



EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Jan Linzenich, Baptiste Meunier

Nowcasting made easier: a toolbox for economists

No 3004

Disclaimer: This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

Electronic copy available at: <https://ssrn.com/abstract=5066456>

Abstract

We provide a versatile nowcasting toolbox that supports three model classes (dynamic factor models, large Bayesian VAR, bridge equations) and offers methods to manage data selection and adjust for Covid-19 observations. The toolbox aims at simplifying two key tasks: creating new nowcasting models and improving the policy analysis. For model creation, the toolbox automatizes testing input variables, assessing model accuracy, and checking robustness to the Covid period. The toolbox is organized along a structured three-step approach: variable pre-selection, model selection, and Covid robustness. Non-specialists can easily follow these steps to develop high-performing models, while experts can leverage the automated tests and analyses. For regular policy use, the toolbox generates a large range of outputs to aid conjunctural analysis like news decomposition, confidence bands, alternative forecasts, and heatmaps. These multiple outputs aim at opening the "black box" often associated with nowcasts and at gauging the reliability of real-time predictions. We showcase the toolbox features to create a nowcasting model for global GDP growth. Overall, the toolbox aims at facilitating creation, evaluation, and deployment of nowcasting models. Code and templates are available on GitHub: https://github.com/baptiste-meunier/Nowcasting_toolbox.

JEL codes: C22, C51, C52, C53, C55.

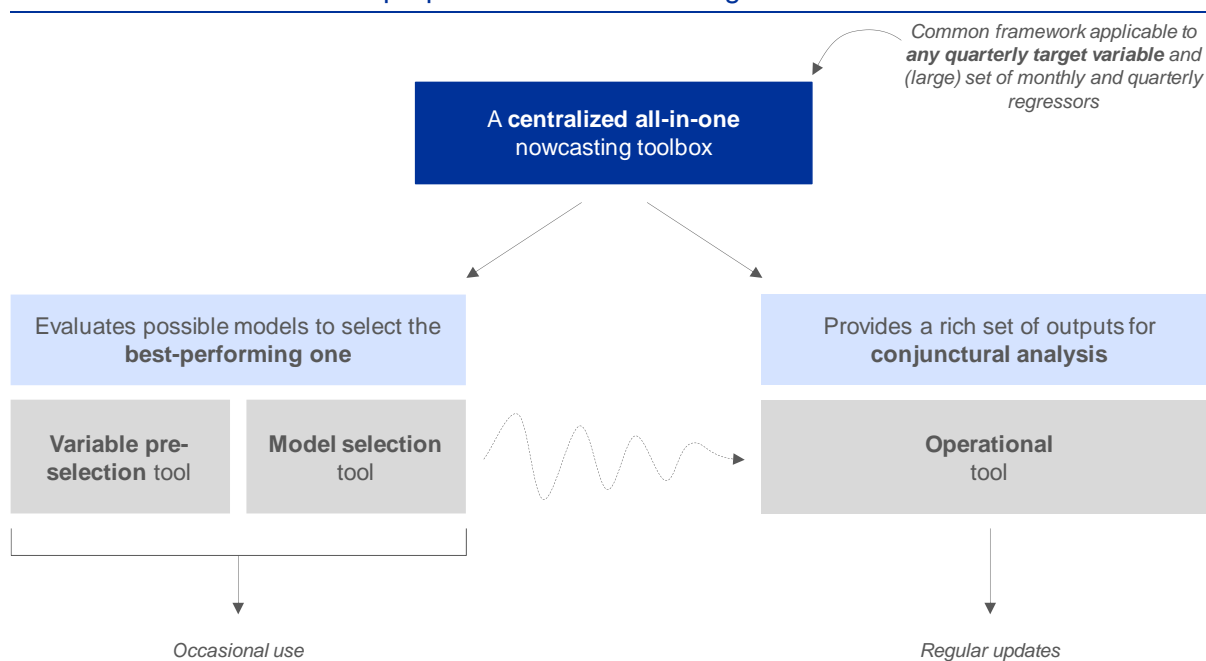
Keywords: Dynamic factor model, Bayesian VAR, bridge equation, large dataset, forecasting.

Non-technical summary

Nowcasting has become essential for policy institutions, but existing tools often lack flexibility while generally focusing on a single model class and often putting a limited emphasis on the policy dimension. Against this background, this paper offers a user-friendly nowcasting toolbox designed to assist applied economists in creating, evaluating, and using nowcasting models. The toolbox simplifies two key tasks: building new nowcasting models from scratch and enhancing the policy analysis through detailed insights into model predictions (**Figure N1**). The toolbox builds on a long-standing effort in the External Developments division of the ECB to design a nowcasting environment that is adaptable, easy to use, and provide model-based information that are relevant for policy. This work involved many ECB colleagues.¹

Figure N1

Schematic overview of main purposes of the nowcasting toolbox



Source: Authors.

For policy analysis, the toolbox offers a rich set of outputs to open the “black box” of nowcasts: point forecasts, news decomposition, contributions, and a heatmap of input variables. These outputs aim at better informing the conjunctural narrative by shedding light on the underlying

¹ Notably S. Delle Chiaie, F. Kurcz, and G. Perez-Quirós who conducted work on previous iterations of nowcasting tools in the division.

drivers of the model predictions. The toolbox provides another set of outputs centred on gauging the reliability of the predictions: confidence intervals, index of directional accuracy, and range of alternative models.

The second mode of the nowcasting toolbox relates to the creation of nowcasting models from scratch. The toolbox simplifies this task by automatizing 1) tests of the predictive power of input variables, 2) tests of the out-of-sample accuracy of model specifications, and 3) checks of robustness to the Covid period. For model creation, the toolbox is organized along a structured three-step approach of: variable pre-selection, model selection, and Covid robustness. This approach, which formalizes the steps generally followed by forecasters, offers a practical start-to-finish and step-by-step guide for *non-specialists* to create a tailored high-performing model. *Specialists* can still benefit from the automation and the flexibility of the toolbox.

The toolbox is designed as a versatile tool encompassing various model classes and methods for data management. The toolbox supports the three most prominent model classes for nowcasting: dynamic factor models (Bánbura and Modugno, 2014), large Bayesian vector auto-regression (Cimadomo et al., 2022), and combination of bridge equations (Bánbura et al., 2023). In addition, the toolbox also includes three methods for variable selection: correlation-based, regression-based, and an iterative forward selection algorithm. It also provides three possible ways to deal with Covid-19: dummies, outlier-correction, and deletion of related observations. Importantly, the toolbox can deal with big data since pre-selection methods allow to discriminate the most informative regressors, and since model classes can accommodate for many input variables.

The paper presents an example application of the toolbox to create a nowcasting model for quarterly global GDP growth. Starting from a set of 540 variables, it shows how following the structured three-step approach leads to a model with high performances.

Code and templates are available: https://github.com/baptiste-meunier/Nowcasting_toolbox.² Appendices to this paper provide a detailed guide on how to use them. We intend to extend the toolbox towards new models and capabilities in future releases – in that respect, users are invited to provide feedback.

² When using models *via* the nowcasting toolbox, users are kindly requested to cite the original papers: Bańbura and Modugno (2014) for the dynamic factor model; Cimadomo et al. (2022) for the large Bayesian vector auto-regression, and Bańbura et al. (2023) for the combination of bridge equations.

Introduction

“All models are wrong; some models are useful.”

(George Box)

Nowcasting has emerged as a key tool for policy institutions, enabling them to obtain early estimates of key economic indicators before official data is released. This becomes especially valuable during crisis, where timely policy responses are critical. However, as the economy evolves and new data becomes available, nowcasting models require periodic revisions to maintain accuracy – as seen in the recent update of the New York Fed Staff Nowcast following a two-year hiatus (Almuzara et al., 2023). Despite existing replication codes, there is a lack of a comprehensive toolkit that combines the ability to produce regular policy updates with the option to create, revise and evaluate nowcasting models.

To address these limitations, this paper introduces a nowcasting toolbox aimed at providing a unified, adaptable, and user-friendly tool for building and refining nowcasting models. The toolbox simplifies two key nowcasting duties: creating new models from scratch and improving policy analysis. In the first duty, the toolbox helps building high-performing models by automating tasks like variable selection, testing of model specifications, and robustness checks. For policy analysis, the toolbox delivers a rich set of outputs that aims at clarifying model predictions and their reliability. A first set of outputs open the “black box” of nowcasts and provide insights on the economic forces behind model predictions. Another set of outputs informs the user on the degree of reliability to attach to predictions. The tool can be adapted to any mixed-frequency model with a quarterly target variable and a set of (many) potential regressors at monthly and quarterly frequencies. The toolbox builds on a long-standing effort from ECB staff in the External Developments division of the ECB, over several years, to create a nowcasting tool tailored for their own needs.³ In that sense, this is a long-matured toolbox created by economists for economists.

The toolbox integrates various model classes and methods to manage data selection or adjust for Covid-19 observations in a unified and standardized code. It supports three model classes:

³ Notably S. Delle Chiaie, F. Kurcz, and G. Perez-Quirós who conducted work on previous iterations of nowcasting tools in the division.

dynamic factor model (Bánbura and Modugno, 2014), large Bayesian vector auto-regression (Cimadomo et al., 2022) and a combination of bridge equations *à la* Bánbura et al. (2023). The toolbox standardizes inputs and outputs with each model, allowing the user to switch easily across models. The toolbox also incorporates three pre-selection techniques (Efron et al., 2004); Bair et al., 2006; Fan and Lv, 2008) and three options for correcting for the Covid period (outlier correction, dummy variables, and deleting Covid-related observations). These various tools make the toolbox easily adaptable to various frameworks: notably, it can accommodate for big data as it offers tools to pre-select the most informative regressors and model classes that can work with many input variables.

The primary contribution of this paper is to provide the toolbox for public use with codes and templates available on GitHub: https://github.com/baptiste-meunier/Nowcasting_toolbox.⁴ Compared with other codes available in the literature (e.g., Anesti et al., 2018; Bok et al., 2018; de Valk et al., 2019; Hopp, 2021; Mosley et al., 2023), this toolbox combines the ability to produce regular policy updates with the option to create, revise and evaluate models. While the literature generally offers only limited intuition into how to build a model from scratch, one key contribution of the toolbox is to organize model creation along a structured three-step approach of 1) variable pre-selection, 2) model selection, and 3) Covid robustness. This approach formalizes the process that a forecaster generally follows when creating a model. It provides a step-by-step methodology to guide users – especially *non-specialists* – through the development of performing models. *Specialists* can take advantage of the time-saving automated analyses and tests. In addition, the toolbox puts the emphasis on policy use and provides a richer set of outputs than available in other codes of the literature. Finally, the toolbox offers a one-stop shop for various model classes, pre-selection techniques, and Covid corrections – where codes for the different techniques are generally scattered across different programs. This effort of standardization allows to easily run comparison across model classes as well as assessing in an automatic way if corrections for Covid period improve performances.

We apply our toolbox and structured approach to create a nowcasting model for the quarterly growth rate of global GDP. Starting with a large dataset of 540 series covering all aspects of the economy, we show how the toolbox assists in pre-selecting variables with the highest

⁴ When using models *via* the nowcasting toolbox, users are kindly requested to cite the original papers: Bánbura and Modugno (2014) for the dynamic factor model; Cimadomo et al. (2022) for the large Bayesian vector auto-regression, and Bánbura et al. (2023) for the combination of bridge equations. We intend to extend the toolbox towards new models and capabilities in future releases – in that respect, users are invited to provide feedback.

predictive power and better timeliness. We further show test many specifications across three model classes and check if correcting for the Covid period yield better performances for pseudo real-time out-of-sample forecasts after 2021. We build a *new* nowcasting model that improves markedly over an *existing* model that was created in a more ad-hoc fashion, lowering the Root Mean Squared Error (RMSE) by 66% and increasing the Forecast Directional Accuracy (FDA) by 12 percentage points.

In summary, we provide an all-in-one tool for applied economists interested in building and deploying nowcasting models. **Section 1** presents the main purposes of the toolbox, introducing the structured approach to build nowcasting models and discussing the outputs for policy use. **Section 2** details the technical features of the toolbox. **Section 3** describes the outputs for policy analysis, giving examples of the insights each can provide. We apply the toolbox to build building a nowcasting model for quarterly global GDP growth in **section 4**. In addition, **Appendix A** contains a step-by-step guide on how to use the toolbox for model building; **Appendix B** details the procedure to run the toolbox for regular policy updates; **Appendix C** provides supplementary material.

1 Main purposes

The toolbox provides an all-in-one framework for nowcasting. The term “nowcasting” relates to the prediction of the present, the very near future and the very recent past. It has been crafted by Bańbura et al. (2011) when this topic started to gain traction in academic research following seminal papers by Evans (2005) and Giannone et al. (2008). Besides academia, nowcasting has been actively used in many policy institutions to assess economic developments in real time (Angelini et al., 2011; Bok et al., 2018; Dauphin et al., 2022). The underlying idea of nowcasting is to use timely information to obtain an early estimate of key economic variables before the official data is released. The approach is meant for quarterly variables published with long delays, typically Gross Domestic Product (GDP).⁵ The toolbox is designed for mixed-frequency set-ups where the target variable is at quarterly frequency and regressors can be monthly or quarterly – as is usually the case in nowcasting (e.g., Giannone et al., 2008; Angelini et al., 2011; Schorfheide and Song, 2015).

The toolbox serves two primary purposes: 1) build new nowcasting models from scratch, and 2) provide a rich set of outputs to aid policy analysis. The first purpose is centred on the evaluation of model performances across a large set of possible input variables and model parameters. The toolbox provides a set of automated tools to assess the predictive power of input variables, the out-of-sample accuracy of model specifications, and the potential improvements from Covid-specific corrections. This facilitates the selection of high-performing models by the user. This use is generally occasional, whenever the user wants to build a new nowcasting model – or to review a current model as the toolbox can be used to evaluate an existing model. The second use of the toolbox is for regular updates of the predictions of the nowcasting model. In this operational use, the toolbox delivers a rich set of outputs designed to better understand model predictions and their reliability. The two purposes of the toolbox are intertwined, as the model built *via* the first leg should be the one used for regular predictions in the second leg. As such, the toolbox can serve as a one-stop shop for creating, evaluating, and deploying a nowcasting infrastructure.⁶

⁵ For example, the first official estimate of GDP in the US or in the UK is published approximately one month after the end of the reference quarter.

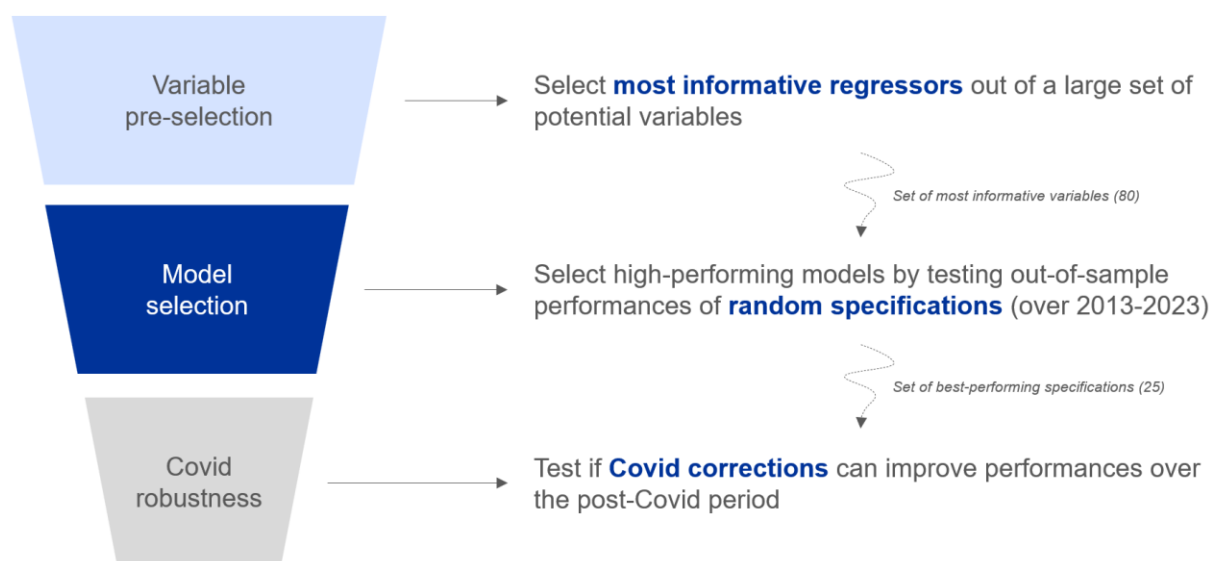
⁶ The first use (model building) is carried out occasionally and is typically triggered by major events – e.g., the Covid-19 crisis that pushed the NY Fed to review its nowcasting framework, see Almuzara et al. (2023) – or the emergence of new datasets – e.g., high-frequency indicators (Bricongne et al., 2020) or satellite data (Bricongne et al., 2021; d’Aspremont et al., 2024). In the second use (policy outputs), updates of model predictions are

1.1 Build a nowcasting model

A key interest of the toolbox is task automation. The underlying idea of model building with the toolbox is to follow an empirically driven process where the toolbox automatically tests which regressors and model specifications perform best, along user requirements. This implies the process is steered by the user to achieve a model tailored to specific needs. While *non-experts* can follow empirical results, the toolbox also provides complementary information to aid the user from judgment calls. For example, model selection might not only consider the predictive accuracy, but also that input variables account for all aspects of the economy. *Expert* users can therefore still benefit from the automation provided by the toolbox, while steering more strongly the construction of a model.

Figure 1

The proposed three-step approach for building nowcasting models



Source: Authors.

Notes: Values in brackets relate to the number of variables, number of high-performing models, and out-of-sample periods selected when building a nowcasting model for world (excl. EA) GDP growth. See [section 4](#) for more details.

typically carried out every few days: e.g., [New York Fed Staff Nowcast](#) and the [UNCTAD Nowcast](#) are updated every week.

The toolbox offers a structured three-step approach to building a nowcasting model.

The three steps of variable pre-selection, model evaluation, and Covid-19 robustness are illustrated in [Figure 1](#) and detailed below (see [Appendix A](#) for a more detailed guide):

Step 1. Variable pre-selection: to start, the user gathers a target variable, like GDP, and a wide range of potential predictors. The toolbox then helps narrow down the dataset to the variables with the highest predictive power – based on three established methods from the literature (see [section 2.1](#)). The toolbox provides complementary information on the type, timeliness, and frequency of the predictors, that help the user decide which variables to keep.

Step 2. Model selection: using the refined set of variables from step 1, the toolbox automatically evaluates various model specifications by testing their *pseudo* real-time out-of-sample predictive performances.⁷ Users can either define model settings or let the toolbox automatically vary the settings – in this case, settings are drawn within bounds defined by the user. Settings relate to both inputs (e.g., start date of estimation, variables) and model parameters (e.g., number of factors and lags – see [section 2.2](#)).⁸ For each specification, the toolbox computes performance metrics (Root Mean Squared Error, RMSE, and Forecast Directional Accuracy, FDA) over the evaluation period set by the user.⁹ These metrics, together with additional information provided by the toolbox (e.g., type of variables, accuracy by sub-period) help the user to identify the best model(s).¹⁰

⁷ A real-time set-up recreates the information set that would have been available to a forecaster at a specific point in the past, mirroring which variables were available and their values. This last point requires historical versions (vintages) to account for revisions. When vintages are not available, the set-up is *pseudo* real-time – which is the case in most nowcasting exercises and this toolbox.

⁸ The list of settings that can be changed is provided in [Appendix C](#). To some extent, model selection pertains to parameter uncertainty – which the literature has shown can be addressed *via* Bayesian techniques (e.g., Giannone et al., 2015). But it should be noted that model selection goes beyond parameter uncertainty and tackles settings related to input data (e.g., start date of the estimation and variables) which cannot be covered by Bayesian techniques. In addition, some model parameters are generally not tackled by Bayesian techniques but rather set by the user (e.g., number of lags in a BVAR).

⁹ Formally, the Root Mean Squared Error is $RMSE = \sqrt{\sum_{t=1}^N (y_t - \hat{y}_t)^2 / N}$ where y_t and \hat{y}_t are respectively the actual value and the predicted value for observation t , and N is the total number of observations. The RMSE measures the accuracy of point forecast, i.e., on average how far does the model prediction falls from the actual value. Formally, the Forecast Directional Accuracy is $FDA = \sum_{i=1}^N I(t) / N$ where $I(t)$ takes value 1 if $(y_t - y_{t-1})(\hat{y}_t - \hat{y}_{t-1}) > 0$ and 0 otherwise. The FDA measures how often the model correctly predicts the direction of the target variable (increase or decrease).

¹⁰ In addition to RMSE and FDA over the full sample, the toolbox also provides RMSE and FDA by sub-periods (pre-Covid, Covid, and post-Covid). RMSE and FDA are also provided for each month of the quarter (i.e., for predictions made in the 1st, 2nd, and 3rd months of the quarter). Finally, while RMSE and FDA are averaged over periods, the toolbox also provides predictions for each out-of-sample observation.

Step 3. Covid robustness: using the best model(s) identified in step 2, the toolbox tests how the model(s) handle(s) Covid observations, as this period presents unique challenges for estimation. The toolbox automatically applies three Covid-specific corrections (see [section 2.3](#)) and compares the predictive performances of the model(s) with and without corrections. While step 2 should explore a long sample (typically a full business cycle), this step focuses on the post-Covid period with a view of ensuring that ensures that the model(s) retain(s) their predictive ability even after pandemic disruptions.

The approach formalizes the steps that a forecaster would generally follow when building a nowcasting model. A forecaster would start by choosing variables relevant for nowcasting. The variable pre-selection step automatizes it by quantifying the predictive power of each potential predictor, as well as providing information on the timeliness and type of variable. Once variables are pre-selected, the forecaster would test a few model specifications to assess which one performs best. The model selection step caters for this, by running many tests in an automatic way – this is especially potent when the user can use virtual machines to let the code running during a long time on distant servers.¹¹ After that, the forecaster would check how the model performs on the recent period, given how the Covid-19 period disrupted the estimation of econometric models (Lenza and Primiceri, 2022) and the signalling power of some variables (de Bondt and Saiz, 2024). The Covid robustness step is designed to check in whether correcting for the Covid period improves performances or not in an automated way.

There are possible shortcuts to the suggested approach as variable pre-selection and Covid robustness steps both aim at improving performances but can be skipped. In other words, only step 2 is required. The purpose of steps 1 and 3 is to facilitate the process and improve the performances of the model. Variable pre-selection is grounded in the literature showing that model forecasts are more accurate when selecting fewer but more informative predictors (Boivin and Ng, 2006; Rünstler, 2016). The empirical literature shows that pre-selection can enhance the out-of-sample accuracy of factor models by around 20% (Bai and Ng, 2008; Chinn et al., 2023). In the structured approach, pre-selection also aims at easing model selection by focusing on fewer but more relevant regressors. As mentioned above, the

¹¹ The idea of the model selection step is not to cover all possible models but rather to automatize the process followed naturally by a forecaster. In this view, it relates to an automatized “trial-and-error” approach rather than an extensive grid search. In general, model selection should be run iteratively: a first run should be based on a broad range of possible settings; then, based on the results of the first run, the second run should narrow the range of possible settings; and so forth and so on until no improvements are visible.

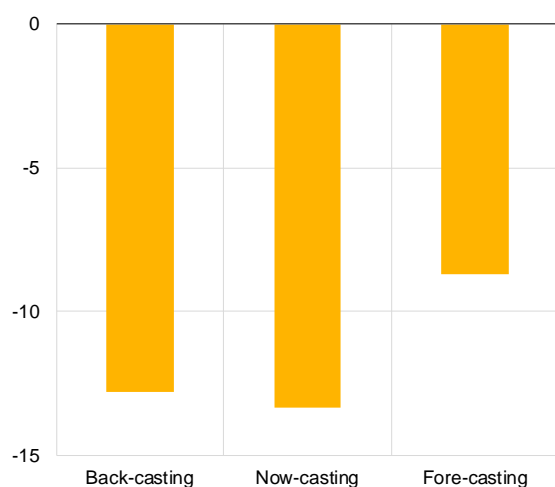
model selection step will vary input variables across the different specifications: pre-selection ensures that such variables are drawn from a pool of highly relevant regressors.¹² This sequential process, where the dataset is first narrowed down in pre-selection before being further refined in model selection (i.e., a “funnel strategy”) is empirically found to improve best performances by around 10% compared with a “strict pre-selection” where the set of input variables is fixed after pre-selection (**Figure 2**, panel a). As regards Covid correction, the underlying idea is that some models might experience difficulties when the peculiar Covid period is included in the estimation sample (Zhang et al., 2021; Lenza and Primiceri, 2022). Empirical tests show that Covid corrections can improve out-of-sample performances over the post-Covid period (2021-2023) by 10% to 30% on average (**Figure 2**, panel b) although with a degree of heterogeneity and with caveats due to the short sample considered (2021-2023).

Figure 2

Accuracy gains from variable pre-selection and Covid corrections

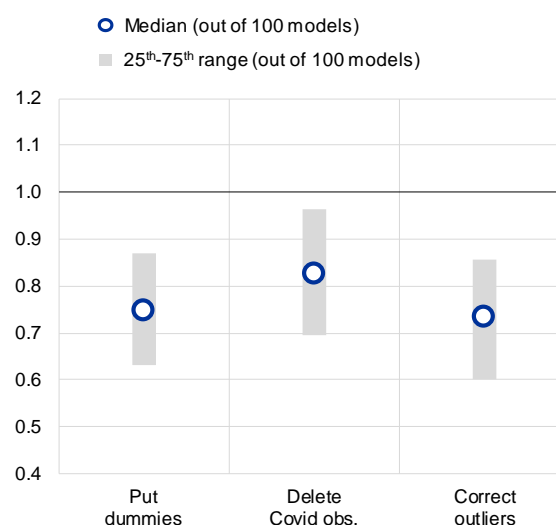
a) Variable pre-selection

(out-of-sample RMSE over 2015-2023 of the best model based on *funnel strategy* relative to the best model based on *strict pre-selection*)



b) Covid corrections

(out-of-sample RMSE over 2021-2023 of models *with* Covid correction relative to models *without* Covid correction)



Sources: Bloomberg, Haver, S&P Global, and authors' calculation.

Notes: Both panels relate to pseudo real-time predictions of global GDP growth. In panel a), the *funnel strategy* relates to the sequential process of 1) narrowing down a set of potential predictors to the 40 with highest predictive power, and 2) extracting 10 (random) out of those 40. The *strict pre-selection* relates to selecting directly the 10 variables with highest predictive power. Results are based on running 100 random dynamic factor model specifications on the 10 variables obtained via *funnel strategy* and *strict pre-selection*. The chart shows the accuracy of the best model obtained with *funnel strategy* relative to the best model obtained with *strict pre-selection*. Negative value indicates that *funnel strategy* produces a more accurate best model than *strict pre-selection*. Panel b) is based on 100 random dynamic factors model (DFM) specifications run *with* and *without* Covid correction. For each DFM specification, a relative out-of-sample RMSE is computed as the ratio of RMSE of the model *with* Covid correction to the RMSE of the same model *without* Covid correction. The chart shows the distribution (median, 25th and 75th percentiles) of the 100 relative out-of-sample RMSEs. A value below 1 indicates that models *with* correction outperforms the same model *without* correction. Values refer to a nowcasting horizon (i.e., a prediction of current quarter).

¹² The underlying idea is that pre-selection techniques can be flawed, notably as they don't consider timeliness, and are then used for pre-screening rather than for a strict pre-selection.

1.2 Provide outputs for policy analysis

The toolbox aims at providing a rich set of complementary outputs to open the “black box” of nowcasting. When used of regular policy updates, the toolbox can be used easily to generate model predictions (see [Appendix B](#)). The purpose is to facilitate policy analysis, with the underlying idea to offer insights on what’s happening “beneath the surface” of the model. The outputs can be grouped in two broad categories serving different goals: 1) providing the narrative underlying model predictions, and 2) assessing the reliability of model predictions. In details (see also [section 3](#) for illustrated examples):

- **Outputs aimed at providing the narrative underlying model prediction:** they are designed to understand the economic drivers shaping model forecasts, and are composed of three types of output:
 - **News decomposition:** the toolbox quantifies the impact of each recent data releases, showing how predictions change as new data becomes available. The underlying idea is that predictions are based on extrapolated values of the input variables if they are not yet released: whenever the *ex-post* data release differs from the *ex-ante* extrapolation, the “news” affects the prediction.¹³ It follows the seminal work of Bańbura and Modugno (2014) and formalizes how forecasters traditionally predict, by monitoring many economic series, forming expectations about them, and then revising the assessment of the state of the economy whenever new input data differ from their previous expectations.
 - **Contributions:** while news decomposition offers a *dynamic* view of how data are affecting forecasts, contributions give a *static* view. If exact contributions cannot be extracted – for instance when Kalman filtering techniques are used, we approximate the contributions by computing the impact of all data releases over recent months. With all due caveats due to this approximation, contributions provide insights of how input variables drive the forecast.

¹³ For example, in a dynamic factor model, Kalman filtering techniques will provide forecasts not only for the target variable but also for each input variable. Please refer to [section 2.2](#) for more details.

- **Heatmap of input variables:** this provides a bird's-eye view at how input variables deviate from their long-term mean.¹⁴ It can be particularly valuable if model predictions seem off – a situation that will arise no matter how accurate the model is (Engle and Brown, 1986; Makridakis et al., 2009). In this situation, the heatmap offers a possibility to ignore the model and go directly to the raw source (i.e., input variables) to understand the state of the economy.
- **Outputs aimed at assessing the reliability of model predictions:** they help the user to determine the degree of confidence attached to a prediction. It is particularly helpful as the user must generally form an opinion about the state of the economy amid multiple sources of information that often provide mixed, and at times contradictory, signals. It is composed of three outputs:
 - **Confidence band:** the toolbox computes the confidence band associated with the point forecast based on the mean absolute error of past forecasts over the last ten years in line with Reifschneider and Tulip (2019). The confidence bands are re-assessed dynamically, meaning the confidence attached to a prediction made on the 1st day of a given month will be different than the confidence for a prediction on the 25th day of the same month. In addition to confidence bands, the toolbox computes the FDA over the past ten years which allows quantifying the uncertainty surrounding the *directional* forecast – while confidence bands quantify the uncertainty on the *point* forecast.
 - **Share of available data:** another way to gauge the uncertainty surrounding a prediction is whether this prediction is based on actual data or on data that are extrapolated by the model. Intuitively, a lower share of actual data means that the prediction is likely to be more affected by data releases.
 - **Range of alternative models:** the toolbox computes forecasts from alternative models that slightly differ from the main nowcasting model. Those alternative models are derived automatically by the toolbox, by disconnecting one or two group(s) of variables from the information set. The interest of those alternative forecasts is to assess what would be the prediction if some variables are excluded. This is particularly useful when some variables provide an erroneous

¹⁴ Formally, the heatmap present the z-scores of input variables, i.e., how much the current observation differs from the long-term average – expressed in standard deviations.

signal of economic conditions – as was somewhat the case with Purchasing Managers' Indices during the pandemics (de Bondt and Saiz, 2024).¹⁵

The goal of offering many outputs is to maintain the policy relevance of the nowcasting toolbox, regardless of how reliable model predictions may be. Even a model that display good nowcasting performances on average can be off at times due to the complexity of economic systems and unforeseeable events (Engle and Brown, 1986; Makridakis et al., 2009). With many outputs, some of them could still prove useful even when the model predictions are off. Notably, the heatmap of input variables does not rely on the model and can therefore help users understand the state of the economy when forecasts seem unreliable. In addition, a reason for odd predictions is generally that a few key input series are giving an erroneous signal: in this case, the alternative models allow to exclude faulty variables. All-in-all, the toolbox intends to provide the maximum of information to the user for policy analysis.

¹⁵ The range of alternative models and the confidence band are computationally intensive. The toolbox includes the option to turn off these outputs to better control the computation time.

2 Technical features

Another key interest of the toolbox is to incorporate various techniques for modelling and data management in a unified code, standardizing inputs outputs of each model. It supports the three most prominent model classes in nowcasting (**section 2.1**): dynamic factor model (Bánbura and Modugno, 2014), large Bayesian vector auto-regression (Cimadomo et al., 2022) and a combination of bridge equations à la Bánbura et al. (2023). The toolbox also incorporates three established pre-selection techniques from the literature (**section 2.2**): the Least Angle Regression (Efron et al., 2004), a t-stat-based method (Bair et al., 2006), and the Sure Independence Screening (Fan and Lv, 2008). The toolbox finally accommodates for three options for correcting for the Covid period (**section 2.3**): outlier correction, dummy variables, and deleting Covid-related observations.

The wide range of techniques makes the toolbox flexible and adaptable to specific contexts and user preferences. For example, it appears in the literature that the best-performing model class depend on the context: dynamic factor models outperform bridge equations in some contexts (Chernis and Sekkel, 2017; Guichard and Rusticelli, 2011) while the opposite can also apply (Bánbura and Saiz, 2020; Soybilgen and Yazgan, 2018). Similarly, the most accurate variable pre-selection and Covid correction methods can depend on the dataset under consideration (Schorfheide and Song, 2021; Chinn et al., 2023). Importantly, the toolbox can accommodate for big data as it offers tools to pre-select the most informative regressors and model classes that can work with many input variables.

2.1 Variable pre-selection methods

The variable pre-selection step narrows the set of potential regressors to those with highest predictive power. The rationale for that the literature has shown that when forecasting with a high-dimensional dataset, factor models are significantly more accurate when selecting fewer but more informative predictors (Bai and Ng, 2008). On a theoretical ground, Boivin and Ng (2006) find that larger datasets lead to poorer forecasting performances when idiosyncratic errors are cross-correlated or when the variables with higher predictive power are dominated. This is confirmed in empirical studies (Jardet and Meunier, 2022; Barbaglia et al., 2023; Chinn et al., 2023; Wang et al., 2023).

The basic principle of variable pre-selection is to rank all potential regressors based on their (assessed) predictive power. Formally, the initial dataset is $X_t = (x_{1,t}, x_{2,t}, \dots, x_{N,t})$ with N , the number of variables, relatively large. The toolbox offers three techniques to rank regressors in X_t by their predictive power:

- **Sure Independence Screening** (SIS) of Fan and Lv (2008): regressors are ranked based on their marginal correlation with the target predictor. Fan and Lv (2008) provide theoretical ground for their approach by demonstrating that it has the sure screening property that *“all important variables survive after applying a variable screening procedure with probability tending to 1”*. This approach has been used for nowcasting in Ferrara and Simoni (2019) or Proietti and Giovannelli (2021).
- **T-stat-based** of Bair et al. (2006): each regressor $x_{i,t}$ is ranked based on the absolute value of the t-statistic associated with its coefficient estimates in a univariate regression of $x_{i,t}$ on the target variable y_t . The univariate regression also includes four lags of the dependent variable to control for the endogenous dynamics of the target variable. While originating in genetic studies, this technique has found its way to economics for example in Jurado et al. (2015).
- **Least-Angle Regression** (LARS) of Efron et al. (2004): while the two methods above are based on univariate relationships of the regressors with the target variable, LARS accounts for the presence of other predictors.¹⁶ LARS is an iterative forward selection algorithm. Starting with no predictors, it adds the predictor x_i most correlated with the target variable y and then move the coefficient β_i in the direction of its least-squares estimate so that the correlation of x_i with the residual $(y - \beta_i x_i)$ gets lower. It does so until another predictor x_j has similar correlation with $y - \beta_i x_i$ than x_i . At this point, x_j is added to the active set and the procedure continues, now moving both coefficients β_i and β_j in the direction of their least-squares estimates until another predictor x_k has as much correlation with the residual (now $y - \beta_i x_i - \beta_j x_j$). This has been used in various nowcasting set-ups by Schumacher (2010), Bulligan et al. (2015) or Falagiardia and Sousa (2015).

¹⁶ Other multivariate methods could also be considered such as Jarociński and Maćkowiak (2017).

2.2 Model classes

Dynamic Factor Model (Bańbura and Modugno, 2014)¹⁷

Dynamic Factor Models (DFMs) have become central in nowcasting as they offer a parsimonious approach to understand underlying forces shaping the economy. At its core, the DFM is designed to distil the economic data into a few underlying factors, reducing the dimensionality of large datasets. This model posits that a set of economic variables $Y_t = (y_{1,t}, y_{2,t}, \dots, y_{n,t})$ can be decomposed into a structured component driven by few latent factors $F_t = (f_{1,t}, f_{2,t}, \dots, f_{r,t})$ and an idiosyncratic noise component ϵ_t as in equation 1. Λ is the loading matrix quantifying the relationship between the observable variables and the latent factors. DFMs are a way to address the large p , small n paradigm by providing a parsimonious representation of the time series since r , the number of factors in F_t , is generally much smaller than n , the number of variables in Y_t . This approach is grounded on the insight that economic fluctuations are driven by a few common sources, a concept dating back to the seminal work of Burns and Mitchell (1946).

$$(1) \quad Y_t = \Lambda F_t + \epsilon_t$$

A notable challenge for the estimation of DFMs is the accurate inference of the unobserved factors and loadings from the observed data.¹⁸ A pivotal advancement has been adoption of a state-space representation (Giannone et al., 2008). Equations 2 and 3 are *transition* equations that describe the dynamics of the model, accounting for serial correlation and persistence.¹⁹ Together with the *measurement* equation (equation 1) that links data to the unobserved factors, they form a state-space representation. This form allows the use of the Kalman filter to estimate latent factors.²⁰ The state-space form makes DFMs especially suited for nowcasting as Kalman filtering techniques are suited to handle data irregularities such as

¹⁷ Users are kindly requested to cite the original paper of Bańbura and Modugno (2014) when using this model. Please refer directly to the paper for an extensive description.

¹⁸ Early factor models (Forni et al., 2000; Stock and Watson, 2002) were based on principal components. Key methodological advancements came from using likelihood-based methods and the expectation-maximization algorithm for iterative estimation (Doz et al., 2011; Doz et al., 2012). Notably, this allowed to handle larger datasets: Doz et al. (2012) showed that maximum likelihood estimators are consistent in high-dimensional datasets and robust to cross-sectional misspecifications and serial correlation of the idiosyncratic components.

¹⁹ u_t and e_t are independently and identically distributed (i.i.d.) and drawn from multivariate normal distributions with mean 0. To maintain a parsimonious empirical specification, idiosyncratic components are generally assumed to be cross-sectionally orthogonal.

²⁰ The state-space form is crucial for the Kalman filter, which is designed to estimate the state of a linear dynamic system from a series of incomplete and noisy measurements.

mixed frequencies and the non-synchronicity of data releases.²¹ In addition, the Kalman filter handle the uncertainties and noise present in real-world data by weighing the uncertainty in previous estimates and the uncertainty in new data to minimize the overall estimation error. Another interesting feature is that Kalman filtering treats the latent factors as unobserved state variables that evolve over time and can update dynamically the estimates as new information becomes available (Bańbura et al., 2013).²² The DFM in the toolbox also allows to use block structure to identify the factors as in Delle Chiaie et al. (2022).

$$(2) \quad F_t = A_1 F_{t-1} + A_2 F_{t-2} + \dots + A_p F_{t-p} + u_t$$

$$(3) \quad \epsilon_t = \Phi \epsilon_{t-1} + e_t$$

The DFM has become a cornerstone in the literature and has been used extensively in nowcasting exercises. This has been the case for example for GDP in the US (Liebermann, 2014; Bok et al., 2018), the euro area (Bańbura et al., 2011; Angelini et al., 2011; Dauphin et al., 2022), the UK (Anesti et al., 2018), Japan (Bragoli, 2017), as well as emerging economies like Brazil (Dahlhaus et al., 2018), India (Bragoli and Fosten, 2018), Mexico (Caruso, 2018), China (Yiu and Chow, 2010), and South Africa (Kabundi et al., 2016). Besides GDP, DFMs have also been used to nowcast world trade (Guichard and Rusticelli, 2011; Martínez-Martín and Rusticelli, 2021), household consumption (Tarsidin et al., 2018), inflation expectations (Palardy and Ovaska, 2015), or commodity prices (Kagraoka, 2016). This highlights DFMs' adaptability to various data structures and economic contexts.

Large Bayesian Vector Auto Regression (Cimadomo et al., 2022)²³

Vector Auto Regression (VAR) models offer an alternative approach. They share many of the advantages of DFM – notably the capacity to be cast in a state-space form, allowing to

²¹ To handle mixed-frequency data, the principle is to write the state-space system at the highest available data frequency and treat the lower-frequency data as high-frequency data that are periodically missing. In our mixed-frequency set-up with quarterly and monthly data, frequency aggregation is based on Mariano and Murasawa (2003) that reconstruct quarterly growth rates from monthly growth rates of the latent variable.

²² In the DFM included in the toolbox, estimates are based on the expectation-maximization (EM) algorithm and Kalman smoother. The algorithm is initialized by computing principal components, with parameters of the model estimated by OLS. Given estimated parameters, common factors are updated using the Kalman smoother. The maximum likelihood estimates are obtained by iterating until convergence. For details on the EM algorithm used in the DFM of the toolbox, please refer to Bańbura and Modugno (2014).

²³ Users are kindly requested to cite the original paper of Cimadomo et al. (2022) when this model. Please refer directly to the paper for an extensive description: the model used in the toolbox uses blocking (B-BVAR in Cimadomo et al., 2022).

handle missing data *via* the Kalman filter. In addition, they provide a joint model of all variables. A VAR model is formed by equation 4 where $Y_t = (y_{1,t}, y_{2,t}, \dots, y_{n,t})$ is a set of economic variables, A_i a set of matrices with the parameters, and ε_t white noises. While nowcasting with VAR had long been limited to only few input variables (e.g., Mitnik and Zdrozny, 2005; Giannone et al., 2009), the literature has shown that Bayesian shrinkage can make Bayesian VAR (BVAR) models suitable for high-dimensional problems (Bańbura et al. 2010, Giannone et al. 2015). It relies on the idea of using a parsimonious naive prior to discipline the estimation of such densely parameterized models.

$$(4) \quad Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + \varepsilon_t$$

The model in the toolbox is the blocking BVAR (B-BVAR) of Cimadomo et al. (2022), a VAR specifically adapted to handle mixed frequencies and ragged edges.²⁴ The idea of B-BVAR is to align all frequencies to the lowest one by treating the higher frequency variables (monthly in our case) as multiple observations at lower frequency (quarterly in our case). The BVAR is written at quarterly frequency and the monthly variables are incorporated as three separate series, one for each month of the quarter, and stacked with other quarterly variables. The system is then estimated with Bayesian methods based on Giannone et al. (2015), with Bayesian shrinkage helping to handle the larger system implied by the blocking approach.²⁵

While DFM have been the preferred nowcasting technique, BVAR are increasingly used in the literature. One of the main drawbacks of using BVAR for nowcasting with big data had been the proliferation of parameters in a high-dimensional setting. Recent literature showed that Bayesian shrinkage can control the high estimation uncertainty in this case and offer an alternative to factor models (De Mol et al., 2008; Bańbura et al., 2010). In addition, the use of Bayesian estimation accounts for the uncertainty coming from modelling choices notably in the model parameters. This is why BVAR have been increasingly used for nowcasting for example in Schorfheide and Song (2015), McCracken et al. (2015), Brave et al. (2019), Knotek and Zaman (2019), Schorfheide and Song (2021), and Consolo et al. (2023).

²⁴ While Cimadomo et al. (2022) presents three different methodologies for nowcasting with large BVAR, their results show that performances are very close across methods (B-BVAR, L-BVAR, and C-BVAR). We opt for the B-BVAR due to its lower computational time and more straightforward interpretability.

²⁵ The model is estimated using a Normal-Inverse Wishart prior. The covariance matrix of the ε_t is estimated using an inverse Wishart. The constant terms (in A_0) are estimated using a flat prior. The other model parameters (in A_i with $i > 0$) are estimated by combining a Minnesota prior (Litterman, 1979) and the sum-of-coefficients prior (Doan et al., 1984). The methodology uses diffuse priors as in Giannone et al. (2015).

Combination of bridge equations (Bańbura et al., 2023)²⁶

A bridge equation links the quarterly target variable to monthly indicators. The latter are aggregated to quarterly frequency to estimate an OLS equation at this frequency and compute the forecast for the target variable as in equation 5 where y_t is the target variable, $x_{i,t}^Q$ is the monthly variable $x_{i,t}$ aggregated at quarterly frequency, and ε_t are error terms. The relationship includes p lags of the dependent variable, as well as q lags of the predictors. Since there is a contemporaneous link between the target variable and the monthly predictors in equation 5, monthly predictors might need to be extrapolated for (some of) the months of the quarter before using the bridge equation. In the toolbox, this is done using an auxiliary BVAR with 6 lags as in Bańbura et al. (2023).

(5)
$$y_t = \beta_0 + \sum_{k=1}^p \alpha_k y_{t-k} + \sum_i \sum_{l=0}^q \beta_{il} x_{i,t-l}^Q + \varepsilon_t$$

The toolbox includes a combination of bridge equations (BEQ) in the spirit of Bańbura et al. (2023). OLS equations tend to have poor predictive performances in a high-dimensional set-up: hence the model in the toolbox follows Bańbura et al. (2023) and uses a combination of simple bridge equations based on one or two monthly predictors and one (or none) quarterly predictors. With n monthly predictors, the nowcasting model will then choose 2 choose n (i.e., all combinations of 2 monthly variables among the n possible) plus n (all models with a single monthly variable) bridge equations. When quarterly predictors are available, the list of models is expanded to include all combinations with one or two monthly regressors and none or one quarterly predictor. For example, with 3 monthly (x_i^m) and 1 quarterly variables (x_i^q), the toolbox estimates the 12 bridge equations in **Table 1**.

Table 1
Combination of bridge equations with three monthly and one quarterly predictors

	Monthly predictor(s)	Quarterly predictor
1	x_1^m	

²⁶ Users are kindly requested to cite the original paper of Bańbura et al. (2023) when using this model. Please refer directly to it for an extensive description. It should be noted that the toolbox uses a simplified version of Bańbura et al. (2023) where it relies only on one type of extrapolation of input variables (see **Appendix C** for more details) and take a more brute-force approach with all possible bridge equations in the combination.

2		x_2^m	
3		x_3^m	
4	x_1^m	x_2^m	
5	x_1^m	x_3^m	
6	x_2^m	x_3^m	
7		x_1^m	x_1^q
8		x_2^m	x_1^q
9		x_3^m	x_1^q
10	x_1^m	x_2^m	x_1^q
11	x_1^m	x_3^m	x_1^q
12	x_2^m	x_3^m	x_1^q

Notes: Blank space denotes no variables. Monthly variables are indicated with a superscript m , and quarterly variables with a superscript q .

Due to their simple setup, bridge equations have long been used in nowcasting. In some cases, bridge equations have been proven to perform better than more sophisticated models (Bańbura and Saiz, 2020; Soybilgen and Yazgan, 2018), with one possible reason being that they tend to be more robust to outliers and structural changes (Bańbura et al., 2023). They have long been used for nowcasting, including at the European Central Bank (Parigi and Schlitz, 1995; Rünstler and Sedillot, 2003; Baffigi et al., 2004; Hahn and Skudelny, 2008; Bańbura and Saiz, 2020; Bańbura et al., 2023).

2.3 Covid corrections

The toolbox finally allows to apply specific corrections on the Covid period. The underlying idea is that econometric frameworks might require some specific adjustments when handling a sequence of extreme observations such as the Covid-19 period (Carreiro et al., 2022; Lenza and Primiceri, 2022; Almuzara et al., 2023). It remains an open question whether the pandemic data should be treated as conventional or will distort the parameter estimates of the models. Nevertheless, the impact of Covid-19 observations on time series model estimations might depend on the type of model and on the input data (e.g., Schorfheide and Song, 2021; Zhang et al., 2021; Bobeica and Hartwig, 2023), leaving this question as a mostly empirical one. Against this background, the toolbox allows for the three Covid-19 corrections in **Figure 3**. The first (panel a) includes dummies for 2020 Q2 and 2020 Q3 as is standard in

the literature when correcting for idiosyncratic events (e.g., Wooldridge, 2015);²⁷ the second (panel b) deletes the observations related to the Covid-19 period as in Schorfheide and Song (2021);²⁸ and the third (panel c) corrects for outliers as in Chen and Liu (1993).²⁹

Figure 3

Possible Covid-specific corrections in the toolbox

a) Include dummies	b) Delete observations	c) Correct for outliers
x_1 x_2 x_3 $d20q2$ $d20q3$	x_1 x_2 x_3	x_1 x_2 x_3
1995 M1		
1995 M2		
1995 M3		
...		
2019 M12		
2020 M1		
2020 M2		
2020 M3		
2020 M4		
2020 M5		
2020 M6		
2020 M7		
2020 M8		
2020 M9		
2020 M10		
2020 M11		
2020 M12		
...		
2023 M5		
2023 M6		
2023 M7		

Source: Authors.

Notes: Covid corrections are provided for an example dataset containing three variables (x_1, x_2, x_3). A red cross indicates that the observation is deleted and left blank (NaN in Matlab) in the input dataset.

²⁷ The toolbox also allows for an alternative setting with dummies rather for 2020 Q1 and 2020 Q2. Alternative settings with either monthly dummies for all months of 2020 Q2 and Q3, or with full quarterly dummies (i.e., taking the value 1 on all three months of the quarter, instead of only the last month of the quarter as in Figure 3) have been tested but were found empirically to perform worst.

²⁸ Observations are deleted from February to September 2020 (both included). This setting is empirically verified by the prevalence of Covid cases over these months and by these months being the period where the greatest number of outliers is detected.

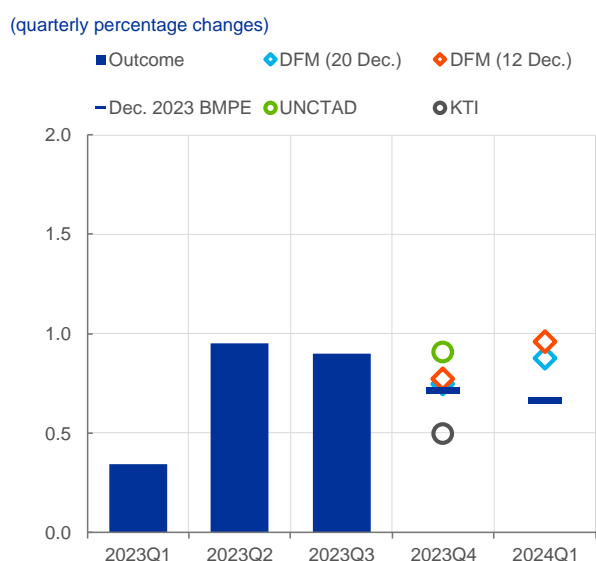
²⁹ Outliers are detected as those exceeding the median plus / minus four times the inter-quintile distance as is standard in the literature (Rousseeuw and Croux, 1993).

3 Outputs for policy analysis

The toolbox offers a range of complementary outputs. The suite of outputs provided by the toolbox serves two purposes: 1) opening the “black box” of predictions to provide insights on the economic drivers behind nowcasts, and 2) assessing the degree of reliability that can be attached to model predictions. In addition, a side effect of delivering multiple outputs is to provide the user with informative content even if point forecast seems off.

3.1 Point forecasts

Figure 4
Short-term predictions



Sources: Bloomberg, Haver, S&P Global, and authors' calculation.
Notes: "KTI" relates to the Kiel Trade Indicator of Stamer (2024); the quarterly growth rate is extrapolated from the month-on-month growth rates; "UNCTAD" relates to the nowcast predictions based on the neural network of Hopp (2021).

The toolbox gives point forecasts for the two quarters after last available data.³⁰

Figure 4 shows a typical way to display the predictions with point forecasts shown as diamonds and compared with ECB staff projections (blue lines) as well as external predictions such as KTI (grey circle) and UNCTAD (green circle).³¹ Points forecasts provide a guidepost for the growth of the target variable in the next quarters: in the example of **Figure 4**, the nowcasting model would predict the target variable (global real trade growth) at 0.7% in 2023 Q4 and 0.9% in 2024 Q1. On top of the *level* of the target variable, they also provide information on its *direction* (increase, decrease, stability)

from one quarter to the other. In the example of **Figure 4**, model-based predictions point to a

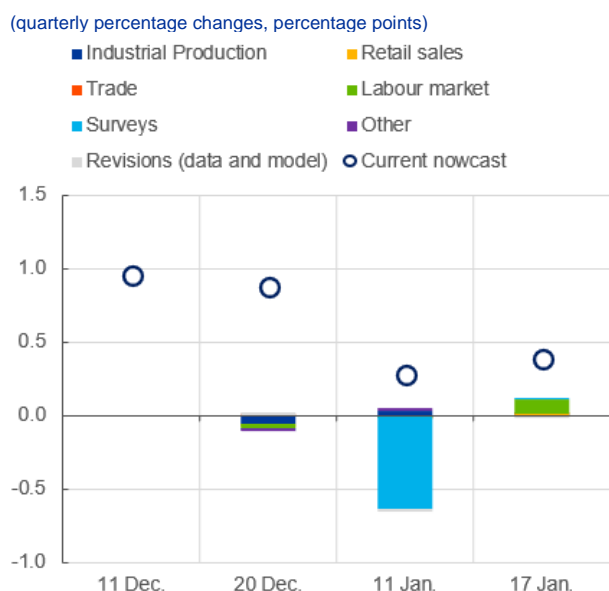
³⁰ The two predictions are either 1) back-cast (previous quarter) and nowcast (current quarter) if the target variable for the previous quarter has not yet been released, and 2) nowcast and forecast (one quarter ahead) otherwise. Forecasts for up to 6 months after the date at which the code is run can however be recovered in the *Matlab* output. Longer horizons can be obtained with slight modifications to the toolbox.

³¹ Predictions from the Kiel Trade Indicator (KTI) are available on the [website of the Kiel Institute for the World Economy](#); they are based on Stamer (2024). Predictions from the UNCTAD are available on the [website of the United Nations Conference for Trade And Development](#); they are based on Hopp (2021).

deceleration of the target variable (global real trade growth) in 2023 Q4 before a recovery in the next quarter. Finally, the figure also indicates the evolution of nowcasts between two runs of the nowcasting model. In the example of [Figure 4](#), the evolution between two predictions suggests that growth prospects for 2024 Q1 have slightly deteriorated between 12 December (red diamond) and 20 December (blue diamond).

3.2 News decomposition

Figure 5
Impact of news releases



Sources: Bloomberg, Haver, S&P Global, and authors' calculation.
Notes: Dots represent the evolution of short-term estimates. Bars are contributions from news to revisions between two consecutive rounds.

News decomposition tracks the impact of data releases on revisions of point forecasts.

Model predictions are based on forecasted values for input variables that are not yet released. When the *ex-post* release of the input variable differs from the *ex-ante* extrapolation, the “news” affects the nowcast. News might be interpreted as the “surprise” relative to the model’s forecast. Only this unexpected component affects the nowcast. The toolbox follows the framework of Bańbura and Modugno (2014) to extract the impact of news releases from models casted in state-space form, based on the expectations that can be derived from the Kalman filter. News, understood in this

context, do not refer to the news release itself, but the component of the news release that is “unexpected” by the model. Formally, this may be expressed as equation 6 where the $E[y_{k,t_k}|\Omega_{v+1}]$ is the new forecasts based on the “new” data vintage Ω_{v+1} . The new forecasts can be expressed as a linear combination of the old forecast $E[y_{k,t_k}|\Omega_v]$, based on the “old” data vintage Ω_v , and a revision component $E[y_{k,t_k}|I_{v+1}]$, based on the news I_{v+1} . The news I_{v+1} represent that part of the news release, which is unexpected, i.e., containing information not already included in Ω_v . This news component may be obtained from the Kalman filter as a weighted average of the news, *weighted* by the importance of each news release for the

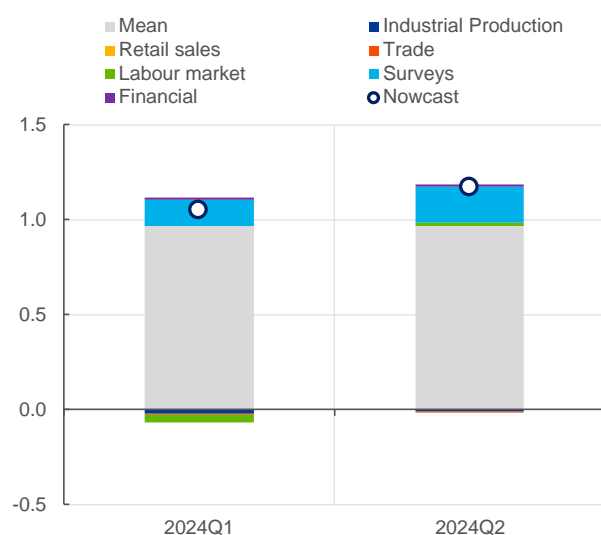
variables contained in the model. From this, it is possible to decompose the forecast revision into contributions from the news component of releases or revisions in individual input variables or groups of input variables (Bánbura and Modugno, 2014). **Figure 5** presents a standard way to portray news decomposition, with the news aggregated by group of variables.³² In the example of **Figure 5**, the nowcast was strongly revised downward between 20 December and 11 January due to the releases of survey data that came more negative than forecasted *ex-ante* by the model; then positive news from the labour market between 11 and 17 January pushed the nowcast up again. In line with the above definition of the news decomposition, it should be kept in mind that *positive* news should be interpreted with respect to what the model forecasted before the actual release.

$$(6) \quad E[y_{k,t_k} | \Omega_{v+1}] = E[y_{k,t_k} | \Omega_v] + E[y_{k,t_k} | I_{v+1}],$$

3.3 Approximate contributions

Figure 6
Contributions to short-term predictions

(quarterly percentage changes, percentage points)



Sources: Bloomberg, Haver, S&P Global, and authors' calculation.
Note: Contributions are approximated as news from all data releases from the two years prior to the target date.

The toolbox also computes the (proxy) contributions of input variables. This

aims at providing insights as to which inputs are driving the model predictions at one point in time. For most models, notably the DFM, the contributions cannot be obtained directly from the model; they are proxied by the impact of all data releases from two years prior to the target date.³³

Figure 6 shows the contributions to the point forecast (blue hollow dot) by groups of variables. The mean (grey bar) is the value to which the model reverts in the absence of input data; while it is in principle close to the actual long-term

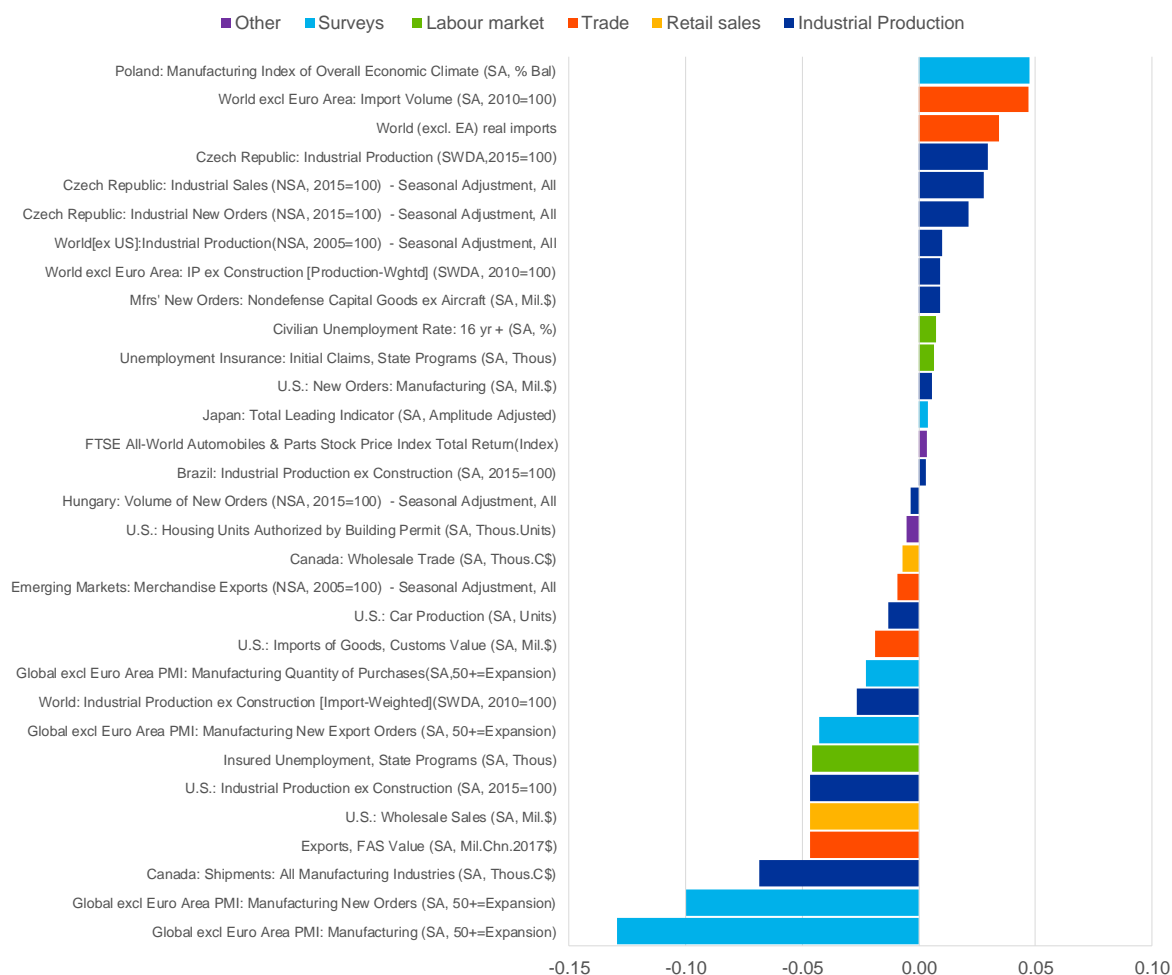
³² The toolbox also provides the impact of news for each individual input variable in the *Excel* output.

³³ Some methods have aimed at extracting exact contributions from dynamic factor models, for example Bańbura and Rünstler (2011) using an algorithm by Koopman and Harvey (2003) to get the weights of the individual

mean of the target variable, it can include some persistent effects of the input variables. In the example of **Figure 6**, surveys (light blue bar) are driving up the predictions for 2024 Q1 and to a stronger extend in 2023 Q4. In addition to *grouped* contributions, the toolbox also produces the contributions of *individual* input series as in **Figure 7**. On top of the name of the variable, the different colours indicate the group to which the variable belong. For instance, this chart shows that the negative contribution of surveys (light blue bar) is mostly driven by the global (excl. EA) manufacturing PMI.

Figure 7
Contributions to short-term predictions – by individual variable

(percentage points)



Source: Bloomberg, Haver, S&P Global, and authors' calculation.

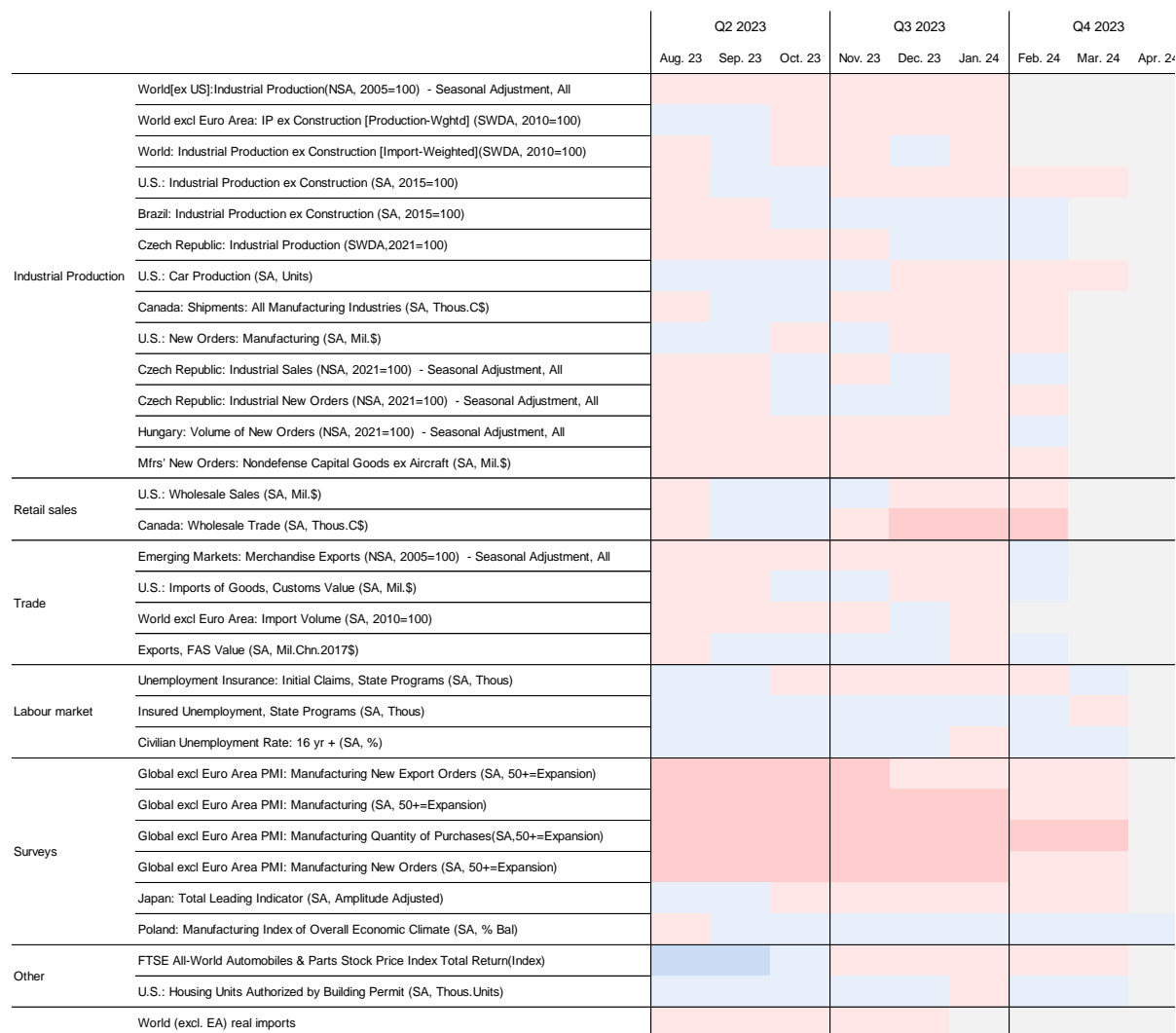
Notes: In such predictions from dynamic factor models, contributions are approximated as news from all data releases from two years prior to the target date.

observations in the state vector. For reasons of computational time, these methods are not implemented in this toolbox. The fact that contributions are proxy should be kept in mind for policy analysis.

3.4 Heatmap

Figure 8
Heatmap of individual input variables

(z-scores: number of standard deviations from the long-term mean)



Sources: Bloomberg, Haver, S&P Global, and authors' calculation.

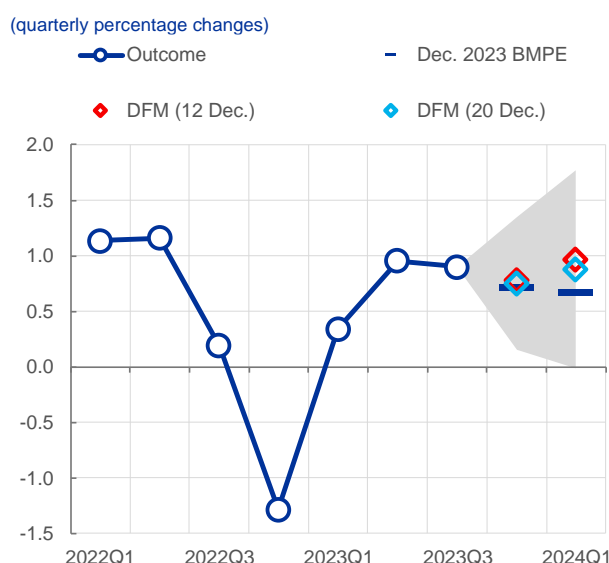
Notes: Z-scores are computed by subtracting the mean and dividing by the standard deviation. For monthly series, values are smoothed over five months as per the approximation of quarterly growth rates with monthly growth rates (Mariano and Murasawa, 2003).

The toolbox provides a heatmap of input variables to directly observe the input variables entering the model. Figure 8 shows the representation produced by the toolbox. The heatmap gives z-scores for each input variable (distance of the current observation to the long-term mean, expressed in standard deviations). Negative distances from the long-term mean are represented in red, and positive distances in blue. Father distances are indicated by darker shades of these colours. Grey means that data are not yet released. The heatmap can

indicates how the economy fares at the current juncture. The heatmap can also be a last resort: at times when the model gives unreliable predictions, the heatmap is a way to look at the raw source (i.e., input variables) to assess the state of the economy.³⁴

3.5 Confidence intervals

Figure 9
Short-term predictions with confidence bands



Sources: Bloomberg, Haver, S&P Global, and authors' calculation.

Notes: The grey range represents the 57.5% confidence interval. It is computed based on the mean absolute errors of prediction errors over the past 10 years, following Reifschneider and Tulip (2019), and adjusted for outliers to represent uncertainty under normal circumstances as in ECB (2009).

To gauge the level of confidence that can be associated with point forecasts, the toolbox produces uncertainty bands.

They are based on Reifschneider and Tulip (2019), using the rolling-window Mean Absolute Error (MAE) of the pseudo real-time projections over the past 10 years.³⁵ The MAE are computed for all horizons (back-cast, nowcast, forecast). In line with ECB (2009), the past projection errors are adjusted for outliers to represent the uncertainty in “normal times”, assuming that future shocks will be of similar magnitude to past shocks. The uncertainty bands are constructed as plus / minus one MAE as in ECB (2009) which represents the 57.5% confidence interval.³⁶ **Figure 9** shows an

example of uncertainty bands (grey areas): it suggests that a higher uncertainty is associated with the prediction for 2024 Q1. Finally, the MAEs are re-estimated at each run of the toolbox to reflect the fact that uncertainty associated with projections changes depending on the

³⁴ The toolbox also offers a heatmap for variable groups, based on the un-weighted average of z-scores.

³⁵ While some of the models (e.g., BVAR) provide densities, their uncertainty bands account generally only for parameters uncertainty. However, sources of uncertainty go beyond parameters (e.g., specification, data, shocks) and are often hardly quantifiable. The literature generally opts for using *ex-post* forecast errors as a catch-all proxy for *ex-ante* forecast uncertainty (Wallis, 1989). This is the practice in many institutions like the Bank of England (Britton et al., 1998), the Fed (Reifschneider and Tulip, 2019), the IMF (Elekdag and Kannan, 2009), the ECB (ECB, 2009), and the Deutsche Bundesbank (Knüppel, 2014).

³⁶ For a normal distribution, the link between the MAE and the standard deviation (σ) is given by $MAE = \sqrt{2/\pi} \times \sigma$. This corresponds to the 57.5% central confidence interval. The computation of uncertainty bands therefore assumes that projection errors are normally distributed.

moment of the forecast: e.g., uncertainty for a nowcast on 1st day of the 1st month of the quarter, when no data are available, will be greater than on the 30th day of the 3rd month. The dynamic computation of the MAE reflects this time-varying uncertainty.³⁷

The toolbox offers two additional metrics to gauge the confidence that can be associated with model's predictions. The first is the share of input variables that have been released for the quarter under consideration: a higher number implies that the model prediction are mostly based on actual data – and *vice versa*, a lower number implies that predictions rely mostly on extrapolated data. The second is the Forecast Directional Accuracy (FDA) computed as the share of occurrences when the model would have correctly predicted the direction (increase / decrease) of the target variable. As for the MAE, this metric is computed on pseudo real-time projections over the past 10 years. It lies between 0 and 100%: a higher number indicates that the model would have correctly predict the direction of the target variable – suggesting that a higher credibility of the direction of the projections.

3.6 Range of alternative models

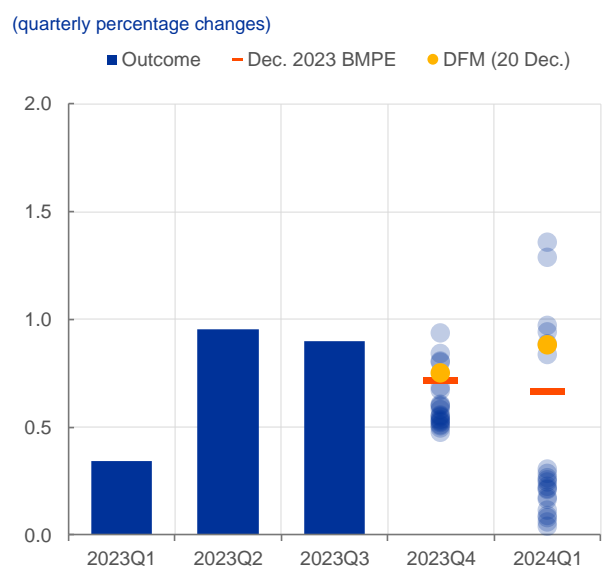
The toolbox computes projections for a range of alternative models, which can serve to assess the impact of excluding some variables. Alternative models are computed by removing one or two groups of variables from the information set at a time.³⁸ In this sense, they provide a picture of what predictions would have been if some variables would be ignored. This can prove useful when some variables are deemed by the user to provide a poor signal of economic conditions at some point. For instance, this can be the case when a key input variable is affected by an idiosyncratic factor (e.g., retail sales affected by front-loading of spending ahead of VAT hike, not signalling some improved economic fundamentals but a specific event). This can be also useful when some variables, historically highly correlated with the target variable, provide a poor signal of economic conditions: one example are PMIs disconnecting somehow from GDP growth during the pandemics (de Bondt and Saiz, 2024) or trade in goods decorrelating from trade in goods and services (Attinasi et al., 2024). **Figure 10** presents the way predictions from alternative (transparent blue dots) are presented in the

³⁷ Dynamic re-estimation is an option in the toolbox as it adds significant computational time (**Box B** in **Appendix B**).

³⁸ Therefore, the number of alternative models depends on the number of groups. The range of alternative models is an option in the toolbox, which adds significant computational time (**Box B** in **Appendix B**).

toolbox output alongside the main model predictions (yellow dot). The toolbox provides the predictions of each alternative model, specifying which variable group(s) has been removed.³⁹ This way, the user can clearly identify what are the predictions when the faulty variables are removed from the information set.⁴⁰ Even in the absence of faulty variables, **Figure 10** can illustrate how some variables disproportionately affect model predictions. For instance, in **Figure 10**, models on the upper bound of the range in 2024 Q1 are those which exclude surveys, suggesting they are weighing heavily on predictions. If surveys are known to be pessimistic at this juncture, the user can put more credibility into the models on the lower bound.⁴¹

Figure 10
Alternative short-term predictions



Sources: Bloomberg, Haver, S&P Global, and authors' calculation.
Notes: Light blue dots represent predictions by alternative models obtained by removing one or two groups of variables from inputs.

The range of alternative models can also be used to further assess uncertainty, loosely in the spirit of thick modelling.

For instance, **Figure 10** somehow shows that some large uncertainty surrounds the predictions for 2024 Q1 with forecasts going from 0% to 1.5%. By contrast, the uncertainty for 2023 Q4 appears lower as most predictions are centred around the main DFM prediction at 0.7% (yellow dot). On top of uncertainty, the range can also provide insights on the direction of risks: for instance, risks appear to the downside in 2024 Q1 as most of the predictions from alternative models are concentrated in the lower end, between 0% and 0.5%, way

below the main prediction of 0.9% (yellow dot). This approach can be loosely related to the thick modelling approach coined by Granger and Jeon (2004). In this approach, slightly different set of variables are used in similar model specifications to provide forecasts. In this spirit, the range of alternative models provide predictions when model settings are kept the

³⁹ Groups of variables are specified by the user in the input dataset.

⁴⁰ Building on this idea, an alternative way to present **Figure 10** is by cherry-picking the alternative model(s) that exclude faulty or unreliable variables.

⁴¹ In this spirit, the toolbox also allows to eliminate a few variables from the information set (see **Appendix B**).

same, but the information set slightly differ – which can be used to quantify uncertainty on the state of the economy when different variables can be giving mixed signals.

4 Example: building a nowcasting model for global GDP growth

The following section presents an example of how the nowcasting toolbox can be used to create a nowcasting model from scratch. This example focuses on nowcasting quarter-on-quarter global GDP growth. We also compare the performances of the new model, which is built following the proposed three-step approach, with an existing model, which was built on a more *ad hoc* procedure. Using the toolbox and the structured approach presented in [section 1.1](#) allowed us to significantly increase model performances over 2013-2023.

4.1 Variable pre-selection

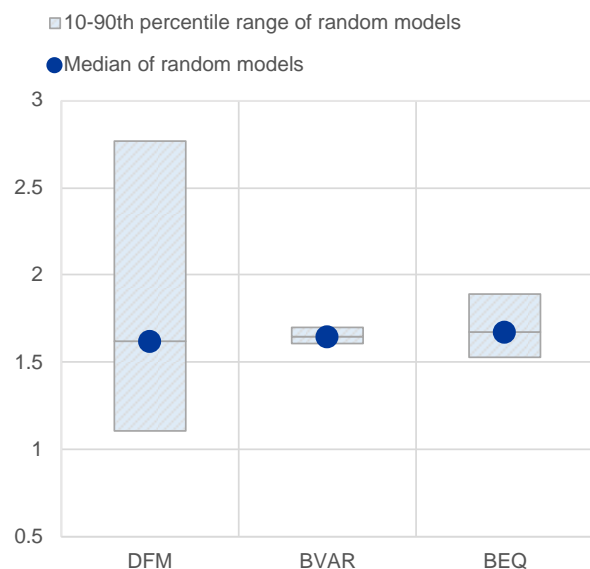
As a first step, a set of 540 potential regressors was narrowed down using the pre-selection methods available in the toolbox. Our initial dataset includes 540 potential regressors for global GDP growth; they aim at covering various aspects of the economy such as industrial production, consumer spending and sentiment, labour markets, PMI surveys, financial markets, commodity prices, trade variables, housing, and consumer prices. Methods for pre-selection (SIS, t-stat based, and LARS) are used to assess the predictive power of these potential regressors with respects to the target variable.⁴² Our pre-selection of variables rely not only on this score of predictive power but also considers timeliness and the groups of variables. For instance, we remove variables that have a long publication delay (more than three months). We also adjust the pre-selection to include a set of input variables relatively balanced across variable groups – even though adjustments on this dimension were relatively limited since the pre-selection naturally tended towards a balanced pre-selected in our case. Our final preselection consisted of 79 monthly and 2 quarterly series, where industrial activity (31%), retail sales (20%) and surveys (12%) represented the largest shares.

⁴² To avoid over-reliance on one technique, we compute a weighted score based on the ranking of variables provided by the three techniques. The aggregate score gives a higher weight to LARS, which has been shown to perform better than other techniques in the literature. More information can be found in the template for pre-selection provided with the code of the toolbox.

4.2 Model selection

Figure 11
Accuracy across model classes

(out-of-sample RMSE over 2013-2023)



Sources: Bloomberg, Haver, S&P Global, and authors' calculation.
Notes: Out-of-sample predictions are done recursively using a pseudo real-time set-up over 2013-2023. Values refer to the nowcasting horizon (current quarter). Results are based on 292 DFM, 780 BVAR, and 170 BEQ.

Model selection proceeded by automatically testing 1,242 model specifications across the three different model classes. The toolbox automatically estimates models of the different classes (DFM, BVAR, and BEQ) with random model parameters and a random subset of variables. Bounds for model parameters and the number of regressors were chosen based on values provided in [Table A1](#) in [Appendix A](#). The starting date of the estimation sample was also drawn randomly by the toolbox between 2002 and 2012. Models were evaluated out-of-sample on the 2013-2023 period – to capture a full business cycle.

The toolbox estimated performed the out-of-sample evaluation in a pseudo real-time forecasting exercise. To this end, the dataset was frozen around 25 September 2023. To conduct the pseudo real-time exercise, the toolbox automatically reconstructs the “ragged edge” pattern of the data for each month of the out-of-sample period.⁴³ At each month of the out-of-sample period, the toolbox also produces the pseudo real-time back-, now- and forecasts. From these predictions, the RMSE and the FDA are calculated.

The comparison across model classes suggests similar average performances, but that the DFMs produced best-performing models. Nowcasting performances of DFM, BVAR and BEQ models are compared in [Figure 11](#): while median performances are very close with a RMSE of around 1.5%, the dispersion of performances is far greater for the DFMs. Hence, several DFMs exhibit better performances than their BVAR and BEQ counterparts. As

⁴³ The toolbox produces predictions for each month of the out-of-sample period, assuming the ragged-edge pattern remains the same. For example, if a series has the last 2 observations missing in the dataset, the toolbox assumes that at any month in the past, the series also had the last 2 observations missing. The underlying assumptions are that 1) predictions for past months are done at the same day of the month, and 2) publication delays remain unchanged.

mentioned above, this finding cannot be generalized as the literature has shown that models that performs best depend on the context and input variables (Guichard and Rusticelli, 2011; Chernis and Sekkel, 2017; Soybilgen and Yazgan, 2018; Bańbura and Saiz, 2020). Nevertheless, we focused the rest of our exercise on the DFMs.

The randomization already produced many models improving upon the previous ad hoc model.⁴⁴ **Figure 12** (panel a) shows RMSE for different horizons, comparing the performances of the range of random models (10th to 90th percentiles in the light blue band) with the previous ad hoc model (blue diamond) and the new model (red diamond). For the back-casting and nowcasting horizons, the previous model ranks among the lowest-end of the random models. This suggests that making an appropriate pre-selection – rather than an ad hoc selection of variables – already yields higher accuracy, regardless of the exact model settings. Forecasting performances of the previous model are more in line with the range of random models, which highlights the difficulties to produce relevant predictions for this horizon. Based on these performance metrics, an overall score was attributed to each model by weighting the RMSE and FDA at the different horizons. More weight was placed on the nowcasting horizon to reflect that the main purpose of the model lies in the current quarter. In addition, a higher weight was placed on RMSE rather than FDA, as our intention was to focus more on point accuracy rather than directional accuracy.

4.3 Covid robustness

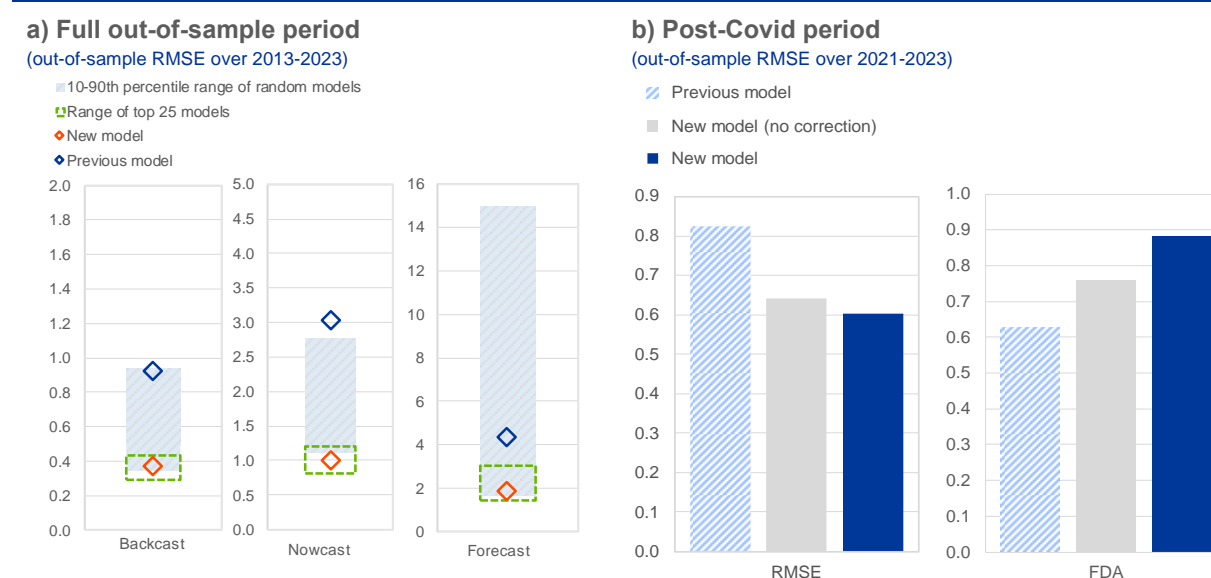
Empirically, Covid correction is found to further improve the performances of our best-performing model. From the models evaluated on the 2013-2023 out-of-sample period, we extracted the 25 models with best aggregate performances, whose RMSE are represented in the green rectangle on **Figure 12** (panel a). These models were then re-evaluated over the 2021-2023 horizon, applying each of the three Covid-correction methods at each model (dummy variables for 2020Q1 and Q2; deleting Covid-19 observation; and correcting for outliers – see **section 2.3**). Model performances were then compared to their benchmark without Covid correction over the 2021-2023 horizon. This allowed to assess the effectiveness of correcting for the Covid period. **Figure 12** (panel b) shows how the final nowcasting model

⁴⁴ The previous ad-hoc model was constructed by taking PMI, retail sales, industrial production, and employment across 7 major global economies (US, UK, Japan, Brazil, Russia, China, India).

with Covid correction (outlier correction) compares against the same model without a correction. Using outlier correction, out-of-sample directional accuracy increased by around 10 p.p. over the post-Covid period. The out-of-sample RMSE also slightly improves.

Figure 12

Accuracy of nowcasting models



Sources: Bloomberg, Haver, S&P Global, and authors' calculation.

Notes: Out-of-sample predictions are done recursively in a pseudo real-time set-up. In panel a) the new model is without Covid-19 correction until Dec. 2020 and with correction afterwards. In panel b) charts relate to the nowcasting horizon (i.e., prediction for the current quarter).

4.4 Results

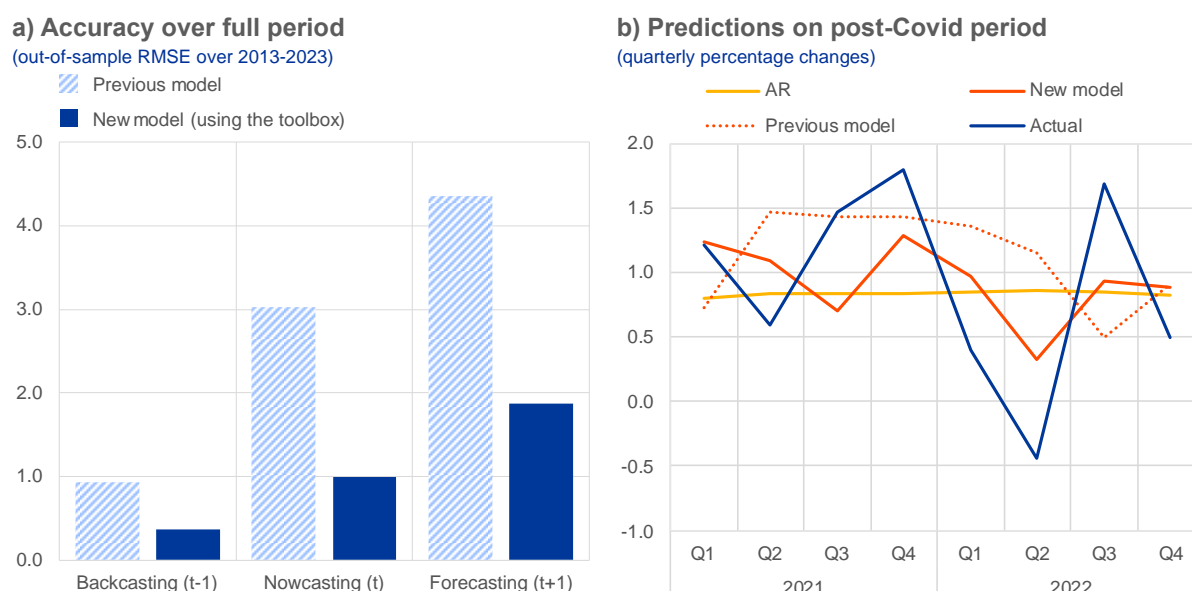
The final global GDP nowcasting model is a four-factor DFM with 25 monthly variables, representing all variable groups from the preselection. The final model selection is based on model performances, as well as other considerations such as the representativity of the variables included. Unsurprisingly, the distribution of variables into the different groups closely follows the distribution of the preselection with a high share of variables pertaining to industrial activity, retail sales, and PMI surveys.

Overall, we were able to achieve a significant increase in prediction accuracy. Figure 13 (panel a) compares the point accuracy of the revised model with the previous model. Across all horizons, we achieve a significant reduction in out-of-sample RMSE over 2013-2023. For our main target horizon (nowcasting) we reduce RMSE by 66%. For both back- and forecasting, we can more than half the prediction error. While the gains are less pronounced

for directional accuracy, we managed to improve by 12 p.p. for the nowcasting horizon. For the back-casting horizon, we matched the already high 85% directional accuracy of the previous model – meaning the model provides the actual direction more than 8 times out of 10. Though it was less a focus of our exercise, we also roughly matched the directional error of the previous model on the forecasting horizon. **Figure 13** (panel b) compares nowcasts over the post-Covid period with the actual outcome. The new model tracks the movements of global GDP growth much more accurately than previous model and the AR benchmark. This showcases the reduced RMSE and the increased directional accuracy. Improvements are similar across other target variables: **Appendix C** presents results for global trade and real GDP growth in the US.

Figure 13

Accuracy of the best-performing model



Sources: Bloomberg, Haver, S&P Global, and authors' calculation.

Notes: Out-of-sample predictions are done recursively in a pseudo real-time set-up. In panel a) the new model is without Covid-19 correction until Dec. 2020 and with correction afterwards. In panel b) predictions are done in the data at the 3rd month of the quarter.

Conclusion

While nowcasting is a key tool for policy decisions, existing publicly available tools suffer from a lack of adaptability. We provide a user-friendly and all-in-one toolbox that combine two workstreams in a centralized toolbox: 1) model building and evaluation to be done periodically, and 2) regular for conjunctural policy analysis. The toolbox is geared towards policy analysis and provides a plethora of detailed outputs to open the “black box” of nowcasting. Another key interest is to provide a step-by-step approach to building a nowcasting model, that can provide a valuable framework for forecasters (above all if non-specialized in nowcasting) willing to build a model from scratch. In addition, the toolbox sets an all-in-one code that combines various model classes, pre-selection methods, and Covid-19 corrections in a single infrastructure.

Applying the toolbox and the structured approach to the revision of a model for quarterly global GDP growth, we showcase the increases in model performances. We further verify that Covid-19 corrections, namely outlier correction, deleting Covid observations and dummy variables can increase model out-of-sample performances over the post-Covid period, thus showing how to prepare nowcasting models for the post-Covid period.

Despite the various features included in the toolbox, it should not be considered a martingale to create the best-in-class model. While the toolbox can automatize model selection and make it easier, it does not perform a full grid-search but rather relies on a “trial-and-error” approach – meaning *even better* models can exist. The outputs of the toolbox should be enjoyed with some caveats, notably the approximate contributions. The automated nature of the toolbox does not – and should not – replace user judgement: it only aims at assisting economists in arriving at an accurate, and most importantly, useful nowcasting model.

Finally, the toolbox should be seen as a living thing that the authors will try to improve and develop further. The toolbox encompasses the most popular models, but other methods might be added only in future versions, for example MIDAS (Ghysels et al., 2004) or recent models like neural networks (Hopp, 2021) or random forests (Lenza et al., 2023). Other methods for pre-selection or Covid correction could be added such as Jarociński and Maćkowiak (2017) or Cascaldi-Garcia (2022). The toolbox might also be developed to accommodate for higher frequency indicators, in the vein of Delle Chiaie and Perez-Quirós (2021).

References

- Almuzara, M., Baker, K., O’Keeffe, H., and Sbordone, A. (2023). “The New York Fed Staff Nowcast 2.0”, *New York Fed Staff Nowcast Technical Paper*
- Anesti, N., Galvão, A., and Miranda-Agrippino, S. (2018). “Uncertain Kingdom: nowcasting GDP and its revisions”, *Working Papers*, No 764, Bank of England
- Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L., and Rünstler, G. (2011). “Short-term forecasts of euro area GDP growth”, *The Econometrics Journal*, 2011, 14(1), 25–44
- Attinasi, M-G., Boeckelmann, L., Hespert, L., Linzenich, J., and Meunier, B. (2024). “Global trade in the post-pandemic environment”, *Economic Bulletin Boxes*, 1, European Central Bank
- Baffigi, A., Golinelli, R., and Parigi, G. (2004). “Bridge models to forecast the euro area GDP”, *International Journal of Forecasting*, 20, 447–460
- Bai, J., and Ng, S. (2008). “Forecasting economic time series using targeted predictors”, *Journal of Econometrics*, 146(2), 304–317
- Bair, E., Hastie, T., Paul, D., and Tibshirani, R. (2006). “Prediction by supervised principal components”, *Journal of the American Statistical Association*, 101(473), 119–137
- Bańbura, M., Belousova, I., Bodnár, K., and Tóth, M. B. (2023). “Nowcasting employment in the euro area”, *Working Paper Series*, No 2815, European Central Bank
- 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 *Oxford Handbook on Economic Forecasting*, Clemens, M., and Hendry, D. (eds.), 193–224.
- Bańbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). “Chapter 4 – Now-Casting and the Real-Time Data Flow”, in *Handbook of Economic Forecasting*, Elliott, G., and Timmermann, A. (eds), Elsevier, 2(A), 195–237

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(11), 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

Bańbura, M., and Saiz, L. (2020). “Short-term forecasting of euro area economic activity at the ECB”, *Economic Bulletin*, 2, European Central Bank

Barbaglia, L., Frattarolo, L., Onorante, L., Pericoli, F. M., Ratto, M., and Tiozzo Pezzoli, L. (2023). “Testing big data in a big crisis: Nowcasting under Covid-19”, *International Journal of Forecasting*, 39(4), 1548–1563

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*, 132, 169–194

Bok, B., Caratelli, D., Giannone, D., Sbordone, A., and Tambalotti, A. (2018). “Macroeconomic Nowcasting and Forecasting with Big Data”, *Annual Review of Economics*, 10(1), 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

Brave, S. A., Butters, R. A., and Justiniano, A. (2019). “Forecasting economic activity with mixed frequency BVARs”, *International Journal of Forecasting*, 35, 1692–1707

Bricongne, J-C., Coffinet, J., Delbos, J.-B., Kaiser, V., Kien, J.-N., Kintzler, E., Lestrade, A., Meunier, B., Mouliom, M., and Nicolas, T. (2020). “Tracking the economy during the Covid-19 pandemic: The contribution of high-frequency indicators”, *Bulletin de la Banque de France*, 231

- Bricongne, J.-C., Meunier, B., and Pical, T. (2021). “Can satellite data on air pollution predict industrial production?”, *Working papers*, No 847, Banque de France
- Bulligan, G., Marcellino, M., and Venditti, F. (2015). “Forecasting economic activity with targeted predictors”, *International Journal of Forecasting*, 31(1), 188–206
- Burns, A. F., and Mitchell, W. C. (1946). *Measuring Business Cycles*, NBER Book Series Studies in Business Cycles
- Caruso, A. (2018). “Nowcasting with the help of foreign indicators: The case of Mexico”, *Economic Modelling*, 69, 160–168
- 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*, 1–38
- Cascaldi-Garcia, D. (2022). “Pandemic Priors”, *International Finance Discussion Papers*, No 1352, Board of Governors of the Federal Reserve System
- Chen, C., and Liu, L.-M. (1993). “Joint Estimation of Model Parameters and Outlier Effects in Time Series”, *Journal of the American Statistical Association*, 88(421), 284–297
- Chernis, T., and Sekkel, R. (2017). “A dynamic factor model for nowcasting Canadian GDP growth”, *Empirical Economics*, 53, 217–234
- Chinn, M. D., Meunier, B., and Stumpner, S. (2023). “Nowcasting world trade with machine learning: a three-step approach”, *Working Paper Series*, No 2836, European Central Bank
- 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
- Consolo, A., Foroni, C., and Martínez Hernández, C. (2023). “A Mixed Frequency BVAR for the Euro Area Labour Market”, *Oxford Bulletin of Economics and Statistics*, 85(5), 1048–1082
- Dahlhaus, T., Guénette, J.-D., and Vasishtha, G. (2017). “Nowcasting BRIC+M in real time”, *International Journal of Forecasting*, 33(4), 915–935
- d’Aspremont, A., Ben Arous, S., Bricongne, J.-C., Lietti, B., and Meunier, B. (2024). “Satellites turn “concrete”: tracking cement with satellite data and neural networks”, *Working Paper Series*, No 2900, European Central Bank

- Dauphin, J-F., Dybczak, K., Maneely, M., Taheri Sanjani, M., Suphaphiphat, N., Wang, Y., and Zhang, H. (2022). “Nowcasting GDP – A Scalable Approach Using DFM, Machine Learning and Novel Data, Applied to European Economies”, *Working Papers*, No 2022/052, International Monetary Fund
- de Bondt, G., and Saiz, L. (2024). “Is the PMI a reliable indicator for nowcasting euro area real GDP?”, *Economic Bulletin Box*, 1, European Central Bank
- Delle Chiaie, S., and Pérez-Quirós, G. (2021). “High frequency indicators. Why? When? And how? A users’ guide”, *unpublished manuscript*
- Delle Chiaie, S., Ferrara, L., and Giannone, D. (2022). “Common factors of commodity prices”, *Journal of Applied Econometrics*, 37(3), 461–476
- 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
- de Valk, R., de Mattos, D., and Ferreira, P. (2019). “Nowcasting: An R Package for Predicting Economic Variables Using Dynamic Factor Models”, *The R Journal*, 11(1), 230–244
- Doan, T., Litterman, R., and Sims, C. (1984). “Forecasting and conditional projection using realistic prior distributions”, *Econometric Review*, 3(1), 1–100
- 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(1), 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
- ECB (2009). “New Procedure for constructing Eurosystem and ECB staff projection ranges”, <https://www.ecb.europa.eu/pub/pdf/other/newprocedureforprojections200912en.pdf>
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). “Least angle regression”, *Annals of Statistics*, 32(2), 407–499
- Elekdag, S., and Kannan, P. (2009). “Incorporating Market Information into the Construction of the Fan Chart”, *Working Papers*, No 2009/178, International Monetary Fund

- Engle, R. F., and Brown, S. J. (1986). "Model selection for forecasting", *Applied Mathematics and Computation*, 20(3-4), 313–327
- Evans, M. (2005). "Where Are We Now? Real-Time Estimates of the Macroeconomy", *International Journal of Central Banking*, 1(2)
- Falagiarda, M., and Sousa, J. (2015). "Forecasting euro area inflation using targeted predictors: is money coming back?", *Working Paper Series*, No 2015, European Central Bank
- Fan, J., and Lv, J. (2008). "Sure independence screening for ultrahigh dimensional feature space", *Journal of the Royal Statistical Society Series B*, 70(5), 849–911
- Ferrara, L., and Simoni, A. (2019). "When are Google data useful to nowcast GDP? An approach via pre-selection and shrinkage", *CREST Working Papers*, No 2019-04
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). "The Generalized Dynamic-Factor Model: Identification and Estimation", *The Review of Economics and Statistics*, 82(4), 540–554
- Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). "The MIDAS touch: Mixed data sampling regression models", *CIRANO Working Paper*, No 2004-20
- Giannone, D., Reichlin, L., and Simonelli, S. (2009). "Nowcasting euro area economic activity in real time: the role of confidence indicators", *National Institute Economic Review*, 210, 90–97
- Giannone, D., Reichlin, L., and Small, D. (2008). "Nowcasting: the real-time informational content of macroeconomic data", *Journal of Monetary Economics*, 55, 665–676
- Giannone, D., Lenza, M., and Primiceri, G. E. (2015). "Prior Selection for Vector Autoregressions", *The Review of Economics and Statistics*, 97(2), 436–451
- Granger, C., and Jeon, Y. (2004). "Thick modelling", *Economic Modelling*, 21(2), 323–343
- Guichard, S., and Rusticelli, E. (2011). "A Dynamic Factor Model for World Trade Growth", *Working Papers*, No 874, OECD Economics Department
- 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*, No 975, European Central Bank

- Hastie, T., Tibshirani, R., and Friedman, J. (2008). *The Elements of Statistical Learning*, Springer Series in Statistics, 2nd edition
- Hopp, D. (2021). “Economic Nowcasting with Long Short-Term Memory Artificial Neural Networks (LSTM)”, *UNCTAD Research Paper*, No 62, United Nations Conference on Trade and Development
- Jardet, C., and Meunier, B. (2022). “Nowcasting World GDP Growth with High-Frequency Data”, *Journal of Forecasting*, 41(6), 1181–1200
- Jarociński, M., and Maćkowiak, B. (2017). “Granger Causal Priority and Choice of Variables in Vector Autoregressions”, *The Review of Economics and Statistics*; 99(2), 319–329
- Jurado, K., Ludvigson, S., and Ng, S. (2015). “Measuring Uncertainty”, *American Economic Review*, 105(3), 1177–1216
- Kabundi, A., Nel, E., and Ruch, F. (2016). “Nowcasting Real GDP growth in South Africa”, *Working Papers*, No 581, Economic Research Southern Africa
- Kagraoka, Y. (2016). “Common dynamic factors in driving commodity prices: Implications of a generalized dynamic factor model”, *Economic Modelling*, 52(B), 609–617
- Knüppel, M. (2014). “Efficient estimation of forecast uncertainty based on recent forecast errors”, *International Journal of Forecasting*, Elsevier, vol. 30(2), 257–267
- Knotek, E., and Zaman, S. (2019). “Financial nowcasts and their usefulness in macroeconomic forecasting”, *International Journal of Forecasting*, 35(4), 1708–1724
- Koopman, S. J., and Harvey, A. (2003). “Computing observation weights for signal extraction and filtering”, *Journal of Economic Dynamics and Control*, 27(7), 1317–1333
- Lenza, M., Moutachaker, I., and Paredes, J. (2023), “Density Forecasts of Inflation: A Quantile Regression Forest Approach”, *Working Paper Series*, No 2830, European Central Bank
- Lenza, M., and Primiceri, G. E. (2022). “How to estimate a vector autoregression after March 2020”, *Journal of Applied Econometrics*, 37(4), 688–699
- Liebermann, J. (2014). “Real-Time Nowcasting of GDP: A Factor Model vs. Professional Forecasters”, *Oxford Bulletin of Economics and Statistics*, 76(6), 783–811

- Litterman, R. (1979). “Techniques of Forecasting Using Vector Autoregressions”, *Technical report*
- Makridakis, S., Hogarth, R. M., and Gaba, A. (2009). “Forecasting and uncertainty in the economic and business world”, *International Journal of Forecasting*, 25(4), 794–812
- Mariano, R., and Murasawa, Y. (2003). “A new coincident index of business cycles based on monthly and quarterly series”, *Journal of Applied Econometrics*, 2003, 18(4), 427–443
- Martínez-Martín, J., and Rusticelli, E. (2021). “Keeping track of global trade in real time”, *International Journal of Forecasting*, 37(1), 224–236
- McCracken, M. W., Owyang, M. T., and Sekhposyan, T. (2015). “RealTime Forecasting with a Large, Mixed Frequency, Bayesian VAR”, *Working Papers*, No 201530, Federal Reserve Bank of St. Louis
- Mittnik, S., and Zadrozny, P. (2005). “Forecasting Quarterly German GDP at Monthly Intervals Using Monthly Ifo Business Conditions Data”, in Sturm, J-E., and Wollmershäuser, T. (eds), *Ifo Survey Data in Business Cycle and Monetary Policy Analysis*, 19-48
- Mosley, L., Chan, T-K., and Gibberd, A. (2023). “sparseDFM: An R Package to Estimate Dynamic Factor Models with Sparse Loadings”, *arXiv*, DOI: <https://arxiv.org/abs/2303.14125>
- Palardy, J., and Ovaska, T. (2015). “Decomposing household, professional and market forecasts on inflation: a dynamic factor model analysis”, *Applied Economics*, 47(20), 2092–2101
- 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
- Proietti, T., and Giovannelli, A. (2021). “Nowcasting monthly GDP with big data: A model averaging approach”, *Journal of the Royal Statistical Society: Series A*, 184(2), 683–706
- Reifschneider, D., and Tulip, P. (2019). “Gauging the uncertainty of the economic outlook using historical forecasting errors: The Federal Reserve’s approach”, *International Journal of Forecasting*, 35(4), 1564–1582
- Rousseeuw, P., and Croux, C. (1993). “Alternatives to the Median Absolute Deviation”, *Journal of the American Statistical Association*, 88(424), 1273–1283

- Rünstler, G. (2016). "On the design of data sets for forecasting with dynamic factor models", *Working Paper Series*, No 1893, European Central Bank
- Rünstler, G., and Sedillot, F. (2003). "Short-term estimates of euro area real GDP by means of monthly data", *Working Paper Series*, No 276, European Central Bank
- Schorfheide, F., and Song, D. (2015). "Real-Time Forecasting with a Mixed-Frequency VAR", *Journal of Business & Economic Statistics*, 33(3), 366–380
- Schorfheide, F., and Song, D. (2021). "Real-Time Forecasting with a (Standard) Mixed-Frequency VAR During a Pandemic", *Working Papers*, No 29535, National Bureau of Economic Research
- Schumacher, C. (2010). "Factor forecasting using international targeted predictors: The case of German GDP", *Economics Letters*, 107(2), 95–98
- Soybilgen, B., and Yazgan, E. (2018). "Evaluating nowcasts of bridge equations with advanced combination schemes for the Turkish unemployment rate", *Economic Modelling*, 72, 99–108
- Stamer, V. (2024). "Thinking outside the container: A sparse partial least squares approach to forecasting trade flows", *International Journal of Forecasting*
- Stock, J. H., and Watson, M. W. (2002). "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economics Statistics*, 20, 147–162
- Tarsidin, Idham, and Nur Rakhman, R. (2018). "Nowcasting Household Consumption And Investment In Indonesia", *Bulletin of Monetary Economics and Banking*, 20(3), 1–30, Bank of Indonesia
- Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso", *Journal of Royal Statistical Society Series B*, 58(1), 267–288
- Wallis, K. F. (1989). "Macroeconomic forecasting: a survey", *The Economic Journal*, 99, 28–61
- Wang, Z., Zhu, Z., and Yu, C. (2023). "Variable Selection in Macroeconomic Forecasting with Many Predictors", *Econometrics and Statistics*
- Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*, Cengage Learning

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

Zarnowitz, V. (1987). "The Regularity of Business Cycles", *Working Papers*, No 2381, National Bureau of Economic Research

Zhang, H., Song, H., Wen, L., and Liu, C. (2021). "Forecasting tourism recovery amid COVID-19", *Annals of Tourism Research*, 87

Zou, H., and Hastie, T. (2005). "Regularization and variable selection via the elastic net", *Journal of Royal Statistical Society Series B*, 67(2), 301–320

Appendix A: Step-by-step guide for running the toolbox to build a nowcasting model

Step 1: Variable pre-selection

The first step consists in building a large dataset with all potential regressors (**Box A1**). Regressors can be monthly or quarterly; it must also include a quarterly target variable (e.g., US GDP). This can be done by gathering all variables relevant for the conjectural monitoring of the target (e.g., PMI surveys, retail sales, industrial production, financial variables). If the target variable relates to a given country, the dataset can also include variables for the external environment (e.g., global cycle, imports of main trading partners). A particular attention should be paid on the easiness for updates – as the goal is to build a nowcasting model to be updated regularly.

Box A1

Building the input *Excel* file in practice

Templates are provided in the folder ‘*dataset*’. The dataset must contain:

- A sheet ‘*Monthly*’ containing monthly potential regressors.
- A sheet ‘*Quarterly*’ containing quarterly potential regressors and the target variable. Please remember to always put the target variable in the rightmost column of ‘*Quarterly*’.

In both sheets, the code for the transformation should be entered on row 1. Please refer to sheet ‘*ReadMe*’ for the meaning of transformation codes.

In both sheets, the group number should be entered on row 2. The group numbers refer to group names that should be entered in the sheet ‘*Groups*’.

The dataset must be named ‘*data_XXX.xlsx*’ with XXX being the country ID. This country ID will be used in the variable pre-selection and model selection tools to automatically read the appropriate dataset.

NB: No need to fill the sheet ‘blocks’ at this stage. It should only be done after variable pre-selection.

The variable pre-selection step is based on the variable pre-selection tool which assesses the predictive power of the potential regressors. The literature on forecasting with factor models generally concludes that predictions are significantly more accurate when selecting fewer but more informative predictors (Bai and Ng, 2008). On a more theoretical ground, Boivin and Ng

(2006) find that larger datasets lead to poorer forecasting performances when idiosyncratic errors are cross-correlated or when the variables with higher predictive power are dominated. Against this background, pre-selection consists in keeping only regressors with the highest predictive power. The predictive power of regressors can be assessed by the variable pre-selection tool using three different techniques taken from the literature, in short:

- “SIS” (Sure Independence Screening; Fan and Lv, 2008) ranks regressors based on their pairwise correlation with the target variable.
- “T-stat-based” (Bair et al., 2006) remains a univariate method but adds in the dynamics of the target variable since the predictive power is tested *via* a regression of the target variable on the potential regressor and four lags of the target variable.
- “LARS” (Least Angle Regression; Efron et al., 2004 – see [Appendix C](#)) is a multivariate method that works as an iterative forward selection algorithm. It accounts for complementarities across regressors: said otherwise, while 10 similar regressors would end up having similar rankings in the first two methods, LARS would only select one and discard the others. This method is also generally found to perform better empirically (Jardet and Meunier, 2022; Chinn et al., 2023).

Box A2

Running the variable pre-selection tool in practice

Templates are provided in the folder ‘dataset’. The dataset must contain:

The code to run is *Variable_selection_v2.R* (located in the main folder). It requires:

- An *Excel* with potential regressors and the target variable in the folder ‘dataset’.
- Parameters on lines 48 to 56 of the code (country, target, evaluation period, number of lags or leads, method – a more extensive description can be found in the *R* code).
- A sub-folder named after the country in the folder ‘eval’.

The output of the variable pre-selection tool will automatically appear in the country sub-folders of the folder ‘eval’.

- It provides a csv with the ranking of regressors based on the method selected. The name of the csv contains information about the method used, the number of lags/leads (if any), and the evaluation period.
- On top of the ranking of regressors (from 1 to N), the csv contains information about the publication delay, the frequency, and the group of each variable.

- This complementary information is meant to help user's judgment for the pre-selection – e.g., to have a balanced mix of variables across groups or to discard variables with long publication delays.
- Publication delay is provided relative to the timeliest regressor, whose delay is set to 0. For other series, the publication delay is the lag (in months) relative to it.

A template Excel is available in the sub-folder '*0 – Tools for eval – templates and code*' to help with pre-selection, notably by constructing an aggregate score based on rankings across the different methods. Please refer to sheet '*ReadMe*'.

NB: The number of variables in the pre-selected set is decided by the user. As a rule of the thumb, it can be twice the maximum number of variables envisaged in the final model: as dynamic factor models typically have 10 to 30 variables, pre-selecting 40-60 variables is a fair starting point.

Pre-selection consists in narrowing the initial dataset with all potential regressors down to a subset containing only the regressors with highest predictive power. Pre-selection is informed by statistical methods, as well as by user's judgment on timeliness and variable groups. The variable pre-selection tool provides the ranking of all regressors by their predictive power (see **Box A2**). The choice of method(s) to run (SIS, t-stat-based, LARS) is left to the user. One possibility is to run all three methods and to compute an aggregate score based on the ranking in each method. The aggregate score can give different weights to the different methods, typically to give more weight to multivariate methods (LARS) that are found to perform better in the literature. Pre-selection can also consider rankings across different sub-periods (e.g., pre-GFC, post-Covid) and combine them. Importantly, pre-selection needs also to consider the publication delay: variables with earlier publication dates should be preferred to provide an advanced signal to the nowcasting models. Pre-selection should also consider selecting a balanced mix of variables to cover all relevant aspects of the economy. Multivariate methods such as LARS should in principle provides a balanced subset of variables, but the user could also adjust. After pre-selection, the user should prepare a *post-preselection Excel* file to be used in model selection (see **Box A3**).

Box A3

Preparing the post-pre-selection input *Excel* file in practice

After pre-selection, the dataset should only contain variables that survived pre-selection. It should be in the '*dataset*' folder. The dataset could be also adjusted in terms of groups (row 3 in '*Monthly*' and '*Quarterly*' and names in '*Groups*'). Remind to always put the target variable in the rightmost column of '*Quarterly*'.

This dataset should contain a sheet '*blocks*' for use in the model selection tool. On this sheet:

- Variable names should be on the rows and the different blocks on the columns.
- A 1 in the matrix means that the variable belongs to the block – or as often put, that the variable loads on the block factor.
- Variables (in rows) should be in the same order as in '*Monthly*' and '*Quarterly*'.
 - Monthly variables on top rows, then quarterly ones.
 - The model selection toolbox will automatically assume that first row relates to first monthly variable, second row to second monthly variable, and so on.
 - The target variable should also be included. As it is in the rightmost column of '*Quarterly*', it will be on the last row in sheet '*blocks*'.
- Generally, there should be a block '*Global*' containing all variables.
- Variables can belong to several blocks. In general, they belong to 'Global' and one other block, but they can also belong to multiple others.
 - As a rule of thumb, blocks should have at least 5 variables.
 - This often involves creating a catch-all block called '*Other*'.
- Groups (sheet '*Groups*') can be, in principle, different from blocks (sheet '*blocks*'). Groups are used to aggregate the impact of news releases and contributions (see **section 3**) while blocks are designed for the estimation of the model.

The new Excel file should be called '*data_XXX.xlsx*' (XXX the country ID) and replace the *Excel* file created for pre-selection.

Step 2: Model selection

Once pre-selection is achieved, the goal is to elect a nowcasting model that performs well in terms of the accuracy of the point forecasts and of the reliability of the forecast direction. The first objective is measured by the Root Mean Squared Error (RMSE) and the second by the Forecast Directional Accuracy (FDA). To do so, forecasters need to set model specifications to maximize such objectives. While this is generally done by forecasters through testing by hand numerous specifications – which can quickly become cumbersome, the model selection tool is versatile enough to automatically try a high number of model specifications and return their performances (see **Box A4**). The toolbox is highly flexible and allows the user to test a large variety of model specifications and compare their out-of-sample performances in a simple way, over a period selected by the user.

The model selection toolbox allows to automatically test many random model specifications, within the bounds defined by the user. Three types of model classes are available: Dynamic Factor Models of B  nbur   and Modugno (2014), large Bayesian Vector Auto-Regression of Cimadomo et al. (2022), and the combination of bridge equations    la B  nbur   et al. (2023).⁴⁵ For all models, bounds include the number of variables and the start of the estimation (in-sample) period. The user set also bounds for model-specific parameters: number of lags, and factors for the DFM, number of lags for the BVAR, and number of monthly and quarterly lags for the bridge equations – please see suggested values for these bounds in **Table A1**. Once the user has defined bounds for model specifications, the model selection tool tests many different models by picking random specifications within the bounds set by the user. For each model, the model selection computes the accuracy metrics (RMSE and FDA) for out-of-sample predictions of the quarter before (back-casting), the current quarter (nowcasting), and the next quarter (forecasting). Accuracy metrics are computed at the different months of the quarter (1st, 2nd, and 3rd months of the quarter) and across different sub-periods (full sample, pre-Covid, Covid defined as the year 2020, post-Covid, and full sample excl. Covid).

Table A1
Suggested parameter boundaries for model selection

Model class	Parameter	Suggested boundaries
Dynamic Factor Model	Number of variables	10 to 30
	Number of lags	1 to 4
	Number of factors	2 to 6
Large BVAR	Number of variables	8 to 20
	Number of lags	1 to 6
Bridge equations	Number of variables	8 to 20
	Number of lags (monthly series)	1 to 4
	Number of lags (quarterly series)	1 to 4
	Number of lags (endogenous var.)	1 to 4

Notes: Values are suggestions based on the literature and empirical tests. In bridge equations, the number of lags for monthly regressors is expressed at quarterly frequency.

⁴⁵ Among the three BVAR methods proposed in Cimadomo et al. (2023), the toolbox includes the blocking / stacking method (B-BVAR in Cimadomo et al., 2022) which consists in treating the higher-frequency data as multiple lower-frequency variables. This method is chosen over the two others (L-BVAR and C-BVAR) due to lower computation costs. In terms of performance, Cimadomo et al. (2022) establish that all three methods have very close out-of-sample accuracy.

Rather than running over random models, the toolbox also allows to test over user-specified models. The user can define the list of model specifications that the toolbox should test, following the instructions provided in [Box A6](#).⁴⁶

The comparison of predictions is based on pseudo-out-of-sample monthly predictions, over a period that is defined by the user. The process is recursive. For each month of the out-of-sample period (defined by the user, see [Box A4](#)), the toolbox reproduces the structure of missing data to create a pseudo real-time dataset;⁴⁷ then, model parameters are estimated on this dataset and (pseudo) out-of-sample predictions are computed. The toolbox then computes the accuracy metrics (RMSE and FDA) over the out-of-sample period set by the user. In general, the out-of-sample period should be sufficiently long to avoid a potential overfitting of the model on a specific period.⁴⁸

Upon testing across all the desired models, the toolbox creates a summary file with accuracy scores across all models. From there, the user can select the best-performing models based on the metrics (and sub-periods, months of the quarter, and horizons) that are deemed of importance.

Box A4

Running the model selection toolbox in practice

The code to run is *Nowcast_main.m* (located in the main folder). Parameters to be adjusted are:

- *do_eval* should be set to 1.
- *do_loop* should be set to 1.
- *country.name* should be set to the desired country ID.
- *country.model* should be set to the desired model class ('DFM', 'BVAR' or 'BEQ').
 - For DFM, *Par.block_factors* should be set to 1 to run a model with block factors, and to 0 otherwise. Please note that DFM with block factors takes much longer to run.
- The evaluation period (out-of-sample) is defined with:
 - *Eval.eval_startyear* = starting year.

⁴⁶ The only difference relates to the fact that the user would have to disconnect the option to test across all possible Covid corrections (see [Box A6](#)).

⁴⁷ The dataset created this way is based on the implicit assumption that publication delays are constant over time. It is qualified of pseudo real-time because it does not account for data revisions.

⁴⁸ As a rule of the thumb, the 2023 revision of the nowcasting models has been conducted over a period of 10 years, covering a period longer than a typical business cycle (Zarnowitz, 1987).

- *Eval.eval_startyear* = starting month.
- *Eval.eval_endyear* = ending year.
- *Eval.eval_endmonth* = ending month.
- *Eval.gdp_rel* should be set to the month (of the following quarter) where the target variable becomes available. For example, 2 means that the target variable is available on the 2nd month of the quarter – e.g., US GDP for 2023Q1 is available in April 2023 (2nd month of Q2).
- In evaluation mode, the dataset should be “frozen” to ensure that evaluations of different models are performed on the same information set. The user must enter the date at which the dataset has been frozen with:
 - *Eval.data_update_lastyear* = year of the update
 - *Eval.data_update_lastmonth* = year of the update
 - The reason is that the toolbox fills missing observations based on when it is run (i.e., taking today’s date). The parameters *Eval.data_update_lastyear* and *Eval.data_update_lastmonth* overwrite today’s date to avoid an unwanted filling of missing observations.
- Parameters for the models to be automatically tested are set in the Loop structure. Common parameters for any model class are:
 - *Loop.n_iter* is the number of models to be (automatically) tested.
 - *Loop.name_loop* is the ID of the loop.
 - *Loop.min_startyear* and *Loop.max_startyear* are the bounds (respectively minimum and maximum) of the start year for the estimation (in-sample).
 - *Loop.startmonth* is the start month for estimation, for any year.
 - *Loop.min_var* and *Loop.max_var* are the bounds for the number of regressors to be included in the model.
- Parameters specific to the DFM are:
 - *Loop.min_p* and *Loop.max_p* for the number of lags.
 - *Loop.min_r* and *Loop.max_r* for the number of factors.
- Parameters specific to the bridge equations are:
 - *Loop.min_lagM* and *Loop.max_lagM* are the bounds for the number of lags for the monthly regressors.
 - *Loop.min_lagQ* and *Loop.max_lagQ* are the bounds for the number of lags for the quarterly regressors.
 - *Loop.min_lagY* and *Loop.max_lagY* are the bounds for the number of lags for the endogenous variable (the target variable).

- Parameters specific to the BVAR are *Loop.min_bvar_lags* and *Loop.max_bvar_lags*, the bounds for the number of lags.
- *Loop.do_random* is a switch on whether to randomize completely the model specifications. It should be generally set to 1. When set to 0, two different runs of the code will provide the same model specifications. It should therefore not be set to 0 for testing many different models. But it is intended to be used to check the impact of a parameter outside of the Loop structure (e.g., *Par.block_factors* or *do_Covid*) on the same set of random specifications.

The outputs of the model selection tool will automatically appear in the country sub-folders of the folder 'eval'. The code delivers two types of *Excel* files:

- One for each tested model with all out-of-sample predictions compared to the actual outcome and to the predictions of an AR model.
 - It includes three separate sheets: 'Bac' for back-casting (predictions made for the past quarter = t-1), 'Now' for nowcasting (quarter t), and 'For' for forecasting (quarter t+1).
 - At the bottom of each sheet are the accuracy metrics (RMSE and FDA) for the model.
 - It also includes a sheet 'Parameters' with model specifications.
- A summary file containing all specifications tested and their accuracy metrics (RMSE and forecast directional accuracy). Each row corresponds to a different model. This is the main file to be used for selecting the best-performing model.

NB: At this stage, the parameter do_Covid can be left to 0 (no correction) as it will be tested at a later stage (Box A5). Users willing to test Covid corrections can nevertheless set it to different values (1 to 3).

Box A5

Selecting best-performing model(s) in practice

The first step is to gather accuracy metrics from all models tested.

- As mentioned in **Box A4**, a summary *Excel* file containing all specifications and their accuracy metrics is produced at each run of the model selection toolbox.
- Some loops might crash due to *Matlab*, virtual machine stoppages, or wrong specifications. In this case, no summary *Excel* file is produced but the additional *Matlab* function 'Get_interrupted_loops.m' allows to recover a summary of models that had run prior to the crash. This function is in the sub-folder 'Tools for eval (templates and code)'. Please refer to the instructions on top of the code.

The second step is to compare models. A template is available in the sub-folder 'Tools for eval (templates and code)' to help the comparison: it helps ranking models based on the criteria that are important to the user. There is one template per model class, for example the one for DFM is called 'XXX_DFM_ALL_template.xlsx'.

- The user only has to copy-paste the accuracy metrics from the summary *Excel* file(s) into the corresponding sheets of the template.
- The template then computes a score for every model by comparing the accuracy to a benchmark. Scores above 0 mean that the model outperform the benchmark. A higher score means a better model. Please refer to the 'Readme' sheet for more details on how to use the file.
- The template offers the possibility to adjust the weights given to the different metrics (RMSE or FDA), the month of forecasts (overall, first month, second month, third month), or the sub-periods (full, pre-Covid, Covid, post-Covid, no Covid). The weight given to each horizon (back-casting, now-casting, and forecasting) can also be adjusted. These settings are entered in the '*All_horizons*' sheet.

The last step consists in using the sheet '*All_horizons*' to rank models and elect the best-performing one(s).

- A mechanical ranking can be obtained by filtering on the column 'Agg. score' (column S) in the sheet '*All_horizons*' of the template.
- The mechanical ranking should be complemented with other considerations such as an appropriate mixture across variable groups, decent metrics across all sub-periods, etc. One possibility can be to identify key variables that should always be included. Then, going through the list of best-performing models, the user can exclude the specifications that did not have those key variables.⁴⁹

*NB: A similar procedure should be applied for selecting the final model after Covid robustness, see **Box A7** for more details.*

Step 3: Covid robustness

Once the user has selected a few model specifications that perform best on the full sample, the last step consists in testing if Covid-specific adjustments can further enhance the performances on the post-Covid sample. The rationale is that Covid observations, characterized by dramatic values, might alter the functioning of the models (Carreiro et al., 2022; Lenza and Primiceri, 2022; Almuzara et al., 2023).

The process consists in running a model selection in the vein of step 2, but this time (i) restricted to only a few user-defined specifications, (b) focused on the post-Covid sample period, and (c) where Covid corrections are tested. The few user-defined specifications are the ones selected after the initial model selection (**Box A5**). On top of testing through random

⁴⁹ Another possibility, rather than discarding the models without the key variables, can be to alter slightly their specification by putting the key variable into the specification. Since specifications are to be tested again in the last step (post-Covid model selection), these augmented models will be tested again. In case the addition of the new variable worsens the performances, the specification can then be discarded at a later stage.

specifications as in step 2, the toolbox allows to test over a set of model specifications defined by the user. The toolbox runs each user-defined specification four times: once without any Covid correction – as a benchmark, and once with each Covid correction. Like for model selection in step 2, the toolbox produces a summary file with all specifications and their accuracy metrics. The procedure is described in **Box A6**.

Once the final model selection is run, the user should be able to select the best performing model. The process to select the final model remains the same as described in **Box A5**. *In fine*, the initial model selection ensures that the model performs well in general – over the full sample, while the final model selection ensures that the model remain highly performant after Covid.

Box A6

Covid robustness checks in practice

Get a set of best-performing models (about 25) based on the procedure described in **Box A5**.

- For use by the toolbox, the list of specifications should be put in the template '*Eval_list_.xlsx*' located in the sub-folder '*Tools for eval (templates and code)*'. Please note there are different templates for each model class. Please refer to the '*Readme*' sheet of the template for more detailed information.

The code to run is *Nowcast_main.m* as in the initial model selection (**Box A4**), but this time after selecting the option *do_loop* should be set to 2. Other parameters to adjust are:

- *Loop.name_customloop* is the name of the custom loop.
- *Loop.list_name* is the name of the Excel file containing the list of potential models. Kind reminder that the list of specifications should be put in the dedicated template.
- *Loop.alter_Covid* should be set to 1 to test all possible Covid corrections.
 - When set to 1, the toolbox tests each specification from the list four times: once with no Covid correction, and once with each Covid correction.
 - When the toolbox is used to test across user-defined specifications, this option should be set to 0 (while still leaving *do_loop* set to 2 as indicated above). Then each specification is tested only once, based on the parameter *do_Covid* set by the user in the code.

As was the case for the initial model selection (**Box A4**), outputs will automatically appear in the country sub-folders of the folder '*eval*'. The code again delivers two types of *Excel* files:

- One for each tested model with all out-of-sample predictions compared to the actual outcome and to the predictions of an AR model.

- It includes three separate sheets: '*Bac*' for back-casting (predictions made for the past quarter = $t-1$), '*Now*' for nowcasting (quarter t), and '*For*' for forecasting (quarter $t+1$).
- At the bottom of each sheet are the accuracy metrics (RMSE and FDA) for the model.
- It also includes a sheet '*Parameters*' with model specifications.

*NB: Once the Covid robustness checks are run, the selection of the best model follows the steps showed in **Box A5**.*

Appendix B: Step-by-step guide for running the toolbox to get policy outputs

The update procedure is straightforward and starts with updating the input dataset. The next step is to run the code (**Box B1**). Results are automatically put in sub-folder “*output/your country*”. It creates *Matlab* and *Excel* vintages at each run of the code. The *Excel* file contains the outputs detailed in **section 3**.

Box B1

Updating the model nowcasts in practice

Open the *Matlab* file ‘*Nowcast_Main_v10.m*’. Update the option ‘*country.name*’ with the country / aggregate you wish to nowcast.

In addition, two options can be activated to provide further outputs. Both can be de-activated (set to 0) to produce a simple nowcast update. They are:

- ‘*do_range*’ computes a range of alternative models (see **section 3.5**). Activating this option increases the computation time by around 15-20 minutes.
- ‘*do_mae*’ computes the mean absolute error over past predictions (see **section 3.4**) to produce uncertainty bands. Activating this option increases the computation time by around 15-20 minutes. If the option is de-activated, uncertainty bands will be based on user-specified values.

Make sure that:

- ‘*country.model*’ is set to the model class corresponding to the country.
- ‘*do_eval*’ is set to 0 – otherwise it evaluates the model but do not produce nowcasts.
- ‘*do_loop*’ is set to 0 – otherwise it loops on models and evaluates them.
- ‘*do_subset*’ is set to 0 – otherwise it takes only a subset of input variables.
 - The option ‘*do_subset*’ should be set to 1 only when the user wants to exclude some specific variables from the information set. The selection of variables to be kept in the model should be specified by the user in the unidimensional array ‘*var_keep*’.

Simply click the ‘*Run*’ button to start the estimation. Depending on the options selected and the size of the model, fully running the code may take up to 45 minutes.

A special case of running the toolbox relates to changes of models. While the toolbox can tackle model changes, some outputs are mechanically de-activated (**Box B2**).

Box B2

What happens if the model changes between two runs

In case the model changes (e.g., addition of new input variables), no news decomposition will be produced. All other model outputs such point forecasts, errors, heatmap, etc. are left intact. Note that the sheets named after quarters (such as ‘2023Q4’) will indicate when a change of model occurred.

When adding variables, make sure the transformation and group are up to date. Make sure also that they are reflected in the ‘*blocks*’ sheet – even if it might not use blocks. In the ‘*Quarterly*’ sheet, the target variable should always be the rightmost column.

Appendix C: Supplementary material

1. Model settings

The underlying idea of the (automated) model selection tool is to test across a range of model specifications. Model specifications are either randomized, within bounds that are defined by the user, or pre-defined by the user. In this exercise, the toolbox (or the user) chooses the different settings in **Table C1**.

Table C1

Model settings randomized or set by the user in model selection

Model class	Setting
Common across all classes	Number of variables
	Start date of the estimation period (in-sample)
Dynamic Factor Model	Number of lags
	Number of factors
Large BVAR	Number of lags
Bridge equations	Number of lags (monthly regressors)
	Number of lags (quarterly regressors)
	Number of lags (endogenous variable)

Note: In bridge equations, the number of lags for monthly regressors is expressed at quarterly frequency.

It should be noted that other settings can be changed by the user in the toolbox; however, they are common settings that will apply identically across all the specifications.⁵⁰ They include the Covid correction (see **section 2.3**) as well as model-specific parameters:

- For the DFM: the estimation of block factors, the assumptions on the idiosyncratic components (AR or i.i.d.), the threshold for convergence in expectation-maximization (EM) algorithm, and the maximum number of iterations in EM algorithm.

⁵⁰ The distinction between parameters that change across specifications and parameters that remain common across all specifications is based on which ones relate rather to the estimation strategy or to the model specification. For example, in the DFM, the convergence threshold and number of iterations in the expectation-maximization algorithm relate more to the estimation strategy – and are then common across all specifications. By contrast, the number of lags and factors rather relate to the model specification – and are then modified between different specifications.

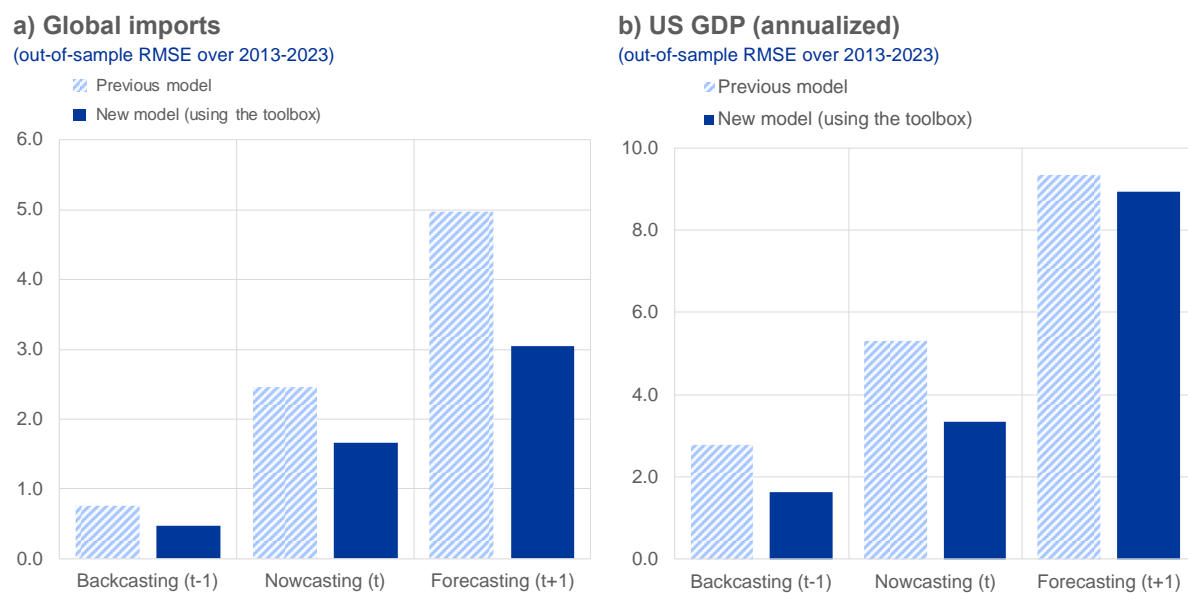
- For the large BVAR: the threshold for convergence in expectation-maximization (EM) algorithm and the maximum number of iterations in EM algorithm.
- For the bridge equations: quarterly dummies and the type of extrapolation used for the regressors. For the latter, the extrapolation always uses a BVAR but the methods in the toolbox depend on the scope of variables considered to run the extrapolation. More precisely, it uses either:⁵¹
 1. All regressors together: in this case the extrapolation is run once, and all bridge equations use the same extrapolated set.
 2. Only regressors that enter in each bridge equation: in this case, a different extrapolation is run for each bridge equation. This means one variable can have different extrapolated values for different bridge equations.
 3. Univariate (i.e., one regressor at a time): in this case, the extrapolation is run once for each variable and all bridge equations will use the same extrapolated set.
 4. A combination of the three above, meaning that the same bridge equations are run three times, one with each extrapolation strategy.

⁵¹ Empirically, the first method (all regressors together) is found to yield the highest performances. It is the default setting in the code but can be changed by the user: this is the parameter *Par.type* (set to 901 by default).

2. Additional target variables

Figure C2

Accuracy of the best-performing model compared with previous model



Sources: Bloomberg, Haver, S&P Global, and authors' calculation.

Notes: Out-of-sample predictions are done recursively in a pseudo real-time set-up. The new model is without Covid-19 correction until Dec. 2020 and with correction afterwards.

Acknowledgements

We are very grateful to M-G. Attinasi, L. Boeckelmann, D. Brignone, S. Delle Chiaie, A. Dieppe, J. Doleschel, R. Gerinovics, R. Gomez Salvador, K. Ka, S. Makridakis, A. Schmidt, B. Schnatz, and M. Tirpak for their useful comments and contributions during the coding and first use of the nowcasting toolbox. We warmly thank S. Delle Chiaie, F. Kurcz, and G. Perez-Quirós who coded previous iterations of nowcasting tools in the division. We are indebted to the original authors of the code for the various techniques in the toolbox: M. Bańbura and M. Mudugno for the dynamic factor model; J. Cimadomo, D. Giannone, M. Lenza, F. Monti, and A. Sokol for the large Bayesian vector auto-regression; M. Bańbura, K. Bodnar, M. B. Toth, and I. Belousova for the combination of bridge equations; and M. Chinn and S. Stumpner for variable selection methods. We are also very grateful to B. Bok, D. Caratelli, D. Giannone, A. Sbordone, and A. Tambalotti for making available the underlying code of the New York Fed nowcasting model.

Jan Linzenich

European Central Bank, Frankfurt am Main, Germany; email: jan.linzenich@ecb.europa.eu

Baptiste Meunier

European Central Bank, Frankfurt am Main, Germany; Aix-Marseille School of Economics (AMSE), Marseille, France;
email: baptiste.meunier@ecb.europa.eu

© European Central Bank, 2024

Postal address 60640 Frankfurt am Main, Germany

Telephone +49 69 1344 0

Website www.ecb.europa.eu

All rights reserved. Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors.

This paper can be downloaded without charge from www.ecb.europa.eu, from the [Social Science Research Network electronic library](#) or from [RePEc: Research Papers in Economics](#). Information on all of the papers published in the ECB Working Paper Series can be found on the [ECB's website](#).

PDF

ISBN 978-92-899-6942-0

ISSN 1725-2806

doi:10.2866/7091049

QB-01-24-045-EN-N