# The OECD Weekly Tracker of activity based on Google Trends

*The Banca d'Italia and Federal Reserve Board joint virtual conference, 12 Nov 2020*

**Nicolas Woloszko**
OECD Economics Department & Innovation LAB

# DATA & PREPROCESSING

# Google Trends for Nowcasting

Extensive country coverage: **46 OECD countries + G20 + Partners**

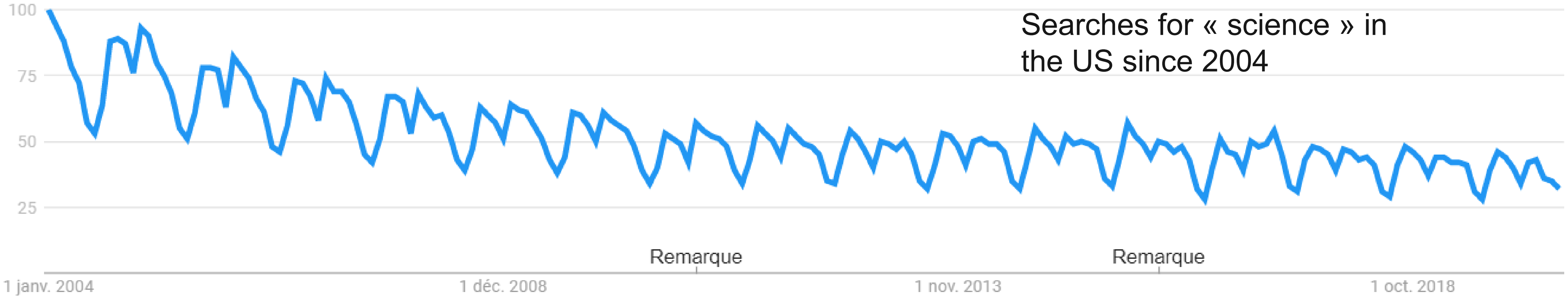
Search Intensity Indices for categories + topics of search keywords

Covering multiple economic sectors:

- ***Consumption***. Eg: Food & Drink, Tourist Destinations, Vehicle Brands, Home Appliances
- ***Labour market***. Eg: Unemployment benefits, Developer Jobs, Temporary jobs
- ***Housing & Debt***. Eg: Real Estate Agencies, Credit & Lending, Forbearance
- ***Business services & dynamism***. Eg: Venture Capital, Commercial Vehicles, Bankruptcy
- ***Industrial activity***. Eg: Maritime Transport, Agricultural Equipment, Agrochemicals
- ***Economic sentiment***. Eg: Economic crisis, Public debt, Economy News
- ***Poverty***. Eg: Food bank, Social Services

# Data preprocessing

- **Multiple sampling** using the API in order to reduce sampling variance.
- **Addressing breaks** in the data due to changes in the data collection process in January 2011 and January 2016.
- Use of **log difference with the same month of the past year** in order to get rid of seasonality
- The **bias caused by changes in the denominator** is addressed using a specific method




Searches for « science » in the US since 2004

# MODEL & SIMULATIONS

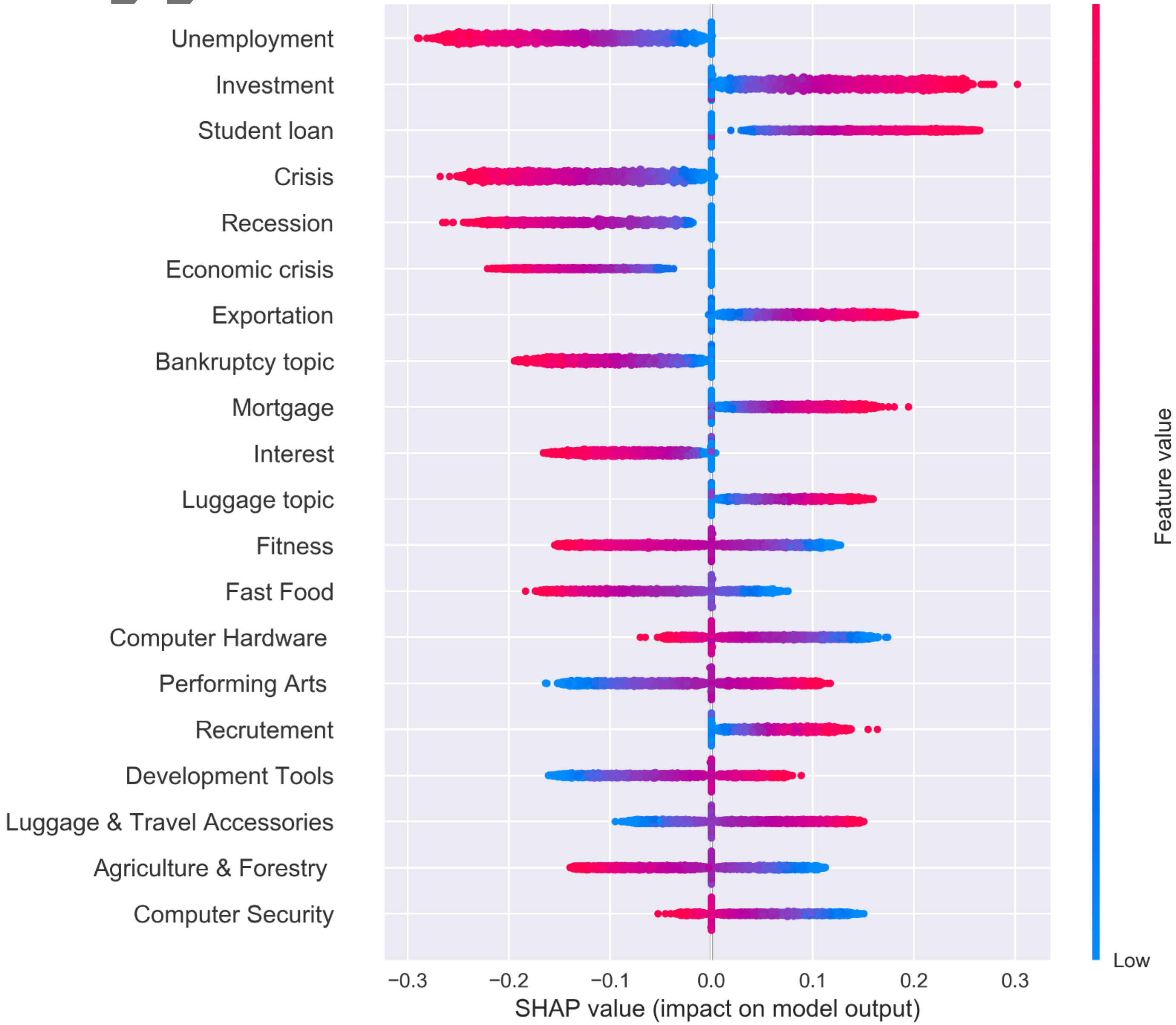
# A neural panel model of GDP growth for 46 countries

- Modelling the relationship $y_{iq} = \boldsymbol{f}\left(GT_{iq}, \mathbb{I}_i\right)$. Two steps:

  1) $\boldsymbol{f}$ is *estimated* using a **neural network** trained on quarterly observations
  2) $\boldsymbol{f}$ *applied* to weekly Google Trends series and predicts estimates of **YoY growth rate in weekly GDP**.

- **Why a neural panel model ?**
  - Country-specific model would have more variables (250) than observations (60)
  - Linear panel models may suffer from bias caused by cross-country heterogeneity
  - **Our neural panel model with country dummies exploits a large sample (2700 quarterly obs.) while capturing cross-country heterogeneity**

# Variable importance




Model interpretability techniques can provide insights into the inner workings of the models.

A recent method ('SHAP') (Lundberg and Lee, 2017), decomposes the predictions made by any algorithm into variable contributions (their "SHAP values") using Shapley Values (from the field of Game Theory)

With a linear model, the SHAP value for observation i and variable $x_j$ is simply equal to $\beta_j * x_{ij}$.

# Capturing cross-country heterogeneity

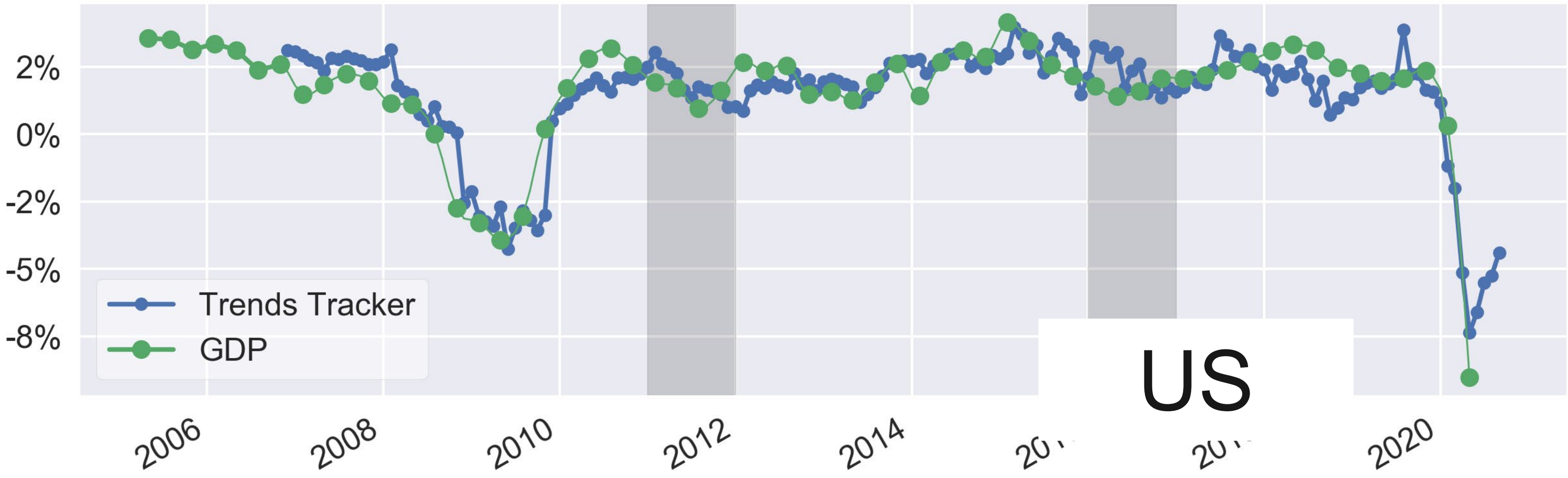

**Partial dependency plot**: variable contribution (using Shapley Values) against variable value.

In a linear setting: it would show the $X\beta$ straight line against the value of $X$.

**The algorithm captures cross-country heterogeneity**: the Weekly Tracker more sensitive to searches for « unemployment » in Argentina than in the UK.

# FORECAST SIMULATIONS

# The model performs well in forecast simulations

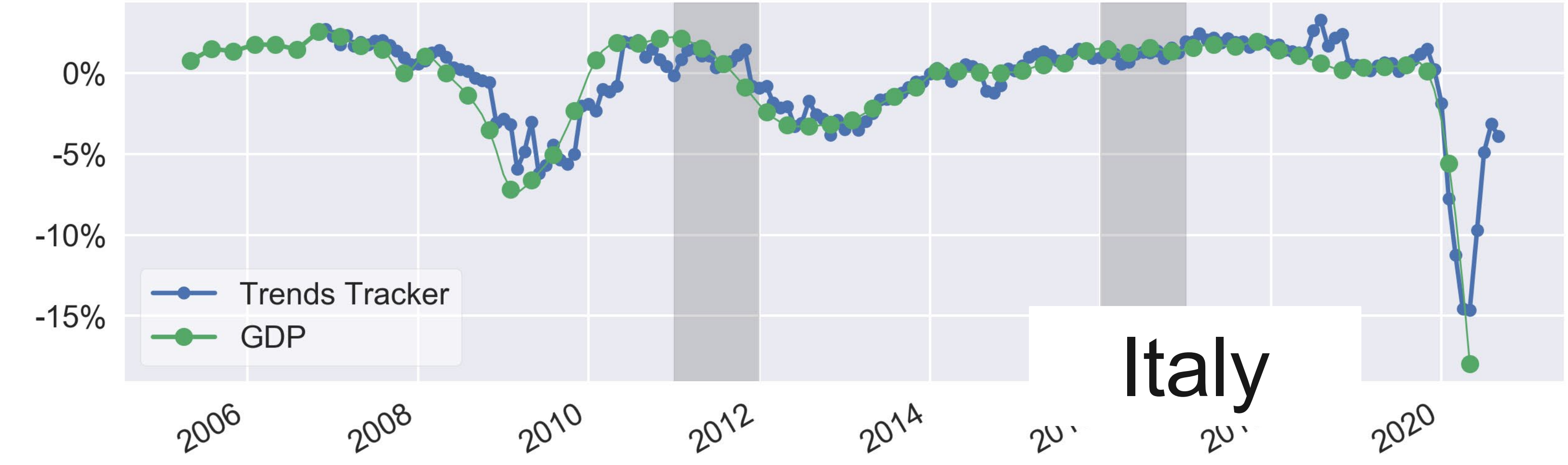


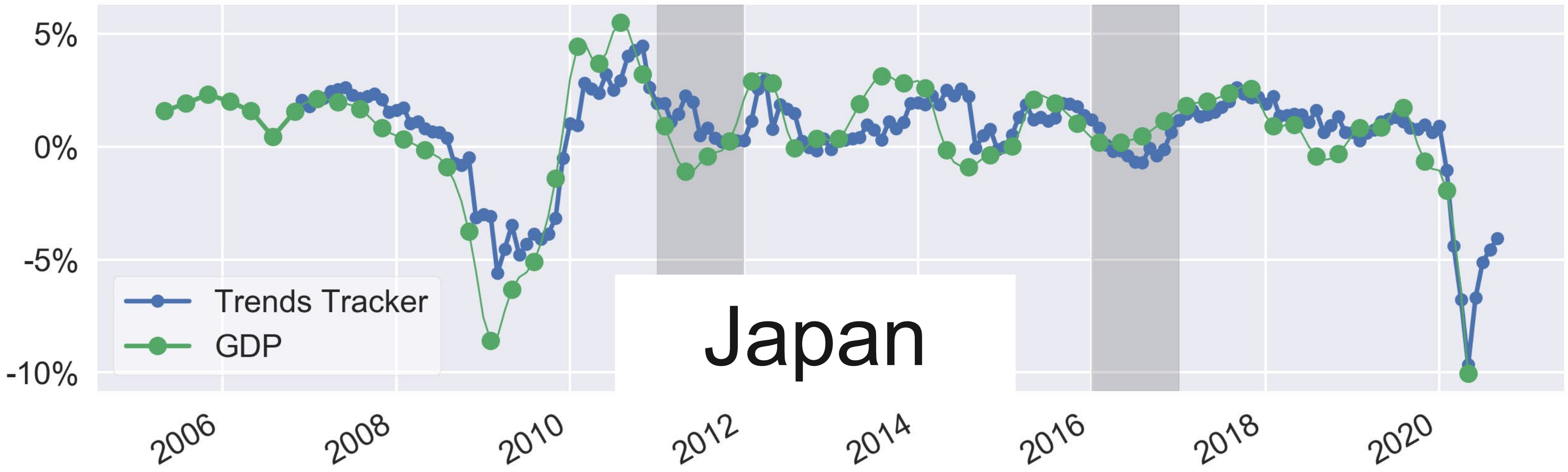


Simulations are ran in pseudo-real time (**out-of-sample**) at M-1

On average, **the Trends tracker is 37% more accurate than an AR(4)** across countries

# Forecast simulations: quarterly model



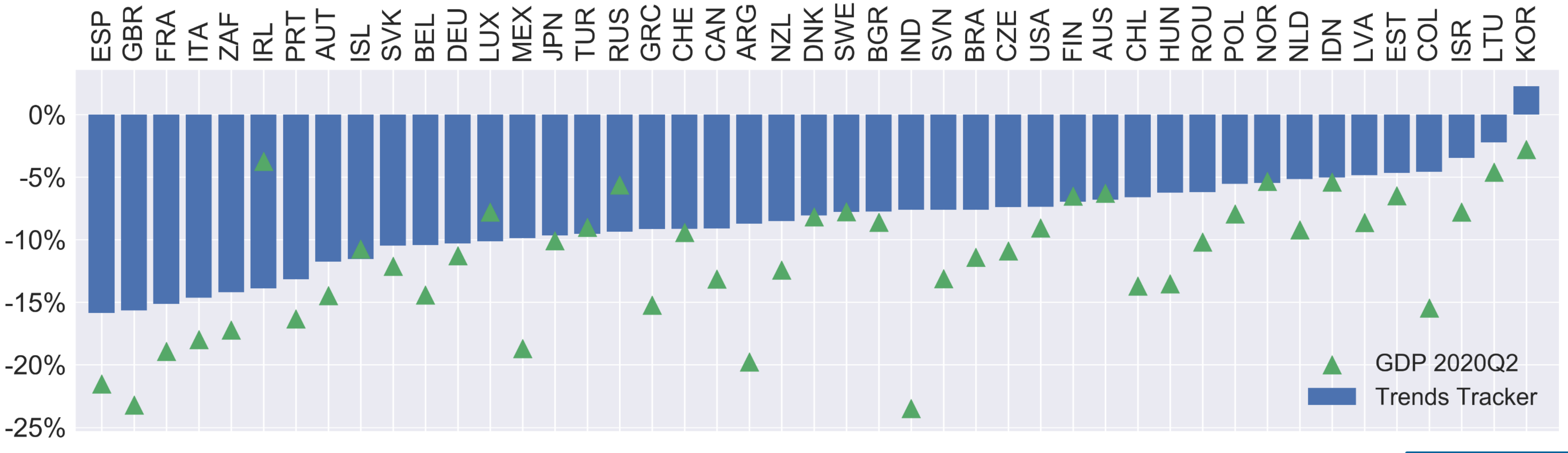




| G7 | RMSE relative to AR(4) |
|---|---|
| France | 0.20 |
| United Kingdom | 0.31 |
| Italy | 0.33 |
| United States | 0.46 |
| Canada | 0.47 |
| Germany | 0.54 |
| Japan | 0.58 |

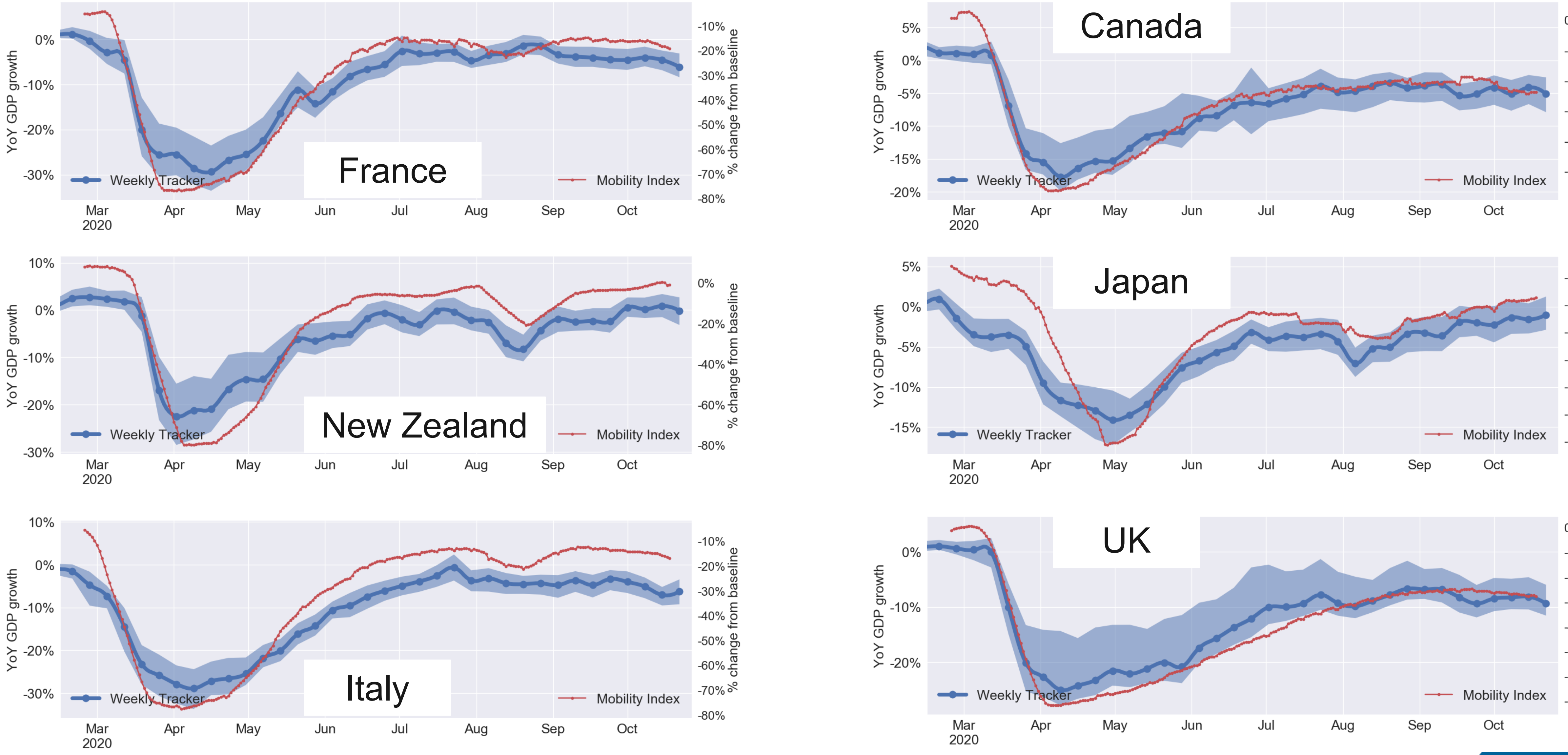
# Focus on 2020 Q2

**The tracker captures the downturn in all 46 countries but one (KOR)** :

- 30 (out of 46) projections are « accurate » (by a 4pp margin)
- Mean absolute forecast error is **3.8pp** (= 34% of the mean value)

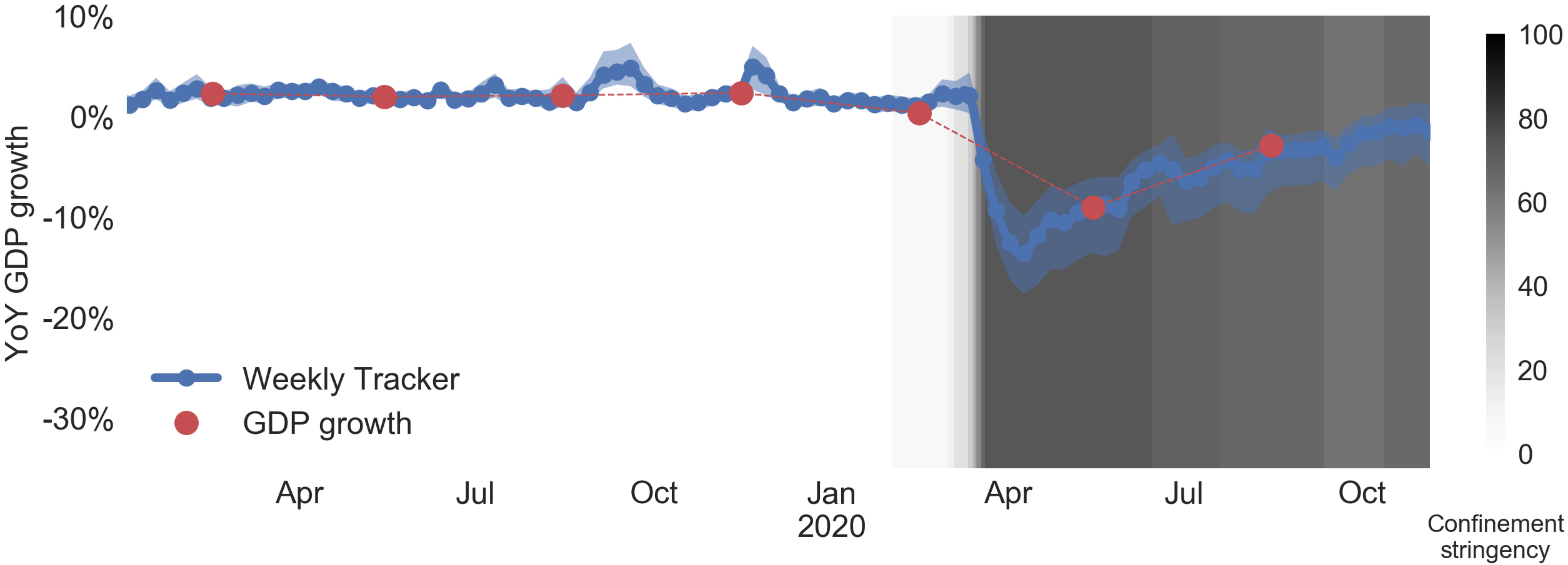


# Benchmarking against another high-frequency indicator: Google Mobility reports

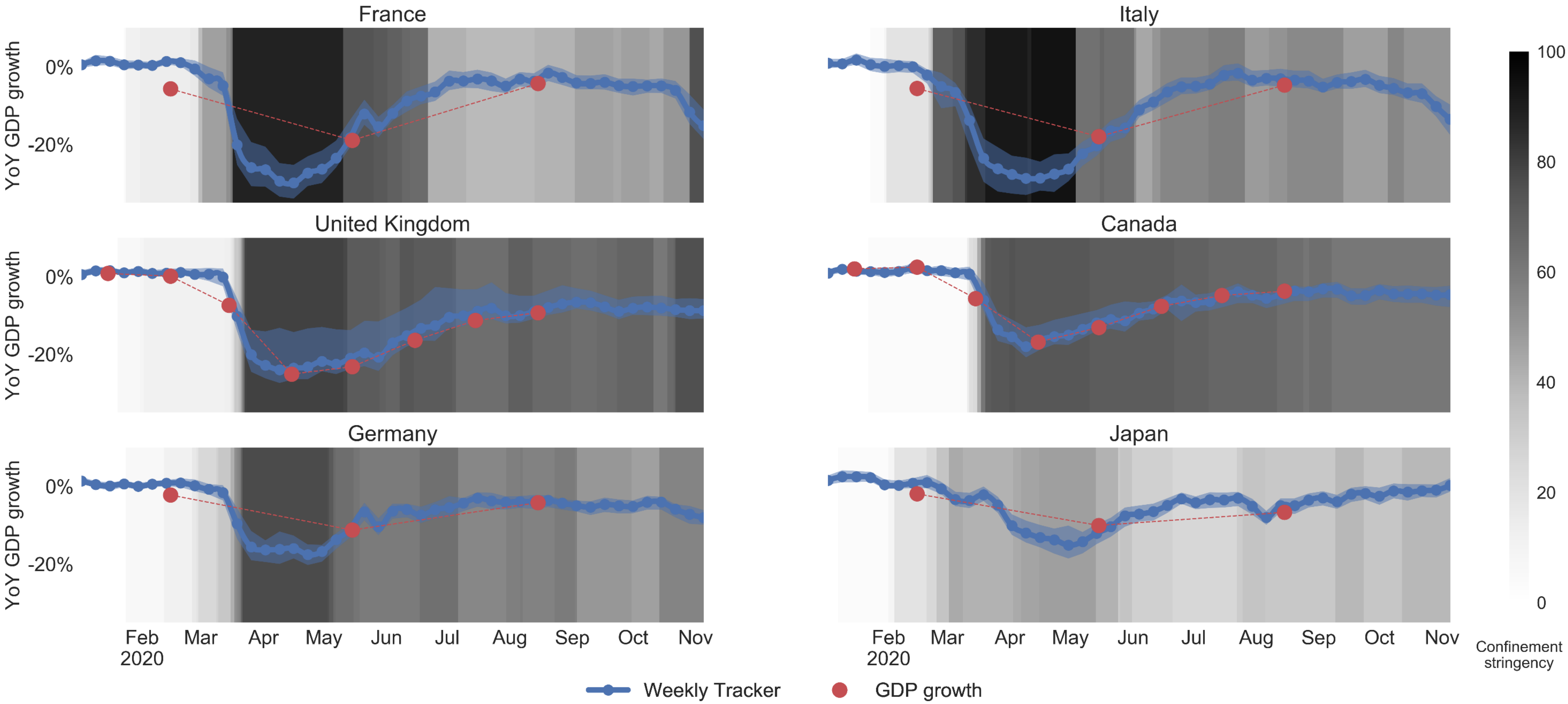


# THE OECD WEEKLY TRACKER

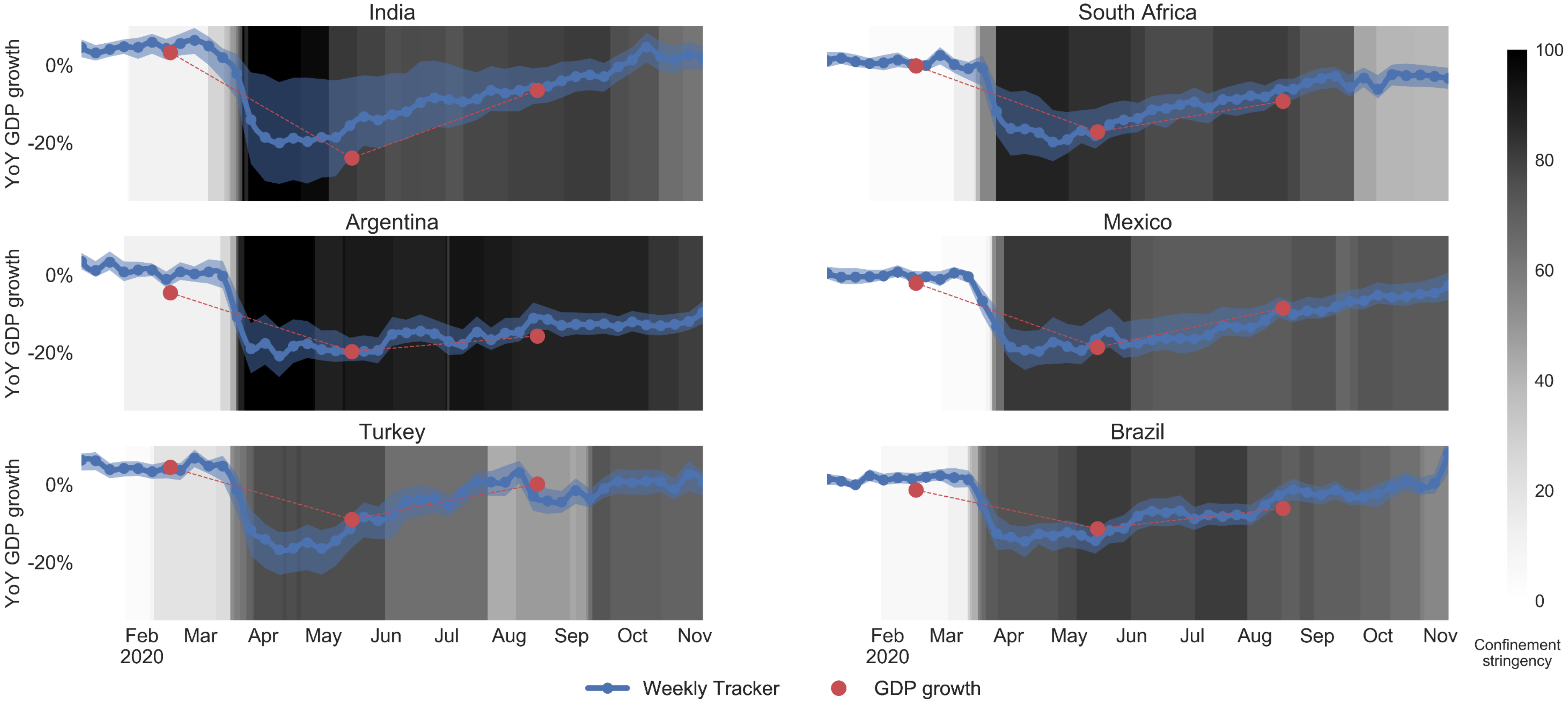
# Weekly Tracker: USA



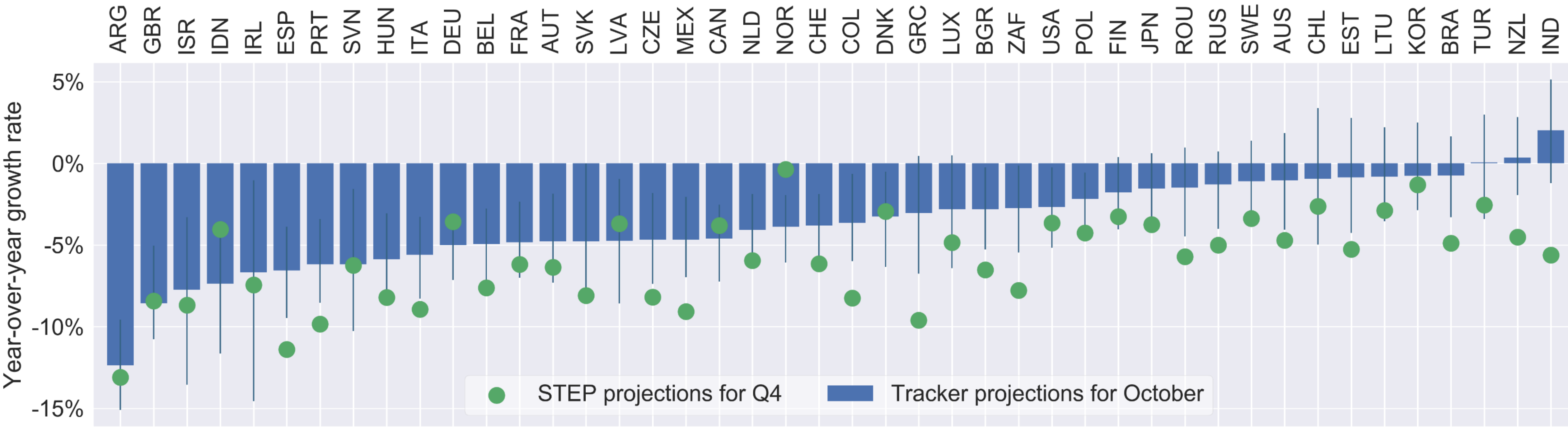
# Weekly Tracker: selected AMEs



# Weekly Tracker: selected EMEs

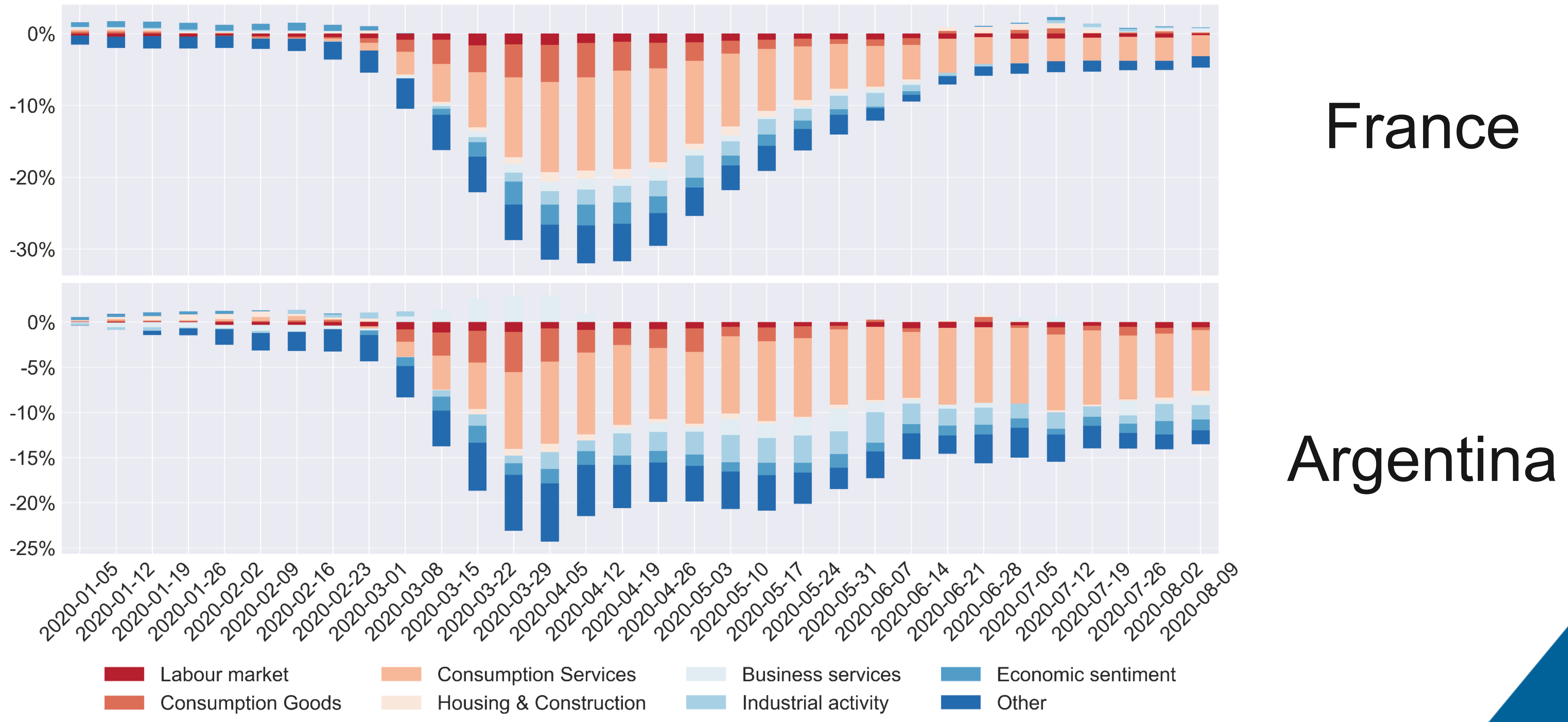


# Early insights for Q4

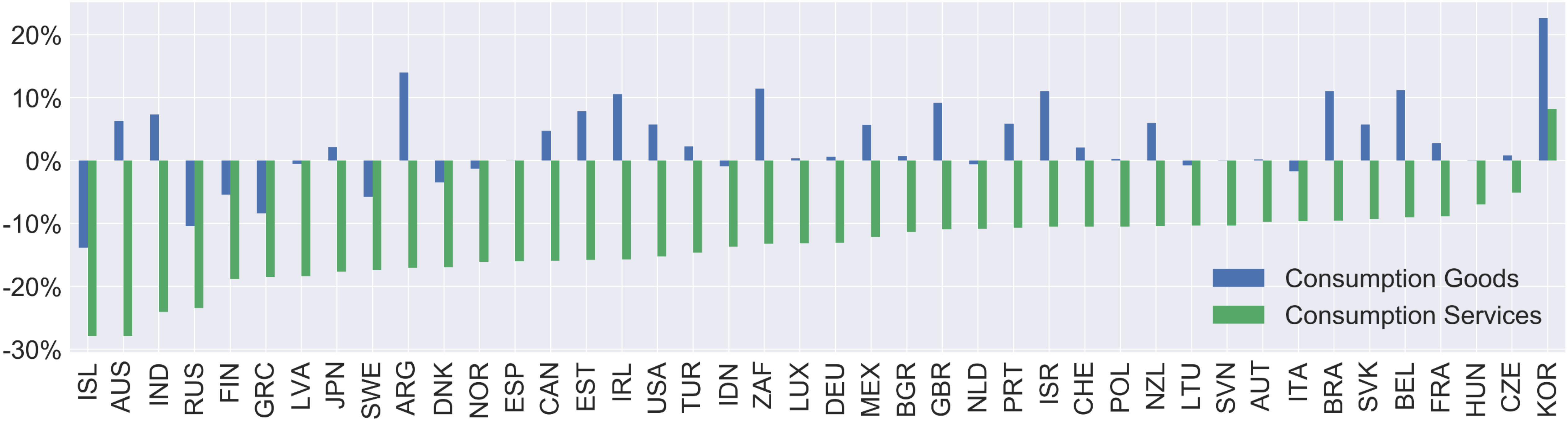


# SECTORAL ANALYSIS

# Variable contributions to GDP growth prediction



# Search Intensities: yoy growth rate in 2020-08

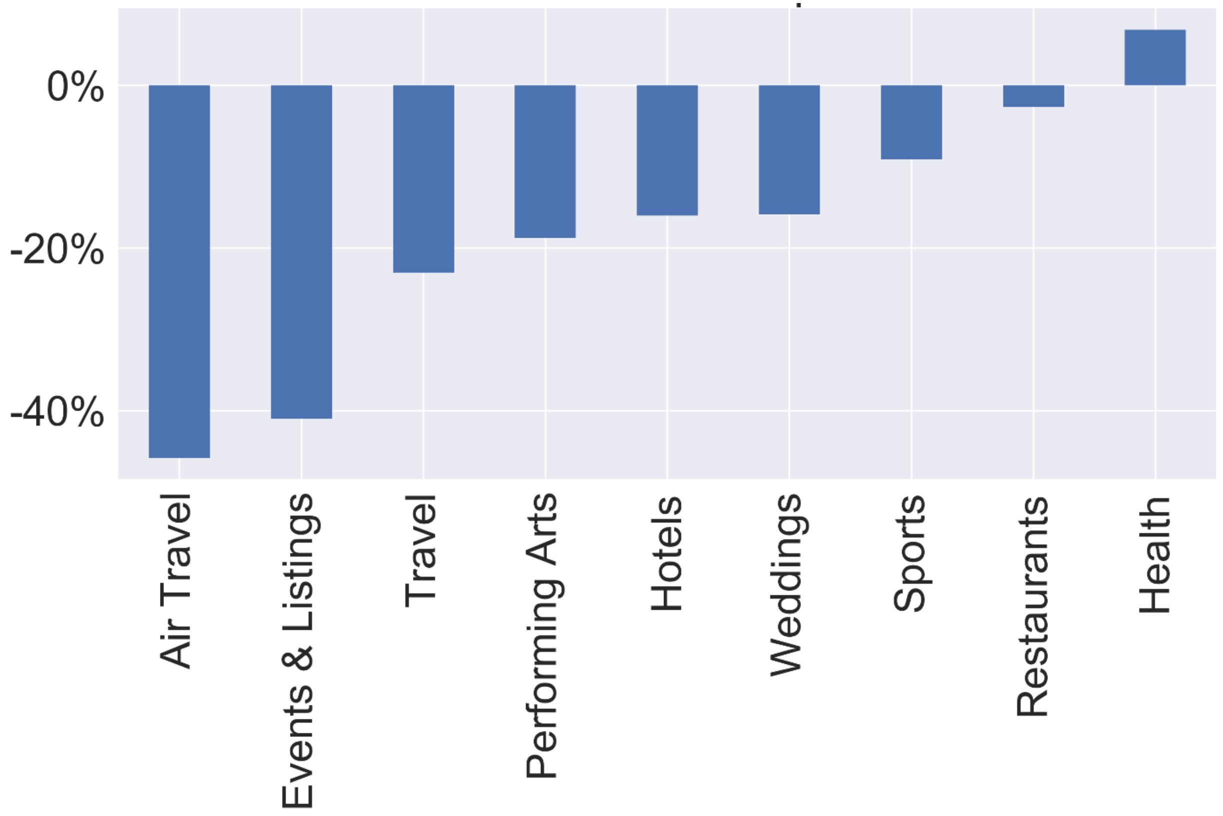



Simple means in yoy growth rates of GT categories picked up by the model as strong GDP predictors

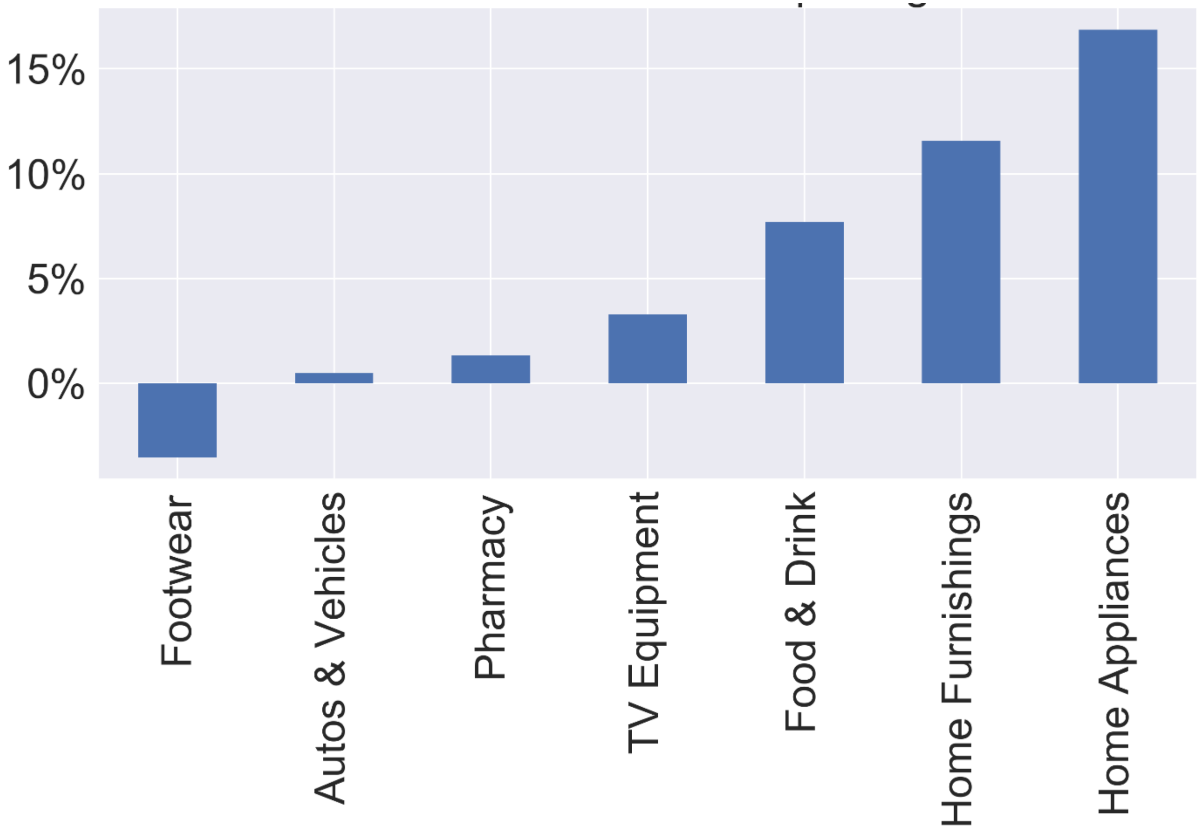
- **Consumption goods**: Food & Drink, Vehicle Brands, Energy & Utilities, Vehicle Shopping, Camera & Photo Equipment, Music Equipment & Technology, Home Appliances, etc
- **Consumption services**: Performing Arts, Travel, Sports, Restaurants, Arts & Entertainment, etc

# Search intensity for selected consumption items 2020-08, yoy, OECD+G20

## A. Selected consumption **services**




## B. Selected consumption **goods**



MANY THANKS !

# Extracting the common trend

1) **Log transformation**: $svi_{ct} = sv_{ct} - svt_t + C_c$.

2) **HP Filter** in order to extract long-term trends.

3) **PCA** is used to extract the common term, that is the log total search volume

The resulting common term is substracted from each series