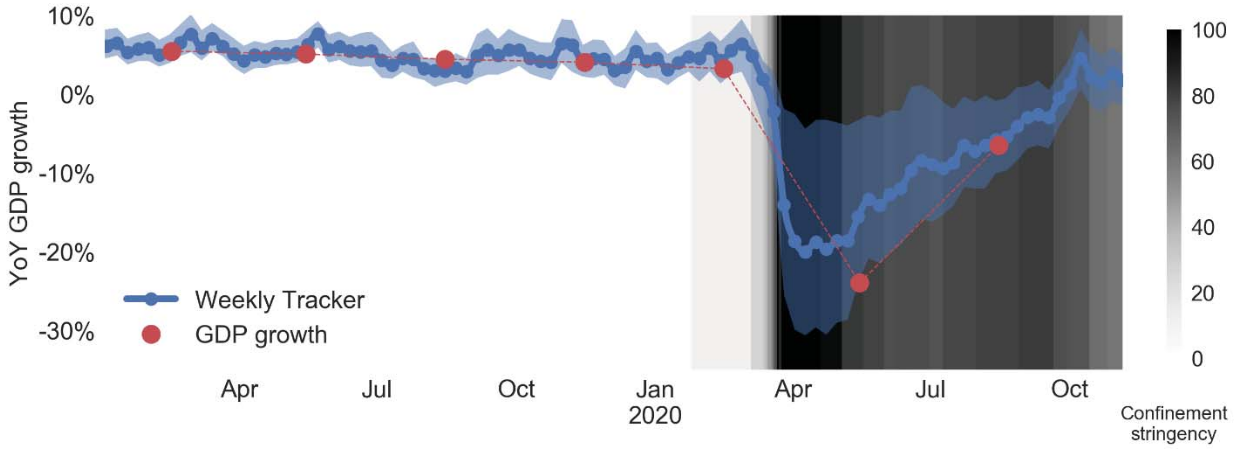


Figure 9. Weekly Tracker: selected emerging economies (*contd.*)

F. Indonesia



G. India



Note: The blue confidence band shows 95% confidence intervals. Red dots representing GDP growth are official outturns except 2020 Q3 for India, Brazil, Turkey, South Africa and Argentina, which is the *Economic Outlook* projection. Monthly GDP growth series are used when available (for the United Kingdom and Canada). The darkness of the grey background reflects the strictness of the confinement measures based on the OECD COVID-19 tracker (darker = stricter).
Source: OECD Economic Outlook 108 database; OECD Weekly Tracker; UK Office for National Statistics; StatCan; and Oxford COVID-19 Government Response Tracker.

7.2. Latest insights from the Weekly Tracker: a stalling recovery below 2019 levels

49. The OECD Weekly Tracker indicates that the rebound started to slow in June, with the most recent estimates implying that activity was stagnating in the third quarter well below levels achieved in 2019 for most countries (Figure 10, Panel A). The out-of-sample performance of the Weekly Tracker for Q3 appears creditable when compared to available GDP outturns for Q3, given the very volatile environment. Across the 22 countries where GDP growth for Q3 had been released at the time of finalising this note, the mean average error in predicting year-on-year GDP growth was around one percentage point with no evidence of systematic bias, compared with actual falls in GDP for the median country of nearly 5 percentage points and variation in quarter-on-quarter growth of between 2 and 18 percentage points across countries. On the basis of the Weekly Tracker, the rebound was particularly weak in Argentina, where activity in Q3 is

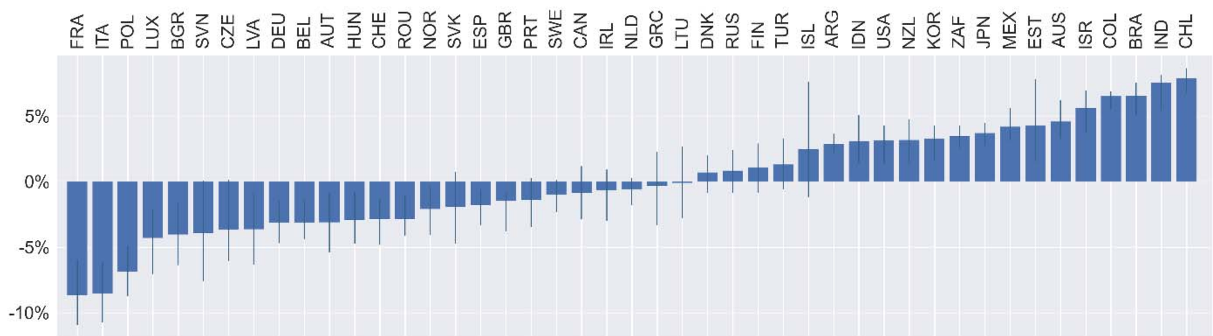
estimated to be around 15% lower than its 2019 level, as well as Mexico, the United Kingdom, Colombia and Spain, with activity estimated around 8-10% lower than 2019 levels.

50. The OECD Weekly Tracker also provides some insight as to which countries have strongest momentum on activity in the fourth quarter, based on results from the Tracker up until the second week of November (Figure 10, Panel B). It suggests that quarterly growth will be negative in many European countries where the stringency of lockdown measures has recently been tightened. By contrast, the Tracker suggests that many non-European G20 countries will have positive growth at least over the first half of the quarter, reflecting some loosening of lockdown stringency, especially in Chile, Argentina, Brazil, India and South Africa, or maintenance of a low level of lockdown stringency. In some countries, including Chile, India, Brazil and Korea, this rebound is predicted to result in the level of GDP in mid-November being higher than a year earlier.

Figure 10. Most recent predictions of the OECD Weekly Tracker

Panel A. Tracker prediction of year-on-year GDP growth for Q3

Panel B. Most recent evolution of the Tracker prediction: change between Q3 and mid-November



Note: In Panel A, the blue bars represent out-of-sample model projections of year-over-year GDP growth based on Google Trends and the black lines represent 95% confidence intervals around them. The triangles are GDP outturns for Q3. In panel B, the blue bars represent the difference between the average Tracker value over the first two weeks of November and Q3, while the black lines represent 95% confidence around them.
Source: *OECD Economic Outlook* 108 database; Google Trends; and OECD Weekly Tracker.

7.3. Consumption volume remains subdued while its composition has shifted

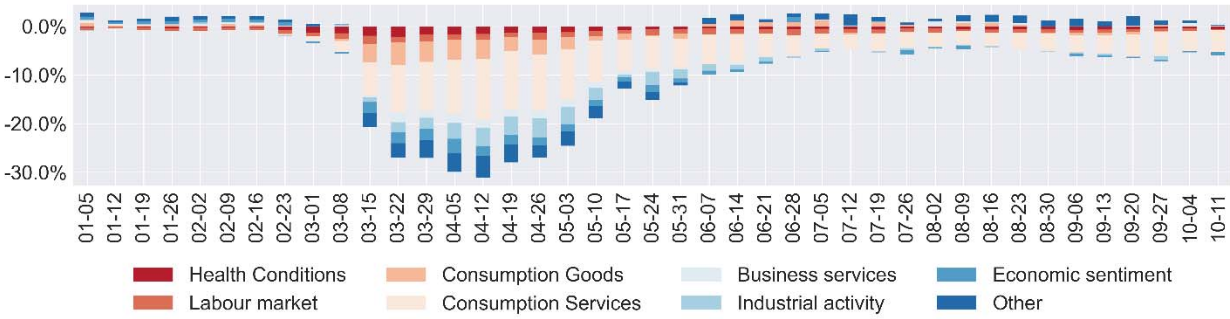
51. Sectoral insights can be derived from decomposition in variable contributions. The tracker predictions are decomposed into Shapley Values (see section 6). Variables with largest Shapley Values are aggregated to form economically relevant groups reflecting large economic sectors. These decompositions explain mechanically contributions to the indicator rather than forecasting the composition

of GDP. They are biased towards consumption as Google Trends provides more information relevant to GDP prediction about consumption than other GDP components. In a linear model, a noisier signal about, say, investment would also result in a weight (β) smaller than its share in GDP. Still, the evolution of contributions over time are informative about economic developments.

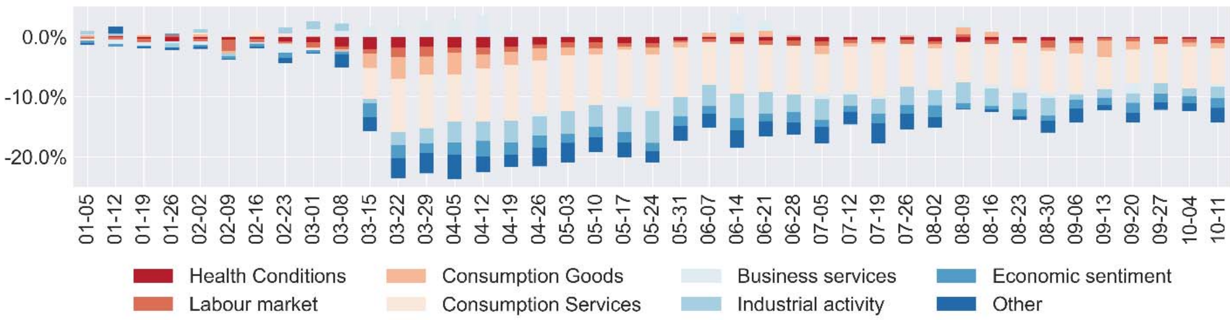
52. This exercise suggests that the tracker is mostly driven by searches around various consumption services with searches for consumption goods also continuing to weigh on the indicator, where there are small positive effects from credit and housing (Figure 11). Searches for consumption goods had a negative impact around the trough in both France and Argentina but much less since May. These two countries are illustrative of a U- and L-shaped recoveries, respectively. The differences in annual growth rates between France and Argentina since then seem mostly explained by searches for consumption services, which have a small negative contribution in France and a large negative contribution in Argentina. These results suggest that households curbs in services consumption reflects virus circulation and fear of the virus.

Figure 11. Drivers of the recovery: aggregated Shapley Values

A. France



B. Argentina



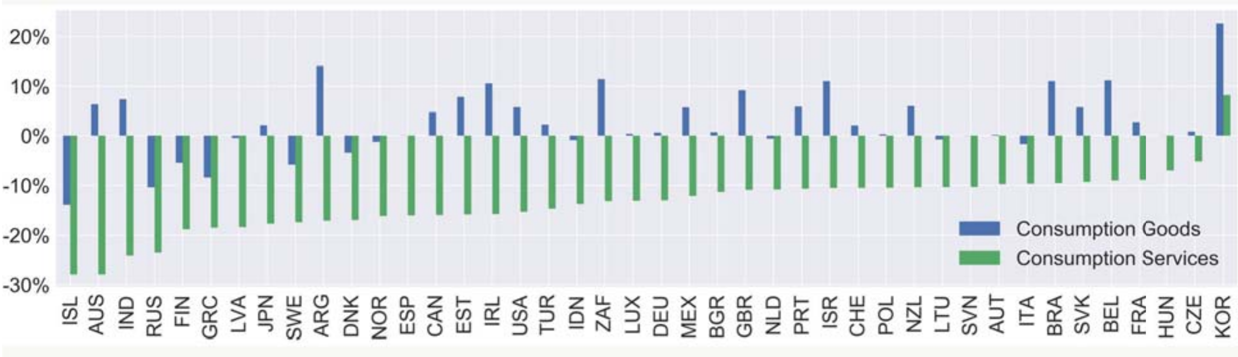
Note: The weekly tracker predictions are decomposed into variable contributions using Shapley Values. Variable contributions are aggregated into economically relevant subgroups reflecting key economic sectors. These contributions explain mechanically contributions to the indicator rather than forecasting the composition of GDP.
Source: Google Trends and OECD.

53. A dive into Google Trends search intensities suggests that consumption has decreased and shifted from services to goods. Figure 12 shows aggregated search intensities in yearly growth rates for consumption goods (including Food & Drink, Vehicle Brands, Energy & Utilities, Vehicle Shopping, Camera & Photo Equipment, Music Equipment & Technology, Home Appliances) and some consumption services (*e.g.*, Performing Arts, Travel, Sports, Restaurants, Arts & Entertainment). Search intensity for consumption services remains lower than a year earlier by 15% on average across OECD and G20

countries, while search intensity for consumption goods is higher than a year ago in most OECD and G20 countries (with a few exceptions).

Figure 12. Consumption has decreased overall and shifted towards new patterns

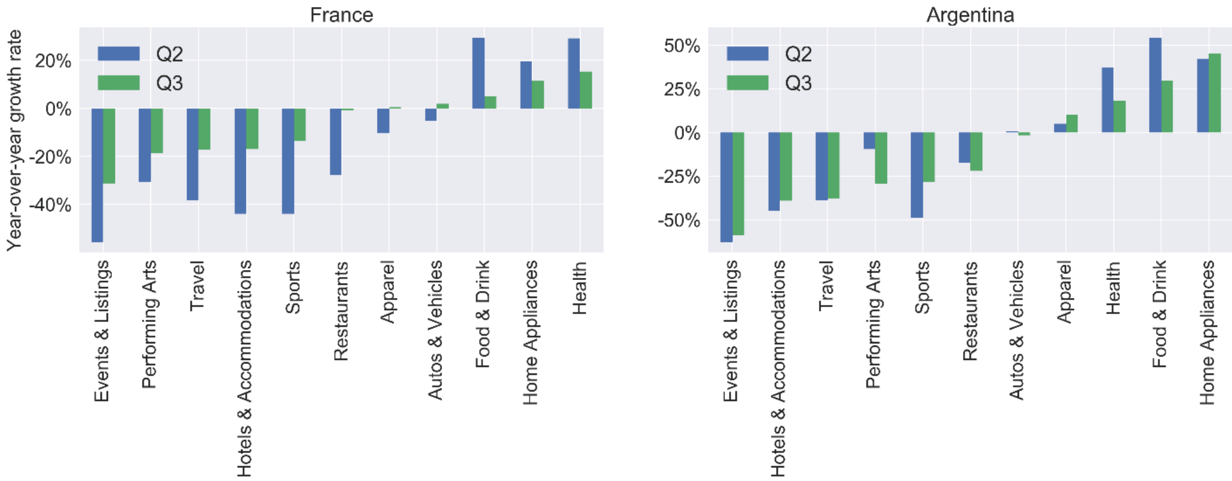
Aggregate search intensities for consumption goods and services in mid-August 2020, yearly growth rate



Note: Annual growth rates in aggregated search intensities for consumption goods and services observed in mid-August.
Source: Google Trends and OECD calculations.

54. The shift in consumption patterns can be documented at a more granular level by looking at the growth rate in search intensities of individual Google Trends variables included into the consumption goods and services groups. Figure 13 highlights the role of the fall of consumption of certain services in explaining the overall weakness in activity in France and Argentina, where the rebounds were particularly strong and weak respectively. In Q2, both countries experienced a strong shift in consumption patterns whereby search interest for interaction-based services (including events, performing arts, travel, hotels, sports and restaurants) decreased by around 30% while searches for food and drinks, household appliances and health-related issues increased by around 20%. Lower services consumption was only partially replaced by additional goods consumption resulting in lower overall spending, helping to explain negative model-estimates of year-over-year GDP growth. This pattern of distortion partly fades away in France in Q3, but not in Argentina, consistent with the different pace at which containment measures were relaxed. The potentially lasting effects of the virus circulation and mobility restrictions may thus explain part of the much weaker rebound in Argentina.

Figure 13. Google search intensities per spending categories



Note: Year-over-year growth rates in search intensities for selected search categories corresponding to spending categories (median over G20 and OECD countries).
Source: Google Trends and OECD calculations.

8. Conclusion

55. This paper describes the construction of the OECD Weekly Tracker of economic activity for 46 OECD, G20 and partner countries using Google Trends and a neural network algorithm. Simulations in pseudo-real time show that the tracker is a reliable predictor of business cycles in most countries. The Tracker is particularly useful around recessions and captures the COVID-19 downturn and subsequent rebound well. Looking inside the model with recent interpretability tools shows that predictions are based on patterns consistent with economic intuition. The algorithm captures non-linear patterns that related papers have shown to be especially important around crises. Using a panel specification allows for the use of complex algorithms such as neural networks. The paper also introduces a new method to address the downward long-term trend common to many Google Trends variables.

56. The paper sheds new light on the current crisis. The Tracker captures the COVID-19 recession in most countries, although it underestimates the depth of the downturn in the countries that endured the worst declines. It provides unique information on the timing of the crisis and on the magnitude of the rebound. The fall in activity leads the lockdowns in the United States, United Kingdom, Germany and Canada, while activity falls down at the exact time tighter lockdown measures were implemented in France and Italy. The rebound seems particularly strong in France, Germany and Eastern European countries and much slower in Spain, Japan and Italy.

57. Further research could explore the inclusion of other variables. Additional predictors could include other high-frequency time series, such as financial variables, commodity prices or electricity consumption. Another option would be to include lower-frequency variables (such as PMIs, which are monthly series), which would raise modelling questions on how to handle variables with mixed-frequencies on the right hand side of the equation when the goal is to make high-frequency predictions. It could be possible to build one model per week in a month. More refined options could leverage a single model that would capture the fact that the monthly PMI indicator should become less and less relevant as time increases since its last release.

References

Abay, K., K. Tafere and A. Woldemichael (2020), “Winners and Losers from COVID-19: Global Evidence from Google Search”, *World Bank Policy Research Working Paper*, No. 9268, World Bank, Washngton D.C., https://papers.ssrn.com/abstract=3617347 (accessed on 1 September 2020). [24]

Askitas, N. and K. Zimmermann (2009), “Google Econometrics and Unemployment Forecasting”, *Applied Economics Quarterly*, Vol. 55/2, pp. 107-120, http://dx.doi.org/10.3790/aeq.55.2.107. [21]

Baker, S., N. Bloom and S. Davis (2013), *Measuring Economic Policy Uncertainty*, http://papers.ssrn.com/abstract=2198490. [2]

Baker, S. and A. Fradkin (2017), “The impact of unemployment insurance on job search: Evidence from google search data”, *Review of Economics and Statistics*, Vol. 99/5, pp. 756-768, http://dx.doi.org/10.1162/REST_a_00674. [15]

Balakrishnan, A. and K. Dixit (2013), *Predicting Market Volatility Using Semantic Vectors and Google Trends*, http://developer.nytimes.com/docs/read/article (accessed on 19 February 2020). [32]

Bank of England (2020), *How are we monitoring the economy during the Covid-19 pandemic? | Bank of England*, https://www.bankofengland.co.uk/bank-overground/2020/how-are-we-monitoring-the-economy-during-the-covid-19-pandemic (accessed on 16 October 2020). [4]

Barigozzi, M. and M. Luciani (2017), “Common Factors, Trends, and Cycles in Large Datasets”, *Finance and Economics Discussion Series*, No. 111, Board of Governors of the Federal Reserve System, Washington, http://dx.doi.org/10.17016/FEDS.2017.111. [49]

Benatti, N. et al. (2020), *High-frequency data developments in the euro area labour market*, ECB, https://www.ecb.europa.eu/pub/economic-bulletin/focus/2020/html/ecb.ebbox202005_06~a8d6c566d3.en.html (accessed on 16 October 2020). [3]

Burdeau, E. and E. Kintzler (2017), *Assessing the use of Google Trends to predict credit developments*. [34]

Butler, D. (2013), *When Google got flu wrong*, http://dx.doi.org/10.1038/494155a. [56]

Carrière-Swallow, Y. and F. Labbé (2010), *Nowcasting With Google Trends in an Emerging Market*, https://ideas.repec.org/p/chb/bcchwp/588.html. [11]

Chen, S. et al. (2020), “Tracking the Economic Impact of COVID-19 and Mitigation Policies in Europe and the United States”, *IMF Research*, IMF. [8]

Chen, T. et al. (2015), “The 2007-2008 U.S. Recession: What Did The Real-Time Google Trends Data Tell the United States?”, *Contemporary Economic Policy*, Vol. 33/2, pp. 395-403, http://dx.doi.org/10.1111/coep.12074. [12]

Combes, S., Bortoli and Clément (2016), “Nowcasting with Google Trends, the more is not always the better”, http://dx.doi.org/10.4995/CARMA2016.2016.4226. [26]

Cournède, B., V. Ziemann and F. De Pace (2020), *Housing amid Covid-19: Policy responses and challenges*, OECD, OECD Policy Responses to Coronavirus (COVID-19). [23]

Craven, M. and J. Shavlik (1996), “Extracting tree-structured representations of trained networks”, *Advances in neural information*, http://papers.nips.cc/paper/1152-extracting-tree-structured-representations-of-trained-networks.pdf (accessed on 30 January 2019). [43]

Csáji, B. (2001), *Approximation with Artificial Neural Networks*, Faculty of Sciences, Etvs Lornd University, Hungary. [30]

D’Amuri, F. et al. (2012), “The predictive power of Google searches in forecasting unemployment”. [17]

Doerr, S. and L. Gambacorta (2020), “Identifying regions at risk with Google Trends: the impact of Covid-19 on US labour markets”, *BIS Bulletins*, https://ideas.repec.org/p/bis/bisblt/8.html (accessed on 1 September 2020). [25]

Fernández-Villaverde, J. and C. Jones (2020), *Macroeconomic Outcomes and COVID-19: A Progress Report*, National Bureau of Economic Research, Cambridge, MA, http://dx.doi.org/10.3386/w28004. [37]

Ferrara, L. and A. Simoni (2019), “When Are Google Data Useful to Nowcast GDP? An Approach Via Pre-Selection and Shrinkage”, *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.3370917. [14]

Fetzer, T. et al. (2020), *Coronavirus Perceptions And Economic Anxiety*. [29]

Fondeur, Y. and F. Karamé (2013), “Can Google data help predict French youth unemployment?”, *Economic Modelling*, Vol. 30/1, pp. 117-125, http://dx.doi.org/10.1016/j.econmod.2012.07.017. [16]

Giannone, D., M. Lenza and G. Primiceri (2018), “Economic Predictions with Big Data: The Illusion of Sparsity”, *CEPR Discussion Paper*, No. DP12256, CEPR, https://www.newyorkfed.org/research/staff_reports/sr847.html. [35]

Ginsberg, J. et al. (2009), “Detecting influenza epidemics using search engine query data”, *Nature*, Vol. 457/7232, pp. 1012-1014, http://dx.doi.org/10.1038/nature07634. [55]

Gonzales, F., A. Jaax and A. Mourougane (2020), “Nowcasting aggregate services trade. A pilot approach to providing insights into monthly balance of payments data”, OECD, Paris. [19]

Goolsbee, A. and C. Syverson (2020), “Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020”, *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.3631180. [52]

Haugen, M., B. Rajaratnam and P. Switzer (2015), "Extracting Common Time Trends from Concurrent Time Series: Maximum Autocorrelation Factors with Application to Tree Ring Time Series Data", http://arxiv.org/abs/1502.01073 (accessed on 10 June 2020). [48]

Havranek, T. and A. Zeynalov (2019), "Forecasting tourist arrivals: Google Trends meets mixed-frequency data", *Tourism Economics*, http://dx.doi.org/10.1177/1354816619879584. [51]

IMF (2020), *World Economic Outlook: A Long and Difficult Ascent*, International Monetary Fund, Washington, DC. [38]

INSEE (2020), *Les données « haute fréquence » sont surtout utiles à la prévision économique en période de crise brutale − Points de conjoncture 2020 | Insee*, Note de Conjoncture, https://www.insee.fr/fr/statistiques/4513034?sommaire=4473296 (accessed on 16 October 2020). [5]

Joseph, A. (2019), "Shapley Regressions: A Framework for Statistical Inference on Machine Learning Models", *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.3351091. [40]

Kliesen, K. (2020), *Tracking the U.S. economy and financial markets during the COVID-19 outbreak | FRED Blog*, The FRED Blog, https://fredblog.stlouisfed.org/2020/03/tracking-the-u-s-economy-and-financial-markets-during-the-covid-19-outbreak/ (accessed on 16 October 2020). [6]

Knotek, E. and S. Zaman (2020), "Real-Time Density Nowcasts of US Inflation: A Model-Combination Approach", *Working paper (Federal Reserve Bank of Cleveland)*, Federal Reserve Bank of Cleveland, http://dx.doi.org/10.26509/frbc-wp-202031. [7]

Laugel, T. et al. (2019), *The dangers of post-hoc interpretability: Unjustified counterfactual explanations*, International Joint Conferences on Artificial Intelligence, http://dx.doi.org/10.24963/ijcai.2019/388. [44]

Lewis, D., K. Mertens and J. Stock (2020), "Monitoring Real Activity in Real Time: The Weekly Economic Index", *Liberty Street Economics*. [53]

Lundberg, S. and S. Lee (2017), "A Unified Approach to Interpreting Model Predictions", *Advances in Neural Information Processing Systems*, Vol. 30, pp. 4765-4774, http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions. [45]

Molnar, C. (2020), *Interpretable Machine Learning*, https://books.google.fr/books?hl=en&lr=&id=jBm3DwAAQBAJ&oi=fnd&pg=PP1&dq=Interpretable+Machine+Learning.+A+Guide+for+Making+Black+Box+Models+Explainable.&ots=EfAX-iFDY3&sig=jPrA7RHxQ5j7Bura6KZnPCMZrj8&redir_esc=y#v=onepage&q=Interpretable%20Machine%20Learning.%20A%20Guide%20for%20Making%20Black%20Box%20Models%20Explainable.&f=false (accessed on 28 August 2020). [54]

Morgavi, H. (2020), "A GARCH model to now-cast private consumption using Google trends data", *OECD Economics Department Working Paper*, OECD, Paris. [18]

Narita, F. and R. Yin (2018), "In Search of Information: Use of Google Trends' Data to Narrow Information Gaps for Low-income Developing Countries", *IMF Working Papers*, No. 18/286, IMF, Washington DC. [13]

OECD (2020), *OECD Economic Outlook, Interim Report September 2020*, OECD Publishing, Paris, https://dx.doi.org/10.1787/34ffc900-en. [9]

Park, S., J. Lee and W. Song (2017), "Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data", *Journal of Travel and Tourism Marketing*, Vol. 34/3, pp. 357-368, http://dx.doi.org/10.1080/10548408.2016.1170651. [50]

Pisu, M., H. Costa and H. Hwang (2020), "Digital platforms and the COVID-19 crisis", *OECD Economic Department Working Paper*, OECD, Paris. [20]

Renard, X. et al. (2019), "Concept Tree: High-Level Representation of Variables for More Interpretable Surrogate Decision Trees", http://arxiv.org/abs/1906.01297 (accessed on 9 September 2019). [39]

Ribeiro, M., S. Singh and C. Guestrin (2016), *Model-Agnostic Interpretability of Machine Learning*, https://arxiv.org/abs/1606.05386 (accessed on 25 June 2020). [41]

Scott, S. and H. Varian (2014), "Bayesian variable selection for nowcasting economic time series", in *Economic Analysis of the Digital Economy*, University of Chicago Press, http://www.nber.org/chapters/c12995. [33]

Scott, S. and H. Varian (2014), "Predicting the present with bayesian structural time series", *International Journal of Mathematical Modelling and Numerical Optimisation*, Vol. 5/1-2, pp. 4-23, http://www.inderscienceonline.com/doi/abs/10.1504/IJMMNO.2014.059942. [27]

Stephens-Davidowitz, S. and H. Varian (2015), *A Hands-on Guide to Google Data*, Google, Inc., http://people.ischool.berkeley.edu/~hal/Papers/2015/primer.pdf (accessed on 19 February 2020). [47]

Suhoy, T. (2009), "Query Indices and a 2008 Downturn: Israeli Data", *Bank of Israel Working Papers*. [28]

Tiffin, A. (2019), "Machine Learning and Causality: The Impact of Financial Crises on Growth", *IMF Working Paper*, No. 19/228, International Monetary Fund, http://dx.doi.org/9781513518305/1018-5941. [46]

Varian, H. and H. Choi (2009), "Predicting the Present with Google Trends", *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.1659302. [10]

Vermeulen, P. (2012), "Quantifying the qualitative responses of the output purchasing managers index in the US and the Euro area", *ECB Working Paper*, No. 1417, ECB, Frankfurt am Main, http://www.ecb.europa.euFax+496913446000http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html (accessed on 16 October 2020). [1]

Vosen, S. and T. Schmidt (2011), "Forecasting private consumption: survey-based indicators vs. Google trends", *Journal of Forecasting*, Vol. 30/6, pp. 565-578, http://dx.doi.org/10.1002/for.1213. [31]

Wu, L. and E. Brynjolfsson (2015), "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales", in *Economic Analysis of the Digital Economy*, University of Chicago Press, http://dx.doi.org/10.7208/chicago/9780226206981.003.0003. [22]

Zhao, Q. and T. Hastie (2019), "Causal Interpretations of Black-Box Models", *Journal of Business & Economic Statistics*, pp. 1-10, http://dx.doi.org/10.1080/07350015.2019.1624293. [42]

Zou, H. and T. Hastie (2005), “Regularization and variable selection via the elastic net”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 67/2, pp. 301-320, http://dx.doi.org/10.1111/J.1467-9868.2005.00503.X@10.1111/(ISSN)1467-9868.TOP_SERIES_B_RESEARCH. [36]

Annex A. Data pre-processing and data issues

Google Trends data are extremely rich and offers a vast research horizon. The data also have limits and need to be adjusted to be suitable to economic analysis. This annex describes pre-processing steps designed to address long-term bias, sampling and data variability and other minor issues.

Multiple sampling

Google Trends search indices are computed on the basis of a sample of the universe of Google searches, for the sake of computational tractability. As a result, some indices with low volumes may suffer from too large sampling variance (Combes, Bortoli and Clément, 2016[26]), especially in countries with lower market penetration. Six samples are drawn in order to alleviate this issue, by spacing requests by 10 minutes with the Google Trends API. The SVIs used to build the tracker are then averaged over the ten queries. Multiple sampling allows standard deviations to be computed and indices with too large variance to be excluded. This can occur with some topics or categories including too little searches, which can result in highly variant and discontinuous time series, as SVIs are bottom coded.

Extracting the common time trend

Google Trends variables exhibit a downward trend, reflecting the increasing number of Google Search users since 2004. Google Trends data are Search Volume Indices (SVI) based on search ratios: the initial search volume (SV) for a category or topic at a given time is divided by the total number of searches at that date. This ratio is multiplied by a constant in order to result in a time series index with maximum over the period equal to 100[15]:

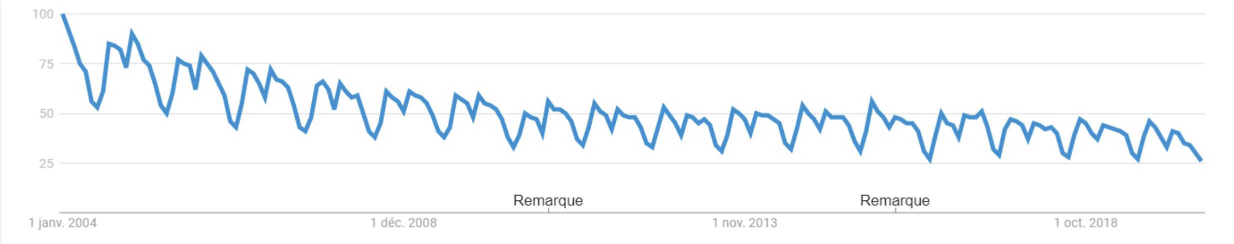
$$SVI_{ct} = \frac{SV_{ct}}{SVT_t} * C_c \qquad [1]$$

Changes in the denominator (total searches, $\boldsymbol{SVT_t}$) can induce biases as use of the internet has evolved since 2004. As shown by Stephens-Davidowitz and Varian (2015[47]), the search intensity for “science” in the United States displays a downward trend, as the scope of Google users has broadened from an expert community to the general population (see Figure A. A.1) This bias is not linear and can alter the economic predictive power of yearly growth rates.

[15] The constant C_c is defined as $(\max_t \frac{SV_{ct}}{SVT_t})^{-1}$

Figure A A.1. Downward long-term bias

Search Ratio Index for “science” in the United States



Source: Google Trends.

This paper introduces an approach to address this downward bias based on the fact that it is common to all series. The long-term in the denominator is a common component to the indices for each search category. This common component is linearly decomposable from the rest when the Search Volume Indices are taken in log:

$$svi_{ct} = \log(SVI_{ct}) = sv_{ct} - svt_t + C_c \qquad [2]$$

The log numerator sv_{ct} and the log multiyplicative constant C_c are indexed by the search category c, but the log denominator svt_t is common to each series and indexed only by time.

Extracting a common component from concurrent time series can be done with various methods (Haugen, Rajaratnam and Switzer, 2015[48]; Barigozzi and Luciani, 2017[49]). The approach adopted in this paper is simple and relies on the intuition that the common term can be considered a common factor with unit loadings for each series. It is extracted with Principal Component Analysis (PCA) on the log-SVI series long-term trends filtered out using an HP filter. The first component from the PCA applied to the series long-term trends is rescaled to have the same mean and standard deviation as the log-SVIs average. It is then subtracted from the log-SVIs[16].

The rescaled first component obtained from the long-term log-SVIs is assumed to capture the common long-term trend. There can be common economic shocks affecting series on the short-run, but given the number and variety of search categories (ranging from “animal products” and “mental health” to “rail transport” or “public safety”), the common term computed with PCA is likely to capture the denominator effect (Figure A.A.2.) shows a number of series in the United States before and after the common trend is removed.

[16] As an alternative to the approach based on HP-filtering and PCA, the common component can be extracted using category fixed-effects panel regressions. For each country, the SVIs can be stacked and the following model can be estimated:

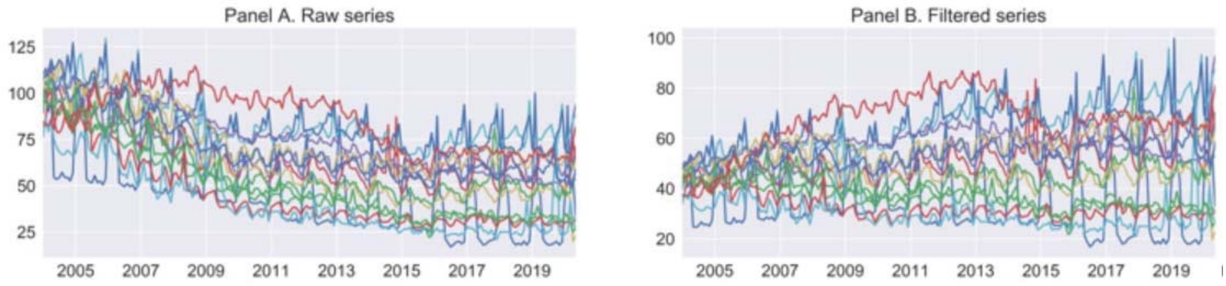
$$svi_{ct} = \alpha_c + P(t)\beta + \epsilon_{ct} \qquad [3]$$

Where $P(t)$ is a polynomial in time of order 5, so that $P(t)\beta$ captures the time trend common to all log SVI. The category fixed effects α_c will capture both the multiplicative constant and the mean log search volume, while the error term will capture the demeaned log search volume SV_{ct}.

This approach yields broadly similar results to the one based on PCA. The latter was preferred as the former requires arbitrarily choosing a parameter for the order of the time trend polynomial.

Figure A A.2. Filtering out the common trend

Selected SVIs, United States



Note: The left-panel shows the raw data for series corresponding to 15 selected categories. The right panel shows the same series after the common bias has been filtered out.
Source: Google Trends and OECD computations.

Addressing seasonality

Most variables exhibit strong seasonal patterns. Consequently, category-based variables are transformed using the log difference with the same month of the past year. Topic-based categories are less sensitive to seasonality, and often react to specific events, and are thus taken in log form, but not differenced. The log-difference and log transformations are performed on the series after the common long-term trend has been filtered out.

Addressing breaks

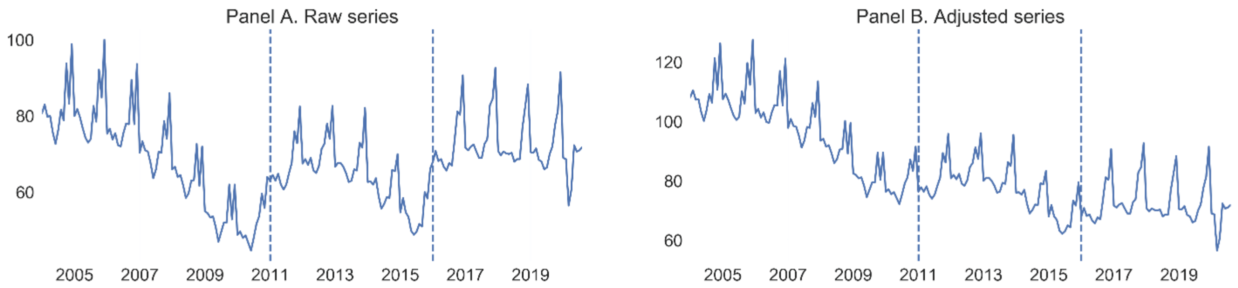
Lastly, changes in the data collection process induce breaks in January 2011 and January 2016. The breaks are briefly documented on the Google Trends website: the "process for geographic localisation" changed in January 2011, and the "data collection system" was "improved" in January 2016. These breaks often appear as minor issues in the literature if at all. This paper takes a different stance and these breaks are addressed with great care:

The post-break series are translated in order to close the gap between the January 2011 (2016) and January 2010 (2015) for each variable (see Figure A.A.3)

For each variable, the difference between January 2011 (2016) and January 2010 (2015) is subtracted from observations after January 2011 (2016) inclusive. This translation arbitrarily sets the yearly growth rate of each variable in January 2011 (2016) to zero: by doing so, it prevents the occurrence of outliers in the growth rate series.

- Setting the growth rate to zero at break points reduces outliers but does not solve the problem that the true growth rates over the years 2011 and 2016 are unknown. At any month in 2011 (2016), the true difference (or log-difference) with the same month of the past year is unknown as the geographic localisation system (data collection process) changed in the meantime. This is a significant problem as these two years represent approximately 12% of the data. Therefore, the tracker is unavailable for the years 2011 and 2016, that will appear under a grey-shaded area in the charts below and that will be ignored when computing performance metrics.

Figure A A.3. Adjusting for breaks in Google Trends series



Note: Search Volume Index for the "Apparel" category in the US raw (Panel A) and adjusted for breaks in data collection occurring in January 2011 and January 2016 (Panel B).
Source: Google Trends and OECD calculations.

Time window and frequencies

Google Trends data are made available at various frequencies depending on the selected time window. Weekly series are available over the five year, and monthly series back to 2004. The model estimation is performed in two-steps, first using the monthly series and thus exploiting history back to 2004, and second using the weekly series to produce a weekly tracker.

Weekly series need to be calibrated on the monthly series. Each SVI is provided with max = 100, which can explain discrepancies between the monthly and weekly index for a given keyword or category. Formally, the monthly SVI is equal to the search volume ratio (SVR) multiplied by a constant scaling factor $c_m = \frac{100}{\max_{2004-2020}(SVR)}$, whereas the weekly SVI is equal to the SVR multiplied by a different scaling factor: $c_w = \frac{100}{\max_{2015-2020}(SVR)}$. The calibration is thus performed by first intrapolating monthly series at the weekly frequencies and, second, multiplying the weekly series by the mean ratio of the monthly SVI divided by the weekly SVI over the 2015-2020 segment.

Yearly log differences are computed from the weekly series. The log SVI at $y-1$ is obtained by intrapolating the whole series at a daily frequency. The log difference for, say, 03-01-2020, is obtained by taking the difference between the $svi_{03-01-2020}$ and the log of a weighted average of the closest known values before and after 03-01-2019, that is 31-12-2018 and 07-01-2019.

Annex B. Additional details

Algorithm details

The neural network model used in this paper uses a standard architecture with two hidden layers of 300 and 10 neurons, an Adam solver and “relu” activation functions. The level of fit is determined at each fit by early stop. Data are first transformed using their standard normalisation rather than levels. The python code is provided in Figure A C.1.

Figure A B.1. Algorithm python code

```
network = MLPRegressor(hidden_layer_sizes=(300, 10),
                                 solver = "adam",
                                 activation = "relu",
                                 learning_rate_init = 0.001,
                                 tol=1e-4,
                                 max_iter=7000,
                                 early_stopping = True,
                                 random_state=0)
mlp = Ensemble(learner = network, size_ensemble = 5, name_param_seed = "random_state")
```
Source: OECD

Variable groupings

'Crisis / Recession': "Economic crisis", "Crisis", "Recession", "Financial crisis", 'Krach'

'Unemployment / unemployment benefits': "Unemployment", "Unemployment benefits", 'Welfare & Unemployment'

"Credit & Loans" : 'Student loan', 'Credit & Lending', 'Loan', 'Interest', 'Mortgage', 'Auto Financing'

'Consumption items' : 'Food & Drink', 'GPS & Navigation', 'Performing Arts','Luggage topic', 'Vehicle Brands', 'Birthday', 'Travel', 'Energy & Utilities', 'Vehicle Shopping', 'Tobacco Products', 'Health', 'Pharmacy', 'Carpooling & Ridesharing', 'Sports', 'Animal Products & Services', 'Fitness', 'Weddings', 'Car Rental & Taxi Services', 'Autos & Vehicles', 'Tourist Destinations', 'Home & Garden', 'Events & Listings', 'Grocery & Food Retailers', 'Vehicle Licensing & Registration', 'Timeshares & Vacation Properties', 'Home Appliances', 'Mass Merchants & Department Stores', 'Car Electronics', 'Fashion & Style', 'Trucks & SUVs', 'Home Furnishings', 'Footwear', 'Cruises & Charters', 'Hotels & Accommodations', 'Luggage & Travel Accessories', 'Fast Food', 'Book Retailers', 'Veterinarians', 'Spas & Beauty Services', 'Acting & Theater', 'Travel Agencies & Services'

"Jobs" : 'Waiter', 'Job Listings', 'Resumes & Portfolios', 'Jobs topic', 'Temporary jobs', 'Private employment agency', 'Recruitement', 'Developer Jobs', 'Job search'

"Bankruptcy" : 'Bankruptcy topic', 'Judicial Liquidation', 'Bankruptcy'

"Housing" : 'Affordable housing', 'House price index', 'Apartments & Residential Rentals', 'Home Insurance', 'Home Improvement'

"**News & Politics**" : 'Economy News', 'Business News', 'World News', 'Politics', 'Newspapers'

"**Construction**" : 'Flooring', 'Construction Consulting & Contracting', 'Swimming Pools & Spas', 'Civil Engineering', 'Construction & Maintenance'

'Personal Finance' : 'Investment', 'Investing', 'Financial Planning'

'Business services' : 'Data Management', 'Enterprise Technology', 'Accounting & Auditing', 'CAD & CAM', 'Development Tools', 'Customer Relationship Management (CRM)', 'Printing & Publishing', 'Corporate Events', 'Computer Security', 'Outsourcing', 'Distribution & Logistics', 'Computer Servers', 'Consulting', 'Web Hosting & Domain Registration', 'Enterprise Resource Planning (ERP)', 'Business Operations', 'Commercial Vehicles'

'Industrial activity' : 'Agriculture & Forestry', 'Agrochemicals', 'Aviation', 'Business & Industrial', 'Chemicals Industry', 'Textiles & Nonwovens', 'Coatings & Adhesives', 'Food Production', 'Dyes & Pigments', 'Freight & Trucking', 'Transportation & Logistics', 'Mail & Package Delivery', 'Manufacturing'

List of topics and categories

<u>**List of topics**</u>:

"Birthday", "Private employment agency" , "House moving", "Unemployment benefits", "Recruitment", "Investment", "Lawyer", "Jobs", "Economic crisis", "Unemployment", "Financial crisis", 'Public debt', "Office space", "Job search", "Temporary jobs", "Housing bubble", "House price index", "Mortgage", "Crisis", "Loan", "Interest", "Student loan", "Affordable housing", "Recession", "Krach", "Bank", "Bankruptcy", "Exportation", "Commercial Building", "Luggage', "Judicial Liquidation", "Foreclosure".

<u>**List of categories**</u>:

Events & Listings
Performing Arts
Autos & Vehicles
Vehicle Brands
Vehicle Licensing & Registration
Health Insurance
Food & Drink
Restaurants
Doctors' Offices
Hospitals & Treatment Centers
Emergency Services
Mental Health
Home & Garden
Real Estate
Real Estate Agencies
Shopping
Travel
Hotels & Accommodations
Computers & Electronics
Apparel
Consumer Electronics
Luxury Goods

Business Services
Recruitment & Staffing
Office Supplies
Bankruptcy
Credit & Lending
Commercial Lending
College Financing
Home Financing
Auto Financing
'Agriculture & Forestry'
'Forestry'
'Aquaculture'
'Grocery & Food Retailers'
'Tobacco Products'
'Footwear'
'Office Supplies'
'Printing & Publishing'
'Fuel Economy & Gas Prices'
'Chemicals Industry'
'Pharmacy'
'Computer Hardware'
'Industrial Materials & Equipment'

'Entertainment Industry'
'Internet & Telecom'
'Programming'
'Finance'
'Insurance'
'Real Estate'
'Legal Services'
'Architecture'
'Advertising & Marketing'
'Veterinarians'
'Business Services'
'Travel Agencies & Services'
'Fire & Security Services'
'Government'
'Education'
'Medical Facilities & Services'
'Social Services'
'Performing Arts'
'Sports'
'Professional & Trade Associations'
'Consumer Electronics'
'Unwanted Body & Facial Hair Removal'

Agricultural Equipment	'Boats & Watercraft'	Alcoholic Beverages
Construction & Maintenance	'Retail Trade'	Building Materials & Supplies
Pharmaceuticals & Biotech	'Freight & Trucking'	Civil Engineering
Transportation & Logistics	'Maritime Transport'	Construction Consulting & Contracting
Distribution & Logistics	'Aviation'	Home Improvement
Jobs	Import & Export	Entertainment Media
Developer Jobs	Rail Transport	Gifts & Special Event Items
Food Production	'Mail & Package Delivery'	Home Appliances
		Home Furnishings

Annex C. Additional results

Figure A C.1 Forecast simulations

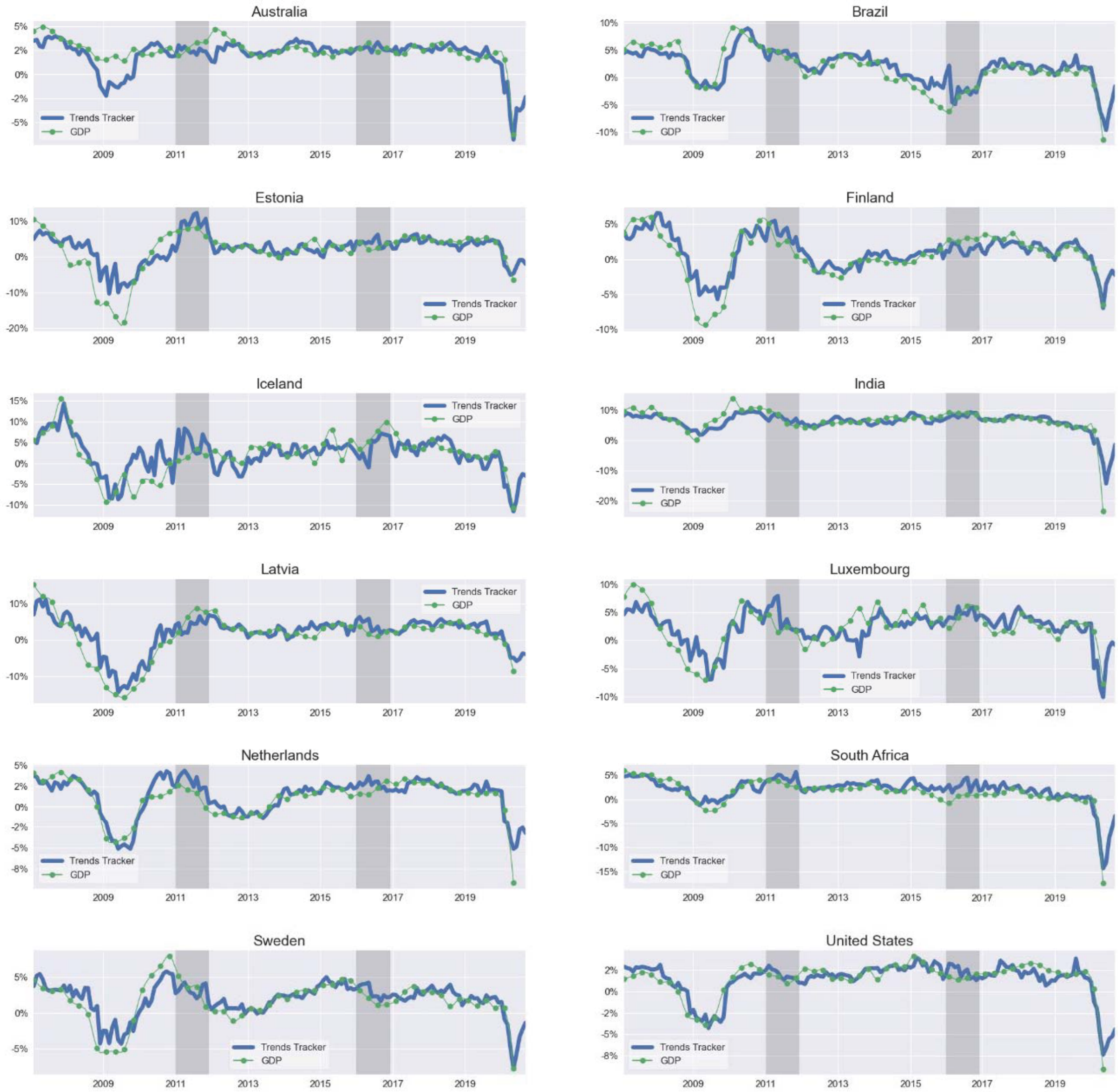


Figure A C.2 Forecast simulations (*contd.*)

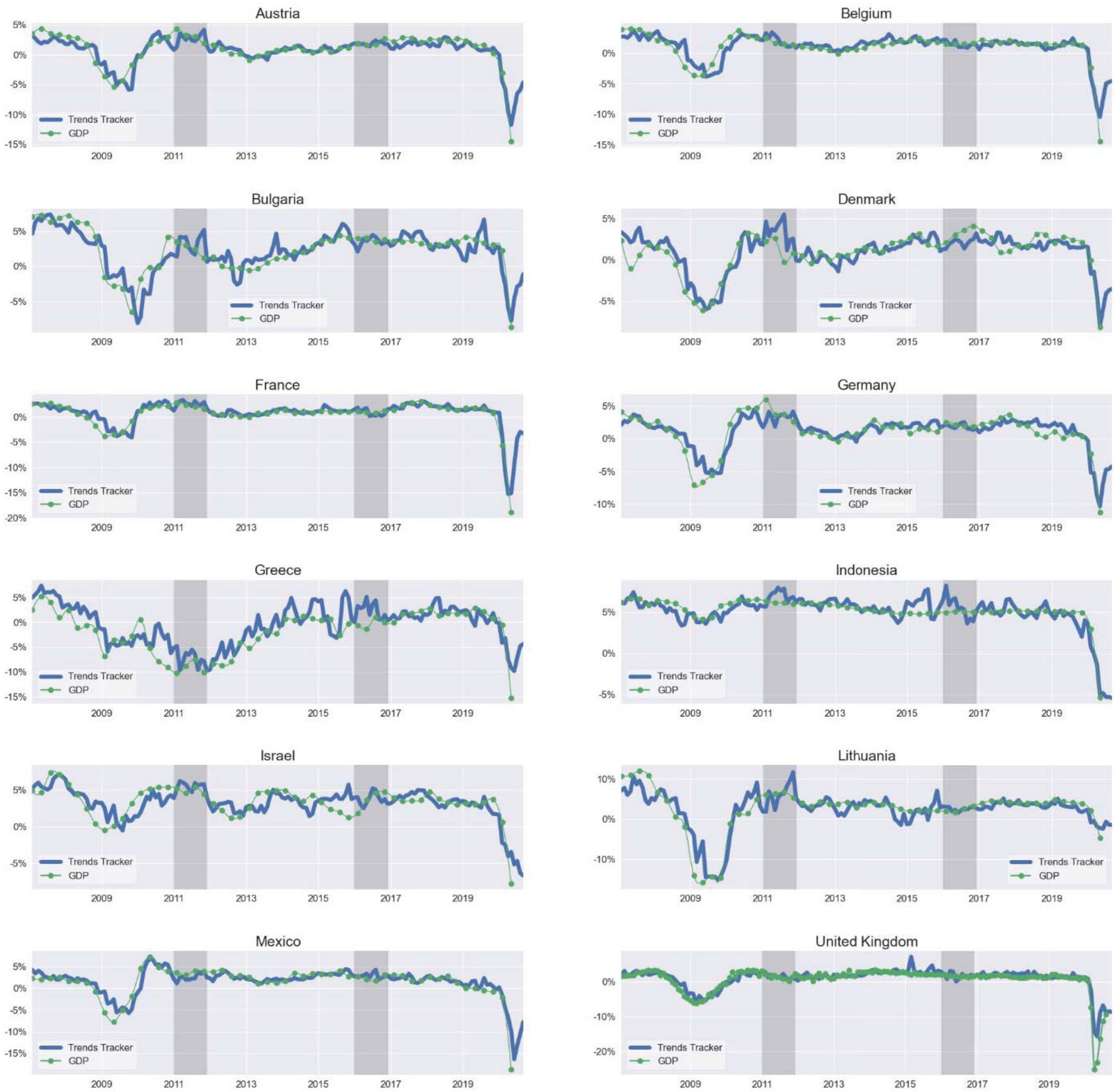


Figure A C.3 Forecast simulations (*contd.*)

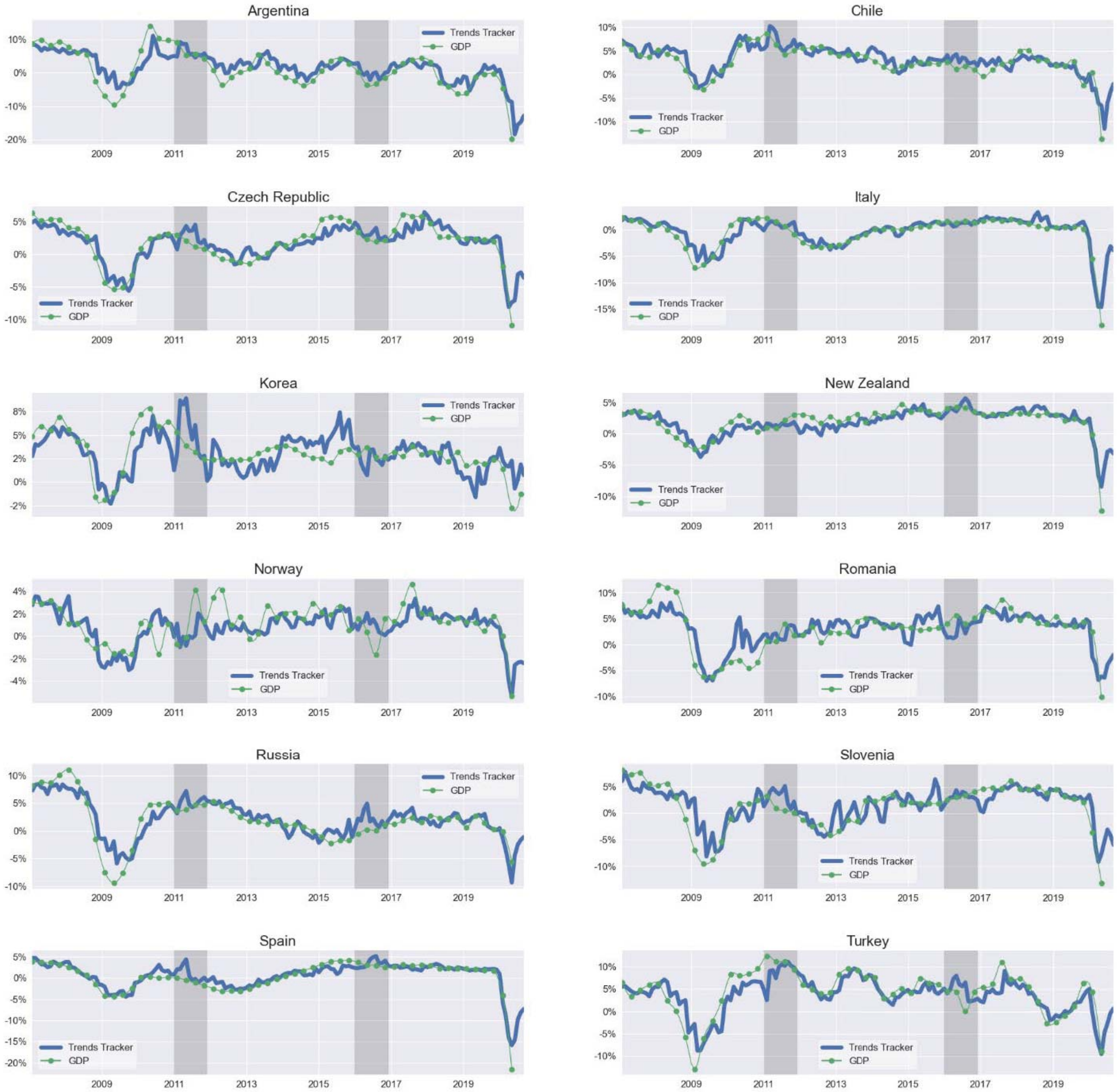
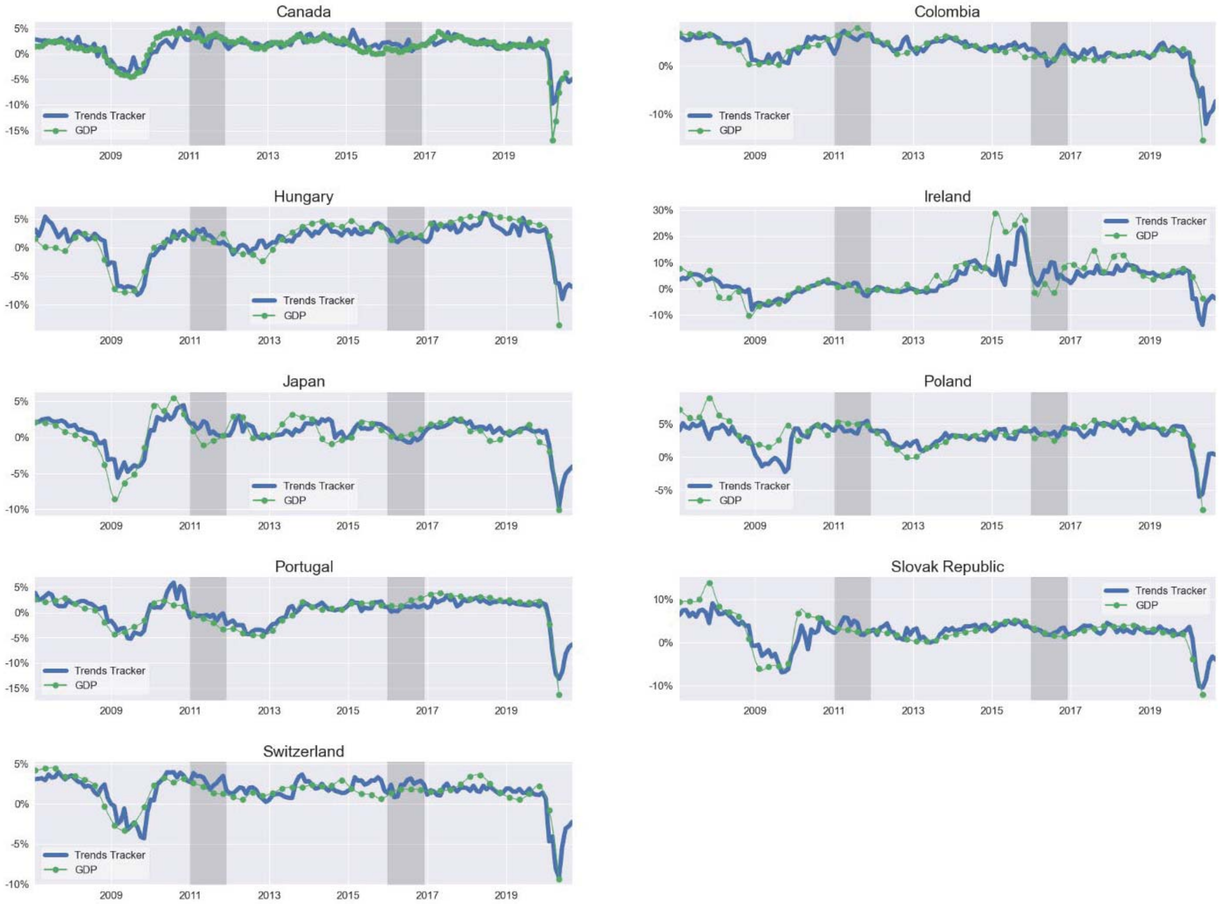


Figure A C.4 Forecast simulations (*contd.*)



Note: The quarterly model is applied to 3-month moving averages of Google Trends series and yields monthly estimates that can be compared to quarterly GDP growth for February (Q1), May (Q2), August (Q3) and November (Q4). Shaded areas in 2011 and 2016 are years when the tracker is unavailable due to structural breaks in Google Trends data preventing the calculation of year-on-year growth rates in search intensities. Simulations are based on the latest GDP data, not the real-time vintages. For each quarter, the forecast is made 5 days after the end of the month, so 3-7 weeks before the GDP is published.
Source: Google Trends; OECD Quarterly National Accounts; and OECD calculations.

Figure A C.5. The OECD Weekly Tracker

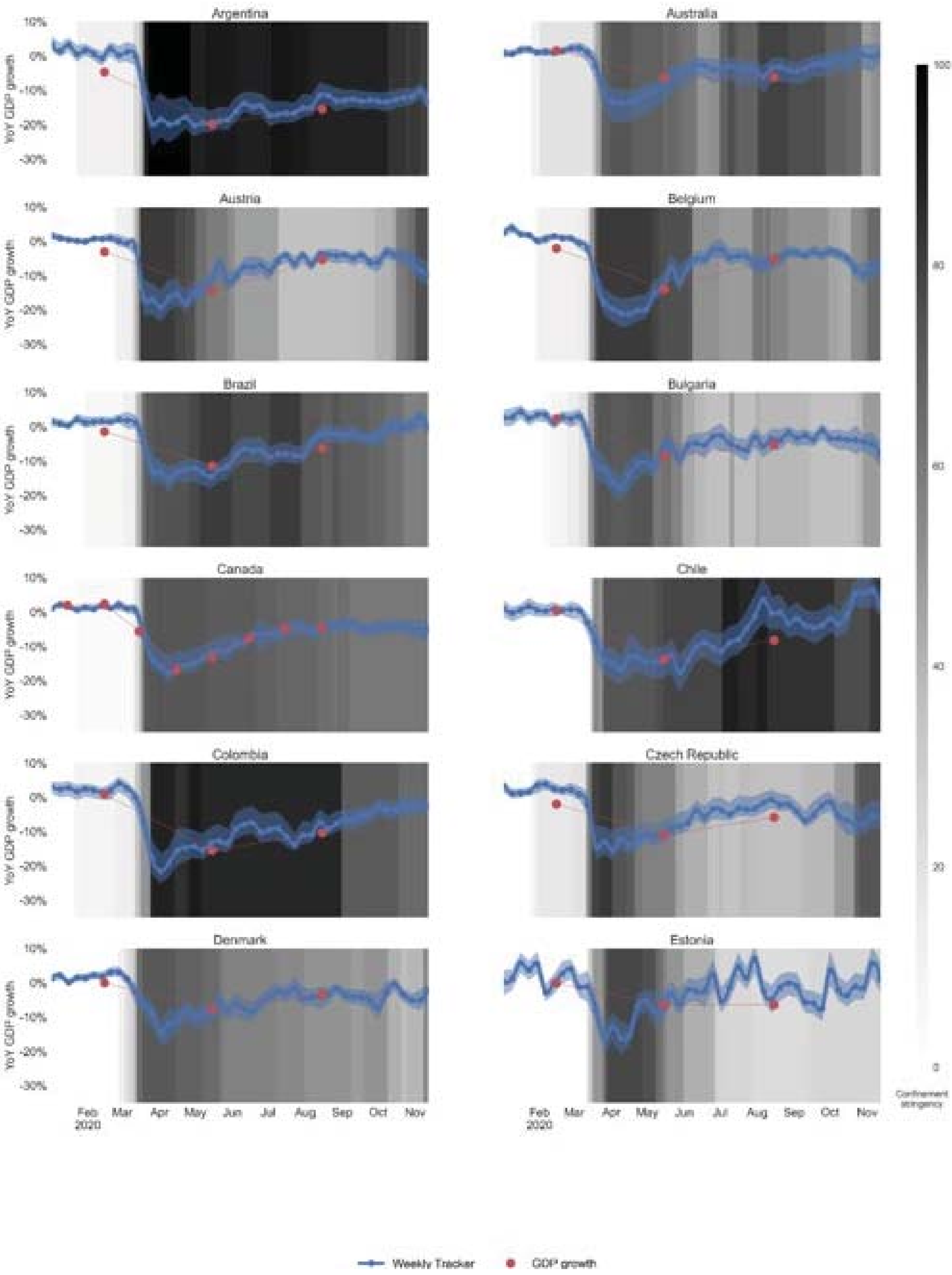


Figure A C.6. The OECD Weekly Tracker *(contd.)*

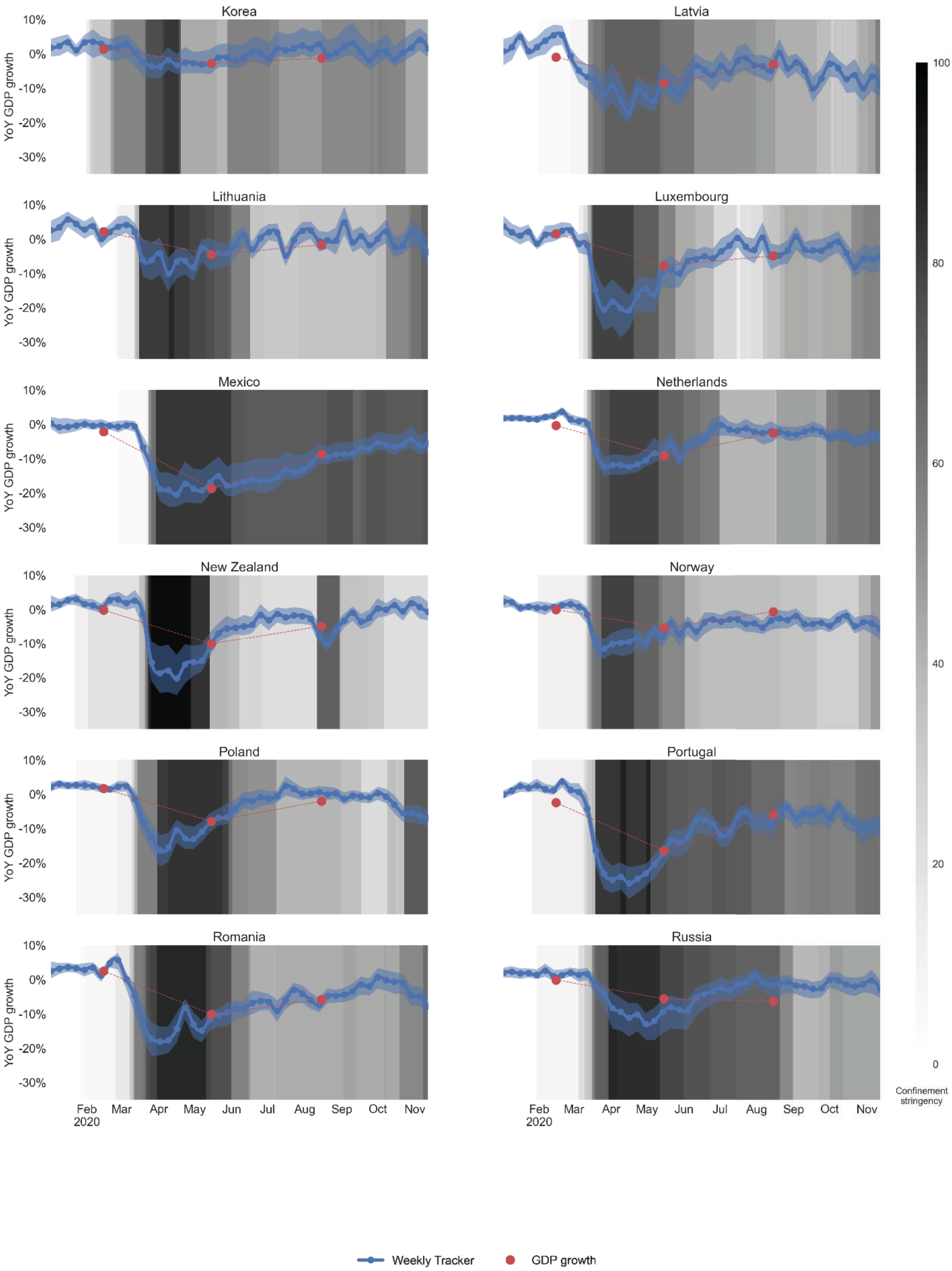


Figure A C.7. The OECD Weekly Tracker (*contd.*)

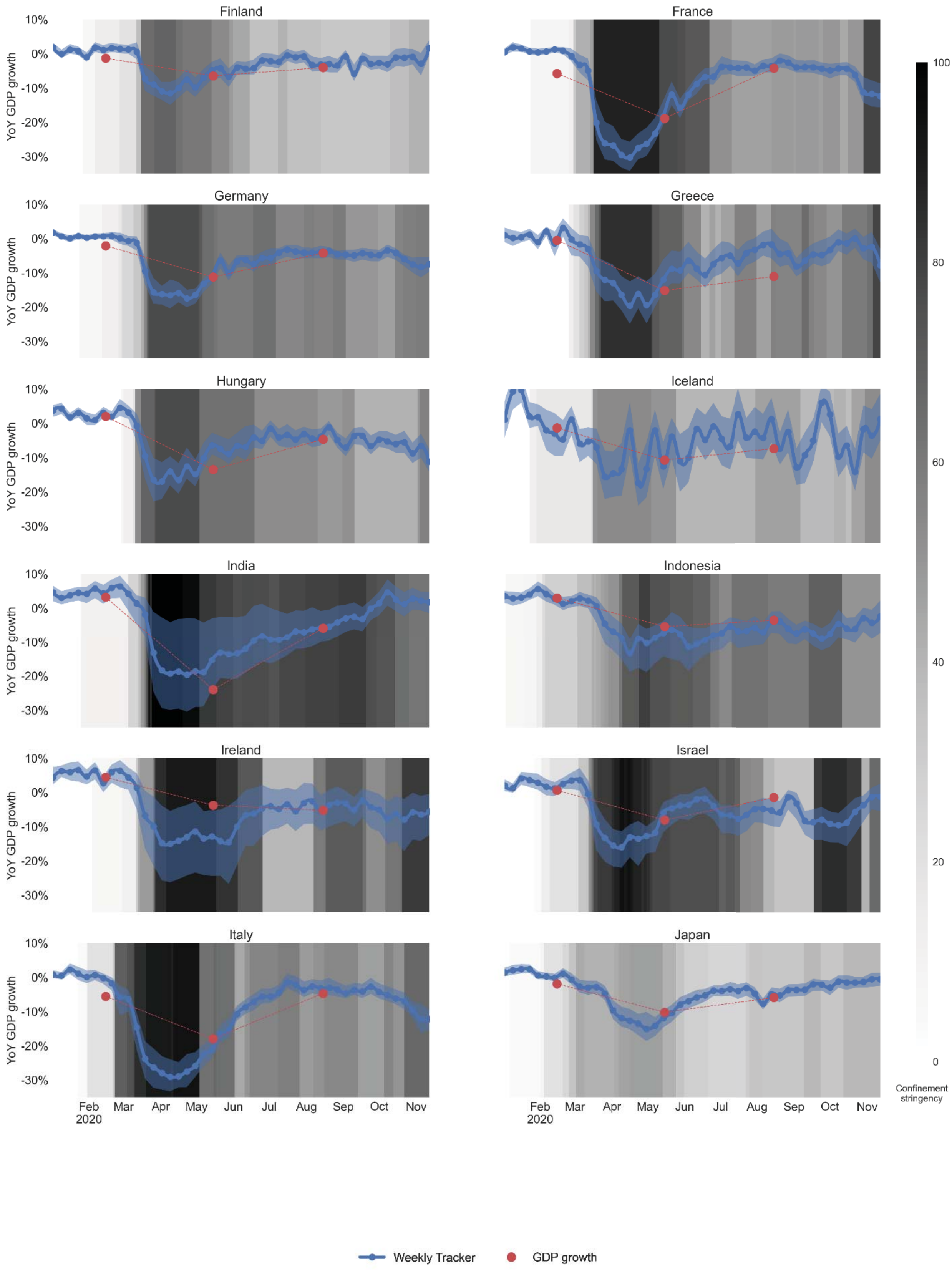
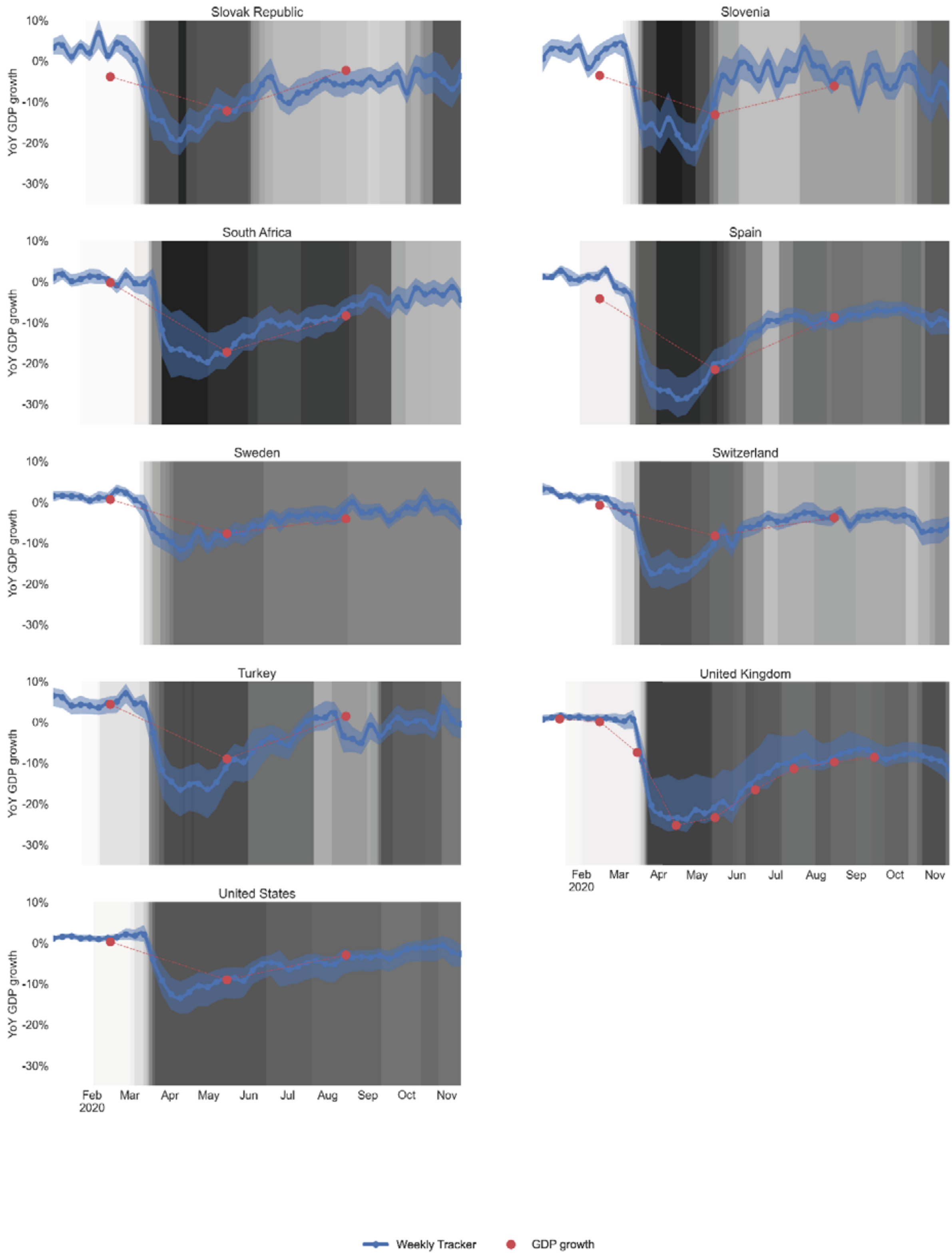


Figure A C.8. The OECD Weekly Tracker *(contd.)*



Note: The confidence band shows 95% confidence intervals. Red dots representing GDP growth are official statistics except for 2020 Q3 where they are either Economic Outlook projections or the outturns when the latter are available (for Indonesia and Mexico).
Source: *OECD Economic Outlook* 108 database; and OECD Weekly Tracker.