The Nowcasting Lab: Live Out-of-Sample Forecasting and Model Testing*

Philipp Kronenberg†, Heiner Mikosch†, Stefan Neuwirth†‡, Matthias Bannert†, and Severin Thöni†

†KOF Swiss Economic Institute, ETH Zurich
‡State Secretariat for Economic Affairs SECO

February 9, 2023

Abstract

The Nowcasting Lab is an automated code-database-website environment for GDP forecasting. It generates nowcasts and one-quarter ahead forecasts for quarterly GDP growth of the United States, the euro area, and currently 14 European economies using several forecasting models and a large amount of data. The predictons are updated daily and released on a website together with detailed additional information. forecasting practitioners can use the website to support their own work. Researchers can use the lab to monitor and test the performance of forecasting models in a live out-of-sample environment. All predictions and input data are stored in a daily vintage database which can be used for real-time forecasting studies. As an application, we analyze the live out-of-sample now- and forecast performance of four common mixed-frequency forecasting models crisis. The models failed to predict the large fluctuations in GDP growth in the year 2020. For the year 2021, there are large differences in prediction accuracy depending on the considered economy and model used.

JEL Classifications: C53, C82, E37

Keywords: Forecasting, nowcasting, GDP, automated, live, real time

*Authors' contact: kronenberg@kof.ethz.ch, mikosch@kof.ethz.ch, stefan.neuwirth@seco.admin.ch, bannert@kof.ethz.ch, and s.thoeni@gmx.ch. The authors thank Maurizio Daniele, Rebecca Gerosa, Frédéric Pellet, Tim Reinicke, Corinne Schibli-Lozano, Flurina Schneider, Stefanie Siegrist, Anne Stücker, and Severin Schweizer for valuable contributions to the creation of the Nowcasting Lab.

1 Introduction

The COVID-19 crisis has raised awareness that the field of macroeconomic forecasting faces several challenges. First, policymakers and the public demand constantly updated forecasts, but most public policy organizations publish their forecasts only quarterly or semiannually. Second, academic forecasting research is constantly producing new insights, but it often takes a long time for new methods to be used in applied forecasting practice. Third, research papers usually demonstrate the utility of new methods using out-of-sample forecast exercises, but often the external validity of new methods remains unresolved, i.e., whether they also achieve good forecasting results for data other than those studied.

Against this background, this paper introduces the Nowcasting Lab, an automated code-database-website environment for live out-of-sample testing of macroeconomic short-term forecasting models and for high-frequency publication of forecast updates. The lab produces nowcasts and one-quarter ahead forecasts for quarter-on-quarter GDP growth for currently the following economies: the United States, the euro area, Germany, the United Kingdom, France, Italy, Spain, the Netherlands, Switzerland, Poland, Sweden, Belgium, Austria, Portugal, Romania, and Bulgaria.[1] The environment can be extended to predicting additional economies and additional variables, e.g., inflation. The following forecasting models are currently implemented: a mixed-frequency dynamic factor model (MF-DFM) following Bok et al. (2018), a mixed data sampling (MIDAS) model along the lines of Ghysels et al. (2007), the unrestricted MIDAS (U-MIDAS) model of Foroni et al. (2015), a bridge equation model as presented in Schumacher (2016), and several univariate models (random walk, direct and iterative AR, rolling mean, ARIMA). In addition, pooled predictions are produced using forecast combination methods described in, e.g., Timmermann (2006). The integration of additional models and methods is easily possible. The models are fed with 250–350 daily, weekly, and monthly macroeconomic and financial predictor variables for each economy. In total, the data set currently comprises 3640 variables. An elastic net variable selection mechanism following Zou and Hastie (2005) selects the most relevant predictor series for each economy. Each night, the input data as well as the GDP growth predictions are updated and written into the real-time vintage database together with accompanying information. It is also possible to update more frequently or even continuously. The database grows with each update, resulting over time in a

[1] Portugal has been integrated on behalf of the Portuguese Finance Ministry, and Romania and Bulgaria have been added on behalf of the Directorate-General for Economic and Financial Affairs of the European Commission.

large real-time database for macroeconomic, sectoral, and financial time series of the aforementioned economies. All prediction updates immediately appear on the website `https://nowcastinglab.org`. The website also shows the update history, the impact of new data releases on the predictions, past prediction errors, forecast distributions, the contribution of different variable categories to the predictions, uncertainty and risk measures, and other details. The information is presented in graphs, charts, and tables to make it easily digestible.

The Nowcasting Lab intends to contribute to the advancement of the field of macroeconomic forecasting in two ways. First, a generally accepted practice in the research community is to test the out-of-sample predictive accuracy of models based on historical data sets that the researcher already knows beforehand. As a consequence, models are often, maybe even unintentionally, designed to provide good forecast performance for precisely those data on which they are then tested in out-of-sample exercises. In contrast, external validity is not always given sufficient attention in the construction of the models. True out-of-sample testing, i.e., the practice of researchers testing the predictive accuracy of their models with data that they did know at the time they built the models, is rare.[2] Perhaps the cleanest version of true out-of-sample testing is "live" out-of-sample testing, i.e., the predictive accuracy of a model is tested using newly published data that did not yet exist when the model was created and implemented.[3] However, live testing procedures are lengthy, and their implementation is costly. Against this background, we invite researchers to use the Nowcasting Lab for live out-of-sample testing of their new macroeconomic forecasting models. The Nowcasting Lab code is modular, thus, models in different software can be added to the lab. Remote access to the lab code is available. The daily model outcomes can be either published on the website or sent to a specific group of persons only (background testing). Further, the real-time database of the lab can be used for research purposes. For many of the included economies, macroeconomic real-time databases do not exist so far.

Second, the demand of policymakers, analysts, journalists, and others for transparent high-frequency information on the current and future state of the economy is high. However, public policy organizations mostly use their forecasting models internally. The resulting predictions are usually published at a quarterly or half-yearly

[2]Forecasting competitions, such as the M competitions, are true out-of-sample forecast tests (see Hyndman, 2020 for a history of forecasting competitions and Makridakis et al., 2022 on the recent M5 competition). While most of these competitions use historical data, the data are not known to the participants beforehand.

[3]Notably, live testing is necessarily real-time in the terminology of, e.g., Giannone et al. (2008).

interval only. In contrast, the Nowcasting Lab provides daily updated predictions on current and future GDP growth in various economies based on a broad data set. The predictions are presented together with meta-information and additional statistics that are helpful for interpretation. Practitioners can consult the website to support their own work. The starting page of the Nowcasting Lab website gives an easily digestible overview of the most important information, and the detail pages provide expert information.

As an empirical use case, we examine the revisions to quarterly GDP during the COVID-19 crisis. For most economies, GDP revisions were massive, especially for 2020, suggesting that the first official GDP estimates were not very accurate. Next, we analyze the live out-of-sample now-/forecast performance of the aforementioned models for quarterly GDP growth during the COVID-19 crisis. To our best knowledge, we are the first paper to study the performance of the models during the crisis for many economies. It turns out that none of the models adequately captured the strong GDP fluctuations during 2020. In the model comparison, the MF-DFM performed best, the U-MIDAS model second best. But even these two models generated large prediction errors. One reason for this is that the AR component of the models distorted the predictions due to the abrupt GDP fluctuations, whereas the inclusion of an AR process normally improves the prediction accuracy. For 2021, the models' quarterly GDP growth predictions improved overall, with high forecast accuracy for some economies and quarters, but not for others. On average across all economies, either the MF-DFM or the U-MIDAS model performed best, depending on whether one uses the first or the last available GDP release for the forecast error calculation. However, the predictive accuracy of the models varied widely by economy. This suggests that it can be problematic to draw general conclusions from testing the predictive accuracy of models only for one economy (missing external validity). Further, our analysis reveals that GDP growth now-/forecasts for 2020Q1–2021Q4 improve when the COVID-19 crisis period is excluded from the models' estimation sample, instead of always extending the estimation sample to the most recent observation. Thus, smart handling of the COVID-19 crisis in model estimation is important for future research.

The remainder of the paper is structured as follows. Section 2 discusses the needs of different groups in the field of macroeconomic forecasting and how the Nowcasting Lab meets these needs. The section also reports on previous work on automated macroeconomic now- and forecasting. Section 3 describes the different parts of the Nowcasting Lab, i.e., the database, the forecasting models and methods, the website, the workflow management, etc. Section 4 presents the empirical application and Section 5 concludes.

2 Motivation and design principles

2.1 Challenges and needs

The demand from researchers, policymakers, practitioners, journalists, and others for real-time information on the current and future state of the economy is high. However, the real-time production and publication of macroeconomic now- and forecasts is a great challenge.

Although many **academic researchers** have the knowledge to implement state-of-the-art nowcast and short-term forecast models, it would be often difficult for them to periodically re-run their models and publish the results timely and in an accessible way for the public. They often lack the IT infrastructure and manpower to automatize their models and data procedures and to establish automated forecast releases. Furthermore, researchers often build highly sophisticated models, which are designed or optimized for a specific academic application but less for general use.

Applied researchers in, e.g., central banks, economics ministries, research institutes, and international organizations update their workhorse nowcast and short-term forecast models periodically. However, with few exceptions (see Section 2.2) the prediction updates are only disseminated internally. They usually enter a broader forecasting process and the results are published in reports only a few times a year, without reference to the employed models. Another issue is that applied researchers often lack the time to implement new models from academic research and to maintain serious models and rich data sets for other economies than their respective national economy. **Forecasting practitioners** in the financial sector tend to generate and publish forecast updates at shorter intervals than public institutions. However, they often rely on simple models or judgemental forecasting based on visual data inspection.

Analyzing the demands and constraints of the different groups, which demand macroeconomic predictions or which are directly involved in macroeconomic forecasting, we identify several **unsatisfied needs**:

1. There is a general desire about timeliness of forecasts. The forecasts should always incorporate the most recent data available and should be published at

high frequency.

2. Transparency is very important. People are dissatisfied with black box solutions and want to know what is behind a forecast or a forecast update. It should always be made transparent, which data update has led to a change in the predictions.

3. Replication should be possible. The employed data, models, and model specifications should be communicated transparently, such that the results can, in principle, be reproduced.

4. Only results that can be easily accessed and quickly understood will be used. It is therefore of utmost importance to present the results in a form that is easily digestible for the various stakeholders.

5. The aforementioned requirements must be met without high recurring effort. This can only be achieved by fully automatizing the high-frequency updating of the input data, the forecasting models, and the forecast releases. A tailor-made technical infrastructure is needed for this. Once the process is up and running, humans only play a role in monitoring the process.

2.2 Previous work

The Nowcasting Lab is not the first project that presents up-to-date model-based macroeconomic short-term predictions on a website. Pioneering work has been done by the Federal Reserve Bank of Atlanta and the Federal Reserve Bank of New York. Since 2014 the Atlanta Fed publishes on www.atlantafed.org/cqer/research/ gdp-now a nowcast of US GDP and accompanying information based on its GDPNow model (Higgins, 2014). This model synthesizes bridge equations, which relate GDP subcomponents to monthly time series, with the DFM of Giannone et al. (2008). Currently, the nowcast is updated six or seven times a month. From 2016 to 2021, the New York Fed published on www.newyorkfed.org/research/policy/now-cast the Nowcasting Report which comprised a nowcast and accompanying information for US GDP. The underlying forecast model was a MF-DFM (Bok et al., 2018).[4] Until September 2021, the nowcast was updated each week. Both the Atlanta Fed GDP-Now and the New York Fed Nowcasting Report provide(d) GDP nowcasts for the United States only. Also, they do not provide model plurality, but rely on only one model.

[4]A slightly modified version of the MF-DFM of Bok et al. (2018) is also included in the Nowcasting Lab (see Section 3.2).

Further, the website www.euroareanowcast.com, which went online in 2021, provides weekly updates of GDP nowcasts for the euro area, Germany, France, and Italy (Cascaldi-Garcia et al., 2021). Moroever, the Finnish research institute ETLA publishes on www.etla.fi/en/etlanow nowcasts and 1-to-4-month ahead forecasts for unemployment of all EU member economies (Anttonen, 2018). The forecasts are updated daily, and the employed model is a Bayesian vector autoregression (BVAR) that uses data from Google Trends, among other sources.

In the wake of the of the COVID-19 crisis and the urgent demand for timely information about the state of the economy, new progress has been made in providing real-time GDP forecasts to the public. Many forecasting institutions started to create high-frequency business cycle indices or forecasts by collecting alternative high-frequency data. Some of them published their predictions with regular updates on their websites. Examples are Lewis et al. (2021) for the United States, Eraslan and Götz (2021) for Germany, Wegmüller et al. (2021) and Eckert et al. (2020) for Switzerland, Lourenço and Rua (2020) for Portugal, Fenz and Stix (2021) for Austria, Adam et al. (2021) for the Czech Republic, and Woloszko (2020) for all OECD economies. This development proved, on the one hand, the increasing demand for timely forecasts and, on the other hand, the feasibility to produce them in an easily accessible and automatized way. Most of the models are specified explicitly for the COVID-19 crisis and use data that are specifically useful for tracking this crisis. It remains to be seen how many of these COVID-19 related high-frequency indices and forecasts will still be published after the end of the crisis.

2.3 Our solution

The Nowcasting Lab extends existing nowcasting projects in several ways. First, the input data and the forecasting model outputs are **updated daily** and they are displayed on the website immediately afterwards. The daily update guarantees that the user is quickly informed about changes in the predictions. Intra-day updates are possible to see the effect of important data releases immediately. In contrast, other nowcasting projects provide only monthly or weekly updates (see, e.g., the Atlanta Fed GDPNow and the New York Fed Nowcasting Report cited in Section 2.2). Second, instead of relying on only one model, the Nowcasting Lab comprises **several widely accepted forecasting models** which are used as workhorse models in, e.g., central banks and research institutes. The user can compare the predictions and the past performance of the models with each other. Third, and related to this, additional models can be easily included due to the **modularity** of the code.

This makes the lab suitable as a **real-time testing facility for new forecasting models**. We invite researchers to integrate newly developed models into the lab. They can use the lab as a facility to test models in a live out-of-sample environment and to show them to other researchers.[5] The ultimate goal is to have a **plurality of models** and techniques in the lab. Fourth, the lab produces GDP nowcasts for **several economies** instead of only one. Additional economies can be easily integrated since the code-database-website environment is modular. Fifth, the lab comes with a **daily vintage real-time database** for macroeconomic and financial time series.[6] Each day, a new vintage, which includes the newest data releases, is added to the database. The database allows to replicate real-time conditions for, e.g., forecasting exercises. For many of the included economies, no real-time data set has been made available so far. We make the database accessible to researchers in our Data Center.

Next to the aforementioned contributions, we would like to highlight several further important aspects of the Nowcasting Lab: First, the whole update and forecasting process, including recurring model re-estimation and variable selection, is **completely machine-driven**. Once the process is set up, no human judgment is involved in the process. Instead, human action is limited to monitoring the process. Second, the Nowcasting Lab addresses **different target groups**. The website presents the most important forecasting information in a digestible form for the broader public or for practitioners who need quick information (starting page). In addition, the website provides more complex and detailed information for forecasting experts (separate detail pages). Third, **transparency** is ensured by info buttons that explain the underlying routines, models, and techniques. Fourth, in view of the continuous development of the code, the use of the distributed version control system Git guarantees **reproducibility** of the results.

[5] Models do not necessarily need to be shown on the website. Instead, their real-time performance may be tested in the background. Notably, due to the modular set-up, models written in different software languages can be incorporated into the Nowcasting Lab. Thus, each researcher can work in the language of her choice.

[6] This relates our project to real-time database projects. For instance, the Federal Reserve Bank of Philadelphia provides the Real-Time Data Set for Macroeconomists (RTDSM), a large-scale database with historical vintages (Croushore and Stark, 2001). The Federal Reserve Bank of Saint Louis runs the ArchivaL Federal Reserve Economic Data (ALFRED). This real-time database is based on archived series from the FRED database (see a discussion in Anderson, 2006).

3 Structure of the Nowcasting Lab

The Nowcasting Lab consists of three interrelated parts: the code that contains the forecasting models; the real-time database from which the forecast models are fed and into which the prediction output is read; and the website that presents the prediction output and various meta-information. In this section, we present the three parts. In addition, we describe the interfaces between the parts and the overall automatic workflow management provided by an Apache Airflow process. Figure 1 provides an initial overview of the Nowcasting Lab structure.

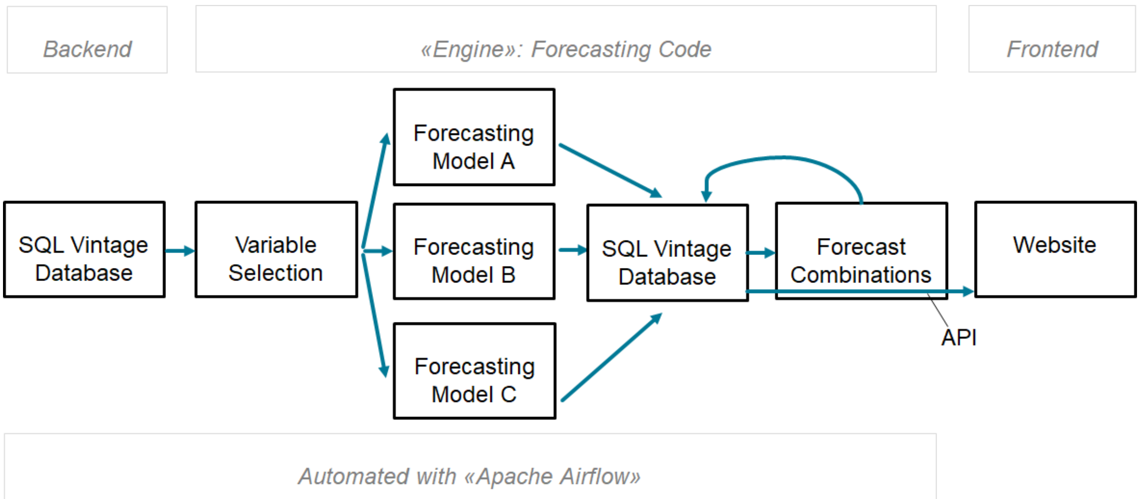

Figure 1: Structure of the Nowcasting Lab.

3.1 Data

Variable set: We compile a large set of daily, weekly, and monthly macroeconomic, sectoral, and financial time series ("predictor variables") for forecasting quarterly GDP growth.[7] Table 1 displays the variable categories. We currently use 3640 predictor variables in total, i.e., an average of 303.33 variables for each of the 12 economies that are currently included in the lab. All data come from Macrobond, a meta-provider that compiles a wide range of data sources. By using a single data provider, we ensure an efficient and automatic data update process.

Variable transformations and pre-selection: We boost the number of series in the data set by applying multiple transformations to all variables: level, month-on-month growth rate, 3-month-on-3-month growth rate, year-on-year growth rate, 1-month absolute change, 3-month absolute change, and 1-year absolute change.

[7]While we include weekly and daily variables in the database, they are for now used in monthly frequency only. On each day, we fill up each daily/weekly variable until the end of the month using a random walk forecast. We then aggregate the daily/weekly values to monthly frequency.

Table 1: Data overview

Category	Variable example	Number
Manufacturing	Industrial production	18
Construction	Residental construction orders	2
International trade	Imports	38
Retail and consumption	New car registrations	13
Labor market	Vacancies	12
Business surveys	Purchasing managers index (PMI)	67
Consumer surveys	Consumer climate	11
Prices and monetary variables	M3	34
Financial variables	Stock market index	43
Foreign variables	PMI of trading partner economy	66

Notes: The third column ("Number") displays the average number of time series per economy.

The data set is further doubled by including trend-adjusted variants of all the aforementioned transformations. The trend adjustment is done by subtracting the moving average over the past year from each time series observation. Subsequently, the variables or variable transformations are checked for stationarity. This is done by applying the augmented Dickey-Fuller test adjusted for the effective sample size as well as the Phillips-Perron test (Dickey and Fuller, 1979 and Phillips and Perron, 1988).[8] The tests are performed for the same period of time for which the models are estimated afterwards. We keep a variable or variable transformation only if both tests reject the null hypothesis that there is a unit root. Afterwards, we sort out variables or variable transformations with presumably weak predictive content by applying the elastic net method following Zou and Hastie (2005). Appendix A provides details.

Real-time database: For data storage, we use a PostgreSQL database, an open-source relational database management system. The database is equipped with a vintage data system to store real-time data. All time series that we use to produce the forecasts are stored in the real-time database, along with meta-information such as when the series were updated. Previous data states are kept in the vintage data system. This makes it possible to replicate the data state that a forecaster has had each day since data collection began, both in terms of data availability and data revisions. Thus, a large and ever-increasing real-time database for daily, weekly, and monthly macroeconomic, sectoral, and financial time series, in addition to the quarterly GDP series, is created over time.

[8]For the Dickey-Fuller test, the optimal lag selection is chosen according to the Bayesian information criteria. For the Phillips-Perron test, the appropriate number of autocovariance lags to include in the Newey-West estimator of the long-run variance is chosen by lag $L_t = 4 * (T/100)^{1/4}$ with T being the length of the respective time series.

Reproducibility: To allow replicability of the forecasts, in addition to the daily input data storage, the code used to generate the forecasts is stored on a repository of GitLab, an internet hosting service for software development and distributed version control. This makes it possible to restore old versions of the code. Last, the forecast output is also stored in the vintage database. Thus, past predictions can be evaluated quickly and easily.

3.2 Forecasting models

Workhorse models: According to our reading of the literature, at least four classes of mixed-frequency GDP nowcasting and short-term forecasting models are nowadays standard both among forecasting practitioners as well as in academic research. The model classes are mixed-frequency dynamic factor models (MF-DFMs), mixed-data sampling (MIDAS) models, bridge equation models, and unrestricted mixed-frequency models.[9] The Nowcasting Lab implements four models, one of each class: the MF-DFM of Bok et al. (2018) in a single-factor version, the autoregressive single-predictor MIDAS model along the lines of Ghysels et al. (2007) and Clements and Galvão (2008), the single-predictor bridge equation model as described in Schumacher (2016), and the autoregressive single-predictor unrestricted MIDAS (U-MIDAS) model of Foroni et al. (2015). Appendix B describes the model specifications used in the Nowcasting Lab in detail. Since the Nowcasting Lab is coded in a modular form it is easy to integrate new forecasting models in the future. It is also possible to integrate and run models without publishing them on the website. Selected users can then monitor these models and their performance via automated daily reports, for example. Storing the model outputs in the database ensures that the real-time performance of the models during different time periods can be analyzed retrospectively at any time.

[9] A fifth model class is mixed-frequency VARs, which are increasingly used for nowcasting. Foroni and Marcellino (2013) and Camacho et al. (2013) provide literature surveys on mixed-frequency GDP forecasting models. Notably, some models combine elements of two or more classes, such as, e.g., factor MIDAS models (Marcellino and Schumacher, 2010). Also, a number of models include additional important features and could therefore be considered separate model classes (e.g., models with regime switching or models with time-varying parameters). In recent years, machine learning algorithms are increasingly applied to forecasting GDP, or machine learning techniques are combined with models from the aforementioned classes (e.g., Masini et al., 2020 for an overview). The elastic net technique that we use in the lab (see Section 3.1) originates from machine learning. The integration of mixed-frequency VARs and further machine learning techniques or models into the Nowcasting Lab is a goal for the future. The COVID-19 pandemic has stimulated research on new model variants that can better capture strong and sudden GDP fluctuations (see references in Section 1). However, the classification of mixed-frequency GDP forecasting models into the aforementioned classes still seems reasonable.

Forecast combinations: Following common practice in the literature, we refrain from (U-)MIDAS models and bridge equation models with multiple predictor variables. Instead, we include always only one predictor variable at a time into a model and then pool the forecasts stemming from all individual models using a forecast combinations procedure. Appendix C explains this procedure in detail.

Benchmarks models: In addition to the aforementioned workhorse models, the Nowcasting Lab currently includes the following benchmark forecasting models: a random walk forecast, a direct as well as an iterative autoregressive (AR) model with the optimal lag length determined according to the Bayesian Information Criterion (BIC), a rolling in-sample mean forecast with the optimal in-sample window length selected according to the root mean square forecast error (RMSFE), and an autoregressive integrated moving-average (ARIMA) model with the number of autoregressive lags, moving average lags and the order of differencing being determined according to the BIC. Appendix B provides a more detailed description of the models. The predictions resulting from the benchmark models do not enter the forecast combinations procedure. Instead, they are shown separately on the Nowcasting Lab website.

3.3 Website

The website `www.nowcastinglab.org` presents for the United States, the euro area, and several European economies daily updated quarter-on-quarter GDP growth forecasts for the current quarter (nowcast) and the upcoming quarter (one-quarter ahead forecast) as well as various additional information.The prediction of a quarterly GDP observation always starts 2 quarters before its release, i.e., the forecast for GDP_{t+2} starts on the release day of GDP_t. The following 16 economies are currently included: the United States, the euro area, Germany, the United Kingdom, France, Italy, Spain, the Netherlands, Switzerland, Poland, Sweden, Belgium, Austria, Portugal, Romania, and Bulgaria. A direct linkage of the website with the database ensures that forecast updates appear immediately on the website (see Section 3.4). The website is divided into two parts, which will be discussed in turn.

Home page: The home page of the website provides a quick overview. It presents in graphical form the current GDP growth prediction for the aforementioned economies (see Figure 2). A drop-down menu allows the selection of the target period of the prediction, i.e., either the nowcast or the one-quarter ahead forecast. The home page further includes a graph showing the history of a selected economy's GDP growth prediction. The economy can be selected either from the drop-down

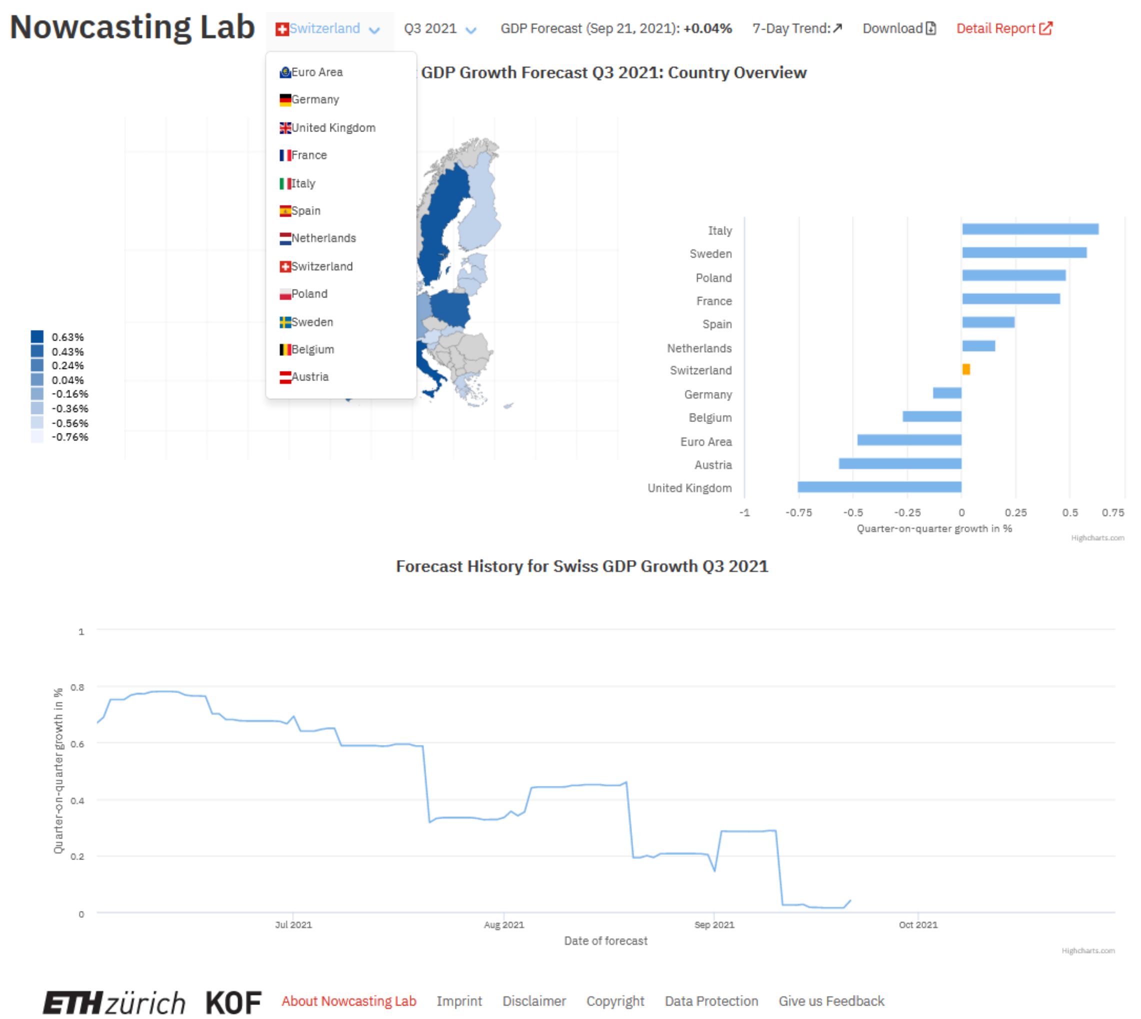

Figure 2: Home page of the Nowcasting Lab website. The figure shows the overview part of the website `www.nowcastinglab.org`.

menu or by clicking on it on the map.

Detail page: The user can choose on the overview page for which economy and for which target period she wants to open a detail page (see Figure 3). The detail page is designed for forecasting practitioners, analysts, and researchers. It contains the following elements:

1. Headline elements showing the current prediction along with a precision measure (mean squared forecast error of the last 12 periods), a forecast interval, and a measure of upside/downside risk (Pearson skewness measure).

2. A time series plot showing how the prediction has evolved over time since the start of the prediction 2 quarters prior to the release of GDP for the selected target period.

3. A stacked bar chart showing either the forecast contributions or the forecast

impacts (i.e., the changes in contributions from the previous day) of different variable categories. A switch allows to change from the contribution chart to the impact chart and vice versa. Note that not all models provide all of those measures. For example, the DFM does not provide contributions, but only impacts. As a consequence, the pooled predictions, which include the prediction of the DFM, do not come with forecast contributions either.

4. A histogram showing the frequency distribution of the predictions across all individual predictor variables. Pooling of the individual forecasts then yields the overall forecast for each model (see Section 3.2).

5. A release calendar that shows the new data releases for each day and displays their impact on the prediction. By clicking on a data release, a pop-up window opens, which provides a detailed description of the series (title, variable description, category, source, release date, reference period, unit, current value, previous value, impact).[10]

6. A graph showing the historical GDP development as well as the current and past predictions for the selected economy.

In addition, the detail page includes a drop-down menu, where the user can select either one of the model types (MF-DFM, MIDAS, bridge equation, U-MIDAS, ARIMA, iterative AR, direct AR, random walk, rolling in-sample mean) or the pooled predictions across MF-DFM, MIDAS, bridge equation, and U-MIDAS. The six aforementioned elements are displayed for all model types as far as applicable. This feature allows to compare the predictions of the different models.

3.4 Workflow management

Automated workflow management: The entire code-database-website environment is managed by an Apache Airflow pipeline. It automatically checks for new data releases once per day.[11] Once a new data release is detected, three consecutive steps are executed: the updated input series are written to the real-time PostgreSQL database as a new data vintage, the full forecast algorithm is executed, and the updated predictions are also stored in the new data vintage. Depending on how many variables were updated, the whole process takes between 5 to 30 minutes for each

[10] Note that, depending on the elastic net selection, some variables might enter the forecast with more than one transformation (see Section 3.1). The displayed impact of the variable is then the sum of all impacts from all selected transformations of this variable.

[11] Intra-day updates can be executed manually to see the effect of important data releases immediately.

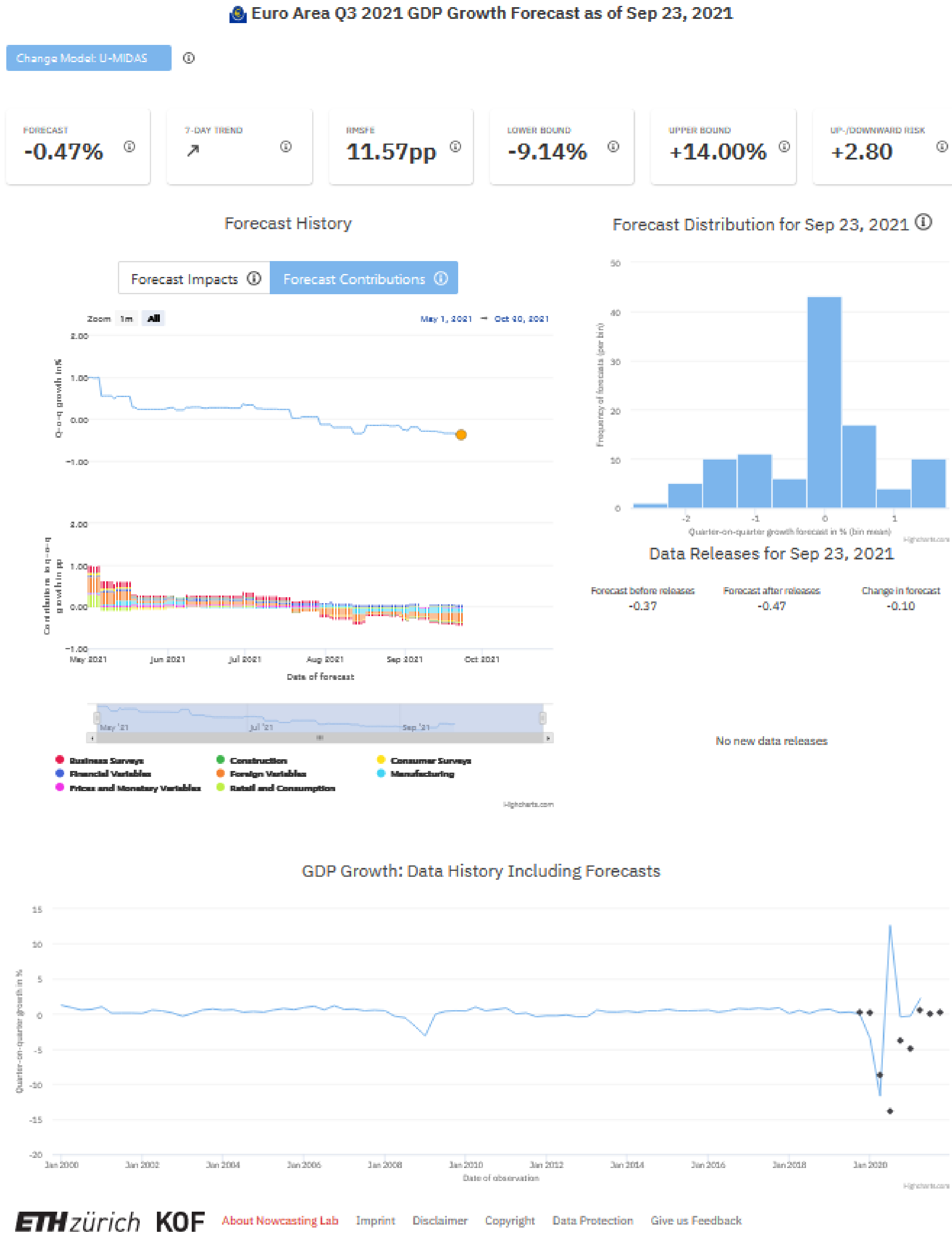

Figure 3: Detail page of the Nowcasting Lab website. The figure shows the detail part of the website `www.nowcastinglab.org`.

economy. The website shows always the latest vintage from the database. Thus, the user can see prediction updates shortly after a new data release.

Interfaces: The real-time PostgreSQL database, the forecasting algorithm, and the website communicate with each other via interfaces. The raw data and accompanying meta-information are loaded from the database into the forecasting algorithm via SQL queries. The output of the forecast algorithm is then written back to the database in the form of tables. These tables contain the necessary information for the different parts (graphs, figures, etc.) of the website. The communication between the website and the database is fully automatic, using Application Programming Interfaces (APIs). Whenever a user opens the website, the latest vintage of the database is delivered to the website in a machine-usable form via the APIs.

Sanity checks: While the code-database-website algorithm is in principle fully automatic, in our experience there are numerous reasons why the run of the pipeline could abort. To prevent this, various sanity checks were implemented. For example, each series is checked to see if the series was properly updated and if the update contains any NA or Inf values. If so, the series is skipped. This ensures that the algorithm continues despite data errors. Based on the sanity checks, we created a log file reporting system to quickly locate errors or other unplanned events.

4 Application

The Nowcasting Lab can be used for real-time data analysis and for live out-of-sample testing of forecasting models. This section presents a use case for the time period of the COVID-19 pandemic. We start by reviewing quarterly GDP growth during the years 2020 and 2021. Next, we examine the revisions of GDP growth which were comparatively strong in 2020. After that, we analyze how the lab's live out-of-sample GDP growth predictions were revised over time and which predictor variables drove these revisions. Further, we compare the lab's live out-of-sample forecast performance of the different models and model specifications. The use case analysis comprises all 12 economies that were already integrated into the lab at the beginning of 2020. The Nowcasting Lab makes it easy to iterate the same analysis for multiple economies. Forecast performance tests of models can thus take place on a broad data basis.

4.1 GDP growth during the pandemic

Figure 4 shows quarterly GDP both in quarter-on-quarter growth rates and in levels (indexed to 2019Q4 = 100) for each economy during 2020–2021. GDP in Spain (ES), France (FR), and the United Kingdom (GB) fell comparatively strongly in 2020Q2 as a consequence of the first COVID-19 wave. In contrast, GDP in Switzerland (CH), the Netherlands (NL), Poland (PL), and Sweden (SE) declined comparatively little. These cross-economy differences can be explained by variations in, e.g., the COVID-19 infection rates, the strength of containment measures, and the sector structures (Muggenthaler et al., 2021). For instance, economies with a comparatively large consumer service sector were affected more strongly. In addition, differences in GDP accounting practices for non-market production, especially education, and healthcare, played a role (Dey-Chowdhury et al., 2021). By 2021Q3 or 2021Q4, most economies returned to, or at least came close to, their respective pre-crisis GDP levels. Only Germany (DE), Spain (ES), and the United Kingdom (GB) lacked behind, whereas the Netherlands (NL) and Poland (PL) already exceeded the pre-crisis GDP level. These differences were due, on the one hand, to persistent differences in the above-mentioned factors and, on the other hand, were related to differential effects of international supply bottlenecks (e.g., Celasun et al., 2022).

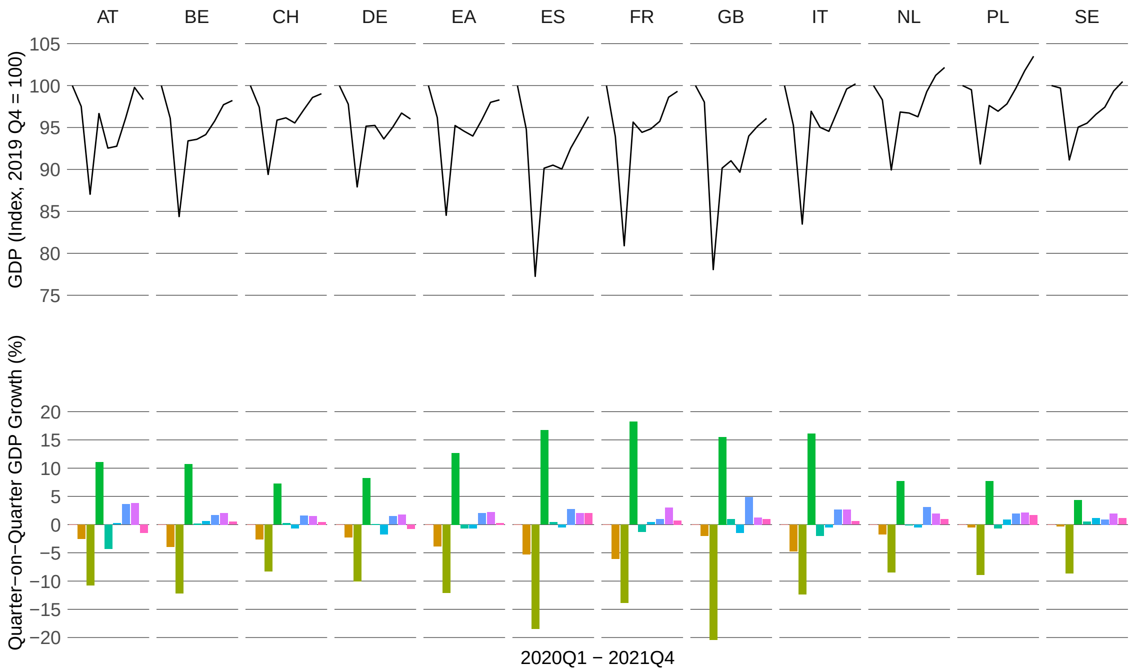

Figure 4: Quarterly GDP during 2020Q1–2021Q4. The figure shows, for each considered economy, the first GDP release for each quarter in 2020 and 2021. Country codes: AT: Austria, BE: Belgium, CH: Switzerland, DE: Germany, EA: euro area, ES: Spain, FR: France, GB: Great Britain, IT: Italy, NL: Netherlands, PL: Poland, SE: Sweden.

4.2 GDP revisions

The real-time data we collect as part of the Nowcasting Lab project reveal that GDP was revised significantly stronger for the first phase of the COVID-19 pandemic than during regular times (e.g., the comparison with the results of Faust et al., 2005). Figure 5 shows, for the four quarters in 2020, the revisions to quarterly GDP growth in percentage points from the respective first GDP release. We make five observations. First, the number of revisions to GDP varies substantially from economy to economy. For example, as of December 5, 2022, French (FR) GDP for 2020Q1 has been revised 20 times since its first release, while Dutch (NL) GDP for 2020Q1 has been revised only eight times.

Second, for most economies, the revisions were quite large for the first three quarters of 2020. For example, the GDP growth estimate for Sweden (SE) for 2020Q3 was revised upward by around 3.2 percentage points between the first and the last vintage. For 2020Q4, with the exception of Austria (AT), revisions were comparatively smaller than for the first three quarters of 2020. GDP revisions tended to be larger in those economies that experienced a higher number of COVID-19 infections. This finding reflects the difficulty of measuring GDP during turbulent times in general and during the COVID-19 crisis in particular (e.g., Dey-Chowdhury et al., 2020b; Jordà et al., 2020).

Third, compared to the first GDP growth estimates for 2020Q1 and 2020Q2, the majority of the national statistical agencies have revised their subsequent estimates upward. In other words, taking the current GDP growth estimates as a benchmark, the agencies initially overestimated the contraction in 2020Q1 and 2020Q2. Notable exceptions are the United Kingdom (GB) and Italy (IT), where the contraction in 2020Q1 was initially underestimated. One explanation is that these economies faced the pandemic and a strict lockdown comparatively early when there was still little experience with measuring GDP correctly and consistently during these unusual times. The statistical agencies of the subsequently affected economies have learned from the experience in the United Kingdom and Italy and even initially tended to overestimate the impact of the crisis on GDP in their respective economies (e.g., Dey-Chowdhury et al., 2020a). No clear revision pattern can be observed for 2020Q3 and 2020Q4. Some statistical agencies initially strongly underestimated the rebound in 2020Q3 and later revised GDP growth upward (e.g., Belgium (BE), Germany (DE), the United Kingdom (GB), and Sweden (SE)). In contrast, other agencies revised GDP growth for 2020Q3 downward (e.g., the Netherlands (NL)) or revised GDP growth slightly as compared to the first release.

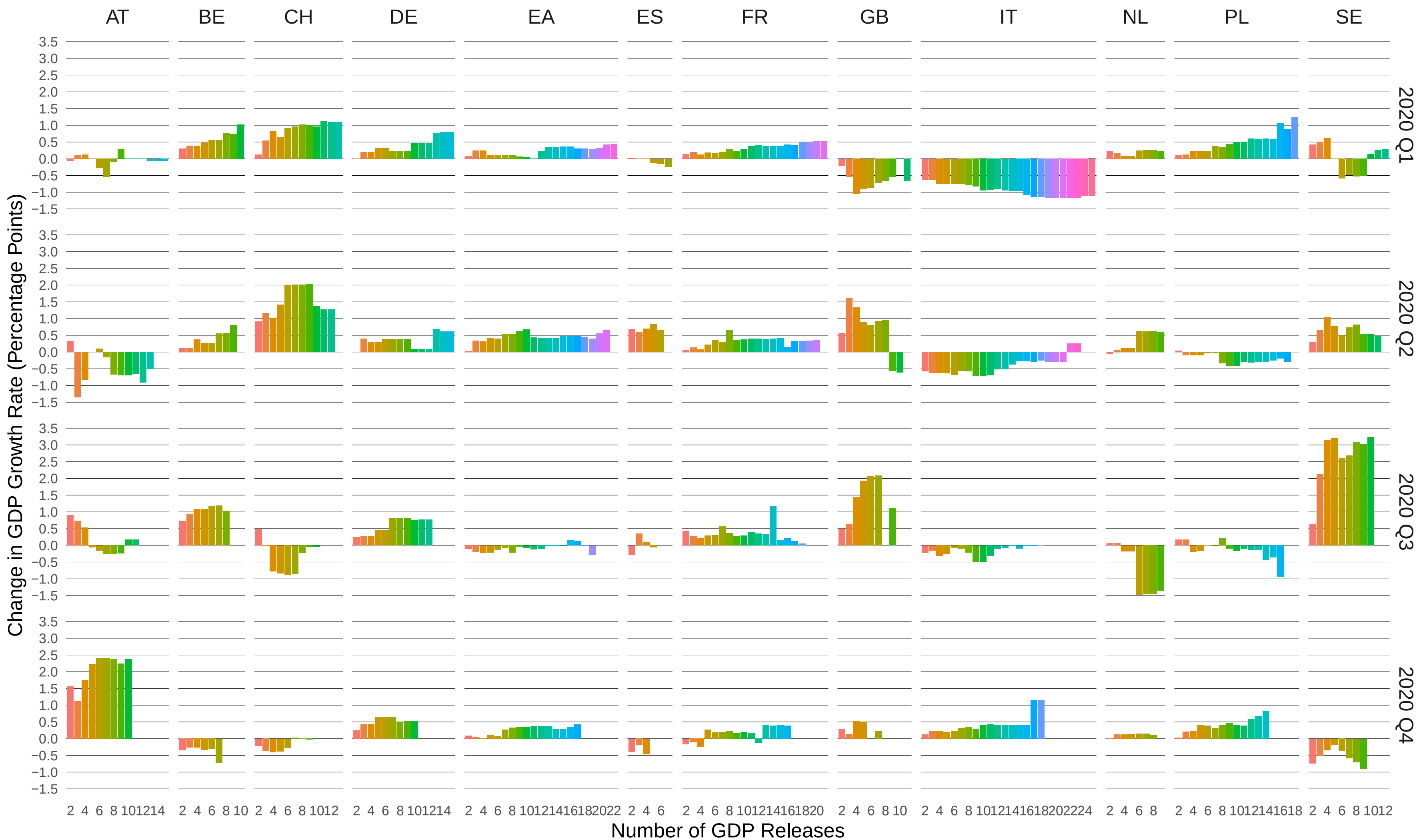

Figure 5: Revisions to quarterly GDP during 2020Q1–2020Q4. The figure shows, for the four quarters in 2020, the revisions (in percentage points) to the first release of the quarter-on-quarter GDP growth rate stemming from the subsequent GDP releases (data considered until 05/12/2022). The number of revisions varies across economies and quarters. See Figure 4 for the country codes.

Fourth, interestingly, for particular quarters, some statistical agencies first revised GDP growth substantially upward and later substantially downward (or the other way around). For instance, Swedish GDP growth for 2020Q1 was revised up by about 0.5 percentage points in the second release, then revised down by 0.5 percentage points in the fifth release and the sixth release respectively, and then revised up again in the tenth release. In the latest vintage, it was close to the GDP growth estimate of the initial release. Another example is the GDP growth for 2020Q2 in the United Kingdom (GB), which was revised sharply upward in the second and third releases, only to be partly reversed afterward. In the ninth release, GDP growth underwent a massive downward revision (–1.5 percentage points compared with the eighth release). Thus, while the contraction in GDP in 2020Q2 appeared to have been initially overestimated, it now seems to have been underestimated. The reverse revision pattern occurred in Italy (IT) for 2020Q2: Until the 21st release, the drop in GDP appeared to have been initially underestimated. In contrast, the 22nd and 23rd GDP releases suggest that the collapse has been initially overestimated.

Fifth, the fact that GDP was strongly revised during the first phase of the COVID-19 pandemic must be taken into account for the forecast evaluation in the following sections. The predictive performance of the models may depend on which GDP release is used as a benchmark in the evaluation.

4.3 Forecast revisions over time

The Nowcasting Lab predicts GDP growth of quarter t each day beginning two quarters before the first GDP release for quarter t. During the first quarter, the predictions are referred to as one-quarter ahead forecasts; during the second quarter, they are referred to as nowcasts. Figure 6 shows how the GDP growth predictions during the COVID-19 pandemic changed from day to day as new data got released over time (see Section 3.1 on the data used). Specifically, the lines in the figure show, for each quarter of 2020 and 2021, the daily development of the pooled GDP growth predictions, starting two quarters before the respective GDP release and ending on the last day before the release. The pooled predictions result from equally weighting the predictions from the MF-DFM, the MIDAS model, the bridge equation model, and the unrestricted mixed-frequency model (see Section 3.2 and Appendices B and C for details). In addition, the dots in the figure show the respective GDP growth realization according to its first release.[12] To avoid clutter, only the euro area (EA)

[12]If we take a later GDP release vintage instead, the general picture does not change, albeit with differences for some economies and quarters.

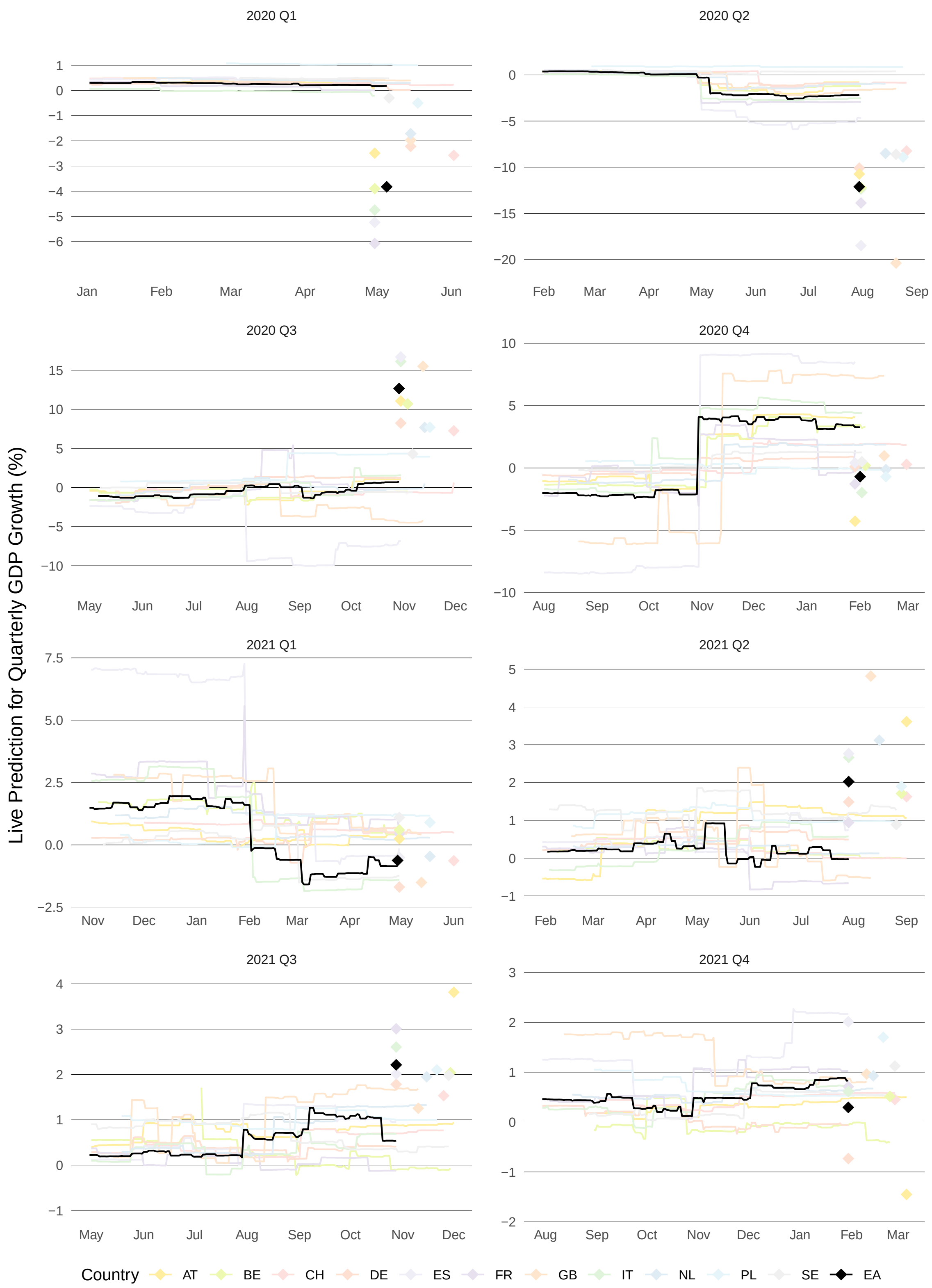

Figure 6: Development of GDP growth live predictions. The figure shows, separately for each economy and quarter in 2020–2021, the development of the daily pooled predictions for quarterly GDP growth (lines) together with the respective GDP growth realizations (dots) according to their first release. The predictions for GDP growth of quarter t start on the first day after the release of GDP of $t-2$ and end on the last day before the release of GDP of t. The pooled predictions are equally weighted averages of the predictions resulting from the workhorse models (see Appendix C). See Figure 4 for the country codes.

is highlighted in Figure 6.[13] Figure 8 in Appendix E shows separate subfigures for each country.

For 2020Q1, the GDP growth predictions for most economies were relatively stable and above or around zero from January to May 2020 and failed to capture the negative growth realizations (see the upper left sub-figure of Figure 6). For 2020Q2, the predictions for most economies were strongly revised downward from May 2020 onward. For instance, the GDP growth prediction for France dropped from around zero to nearly –4% in May 2020. Despite these downward revisions, the massively negative growth realizations for 2020Q2 were anticipated in none of the economies. Further, the massive growth rebound in 2020Q3 and the mostly moderately positive or negative growth rates in 2020Q4 were equally missed by the predictions. Notably, the predictions for 2020Q4 were on track for most economies until November 2020. Then, the release of the very high GDP growth figures for 2020Q3, which affected the GDP growth predictions for 2020Q4 via the autoregressive components of the models, resulted in much too strong upward revisions of the predictions and high forecast errors.

For 2021Q1, the GDP growth predictions for all economies except Germany moved in the direction of the respective realized GDP growth values. For the euro area, Austria, Belgium, Spain, France, Italy, and Poland, the predictions turned out to be quite exact. For the other economies, the prediction errors were at least not as high as in 2020. For 2021Q2 to 2021Q4, the predictions for most economies again failed to capture the positive growth realizations. However, there exist exceptions. For instance, the 2020Q4 predictions for Spain, the United Kingdom, and Italy are remarkably exact.

Overall, the models were unable to capture the large fluctuations in GDP growth in 2020. The results for 2021 are more mixed, with partly accurate and partly inaccurate predictions. The reasons why the forecast models failed will be discussed in the next section.

4.4 Forecast drivers

The previous section has shown that the GDP growth predictions for most quarters in 2020–2021 changed significantly over time. This is because new data were published gradually and affected the predictions. We want to determine how different

[13]We predict the GDP growth of the euro area directly, and therefore, we treat the euro area as a separate economy for this analysis. We currently work on a bottom-up construction of the euro area GDP growth prediction from the predictions of the individual economies.

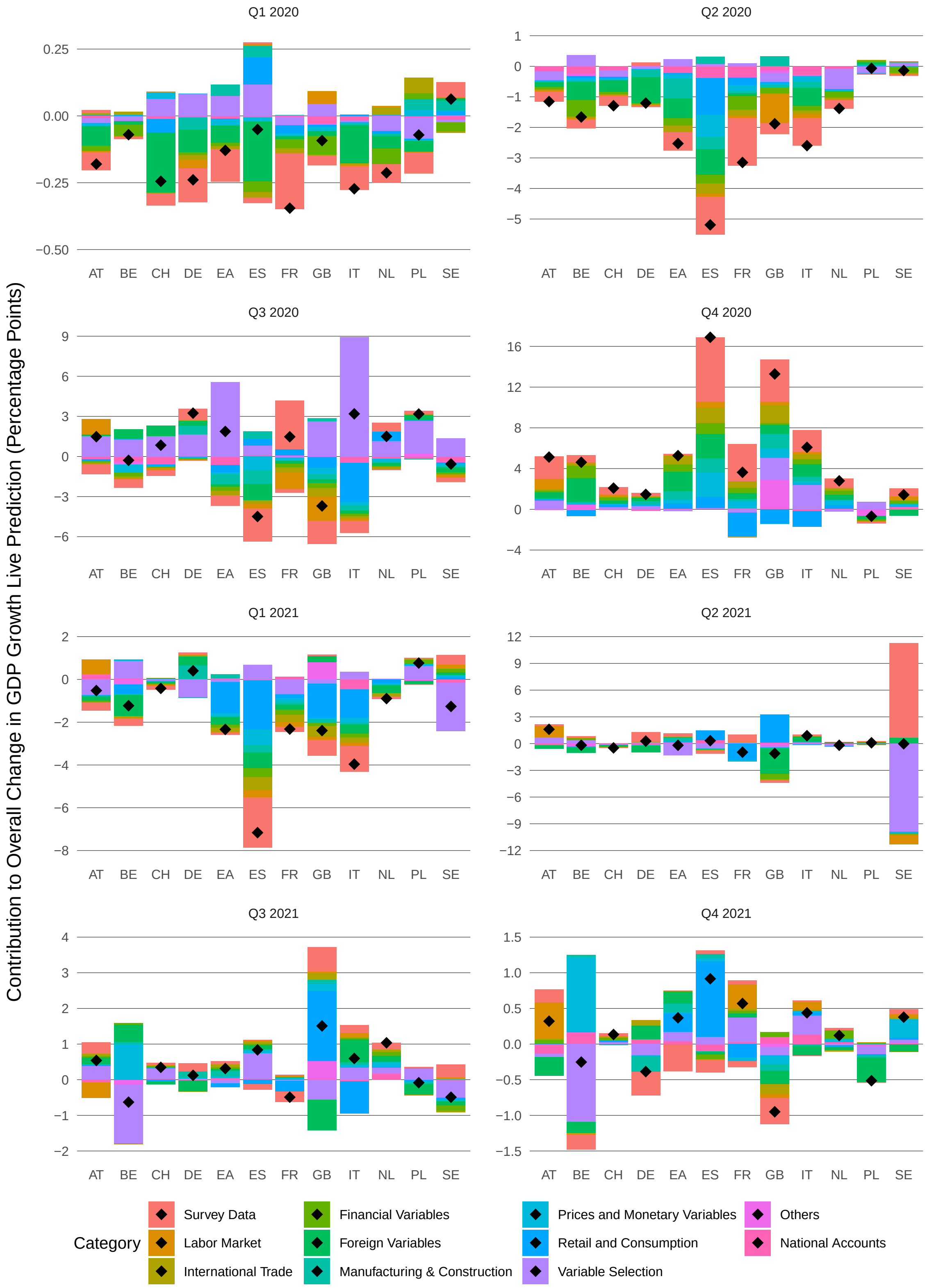

Figure 7: Impacts of variable categories on GDP growth predictions. The black diamonds in the figure show, separately for each economy and quarter in 2020–2021, the overall change in the pooled GDP growth prediction, i.e., the difference between the first pooled GDP growth forecast (2 quarters before GDP release) and the last pooled GDP growth nowcast (1 day before GDP release). The colored bars depict the *impacts* of different variable categories, i.e., their contributions to the aforementioned change. As compared to Table 1, business surveys and consumer surveys are combined into the category survey data, and manufacturing and construction are also combined into one category. See Figure 4 for the country codes.

variable categories contributed to the prediction changes. The black diamonds in Figure 7 show, separately for each economy and quarter in 2020–2021, the overall change in the prediction, i.e., the difference between the first GDP growth forecast (2 quarters before GDP release) and the last GDP growth nowcast (1 day before GDP release). The colored bars depict the *impacts* of different variable categories, i.e., their contributions to the aforementioned change.

The figure reveals that the relative importance of the individual variable categories varies strongly by quarter and economy. Three categories stand out: survey data (red bars), variable selection (purple bars), and retail and consumption (blue bars). The survey data category includes time series such as purchasing manager indices (PMIs) or consumer climate.[14] They traditionally play an important role in predicting GDP growth, as the surveyed assessments correlate with important subsectors and components of GDP and are usually published monthly with a short lag.

The variable selection category captures changes in the predictions that happen due to changes in the composition of the pre-selected variables: As part of the automated Nowcasting Lab process, the elastic net variable pre-selection algorithm is regularly re-run (see Section 3.1 and Appendix A). If, for example, a PMI variable drops out of the selected variable set and an interest rate variable enters instead, the resulting change in the GDP growth prediction is attributed to the variable selection category. In normal times, the impacts of the variable selection category are relatively small. Its important role during the COVID-19 crisis is due to the fact that the drivers and composition of GDP changed rapidly and abruptly several times (e.g., IMF, 2022). In line with these changes, the automated variable selection mechanism significantly adjusted the pool of selected variables.

The importance of the retail and consumption category during the COVID-19 crisis is also remarkable. During the past decades, the cyclical fluctuations in GDP were mostly driven by investment and, for small open economies, international trade. Accordingly, changes in GDP growth predictions over time were often due to news in predictor variables related to these two GDP components. Consumption, on the other hand, has a high share of GDP but usually fluctuates relatively little. Consequently, changes in consumption and retail variables typically do not play an important role either for changes in GDP growth predictions. During the COVID-19 crisis, however, this was completely different. As the virus spread, the

[14] Survey data are also referred to as soft data, as they reflect subjective assessments and intentions. Sales and turnover data, on the other hand, are referred to as hard data.

population reduced, voluntarily or forced by government restrictions, its mobility and consumption (see, e.g., Eckert et al., 2020 for the case of Switzerland). After the virus receded, mobility and consumption rebounded until the cycle began anew with a new wave of the virus.

4.5 Forecast comparison across models

The performance of the pooled quarterly GDP growth predictions was relatively poor during the year 2020 and mixed during 2021 (see Section 4.3). We want to know whether some forecasting models performed better than others. Tables 2 and 3 show the live out-of-sample root mean square forecast errors (RMSFEs) for all economies and models for 2020Q1–2020Q4 and 2021Q1–2021Q4, respectively. The figures in the last row ("Average") show, for each model, the unweighted average of the RMSFEs of all economies. The non-italic figures in the tables (i.e., always the first row in each economy section) display absolute RMSFEs, and the italic figures show relative RMSFEs to the random walk model, i.e., the RMSFE of a model divided by the RMSFE of the random walk model (abbreviated by RW in the tables). We chose the random walk model to calculate the relative RMSFEs since it is the most simple of all benchmark models. The absolute RMSFEs are calculated as follows: The Nowcasting Lab generates daily GDP growth predictions for each economy, each model, and each quarter t starting two quarters before the respective GDP release and ending on the last day before the release. We take the average over all daily GDP growth predictions for quarter t and subtract the actual GDP growth realization to get the average prediction error for quarter t. Then, we calculate, for each economy and each model, the RMSFE from the four average prediction errors of quarters $t = 2020\text{Q}1, \ldots, 2020\text{Q}4$ and $t = 2021\text{Q}1, \ldots, 2021\text{Q}4$, respectively.[15] We always use the first publication of realized GDP growth for the RMSFE calculation. Below we discuss results based on the latest available GDP release.

As can be seen in the last row ("Average") of Table 2, on average over all considered economies, the MF-DFM predictions (abbreviated DFM in the tables) performed best during 2020Q1–2020Q4 with absolute RMSFE of 5.5 percentage points. They outperformed the random walk predictions by about 36% (relative RMSFEs of 0.64). The bridge equation model predictions, the MIDAS model predictions, and the U-MIDAS model predictions (abbreviated BE, M, and UM respectively), out-

[15]For reference, Figure 9 in Appendix E shows the GDP growth prediction errors from the different forecasting models separately for each quarter from 2020Q1–2021Q4.

Table 2: RMSFE comparison across models for 2020Q1–Q4

Economy	Workhorse models					Benchmark models				
	DFM	M	UM	BE	Pooled	ARIMA	AR_{it}	AR_{di}	RM	RW
AT	6.00	7.74	8.00	8.10	7.46	8.37	8.70	8.14	7.04	8.67
	0.69	*0.89*	*0.92*	*0.93*	*0.86*	*0.96*	*1.00*	*0.94*	*0.81*	*1.00*
BE	5.86	7.80	6.73	7.58	6.99	11.19	8.57	9.48	9.12	8.50
	0.69	*0.92*	*0.79*	*0.89*	*0.82*	*1.32*	*1.01*	*1.11*	*1.07*	*1.00*
CH	3.84	5.34	4.67	5.21	4.72	6.81	6.61	6.08	6.29	5.98
	0.64	*0.89*	*0.78*	*0.87*	*0.79*	*1.14*	*1.11*	*1.02*	*1.05*	*1.00*
DE	3.82	5.41	5.32	5.50	4.96	6.31	6.04	6.37	7.12	6.70
	0.57	*0.81*	*0.79*	*0.82*	*0.74*	*0.94*	*0.90*	*0.95*	*1.06*	*1.00*
EA	5.24	8.47	7.68	8.62	7.50	12.48	10.22	10.69	9.29	9.05
	0.58	*0.94*	*0.85*	*0.95*	*0.83*	*1.38*	*1.13*	*1.18*	*1.03*	*1.00*
ES	8.53	12.03	12.05	12.08	11.03	19.99	18.67	18.61	13.86	12.85
	0.66	*0.94*	*0.94*	*0.94*	*0.86*	*1.56*	*1.45*	*1.45*	*1.08*	*1.00*
FR	7.02	10.92	10.15	10.88	9.69	13.32	16.72	14.43	12.22	12.17
	0.58	*0.90*	*0.83*	*0.89*	*0.80*	*1.10*	*1.37*	*1.19*	*1.00*	*1.00*
GB	8.65	10.88	10.56	11.01	9.97	19.67	14.92	17.14	12.97	12.35
	0.70	*0.88*	*0.86*	*0.89*	*0.81*	*1.59*	*1.21*	*1.39*	*1.05*	*1.00*
IT	7.36	10.19	8.76	10.00	9.08	12.43	10.71	11.44	10.43	10.78
	0.68	*0.95*	*0.81*	*0.93*	*0.84*	*1.15*	*0.99*	*1.06*	*0.97*	*1.00*
NL	3.24	4.97	4.85	5.26	4.58	6.43	5.43	5.98	6.05	5.71
	0.57	*0.87*	*0.85*	*0.92*	*0.80*	*1.13*	*0.95*	*1.05*	*1.06*	*1.00*
PL	3.20	4.48	5.19	4.57	4.36	4.67	4.64	4.59	5.64	5.55
	0.58	*0.81*	*0.94*	*0.82*	*0.79*	*0.84*	*0.84*	*0.83*	*1.02*	*1.00*
SE	3.45	3.81	3.72	3.72	3.53	4.49	4.41	4.66	4.39	4.72
	0.73	*0.81*	*0.79*	*0.79*	*0.75*	*0.95*	*0.93*	*0.99*	*0.93*	*1.00*
Average	5.50	7.70	7.30	7.70	7.00	10.50	9.60	9.80	8.70	8.60
	0.64	*0.90*	*0.85*	*0.90*	*0.81*	*1.22*	*1.12*	*1.14*	*1.01*	*1.00*

Notes: The table shows RMSFEs for 2020Q1–2020Q4 by economy and model. The figures in "Average" are the unweighted averages of the RMSFEs of all economies. The non-italic figures, i.e., the figures in the first row of each economy section, are absolute RMSFEs. The italic figures are relative RMSFEs to the random walk model (RW), i.e., the RMSFE of a model divided by the RMSFE of the random walk model. The absolute RMSFEs are calculated from the average of the daily predictions for each target period $t \in \{2020Q1, \ldots, 2020Q4\}$ minus the actual GDP growth realization (first release) for t. These four prediction errors are then used to calculate the RMSFE in 2020. See Figure 4 for the country codes. Model abbreviations: DFM: mixed-frequency dynamic factor model, M: MIDAS model, UM: U-MIDAS model, BE: bridge equation model, Pooled: unweighted average of all workhorse model predictions, ARIMA: ARIMA model, AR_{it}: Iterative AR model, AR_{di}: direct AR model, RM: in-sample rolling mean, RW: random walk model.

performed the random walk model by about 10 to 15% on average over all economies (relative RMSFEs of 0.90, 0.85 and 0.90, respectively). The pooled model predictions, which result from equally weighting the predictions of all workhorse models as explained in Section 3.2, outperformed the random walk predictions by around 20%. The predictions from the ARIMA model, the iterative AR model (abbreviated by AR_{it}), the direct AR (abbreviated by AR_{di}), and the rolling in-sample mean (abbreviated by RM) were worse than the random walk predictions. Notably, the MF-DFM delivered the lowest RMSFE for all individual economies, including the euro area. The U-MIDAS model ranked second for most economies. The better forecast performance of the MF-DFM relative to the other models for the 2020Q1-2020Q4 period is noteworthy. One reason for this outcome is the different specification of the elastic net variable pre-selection (see Appendix A). For the MF-DFM, the selection is limited to a maximum of 20 variables since, according to our prior experience, the one-factor version of the MF-DFM that we mainly use in the Nowcasting Lab performs particularly well for a medium-sized set of variables. In contrast, we impose no a priori restriction on the number of variables for the MIDAS, the U-MIDAS, and the bridge equation model. Accordingly, the elastic net algorithm usually selects between 70 and 90 variables for these models. As the MF-DFM relies on relatively few predictive variables, it is more aggressive in indicating turning points, but also more volatile and, hence, more prone to false signals. In contrast, the other models deliver more balanced predictions, but are, on the other hand, also more sluggish when it comes to capturing cyclical swings.

For the 2021Q1–2021Q4 period, the RMSFEs were for all considered economies much smaller than for the 2020 period (see Table 3). On average over all economies, the MF-DFM delivered the best predictions in terms of RMSFE (relative RMSFE of 0.46), closely followed by the rolling in-sample mean and the U-MIDAS model (relative RMSFEs of 0.47 and 0.48, respectively). However, unlike during the year 2020, the MF-DFM is the best-performing model in only some economies. It has the lowest prediction errors among all models for Austria (AT), the euro area (EA), Spain (ES), the United Kingdom (GB), Italy (IT), and the Netherlands (NL). In contrast, the MIDAS model performs best for Switzerland (CH), and Poland (PL), the U-MIDAS model for Germany (DE), the bridge equation model for Sweden (SE), and the rolling in-sample mean for Belgium (BE) and France (FR). Appendix D shows that the results of Tables 2 and 3 are robust to calculating the RMSFEs based on the latest available GDP release for each quarter instead of the first GDP release. However, there is a notable exception: the U-MIDAS model and also the pooled quarterly GDP growth predictions now outperform the MF-DFM in terms of RMSFE on average over all economies.

Table 3: RMSFE comparison across models for 2021Q1–Q4

Economy	Workhorse models					Benchmark models				
	DFM	M	UM	BE	Pooled	ARIMA	AR_{it}	AR_{di}	RM	RW
AT	1.32	2.58	2.01	2.48	1.93	1.83	2.15	2.34	2.24	4.32
	0.31	*0.60*	*0.46*	*0.57*	*0.45*	*0.42*	*0.50*	*0.54*	*0.52*	*1.00*
BE	2.00	1.11	1.25	1.21	1.26	1.26	2.45	2.14	0.95	2.13
	0.94	*0.52*	*0.59*	*0.57*	*0.59*	*0.59*	*1.15*	*1.00*	*0.45*	*1.00*
CH	1.13	0.86	0.90	1.13	0.92	0.88	0.85	0.66	0.87	2.09
	0.54	*0.41*	*0.43*	*0.54*	*0.44*	*0.42*	*0.41*	*0.32*	*0.42*	*1.00*
DE	1.12	1.61	1.01	1.48	1.30	1.33	1.26	1.34	1.44	3.13
	0.36	*0.52*	*0.32*	*0.47*	*0.42*	*0.43*	*0.40*	*0.43*	*0.46*	*1.00*
EA	1.09	1.79	1.17	1.81	1.19	1.21	1.22	1.81	1.18	3.17
	0.34	*0.56*	*0.37*	*0.57*	*0.38*	*0.38*	*0.39*	*0.57*	*0.37*	*1.00*
ES	0.89	2.53	2.24	2.62	2.07	5.41	5.53	22.77	1.74	3.30
	0.27	*0.77*	*0.68*	*0.80*	*0.63*	*1.64*	*1.68*	*6.91*	*0.53*	*1.00*
FR	1.98	1.87	1.29	1.56	1.36	8.47	2.41	17.24	1.04	3.38
	0.59	*0.55*	*0.38*	*0.46*	*0.40*	*2.50*	*0.71*	*5.10*	*0.31*	*1.00*
GB	1.26	2.83	1.53	2.78	2.01	2.24	4.07	2.55	1.95	4.52
	0.28	*0.63*	*0.34*	*0.62*	*0.45*	*0.50*	*0.90*	*0.57*	*0.43*	*1.00*
IT	1.54	2.00	1.73	1.96	1.48	1.74	2.92	4.18	1.61	3.73
	0.41	*0.54*	*0.46*	*0.53*	*0.40*	*0.47*	*0.78*	*1.12*	*0.43*	*1.00*
NL	1.33	1.52	1.62	1.38	1.42	1.59	1.48	1.52	1.46	2.55
	0.52	*0.59*	*0.63*	*0.54*	*0.56*	*0.62*	*0.58*	*0.60*	*0.57*	*1.00*
PL	0.81	0.67	1.02	0.82	0.80	0.87	1.19	0.86	0.80	1.34
	0.60	*0.50*	*0.76*	*0.61*	*0.59*	*0.65*	*0.88*	*0.64*	*0.60*	*1.00*
SE	1.31	1.22	0.91	0.77	0.93	0.85	0.88	0.84	0.82	0.76
	1.73	*1.61*	*1.20*	*1.01*	*1.22*	*1.12*	*1.16*	*1.11*	*1.08*	*1.00*
Average	1.31	1.72	1.39	1.67	1.39	2.31	2.20	4.86	1.34	2.87
	0.46	*0.60*	*0.48*	*0.58*	*0.48*	*0.80*	*0.77*	*1.69*	*0.47*	*1.00*

Notes: See Table 2, but for 2021Q1–2021Q4.

The RMSFEs in the previous tables were calculated from the quarterly averages of the daily GDP growth predictions. Instead, Table 4 presents RMSFEs calculated from the prediction errors at specific forecast horizons (2-quarter ahead, 1-quarter ahead, and 1-day ahead of GDP release). The table displays the average RMSFEs across all economies. Tables 8 to 13 in Appendix E show the results for the individual economies. For the 2020Q1–2020Q4 period, the U-MIDAS model generated slightly better forecasts than the MF-DFM at the 2-quarter ahead horizon (RMSFE of 7.7 versus RMSFE of 7.8). However, the forecast quality of the MF-DFM improved sig-

Table 4: RMSFEs (economy averages) for different forecast horizons

Horizon	Workhorse models					Benchmark models				
	DFM	M	UM	BE	Pooled	ARIMA	AR_{it}	AR_{di}	RM	RW
2020Q1–2020Q4										
2-quarter ahead	7.80	8.00	7.70	8.00	7.70	11.60	8.40	8.60	9.70	10.70
	1.42	*1.04*	*1.05*	*1.04*	*1.10*	*1.10*	*0.88*	*0.88*	*1.11*	*1.24*
1-quarter ahead	4.80	9.10	7.80	9.10	7.50	10.80	12.20	12.20	8.60	11.90
	0.87	*1.18*	*1.07*	*1.18*	*1.07*	*1.03*	*1.27*	*1.24*	*0.99*	*1.38*
1-day ahead	4.40	9.00	7.60	9.10	7.30	10.50	12.30	12.30	8.50	11.90
	0.80	*1.17*	*1.04*	*1.18*	*1.04*	*1.00*	*1.28*	*1.26*	*0.98*	*1.38*
2021Q1–2021Q4										
2-quarter ahead	1.30	2.30	1.60	2.20	1.70	2.70	3.20	6.70	1.40	4.60
	0.99	*1.34*	*1.15*	*1.32*	*1.22*	*1.17*	*1.45*	*1.38*	*1.04*	*1.60*
1-quarter ahead	1.60	1.30	1.40	1.30	1.20	2.20	3.30	3.30	1.30	1.60
	1.22	*0.76*	*1.01*	*0.78*	*0.86*	*0.95*	*1.50*	*0.68*	*0.97*	*0.56*
1-day ahead	1.80	1.30	1.40	1.30	1.30	2.20	3.30	3.30	1.30	1.60
	1.37	*0.76*	*1.01*	*0.78*	*0.94*	*0.95*	*1.50*	*0.68*	*0.97*	*0.56*

Notes: The non-italic figures in the table show, for 2020Q1–2020Q4 and for 2021Q1–2021Q4, absolute RMSFEs from different models using for each quarter the GDP growth prediction at the first day of the prediction period (2-quarter ahead of GDP release), in the middle of the prediction period (1-quarter ahead of GDP release), and at the last day of the prediction period (1-day ahead of GDP release). The displayed figures are unweighted averages of the RMSFEs of all economies (see Appendix E for the economy-specific RMSFEs). The italic figures are the aforementioned absolute RMSFEs divided by the absolute RMSFEs calculated from the average of all daily predictions starting 2 quarters before the GDP release and ending 1 day before the release. Model abbreviations: DFM: mixed-frequency dynamic factor model, M: MIDAS model, UM: U-MIDAS model, BE: bridge equation model, Pooled: unweighted average of all workhorse model predictions, ARIMA: ARIMA model, AR_{it}: Iterative AR model, AR_{di}: direct AR model, RM: in-sample rolling mean, RW: random walk model.

nificantly thereafter (RMSFEs of 4.8 and 4.4 at the 1-quarter and 1-day ahead horizons, respectively). In contrast, no improvement was observed for the U-MIDAS model (RMSFEs of 7.8 and 7.6 at the 1-quarter and 1-day ahead horizons, respectively). One reason is that the AR component of the model provided false signals for the latter two horizons. The same applies to the two AR models and the random walk prediction, whose RMSFE was worse at the 1-quarter and 1-day ahead horizons than at the 2-quarter ahead horizon. The MIDAS model and the bridge equation model also suffered from the false signals of the AR component. The period 2021Q1–2021Q4 provided a different pattern. Here, the MF-DFM performed better than the U-MIDAS model at the 2-quarter horizon (RMSFE of 1.3 versus RMSFE of 1.6). Thereafter, however, the predictive accuracy of the U-MIDAS model – as well as the MIDAS and bridge equation models – improved substantially. In contrast, the predictive accuracy of the MF-DFM declined and fell behind the accuracy of the other workhorse models. This is where the risk of aggressive specifications becomes apparent: As described above, the MF-DFM contains 20 variables that are mostly very forward looking. These variables indicated for many economies sharp declines and rebounds for some quarters in 2021, just as they did for 2020. However, GDP growth during 2021 was much less volatile than in 2020 (see Figure 4). In contrast, the other workhorse models are based on 70–90 variables and provided more balanced GDP growth predictions. The variable pre-selection algorithm also played a role: it selected variables based on the development in 2020, which were then less informative in 2021. With a set of only 20 variables, this selection played a greater role than with 70–90 variables.

4.6 Forecast performance and choice of estimation sample

As the COVID-19 crisis gathered momentum in early 2020, we conjectured that the correlations between GDP and many higher-frequency variables would soon be subject to instabilities and structural breaks. Therefore, we decided to freeze the estimation sample for the MIDAS, the U-MIDAS, and the bridge equation model on the day of the GDP release for 2019Q4. Sections 4.3 to 4.5 reported the forecast results for this frozen estimation window specification. In addition, we continued to run the models based on an expanding window estimation as described in Section 3.2. The forecast results are shown in Table 5. For 2020Q1–2020Q4, the RMSFEs of the expanding estimation window specification turned out to be higher for all economies except Switzerland (CH) and Poland (PL) than the RMSFEs of the frozen estimation window specification. For 2021Q1–2021Q4, both specifications performed similarly in terms of RMSFE. However, the expanding estimation window specification produced more forecast outliers. The results are robust to calculating the forecast error of each quarter using the latest available GDP release (as of 05/12/2022) instead of using the first GDP release (see Table 14 in Appendix E). Our findings are in line with Lenza and Primiceri (2020), Carriero et al. (2021), and Schorfheide and Song (2021) who find that treating the COVID-19 period as missing observations when estimating a BVAR leads to better forecast results.

Table 5: **RMSFEs with alternative estimation**

	2020Q1–2020Q4			2021Q1–2021Q4		
Economy	M	UM	BE	M	UM	BE
AT	7.26	7.19	9.01	2.05	2.19	2.71
	0.94	*0.90*	*1.11*	*0.79*	*1.09*	*1.09*
BE	9.73	9.44	12.14	2.61	2.32	1.02
	1.25	*1.40*	*1.60*	*2.36*	*1.86*	*0.84*
CH	5.82	5.61	7.02	0.54	0.63	0.93
	1.09	*1.20*	*1.35*	*0.63*	*0.70*	*0.82*
DE	6.09	6.11	6.22	1.18	1.24	1.38
	1.13	*1.15*	*1.13*	*0.73*	*1.23*	*0.93*
EA	9.52	9.33	12.30	2.10	1.56	1.36
	1.12	*1.21*	*1.43*	*1.17*	*1.33*	*0.75*
ES	14.02	13.48	20.40	11.96	10.36	19.23
	1.17	*1.12*	*1.69*	*4.72*	*4.62*	*7.33*
FR	11.96	11.78	13.67	7.86	7.04	14.79
	1.10	*1.16*	*1.26*	*4.21*	*5.45*	*9.50*
GB	14.75	14.66	23.39	2.33	2.12	1.97
	1.36	*1.39*	*2.12*	*0.82*	*1.39*	*0.71*
IT	10.02	10.02	12.96	3.83	3.08	2.51
	0.98	*1.14*	*1.30*	*1.92*	*1.78*	*1.28*
NL	5.65	5.70	6.16	1.42	1.48	1.59
	1.14	*1.17*	*1.17*	*0.94*	*0.92*	*1.15*
PL	4.70	4.54	4.56	0.90	0.84	1.04
	1.05	*0.87*	*1.00*	*1.35*	*0.82*	*1.26*
SE	4.08	4.09	4.05	0.51	0.53	0.70
	1.07	*1.10*	*1.09*	*0.42*	*0.58*	*0.91*
Average	8.60	8.50	11.00	3.10	2.80	4.10
	1.12	*1.16*	*1.43*	*1.80*	*2.01*	*2.46*

Notes: See Tables 2 and 3, but the full sample including the period after 2019Q4 is used for the parameter estimation of the models (expanding estimation window as described in Section 3.2). The italic figures are relative RMSFEs to the corresponding model specifications with parameter estimation excluding the period after 2019Q4, i.e., RMSFE of the model specifications from this table divided by RMSFE of the corresponding model specifications from Tables 2 and 3.

5 Conclusion

This paper presented a code-database-website environment called Nowcasting Lab. It produces and publishes model-based now-/forecasts for quarterly GDP growth of currently 16 economies including the United States and the euro area. The predictions are automatically updated daily and released immediately afterwards on the website `https://nowcastinglab.org` together with accompanying statistics and information. The input data, i.e, currently 3640 times series, the predictions, and other model outputs are stored in a daily vintage database that we make available to researchers for real-time analyses.

The Nowcasting Lab has two main purposes. First, forecasting researchers and practitioners in, e.g., central banks, ministries, international organizations, research institutes, and the financial sector often have a great deal of knowledge about forecasting models and data. However, they often lack the time to maintain several models and rich data sets for different economies and to implement new models from academic research. Against this background, practitioners can consult the Nowcasting Lab website to support their own forecasting tasks. Additional economies and predictor variables can be implemented on request. Also, the environment can be extended to predicting additional variables, e.g., inflation.

Second, the Nowcasting Lab can be used to test the live out-of-sample predictive accuracy of macroeconomic nowcasting and short-run forecasting models. We invite forecasting researchers to implement new models and methods in the lab. The model outcomes can be published on the website to increase their visibility in the community. Alternatively, the models can be run in the background, i.e., their results and performance are only accessible to certain people. The use of Git repositories makes it easy to work on the lab code remotely. Since the Nowcasting Lab code is modular, models in different software can be added to the lab. The researchers can thus work in their preferred programming languages.

As an empirical use case, we have analyzed the live out-of-sample nowcasts and one-quarter ahead forecasts for quarterly GDP growth of several economies generated by the Nowcasting Lab during the COVID-19 crisis, i.e., the period 2020Q1–2021Q4. At that time, the euro area, Germany, the United Kingdom, France, Italy, Spain, the Netherlands, Switzerland, Poland, Sweden, Belgium, and Austria were included. The model set comprised models that are often used in, e.g., central banks for nowcasting and short-term forecasting: a mixed-frequency DFM (Bok et al., 2018), a MIDAS model (Ghysels et al., 2007), an unrestricted mixed-frequency model (Foroni et al., 2015), a bridge equation model (Schumacher, 2016) as well as several simple univariate models. Further, an elastic net variable pre-selection mechanism following Zou and Hastie (2005) was employed to shrink the set of relevant predictor series for each individual model.

We found that none of the considered models delivered accurate live out-of-sample forecasts for quarterly GDP growth in the year 2020. The MF-DFM performed best in terms of RMSFE for all economies and all considered forecasting models. The U-MIDAS model ranked second. The sharp ups and downs in GDP in 2020 caused the AR term in the U-MIDAS, MIDAS, and bridge equation models, the inclusion of which typically improves the model's predictive accuracy, to bias the quarterly GDP growth predictions. For the four quarters in 2021, the forecast accuracy of most models improved strongly compared to the quarters in 2020. On average over all economies, the MF-DFM and the U-MIDAS model performed similarly, with a slight advantage for the former (the latter) when the RMSFEs are calculated based on the first (latest) GDP releases. However, the performance of the different forecasting models varied greatly depending on the individual economy. Thus, an analysis based on only one or a few economies might have led to a misleading assessment of the predictive accuracy of the models. Further, our analysis revealed the advantages and disadvantages of different model specifications for GDP nowcasting: The MF-DFM is based on relatively few, often forward-looking variables. In contrast, the U-MIDAS, MIDAS and bridge equation models are based on a relatively broad set of variables. As a consequence, the MF-DFM nowcasted the strong ups and downs of GDP in 2020 better than the other models, but indicated too strong GDP fluctuations in 2021. In contrast, the other models responded too sluggishly in 2020, but provided relatively balanced GDP growth nowcasts for 2021 and outperformed the MF-DFM.

In addition, we have examined how different variable categories contributed to the GDP predictions for the 2020Q1–2021Q4 period changing over the forecast horizon. The relative importance of the variable categories varied greatly from quarter to quarter and economy to economy, but three categories played an important role overall: survey data, variable selection, and retail and consumption. The variable selection category records changes in the predictions that are due to changes in the composition of the pre-selected variables. Its unusually important role during the COVID-19 crisis comes from the fact that the drivers and composition of GDP changed rapidly and abruptly several times. The importance of the consumption and retail category is also in marked contrast to normal times. Consumption usually fluctuates little relative to other GDP components, so consumption and retail indicators are usually less important than variables related to, e.g., investment. However, the COVID-19 crisis was different.

Further, our analysis yielded that the GDP growth predictions for the 2020Q1–2021Q4 period improve when the the crisis period is excluded from the estimation sample of the models. The finding suggests that future research might consider treating the COVID-19 period special in model estimation, either through flexible parameterization or simply through exclusion or outlier correction.

Last, using the real-time data collected in our database, we have analyzed the revisions to quarterly GDP for the period 2020Q1–2021Q4. For most economies, the GDP revisions were very large, especially for 2020. This indicates that the first official GDP estimates were not very accurate in many economies. The fact that most national statistical agencies initially substantially overestimated the GDP slump in the first half of 2020 and underestimated the subsequent recovery is problematic from a policy perspective. This is because the 2020 stimulus programs, which tore deep holes in government budgets and contributed to the current rise in core inflation, would otherwise probably have been less massive in some economies.

References

Adam, T., Michálek, O., Michl, A., and Slezáková, E. (2021). The Rushin index: A weekly indicator of Czech economic activity. *CNB Working Paper*, 2021(4). Czech National Bank (CNB).

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, 33(1):178–196.

Anderson, R. G. (2006). Replicability, real-time data, and the science of economic research: Fred, Alfred, and VDC. *Review Federal Reserve Bank of Saint Louis*, 88(1):81.

Andreou, E., Ghysels, E., and Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2):246–261.

Andreou, E., Ghysels, E., and Kourtellos, A. (2011). Forecasting with mixed-frequency data. In Clements, M. P. and Hendry, D. F., editors, *Oxford Handbook on Economic Forecasting*, pages 225–245. Oxford University Press: Oxford.

Andreou, E., Ghysels, E., and Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, 31(2):240–251.

Angelini, E., Camba-Méndez, G., Giannone, D., Rünstler, G., and Reichlin, L. (2011). Shortterm forecasts of euro area GDP growth. *The Econometrics Journal*, 14(1):C25–C44.

Anttonen, J. (2018). Nowcasting the unemployment rate in the EU with seasonal BVAR and Google search data. *ETLA Working Papers*, 62. Research Institute of the Finnish Economy (ETLA).

Babii, A., Ghysels, E., and Striaukas, J. (2021). Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, pages 1–23.

Baffigi, A., Golinelli, R., and Parigi, G. (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting*, 20(3):447–460.

Bańbura, M., Giannone, D., and Reichlin, L. (2011). Nowcasting. In Clements, M. P. and Hendry, D. F., editors, *The Oxford Handbook on Economic Forecasting*, chapter 2, pages 193–224. Oxford University Press, Oxford.

Bańbura, M. and Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160.

Bergmeir, C., Hyndman, R. J., and Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120:70–83.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., and Tambalotti, A. (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics*, 10:615–643.

Bulligan, G., Golinelli, R., and Parigi, G. (2010). Forecasting monthly industrial production in real-time: From single equations to factor-based models. *Empirical Economics*, 39(2):303–336.

Bulligan, G., Marcellino, M., and Venditti, F. (2015). Forecasting economic activity with targeted predictors. *International Journal of Forecasting*, 31(1):188–206.

Camacho, M., Pérez-Quirós, G., and Poncela, P. (2013). Short-term forecasting for empirical economists: A survey of the recently proposed algorithms. *Foundations and Trends in Econometrics*, 6(2).

Carriero, A., Clark, T. E., Marcellino, M. G., and Mertens, E. (2021). Addressing COVID-19 outliers in BVARs with stochastic volatility. *Federal Reserve Bank of Cleveland Working Paper*, 21-02. Federal Reserve Bank of Cleveland.

Cascaldi-Garcia, D., RT Ferreira, T., Giannone, D., and Modugno, M. (2021). Back to the present: Learning about the euro area through a now-casting model. *International Finance Discussion Paper*, (1313). Board of Governors of the Federal Reserve System.

Celasun, O., Hansen, N.-J., Mineshima, A., Spector, M., and Zhou, J. (2022). Supply bottlenecks: Where, why, how much, and what next? *IMF Working Paper*, 2022/031. International Monetary Fund.

Chow, G. C. and Lin, A.-l. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53(4):372–375.

Clements, M. P. and Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business & Economic Statistics*, 26(4):546–554.

Clements, M. P. and Galvão, A. B. (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, 24(7):1187–1206.

Croushore, D. and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, 105(1):111–130.

Dey-Chowdhury, S., Drew, S., Kent-Smith, R., Lee, P., Mahajan, S., Mills, K., Scruton, J., Wales, P., and Walton, A. (2020a). Coronavirus and the effects on UK GDP. *Office for National Statistics Online Publications*. `https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/articles/coronavirusandtheeffectsonukgdp/2020-05-06#practical-challenges`.

Dey-Chowdhury, S., Khaliq, M., and Martin, J. (2020b). Coronavirus and the changing need for economic statistics and analysis. *Office for National Statistics Online Publications*. `https://www.ons.gov.uk/economy/economicoutputandproductivity/output/articles/coronavirusandthechangingneedforeconomicstatisticsandanalysis/2020-07-08`.

Dey-Chowdhury, S., McAuley, N., and Walton, A. (2021). International comparisons of GDP during the coronavirus (COVID-19) pandemic. *Office for National Statistics Online Publications*. `https://www.ons.gov.uk/economy/grossdomesticproductgdp/articles/internationalcomparisonsofgdpduringthecoronaviruscovid19pandemic/2021-02-01`.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431.

Diron, M. (2008). Short-term forecasts of euro area real GDP growth: An assessment of real-time performance based on vintage data. *Journal of Forecasting*, 27:371–390.

Doz, C., Giannone, D., and Reichlin, L. (2012). A quasi–maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics*, 94(4):1014–1024.

Eckert, F., Kronenberg, P., Mikosch, H., and Neuwirth, S. (2020). Tracking economic activity with alternative high-frequency data. *KOF Working Papers*, 488:1–35. KOF Swiss Economic Institute, ETH Zurich.

Engle, R. and Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, 76(376):774–781.

Eraslan, S. and Götz, T. (2021). An unconventional weekly economic activity index for Germany. *Economics Letters*, 204:109881.

Faust, J., Rogers, J. H., and Wright, J. H. (2005). News and noise in G-7 GDP announcements. *Journal of Money, Credit and Banking*, 37(3):403–419.

Fenz, G. and Stix, H. (2021). Monitoring the economy in real time with the weekly OeNB GDP indicator: Background, experience and outlook. *Monetary Policy & the Economy*, Q4/20-Q1/21:17–40. Oesterreichische Nationalbank (Austrian Central Bank).

Ferrara, L., Marsilli, C., and Ortega, J.-P. (2014). Forecasting growth during the Great Recession: Is financial volatility the missing ingredient? *Economic Modelling*, 36:44–50.

Foroni, C. and Marcellino, M. (2013). A survey of econometric methods for mixed-frequency data. *Norges Bank Research Working Paper*, 2013-06. Norges Bank, Norway.

Foroni, C. and Marcellino, M. (2014). A comparison of mixed frequency approaches for nowcasting euro area macroeconomic aggregates. *International Journal of Forecasting*, 30(3):554–568.

Foroni, C., Marcellino, M., and Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A*, 178(1):57–82.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In Aigner, D. J. and Goldberger, A. S., editors, *Latent variables in socio-economic models*, chapter 19. Amsterdam, North-Holland.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Golinelli, R. and Parigi, G. (2007). The use of monthly indicators to forecast quarterly GDP in the short run: An application to the G7 countries. *Journal of Forecasting*, 26:77–94.

Heinisch, K. and Scheufele, R. (2018). Bottom-up or direct? Forecasting German GDP in a data-rich environment. *Empirical Economics*, 54(2):705–745.

Higgins, P. C. (2014). GDPNow: A model for GDP "nowcasting". *FRB Atlanta Working Paper*, 2014(7). Federal Reserve Bank of Atlanta.

Hyndman, R. J. (2020). A brief history of forecasting competitions. *International Journal of Forecasting*, 36(1):7–14.

IMF (2022). Rising caseloads, a disrupted recovery, and higher inflation. *World Economic Outlook*, January 2022. International Monetary Fund.

Ingenito, R. and Trehan, B. (1996). Using monthly data to predict quarterly output. *Federal Reserve Bank of San Francisco Economic Review*, 3:3–11.

Jordà, O., Kouchekinia, N., Merrill, C., and Sekhposyan, T. (2020). The fog of numbers. *FRBSF Economic Letter*, 2020-20. Federal Reserve Bank of San Francisco.

Kuzin, V., Marcellino, M., and Schumacher, C. (2013). Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries. *Journal of Applied Econometrics*, 28(3):392–411.

Lenza, M. and Primiceri, G. (2020). How to estimate a VAR after March 2020. *NBER Working Paper*, 27771. National Bureau of Economic Research.

Lewis, D. J., Mertens, K., Stock, J. H., and Trivedi, M. (2021). Measuring real activity using a weekly economic index. *Journal of Applied Econometrics*, 37(4):667–687.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

Lourenço, N. and Rua, A. (2020). The DEI: Tracking economic activity daily during the lockdown. *Banco de Portugal Working Papers*, 13. Banco de Portugal.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4):1346–1364.

Marcellino, M. and Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, 72(4):518–550.

Mariano, R. S. and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18(4):427–443.

Masini, R. P., Medeiros, M. C., and Mendes, E. F. (2020). Machine learning advances for time series forecasting. *arXiv preprint arXiv:2012.12802*.

Mikosch, H. and Solanko, L. (2019). Forecasting quarterly Russian GDP growth with mixed-frequency data. *Russian Journal of Money and Finance*, 78(1):19–35.

Mikosch, H. and Zhang, Y. (2014). Forecasting Chinese GDP growth with mixed frequency data: Which indicators to look at? *KOF Working Paper*, 359. KOF Swiss Economic Institute, ETH Zurich.

Mogliani, M. and Simoni, A. (2021). Bayesian MIDAS penalized regressions: Estimation, selection, and prediction. *Journal of Econometrics*, 222(1):833–860.

Muggenthaler, P., Schroth, J., and Sun, Y. (2021). The heterogeneous economic impact of the pandemic across euro area countries. *ECB Economic Bulletin*, 2021(5). European Central Bank.

Phillips, P. C. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.

Rünstler, G., Barhoumi, K., Benk, S., Cristadoro, R., Den Reijer, A., Jakaitiene, A., Jelonek, P., Rua, A., Ruth, K., and Van Nieuwenhuyze, C. (2009). Short-term forecasting of GDP using large datasets: A pseudo real-time forecast evaluation exercise. *Journal of Forecasting*, 28(7):595–611.

Sargent, T. J. and Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New methods in Business Cycle Research*, 1:145–168.

Schorfheide, F. and Song, D. (2021). Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic. *NBER Working Paper*, 29535. National Bureau of Economic Research.

Schumacher, C. (2016). A comparison of MIDAS and Bridge equations. *International Journal of Forecasting*, 32(2):257–270.

Siliverstovs, B. (2017). Short-term forecasting with mixed-frequency data: A MIDASSO approach. *Applied Economics*, 49(13):1326–1343.

Stock, J. H. and Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, 4:351–394.

Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):44–47.

Timmermann, A. (2006). Forecast combinations. In Elliott, G., Granger, C. W. J., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, pages 135–196. Elsevier, North Holland.

Wegmüller, P., Glocker, C., and Guggia, V. (2021). Weekly economic activity: Measurement and informational content. *International Journal of Forecasting*, forthcoming.

Woloszko, N. (2020). Tracking activity in real time with Google trends. *OECD Economics Department Working Papers*, 1634. Organisation for Economic Co-operation and Development (OECD).

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 167(2):301–320.

The Nowcasting Lab: Live Out-of-Sample Forecasting and Model Testing – Online Appendix –

P. Kronenberg, H. Mikosch, S. Neuwirth, M. Bannert, and S. Thöni

A Elastic net variable selection

For the variable selections, we use the elastic net algorithm following Zou and Hastie (2005). The elastic net arose from criticism of LASSO (Least Absolute Shrinkage and Selection), whose variable selection may depend too much on data and can therefore be unstable. The solution is to combine the L2 penalty of ridge regressions and the L1 penalty of LASSO to get the best of both worlds. Elastic net aims at minimizing the following loss function for regression with m predictors:

$$L_{enet}(\hat{\beta}) = \frac{\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2}{2n} + \lambda(\frac{1-\alpha}{2}\sum_{j=1}^{m}\hat{\beta}_j^2 + \alpha\sum_{j=1}^{m}|\hat{\beta}|_j)$$

where α is the mixing parameter between ridge ($\alpha = 0$) and LASSO ($\alpha = 1$). In our application, we restrict α to 0.5 to get a combination of both approaches. The λ is determined by optimizing the model.

Usually, the parameters for the elastic net algorithm are optimized using cross-validation. As Bergmeir et al. (2018) explain, cross-validation (CV) is not straightforward for time-series data due to their inherent serial correlation and potential non-stationarity. For the Nowcasting Lab, we follow the approach of Bergmeir et al. (2018), i.e., we check for serial correlation using the Ljung-Box test by Ljung and Box (1978) and include the appropriate number of lags of the target variable in the variable selection matrix $\mathbf{X}$. The errors of each selected combination are then retested for serial correlation, and only those that pass the test are used in the subsequent optimization.

As the criteria for optimizing the elastic net algorithm, we use the Akaike Information Criteria (AIC). As shown by Stone (1977), minimizing the AIC approximates the leave-one-out CV method. This approach is computationally less demanding and ensures the stability of the selected variables over time. Depending on the model type, a different variable selection is run.

For the bridge equation models, which are estimated at a quarterly frequency, the variable selection is also conducted at a quarterly frequency. All variables are aggregated to the quarterly frequency and transformations calculated subsequently (see Section 3.1). The optimal variables or variable transformations according to the elastic net selection are then used for the model combination.

For the MF-DFM, the variable selection is also run at a quarterly frequency. According to our prior experience, the one-factor version of the Bok et al. (2018) MF-DFM that we use performs particularly well for a medium-sized set of variables, whereas too many variables can lead to unstable results and long computation times. Thus, we decided to limit the maximum number of variables for the one-factor MF-DFM version, which we show on the Nowcasting Lab website and analyze in this paper, to 20 (plus GDP growth).

The variable selection for the MIDAS and U-MIDAS models is modified to consider the models' particularities. First, the forecasts of these models are done directly (contrasting with the iterative approach of the bridge equation model). As a consequence, the variable selection is conducted for each forecast horizon, as the variables might not be equally useful for each forecast horizon. Second, the variable selection is done in mixed frequency, as the models are estimated in mixed frequency. For this, we adapt the MIDASSO approach by Siliverstovs (2017), which combines the elastic net algorithm by Zou and Hastie (2005) with the U-MIDAS approach. Concretely, each monthly predictor variable is split into three quarterly series containing the first, second, and third months of a quarter (blocking or skip-sampling approach). The three series are then included in the variable selection matrix. If one or more of the three series are selected in the optimal combination, the monthly variable will enter the MIDAS model combination and the U-MIDAS model combination.

B Model specifications

This section describes the forecasting models currently used in the Nowcasting Lab.

B.1 Mixed-frequency dynamic factor model

Dynamic factor models are widely used in macroeconomic forecasting. The main idea is to synthesize common information from many variables by extracting a few latent factors that can, among other things, be used for forecasting. In that sense, factor analysis is a dimension reduction technique to tackle the "curse of dimensionality" by summarizing the main sources of variation in the data set. Early examples of factor models in macroeconomics are Sargent and Sims (1977), Geweke (1977), Engle and Watson (1981), and Stock and Watson (1989).

The mixed-frequency dynamic factor model (MF-DFM) implemented in the Nowcasting Lab follows Bok et al. (2018). This model also stands behind the Nowcasting Report of the Federal Reserve Bank of New York (www.newyorkfed.org/research/policy/nowcast). The model assumes that a large data set of observed time series $\mathbf{x}_t = (x_{1,t}, \ldots, x_{n,t})'$ can be described by a few unobserved common factors $\mathbf{f}_t = (f_{1,t}, \ldots, f_{r,t})'$, which capture common movements in the data, and an idiosyncratic component $\mathbf{e}_t = (e_{1,t}, \ldots, e_{n,t})'$, which captures variable-specific variation as well as measurement errors.[1] We follow the notation of Bok et al. (2018) that t is the time index, n is the number of variables in the data set, and r is the number of latent factors. The following measurement equation describes the link between the data and the unobserved factors:

$$\mathbf{x}_t = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_t + \mathbf{e}_t, \tag{1}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ are the unconditional means of the observed variables and $\boldsymbol{\Lambda} = (\lambda_1, \ldots, \lambda_n)'$ is a matrix that contains the time-invariant factor loadings. For the estimation, the variables in vector $\mathbf{x}_t$ have been transformed to satisfy the condition of stationarity (see Section 3.1 for the transformations applied) and are demeaned and normalized such that the vector $\boldsymbol{\mu}$ can be removed from the equation. As a consequence, the factors are assumed to have mean zero as well. For the model shown on the Nowcasting Lab website, only one common factor is used.[2]

The law of motion of the factors is described by the transition equation for the factors, which follows a vector autoregression (VAR) process of order 1 and is defined as

$$\mathbf{f}_t = \mathbf{A}\mathbf{f}_{t-1} + \mathbf{u}_t, \qquad \mathbf{u}_t \sim iid\, \mathcal{N}(\mathbf{0}, \mathbf{Q}), \tag{2}$$

where matrix $\boldsymbol{A}$ contains the autoregressive coefficents. The error term $\mathbf{u}_t$ follows a normal distribution with mean zero and variance-covariance matrix $\mathbf{Q}$. Further, the

[1] Bok et al. (2018) estimate a common factor ("global factor"), on which all variables load, and "local factors", on which only specific variable groups can load. The existence of local factors can improve inference. However, the authors also state that estimation is robust to the presence of local correlations among idiosyncratic components. For simplicity reasons, we refrain from including local factors in our model.

[2] In addition, several alternative specifications with more than one factor are run in the background. The output of these specifications is stored in the Nowcasting Lab database. This allows us to test, at a later stage, whether the parsimonious version with only one common factor delivers better forecasts than the more complex models.

transition equation for the dynamics of the idiosyncratic component follows a VAR process of order 1 and is described by

$$\mathbf{e}_t = \boldsymbol{\rho}\mathbf{e}_{t-1} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim iid\,\mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right), \tag{3}$$

where the error term follows a normal distribution with mean zero and variance-covariance matrix $\boldsymbol{\Sigma}$. Since the dynamic factors account for the majority of the cross-correlation, it can be assumed that the idiosyncratic components are cross-sectionally uncorrelated and, thus, the covariance matrix $\boldsymbol{\Sigma}$ is diagonal with $\mathbf{E}[\epsilon_{i,t}\epsilon_{j,s}] = 0$ for $j \neq i$.

As it is possible to write the whole system in state-space form (Bańbura et al., 2011; Bańbura and Modugno, 2014), the model can be estimated by quasi-maximum likelihood (QML) estimation using an expectation-maximization (EM) algorithm as proposed by Doz et al. (2012) and extended for missing data by Bańbura and Modugno (2014). First, the latent factors estimates and the factor loading estimates, $\hat{\mathbf{f}}_t$ and $\hat{\boldsymbol{\Lambda}}$, are initialized by principal component analysis. The idiosyncratic errors in the measurement equation are calculated by $\hat{\mathbf{e}}_t = \mathbf{x}_t - \hat{\boldsymbol{\Lambda}}\hat{\mathbf{f}}_t$ with its associated empirical variance $\hat{\sigma}_e^2$. Estimation of the VAR parameters $\hat{\mathbf{A}}$ and $\hat{\rho}$ in the transition equations is done line by line by ordinary least squares (OLS). Again, the error term in the transition equations is calculated by $\hat{\mathbf{u}}_t = \hat{\mathbf{f}}_t - \hat{\mathbf{A}}\hat{\mathbf{f}}_{t-1}$ with its associated empirical variance $\hat{\sigma}_u^2$ and with $\hat{\boldsymbol{\epsilon}}_t$ being set as a diagonal with a fixed variance. Second, the expectation of the log-likelihood is calculated by updating the parameters using a Kalman filter and smoother conditional on the estimated parameters. Third, the parameters are re-estimated through the maximization of the log-likelihood. The three steps are iterated in the EM algorithm until convergence of the log-likelihood estimator is obtained.[3] Finally, as the dynamics of the factors are explicitly captured, the estimates $\hat{\mathbf{f}}_t$ can be used to recursively calculate $\hat{\mathbf{f}}_{T+h|T}$ using the transition equation of the factors, while the measurement equation can be used to obtain $\hat{\mathbf{x}}_{T+h|T} = \hat{\boldsymbol{\Lambda}}\hat{\mathbf{f}}_{T+h|T}$ as a forecast. Further, the forecast impacts ("news") can be easily retrieved from the Kalman smoother (Bańbura and Modugno, 2014).

The representation of the model in state space form further allows the handling of missing observations which exist due to different release lags, different lengths of the data, reporting errors as well as mixed frequencies. For the estimation of the initial parameters, a balanced panel is created. First, missing variable observations within the panel are either dropped or proxied by linear interpolation. Then, the

[3]The threshold for convergence in the EM algorithm is set to 10^{-4} and the maximum iteration steps is set to 5'000.

missing observations are updated by expectations which are estimated conditional on the existing variable observations, the factors, and the factor loadings from the previous Kalman filter iteration by using a selection matrix following Bańbura and Modugno (2014).

In order to handle variables of different frequencies, the model applies time-aggregation functions that translate the lower-frequency variable observations to latent high-frequency variable observations and vice versa. The type of each variable determines the form of the corresponding function. For stock variables, the (latent) high-frequency variable observations are averaged over each low-frequency period. In contrast, for flow variables, the (latent) high-frequency variable observations are summed over each low-frequency period. For flow variables that enter the model in growth rates, such as GDP, the exact temporal aggregation would require a non-linear state space model. In order to preserve a linear model, Mariano and Murasawa (2003) propose an approximate temporal aggregation method using the geometric instead of the arithmetic mean.

In the following, we demonstrate the Mariano and Murasawa (2003) method for the case of the aggregation of GDP from monthly to quarterly frequency. By construction, the quarterly GDP level, GDP_t^Q, is always observed in the third month of a quarter. The relationship between GDP_t^Q and the monthly GDP levels, GDP_t^M, can be expressed by

$$\mathrm{GDP}_t^Q = \mathrm{GDP}_t^M + \mathrm{GDP}_{t-1}^M + \mathrm{GDP}_{t-2}^M, \qquad t = 3, 6, 9, \ldots. \tag{4}$$

We define $Y_t^Q = 100 \log(\mathrm{GDP}_t^Q)$ and $Y_t^M = 100 \log(\mathrm{GDP}_t^M)$. Further, we define the quarterly growth rate $y_t^Q = Y_t^Q - Y_{t-3}^Q$ for $t = 3, 6, 9, \ldots$ and let $y_t = \Delta Y_t^M$, where ΔY_t^M is the latent monthly growth rate of GDP from the model.

Using the geometric mean approximation of Mariano and Murasawa (2003), the quarterly growth rate is given by

$$\begin{aligned}
\bar{y}_t^Q &= Y_t^Q - Y_{t-3}^Q \\
&\approx \left(Y_t^M + Y_{t-1}^M + Y_{t-2}^M\right) - \left(Y_{t-3}^M + Y_{t-4}^M + Y_{t-5}^M\right) \\
&= \left(Y_t^M - Y_{t-3}^M\right) + \left(Y_{t-1}^M - Y_{t-4}^M\right) + \left(Y_2^M - Y_{t-5}^M\right).
\end{aligned}$$

Introducing additional terms into the equation, which cancel out, gives

$$\begin{aligned}\bar{y}_t^Q =&\Big(Y_t^M - Y_{t-1}^M + Y_{t-1}^M - Y_{t-2}^M + Y_{t-2}^M - Y_{t-3}^M\Big) + \\ &\Big(Y_{t-1}^M - Y_{t-2}^M + Y_{t-2}^M - Y_{t-3}^M + Y_{t-3}^M - \log Y_{t-4}^M\Big) + \\ &\Big(Y_{t-2}^M - Y_{t-3}^M + Y_{t-3}^M - Y_{t-4}^M + Y_{t-4}^M - Y_{t-5}^M\Big).\end{aligned}$$

Rearranging terms and expressing log differences as growth rates, the previous equation simplifies to

$$\bar{y}_t^Q = y_t + 2y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4} \qquad t = 3, 6, 9, \ldots. \tag{5}$$

Following Bańbura et al. (2011), the triangular weighting structure in Equation (5) enters $\mathbf{\Lambda}$ in Equation (1) as distributed lag matrices, resulting for the quarterly variables in

$$\mathbf{y}_t = \boldsymbol{\mu}_Q + \mathbf{\Lambda}_Q \mathbf{f}_t + \mathbf{e}_t^Q, \tag{6}$$

$$\mathbf{e}_t^Q = \boldsymbol{\rho}_{\boldsymbol{Q}} \mathbf{e}_{t-1}^Q + \boldsymbol{\epsilon}_t^Q, \qquad \boldsymbol{\epsilon}_t^Q \sim iid\, \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_Q\right). \tag{7}$$

In the state space form, this results in additional columns for the lagged variable coefficients in $\mathbf{\Lambda}$ and additional entries for the lagged factors and error terms of the model.

B.2 Mixed data sampling

Building on earlier research on distributed lag polynomial models, the mixed data sampling (MIDAS) approach has been introduced by Ghysels and co-authors (e.g., Ghysels et al., 2007, Andreou et al., 2010, and Andreou et al., 2011). MIDAS attempts to balance two goals which are usually in a trade-off position to each other. The first goal is model flexibility, that is, allowing the relative importance of any lagged observation compared to any other lagged observation to be determined by the data and not pre-determined by the model itself. The second goal is a parsimoniously parameterized model that prevents parameter proliferation or overfitting. MIDAS models and combinations of MIDAS with other models or methods are widely used in both academic research and applied forecasting work (e.g., Clements and Galvão, 2008, 2009; Andreou et al., 2013; Kuzin et al., 2013; Ferrara et al., 2014; Heinisch and Scheufele, 2018; Mogliani and Simoni, 2021; Babii et al., 2021). For the Nowcasting Lab, we implement the single-predictor MIDAS model with autoregressive terms and an intercept as described in more detail below.[4] This model has proven to deliver good forecasting results in conjunction with forecast combinations (see Section C).

In the MIDAS approach, a low-frequency variable is forecast using a possibly large number of (lagged) observations of a high-frequency variable, with the lag coefficients being modeled as a possibly very flexible non-linear distributed lag function. More formally, let y_t^Q be the quarterly GDP growth of quarter $t = 1, \ldots, T$, and let $x_{t-2/3}^M$, $x_{t-1/3}^M$, and x_t^M be the first, second, and third monthly observations of a monthly variable (e.g., growth in money supply) within quarter t. The single-predictor MIDAS estimation equation with autoregressive terms and an intercept then writes

$$y_t^Q = \alpha_0 + \sum_{i=1}^{I} \alpha_i y_{t-i}^Q + \sum_{j=0}^{J} b(j, \boldsymbol{\theta}) x_{t-h/3-j/3}^M + \epsilon_t^Q, \tag{8}$$

where α_0 is a constant parameter, $\alpha_1, \ldots, \alpha_I$ are the autoregressive parameters, ϵ_i is an error term, and $b(j, \boldsymbol{\theta})$ is a distributed lag function that depends on the lag index j and on the parameters $\theta_0, \ldots, \theta_P$ which are summarized in the parameter vector $\boldsymbol{\theta}$.

Notably, the model in Equation (8) is estimated separately for each forecast horizon h. For instance, h equals 0 if the monthly predictor variable's third monthly observation of quarter t, x_t^M, is already published, and y_t^Q is not yet published. h equals 1 if the monthly predictor variable's second monthly observation of quarter t, $x_{t-1/3}^M$, is published, but x_t^M and y_t^Q are not yet released. And so on. In the Nowcasting Lab, the maximum h is 5 or 6, depending on the exact release difference between quarterly GDP and the respective monthly predictor variable.

For the distributed lag function, we employ the non-exponential Almon lag polynomial originally proposed by Almon (1965):

$$b(j, \boldsymbol{\theta}) = \sum_{p=0}^{P} \theta_p h^p. \tag{9}$$

In a preliminary analysis, the non-exponential Almon lag polynomial led to higher

[4]In the terminology of Andreou et al. (2013), this is the autoregressive distributed lag (ARDL) MIDAS model. The model set of the Nowcasting Lab also comprises single-predictor MIDAS models without autoregressive terms and/or an intercept. While we do not show the predictions resulting from these alternative specifications on the website of the lab, we store them in the lab's database to analyze and compare their forecast performance over time.

forecast accuracy than the exponential Almon lag polynomial or alternative non-linear polynomial schemes. Notably, even though the non-exponential Almon lag polynomial is highly non-linear, MIDAS models with non-exponential Almon lag polynomials can be transformed into a linear form. In turn, ordinary least squares (OLS) estimation of the model parameters $\alpha_0, \dots, \alpha_I$ and $\theta_0, \dots, \theta_p$ is feasible.[5] Usually, the number of monthly variable lags in Equation (8) is chosen to be relatively large (i.e., J is relatively large), and the number of polynomial parameters is chosen to be relatively small (i.e., the order of the polynomial, P, in Equation (9) is relatively small). Thus, MIDAS models can be at the same time rich in information (due to the many monthly lags), flexible (due to the highly non-linear polynomial), and parsimonious (due to the only few polynomial parameters that need to be estimated).

The parameters to be optimized are (1.) the polynomial lags $\theta_0, \dots, \theta_P$, where we let P be 1, 2 or 3, (2.) the constant parameter β_0, and (3.) the number of autoregressive parameters $\alpha_1, \dots, \alpha_I$, where we let I be 1, 2, 3 or 4. Further, we let J be between P and 11. Altogether, this gives $12^4 + 11^4 + 10^4 = 45377$ different model specifications. We choose, separately for each monthly predictor variable and each monthly horizon h, the model specification that yields the lowest Bayesian information criterion (BIC).

B.3 Unrestricted mixed-frequency model

Foroni et al. (2015) propose unrestricted mixed-frequency regressions and call the approach unrestricted MIDAS (U-MIDAS). Here, a low-frequency variable is forecast with (lagged) observations of a high-frequency predictor variable whose lag coefficients are left fully unrestricted and, hence, can be estimated by OLS. For the Nowcasting Lab, we implement the single predictor U-MIDAS model with autoregressive terms and an intercept:

$$y_t^Q = \alpha_0 + \sum_{i=1}^{I} \alpha_i y_{t-i}^Q + \sum_{j=0}^{J} \beta_j x_{t-h/3-j/3}^M + \epsilon_t^Q, \tag{10}$$

where $\beta_1, \dots, \beta_I$ are the parameters attached to the observations of the monthly predictor variable, and all other terms have the same definition as for Equation (8).[6] While MIDAS and U-MIDAS share the same name, we consider them different

[5] See, e.g., Mikosch and Zhang, 2014, Appendix 6.1. This feature distinguishes the non-exponential Almon lag polynomial from alternative non-linear polynomials, which require non-linear least squares estimation.

[6] The Nowcasting Lab model set further comprises single-predictor U-MIDAS models without au-

categories. A common feature is that both approaches organize the mixed-frequency data according to a blocking (or skip-sampling) approach (see, e.g., Section 1 of Siliverstovs, 2017). The major difference is that MIDAS models involve the use of non-linear lag polynomial functions, whereas in U-MIDAS models the individual coefficients are left unrestricted. While extremely flexible, the U-MIDAS model is not parsimonious when the number of high-frequency variable lags is large. Thus, U-MIDAS is only superior to MIDAS when the number of lags is sufficiently small (see the evaluation in Foroni et al., 2015).

We let the number of (lagged) monthly variable observations, $J + 1$, range between 1 and 12, and the number of low-frequency autoregressive variable lags, I, range between 1 and 4. Hence, there exist $12^4 = 20736$ possible model specifications. As for the MIDAS model, we select, separately for each monthly predictor variable and each monthly horizon h, the specification with the lowest Bayesian information criterion (BIC).

B.4 Bridge equations

Bridge equation models are popular in policy institutions because they are transparent and easy to implement. Research applications include Ingenito and Trehan (1996), Baffigi et al. (2004), Golinelli and Parigi (2007), Diron (2008), Rünstler et al. (2009), Bulligan et al. (2010), Bulligan et al. (2015), Angelini et al. (2011), Foroni and Marcellino (2014), Schumacher (2016), and Mikosch and Solanko (2019).

For the Nowcasting Lab, we employ the classic bridge equation procedure as outlined in, e.g., Schumacher (2016): First, we estimate an iterative univariate autoregressive (AR) model with an intercept term for a high-frequency (monthly) predictor variable x_t^M:

$$x_t^M = \phi_0 + \sum_{k=1}^{K} \phi_k x_{t-k}^M + \epsilon_t^M, \tag{11}$$

where we allow the lag order k to be between 1 and 12 and choose the k that minimizes the BIC of the model. Next, the high-frequency variable is aggregated to a lower (quarterly) frequency. Following Chow and Lin (1971), e.g., the time-

toregressive terms and/or an intercept. The output of these model variants is stored in the database of the lab, but is not shown on the website.

aggregation can be formalized by

$$x_t^Q = \omega(L^{1/3})x_t^M = \sum_{r=0}^{R} \omega_r x_{t-r/3}^M, \tag{12}$$

where $\omega(L^{1/3})$ represents a lag polynomial time-aggregation function and ω_r is an aggregation weight. The form of the time-aggregation function is determined by the stock or flow nature of the predictor variable x_t^M, and whether it comes in levels or growth rates. In case x_t^M is a (stationary) flow variable, such as exports, the value for quarter t is equal to the sum over the monthly observations in quarter t:

$$x_t^Q = x_t^M + x_{t-1/3}^M + x_{t-2/3}^M.$$

In case x_t^M is a stock variable, such as money supply, the value for quarter t is equal to the average over the monthly observations in quarter t:

$$x_t^Q = \frac{1}{3}(x_t^M + x_{t-1/3}^M + x_{t-2/3}^M).$$

Further, in case x_t^M is a month-on-month growth rate, the quarter-on-quarter growth rate of quarter t is, according to the geometric mean approximation method proposed by Mariano and Murasawa (2003), given by

$$x_t^Q = \frac{1}{3}x_t^M + \frac{2}{3}x_{t-1/3}^M + x_{t-2/3}^M + \frac{2}{3}x_{t-1}^M + \frac{1}{3}x_{t-4/3}^M.$$

After the temporal aggregation of the high-frequency variable, we estimate the following model for quarterly GDP growth:

$$y_t^Q = \alpha_0 + \sum_{i=1}^{I} \alpha_i y_{t-i}^Q + \sum_{n=0}^{N} \gamma_j x_{t-j}^Q + \epsilon_t^Q, \tag{13}$$

where the quarter t-value and the lagged values of the time-aggregated predictor variable enter on the right-hand side along with the lags of GDP growth and an intercept term.[7] We let the lag lengths of both the quarterly GDP growth and the time-aggregated predictor variable range between 1 and 4, which gives $4^4 = 16$ different model specifications. Again, we choose the model specification that minimizes the BIC.

To illustrate the forecasting process using the estimation and aggregation equations above, suppose that we want to now-/forecast GDP growth of 2022Q3 and 2022Q4. Further, suppose that x_t^M is observed until July 2022 and x_t^Q is observed

[7] The Nowcasting Lab model set also includes a bridge equation model excluding lags of GDP growth. The output of this model is not shown on the website but stored in the database.

until 2022Q2. Using the parameter estimates obtained from estimating Equation (11) by OLS, we first iteratively predict x_t^M until December 2022. Second, based on Equation (12), we build forecasts of x_t^Q for 2022Q3 and 2022Q4 by time-aggregating the predicted and realized monthly values. Third, we iteratively forecast y_t^Q for 2022Q3 and 2022Q4 using the predicted and realized quarterly values of x_t^Q and the parameter estimates obtained from estimating Equation (13) by OLS.

B.5 Benchmark models

Next to the MF-DFM, MIDAS, U-MIDAS, and bridge equation models, the Nowcasting Lab includes several simple forecasting models. The predictions resulting from these models do not enter the forecast pooling procedure described in Section C. Instead, they are displayed separately on the website and can be used as comparative benchmarks for the pooled predictions and the predictions from the workhorse models above.

Random walk model: The nowcast and the one-quarter ahead forecast for GDP growth are set equal to the latest GDP growth realization. This procedure corresponds to a random walk forecasting model.

Rolling mean model: The nowcast and the one-quarter ahead forecast for GDP growth are equal to the rolling in-sample mean of past GDP growth. We choose the in-sample window length (between 2 and 40 quarters) which minimizes the root mean square forecast error (RMSFE) of the past 28 quarters.

Direct AR model: The nowcast and the one-quarter ahead forecast for GDP growth result from a direct autoregressive (AR) model with intercept:

$$y_t^Q = \alpha_0^h + \sum_{i=1}^{I} \alpha_i^h y_{t-h-i}^Q + \epsilon_t^{h,Q}, \tag{14}$$

where the prediction horizon $h \in \{0, 1\}$. Equation (14) is estimated separately for each h (which is why we attach in the equation the h-index to the α-parameters and the error term). The optimal autoregressive lag length is chosen, in a range between 1 and 4, according to the BIC.

Iterative AR model: The nowcast and the one-quarter ahead forecast for GDP growth result from an iterative one-step ahead AR model with the optimal lag length being determined, in a range between 1 and 4, according to the BIC. The model corresponds to Equation (14) with $h = 0$. Based on the parameter estimates, y_t^Q is predicted iteratively into the future.

ARIMA model: The nowcast and the one-quarter ahead forecast for GDP growth result from an iterative one-step ahead autoregressive integrated moving-average (ARIMA) model. We set the degree of differencing to zero and let both the order of the autoregressive model (i.e., the number of autoregressive lags) and the order of the moving average model range between 1 and 4, which gives $4^4 = 16$ alternative model specifications. Once again, the BIC determines the optimal model specification among the 16 alternative specifications.

C The forecast combinations procedure

We use a two-step forecast combination procedure. In the first step, the predictions stemming from all single-predictor equations are pooled separately for the MIDAS model, the bridge equation model, and the U-MIDAS model, respectively. The single-factor MF-DFM only generates one forecast and does not require forecast pooling in the first step. To obtain an overall pooled prediction, the second step is to take an unweighted average of the four (pooled) predictions resulting from the MF-DFM, the MIDAS model, the bridge equation model, and the U-MIDAS model.

For the first step, we build pooled predictions for the MIDAS model, the bridge equation model, and the U-MIDAS model, respectively, by taking the unweighted average over the predictions resulting from the different single-predictor equations of each model. Once the Nowcasting Lab has collected a sufficiently long real-time history of quarterly GDP growth prediction errors, we will change from equal weighting to a forecast combinations procedure based on past prediction errors (e.g., Stock and Watson, 2004; Kuzin et al., 2013). In this variant, we will pool the predictions stemming from the different single-predictor equations of a particular model (MIDAS, bridge equation, or U-MIDAS) by assigning a weight between zero and one to each prediction. All weights must add up to one (weighted averaging). The weights are determined in the following way: We calculate for each single-predictor equation the rolling mean square forecast error (MSFE) using the predictions of the previous four quarters. Each MSFE is then divided by the sum of the MSFEs of all single-predictor equations such that the MSFEs of all equations sum to one (normalization). Finally, the combination weight for each prediction is the inverse of the normalized MSFE divided by the sum of all normalized inverted MSFEs. To provide the intuition behind the procedure: Instead of selecting predictor variables based on in-sample criteria, the predictions resulting from the different predictors are weighted according to their forecast accuracy in the recent past. The better the predictions stemming from a particular predictor in the recent past, the higher the weight of the current prediction of that predictor in constructing the weighted average of the current pooled prediction.

Notably, the two-step forecast combinations procedure is run completely separately for each prediction date (i.e., the number of days until the next GDP release) and each prediction horizon (nowcast and one-quarter ahead forecast).

D Forecast comparison using latest GDP releases

Given the large revisions of quarterly GDP during the COVID-19 period, the predictive power of the models might depend on which GDP release is used to calculate the forecast errors (see Section 4.2). To check this, Tables 6 and 7 iterate the RMSFE calculations shown in Tables 2 and 3 based on the latest available GDP release for each quarter instead of always the first GDP release.[8] For the 2020Q1–2020Q4 period, the forecast errors for all models are slightly lower on average for all economies compared to using the first GDP release. The MF-DFM continues to provide the best forecasts for all economies except Sweden (SE), where the MIDAS model performs slightly better. For the 2021Q1–2021Q4 period, the RMSFEs for all models, except the random walk model, are slightly higher on average over all economies when using the latest instead of the first GDP releases. Notably, the U-MIDAS model and the random walk model now outperform the MF-DFM in terms of RMSFE on average over all economies. However, as previously, no single model clearly dominates the other models in forecast performance.

[8] The cutoff date for the analysis is 05/12/2022. At this date, the quarterly GDP vintages for the 2020–2021 period were still subject to substantial revisions for some economies. The longer a quarter has passed, the weaker the GDP revisions usually tend to be, hence, the more can the latest available GDP release be considered as final (except for benchmark revisions). We will update Tables 6 and 7 in a revised paper version.

Table 6: **RMSFEs for 2020Q1-Q4 using latest GDP releases**

Economy	Workhorse models					Benchmark models				
	DFM	M	UM	BE	Pooled	ARIMA	AR_{it}	AR_{di}	RM	RW
AT	5.59	7.34	7.59	7.69	7.05	9.15	8.29	8.92	7.82	8.27
	0.68	*0.89*	*0.92*	*0.93*	*0.85*	*1.11*	*1.00*	*1.08*	*0.95*	*1.00*
BE	5.84	7.78	6.71	7.56	6.97	10.81	8.50	9.10	8.74	8.12
	0.72	*0.96*	*0.83*	*0.93*	*0.86*	*1.33*	*1.05*	*1.12*	*1.08*	*1.00*
CH	3.23	4.74	4.07	4.61	4.13	6.20	6.00	5.47	5.68	5.37
	0.60	*0.88*	*0.76*	*0.86*	*0.77*	*1.15*	*1.12*	*1.02*	*1.06*	*1.00*
DE	3.79	5.38	5.26	5.23	4.91	6.28	6.01	6.34	7.09	6.67
	0.57	*0.81*	*0.79*	*0.78*	*0.74*	*0.94*	*0.90*	*0.95*	*1.06*	*1.00*
EA	4.79	8.01	7.23	8.17	7.05	12.24	9.98	10.45	9.05	8.60
	0.56	*0.93*	*0.84*	*0.95*	*0.82*	*1.42*	*1.16*	*1.21*	*1.05*	*1.00*
ES	8.46	12.02	11.82	12.08	10.99	19.75	18.67	18.61	13.63	12.61
	0.67	*0.95*	*0.94*	*0.96*	*0.87*	*1.57*	*1.48*	*1.48*	*1.08*	*1.00*
FR	6.71	10.61	10.04	10.57	9.39	13.21	16.41	14.12	12.11	11.86
	0.57	*0.90*	*0.85*	*0.89*	*0.79*	*1.11*	*1.38*	*1.19*	*1.02*	*1.00*
GB	9.19	11.53	11.22	11.55	10.51	20.32	15.46	17.80	13.63	13.01
	0.71	*0.89*	*0.86*	*0.89*	*0.81*	*1.56*	*1.19*	*1.37*	*1.05*	*1.00*
IT	6.87	9.70	8.27	9.51	8.59	12.52	10.79	11.52	10.51	10.28
	0.67	*0.94*	*0.80*	*0.92*	*0.84*	*1.22*	*1.05*	*1.12*	*1.02*	*1.00*
NL	2.67	4.40	4.28	4.69	4.01	5.91	4.91	5.46	5.54	5.19
	0.51	*0.85*	*0.82*	*0.90*	*0.77*	*1.14*	*0.95*	*1.05*	*1.07*	*1.00*
PL	2.82	3.92	4.52	3.91	3.69	4.00	4.03	4.10	5.37	5.22
	0.54	*0.75*	*0.87*	*0.75*	*0.71*	*0.77*	*0.77*	*0.79*	*1.03*	*1.00*
SE	4.29	4.22	4.56	4.40	4.37	4.88	4.80	5.05	4.78	5.11
	0.84	*0.83*	*0.89*	*0.86*	*0.85*	*0.96*	*0.94*	*0.99*	*0.94*	*1.00*
Average	5.35	7.47	7.13	7.50	6.80	10.44	9.49	9.75	8.66	8.36
	0.64	*0.89*	*0.85*	*0.90*	*0.81*	*1.25*	*1.14*	*1.17*	*1.04*	*1.00*

Notes: See Table 2, but the RMSFEs are calculated using the latest available GDP release (as of 05/12/2022) for each quarter instead of using always the first GDP release.

Table 7: **RMSFEs for 2021Q1-Q4 using latest GDP releases**

Economy	Workhorse models					Benchmark models				
	DFM	M	UM	BE	Pooled	ARIMA	AR_{it}	AR_{di}	RM	RW
AT	1.83	3.09	1.93	2.99	2.44	2.34	2.30	2.49	2.72	4.83
	0.38	*0.64*	*0.40*	*0.62*	*0.51*	*0.48*	*0.48*	*0.52*	*0.56*	*1.00*
BE	1.90	1.00	1.53	1.11	1.20	1.54	2.73	2.41	1.23	2.00
	0.95	*0.50*	*0.76*	*0.56*	*0.60*	*0.77*	*1.37*	*1.21*	*0.62*	*1.00*
CH	1.04	0.88	0.82	1.16	0.94	0.91	0.76	0.81	0.87	2.11
	0.49	*0.42*	*0.39*	*0.55*	*0.45*	*0.43*	*0.36*	*0.38*	*0.41*	*1.00*
DE	1.01	1.24	0.72	1.11	0.93	0.96	0.91	0.97	1.07	2.76
	0.37	*0.45*	*0.26*	*0.40*	*0.34*	*0.35*	*0.33*	*0.35*	*0.39*	*1.00*
EA	1.29	1.60	1.37	1.62	1.00	1.13	1.42	2.01	1.10	2.98
	0.43	*0.54*	*0.46*	*0.54*	*0.34*	*0.38*	*0.48*	*0.68*	*0.37*	*1.00*
ES	1.15	2.44	2.16	2.53	1.98	6.17	6.29	23.54	1.65	3.12
	0.37	*0.78*	*0.69*	*0.81*	*0.63*	*1.98*	*2.01*	*7.53*	*0.53*	*1.00*
FR	2.15	2.11	1.27	1.80	1.60	8.37	2.39	17.14	1.10	3.62
	0.59	*0.58*	*0.35*	*0.50*	*0.44*	*2.31*	*0.66*	*4.73*	*0.31*	*1.00*
GB	1.88	3.06	2.21	2.93	2.49	2.88	4.88	3.36	2.59	4.57
	0.41	*0.67*	*0.48*	*0.64*	*0.54*	*0.63*	*1.07*	*0.74*	*0.57*	*1.00*
IT	1.79	1.74	1.64	1.70	1.33	1.86	3.17	4.53	1.66	3.46
	0.52	*0.50*	*0.47*	*0.49*	*0.38*	*0.54*	*0.92*	*1.31*	*0.48*	*1.00*
NL	1.24	1.59	1.53	1.46	1.33	1.50	1.49	1.43	1.37	2.62
	0.47	*0.61*	*0.58*	*0.56*	*0.51*	*0.57*	*0.57*	*0.55*	*0.52*	*1.00*
PL	1.28	1.14	1.48	1.16	1.27	1.28	1.66	1.33	1.27	1.02
	1.26	*1.12*	*1.46*	*1.14*	*1.24*	*1.25*	*1.62*	*1.30*	*1.25*	*1.00*
SE	1.48	1.43	1.07	0.98	1.14	1.06	1.05	1.01	0.98	0.56
	2.64	*2.56*	*1.92*	*1.76*	*2.04*	*1.90*	*1.87*	*1.80*	*1.76*	*1.00*
Average	1.50	1.78	1.48	1.71	1.47	2.50	2.42	5.09	1.47	2.80
	0.54	*0.64*	*0.53*	*0.61*	*0.52*	*0.89*	*0.86*	*1.82*	*0.52*	*1.00*

Notes: See Table 3, but the RMSFEs are calculated using the latest available GDP release (as of 05/12/2022) for each quarter instead of using always the first GDP release.

E Additional tables and figures

Table 8: RMSFEs for 2020Q1–Q4: 2-quarter ahead-of-release horizon

Economy	Workhorse models				Pooled	Benchmark models				
	DFM	M	UM	BE		ARIMA	AR_{it}	AR_{di}	RM	RW
AT	7.46	6.76	7.42	6.93	7.14	8.28	8.04	6.98	7.52	8.41
	0.89	*0.80*	*0.88*	*0.82*	*0.85*	*0.98*	*0.96*	*0.83*	*0.89*	*1.00*
BE	7.78	8.91	7.76	9.20	8.38	13.65	8.96	9.05	11.10	13.20
	0.59	*0.68*	*0.59*	*0.70*	*0.63*	*1.03*	*0.68*	*0.69*	*0.84*	*1.00*
CH	5.04	4.88	4.89	5.41	4.93	6.86	5.88	4.85	6.42	7.47
	0.68	*0.65*	*0.66*	*0.72*	*0.66*	*0.92*	*0.79*	*0.65*	*0.86*	*1.00*
DE	7.45	5.68	5.40	5.28	5.95	6.09	5.26	5.90	7.63	8.27
	0.90	*0.69*	*0.65*	*0.64*	*0.72*	*0.74*	*0.64*	*0.71*	*0.92*	*1.00*
EA	7.99	8.39	7.51	8.38	7.92	13.63	8.07	9.16	9.98	11.04
	0.72	*0.76*	*0.68*	*0.76*	*0.72*	*1.23*	*0.73*	*0.83*	*0.90*	*1.00*
ES	10.50	13.93	13.87	14.22	13.13	23.73	12.12	14.20	15.18	16.37
	0.64	*0.85*	*0.85*	*0.87*	*0.80*	*1.45*	*0.74*	*0.87*	*0.93*	*1.00*
FR	11.00	11.41	10.05	10.90	10.33	17.15	12.47	12.92	13.53	14.26
	0.77	*0.80*	*0.70*	*0.76*	*0.72*	*1.20*	*0.87*	*0.91*	*0.95*	*1.00*
GB	10.19	12.27	12.09	11.99	11.52	21.07	15.19	13.93	13.99	15.41
	0.66	*0.80*	*0.78*	*0.78*	*0.75*	*1.37*	*0.99*	*0.90*	*0.91*	*1.00*
IT	9.61	9.14	9.13	9.09	8.81	12.55	9.89	9.88	11.36	12.02
	0.80	*0.76*	*0.76*	*0.76*	*0.73*	*1.04*	*0.82*	*0.82*	*0.95*	*1.00*
NL	5.97	5.12	5.00	4.98	5.16	6.15	4.94	5.51	6.46	7.20
	0.83	*0.71*	*0.69*	*0.69*	*0.72*	*0.85*	*0.69*	*0.77*	*0.90*	*1.00*
PL	6.24	5.87	5.89	5.99	6.00	6.00	6.29	6.17	7.56	8.50
	0.73	*0.69*	*0.69*	*0.71*	*0.71*	*0.71*	*0.74*	*0.73*	*0.89*	*1.00*
SE	3.96	3.97	3.58	3.87	3.71	4.28	3.58	4.34	5.32	5.76
	0.69	*0.69*	*0.62*	*0.67*	*0.64*	*0.74*	*0.62*	*0.75*	*0.92*	*1.00*
Average	7.80	8.00	7.70	8.00	7.70	11.60	8.40	8.60	9.70	10.70
	0.73	*0.75*	*0.72*	*0.75*	*0.72*	*1.08*	*0.79*	*0.80*	*0.91*	*1.00*

Notes: See Table 2, but the RMSFEs are calculated using for each quarter the GDP growth prediction at the first day of the prediction period, i.e., two quarters ahead of the GDP release.

Table 9: **RMSFEs for 2020Q1–Q4: 1-quarter ahead-of-release horizon**

Economy	Workhorse models					Benchmark models				
	DFM	M	UM	BE	Pooled	ARIMA	AR_{it}	AR_{di}	RM	RW
AT	5.96	8.75	9.07	9.11	8.22	9.46	9.08	9.08	8.67	12.04
	0.50	*0.73*	*0.75*	*0.76*	*0.68*	*0.79*	*0.75*	*0.75*	*0.72*	*1.00*
BE	5.82	9.14	6.62	8.94	7.54	11.02	11.00	11.00	8.82	11.50
	0.51	*0.79*	*0.58*	*0.78*	*0.66*	*0.96*	*0.96*	*0.96*	*0.77*	*1.00*
CH	3.66	6.19	4.50	6.21	5.09	6.84	7.17	7.17	6.08	7.75
	0.47	*0.80*	*0.58*	*0.80*	*0.66*	*0.88*	*0.93*	*0.93*	*0.78*	*1.00*
DE	2.17	5.83	5.46	5.82	4.79	6.60	6.77	6.77	6.61	9.15
	0.24	*0.64*	*0.60*	*0.64*	*0.52*	*0.72*	*0.74*	*0.74*	*0.72*	*1.00*
EA	4.88	10.29	8.19	10.30	7.85	11.69	12.34	12.34	9.10	12.59
	0.39	*0.82*	*0.65*	*0.82*	*0.62*	*0.93*	*0.98*	*0.98*	*0.72*	*1.00*
ES	8.33	16.18	14.56	16.14	13.81	18.70	29.22	29.22	12.65	17.62
	0.47	*0.92*	*0.83*	*0.92*	*0.78*	*1.06*	*1.66*	*1.66*	*0.72*	*1.00*
FR	4.28	12.81	10.30	12.26	9.90	16.30	21.23	21.23	11.72	16.35
	0.26	*0.78*	*0.63*	*0.75*	*0.61*	*1.00*	*1.30*	*1.30*	*0.72*	*1.00*
GB	7.78	14.26	11.03	14.42	11.87	21.73	20.75	20.75	12.42	17.70
	0.44	*0.81*	*0.62*	*0.81*	*0.67*	*1.23*	*1.17*	*1.17*	*0.70*	*1.00*
IT	5.33	12.23	8.56	11.61	9.18	11.76	12.87	12.87	10.36	14.65
	0.36	*0.84*	*0.58*	*0.79*	*0.63*	*0.80*	*0.88*	*0.88*	*0.71*	*1.00*
NL	1.87	5.68	5.31	6.06	4.57	6.68	6.44	6.44	5.84	8.21
	0.23	*0.69*	*0.65*	*0.74*	*0.56*	*0.81*	*0.78*	*0.78*	*0.71*	*1.00*
PL	4.25	4.16	5.54	4.24	3.94	4.44	4.64	4.64	6.22	8.53
	0.50	*0.49*	*0.65*	*0.50*	*0.46*	*0.52*	*0.54*	*0.54*	*0.73*	*1.00*
SE	3.26	3.89	4.14	3.96	3.75	4.90	5.25	5.25	4.58	6.36
	0.51	*0.61*	*0.65*	*0.62*	*0.59*	*0.77*	*0.83*	*0.83*	*0.72*	*1.00*
Average	4.80	9.10	7.80	9.10	7.50	10.80	12.20	12.20	8.60	11.90
	0.40	*0.76*	*0.66*	*0.76*	*0.63*	*0.91*	*1.03*	*1.03*	*0.72*	*1.00*

Notes: See Table 2, but the RMSFEs are calculated using for each quarter the GDP growth prediction in the middle of the prediction period, i.e., one quarter ahead of the GDP release.

Table 10: RMSFEs for 2020Q1–Q4: 1-day ahead-of-release horizon

Economy	Workhorse models					Benchmark models				
	DFM	M	UM	BE	Pooled	ARIMA	AR_{it}	AR_{di}	RM	RW
AT	5.80	8.72	8.36	9.37	7.72	9.64	9.79	9.79	8.84	12.57
	0.46	*0.69*	*0.67*	*0.75*	*0.61*	*0.77*	*0.78*	*0.78*	*0.70*	*1.00*
BE	4.61	9.38	6.61	8.97	7.39	11.07	11.20	11.20	8.82	11.78
	0.39	*0.80*	*0.56*	*0.76*	*0.63*	*0.94*	*0.95*	*0.95*	*0.75*	*1.00*
CH	2.46	5.90	4.34	5.91	4.59	6.33	6.59	6.59	5.91	7.52
	0.33	*0.78*	*0.58*	*0.79*	*0.61*	*0.84*	*0.88*	*0.88*	*0.79*	*1.00*
DE	2.53	5.61	5.48	5.62	4.81	6.54	6.74	6.74	6.56	9.11
	0.28	*0.62*	*0.60*	*0.62*	*0.53*	*0.72*	*0.74*	*0.74*	*0.72*	*1.00*
EA	4.27	9.93	7.74	10.00	7.45	11.50	12.09	12.09	8.92	12.52
	0.34	*0.79*	*0.62*	*0.80*	*0.60*	*0.92*	*0.97*	*0.97*	*0.71*	*1.00*
ES	5.40	15.85	14.08	15.78	12.77	18.24	28.47	28.47	12.56	17.35
	0.31	*0.91*	*0.81*	*0.91*	*0.74*	*1.05*	*1.64*	*1.64*	*0.72*	*1.00*
FR	5.61	12.58	10.30	12.52	9.53	15.46	21.05	21.05	11.72	16.49
	0.34	*0.76*	*0.62*	*0.76*	*0.58*	*0.94*	*1.28*	*1.28*	*0.71*	*1.00*
GB	6.98	14.39	11.54	14.46	11.84	18.82	22.69	22.69	12.37	17.63
	0.40	*0.82*	*0.65*	*0.82*	*0.67*	*1.07*	*1.29*	*1.29*	*0.70*	*1.00*
IT	5.96	12.03	8.35	11.61	8.82	12.20	12.96	12.96	10.39	14.61
	0.41	*0.82*	*0.57*	*0.79*	*0.60*	*0.84*	*0.89*	*0.89*	*0.71*	*1.00*
NL	2.28	5.68	5.30	6.15	4.68	6.78	6.49	6.49	5.79	8.30
	0.27	*0.68*	*0.64*	*0.74*	*0.56*	*0.82*	*0.78*	*0.78*	*0.70*	*1.00*
PL	3.64	4.21	5.50	4.19	3.89	4.56	4.51	4.51	6.22	8.62
	0.42	*0.49*	*0.64*	*0.49*	*0.45*	*0.53*	*0.52*	*0.52*	*0.72*	*1.00*
SE	3.27	3.88	4.15	4.26	3.85	4.77	5.15	5.15	4.19	6.55
	0.50	*0.59*	*0.63*	*0.65*	*0.59*	*0.73*	*0.79*	*0.79*	*0.64*	*1.00*
Average	4.40	9.00	7.60	9.10	7.30	10.50	12.30	12.30	8.50	11.90
	0.37	*0.76*	*0.64*	*0.76*	*0.61*	*0.88*	*1.03*	*1.03*	*0.71*	*1.00*

Notes: See Table 2, but the RMSFEs are calculated using for each quarter the GDP growth prediction at the last day of the prediction period, i.e., one day ahead of the GDP release.

Table 11: RMSFEs for 2021Q1–Q4: 2-quarter ahead-of-release horizon

Economy	Workhorse models					Benchmark models				
	DFM	M	UM	BE	Pooled	ARIMA	AR_{it}	AR_{di}	RM	RW
AT	2.96	3.18	2.60	2.92	2.47	2.16	2.51	2.88	2.24	6.83
	0.43	*0.47*	*0.38*	*0.43*	*0.36*	*0.32*	*0.37*	*0.42*	*0.33*	*1.00*
BE	1.14	1.57	1.33	1.78	1.20	1.02	3.08	2.94	0.98	3.56
	0.32	*0.44*	*0.37*	*0.50*	*0.34*	*0.29*	*0.87*	*0.83*	*0.27*	*1.00*
CH	0.88	0.85	0.92	1.34	0.95	0.92	0.93	0.99	0.88	3.14
	0.28	*0.27*	*0.29*	*0.43*	*0.30*	*0.29*	*0.30*	*0.32*	*0.28*	*1.00*
DE	1.66	1.85	0.81	1.52	1.45	1.42	1.19	1.36	1.43	4.25
	0.39	*0.43*	*0.19*	*0.36*	*0.34*	*0.33*	*0.28*	*0.32*	*0.34*	*1.00*
EA	1.11	2.41	1.63	2.42	1.53	1.14	1.68	2.82	1.20	5.14
	0.22	*0.47*	*0.32*	*0.47*	*0.30*	*0.22*	*0.33*	*0.55*	*0.23*	*1.00*
ES	1.51	3.93	3.02	4.07	3.13	6.07	8.75	33.14	1.74	5.71
	0.27	*0.69*	*0.53*	*0.71*	*0.55*	*1.06*	*1.53*	*5.80*	*0.31*	*1.00*
FR	0.88	2.85	1.83	2.36	1.49	11.57	7.89	24.49	1.05	5.70
	0.15	*0.50*	*0.32*	*0.41*	*0.26*	*2.03*	*1.38*	*4.29*	*0.18*	*1.00*
GB	1.89	3.90	1.69	3.72	2.61	2.52	5.08	1.96	1.97	6.87
	0.28	*0.57*	*0.25*	*0.54*	*0.38*	*0.37*	*0.74*	*0.29*	*0.29*	*1.00*
IT	1.65	3.00	1.61	2.91	2.22	1.55	3.33	7.15	1.66	6.60
	0.25	*0.45*	*0.24*	*0.44*	*0.34*	*0.23*	*0.50*	*1.08*	*0.25*	*1.00*
NL	1.37	1.84	1.64	1.71	1.62	1.55	1.60	1.48	1.48	4.00
	0.34	*0.46*	*0.41*	*0.43*	*0.41*	*0.39*	*0.40*	*0.37*	*0.37*	*1.00*
PL	0.50	0.71	1.33	0.80	0.80	0.73	1.53	0.94	0.82	2.70
	0.18	*0.26*	*0.49*	*0.30*	*0.30*	*0.27*	*0.57*	*0.35*	*0.30*	*1.00*
SE	0.47	1.59	0.98	0.87	0.88	1.52	1.08	0.83	0.85	1.18
	0.40	*1.36*	*0.83*	*0.74*	*0.75*	*1.29*	*0.92*	*0.71*	*0.72*	*1.00*
Average	1.30	2.30	1.60	2.20	1.70	2.70	3.20	6.70	1.40	4.60
	0.28	*0.50*	*0.35*	*0.48*	*0.37*	*0.59*	*0.70*	*1.46*	*0.30*	*1.00*

Notes: See Table 3, but the RMSFEs are calculated using for each quarter the GDP growth prediction at the first day of the prediction period, i.e., two quarters ahead of the GDP release.

Table 12: **RMSFEs for 2021Q1–Q4: 1-quarter ahead-of-release horizon**

Economy	Workhorse models					Benchmark models				
	DFM	M	UM	BE	Pooled	ARIMA	AR_{it}	AR_{di}	RM	RW
AT	1.66	2.51	2.17	2.59	1.91	2.46	2.70	2.70	2.25	3.34
	0.50	*0.75*	*0.65*	*0.78*	*0.57*	*0.74*	*0.81*	*0.81*	*0.67*	*1.00*
BE	2.09	0.95	1.12	0.84	1.17	1.50	2.05	2.05	0.94	0.84
	2.49	*1.13*	*1.34*	*1.00*	*1.39*	*1.79*	*2.44*	*2.44*	*1.11*	*1.00*
CH	0.88	0.89	0.89	0.93	0.81	0.85	0.89	0.89	0.87	1.09
	0.81	*0.81*	*0.82*	*0.86*	*0.75*	*0.78*	*0.82*	*0.82*	*0.80*	*1.00*
DE	0.97	1.47	1.27	1.49	1.28	1.24	1.34	1.34	1.44	1.94
	0.50	*0.76*	*0.65*	*0.76*	*0.66*	*0.64*	*0.69*	*0.69*	*0.74*	*1.00*
EA	1.04	1.11	1.31	1.16	0.94	1.27	1.35	1.35	1.17	1.21
	0.87	*0.92*	*1.08*	*0.96*	*0.78*	*1.06*	*1.12*	*1.12*	*0.97*	*1.00*
ES	2.40	1.14	1.45	1.15	1.48	5.41	12.82	12.82	1.73	1.24
	1.94	*0.92*	*1.17*	*0.93*	*1.19*	*4.36*	*10.34*	*10.34*	*1.40*	*1.00*
FR	2.70	1.16	0.94	1.31	1.35	5.98	9.56	9.56	1.03	1.65
	1.64	*0.70*	*0.57*	*0.80*	*0.82*	*3.63*	*5.81*	*5.81*	*0.63*	*1.00*
GB	2.13	2.42	2.14	2.45	1.93	2.28	2.66	2.66	1.94	3.16
	0.67	*0.76*	*0.68*	*0.77*	*0.61*	*0.72*	*0.84*	*0.84*	*0.61*	*1.00*
IT	2.07	1.31	1.67	1.30	1.33	1.91	2.61	2.61	1.57	1.67
	1.23	*0.78*	*1.00*	*0.78*	*0.79*	*1.14*	*1.56*	*1.56*	*0.94*	*1.00*
NL	1.26	1.15	1.53	1.09	1.22	1.56	1.55	1.55	1.44	1.53
	0.83	*0.75*	*1.00*	*0.72*	*0.80*	*1.02*	*1.01*	*1.01*	*0.94*	*1.00*
PL	0.83	0.85	0.90	0.84	0.70	0.99	1.05	1.05	0.80	0.80
	1.03	*1.06*	*1.12*	*1.05*	*0.88*	*1.24*	*1.31*	*1.31*	*1.00*	*1.00*
SE	0.79	1.10	0.87	0.88	0.85	1.50	0.75	0.75	0.81	0.69
	1.14	*1.59*	*1.26*	*1.27*	*1.23*	*2.18*	*1.08*	*1.08*	*1.17*	*1.00*
Average	1.60	1.30	1.40	1.30	1.20	2.20	3.30	3.30	1.30	1.60
	1.00	*0.81*	*0.87*	*0.81*	*0.75*	*1.38*	*2.06*	*2.06*	*0.81*	*1.00*

Notes: See Table 3, but the RMSFEs are calculated using for each quarter the GDP growth prediction in the middle of the prediction period, i.e., one quarter ahead of the GDP release.

Table 13: RMSFEs for 2021Q1–Q4: 1-day ahead-of-release horizon

Economy	Workhorse models					Benchmark models				
	DFM	M	UM	BE	Pooled	ARIMA	AR_{it}	AR_{di}	RM	RW
AT	1.56	2.38	1.98	2.55	1.89	2.20	2.37	2.37	2.24	3.28
	0.48	*0.73*	*0.60*	*0.78*	*0.57*	*0.67*	*0.72*	*0.72*	*0.68*	*1.00*
BE	2.14	0.98	1.20	0.65	1.22	1.53	2.03	2.03	0.93	0.78
	2.77	*1.27*	*1.55*	*0.84*	*1.57*	*1.97*	*2.62*	*2.62*	*1.20*	*1.00*
CH	1.17	0.88	0.89	0.93	0.91	0.85	0.89	0.89	0.87	1.09
	1.08	*0.80*	*0.81*	*0.86*	*0.84*	*0.78*	*0.82*	*0.82*	*0.80*	*1.00*
DE	1.60	1.39	1.26	1.39	1.35	1.24	1.34	1.34	1.44	1.98
	0.81	*0.71*	*0.64*	*0.70*	*0.69*	*0.63*	*0.68*	*0.68*	*0.73*	*1.00*
EA	2.31	1.11	1.32	1.06	1.12	1.31	1.37	1.37	1.16	1.12
	2.06	*0.99*	*1.18*	*0.95*	*1.00*	*1.17*	*1.22*	*1.22*	*1.04*	*1.00*
ES	0.78	1.21	1.53	1.20	0.90	6.13	13.21	13.21	1.74	1.33
	0.59	*0.91*	*1.15*	*0.91*	*0.68*	*4.62*	*9.96*	*9.96*	*1.31*	*1.00*
FR	2.89	1.20	0.95	1.27	1.29	5.86	9.41	9.41	1.03	1.79
	1.62	*0.67*	*0.53*	*0.71*	*0.72*	*3.29*	*5.27*	*5.27*	*0.58*	*1.00*
GB	1.93	2.61	2.24	2.65	1.96	2.29	2.99	2.99	1.92	3.38
	0.57	*0.77*	*0.66*	*0.78*	*0.58*	*0.68*	*0.88*	*0.88*	*0.57*	*1.00*
IT	1.77	1.20	1.69	1.22	1.25	1.96	2.63	2.63	1.57	1.50
	1.18	*0.80*	*1.13*	*0.82*	*0.84*	*1.31*	*1.76*	*1.76*	*1.05*	*1.00*
NL	1.10	1.15	1.55	1.04	1.15	1.57	1.57	1.57	1.43	1.81
	0.61	*0.63*	*0.86*	*0.58*	*0.64*	*0.87*	*0.87*	*0.87*	*0.79*	*1.00*
PL	0.87	0.85	0.89	0.88	0.87	1.00	1.04	1.04	0.80	0.80
	1.08	*1.06*	*1.10*	*1.09*	*1.08*	*1.24*	*1.29*	*1.29*	*0.99*	*1.00*
SE	2.90	0.81	0.80	0.61	1.22	0.61	0.71	0.71	0.80	0.83
	3.49	*0.97*	*0.96*	*0.73*	*1.46*	*0.73*	*0.86*	*0.86*	*0.96*	*1.00*
Average	1.80	1.30	1.40	1.30	1.30	2.20	3.30	3.30	1.30	1.60
	1.12	*0.81*	*0.87*	*0.81*	*0.81*	*1.38*	*2.06*	*2.06*	*0.81*	*1.00*

Notes: See Table 3, but the RMSFEs are calculated using for each quarter the GDP growth prediction at the last day of the prediction period, i.e., one day ahead of the GDP release.

Table 14: RMSFEs using latest GDP releases

Economy	2020Q1–2020Q4 M	2020Q1–2020Q4 UM	2020Q1–2020Q4 BE	2021Q1–2021Q4 M	2021Q1–2021Q4 UM	2021Q1–2021Q4 BE
AT	8.05	7.98	9.79	2.55	2.70	3.22
	1.10	*1.05*	*1.27*	*0.82*	*1.40*	*1.08*
BE	9.35	9.06	11.76	2.89	2.61	1.30
	1.20	*1.35*	*1.55*	*2.88*	*1.70*	*1.17*
CH	5.20	4.99	6.41	0.64	0.59	0.91
	1.10	*1.23*	*1.39*	*0.73*	*0.72*	*0.78*
DE	6.06	6.08	6.19	0.81	0.87	1.01
	1.13	*1.16*	*1.18*	*0.65*	*1.21*	*0.91*
EA	9.28	9.09	12.06	2.30	1.75	1.27
	1.16	*1.26*	*1.48*	*1.44*	*1.28*	*0.79*
ES	14.02	13.48	20.17	12.72	11.13	19.99
	1.17	*1.14*	*1.67*	*5.21*	*5.16*	*7.89*
FR	11.83	11.67	13.55	7.76	6.94	14.69
	1.11	*1.16*	*1.28*	*3.68*	*5.45*	*8.18*
GB	15.41	15.32	24.05	3.13	2.93	2.61
	1.34	*1.37*	*2.08*	*1.03*	*1.32*	*0.89*
IT	10.11	10.11	13.05	4.09	3.34	2.76
	1.04	*1.22*	*1.37*	*2.35*	*2.04*	*1.63*
NL	5.14	5.18	5.64	1.33	1.39	1.50
	1.17	*1.21*	*1.20*	*0.84*	*0.91*	*1.03*
PL	4.44	4.27	4.24	1.14	1.28	1.18
	1.13	*0.95*	*1.09*	*1.00*	*0.86*	*1.02*
SE	4.47	4.48	4.44	0.71	0.68	0.86
	1.06	*0.98*	*1.01*	*0.49*	*0.63*	*0.88*
Average	8.60	8.50	10.90	3.30	3.00	4.30
	1.15	1.19	1.45	1.85	2.03	2.51

Notes: See Table 5, but the RMSFEs are calculated using the latest available GDP release (as of 05/12/2022) for each quarter instead of using always the first GDP release.

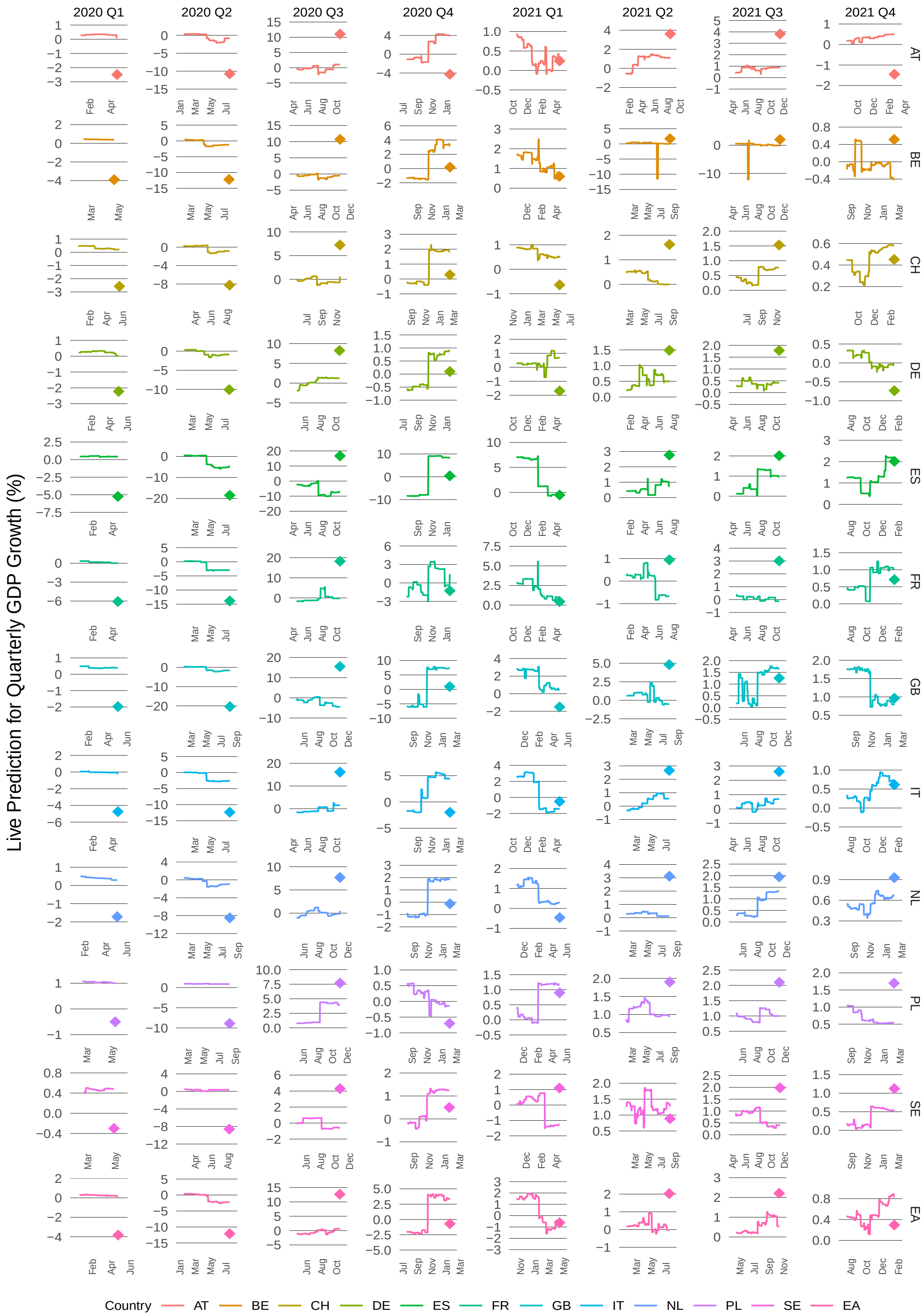

Figure 8: GDP growth live predictions economy by economy. The figure shows, separately for each economy and quarter in 2020–2021, the development of the daily pooled predictions for quarterly GDP growth (lines) together with the respective GDP growth realizations (dots) according to their first release. The predictions for GDP growth of quarter t start on the first day after the release of GDP of $t-2$ and end on the last day before the release of GDP of t. The pooled predictions are equally weighted averages of the predictions resulting from the workhorse models (see Appendix C). See Figure 4 for the country codes.

Figure 9: GDP growth prediction errors for 2020Q1–2021Q4. The figure shows GDP growth prediction errors from different forecasting models separately for each quarter from 2020Q1–2021Q4. The prediction errors are calculated as the difference between the average of the daily predictions for each target period $t \in \{2020Q1, \ldots, 2021Q4\}$ minus the actual GDP growth realization (first release) for t.