Lessons from Nowcasting GDP across the World*

Danilo Cascaldi-Garcia† Matteo Luciani‡ Michele Modugno§

Federal Reserve Board *Amazon.com* *Federal Reserve Board*

December 2023

Abstract

In economics, we need to *forecast the present* because reliable and comprehensive measures of the state of the economy are released with a substantial delay and considerable measurement error. Nowcasting exploits timely data to obtain early estimates of the state of the economy and updates these estimates continuously as new macroeconomic data are released. In this chapter, we describe how the framework used to nowcast GDP has evolved and is applied worldwide.

Keywords: Dynamic factor models, forecasting, nowcasting

JEL codes: C33, C53, E37

*We are grateful to the editors Michael Clements and Ana Beatriz Galvão, one anonymous referee, and stimulating discussions with Domenico Giannone. Any errors are our responsibility. This is a draft chapter. The final version will be available in the Handbook on Macroeconomic Forecasting edited by Michael Clements and Ana Beatriz Galvão, Edward Elgar Publishing Ltd. The material cannot be used for any other purpose without further permission of the publisher, and is for private use only.

†Federal Reserve Board, International Finance Division, Washington DC 20551, USA; E-mail address: danilo.cascaldi-garcia@frb.gov.

‡Amazon.com; E-mail address: luciani.matteo@gmail.com.

§Federal Reserve Board, Financial Stability Division, Washington DC 20551, USA; E-mail address: michele.modugno@frb.gov.

1 Introduction

The term "nowcasting" is a contraction of the words "now" and "forecasting," and it refers to the prediction of the very recent past, the present, and the very near future. This word has been used in meteorology for a long time, and it was introduced into economics by Giannone et al. (2008).

Obtaining a reliable measure of the state of the economy is pivotal to making policy and business decisions. Every day, policy institutions, market analysts, and financial and non financial corporations parse troves of economic data released by statistical agencies, private and public surveys, and other sources to assess the health of the economy. Based on these data, they nowcast the current state of the economy; that is, they create a narrative about where the economy is and where it is headed. The difficulty comes in separating meaningful economic signals from the noise.

In their seminal paper, Giannone et al. (2008) designed a nowcasting model to formalize key features of how market participants and policymakers read data in real time. A few years later, Bańbura and Modugno (2014) enriched the state-of-the-art nowcasting model with the appropriate tool to monitor how multiple and asynchronous data releases change the assessment of the state of the economy. In other words, with this tool, the nowcasting model can mimic the behavior of market participants who revise their assessments whenever a new data release differs from their expectations.

Before Giannone et al. (2008) introduced their nowcasting model, real-time monitoring of macroeconomic conditions was more of an art than a science. The common practice was to use a set of heuristic models and a good dose of judgment to make predictions about the state of the economy. However, judgmental and simplified heuristic procedures are exposed to internal inconsistencies, with the constant risk of putting too much weight on outdated signals or on timely but unreliable releases. In addition, this procedure cannot help interpreting the information content of each data release in a systematic way. Moreover, assessing the state of the economy in real time involves analyzing a large amount of complex information that is continuously released, often with multiple data releases in a single day. Lastly, processes that are not scientific and do not use formal methods cannot be evaluated ex post. In conclusion, updating the assessment of the economy in real time using a procedure that is not entirely automated is costly, risky, and not scalable.

The challenge embraced by the literature was to design an entirely automated platform capable of tracking the state of the economy without relying on any judgment or subjective prior information. Hence, the nowcasting literature developed a formal and internally coherent methodology replicating the experts' judgmental process. To perform this task, an arsenal of tools and methods in econometrics, statistics, and data analysis has been deployed, building upon the nascent developments and insights in big data analytics and taking advantage of improvements in scientific computing, data handling, and visualization. Compared to judgmental predictions, the advantage of having such a platform is that it delivers a transparent monitoring of the economy through a robust methodology and provides a coherent analysis of the links between macroeconomic and cyclical developments.

The first nowcasting model was a dynamic factor model (DFM) equipped with efficient filtering techniques. This model exploits two essential and robust features of business cycle fluctuations. First, macroeconomic data strongly co-move, so a few common factors summarize their dynamics well—in this context, the common factors are typically associated with the unobserved state of the economy. Second, historically, economic booms and busts persist for a considerable period of time, so the past dynamics of such factors should be informative to understand where we are and where we are heading in the near future. Hence, a DFM provides a parsimonious yet suitable representation for the large set of macroeconomic time series.

Formally, DFMs can be written in a state-space form or as a system of two types of equations: measurement equations linking observed series to the unobserved factors, and transition equations describing the dynamics of the unobserved factors. The state-space representation allows the use of Kalman filtering techniques to obtain projections for the observed variables and the unobserved factors. Most importantly, given an estimate of the parameters, the Kalman filter can easily cope with challenging features of the nowcasting information set, such as data observed at different frequencies and with missing data. These can appear either at the end of the sample due to asynchronous data releases (ragged edges) or at the beginning of the sample due to only a recent collection of some data sources.

The use of DFMs, coupled with the Kalman filter, has a long tradition in econometrics. However, for a long time, it was considered infeasible for high dimensional data, as they require estimating a large number of parameters. Doz et al. (2012) challenged this view and, by studying the asymptotic properties of the maximum likelihood estimator when the complexity of the model and the sample size increase, showed that these models are viable for analyzing big datasets. They also refined the estimation procedure to make the computation scalable to high-dimensional problems. However, their procedure was not directly suitable for nowcasting with an information set characterized by data with mixed frequencies and missing data. Bańbura and Modugno (2014) tackled this problem by modifying Doz et al.'s maximum likelihood procedure to efficiently use all the information embedded in incomplete datasets. Subsequently, D'Agostino et al. (2016) provided an alternative solution based on Bayesian inference.

Taking stock of the accumulated experience has shown that the model provides predictions whose accuracy equals or exceeds the accuracy of expert judgment predictions. This performance is why, today, almost every central bank in the world has a nowcasting model. For example, various Federal Reserve branches (Atlanta, Cleveland, and New York) periodically publish their nowcasting models' results, and Bloomberg makes available nowcasts through its platform. All these estimates are widely followed and discussed by the press and analysts at hedge funds, investment banks, and large corporations. Nowcasting has also become an active area of academic research. In a survey article, Stock and Watson (2017) included nowcasting among the 10 most important innovations in time-series econometrics over the previous 20 years, and many papers applying the nowcasting framework to economies all around the world have been published.[1]

In this chapter, we focus on how DFMs are constructed and estimated for nowcasting gross domestic product (GDP).[2] We will briefly discuss the limits of how real-time economic monitoring was conducted before nowcasting and how this framework has overcome those limitations, particularly interpretability. After a brief description of alternative models that have been proposed for real-time monitoring of the economy, we will also discuss how estimation algorithms for DFMs have evolved to efficiently use all the information content embedded in a dataset characterized by a large cross-section of data, with mixed-frequency and mismatched time span coverage. We will then describe the data selection process and conclude with an empirical section that will highlight the performance of some of the nowcasting applications during the Great Financial Crisis and the onset of the COVID-19 pandemic across multiple countries.

2 Models and their interpretability

Key data that describe the current state of the economy are available with a significant delay, particularly those collected quarterly, with GDP being a prominent example. For instance, limiting our attention only to G7 countries, the delay between the publication of the first official estimate of GDP and the end of the reference quarter is approximately four weeks in the United States and the United Kingdom, six weeks in Japan, and eight

[1]The economies include Belgium (de Antonio Liedo, 2015), Brazil (Bragoli et al., 2015), BRICs plus Mexico (Dahlhaus et al., 2017), Canada (Bragoli and Modugno, 2017), China (Yiu and Chow, 2010 and Giannone et al., 2013) ,the Czech Republic (Arnostova et al., 2011; Rusnák, 2016), Ecuador (González-Astudillo and Baquero, 2019), Euro Area (Angelini et al., 2010, Camacho and Perez-Quiros, 2010, Angelini et al., 2011, Bańbura and Rünstler, 2011, Bańbura et al., 2011, Bańbura and Modugno, 2014 Carriero et al., 2019, and Cascaldi-Garcia et al., 2023), European countries (Rünstler et al., 2009 and Jansen et al., 2016), France (Barhoumi et al., 2010 and Bessec and Doz, 2014), Germany (Marcellino and Schumacher, 2010; Andreini et al., 2023), India Bragoli and Fosten, 2018, Indonesia (Luciani et al., 2018), Ireland (D'Agostino et al., 2012), Japan (Bragoli, 2017; Hayashi and Tachi, 2023), Mexico Caruso, 2018, New Zealand (Matheson, 2010), Norway (Aastveit and Trovik, 2012, and Luciani and Ricci, 2014), Switzerland (Siliverstovs, 2012), Turkey (Modugno et al., 2016), the United Kingdom (Anesti et al., 2018), United States (Giannone et al., 2008; Lahiri and Monokroussos, 2013; Bańbura et al., 2013; Bok et al., 2018; Antolin-Diaz et al., 2020). Surveys of the literature on nowcasting are provided by Bańbura et al. (2011, 2013); Luciani (2017); Bok et al. (2018).

[2]Nowcasting models have also been applied to variables other than GDP, such as, among others, inflation (Modugno, 2013) and trade (D'Agostino et al., 2017).

weeks in Canada. France, Germany, and Italy were characterized by a six-week delay until October 2015, four weeks after that.

However, plenty of information (cor-)related to GDP is published at higher frequencies and earlier than the variable of interest: information about the labor market, industrial production, trade, sales, housing, or surveys about the state of the economy. These are frequently available at monthly frequencies and are released before the current quarter figure of GDP.

The first foundational principle of nowcasting was to create a framework that could exploit a large and timely information set (cor-)related to the target variable to generate its early estimates. However, using timely information from various sources has a number of implications regarding the features of the information set: 1) it may be composed of data with different frequencies; 2) data are released in a non synchronous manner and with different degrees of delay, creating the so-called "ragged" or "jagged" edge dataset; 3) data may have a different time availability.

Nowcasting is not the first framework deployed that produces estimates of the state of the economy by handling information sets characterized by mixed frequency, unbalanced, and "ragged edge" data. Central banks and financial market practitioners have long relied on frameworks mainly based on bridge equations and model averaging. However, those frameworks do not provide tools to interpret how new releases of the input variables affect the change in the early estimate of the target variable. This results from their "partial" model nature, as they are not set up to isolate the unpredictable component in the newest data release, conditional on the existing information set. Another drawback of these partial solutions is the need to specify a different model (and, consequently, estimate different sets of parameters) for data vintages with different "ragged edge" structures. As such, it is not possible to interpret the impact of new data releases, which is at odds with the second foundational principle of nowcasting of a framework that links and interprets how new releases of the input variables revise the model estimate of the target variable. This objective can only be reached by using multivariate econometric frameworks that allow isolating from each new data release the innovation that is orthogonal to the entire available information set and linking how this innovation changes the estimate of the target variable—i.e., a multivariate unique framework that can produce a forecast for each variable in the system.

The natural candidates that display this characteristic among the econometric frameworks typically used in macro-econometric analysis are DFMs and vector autoregressive models (VARs). However, when the nowcasting literature started, the estimation of VARs on a large number of variables was thought unfeasible, which restricted the set of available options to the DFMs. In recent years, advances in estimation algorithms for VARs have overcome the curse of dimensionality and made it feasible to estimate these models on large datasets, thus making them a palatable solution for nowcasting. We refer to Cimadomo et al. (2022) for an exhaustive exposition of how VARs can be used for nowcasting.

In the remainder of this section, we start by describing bridge equations. We then take a deep dive into DFMs, exploring different specifications and adaptations to the kind of data explored. Next, we explain how news can be extracted from "joint models." The last subsection is dedicated to a quick overview of alternative models proposed for short-term forecasting but that still cannot interpret the evolution of the assessment of the current conditions. Therefore, we group them in the "partial" model family with bridge equations.

Before starting the description of the models, let us set the basic notation we will use throughout the paper: $y^f_{t,n}$ is variable n, released at frequency f, which describes the value of that variable for period t, where, by convention, t indicates the last day of the reference period.[3]

2.1 Bridge equations

In this type of model, the nowcast and forecasts of $y^q_{t,gdp}$ are obtained via the following regression:

$$y^q_{t,gdp} = \alpha + \beta \overline{y}^q_{t,n} + e_t, \tag{1}$$

where $\overline{y}^q_{t,n}$ is the aggregation of the predictor $y^f_{t,n}$, which can be available at a frequency f higher than the target variable and therefore needs to be aggregated to match the frequency and units of the target variable.[4][5] Hence, the mixed-frequency problem is solved by temporal aggregation of the predictors to the lower frequency. To handle the ragged edge, bridge models resort to auxiliary models, such as autoregressive moving average (ARMA) or VAR, to forecast $y_{t,n}$ and close the target period of interest. This was the "traditional" tool popularly employed at central banks to obtain early estimates of GDP or its components, and the predictors were usually monthly (see, e.g., Kitchen and Monaco, 2003; Parigi and Golinelli, 2007, Parigi and Schlitzer, 1995; and Baffigi et al., 2004).

Equation (1) is typically estimated by ordinary least squares (OLS) and can be further extended to include more predictors or the lags of the dependent variable. If the information set is large, forecast combination is often an alternative (Kitchen and Monaco, 2003; Diron, 2008; Angelini et al., 2011; Rünstler et al., 2009). Bridge equations can also be

[3] E.g., the monthly industrial production of January 2022 will be represented as $y^m_{01/31/2022,ip}$, and the quarterly GDP of 2022:Q1 will be represented as $y^q_{03/31/2022,gdp}$.

[4] From now on, we will use the convention that the time notation will indicate the final highest-frequency finite fraction of the reference period. For example, if the variable of interest is quarterly and the highest frequency is monthly, τ will indicate the last month in the quarter of interest.

[5] The aggregation used in bridge equations follows the same logic as in subsection 2.3 for factor models.

combined in a so-called bottom-up approach where one obtains early estimates of GDP by aggregating the early estimates of its components, exploiting national accounts' identities (see Hahn and Skudelny, 2008; Drechsel and Scheufele, 2012; Baffigi et al., 2004).[6]

2.2 Dynamic factor models

The typical DFM used for nowcasting decomposes every economic indicator into at least two parts: (i) factors common to each indicator in the information set, and (ii) an indicator-specific idiosyncratic component. The main identification assumption behind these models is that the common factors are the only components that explain the co-movement among the economic indicators, while the idiosyncratic components capture indicator-specific variation. This identification assumption is formalized by imposing that the idiosyncratic components are orthogonal to each other and to the common factors at each lead and lag.[7]

More precisely, we specify the model as

$$\boldsymbol{y}_t = \boldsymbol{\Lambda} \cdot \boldsymbol{F}_t + \boldsymbol{e}_t, \tag{2}$$

$$\boldsymbol{F}_t = \boldsymbol{A} \cdot \boldsymbol{F}_{t-1} + \boldsymbol{u}_t, \quad \boldsymbol{u}_t \sim \text{i.i.d. } N(\boldsymbol{0}, \boldsymbol{Q}), \tag{3}$$

$$\boldsymbol{e}_t = \boldsymbol{D} \cdot \boldsymbol{e}_{t-1} + \boldsymbol{v}_t, \quad \boldsymbol{v}_t \sim \text{i.i.d. } N(\boldsymbol{0}, \boldsymbol{R}), \tag{4}$$

where $\boldsymbol{y}_t$ is a vector of $(n \times 1)$ standardized economic indicators; $\boldsymbol{F}_t$ is a vector of $(r \times 1)$ common factors (with $r<n$); $\boldsymbol{e}_t$ are the $(n \times 1)$ idiosyncratic components; $\boldsymbol{\Lambda}$ is a matrix of $(n \times r)$ loadings of the economic indicators on the factors; $\boldsymbol{A}$ is the $(r \times r)$ auto-regressive matrix of the factors (in companion form); $\boldsymbol{D}$ is the $(n \times n)$ diagonal auto-regressive matrix of the idiosyncratic components; $\boldsymbol{Q}$ is the variance-covariance matrix of the common factors; and $\boldsymbol{R}$ is the diagonal variance covariance matrix. $\boldsymbol{D}$ and $\boldsymbol{R}$ are assumed to be diagonal to preserve the cross-orthogonality condition among the idiosyncratic components.

In order to make this model suitable for the Kalman filter, which is a central ingredient for both the estimation of the model's parameter (see section 3) and the production of forecasts, we need its state-space representation. Equations (2) to (4) can therefore be

[6] Note that the model of Giannone et al. (2008) can also be interpreted as "bridging with factors," as the factors extracted with the Kalman filter were plugged into an equation similar to (1) to obtain the nowcasts. The Kalman filter allowed using the ragged edge part of the information set (not data at different frequencies) to update the factors' estimate but not the model's parameter. Once Bańbura and Modugno (2014) and D'Agostino et al. (2016) showed how to efficiently use all the information set for estimating both factors and parameters, "bridging with factors" was outmoded.

[7] The assumption that the idiosyncratic components are not cross-sectionally correlated is a simplified assumption used for exposition purposes. Indeed, in a large macroeconomic dataset, it is most likely the case that these idiosyncratic components are cross-correlated. If those correlations are small, the model can be estimated without additional problems (Doz et al., 2012; Barigozzi and Luciani, 2022).

written as

$$\boldsymbol{y_t} = \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{I_{n\times n}} \end{bmatrix} \begin{bmatrix} \boldsymbol{F_t} \\ \boldsymbol{e_t} \end{bmatrix} \tag{5}$$

$$\begin{bmatrix} \boldsymbol{F_t} \\ \boldsymbol{e_t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0_{r\times n}} \\ \boldsymbol{0_{n\times r}} & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{F_{t-1}} \\ \boldsymbol{e_{t-1}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{u_t} \\ \boldsymbol{v_t} \end{bmatrix} \tag{6}$$

where

$$\begin{bmatrix} \boldsymbol{u_t} \\ \boldsymbol{v_t} \end{bmatrix} \sim \text{i.i.d. } N\left(\begin{bmatrix} \boldsymbol{0_{r\times 1}} \\ \boldsymbol{0_{n\times 1}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{R} & \boldsymbol{0_{n\times r}} \\ \boldsymbol{0_{r\times n}} & \boldsymbol{Q} \end{bmatrix}\right) \tag{7}$$

Most of the nowcasting applications have been successful, in terms of nowcasting accuracy, with the simplest specification of this model—i.e., assuming the existence of one common factor (r=*1*).[8] However, there have been applications where more complicated specifications of the DFM have been deployed due to the nature of the data, the problem under scrutiny, or the need to understand links between specific groups of variables and the GDP.

One example is Cascaldi-Garcia et al. (2023). In this paper, the authors formalize how to monitor the euro-area economy following a multi-country approach inspired by the example of market participants, who track both euro-area aggregate data and largest country-specific data, and policymakers, who build euro-area forecasts from projections for individual countries. To do so, they assume a block structure in which each economic indicator loads only on its economy-specific factor, as in

$$\begin{bmatrix} \boldsymbol{y}_t^{ea} \\ \boldsymbol{y}_t^{fr} \\ \boldsymbol{y}_t^{ge} \\ \boldsymbol{y}_t^{it} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}^{ea} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}^{fr} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Lambda}^{ge} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Lambda}^{it} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_t^{ea} \\ \boldsymbol{f}_t^{fr} \\ \boldsymbol{f}_t^{ge} \\ \boldsymbol{f}_t^{it} \\ \boldsymbol{e}_t^{ea} \\ \boldsymbol{e}_t^{fr} \\ \boldsymbol{e}_t^{ge} \\ \boldsymbol{e}_t^{it} \end{bmatrix}, \tag{8}$$

where the data $\boldsymbol{y_t}$ is partitioned into indicators from the euro-area aggregate ($\boldsymbol{y}_t^{ea}$), Germany ($\boldsymbol{y}_t^{ge}$), France ($\boldsymbol{y}_t^{fr}$), and Italy ($\boldsymbol{y}_t^{it}$) that load on one factor per economy, respectively: euro-area aggregate ($\boldsymbol{f}_t^{ea}$), Germany ($\boldsymbol{f}_t^{ge}$), France ($\boldsymbol{f}_t^{fr}$), and Italy ($\boldsymbol{f}_t^{it}$). Moreover, each variable loads also its own idiosyncratic component, included in vectors $\boldsymbol{e}_t^{ea}$, $\boldsymbol{e}_t^{ge}$, $\boldsymbol{e}_t^{fr}$, and $\boldsymbol{e}_t^{it}$. In this specification, each of the matrices Λ^i, $\boldsymbol{0}$, and $\boldsymbol{I}$ is of dimension $n^i \times 1$, with n^i that differs according to the geographical area $i = ea, ge, fr, it$.

[8] Usually, some restrictions on the loading matrix are imposed in order to coherently model the relation among variables published at different frequencies, as explained in detail in section 2.3.

Another example is Bok et al. (2018), where the authors specify a model with a global factor, which affects all the variables, and a few local factors, which affect only blocks of variables. The scope of the authors is to control for idiosyncrasies in particular subgroups of series. Specifically, to model the local correlations in survey data, they include a soft block, loading only variables representing economic agents' perceptions and sentiments. In a similar vein, they add two more factors, one for the block of real variables and one for labor variables.[9]

2.3 Mixed-frequency and time aggregations

Most of the nowcasting models for GDP have been developed for datasets that contain time series published at monthly and quarterly frequencies. Therefore, in equation (2), $\boldsymbol{y_t} = [\boldsymbol{y_t^q}; \boldsymbol{y_t^m}]$. If those data are seasonally adjusted but are not stationary in mean, monthly and quarterly data are transformed into month-on-month and quarter-on-quarter growth rates. If they are stationary, they are transformed into month-on-month or quarter-on-quarter differences.[10]

To construct a model that explicitly considers the different units of measure within a mixed-frequency dataset, the general strategy has been to assume that the low-frequency variables have a partially observed counterpart with the highest frequency among those included in the dataset.

To be more concrete, let's consider a dataset with several monthly variables and the quarterly GDP. The latter is treated as a partially observed monthly variable in which the quarterly release is assigned to the third month of the respective quarter. This partially observed variable is assumed to be an aggregation of an unobserved monthly growth rate of GDP ($y^m_{t,u}$, whose log-level is $Y^m_{t,u}$ that admits the same factor model representation as the other monthly variables:

$$y^m_{t,u} = \lambda_{gdp} f_t + e_{t,gdp}, \tag{9}$$

where

$$e_{t,gdp} = \rho_{gdp} e_{t-1,gdp} + v_{t,gdp}. \tag{10}$$

To link $y^m_{t,u}$ with the observed GDP growth rate $y^q_{t,gdp}$, let us start considering this growth rate as the difference of the log-levels of the quarterly GDP Y^q_t and then use the

[9]It is also possible to determine the number of factors using statistical tests, which depend on the estimation technique adopted; see Coroneo et al. (2016) for maximum likelihood and Bai and Ng (2002) for principal components.

[10]Surveys are usually included in levels, given that they already express changes. For example, questionnaires behind most of the surveys used in nowcasting models ask how the current conditions (about confidence, consumption, business, etc.) compare to the previous month/quarter.

following triangular aggregation:

$$
\begin{aligned}
y^q_{t,gdp} &= Y^q_{t,gdp} - Y^q_{t-3,gdp} = (1 - L^3)Y^q_{t,gdp} \\
&\approx (1 - L^3)(1 + L + L^2)Y^m_{t,u} = (1 + L + L^2 - L^3 - L^4 - L^5)Y^m_{t,u} \\
&= (1 - L + 2L - 2L^2 + 3L^2 - 3L^3 + 2L^3 - 2L^4 + L^4 - L^5)Y^m_{t,u} \\
&= y^m_{t,u} + 2y^m_{t-1,u} + 3y^m_{t-2,u} + 2y^m_{t-3,u} + y^m_{t-4,u}.
\end{aligned}
$$

Therefore, the quarterly GDP $y^q_{t,gdp}$ can be written as

$$
\begin{aligned}
y^q_{t,gdp} = \lambda_{gdp}(f_t + 2f_{t-1} + 3f_{t-2} + 2f_{t-3} + f_{t-4}) + ... \\
+ e_{t,gdp} + 2e_{t-1,gdp} + 3e_{t-2,gdp} + 2e_{t-3,gdp} + e_{t-4,gdp}
\end{aligned}
\tag{11}
$$

and easily cast in the state-space form

$$
\begin{bmatrix} y^q_{t,gdp} \\ \mathbf{y^m_t} \end{bmatrix} = \begin{bmatrix} \lambda_{gdp} & 2\lambda_{gdp} & 3\lambda_{gdp} & 2\lambda_{gdp} & \lambda_{gdp} & 1 & 2 & 3 & 2 & 1 & \mathbf{0'} \\ \Lambda_m & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \underline{\mathbf{0}} \end{bmatrix} \begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ e_{t,gdp} \\ e_{t-1,gdp} \\ e_{t-2} \\ e_{t-3,gdp} \\ e_{t-4,gdp} \\ \mathbf{e_{t,m}} \end{bmatrix} \tag{12}
$$

However, for some countries, due to the lack of reliable seasonal adjustment techniques, both monthly and quarterly data are published only as year-on-year growth rates. This is the case of China, for which Giannone et al. (2013) propose a time aggregation that takes into account the nature of this data, also adopted by Modugno et al. (2016), Dahlhaus et al. (2017), Bragoli and Fosten (2018), and Barcelona et al. (2022).

Let us again assume that GDP level data for a given quarter is the sum of monthly unobserved contributions and let $y^{m_y}_{t,u}$ denote the unobserved monthly year-on-year GDP growth rate. We assume that $y^{m_y}_{t,u}$ admits the same factor structure of the other year-on-year monthly variables in the dataset, similarly to equations (9) and (10).

The monthly unobserved year-on-year growth rate can then be linked to a partially observed (at every third month of the quarter) quarterly year-on-year growth rate ($y^{q_y}_{t,gdp}$)

using the following formula:

$$
\begin{aligned}
y_{t,gdp}^{q_y} &= Y_{t,gdp}^{q} - Y_{t-12,gdp}^{q} = (1-L^{12})Y_{t,gdp}^{q} \\
&\approx (1-L^{12})(1+L+L^2)Y_{t,u}^{m} = (1+L+L^2)y_{t,u}^{m_y} \\
&= y_{t,u}^{m_y} + y_{t-1,u}^{m_y} + y_{t-2,u}^{m_y},
\end{aligned} \tag{13}
$$

which implies that the quarterly variables are required to load equally on the current and lagged values of the unobserved monthly factor.

Using a similar logic but with some more cumbersome algebra, weekly and daily data can also be included in the dataset for a nowcasting model. Modugno (2013) and Bańbura et al. (2013) show how to modify the state space by including aggregator variables that link flow and stock variables coherently at frequencies from daily to quarterly.

2.4 News

A critical ingredient of a nowcasting model is to interpret how new data releases of indicators included in the information set revise the estimate of the variable of interest. In order to do so, we first need to extract the unpredictable component of each new data release given the available information set—i.e., the innovation, and then the contribution of this innovation to the forecast revision, the so-called "news."

To extract the innovation, we need a framework that jointly models all the variables in the information set. Such a model allows us to compute expectations for each new data release and isolate the innovation, which is the difference between the actual release and its expectation. The innovation can then be linked to the variable of interest through weights that depend on the model parameters, and the product of those innovations and their corresponding weights is news. Having model-based news for all variables allows the obtainment of the revision of the GDP nowcast as the weighted sum of the news. Computing the news is key for understanding the changes in the model assessment of current economic activity over time and helping evaluate the significance of each data publication.

Bańbura and Modugno (2014) explain these ideas more formally. For the sake of simplicity, in what follows, we abstract from data revision and parameter re-estimation, so the new information we consider is only due to new data releases.[11] Let us define $y_{\tau,gdp}^{q}$ as our target variable, e.g., real GDP quarterly growth for a given quarter, which we attribute to the last day of the quarter τ. For each data vintage Ω_v, available in day v, we can produce an estimate of our variable of interest $\mathbb{E}[y_{\tau,gdp}^{q}|\Omega_v]$. The difference between Ω_v and Ω_{v+1} is the new data released between v and $v+1$. For simplicity, let

[11] Hayashi and Tachi (2021) extend the revision analysis by providing a method for breaking down the decomposition of nowcast changes into the new-observations, data-revisions, and parameter-revisions contribution from each individual indicator variables.

$y^m_{t,n}$ and $y^m_{s,m}$ be the only available data, relative to variable n and m and attributed to day t and s, respectively, which have been released between v and $v+1$.

Formally we have

$$\Omega_v \subset \Omega_{v+1} \qquad \text{and} \qquad \Omega_{v+1} \backslash \Omega_v = \{y^m_{t,n}, y^m_{s,m}\}. \tag{14}$$

Hence the information set is "expanding."

With the new releases, and therefore the new information set Ω_{v+1}, a new estimate $\mathbb{E}\Big[y_{\tau,gdp}|\Omega_{v+1}\Big]$ can be generated. Using equation (14) and the properties of conditional expectations as an orthogonal projection operator, the following decomposition holds:

$$\underbrace{\mathbb{E}\Big[y^q_{\tau,gdp}|\Omega_{v+1}\Big]}_{\text{new forecast}} = \underbrace{\mathbb{E}\Big[y^q_{\tau,gdp}|\Omega_v\Big]}_{\text{old forecast}} + \underbrace{\mathbb{E}\Big[y^q_{\tau,gdp}|I_{v+1}\Big]}_{\text{revision}},$$

where

$$I_{v+1} \quad = \quad \begin{bmatrix} y^m_{t,n} - \mathbb{E}\left[y^m_{t,n}|\Omega_v\right] \\ y^m_{s,m} - \mathbb{E}\left[y^m_{s,m}|\Omega_v\right] \end{bmatrix}. \tag{15}$$

I_{v+1} is the part of the releases $y^m_{t,n}$ and $y^m_{s,m}$ that was unpredictable with the information contained in Ω_v (given a specific model), or, more formally, $I_{v+1} \perp \Omega_v$. This is the reason why I_{v+1} is labeled *news*, as it is the new information content available in Ω_{v+1} with respect to Ω_v. Note that it is the news and not the release itself that leads to nowcast revisions. In particular, if the new values in Ω_{v+1} are exactly as predicted, given the information in Ω_v (in other words, "there is no news"), the nowcast will not be revised.

We can further develop the expression for the revision, or the difference between the new and the old nowcast, as

$$\mathbb{E}\left[y^q_{\tau,gdp}|I_{v+1}\right] = \mathbb{E}\left[y^q_{\tau,gdp}I'_{v+1}\right]\mathbb{E}\left[I_{v+1}I'_{v+1}\right]^{-1} I_{v+1}\,. \tag{16}$$

Given the model described in equations (2) to (4), assuming that there is only one factor and abstracting from parameter uncertainty, equation (16) can be decomposed as

$$\mathbb{E}\left[y^q_{\tau,gdp}I'_{v+1}\right] \quad = \quad \begin{bmatrix} \Lambda'_{gdp}\mathbb{E}\Big[\left(\boldsymbol{f_\tau} - \mathbb{E}\left[\boldsymbol{f_\tau}|\Omega_v\right]\right)\left(f_t - \mathbb{E}\left[f_t|\Omega_v\right]\right)\Big]\lambda_n \\ \Lambda'_{gdp}\mathbb{E}\Big[\left(\boldsymbol{f_\tau} - \mathbb{E}\left[\boldsymbol{f_\tau}|\Omega_v\right]\right)\left(f_s - \mathbb{E}\left[f_s|\Omega_v\right]\right)\Big]\lambda_m \end{bmatrix}, \text{ and}$$

$\mathbb{E}\left[I_{v+1}I'_{v+1}\right] =$

$$\begin{bmatrix} \lambda_n\mathbb{E}(f_t - \mathbb{E}\left[f_t|\Omega_v\right])\mathbb{E}\left(f_t - \mathbb{E}\left[f_t|\Omega_v\right]\right)\lambda_n + R_{n,n} & \lambda_n\mathbb{E}\left(f_t - \mathbb{E}\left[f_t|\Omega_v\right]\right)\mathbb{E}\left(f_s - \mathbb{E}\left[f_s|\Omega_v\right]\right)\lambda_m \\ \lambda_m\mathbb{E}\left(f_s - \mathbb{E}\left[f_s|\Omega_v\right]\right)\mathbb{E}\left(f_t - \mathbb{E}\left[f_t|\Omega_v\right]\right)\lambda_n & \lambda_m\mathbb{E}\left(f_s - \mathbb{E}\left[f_s|\Omega_v\right]\right)\mathbb{E}\left(f_s - \mathbb{E}\left[f_s|\Omega_v\right]\right)\lambda_m + R_{m,m} \end{bmatrix},$$

where $\Lambda'_{gdp} = [\lambda_{gdp}\; 2\lambda_{gdp}\; 3\lambda_{gdp}\; 2\lambda_{gdp}\; \lambda_{gdp}]$, $\boldsymbol{f_\tau} = [f_\tau\; f_{\tau-1}\; f_{\tau-2}\; f_{\tau-3}\; f_{\tau-4}]'$, and the appropriate expectations can be extracted by the Kalman filter and smoother.

Naming $\delta_{t,n}^{\tau,gdp} = \mathbb{E}\left[y_{\tau,gdp}^{q} I_{v+1}'\right] \mathbb{E}\left[I_{v+1} I_{v+1}'\right]^{-1}$, we can write

$$\underbrace{\mathbb{E}\left[y_{\tau,gdp}^{q}|\Omega_{v+1}\right] - \mathbb{E}\left[y_{\tau,gdp}^{q}|\Omega_{v}\right]}_{\text{revision}} = \delta_{t,n}^{\tau,gdp}\Big(\underbrace{y_{t,n}^{m} - \mathbb{E}\left[y_{t,n}^{m}|\Omega_{v}\right]}_{\text{news}}\Big) + \delta_{s,m}^{\tau,gdp}\Big(\underbrace{y_{s,m}^{m} - \mathbb{E}\left[y_{s,m}^{m}|\Omega_{v}\right]}_{\text{news}}\Big). \tag{17}$$

In other words, the revision can be decomposed as a weighted average of the news in the latest release. What matters for the revision is both the size of the news as well as its relevance for the variable of interest, as represented by the associated weight $\delta_{t,n}^{\tau,gdp}$. This weight captures the importantance of the update of the factor f given the new information about series n relative to time t for the update of factor f at time τ. Equation (17) can be considered as a generalization of the usual Kalman filter update equation to the case in which new data arrive in a non synchronous manner, or $\tau \neq t$.

As stressed before, this crucial relationship described in equation (17) can be obtained only through "joint models" like DFMs or VARs and in the case of a simultaneous release of several (groups of) variables, bringing the possibility of tracking how single releases have contributed to the forecast revision. While in this review we focus on factor models, most tasks performed with a DFM can also be performed with a VAR. The key is to write the VAR in a state-space form to use the Kalman filter (see, for example, Bańbura et al., 2015). Differently from DFMs, however, the time aggregation in VARs is not exact but regression based (Cimadomo et al., 2022).

2.5 Other "partial" models

Subsection 2.4 has highlighted the importance of relying on joint models for nowcasting. In this subsection, we present advances in "partial" models that are also commonly used for tracking the current state of the economy.

2.5.1 MIDAS-type equations

MIDAS represents an evolution of the "partial" model approach. Here, the predictors are included in the regression at their original observation frequency:

$$y_{\tau,gdp}^{q} = \alpha + \beta\Gamma(L,\theta) y_{t-h_n,n}^{f} + e_{\tau}, \tag{18}$$

where f can be any frequency, $t - h_n$ may also coincide with τ, and $\Gamma(L,\theta)$ is a lag polynomial. Since, for large h_n, many lags of the explanatory variable might be required, the key in this approach is to parsimoniously parameterize $\Gamma(L,\theta)$. Various versions have been proposed (see, for example, Ghysels et al., 2007), including exponential Almon polynomials for which $\Gamma(L,\theta) = \sum_{m=1}^{M} \gamma(m,\theta) L^m$ with $\theta = (\theta_1, \theta_2)$ and

$\gamma(m,\theta) = \frac{\exp(\theta_1 m + \theta_2 m^2)}{\sum_{m=1}^{M}\exp(\theta_1 m+\theta_2 m^2)}$. In contrast to the bridge equations, MIDAS-type regression implies that the temporal aggregation weights are data-driven.

Regarding the problem of ragged edge, the solution in this type of approach can be thought of as re-aligning each time series. The time series with missing observations at the end of the sample are shifted forward to obtain a balanced dataset with the most recent information.[12] The parameters in equation (18) depend on h_n, which is determined by the difference between the forecast target period and the period of the last observation of the predictor. Consequently, separate models must be estimated for different data vintages as the corresponding h_n varies. The case of $t - h_n > \tau$—i.e., when some data referring to the target quarter are available—is sometimes labeled as MIDAS with leads (Andreou et al., 2008).

Applications of this type of model to short-term forecasting include Clements and Galvão (2008, 2009) and Kuzin et al. (2011), who use monthly indicators to forecast GDP, and Andreou et al. (2008), who also include daily financial variables.[13] Given that the MIDAS equations suffer from the curse of dimensionality, a popular strategy for dealing with large information sets is forecast combination (see, e.g., Andreou et al., 2008) or substituting the right-hand side observables $y_{t-h_n,n}$ with factors extracted from a set of monthly predictors as in Marcellino and Schumacher (2010).

As we have already noted, these models are not suited to interpret the impact of new releases on the assessment of the state of the economy. Attempts to circumvent the problem have been based on heuristic procedure, as in Ghysels and Wright (2009), where they construct news using market expectations linked to the change in the forecast by estimating additional auxiliary regressions.

2.5.2 Machine learning

Several papers released in recent years have started to explore the performance of machine learning techniques for short-term forecasting.[14] Among others, Soybilgen and Yazgan (2021) use bagged decision trees, random forests, and stochastic gradient tree boosting models to produce early estimates of U.S. GDP; Richardson et al. (2021) explore ridge, Lasso, elastic net, and support vector machine regression methods other than gradient boosting and neural networks to estimate in real time New Zealand GDP; and Zhang et al. (2023) compare the performance of various machine learning algorithms to DFMs, static factor models, and MIDAS for short-term forecasting the Chinese annualized real

[12]Re-aligning has been a popular strategy to deal with ragged-edge data. See, for example, Altissimo et al. (2001, 2010); de Antonio Liedo and Muñoz (2010).

[13]Clements and Galvão (2008) also show how to add a lag of the low-frequency variable to avoid a seasonal response of the dependent variable to the predictors and use the Broyden-Fletcher-Goldfarg-Shanno method to obtain the estimates of the parameters.

[14]See Goulet Coulombe et al. (2022) for a detailed description of how machine learning techniques can be applied to forecast macroeconomic variables.

GDP growth rate.

Although some of these applications have displayed some encouraging results in terms of accuracy, most of the characteristics of machine learning approaches are classified as "partial," as they are set up in a way that does not allow extracting the unpredictable component of new releases. Therefore, machine learning models cannot interpret how new data releases change the early estimates of the state of the economy.

Moreover, two other characteristics make these models not palatable for nowcasting. First, the "ragged edge" problem is solved by filling the missing values in the quarter of interest with model-based forecasts. Second, the mixed-frequency problem is solved "outside the models"—that is, data are averaged over the quarter to have the predictors at the same frequency as the target variable. These approaches have two side effects. The machine learning models are fed with redundant information—i.e., the information set of these models contains forecasts that are linear combinations of information already contained in the set. Moreover, the aggregation "outside the models" may lag the detection of the early signal that high-frequency variables can deliver, downplaying its effect.

3 Estimation algorithms

We will now describe the algorithms used to estimate DFMs' parameters and to infer their unobserved components in the context of nowcasting applications. The main issue that this part of the literature has tried to overcome is how to exploit all the information when the available data are characterized by different frequencies and by covering dissimilar time spans due to either mismatched historical availability or the asynchronous timing of the releases, thus creating the so-called "ragged" edge.

3.1 Principal components and Kalman filter

One of the most common methods to estimate factor models in the economic literature is the principal component analysis (PCA). PCA estimates the factors and the loadings by finding the pair of $\mathbf{F}_t$ and $\mathbf{\Lambda}$ that minimize the variance of the idiosyncratic component, subject to the constraint that the covariance matrix of the loadings is an identity matrix. Formally,

$$\min_{\{\mathbf{F}_t\}_{t=1}^{T},\mathbf{\Lambda}} \quad \frac{1}{nT}\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{\Lambda}\mathbf{F}_t)'(\mathbf{y}_t - \mathbf{\Lambda}\mathbf{F}_t), \qquad \text{s.t.} \quad \frac{1}{n}\mathbf{\Lambda}'\mathbf{\Lambda} = \mathbf{I}. \tag{19}$$

The solution to this problem yields

$$\widehat{\boldsymbol{\Lambda}} = \mathcal{V}_r \mathcal{D}_r^{1/2} \tag{20}$$

$$\widehat{\mathbf{F}}_t = \frac{1}{n} \widehat{\boldsymbol{\Lambda}}' \mathbf{y}_t, \tag{21}$$

where $\mathcal{D}_r$ is an $r \times r$ diagonal matrix containing the eigenvalues, and $\mathcal{V}_r$ is the $n \times r$ matrix containing the associated eigenvectors of the covariance matrix of $\mathbf{y}_t$. This is equivalent to estimating the factors as a weighted average of the data $\frac{1}{n}\mathbf{w}'\mathbf{y}_t$ where the weights are $\mathbf{w} = \mathcal{V}_r \mathcal{D}_r^{1/2}$. The main intuition of why principal components work is that as the number of variables increases to infinity, the common component survives to aggregation, whereas the idiosyncratic component vanishes.

Estimation of approximate factor models with (static) principal components is studied in Stock and Watson (2002), Bai and Ng (2002), Bai (2003), and Forni et al. (2009). We refer the interested reader to these references for more details.

However, using PCA to estimate loadings and factors in equation (2) does not overcome the complications arising from the mixed frequency and the dissimilar time spans that characterize datasets used for real-time monitoring of the economy. Indeed, PCA can be applied only on the balanced part of the dataset, therefore disregarding historical information, the most recent one, and also all the data series that have a frequency different from the one of the majority of the series in the dataset. Moreover, as a static representation of the data and factors, the PCA specification prevents producing forecasts beyond the sample period.

Doz et al. (2011) made a first step towards overcoming part of these limitations with a two-step procedure. Their idea is to write the factor model in state space, adding equation (3). In the first step, the parameters and the factors of equation (2) are estimated via principal components on the "balanced" part of the information set. The parameters of equation (3) are then estimated via OLS, regressing the estimated factors on their lags. In the second step, factors are re-estimated by applying the Kalman smoother to the part of the information set, including its "unbalanced" part, that contains data with the same frequency. For example, in Giannone et al. (2008), the second step is applied only to the monthly data. Therefore, given the parameters, the factors are also estimated using the unbalanced part of the panel.

The drawbacks of this methodology are twofold. First, data with frequencies different from those prevalent in the information set are either not used to estimate the factors (lower-frequency data) or need to be aggregated outside the model (higher-frequency data). In the example above, quarterly data are disregarded from the estimation of the factors. This can be a material limitation when the information set may contain a large amount of series with a different frequency from the prevalent one. Second, the parameters are estimated only using the "balanced" part of the dataset, while low frequency data are included in the analysis only if they have to be forecasted, usually via a bridge equation that uses the estimated factors.

The literature has proposed overcoming these limitations by estimating the model with a modified version of the expectation-maximization (EM) algorithm or with a Bayesian algorithm. The following two sections will discuss these two algorithms.

3.2 Expectation-maximization algorithm

The EM algorithm is an iterative method to find maximum likelihood estimates of parameters in models with unobserved latent variables. In the case of the model described by equations (2) through (4), at any iteration $\kappa > 0$, in the E-step, given an estimate of the parameters $\widehat{\mathbf{\Lambda}}^{(\kappa-1)}$, $\widehat{\boldsymbol{A}}^{(\kappa-1)}$, $\widehat{\boldsymbol{R}}^{(\kappa-1)}$, and $\widehat{\boldsymbol{Q}}^{(\kappa-1)}$, the factors are extracted using the Kalman filter and the Kalman smoother. Then, given $\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{F_t}|\Omega_v\right]$, in the M-step the parameters are re-estimated. Specifically,

$$\widehat{\mathbf{\Lambda}}^{(\kappa)} = \left(\sum_{t=1}^{T_v}\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{y_t F_t'}|\Omega_v\right]\right)\left(\sum_{t=1}^{T_v}\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{F_t F_t'}|\Omega_v\right]\right)^{-1}, \tag{22}$$

$$\widehat{\mathbf{A}}^{(\kappa)} = \left(\sum_{t=1}^{T_v}\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{F_t F_{t-1}'}|\Omega_v\right]\right)\left(\sum_{t=1}^{T_v}\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{F_{t-1} F_{t-1}'}|\Omega_v\right]\right)^{-1}. \tag{23}$$

from which we can estimate $\boldsymbol{R}$ and $\boldsymbol{Q}$ as follows:

$$\widehat{\boldsymbol{R}}^{(\kappa)} = \operatorname{diag}\left(\frac{1}{T_v}\left(\sum_{t=1}^{T_v}\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{y_t y_t'}|\Omega_v\right]\right] - \widehat{\mathbf{\Lambda}}^{(\kappa)}\sum_{t=1}^{T_v}\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{F_t y_t'}|\Omega_v\right)\right) \tag{24}$$

$$\widehat{\boldsymbol{Q}}^{(\kappa)} = \frac{1}{T}\left(\sum_{t=1}^{T_v}\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{F_t F_t'}|\Omega_v\right] - \widehat{\mathbf{A}}^{(\kappa)}\sum_{t=1}^{T_v}\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{F_{t-1} F_t'}|\Omega_v\right]\right). \tag{25}$$

The algorithm runs until the increase in the likelihood between two consecutive iterations is below a certain threshold. Lastly, the algorithm is initialized by estimating $\mathbf{F}_t$ and $\mathbf{\Lambda}$ by principal components and $\mathbf{A}$ by OLS. For a rigorous treatment of the EM algorithm in DFMs, we refer the reader to Doz et al. (2012) and Barigozzi and Luciani (2022).

However, this algorithm *per se* is also not suited to deal with datasets characterized by different frequencies and covering dissimilar time spans. Indeed, if $\boldsymbol{y_t}$ did not contain missing observations, we would have that

$$\mathbb{E}_\kappa\left[\boldsymbol{y_t y_t'}|\Omega_v\right] = \boldsymbol{y_t y_t'} \quad \text{and} \quad \mathbb{E}_\kappa\left[\boldsymbol{y_t F_t'}|\Omega_v\right] = \boldsymbol{y_t}\mathbb{E}_\kappa\left[\boldsymbol{F_t'}|\Omega_v\right], \tag{26}$$

which can be plugged into the equations above and $\mathbb{E}_{\kappa}\left[\boldsymbol{F_t F_t'}|\Omega_v\right]$, $\mathbb{E}_{\kappa}\left[\boldsymbol{F_t F_{t-1}'}|\Omega_v\right]$, and $\mathbb{E}_{\kappa}\left[\boldsymbol{F_t}|\Omega_v\right]$ can be obtained via the Kalman filter and smoother. Given that in nowcasting applications $\boldsymbol{y_t}$ contains missing observations due to the mixed-frequency nature of the data and the dissimilar time spans availability, the EM algorithm needs to be modified. Bańbura and Modugno (2014) make the EM algorithm suitable to such cases by re-defining the vector of data $\boldsymbol{y_t}$ as

$$\boldsymbol{y_t} = \boldsymbol{W_t y_t} + (\boldsymbol{I} - \boldsymbol{W_t})\boldsymbol{y_t},$$

where $\boldsymbol{W_t}$ is a diagonal matrix with ones corresponding to the non-missing entries in $\boldsymbol{y_t}$ and zeros otherwise. With this change, equations (22) and (24) become

$$\mathrm{vec}\big(\widehat{\boldsymbol{\Lambda}}^{(\kappa)}\big) = \left(\sum_{t=1}^{T_v} \mathbb{E}_{(\kappa-1)}\left[\boldsymbol{F_t F_t'}|\Omega_v\right] \otimes \boldsymbol{W}_t\right)^{-1} \mathrm{vec}\left(\sum_{t=1}^{T_v} \boldsymbol{W_t y_t}\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{F_t'}|\Omega_v\right]\right) \tag{27}$$

and

$$\begin{aligned}
\widehat{\boldsymbol{R}}^{(\kappa)} \; &= \; \mathrm{diag}\Bigg(\frac{1}{T_v}\sum_{t=1}^{T_v}\Big(\boldsymbol{W_t y_t y_t' W_t'} - \boldsymbol{W_t y_t}\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{F_t'}|\Omega_v\right]\widehat{\boldsymbol{\Lambda}^{(\kappa)\prime}}\boldsymbol{W_t} \\
&- \; \boldsymbol{W}_t\widehat{\boldsymbol{\Lambda}}^{(\kappa)}\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{F}_t|\Omega_v\right]\boldsymbol{y}_t'\boldsymbol{W}_t + \boldsymbol{W}_t\widehat{\boldsymbol{\Lambda}}^{(\kappa)}\mathbb{E}_{(\kappa-1)}\left[\boldsymbol{F_t F_t'}|\Omega_v\right]\widehat{\boldsymbol{\Lambda}}^{(\kappa)\prime}\boldsymbol{W}_t \\
&+ \; (\boldsymbol{I} - \boldsymbol{W_t})\widehat{\boldsymbol{R}}^{((\kappa-1))}(\boldsymbol{I} - \boldsymbol{W_t})\Big)\Bigg).
\end{aligned} \tag{28}$$

Therefore, $\boldsymbol{W_t}$ works as a selection matrix that allows us to obtain the expectations in equation (26) and the corresponding remaining parts of the estimates when the data are available.

3.3 Bayesian inference

An alternative way of estimating DFMs on incomplete datasets is to use Bayesian inference. D'Agostino et al. (2016) propose the following model:

$$x_{it} = \sum_{s=0}^{p} \boldsymbol{\lambda}_{is}\mathbf{f}_{t-s} + \sum_{s=1}^{p} \rho_{is}x_{it-s} + e_{it} \tag{29}$$

$$\mathbf{f}_t = \sum_{s=1}^{p} \mathbf{A}_s\mathbf{f}_{t-s} + \mathbf{u}_t \tag{30}$$

where $\mathbf{u}_t \sim \mathcal{N}(0, \mathbf{I}_r)$ and $e_{it} \sim \mathcal{N}(0, \psi_{it})$.

Compared to models (2) through (4), the main difference is that the factors are allowed to be loaded dynamically by the variables through the polynomial $\boldsymbol{\lambda}_i(L) = \sum_{s=0}^{p} \boldsymbol{\lambda}_{is}$. To estimate the large number of parameters in the model, D'Agostino et al. (2016) propose an algorithm based on the following priors:

$$\begin{aligned}
&R_{i,i} \sim IG(1,3), \\
&\Lambda_{i,r,h} \sim N(0, \tau\frac{1}{(h+1)^2}), \\
&A_{r,h} \sim N(0, \tau\frac{1}{h^2}), \\
&Q_{r,r} \sim N(0,1),
\end{aligned}$$

where r indicates the factor and h indicates the lag of the factor to which the coefficient is associated. The prior covariance among coefficients associated with different variables and lags is set to zero. Notice that the variance of the prior is lower for the coefficients associated with more distant lags. The hyperparameter τ controls the scale of all the variances and effectively governs the overall level of shrinkage, and the authors fix it to the conventional value of 0.2. These priors, including the choice of the degree of overall shrinkage, are similar to the Minnesota prior proposed by Litterman (1986) in the context of Bayesian VARs. The inference is conducted using Gibbs sampling techniques. If all data and the common factor were observed, drawing from the posterior of the parameters would have been easy since the prior is conjugate. Conditionally on the parameters and the observable data, the common factors and the missing data can be drawn using simulation smoothers (Carter and Kohn, 1994; De Jong and Shephard, 1995; Durbin and Koopman, 2002). In other words, the Gibbs sampler consists of alternating the following two steps: (i) given a draw of the parameters, draw the missing data and the latent factor conditional on the observations using the simulation smoother; and (ii) given a draw of the full data and the latent factors, draw the parameters from their posterior.

The algorithm is initialized using the parameters associated with principal components computed by fitting missing data by a spline function. This algorithm has been successfully applied to nowcast U.S. GDP by D'Agostino et al. (2016) and Drechsel et al. (2023) and Norwegian GDP by Luciani and Ricci (2014).

4 Data selection

Which variables should we include in the dataset to nowcast GDP? And how many of them? In theory, these should be trivial questions, as when estimating large DFMs, we can add as many variables as we like, even more so because these models are consistently estimated for n growing to infinity. However, it turns out that adding as many variables as we can is not the right recipe. The right recipe consists of adding the right variables; that is, we want to add variables that contain signals and avoid those that contain just noise.[15]

How do we determine which are the right variables? The answer to this question has stimulated an interesting debate among academics and practitioners at central banks and other financial institutions. The literature has broadly conceived two alternative methods to answer this question: expertise-based and statistical-based selection methods.

4.1 Expertise-based selection

The first solution relies on the expertise of the people who monitor the state of the economy daily. Their monitoring activities inform monetary policy decisions at central banks, fiscal policy decisions at governmental agencies, or investment decisions at financial and non financial businesses. Including the data that the experts consider important to assess the state of the economy is a natural choice, as one of the goals of nowcasting is to interpret how new data releases change the model-based assessment of current macroeconomic conditions.

In academic papers, scholars rely on the so-called "market-moving indicators" to infer the experts' preferences about data. For example, news platforms and data providers such as Bloomberg, Forex Factory, and Trading Economics report quantitative or qualitative indexes that indicate the importance of a given data release for their users from which we can identify the market-moving indicators.[16]

As an alternative to identifying "market moving indicators," some papers rely on expert judgment. For example, in their seminal paper, Giannone et al. (2008) constructed the dataset with the help of economists at the Board of Governors of the Federal Reserve System, therefore tailoring the dataset to the internal expertise and interest in specific variables, among them several sectoral disaggregated variables. Another example is Barigozzi and Luciani (2021), who started from a large dataset and then eliminated variables by looking for those with very high idiosyncratic cross-correlation and using judgment.

Through time, the literature has concluded that for nowcasting GDP, the information set should include two categories of data. The first category is so-called hard data—i.e., data collected by statistical agencies based on measurable quantities, like variables about labor markets (e.g., the unemployment rate), the industrial sector (e.g., the index of

[15] Adding noise means adding a variable that is idiosyncratic and contains no information about the common factors or adding a variable that is very correlated with another variable in the dataset and, hence, contain no additional information. When we add a variable very similar to another variable in the dataset, we increase the cross-correlation among idiosyncratic components, which is very problematic, as Boivin and Ng (2006) show that excessive cross-sectional correlation among idiosyncratic components worsens the model's forecasting performance, a result confirmed by the simulations in Luciani (2014).

[16] For example, for each data release, Bloomberg reports a relevance index based on the percentage of Bloomberg users who have set up an automatic alert about that specific release. See Cascaldi-Garcia et al. (2023) for a nowcast application that selects market-moving indicators based on the Bloomberg relevance index.

industrial production, and industrial turnover), the construction sector (e.g., the index of production in construction), private consumption (e.g., retail sales and car registrations), and the external sector (e.g., exports and imports of goods). The second category is so-called soft data—i.e., survey indexes that portray feelings and perceptions of economic agents about current and future economic prospects.[17]

Moreover, the nowcasting literature has concluded that focusing mainly on the headlines of each macroeconomic report while disregarding sectoral disaggregation is a simple and very effective solution. For example, Bańbura and Modugno (2014) and Bańbura et al. (2011) show that the marginal impact on the nowcast precision of disaggregated data is minimal, which is in line with market participants primarily focusing on the headlines of each report. Moreover, the same authors show that the model's nowcasting performance does not deteriorate if the right disaggregated data are included—regardless of relying on market participants or expert judgment—and the factor model is robust; see the empirical analysis in Bańbura et al. (2010) for forecasting at longer horizons, and the simulation studies of Doz et al. (2011, 2012) and Barigozzi and Luciani (2022) for estimation performance. Another result of the nowcasting literature is that daily and weekly indicators, such as financial variables, do not improve the performance of a nowcasting model either during normal times or during recessions because the high-frequency component of these indicators is detached from the real economy (see Bańbura et al., 2013). This result does not imply that higher-frequency indicators are unrelated to real economic activity but that the link is through their low-frequency component. As such, the usefulness of high-frequency indicators for nowcasting is dim.

4.2 Statistical-based selection

The second solution consists of selecting the variables using statistical criteria as suggested by Boivin and Ng (2006) and Bai and Ng (2008)—for example, Bai and Ng (2008) suggest using only the variables informative for forecasting the target variable—which are also at the foundation of the machine learning approach as discussed in Section 2.5.2. However, statistical-based selection does not work well when the data are very correlated, which is the case of macroeconomic data. In particular, De Mol et al. (2008) find no major differences in the forecasting performance between models using statistical-based selection and those with no selection.[18] Most importantly, variable selection is unstable because of collinearity among predictors. In other words, the set of predictors selected from month to month is very sensitive to minor perturbations of the dataset, such as adding new variables or extending the sample length, which makes (revisions to) the forecasts obtained with this method difficult to interpret. Similar instabilities have also been found

[17] Measures of prices and monetary aggregates are usually found not informative for nowcasting GDP. This result goes back to Sargent and Sims (1977).

[18] This result also emerges from a careful reading of the empirical results of Boivin and Ng (2006).

in the context of model selection and model averaging (Ouysse, 2011; Stock and Watson, 2012). Finally, Giannone et al. (2021) find that economic data do not appear informative enough to uniquely identify the relevant predictors when a large pool of variables is available to the researcher.

4.3 Alternative data

Given the growing availability of information and the increased ability to process and store data, information that goes beyond macro data has been tested for nowcasting GDP in recent years. One of the most promising typologies of data is corporate accounting data. Indeed, Abdalla et al. (2021) find that factors based on the real-time flow of accounting data from the corporate financial reports are incrementally relevant for nowcasting and forecasting major components of economic output in the BEA's National Income and Product Accounts.

One of the most-studied sources of big data in macroeconomics is Google Trends, but the literature expresses mixed views about its usefulness for nowcasting. While Choi and Varian (2012) argue that Google Trends data help in forecasting near-term values of several economic indicators and Ferrara and Simoni (2022) find improvements in nowcasting GDP accuracy for several countries, Larson and Sinclair (2022) show that such data do not improve the accuracy of nowcasts of unemployment insurance claims, neither in normal times nor during the COVID-19 pandemic. Moreover, as pointed out by Lazer et al. (2014), Google Trends, like other publicly available indices, are the product of numerous algorithms and decisions made by engineers that are invisible to the user. The problem for forecasters is that these algorithms are not static but are tweaked and adapted as time passes. Therefore, historical values available to us now are not the same as those that were available in the past, and values available in the future may be different again.[19]

5 Empirical application

In this section, we put into practice the nowcasting lessons detailed in this chapter. As highlighted by Cascaldi-Garcia et al. (2023), timely soft data such as surveys and confidence indicators, even if only qualitative, contain important information about the state of the economy, which is paramount for nowcasting economic activity. We test this result further by presenting the nowcasting performance of a mixed-frequency DFM combining soft and hard data for seven advanced economies: the euro area, Germany, France, Italy, Canada, Japan, and the U.K.

[19] Simon van Norden explained these issues in a 2017 post on Econbrowser, available at econbrowser.com/archives/2017/05/guest-contribution-big-data-and-fake-forecasts.

We illustrate the performance of such models by looking at two global events of large deterioration of the economic conditions, namely the 2008 Global Financial Crisis (GFC) and the onset of the COVID-19 pandemic. For the GFC, we use pseudo-real-time data for the 2008–09 period. We provide weekly nowcasts of quarter-on-quarter (QoQ) GDP growth (at an annual rate, a.r.), following the evolution of the model predictions until the eve of the official GDP release. For the onset of the pandemic, we provide the weekly nowcasts of GDP growth (QoQ a.r.) for the second quarter of 2020. By using real-time data, we follow the evolution of the model predictions starting in the first week of January 2020 through the first release of each country's GDP.[20]

For the euro area and its main economies, we follow the model set-up and data selection proposed by Cascaldi-Garcia et al. (2023). For Canada, we follow the model set-up and a similar data selection as proposed by Bragoli and Modugno (2017), who combine soft and hard data not only from Canada but also from the U.S.—the model also highly benefits from the official release of monthly GDP. For Japan, we follow a data selection similar to Carriero et al. (2019) and Hayashi and Tachi (2023). Lastly, for the U.K., we follow a data selection similar to Carriero et al. (2019) and Anesti et al. (2018)—as for the model for Canada, the UK model highly benefits from the official release of monthly GDP.

5.1 The Global Financial Crisis

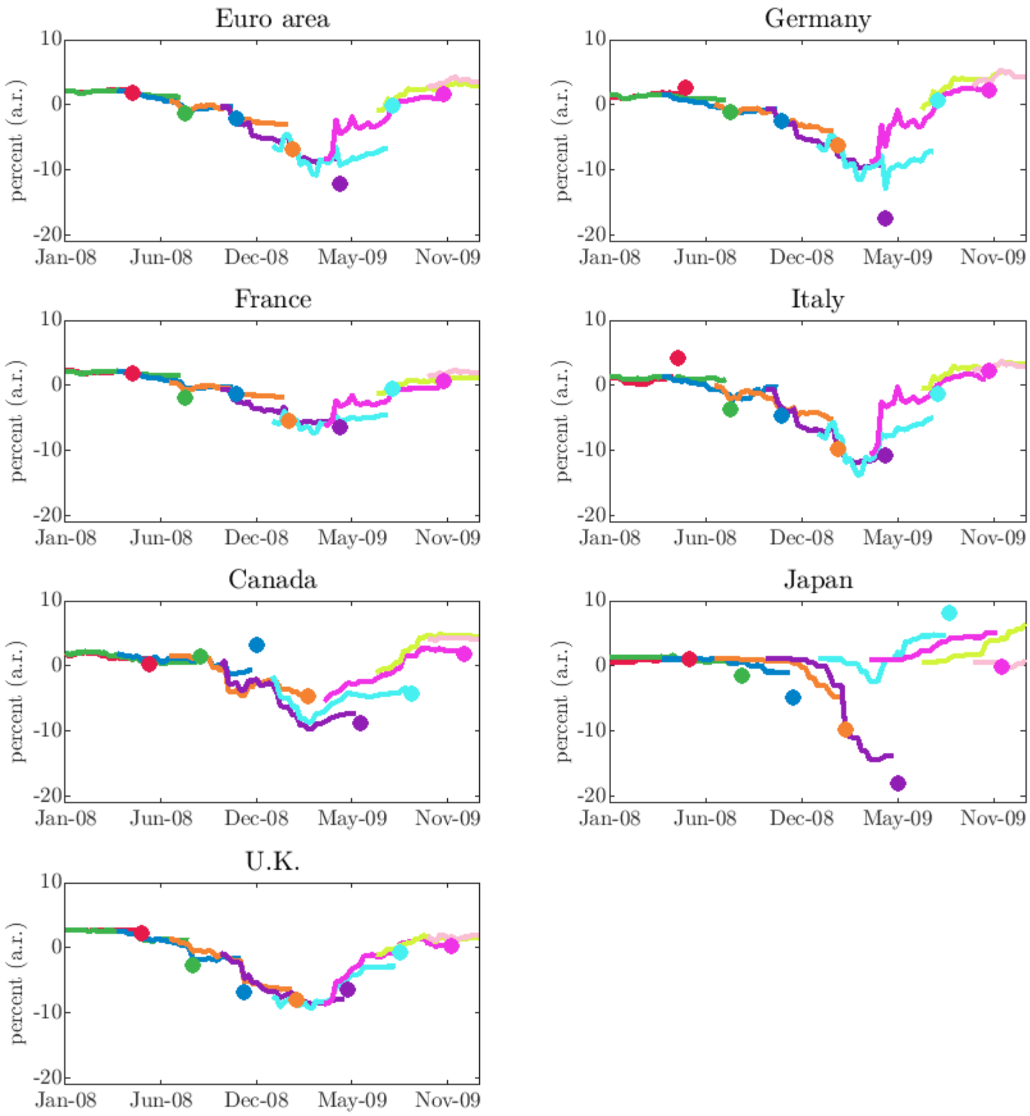
The GFC started in 2008 with financial stresses in the U.S. It quickly spread across the world through the banking system, causing substantial real economic contraction that lasted several quarters. The fast-moving contagion made timely and accurate hard data on the state of the economy more important than ever for policymakers to direct counter-cyclical measures. However, the long delay in the release of such data made policy decisions based on such indicators impracticable. The nowcasting literature has extensively shown that mixed-frequency DFMs enriched with soft data would have been able to track economic developments during the GFC in a timely and precise manner. In this section, we bring additional evidence in support of this finding.

Figure 1 presents the evolution of the nowcast of GDP growth (QoQ a.r.) for seven advanced economies. Each line corresponds to the evolution of the nowcast as new data become available, while the dots correspond to the official GDP release. As shown in the figure, the nowcast evolution tracked quite closely the economic disarray observed through 2008, culminating in a global recession. Moreover, the models were quite accurate even at the height of the crisis, with double-digit contractions.

Let us zoom in on the weekly evolution of the euro-area GDP nowcast for 2009:Q3. Figure 2 shows the great advantage of using "joint" models over "partial" models: the

[20]The first release of 2020:Q2 GDP was published in August for all the countries analyzed.

Figure 1 Nowcast evolution during the Global Financial Crisis



Note: Nowcast evolution from dynamic factor models for seven advanced economies. Red, green, blue, orange, purple, cyan, magenta, yellow, and pink lines follow the nowcasts for the GDP growth (QoQ a.r.) for 2008:Q1 through 2010:Q1, respectively. Dots correspond to the final official GDP release.

possibility of clearly disentangling how releases of all the data included in the information set change the assessment of the current state of the economy. The upper panel shows the weekly evolution of the 2009:Q3 euro-area GDP growth nowcast from the beginning of April 2009 to the end of October 2009. For the euro area, we can decompose the contribution of new releases either by country, middle panel, or by nature of the data (soft or hard), lower panel. Focusing on the lower panel, we can see the pivotal role played by soft data in capturing the current state of the economy, anecdotally confirming that soft data are very important to extract timely signals. However, soft data may be noisy: although the overall signal points to improved macroeconomic conditions in the euro area for this specific quarter, few soft data releases deliver confounding signals. In contrast, hard data consistently point to improved macroeconomic conditions, even though their contribution becomes relevant only in the middle of the quarter of interest.[21]

5.2 The COVID-19 experience

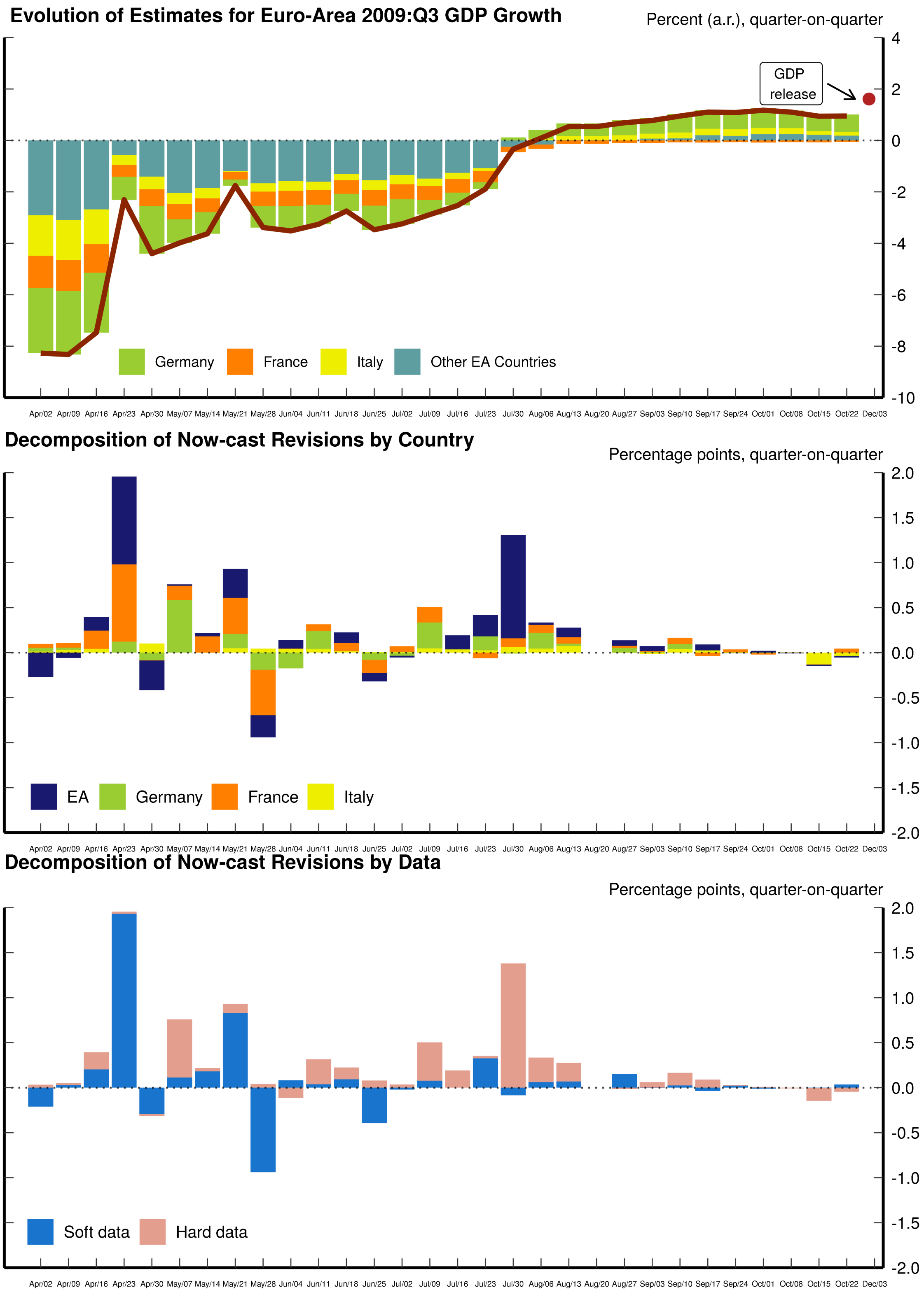
The onset of the COVID-19 pandemic caused a sharp and intense deterioration of the economy that was unparalleled in recent history. The fast-moving developments observed from March 2020 onward posed challenges to the usual tools for nowcasting, making this event a natural laboratory for these methods.

The lockdowns widely imposed across the globe at the onset of the pandemic closed down factories and drastically diminished manufacturing production, services not deemed essential were suspended, and entire sectors such as tourism and air transportation came to a halt. Official quantitative data became difficult to collect and were unreliable, and the long release delay made them close to useless. As such, researchers and policymakers were using the few and unstructured releases of timely sentiment indicators, such as confidence indexes and purchasing managers' indexes, to grasp the magnitude of the damage the pandemic inflicted on the economy. The results in this section show that a DFM would have pointed to the eventual double-digit contraction over the advanced economies as soon as these timely indicators started to be released.

Figure 3 illustrates the weekly evolution of the nowcasts from DFMs for selected advanced economies of 2020:Q2 GDP growth (QoQ a.r.), starting from January 2020. The lockdowns started to be broadly implemented between March and April, so their real quantitative effects would only be manifested once indicators such as industrial production and retail sales were released, which may have a delay of up to 60 days. However, as early as the end of March, confidence indicators already gave hints of an upcoming severe economic contraction. By mid-April, the nowcasts of all selected advanced economies showed predictions close to double-digit contraction in the second quarter of 2020, while almost no hard data had been released yet. Once hard data started to become available,

[21] See Cascaldi-Garcia et al. (2023) for the full description of the model.

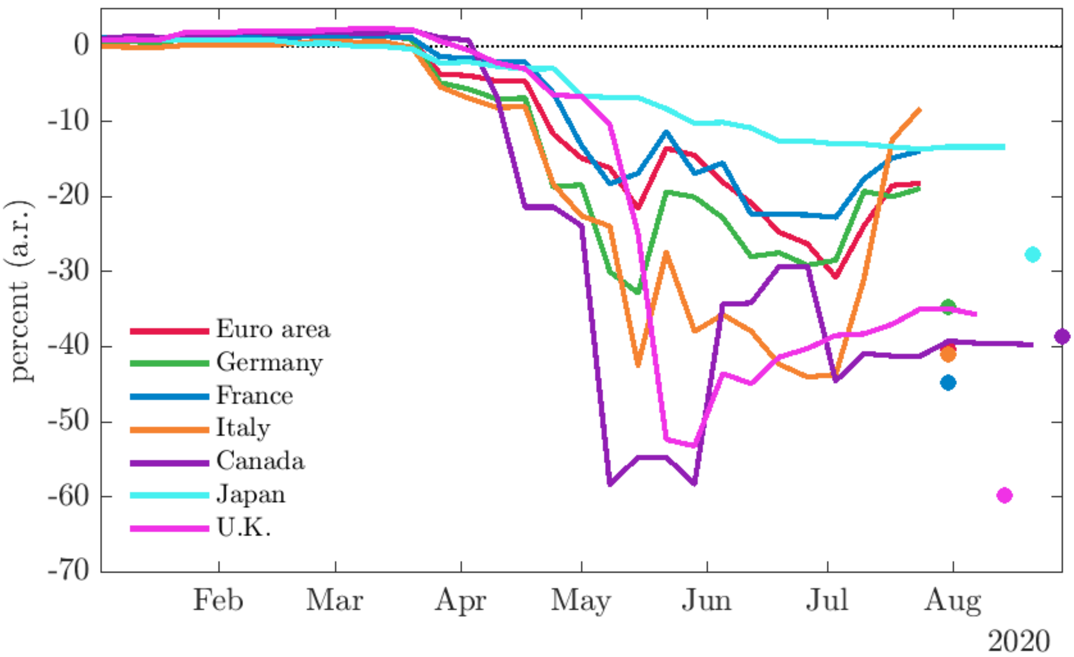
Figure 2 News decomposition



Note: The upper chart shows the weekly evolution of the 2009:Q3 euro area GDP growth (QoQ a.r.) nowcast from April 2, 2009 to December 3, 2009. The middle chart shows the contribution to the nowcast revision of variables grouped by country. The lower chart shows the contribution of variables grouped by hard and soft data.

the model assessed that economic conditions were deteriorating even further, with nowcasts hitting unprecedented marks ranging from -10% to -60% by mid-May, persisting until the eventual first GDP release in August 2020.

Figure 3 Nowcast evolution during the onset of the COVID-19 pandemic



Note: Nowcast evolution from dynamic factor models for seven advanced economies. Solid lines follow the nowcasts for the quarter-on-quarter GDP growth (QoQ a.r.) for 2020:Q2. Dots correspond to the first official GDP release.

In conclusion, two results emerge from our analysis of the COVID-19 pandemic. First, a comprehensive model that summarizes unstructured, unbalanced, and mixed-frequency data can be quite informative about the state of the economy, especially in events of rapid regime changes. Second, timely soft data proved useful to get a first assessment of the magnitude of the impact on the economy from the lockdowns.

5.3 Dealing with the pandemic period in nowcasting models

The DFMs that we used in this section to nowcast the onset of the COVID-19 pandemic are linear Gaussian models. As such, they are not well-equipped to fit such an extreme event unlike anything else in history. Going forward, methodological changes may be needed to improve the nowcasting performance of mixed-frequency models when dealing with extreme events.

Since March 2020, the literature has proposed many interesting econometric enhancements on how to deal with an extreme episode such as the COVID-19 pandemic, both in the Frequentist and Bayesian environments. These enhancements are either already directly implemented in mixed-frequency set-ups, or future research could engineer them to work with the current set of tools available for nowcasting.[22] A non-extensive list of methods proposed include the following:

- Excluding the pandemic observations altogether (Schorfheide and Song, 2020)
- Estimating a common shift and persistence of the volatility of the shocks during the extreme periods of the pandemic (Lenza and Primiceri, 2021)
- Downplaying the importance of extreme observations (Cascaldi-Garcia, 2022)
- Modeling extreme observations either as random shifts in the stochastic volatility (Carriero et al., 2022; Álvarez and Odendahl, 2022), or through non-parametric methods (Huber et al., 2023)
- Estimating the model with t-distributed errors (Bobeica and Hartwig, 2023)
- Augmenting the information set with an exogenous variable capturing the pandemic period (Ng, 2021)
- Modeling outliers directly in the DFM (Antolin-Diaz et al., 2020)

6 Conclusion

Nowadays, nowcasting models are the most popular tools used to assess the current state of the economy at central banks, governmental agencies, and financial and non-financial corporations all around the world. They have also been the topic of a hefty body of academic literature. In this chapter, we have summarized how the nowcasting framework has evolved and is currently applied to 1) efficiently use large information sets characterized by data with mixed frequency and mismatched time spans due to dissimilar historical availability and asynchronous release time, and 2) provide the final users with a tool that helps to interpret why a given data release has changed the model's assessment of the state of the economy.

References

Aastveit, K. A. and Trovik, T. (2012). Nowcasting Norwegian GDP: The role of asset prices in a small open economy. *Empirical Economics*, 42(1):95–119.

Abdalla, A. M., Carabias, J. M., and Patatoukas, P. N. (2021). The real-time macro content of corporate financial reports: A dynamic factor model approach. *Journal of Monetary Economics*, 118:260–280.

Altissimo, F., Bassanetti, A., Cristadoro, R., Forni, M., Hallin, M., Lippi, M., Reichlin, L., and Veronese, G. (2001). Eurocoin: A real time coincident indicator of the euro area business cycle. CEPR Discussion Papers 3108, C.E.P.R. Discussion Papers.

[22] Ho (2023) summarizes the macroeconomic forecasting performance after the pandemic considering several different methods.

Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., and Veronese, G. (2010). New eurocoin: Tracking economic growth in real time. *The Review of Economics and Statistics*, 92(4):1024–1034.

Álvarez, L. J. and Odendahl, F. (2022). Data outliers and Bayesian VARs in the Euro Area. Technical report.

Andreini, P., Hasenzagl, T., Reichlin, L., Senftleben-König, C., and Strohsal, T. (2023). Nowcasting German GDP: Foreign factors, financial markets, and model averaging. *International Journal of Forecasting*, 39(1):298–313.

Andreou, E., Ghysels, E., and Kourtellos, A. (2008). Should macroeconomic forecasters look at daily financial data? Manuscript, University of Cyprus.

Anesti, N., Galvão, A. B., and Miranda-Agrippino, S. (2018). Uncertain Kingdom: Nowcasting GDP and its Revisions. Discussion Papers 1824, Centre for Macroeconomics (CFM).

Angelini, E., Bańbura, M., and Rünstler, G. (2010). Estimating and forecasting the euro area monthly national accounts from a dynamic factor model. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2010/1.

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

Antolin-Diaz, J., Drechsel, T., and Petrella, I. (2020). Advances in nowcasting economic activity: Secular trends, large shocks and new data. Technical report, Working paper.

Arnostova, K., Havrlant, D., Rùžièka, L., and Tóth, P. (2011). Short-Term Forecasting of Czech Quarterly GDP Using Monthly Indicators. *Czech Journal of Economics and Finance (Finance a uver)*, 61(6):566–583.

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

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71:135–171.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70:191–221.

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317.

Bańbura, M., Giannone, D., and Lenza, M. (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of Forecasting*, 31(3):739–756.

Bańbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-Casting and the Real-Time Data Flow. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2 of *Handbook of Economic Forecasting*, chapter 0, pages 195–237. Elsevier.

Bańbura, M., Giannone, D., and Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of applied Econometrics*, 25(1):71–92.

Bańbura, M., Giannone, D., and Reichlin, L. (2011). Nowcasting. In *Handbook on Economic Forecasting*, volume 2, pages 63–90. 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.

Bańbura, M. and Rünstler, G. (2011). A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, 27(2):333–346.

Barcelona, W., Cascaldi-Garcia, D., Hoek, J., and Van Leemput, E. (2022). What Happens in China Does Not Stay in China. *International Finance Discussion Paper*, (1360).

Barhoumi, K., Darné, O., and Ferrara, L. (2010). Are disaggregate data useful for factor analysis in forecasting French GDP? *Journal of Forecasting*, 29(1-2):132–144.

Barigozzi, M. and Luciani, M. (2021). Measuring US aggregate output and output gap using large datasets. *Review of Economics and Statistics*. forthcoming.

Barigozzi, M. and Luciani, M. (2022). Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the em algorithm. arXiv:1910.03821.v4.

Bessec, M. and Doz, C. (2014). Short-term forecasting of French GDP growth using dynamic factor models. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2013(2):11–50.

Bobeica, E. and Hartwig, B. (2023). The COVID-19 shock and challenges for inflation modelling. *International Journal of Forecasting*, 39(1):519–539.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 127:169–194.

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.

Bragoli, D. (2017). Now-casting the Japanese economy. *International Journal of Forecasting*, 33(2):390–402.

Bragoli, D. and Fosten, J. (2018). Nowcasting Indian GDP. *Oxford Bulletin of Economics and Statistics*, 80(2):259–282.

Bragoli, D., Metelli, L., and Modugno, M. (2015). The importance of updating: Evidence from a Brazilian nowcasting model. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2015(1):5–22.

Bragoli, D. and Modugno, M. (2017). A now-casting model for Canada: Do U.S. variables matter? *International Journal of Forecasting*, 33(4):786–800.

Camacho, M. and Perez-Quiros, G. (2010). Introducing the euro-sting: Short-term indicator of euro area growth. *Journal of Applied Econometrics*, 25(4):663–694.

Carriero, A., Clark, T. E., Marcellino, M., and Mertens, E. (2022). Addressing COVID-19 Outliers in BVARs with Stochastic Volatility. *The Review of Economics and Statistics*, pages 1–38.

Carriero, A., Galvão, A. B., and Kapetanios, G. (2019). A comprehensive evaluation of macroeconomic forecasting methods. *International Journal of Forecasting*, 35(4):1226–1239.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81(3):541–553.

Caruso, A. (2018). Nowcasting with the help of foreign indicators: The case of Mexico. *Economic Modelling*, 69:160–168.

Cascaldi-Garcia, D. (2022). Pandemic priors. *International Finance Discussion Paper*, (1352).

Cascaldi-Garcia, D., Ferreira, T. R., Giannone, D., and Modugno, M. (2023). Back to the present: Learning about the euro area through a now-casting model. *International Journal of Forecasting*.

Choi, H. and Varian, H. (2012). Predicting the Present with Google Trends. *The Economic Record*, 88(s1):2–9.

Cimadomo, J., Giannone, D., Lenza, M., Monti, F., and Sokol, A. (2022). Nowcasting with large Bayesian vector autoregressions. *Journal of Econometrics*, 231(2):500–519. Special Issue: The Econometrics of Macroeconomic and Financial Data.

Clements, M. P. and Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data. *Journal of Business & Economics Statistics*, 26:546–554.

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.

Coroneo, L., Giannone, D., and Modugno, M. (2016). Unspanned Macroeconomic Factors in the Yield Curve. *Journal of Business & Economic Statistics*, 34(3):472–485.

D'Agostino, A., Giannone, D., Lenza, M., and Modugno, M. (2016). Nowcasting Business Cycles: A Bayesian Approach to Dynamic Heterogeneous Factor Models. In Hillebrand, E. and Koopman, S. J., editors, *Dynamic Factor Models*, volume 35 of *Advances in Econometrics*, pages 569–594. Emerald Publishing Ltd.

D'Agostino, A., McQuinn, K., and O'Brien, D. (2012). Nowcasting Irish GDP. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2012(2):21–31.

D'Agostino, A., Modugno, M., and Osbat, C. (2017). A Global Trade Model for the Euro Area. *International Journal of Central Banking*, 13(4):1–34.

Dahlhaus, T., Guénette, J.-D., and Vasishtha, G. (2017). Nowcasting BRIC+M in real time. *International Journal of Forecasting*, 33(4):915–935.

de Antonio Liedo, D. (2015). Nowcasting Belgium. *Eurostat Review on National Accounts and Macroeconomic Indicators*, 2:7–48.

de Antonio Liedo, D. and Muñoz, E. F. (2010). Nowcasting Spanish GDP growth in real time: "one and a half months earlier". Banco de España Working Papers 1037, Banco de España.

De Jong, P. and Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, 82(2):339–350.

De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318–328.

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(5):371–390.

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164:188–205.

Doz, C., Giannone, D., and Reichlin, L. (2012). A Quasi–Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models. *The Review of Economics and Statistics*, 94(4):1014–1024.

Drechsel, K. and Scheufele, R. (2012). A comparison of bottom-up approaches and direct forecasts of German GDP in a data-rich environment. Technical report, Halle Institute for Economic Research (IWH). Manuscript.

Drechsel, T., Antolin-Diaz, J., and Petrella, I. (2023). Advances in Nowcasting Economic Activity: The Role of Heterogeneous Dynamics and Fat Tails. CEPR Discussion Papers 17800, C.E.P.R. Discussion Papers.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616.

Ferrara, L. and Simoni, A. (2022). When are Google data useful to nowcast GDP? An approach via preselection and shrinkage. *Journal of Business & Economic Statistics*, pages 1–15.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the Black Box: Structural Factor Models versus Structural VARs. *Econometric Theory*, 25(5):1319–1347.

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

Ghysels, E. and Wright, J. H. (2009). Forecasting professional forecasters. *Journal of Business and Economic Statistics*, 27(4):504–516.

Giannone, D., Agrippino, S. M., and Modugno, M. (2013). Nowcasting China real GDP. Technical report, Mimeo.

Giannone, D., Lenza, M., and Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2409–2437.

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.

González-Astudillo, M. and Baquero, D. (2019). A nowcasting model for Ecuador: Implementing a time-varying mean output growth. *Economic Modelling*, 82:250–263.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., and Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37(5):920–964.

Hahn, E. and Skudelny, F. (2008). Early estimates of euro area real GDP growth - a bottom-up approach from the production side. Working Paper Series 975, European Central Bank.

Hayashi, F. and Tachi, Y. (2021). The nowcast revision analysis extended. *Economics Letters*, 209:110112.

Hayashi, F. and Tachi, Y. (2023). Nowcasting japan's GDP. *Empirical Economics*, 64(4):1699–1735.

Ho, P. (2023). Forecasting in the absence of precedent. *Journal of Economic Surveys*, 37(3):1033–1058.

Huber, F., Koop, G., Onorante, L., Pfarrhofer, M., and Schreiner, J. (2023). Nowcasting in a pandemic using non-parametric mixed frequency VARs. *Journal of Econometrics*, 232(1):52–69.

Jansen, W. J., Jin, X., and de Winter, J. M. (2016). Forecasting and nowcasting real GDP: Comparing statistical models and subjective forecasts. *International Journal of Forecasting*, 32(2):411–436.

Kitchen, J. and Monaco, R. M. (2003). Real-time forecasting in practice: The U.S. Treasury sta.'s real-time GDP forecast system. *Business Economics*, pages 10–19.

Kuzin, V., Marcellino, M., and Schumacher, C. (2011). Midas vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, 27(2):529–542.

Lahiri, K. and Monokroussos, G. (2013). Nowcasting US GDP: The role of ISM business surveys. *International Journal of Forecasting*, 29(4):644–658.

Larson, W. D. and Sinclair, T. M. (2022). Nowcasting unemployment insurance claims in the time of COVID-19. *International Journal of Forecasting*, 38(2):635–647.

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). Big data. The parable of Google Flu: traps in big data analysis. *Science*, 343.

Lenza, M. and Primiceri, G. E. (2021). How to Estimate a VAR after March 2020. *Journal of Applied Econometrics (forthcoming).*

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions—five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38.

Luciani, M. (2014). Forecasting with approximate dynamic factor models: The role of non-pervasive shocks. *International Journal of Forecasting*, 30:20–29.

Luciani, M. (2017). Large-dimensional dynamic factor models in real-time: A survey. In Mazzi, G. L. and Ozyildirim, A., editors, *Handbook on Cyclical Composite Indicators*, pages 429–451. Eurostat.

Luciani, M., Pundit, M., Ramayandi, A., and Veronese, G. (2018). Nowcasting Indonesia. *Empirical Economics*, 55(2):597–619.

Luciani, M. and Ricci, L. (2014). Nowcasting Norway. *International Journal of Central Banking*, 10(4):215–248.

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.

Matheson, T. D. (2010). An analysis of the informational content of New Zealand data releases: The importance of business opinion surveys. *Economic Modelling*, 27(1):304–314.

Modugno, M. (2013). Now-casting inflation using high frequency data. *International Journal of Forecasting*, 29(4):664–675.

Modugno, M., Soybilgen, B., and Yazgan, E. (2016). Nowcasting Turkish GDP and news decomposition. *International Journal of Forecasting*, 32(4):1369–1384.

Ng, S. (2021). Modeling macroeconomic variations after COVID-19. Technical report, National Bureau of Economic Research.

Ouysse, R. (2011). Comparison of Bayesian moving average and principal component forecast for large dimensional factor models. Discussion Papers 2012-03, School of Economics, The University of New South Wales.

Parigi, G. and Golinelli, R. (2007). The use of monthly indicators to forecast quarterly GDP in the short run: an application to the G7 countries. *Journal of Forecasting*, 26(2):77–94.

Parigi, G. and Schlitzer, G. (1995). Quarterly forecasts of the Italian business cycle by means of monthly indicators. *Journal of Forecasting*, 14(2):117–141.

Richardson, A., van Florenstein Mulder, T., and Vehbi, T. (2021). Nowcasting GDP using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting*, 37(2):941–948.

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.

Rusnák, M. (2016). Nowcasting Czech GDP in real time. *Economic Modelling*, 54:26–39.

Sargent, T. J. and Sims, C. A. (1977). Business cycle modeling without pretending to have too much a-priori economic theory. In et al., C. S., editor, *New Methods in Business Cycle Research*. Federal Reserve Bank of Minneapolis.

Schorfheide, F. and Song, D. (2020). Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic. *International Journal of Central Banking (forthcoming)*.

Siliverstovs, B. (2012). Keeping a finger on the pulse of the economy: Nowcasting Swiss GDP in real-time squared. Kof working paper, KOF Swiss Economic Institute, ETH Zurich.

Soybilgen, B. and Yazgan, E. (2021). Nowcasting US GDP using tree-based ensemble models and dynamic factors. *Computational Economics*, 57:387–417.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179.

Stock, J. H. and Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, 30(4):481–493.

Stock, J. H. and Watson, M. W. (2017). Twenty Years of Time Series Econometrics in Ten Pictures. *Journal of Economic Perspectives*, 31(2):59–86.

Yiu, M. S. and Chow, K. K. (2010). Nowcasting Chinese GDP: information content of economic and financial data. *China Economic Journal*, 3(3):223–240.

Zhang, Q., Ni, H., and Xu, H. (2023). Nowcasting Chinese GDP in a data-rich environment: Lessons from machine learning algorithms. *Economic Modelling*, 122:106204.