WORKING paper

BANQUE DE FRANCE
EUROSYSTÈME

A Mixed-Frequency Factor Model for Nowcasting French GDP

Julien André[1] & Marie Bessec[2]

December 2024, WP #975

ABSTRACT

This article presents a new nowcasting model for quarterly real GDP growth in France, developed at the Banque de France. The model is designed to forecast the first release of GDP growth at the end of each month within the quarter in question. The model belongs to the class of targeted factor models and it is estimated using the mixed-frequency three-pass regression filter. We estimate the model on a large set of monthly indicators. The Banque de France survey variables on manufacturing and services are particularly useful for estimating the factors. We extend the formulae for the contributions of the predictors in the mixed-frequency case, and show that, beyond a positive constant level of growth (the intercept), all groups of normalised supply-side and demand-side variables have contributed negatively to GDP growth since the onset of COVID-19 pandemic. A pseudo-real-time evaluation of the method shows the good performance of the model compared to several simple benchmarks and the existing MIBA tool used at the Banque de France, especially during the critical first two months of each quarter. The forecasting combination of the MIBA tool and the new model also performs well at the shortest horizon. In the robustness analysis, we show that this model outperforms a large set of alternative specifications.

Keywords: GDP Nowcasting, Factor Model, Mixed-Frequency

JEL classification: C22, E32, E37

[1] Banque de France, julien.andre@banque-france.fr.
[2] Université Paris Dauphine, Université PSL, LEDa, CNRS, IRD, marie.bessec@dauphine.psl.eu. We are grateful to M. Coueffe, O. Darné, O. de Bandt, C. Doz, E. Jousselin, Y. Kalantzis, M. Marx, M. Mogliani and S. Rivaud for their useful comments and suggestions. The paper has also benefited from discussions with participants at the Banque de France Research Seminar, the 17th International Conference on Computational Financial Econometrics (CFE) in Berlin in 2023 and the 44th International Symposium on Forecasting (ISF) in Dijon in 2024.

NON-TECHNICAL SUMMARY

In this article, we present a new model aimed at nowcasting the first release of French GDP growth at the end of each month within the quarter in question. This new model is called MF-3PRF (Mixed-Frequency Three-Pass Regression Filter) with reference to the estimation procedure. The estimation method was introduced by Kelly and Pruitt (2015), and provides a strategy for extracting the relevant common factors from a large dataset to forecast a target variable, here GDP growth. Hepenstrick and Marcellino (2019) have extended this approach to cope with mixed-frequency data. This extension allows one to consider a target variable that is sampled quarterly and predictors that are available at a monthly frequency, as it is typically the case in short-run forecasting. To estimate this model, we use a database consisting of 60 monthly indicators. The latter contains survey data (EMC data from the Banque de France), hard data (Industrial Production Index, Services Production Index, construction and employment data), financial data (monetary aggregates, stock and price data), international data (Economic Sentiment Index and IPI in Germany and the euro area), and an index of Economic Policy Uncertainty.

Figure N1. Weights of the predictors in the factor

Note: This figure represents the variables according to their weight in the factor estimated for the last forecasting equation. The top 5 variables are taken from the Banque de France survey on manufacturing industry (EVPRO, EVLIV and EVCOM) and services (DETEM and PREVACT).

Codes for the survey variables in manufacturing industry: EVPRO = Change in output compared with previous month; EVLIV = Change in deliveries compared with previous month; EVCOM = Change in overall level of new orders compared with previous month. Codes for the survey variables in services: DETEM = Change in aggregate demand compared with previous month; PREVACT = Expected overall activity over next month.

The main findings of this new approach to forecasting French GDP growth are as follows. First, to assess the importance of each variables in the factor building, we compute their weights as the absolute value of their correlation coefficient with GDP growth. As figure N1 shows, the Banque de France survey variables on manufacturing and services are very useful for forecasting French GDP growth in this framework since their weights are prominent in the estimation of the factors. In fact, the top 5 variables in terms of weight come from the Banque de France survey on manufacturing (EVPRO, EVLIV and EVCOM) and services (DETEM and PREVACT). Second, we have derived a formula to calculate the contributions of each predictor in the mixed-frequency context, and it

allows us to show that, beyond a positive intercept measuring average growth, all groups of normalized supply and demand variables have contributed negatively to GDP growth since the onset of the COVID-19 pandemic. This is evidence of growth below average since the Covid-19 period. We have also conducted some exercises to assess the ability of the model to produce accurate nowcasts of French GDP growth with respect to several simple benchmarks and existing tools at the Banque de France. In particular, we have compared the performance of our model with that of the MIBA model, which performed very well before the pandemic. We find that our new model exhibits better results than MIBA in the first two months of the nowcast quarter. In the third month, the MF-3PRF model is still useful, as it outperforms MIBA taken alone when the forecasts of the two models are combined with a simple arithmetic mean. Finally, we have also conducted a robustness analysis to challenge our reference specification with several variant specifications.

Un modèle à facteurs à fréquence mixte pour le nowcasting du PIB français

RÉSUMÉ

Cet article présente un nouveau modèle de *nowcasting* du taux de croissance trimestriel du PIB réel en France, développé à la Banque de France. Le modèle est destiné à prévoir la première publication de la croissance du PIB à la fin de chaque mois du trimestre concerné. Nous utilisons un modèle à facteurs ciblés estimé à l'aide de la méthode *three-pass regression filter* en fréquence mixte. Le modèle est estimé sur un large ensemble d'indicateurs mensuels. Les variables d'enquête de la Banque de France dans l'industrie manufacturière et les services s'avèrent très utiles dans le calcul des facteurs. Nous étendons les formules des contributions des prédicteurs au cas de données de fréquences mixtes et montrons qu'au-delà du terme constant, mesurant la croissance moyenne, tous les groupes de variables d'offre et de demande centrées et réduites ont pesé négativement sur la croissance du PIB depuis le début de la pandémie. Une évaluation en pseudo-temps réel montre la bonne performance du modèle par rapport à plusieurs benchmarks simples et aux outils existants à la Banque de France, notamment aux deux premiers mois du trimestre. La combinaison des prévisions du modèle ISMA et du nouveau modèle est aussi performante à l'horizon le plus court. Dans l'analyse de robustesse, nous montrons la supériorité du modèle sur un grand nombre de spécifications alternatives.

Mots-clés : nowcasting du PIB, modèle à facteurs, fréquences mixtes

---

1 Introduction

The French economy has recently faced significant shocks, including global events such as the COVID-19 pandemic and conflicts in Ukraine and the Middle East, as well as domestic ones such as the yellow vest movement in 2019 and the pension reform strikes in late 2019 and early 2023. At the end of 2023, the Banque de France uncertainty index, which is based on the comments of respondents to their business survey, remained above pre-COVID levels. In this context, nowcasting tools are particularly useful for policymakers, providing a timely and accurate snapshot of the current state of the economy.

The Banque de France currently employs several approaches for nowcasting quarterly Gross Domestic Product (GDP) growth in France. One of these methods is the MIBA (Monthly Index of Business Activity nowcasting) model and has been developed by Mogliani et al. (2017). The MIBA model is an unconstrained mixed-data sampling (U-MIDAS) specification (Foroni et al. (2015)) with a preselection of variables from the Banque de France's manufacturing industry survey for each month and it is designed to target the initial estimates of GDP growth. This model delivers a nowcast of French GDP growth, that is subsequently published alongside the business surveys. It performed very well in the pre-covid period. Second, the PRISME (Prévision Intégrée Sectorielle Mensuelle) model developed by Thubin et al. (2016) offers an alternative forecast by aggregating the value added forecasts of six sectors: market services, manufacturing, construction, energy, non-market services, and agriculture. The PRISME model is particularly useful in scenarios where there is a divergence between the manufacturing sector and other sectors of the economy. Furthermore, it provides insights into the individual contributions of these sectors to the GDP growth forecast.

In this paper, we consider another approach, a factor model estimated with the mixed-frequency three-pass regression filter (hereafter MF-3PRF). This particular estimation method derives from a large dataset a subset of factors that are useful for forecasting a variable of interest, in our case French GDP growth, while discarding the others, in contrast to principal component based methods. The original method (3PRF) was proposed by Kelly and Pruitt (2015), and Hepenstrick and Marcellino (2019) have extended the method to the case where the dataset contains indicators sampled at a higher frequency than the target variable and possibly ragged edges. The authors show an improvement over the usual factor models, in line with the literature on factor models with targeted predictors since the seminal paper by Bai and Ng (2008). Other applications showing the potential of the three-pass regression filter in forecasting include Guérin et al. (2020) for several variables related to the US economic activity and bilateral exchange rates, Marcellino and Sivec (2021) and Chow et al. (2023) for real GDP growth in Luxembourg and Singapore, Chatelais et al. (2023) for US economic activity with sectoral dividend yield series, and Liang et al. (2024) for stock volatility.

We apply the MF-3PRF approach to nowcast French GDP growth using a dataset of 60 monthly indicators. We derive three nowcasting equations that can be used at the end of each month of the quarter to be nowcast. As in the MIBA model, we explicitly target the first release of French GDP growth. Among the 60 indicators, the Banque de France survey variables play an important role in estimating the factors at the three horizons, confirming their usefulness for nowcasting French GDP growth. The fit of the model improves over the quarter as more information becomes available and it dominates the fit of the MIBA model in the last two months. We extend the formulae for the contributions of the predictors of Kelly and Pruitt (2015) to the mixed-frequency case and we analyze the contributions of different groups of demand and supply-side variables to French growth. We find that all groups of variables have a negative impact on French growth since the outbreak of the pandemic crisis in 2020.

We perform a pseudo-real-time evaluation of the three nowcasting equations over the last decade. We evaluate the performance of the model over the whole period, as well as in the pre- and post-covid periods. We show that the new model performs well compared to simple benchmarks and existing tools at the Banque de France, especially during the first two months of the quarter. The forecasting combination of MIBA and the new model also performs well at the shortest horizon. These results are valid for the different evaluation windows and are robust to the consideration of a large number of variants regarding the specification, the treatment of missing values and the aggregation scheme of the monthly predictors in the first step of the method. The use of mixed-frequency data is also beneficial.

The new model represents an intermediate approach within the range of forecasting tools used at the Banque de France to forecast French GDP growth in the short run. It makes it possible to derive three nowcasts of the first releases of the French GDP growth concomitantly to the publication of the Banque de France's business surveys. Like the PRISME model, it can exploit the information provided by a large database and its forecasting performance is close to that of the MIBA model in the pseudo real-time evaluation. It is also based on ordinary least squares (OLS) regressions and, unlike traditional factor models, it offers ease of use and practicality. In addition, the results derived from the model are easier to interpret and provide a clearer understanding of the underlying factors. Of particular interest is the ability to derive variable contributions that allows a sectoral interpretation, although less detailed than in the PRISME model. In particular, we can assess the role of services sector, an ingredient that was missing with the MIBA model.

The rest of the paper is structured as follows. Section 2 is devoted to the specification and the estimation method. We also derive the contribution of the predictors in the case of mixed frequencies. In Section 3, we evaluate the forecasting performance of the model on French data, both in-sample and out-of-sample, in comparison with various benchmarks, including the MIBA model. Finally, section 4 concludes.

2 Forecasting method

2.1 Model

Let y be the variable to be forecast at horizon h and x be a large set of variables. There is now a long tradition in forecasting of using factor models as a reduction-information strategy. In this paper, we follow Kelly and Pruitt (2015) and assume that the factors driving the target variable y can be a subset of the factors driving the set of predictors x.

Formally, Kelly and Pruitt (2015) consider the following model:

$$y_{t+h} = \beta_0 + \beta' F_t + \eta_{t+h}$$
$$z_t = \lambda_0 + \Lambda F_t + \omega_t$$
$$x_t = \phi_0 + \Phi F_t + \varepsilon_t$$

for $t = 1, \ldots, T-h$ where y is the target variable, $F_t = (f_t', g_t')'$ the $K = K_f + K_g$ common factors to all variables x, $\beta = (\beta_f', 0')'$ so that y only depends on f and not on g, z is a proxy driven by the same underlying forces than y so that $\Lambda = (\Lambda_f, 0)$ and x is a large set of N weakly stationary variables driven by f and g. Since we consider the forecast of the current quarter, the forecast horizon h is set to zero in the following.

The main issue with this specification is that the variation in x_t is driven by more factors than the variation in the target variable y_t. Including more factors than those strictly needed in the first equation can be detrimental to forecasting in small samples or when g_t are strong factors while f_t are weak. Kelly and Pruitt (2015) have proposed a 3-step method for estimating the only part f_t that is useful for forecasting y_t. In contrast to the usual Principal Component Regression (PCR) approach, where the weights used to construct factors are estimated according to the correlations within the predictors, in the 3PRF method the weights used to construct factors are estimated according to their correlation with the target variable. Kelly and Pruitt (2015) have demonstrated the asymptotic optimality of the method. The estimated factors are consistent and the 3PRF forecast $\hat{y}_{t+h}$ is consistent for the infeasible best forecast $\beta_0 + \beta' F_t$ for large T and N. Hepenstrick and Marcellino (2019) have extended the estimation method in the case where the variables x_t are observed at a higher frequency than y_t. We describe this extension below.

2.2 Estimation and nowcast

We present the estimation method extended to mixed-frequency data. In the following, τ denotes the low frequency time unit (quarterly) and t the high frequency time unit (monthly). We consider the case of a quarterly variable y_τ observed on $T/3-1$ quarters that we want to nowcast in quarter $T/3$ with N monthly predictors x_{it} observed on $T-3+m$ months, $m = 1, 2, 3$ depending on the month of the nowcast m. The estimation of the model consists of three steps with OLS regressions.

In the *first step*, we run the following (time series) regression at the quarterly frequency for each variable i, $i = 1, \ldots, N$:

$$x_{i\tau} = \alpha_{0,i} + \alpha_{1,i} z_\tau + \varepsilon_{i,\tau} \quad \tau = 1, \ldots, T/3-1$$

with $T/3-1$ the number of quarters and z_τ a proxy variable driven by the same factors than the target variable y_τ. We store the N estimates $\hat{\alpha}_{1,i}$ in a vector(s). In this step, each predictor x_{it} is converted to the quarterly frequency (e.g., by average or by taking a specific month of each quarter). The choice of aggregation scheme is not neutral, as we will see later.

In the *second step*, we run a (cross-sectional) regression for each month $t = 1, \ldots, T-3+m$:

$$x_{it} = \phi_{0,t} + \hat{\alpha}_{1,i} F_t + \varepsilon_{i,t} \quad i = 1, \ldots, N$$

with $T-3+m$ the number of months. We collect the $T-3+m$ estimates $\hat{F}_t$ in a vector(s). The last observations $T-2,\ldots,T-3+m$ of the factor are used to compute the nowcast.

In the *third step*, we split the monthly factor(s) $\hat{F}_t$ of dimension $T-3+m$ in 3 quarterly factors of dimension T/3 (or T/3-1). The first variable $\hat{F}_{1,\tau}$ consists of the first months of each quarter (that are January, April, July and October), $\hat{F}_{2,\tau}$ of the second months and $\hat{F}_{3,\tau}$ of the third months. Then, we run a (time series) regression of the target variable on the first T/3-1 observations of the factors and their lags. At this level, we use horizon-specific U-MIDAS equations, with the only m contemporaneous months of the factors available at the time of the forecast:

$$y_\tau = \beta_0 + \sum_{i=1}^{p} \gamma_i y_{\tau-i} + \sum_{r=1}^{m} \beta_{r,0} \hat{F}_{r,\tau} + \sum_{j=1}^{q} \sum_{r=1}^{3} \beta_{r,j} \hat{F}_{r,\tau-j} + \eta_\tau$$

for $\tau = \max(p,q)+1,\ldots,T/3-1$. To provide GDP growth nowcasts concomitantly to the publication of the Banque de France's business survey, we consider three alternative specifications with $m = 1,2,3$. The number of lags p and q is chosen with information criteria. We allow a maximum of $p = 4$ autoregressive terms for the target variable and the lags of the factor can cover the whole year with $q_{\max} = 3$. The equation can also include dummy variables to capture extreme variations of y_τ.

The GDP growth forecast for the quarter $T/3$ conditionally to the information available in month m of quarter $T/3$ is finally given by:

$$\hat{y}_{T/3} = \hat{\beta}_0 + \sum_{i=1}^{p} \hat{\gamma}_i y_{T/3-i} + \sum_{r=1}^{m} \hat{\beta}_{r,0} \hat{F}_{r,T/3} + \sum_{j=1}^{q} \sum_{r=1}^{3} \hat{\beta}_{r,j} \hat{F}_{r,T/3-j}$$

At this level, we use the last observations $T-2,\ldots,T-3+m$ of the factor(s) that were discarded in the third step.

As suggested by Kelly and Pruitt (2015), we use the target variable y_τ as a proxy variable z_τ in the first step of the method to derive the first factor (target-proxy 3PRF). To consider additional factors, we follow their automatic procedure, which involves sequentially adding the residuals of the forecast equation in step 3 as additional proxies in z_τ. However, we will show in the following section that the best results are obtained with one factor in our application.

Compared to the original 3PRF method, there are several differences in the implementation of the mixed-frequency variant. The MF-3PRF requires the high frequency indicator to be converted to low frequency in the first step, the second step is implemented at high frequency and, in the last step we use a U-MIDAS specification instead of converting the high frequency factor at quarterly frequency with arbitrary weighting schemes. Hepenstrick and Marcellino (2019) show an improvement over the 3PRF method for forecasting GDP growth for the US and 6 other countries (in the latter, the variables x_t are converted to low frequency and the three steps are implemented at low frequency). We also find an improvement in the French data for the more recent period considered in this paper. Furthermore, we show that the choice of the conversion method used in the first step of the MF-3PRF approach is not neutral and should depend on the forecast horizon and the availability of the indicators. Specifically, we convert the monthly predictors by average for the first two forecast horizons (in month 1 or 2 of the quarter to be forecast). In the final exercise, the quarterly series are made up of the third months of each quarter. In doing so, we emphasise the information at the end of the quarter, as it contains the clearest signals of the current quarter's economic activity.

In practice, it is common for some predictors x_t used in nowcasting to have missing values at the end of the sample due to publication delays and possibly at the beginning of the sample due to different starting dates. Several strategies to deal with this problem have been discussed in the literature. In this paper, missing values at the beginning of the sample are handled using an iterative expectation-maximization (EM) algorithm, and in addition to the EM algorithm, we consider three alternative methods to deal with the last missing observations: estimating a univariate AR model for each time series with ragged edges and then replacing the missing values with the forecast values, filling the missing values with zero (which corresponds to a naive forecast for normalised time series), or shifting the series to fill the missing values (vertical realignment). Note that with these three strategies, we fill the missing values up to the month of the nowcast, i.e. up to January in the first nowcast of the first-quarter GDP growth, up to February in the second nowcast, and up to March in the last nowcast (an alternative in the tradition of the bridge models is to forecast the monthly predictors up to the end of the nowcast quarter before estimating the factor, but as seen later, this strategy worsens the results).

2.3 Predictors' contributions

Kelly and Pruitt (2015) provide a one-step closed form of the 3PRF forecast. This formulae allows the calculation of the contribution of each predictor x_{it} (or group of predictors, e.g.

financial variables in the following) to the forecast and writes as follows:

$$\begin{aligned}\hat{y} &= \iota\overline{y} + \hat{F}\hat{\beta} \\ &= \iota\overline{y} + J_T X_T \hat{\alpha}\end{aligned}$$

with $J_T = I_T - \frac{1}{T}\iota_T\iota_T'$, I_T the T-dimensional identity matrix, ι_T a T-vector of ones, X_T the matrix of predictors and:

$$\hat{\alpha} = W_{XZ}\left(W_{XZ}' S_{XX} W_{XZ}\right)^{-1} W_{XZ}' S_{XY}$$

where $W_{XZ} = J_N X_T' J_T Z$, $S_{XX} = X_T' J_T X_T$ and $S_{XY} = X_T' J_T y$.

In this paper, we extend the formula to the mixed-frequency case. For simplicity, we present the calculations in the case where we want to include the last three months of the estimated factor in the forecasting equation and we omit the AR terms. In matrix form, the forecasting equation estimated in step 3 (with indices indicating the size of the elements) is as follows:

$$y_{T/3} = c + Q^{(1)}_{T/3,T} F_T \beta_1 + Q^{(2)}_{T/3,T} F_T \beta_2 + Q^{(3)}_{T/3,T} F_T \beta_3 + \eta_{T/3}$$

with F_T the monthly factor (vector of dimension T) and $Q^{(1)}_{T/3,T} F_T$ the vector of the contemporaneous values of the factor in month 1, $Q^{(2)}_{T/3,T} F_T$ in month 2 and $Q^{(3)}_{T/3,T} F_T$ in month 3, three elements of length $T/3$. The matrices of temporal aggregation are defined below:

$$Q^{(1)}_{T/3,T} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix}$$

etc.

$$Q^{(3)}_{T/3,T} = . \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix}$$

We show that the MF-3PRF forecast rewrites as follows:

$$\hat{y} = \iota\overline{y} + J_{T/3}Q^{(1)}_{T/3,T}X_T \underbrace{W_{XZ}\left(S_{XZ}'W_{XZ}\right)^{-1}S_{ZZ}\widehat{\beta}_1}_{\widehat{\alpha_1}} + J_{T/3}Q^{(2)}_{T/3,T}X_T \underbrace{W_{XZ}\left(S_{XZ}'W_{XZ}\right)^{-1}S_{ZZ}\widehat{\beta}_2}_{\widehat{\alpha}_2}$$
$$+ J_{T/3}Q^{(3)}_{T/3,T}X_T \underbrace{W_{XZ}\left(S_{XZ}'W_{XZ}\right)^{-1}S_{ZZ}\widehat{\beta}_3}_{\widehat{\alpha}_3}$$

with $\widehat{\beta}_i$ the i-th element of $\widehat{\beta}$ (see Appendix 1 for the demonstration). The contribution of the variable i to the forecast is then given by:

$$\begin{bmatrix} J_{T/3}Q^{(1)}_{T/3,T}X^i_T W_{XZ} & J_{T/3}Q^{(2)}_{T/3,T}X^i_T W_{XZ} & J_{T/3}Q^{(3)}_{T/3,T}X^i_T W_{XZ} \end{bmatrix} \times \left(S_{XZ}'W_{XZ}\right)^{-1} S_{ZZ}\ \widehat{\beta}$$

with X^i_T the X_T matrix whose columns are replaced by columns of zeros except column i. In the following, we will use this last equation to derive the contribution of the main groups of supply-side and demand-side variables in our dataset to the French GDP nowcast.

3 Nowcasting French GDP growth

3.1 Data

We now use the MF-3PRF procedure to nowcast the French GDP growth. The database consists of French GDP growth and 60 monthly predictors. It was downloaded on February 1, 2024 and covers the period from January 1995 to December 2023. In estimation and forecasting, we follow Schorfheide and Song (2021) and exclude the Covid-19 observations.

The dataset consists of survey variables (the manufacturing and services surveys conducted by the Banque de France), real indicators (industrial production index, household consumption of goods, construction variables, etc.), financial and monetary variables (monetary aggregates, interest rates, various stock and price indices) and, finally, international variables (key macroeconomic indicators for Germany and the euro area).[1] In addition to these traditional indicators in nowcasting, we use the media-based indicator of economic policy uncertainty for France (Baker et al. (2016)) and electricity consumption, whose usefulness in tracking real activity in uncertain times is documented by Barbaglia et al. (2023).[2] All variables are seasonally adjusted,

[1] The variables are listed in Appendix 2, together with their publication lag and their transformation. The publication lags are calculated with reference to the release of the business surveys of the Banque de France.

[2] Barbaglia et al. (2023) also consider text-based sentiment indicators, Google Trends, Airbnb review figures, air cargo and air quality statistics, mobility indicators based on mobile phone data, and aviation figures. We

except the price indices and the index of uncertainty.[3] The monthly variables are published with different delays: 25 are available right after the end of the month, while 30 are published with a delay of one month and 5 with a delay of two months. The target variable consists of the first releases of French quarterly real GDP growth, which are now communicated by Insee 30 days after the end of the quarter in question. As with the current MIBA model, we aim to forecast the first GDP figure provided by Insee, which is closely followed by both the media and policymakers.

Several treatments are applied to the data before estimating the model. For the Banque de France survey variables on manufacturing industry and services, we use moving averages, the order of which depends on the available observations on the nowcast quarter at the time of the forecast. We use a 2-month (3-month) moving average in the second (third) month of the quarter in the forecast, while in the first month, we do not average the data and use the raw survey variables. In this way, we make use of the most relevant information on the nowcast quarter in step 3. For housing starts and building permits, which are highly volatile, we use a 3-month moving average. If non-stationary, the series are transformed by taking their growth rate or first difference.[4] Finally, the series are normalised with the mean and standard deviation of the raw variables on the sample excluding the covid period (January 2020 to August 2021).

Nine series have missing observations at the beginning of the sample (e.g., the survey variables in services start in October 2002). These missing values are treated using an iterative expectation-maximization (EM) algorithm.[5] The number of factors is chosen using an information criterion (Bai and Ng (2002)) with a maximum of three factors. The sample only covers the period 1995-2019, to avoid a possible effect of the covid period. As described above, more than half of our series (35) are also not known by the month in which the forecast is released, due to the delay in their publication. For example, the survey variables are known at the end of the month they deal with, while the industrial production index has one missing observation

do not consider these alternative indicators because they have been shown to be more useful for nowcasting pandemic observations (not considered here) and are available for a much shorter time period, e.g., since 2004 for Google Trends data and 2015 for Airbnb reviews.

[3] Electricity consumption published by RTE is not seasonally adjusted. We use the X-13 ARIMA seasonal adjustment procedure in the Eviews software.

[4] Unit root tests are applied from 1995 to 2019, to avoid the detrimental effect of the large break in the series during the pandemic crisis on the performance of the tests. See Appendix 2 for the transformation of the data.

[5] We use the Matlab code provided by McCracken and Ng (2017) at `https://github.com/geoluna/FactorModels`. To initialise the EM algorithm, we fill in the missing values in our set of monthly predictors with their unconditional mean and we run principal components on the updated dataset to obtain an estimate of the factors and the loadings. Then, the EM algorithm consists of the following iterative steps: we update the missing values with the predicted values by the factors, construct a new set of factors from the updated dataset, and repeat these calculations until the predicted values of the new dataset do not change.

and the index of services production is not available for the last two months. In the reference specification, we replace these missing observations with AR forecasts. Again, the autoregressive process is estimated without the observations of the pandemic period and the order of the process for each predictor is selected using the Bayesian Information Criterion (BIC). Other strategies are discussed in the robustness analysis.

Using this dataset, we develop several forecasting equations. At the Banque de France, the nowcasts are typically released together with the business surveys on the 6th working day following the end of the month in question. For instance, nowcasts for the first quarter are published at the beginning of February, March and April. Our primary competitor being the MIBA model, we focus on the forecasts generated by our models, which are based on the information available at the time of the survey release. Therefore, in the following, we develop three nowcasting models, which are designed to be used with the same timing as the MIBA model, one at the end of each month of the quarter (hereafter referred to as the M1, M2 and M3 equations). For the nowcast of the first quarter, M1 refers to early February, M2 to early March, and M3 to early April.

3.2 Estimation results

We present the estimation results of the models for the entire sample period from 1995 to 2023, excluding the extreme observations due to the pandemic.

We consider three forecasting equations to nowcast real GDP growth a few days after the end of each month of the quarter (referred to here as M1, M2 and M3), which is the timing of the Banque de France's forecast using the MIBA model.[6] For the three horizons, we consider the simple case of a single factor. As will be shown later, the inclusion of more factors is detrimental to the forecasting performance of the models. In the first step of the method, the monthly indicators are transformed by averaging the months to develop the forecasting models in M1 and M2, and by taking the third month of each quarter for the model to be used in M3. In this way, the end-of-quarter information is emphasized in the final forecast. Note, however, that in the particular case of the survey variables in our dataset, the information on the whole quarter is taken into account, given the preliminary transformation of the variables (3-month moving averages) in M3.

[6] Note that in practice, the factor can be re-estimated each time an indicator is released in the database. Given the variety of time releases of the predictors, it is possible to update the scenario with a much higher frequency.

Figure 1 displays the factor estimated from January 1995 to December 2023, along with the wordclouds of the variables based on their weights in the factor (the estimated first-pass coefficients in absolute terms). The results are shown for each forecast horizon.[7] Although the transformations of the survey data differ according to the forecast horizon, as well as the aggregation method in step 1 of the filter between the first two months and the third one, the factors estimated for the three months are close (correlation of more than 90% between the three factors). They show a deep trough during the 2008-09 recession and less severe downturns in the 2002-03 and 2012-13 slowdown episodes in France, reflecting well the fluctuations in French GDP.[8]

Looking at the weights of the variables in the wordclouds, it is noticeable that the Banque de France survey variables play an important role at the three horizons, particularly in industry (including key variables also selected in the MIBA model, such as the change in deliveries EVLIV and the expected change in production PREVPRO, in addition to the change in orders EVCOM and the change in production EVPRO) and in services (the expected change in activity and the change in aggregate demand, PREVACT and DETEM). Although their weight is much lower, some variables in construction (housing starts and building permits) also contribute to the factor according to this graph. Thus, unlike the current MIBA model, which is based only on manufacturing survey data, the forecast takes into account information from other sectors. This may be an advantage over the MIBA model in case of a decoupling between the dynamics of manufacturing and the rest of the economy, a concern raised by Thubin et al. (2016). Among other hard data indicators, the IPI for basic metals and fabricated metal products and unemployment variables have a large weight. To a lesser extent, financial and international variables such as interest rates and industrial production in the euro area contribute to the calculation of the factor.

This picture is stable over time. This is shown in Figure 2, which plots the coefficients estimated in step 1 over recursive sample periods. For the economic interpretation of the factor, Figure 2 is a useful complement to Figure 1, which displays the absolute weights of the predictors. As shown in Figure 1, the variables that play a predominant role in the construction of the factor are mainly the survey balances for industry and services whatever the period. The coefficients are positive for activity indicators (survey variables on activity, demand, new

[7] The weights in the wordclouds differ slightly in months 1 and 2, due to the different transformation of the survey variables, and in month 3, where we use a different aggregation rule.

[8] See Aviat et al. (2023) for a comprehensive analysis of the French business cycle.

orders, as well as variables related to consumption, activity in Germany and in the euro area). Conversely, they are negative for counter-cyclical variables such as unemployment, inventories, and the economic policy uncertainty index. In both figures it can be seen that financial variables have a relatively low weight. Therefore, the estimated factor mainly reflects information on economic activity, in particular from survey variables. This conclusion is also supported by the strong correlation between our factor and the business climate indicator for France provided by Insee, as shown in Figure 4.

In Table 1, we present the estimation results of the forecasting equations in months 1, 2 and 3. The sample period ranges from 1995Q1 to 2023Q4, and we have dropped the extreme observations from the pandemic (2020Q1 to 2021Q4).[9] At each forecast horizon, the equation can include the contemporaneous factor(s) available in that month (for example, in month 2, the factor composed of months 2 and 1, $F_{2,t}$ and $F_{1,t}$) and the lagged factors at the three months ($F_{r,t-j}$, $r = 1, 2, 3$ and $j = 1, 2, 3$). We also allow up to 4 autoregressive terms. The choice of lags (factor and autoregressive terms) in each equation is based on the BIC. With this criterion, we retain a very parsimonious specification that contains only the last available contemporaneous factor (i.e., the factor consisting of months 1 in equation M1, months 2 in equation M2, and months 3 in equation M3) and one autoregressive term. In addition, each equation includes two dummy variables: the first for the first quarter of 1996 to account for the recovery in activity following the strikes in France at the end of 1995, and the second for the first quarter of 2009 to capture the trough of the 2008-09 recession.

The explanatory variables in the three equations are statistically significant at the 5% or 10% level. The goodness of fit of the models is similar to that of the MIBA model estimated over the same period in month 1 (adjusted R-squared of 59%) and better in months 2 and 3 with adjusted R-squared of 70% and 73% (see Appendix 3 for the estimation results of the MIBA equations). Furthermore, the explanatory power of the models increases from M1 to M3 as more information about the quarter's activity becomes available. The three equations pass the usual diagnostic tests. The Breusch-Godfrey and Breusch-Pagan-Godfrey tests show no autocorrelation and no heteroskedasticity in the residuals, and the residuals are Gaussian according to the Jarque Bera test. Figure 3 also shows the stability of the forecasting equations. In the top panel, the plots of the cumulated recursive residuals in the CUSUM test show no structural change in the coefficients of the three models over the sample period. The plots in

[9]Due to the AR(1) term, we also remove the last quarter of 2021 in the estimation of the models.

the bottom panel also show the stability of the specification (in terms of the number of lags of the dependent variable and the factor) in recursive estimations.

In Figure 5, we utilize the formulas for the contributions that we derived in the mixed-frequency case to divide the GDP estimate into various blocks (see Appendix 2 for the exact composition of each group). On the supply side, we display the contribution of activity in industry, services, and construction. We also compute the contribution of consumption, sales, and labor data, which together represent the demand side of the economy. The last two blocks reflect the financial and international environment. The financial block consists of the contribution of monetary and financial variables to GDP growth. The international contribution is calculated using activity indicators for Germany and the euro area, showing how international factors affect the domestic economy.

Since the recession of 2008-09 and the euro-zone debt crisis of 2010-12, the main blocks contribute negatively to the estimated GDP growth. The picture improves between 2017 and 2020, especially with positive signals on the demand side. Nevertheless, since the start of the COVID-19 pandemic, all six blocks make negative contributions to estimated GDP growth, including demand-side indicators. Most of the variables are below their long-run average in the most recent period, which results in negative contributions to GDP growth from the six blocks. This could reflect the widespread impact of the pandemic and the high level of uncertainty since then on different facets of the economy.

3.3 Out-of-sample evaluation

We finally turn to the forecast evaluation of the models. As mentioned earlier, we consider three nowcast horizons defined according to the publication schedule of the Banque de France business surveys, from the first month to the last month of the target quarter.

The models are estimated using available observations from 1995. Their forecasting performance is evaluated from 2010Q1 to 2023Q4. To replicate the conditions of the nowcasting exercise, the MF-3PRF models are first estimated from 1995Q1 to 2009Q4 and the first quarter of 2010 is forecast based on the information available in the first days of February, March, and then April 2010. Similarly, we generate three forecasts for the GDP growth rate in the second quarter of 2010 using data from May 2010 to July 2010. These calculations are repeated for each subsequent quarter within the out-of-sample period. Therefore, we obtain three sets of forecasts for the quarters 2010Q1 to 2023Q4, based on the information available at the end of each month during the respective nowcast quarter. The lag length for the factor and the number of AR terms is selected in each recursion using the BIC criterion. We also fill in the missing values of the 35 series at the end of the sample (up to the month of the forecast) in each recursion. We do this using only the information available at the time of the forecast.

Our reference model is a single-factor model (the three steps of the filter are iterated once). Missing values at the beginning of the sample are treated with the EM algorithm, and the ones at the end of the sample are forecast with AR processes estimated for each of the 35 predictors. For example, in the first forecast of first-quarter GDP growth, the industrial production index is available until December and needs to be forecast in January. Again, the AR order is selected using the BIC criterion for each predictor.[10] As mentioned earlier, in the first step of the filter, the monthly predictors are transformed to quarterly frequency by average in M1 and M2 and by taking the third month in M3. The relevance of these choices is examined below.

We compare the performance of this factor model with several benchmarks. First, we consider an autoregressive process with a constant and the two dummy variables for 1996 and 2009. In the case of the AR process, the lag length is selected at each recursion of the out-of-sample exercise using the BIC criterion.[11] Second, we compare the forecasts with the reference tool in Banque de France, the MIBA model. For each month, we use the corresponding MIBA equation. We also consider the performance of the combination of the MIBA and MF-3PRF models, using a simple average of the two forecasts.

We do not use real-time dataset (half of the 60 series are revised) which is a caveat of the analysis. According to Bernanke and Boivin (2003) and Schumacher and Breitung (2008) however, the conclusions about forecasting performance remain largely unchanged when final data is used instead of vintage data. It is also important to note that prior to January 2016, the release of the GDP growth rate occurred 45 days after the end of the corresponding quarter (as compared to 30 days in the current calendar). This posed a challenge to the nowcasting process in the first month, as the previous GDP growth rate was unknown. This complication was particularly challenging when the model incorporated autoregressive terms, as is the case here. In the out-of-sample evaluation, we assume that the new calendar also applies to the period between 2010 and 2016. This allows us to assess the performance of the model within

[10] In the final recursions, the AR processes are estimated without the pandemic period.

[11] The use of the AIC criterion does not change the forecasting performance of the AR equation.

the current release schedule, which is the question of interest here.

The pseudo-real-time forecasts derived in this design with the MF-3PRF and MIBA models in M1, M2, and M3 are plotted against the first release of the GDP growth rate in Figure 6. The forecast accuracy of the factor model and the competitors is then measured by the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE). The criteria are computed for the entire out-of-sample window, excluding the extreme observations due to the pandemic (2020Q1-2021Q4), and for two subperiods, before the pandemic (2010Q1-2019Q4) and after the pandemic (2022Q1-2023Q4).[12] These periods consist of 48, 40 and 8 quarters respectively. Looking at the last window gives an idea of the model's performance in the most recent period, taking into account the potential impact of the pandemic. However, given the very limited number of observations, the results for this last subperiod should be treated with caution. We run the modified version of the Diebold-Mariano test proposed by Coroneo and Iacone (2020) to assess whether the forecasting performance of the models is better than that of the MIBA benchmark. This variant overcomes the small-sample size distortions of standard tests. We use the test statistic with a weighted periodogram estimate and a Daniell kernel.[13] The null hypothesis is that the forecasting accuracy of the models is equal and the alternative is that the competing model performs better. The results for the reference factor model and several variants are shown in Table 2.

In Table 2a (first block), we report the RMSE and MAE of the two benchmarks (AR and MIBA), the factor model (MF3PRF) and its combination with the MIBA model (COMB) in the reference configuration (one factor, AR forecasts to fill the missing values, conversion of the monthly predictors by average in M1-M2 or by taking the third month in M3). The best results at each horizon are in bold and marked with asterisks when the criteria are significantly lower than those of the MIBA model according to the Coroneo and Iacone (2020) test.

Overall, the quality of the forecasts naturally improves throughout the quarter with the gradual arrival of information. The performance of the models declines slightly in the most recent period due to the unexpectedly strong GDP growth in the second quarter of 2023.

[12] The dummy variables, which would contain information not available at the time of the forecast, are not active in these periods.

[13] The authors consider an alternative estimate of the long-run variance of the loss differential d in the computation of the Diebold-Mariano statistic. They use a weighted periodogram estimate with the Daniell kernel $\hat{\gamma}^2_{DAN} = 2\pi \frac{1}{m} \sum_{j=1}^{m} I(\lambda_j)$ with $I(\lambda_j) = \left| \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} d_t e^{-i\lambda_j t} \right|^2$. The modified DM statistic $\sqrt{T}\frac{\bar{d}}{\hat{\gamma}_{DAN}}$ has a t-distribution with degrees of freedom $2m$. We adjust the code of the authors to run a unilateral test. The bandwidth parameter of the weighted periodogram estimate is set at its default value $\left\lfloor T^{1/3} \right\rfloor$.

The comparison of the models then shows that MF-3PRF is a good competitor to the MIBA reference model. MF-3PRF outperforms MIBA in the first two months M1 and M2 and the gain is particularly strong in M1 and still important in M2 (e.g. gain of 16% in M1 and 5% in M2 in terms of RMSE over the full evaluation window). However, the MIBA model is slightly superior in the last forecasting exercise M3. In this last exercise, MF-3PRF is still useful since the combination of MIBA and MF-3PRF slightly outperforms MIBA in M3 (gain of 2% in RMSE). Comparing MF-3PRF and COMB with MIBA using the Coroneo and Iacone (2020) test, the gain with the combination is significant at the 10% or 5% level in M1 and M2 over the long window and before the pandemic, and in M1 over the recent subperiod. The gain is also significant at 10% in M2 with MF-3PRF before the pandemic and in M1 afterwards. MF-3PRF and COMB do not perform significantly better in M3. In summary, the best performing tool is generally the MF-3PRF in M1 and M2 and the combination of MF-3PRF and MIBA in M3.

In the remainder of Table 2, we consider several variants for MF-3PRF and its combination with MIBA and we report the RMSE (or MAE) in relative terms with respect to the reference case discussed above. A ratio less than one indicates a gain relative to the reference case. In the first variants (Table 2a), we examine the impact of the treatment of the missing values of the monthly predictors up to the month of the forecast. In the reference specification, we fill the missing values with the forecast of an autoregressive process, the order of which is selected according to the BIC criterion for each predictor. Alternatively, we consider a naive forecast consisting of replacing the missing values with zeros (since the predictors are normalised), the use of the EM algorithm and the realignment strategy. As with the treatment of missing values in nine series at the beginning of the sample, the number of factors in the EM algorithm is chosen using an information criterion from Bai and Ng (2002) with a maximum number of three. Overall, the AR forecast outperforms the other treatments. The realignment strategy seems to be particularly disadvantageous in our application (e.g., loss of 17%-18% in months 1 and 2 over the entire evaluation period). The EM algorithm gives a slight improvement in M1 in terms of MAE, but worsens the results at the other horizons.

Table 2b examines several other variants. First, we explore alternative methods to convert the monthly factor to the quarterly frequency in the first step of the filter. Using the appropriate aggregation scheme is crucial to take into account the availability of the information at each forecast horizon, as well as the quality of the signals provided by the predictors in each month of the quarter. In the reference specification, we use an average of the three months in M1 and M2. In M3, the quarterly predictors consist of the last month of each quarter. We consider the symmetric configuration, which involves using time series consisting of the first months of each quarter in M1, the second months in M2, and the average of the three months in M3. The models exhibit a significant decrease in performance, with a 7% increase in mean squared error (MSE) for M3 over the full window. The combination also shows a decrease in performance, albeit to a lesser extent.

In the second variant, we investigate whether the results can be improved by pre-selecting the predictors before estimating the factor, as recommended by Bai and Ng (2008) in ordinary DFM. Following Bai and Ng (2008), we use the LARS-EN algorithm with different parameter sets: the number of predictors to select $N_A = \{10, 30, 50\}$ and as a penalty parameter for the l_2 norm of the coefficients vector $\lambda_2 = \{0.1, 0.25, 0.5, 0.75, 1\}$. For parsimony, we only report the case $N_A = 30$ and $\lambda_2 = 0.5$. Overall, we find no gain when the factors are estimated from a reduced dataset. In M2, the loss is 13% in terms of MSE and MAE over the global window. Thus, the MF-3PRF efficiently weights indicators relevant to the forecast of the target variable, making an additional variable preselection procedure unnecessary.[14] It is important to note, however, that the dataset used in this study was limited to well-established predictors of GDP growth, and the elimination of non-significant predictors might have been beneficial in a larger dataset.

Third, we examine the gain from using mixed-frequency data in the estimation method of Hepenstrick and Marcellino (2019) compared to the original approach in Kelly and Pruitt (2015). In the latter, the three steps are performed at a quarterly frequency. In our application, both approaches give similar results in month 1, but there is a clear gain in the following two months (7% in month 2 and 12% in month 3 in terms of MSE for 2010-2023). This confirms the usefulness of considering high-frequency data for GDP nowcasting, as suggested in the MIDAS literature. The next block of results in the table shows that a two-factor model significantly worsens the forecasts in the three months. The loss is particularly large for M1 (11% in terms of MSE over the whole window).

In the next variant, referred to as the *blocking approach* in Table 2b, we assess whether the

[14] As in Hepenstrick and Marcellino (2019), we also consider removing variables that do not have a significant relationship with GDP growth at different levels of significance in the first step of the filter. The results are similar and are available on request.

use of three factors constructed from the correlation between GDP growth and the variables in each month improves the results. In this case, we regress the observations of the variables in month 3 (2 and 1 respectively) on GDP growth in the first step of the filter to derive the factor in month 3 (2 and 1 respectively) in step 2. This approach allows us to give more weight to information in month 1 with a forward-looking content (e.g. firms' production expectations in the Banque de France surveys) and to coincident variables in month 3 (e.g. firms' opinion of their past activity). The three factors obtained and their lags (if selected by the BIC criterion) are then included in the forecasting equation in step 3. This method worsens the results with a loss of up to 7%.

In the final block of results, referred to as the *bridge approach* in Table 2b, we forecast the monthly predictors to the end of the nowcast quarter before estimating the factor. This allows the forecast equation to include the three contemporaneous factors at the three horizons. In the reference approach, for example, the forecast equation in the first month cannot include the non-lagged factor consisting of months 2 and 3. This approach differs from the previous one only in months 1 and 2, but is obviously equivalent to the one we use in M3. Each predictor is forecast with an AR model (the other approaches do not improve the results). The lags of the factor and the number of autoregressive terms in the forecasting equation are again chosen using the BIC criterion. The optimal aggregation method in the first stage regressions is different in this case (we use time series consisting of the specific month of the forecast). As shown in the table, there is generally no gain with this approach in months 1 and 2. All ratios are greater than or equal to one. In month 3, the two approaches are equivalent (ratios equal to one).

Overall, the reference specification outperforms the other variants. In our application to French data, it is preferable to estimate a single factor across the entire dataset, where missing values up to the forecast month are filled with AR forecasts. There is also a gain in using the mixed-frequency version of the filter, with an aggregation scheme in step 1 chosen according to the forecast horizon.

4 Conclusion

In this paper, we have developed a novel model to nowcast the first release of real GDP growth in France. The model is estimated using the three-pass regression filter proposed by Kelly and Pruitt (2015) and extended to the mixed-frequency case by Hepenstrick and Marcellino

(2019). Rather than simply summarizing the information from a large set of monthly indicators, this approach allows us to obtain targeted factors for forecasting the variable of interest, can accommodate ragged edge datasets, and provides a comprehensive and easily interpretable framework.

We estimate this model on a large set of monthly indicators. Among the types of variables, the Banque de France survey variables on manufacturing and services are particularly useful. We extend the formulae for the contributions of the predictors in the mixed-frequency case and show that all groups of supply and demand variables have contributed negatively to GDP growth since the onset of COVID-19 pandemic.

A pseudo-real-time evaluation of the method shows the good performance of the model compared with several simple benchmarks and the existing tools at the Banque de France, especially during the critical first two months of each quarter. The forecasting combination of MIBA and the new model also performs well in the shortest horizon. In the robustness part, we show that this model outperforms a large set of variants.

As part of future research, we could include a volatility parameter in our model, as was done in Lenza and Primiceri (2022) in the context of VAR estimation. This volatility parameter could allow us to take into account the Covid period more accurately.

Overall, this research contributes to the field of nowcasting by proposing an effective model for forecasting French GDP growth, providing guidance to policymakers and economic analysts seeking to make informed decisions in a timely manner.

References

Aviat, A., F. Bec, C. Diebolt, C. Doz, D. Ferrand, L. Ferrara, E. Heyer, V. Mignon, and P.-A. Pionnier (2023). Les cycles économiques de la France: une datation de référence. *Revue économique 74*(2), 5–52.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

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

Baker, S., N. Bloom, and S. Davis (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics 131*(4), 1593–1636.

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

Chatelais, N., A. Stalla-Bourdillon, and M. Chinn (2023). Forecasting real activity using cross-sectoral stock market information. *Journal of International Money and Finance 131*, 102800.

Chow, H., Y. Fei, and D. Han (2023). Forecasting GDP with many predictors in a small open economy: forecast or information pooling? *Empirical Economics*, 1–25.

Coroneo, L. and F. Iacone (2020). Comparing predictive accuracy in small samples using fixed-smoothing asymptotics. *Journal of Applied Econometrics 35*(4), 391–409.

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

Guérin, P., D. Leiva-Leon, and M. Marcellino (2020). Markov-switching three-pass regression filter. *Journal of Business & Economic Statistics 38*(2), 285–302.

Hepenstrick, C. and M. Marcellino (2019). Forecasting gross domestic product growth with large unbalanced data sets: the mixed frequency three-pass regression filter. *Journal of the Royal Statistical Society Series A: Statistics in Society 182*(1), 69–99.

Kelly, B. and S. Pruitt (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics 186*(2), 294–316.

Lenza, M. and G. Primiceri (2022). How to estimate a vector autoregression after march 2020. *Journal of Applied Econometrics 37*(4), 688–699.

Liang, C., L. Wang, and D. Duong (2024). More attention and better volatility forecast accuracy: How does war attention affect stock volatility predictability? *Journal of Economic Behavior & Organization 218*, 1–19.

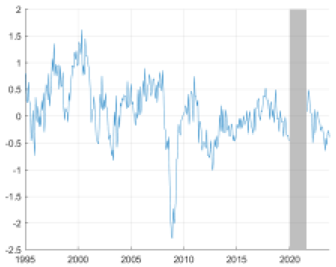
Marcellino, M. and V. Sivec (2021). Nowcasting GDP growth in a small open economy. *National Institute Economic Review 256*, 127–161.

Mogliani, O. M. Darné, and B. Pluyaud (2017). The new MIBA model: Real-time nowcasting of French GDP using the Banque de France's monthly business survey. *Economic Modelling 64*, 26–39.

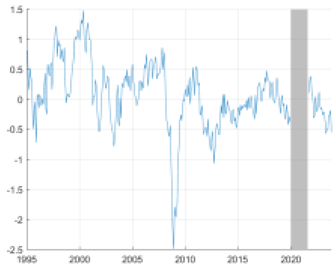
Schorfheide, F. and D. Song (2021). Real-time forecasting with a (standard) mixed-frequency var during a pandemic. Technical report, National Bureau of Economic Research.

Thubin, C., E. Monnet, M. Marx, V. Oung, and T. Ferriere (2016). The PRISME model: can disaggregation on the production side help to forecast GDP? Technical report.

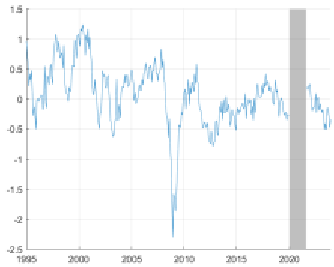
Figure 1: Monthly factor and predictors' weights



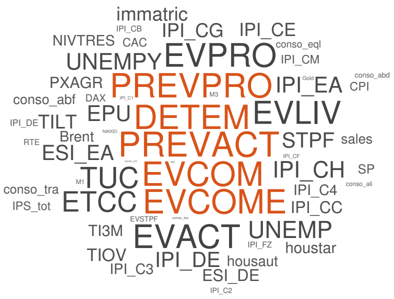
(a) month 1



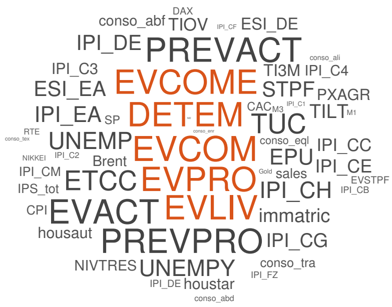
(b) month 2



(c) month 3



(a) month 1



(b) month 2

(c) month 3

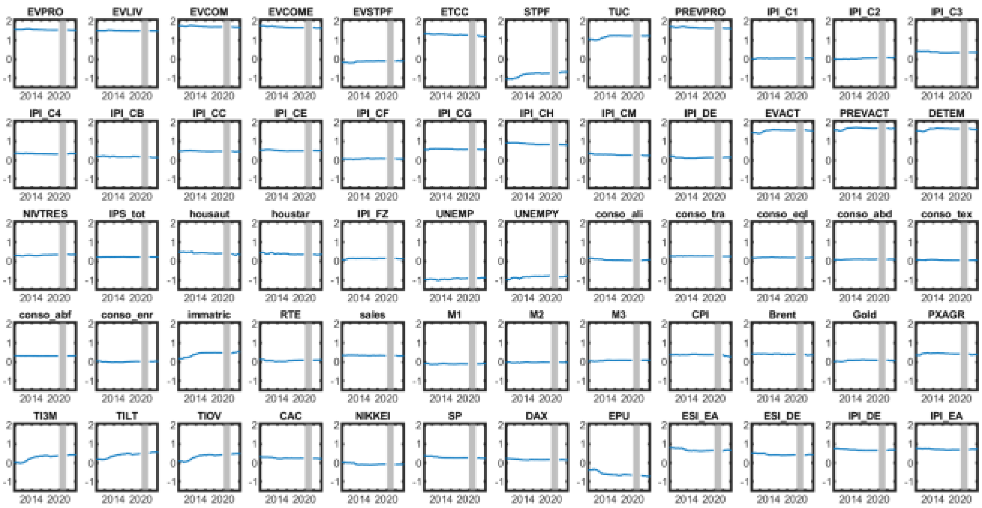
Notes: The graphs in the top panel show the monthly factor derived in the second step of the filter from January 1995 to December 2023. The shaded area corresponds to the covid period. The bottom figures represent the variables according to their weight in the factor (the estimated first-pass coefficient in absolute value). The variable codes are given in Appendix 2. For example, EVPRO refers to the change in output compared to the previous month in the Banque de France survey for industry. The results are reported for months 1, 2 and 3.

Table 1: Estimation results of the MF-3PRF equations (1995Q1-2023Q4)

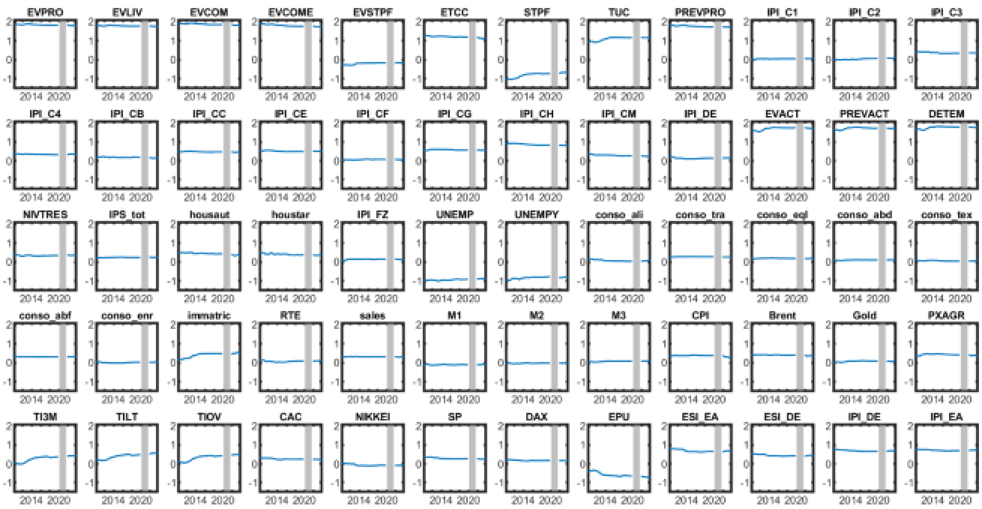
	M1		M2		M3	
Variable	coefficient	t-statistic	coefficient	t-statistic	coefficient	t-statistic
Intercept	0.44	11.99	0.43	13.88	0.47	15.57
Dum96Q1	0.44	1.67	0.60	2.73	0.63	3.03
Dum09Q1	-1.14	-4.15	-0.91	-3.83	-0.91	-4.04
AR(1)	-0.43	-5.16	-0.43	-6.28	-0.45	-6.87
F3,t	-	-	-	-	0.69	14.71
F2,t	-	-	0.63	13.59	-	-
F1,t	0.58	10.34	-	-	-	-
Adj-R2	0.59		0.70		0.73	
Sig_e	0.25		0.21		0.20	
BIC	-2.59		-2.91		-3.01	
AR(4)	0.77		0.48		0.70	
Het	0.89		0.84		0.74	
JB	0.50		0.45		0.14	

Notes: This table reports the estimation results of the forecasting equations in months 1, 2 and 3 for the period 1995Q1 to 2023Q4 (excluding the observations from 2020Q1 to 2021Q4). The table reports the parameter estimates and the corresponding t-statistics. The p-values of the residual tests are reported at the bottom of the table. AR(4) denotes the Breusch-Godfrey test for residual serial correlation up to order four, Het denotes the Breusch-Pagan-Godfrey test for heteroskedasticity, and JB denotes the Jarque Bera test for normality.

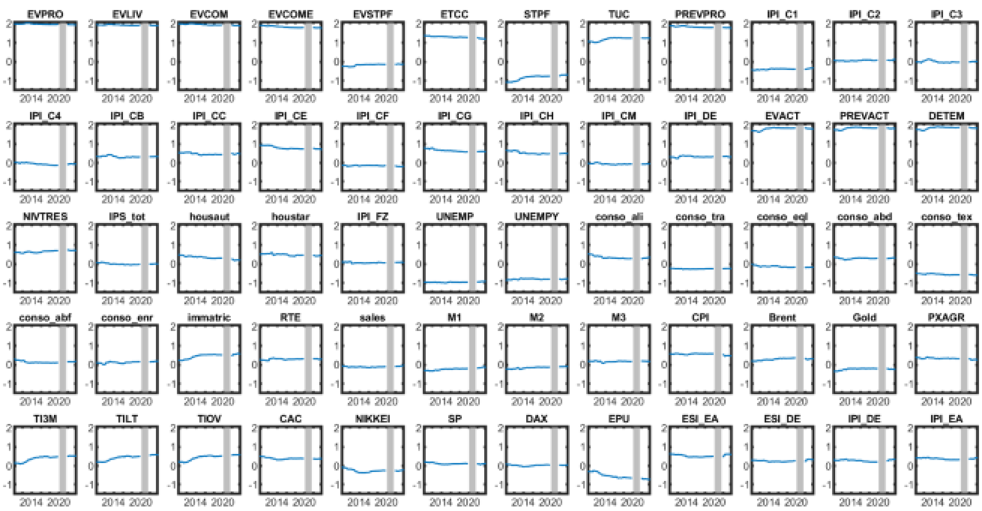
Figure 2: Stability analysis - coefficients in step 1



(a) month 1



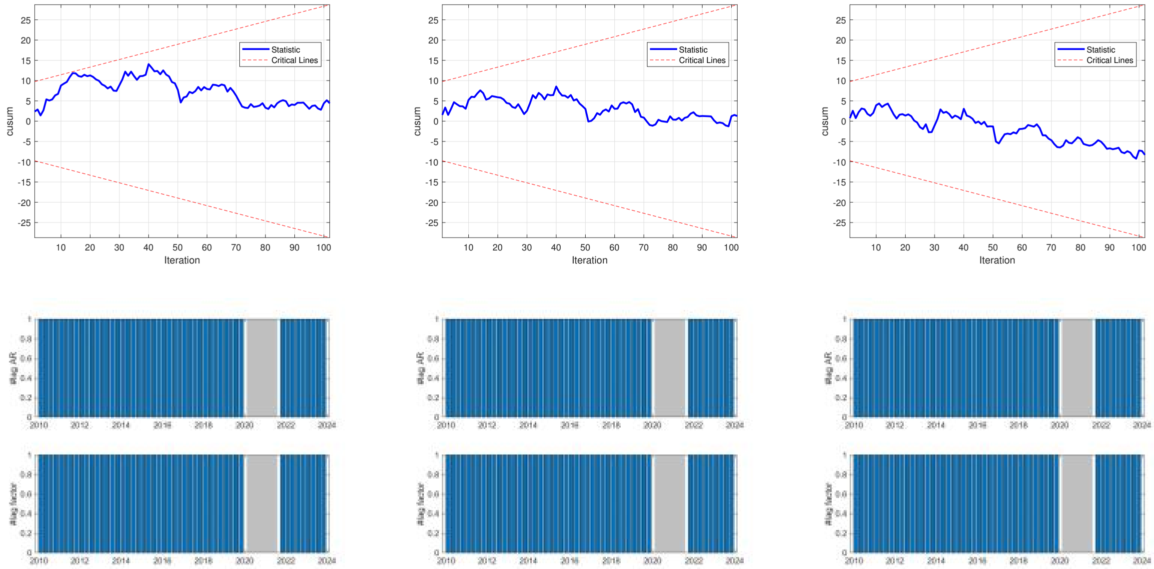
(b) month 2



(c) month 3

Notes: This figure plots the coefficients in step 1 estimated over recursive sample periods, from 1995Q1 to 2010Q1, 1995Q1 to 2010Q2, ..., 1995Q1 to 2023Q4. Results are shown for months 1, 2 and 3. The shaded area corresponds to the covid period.

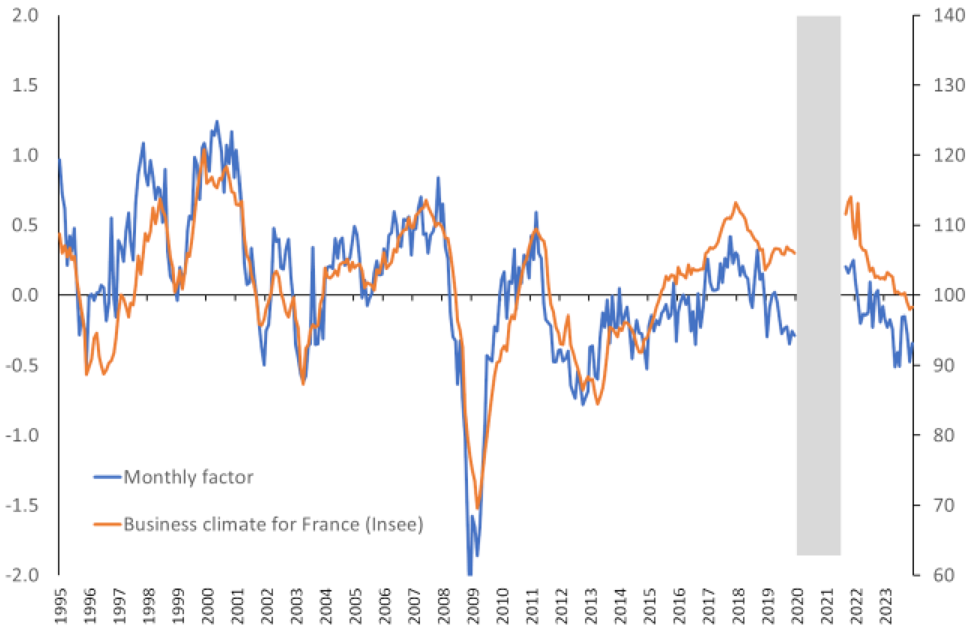
Figure 3: Stability analysis - forecasting equation



(a) month 1 (b) month 2 (c) month 3

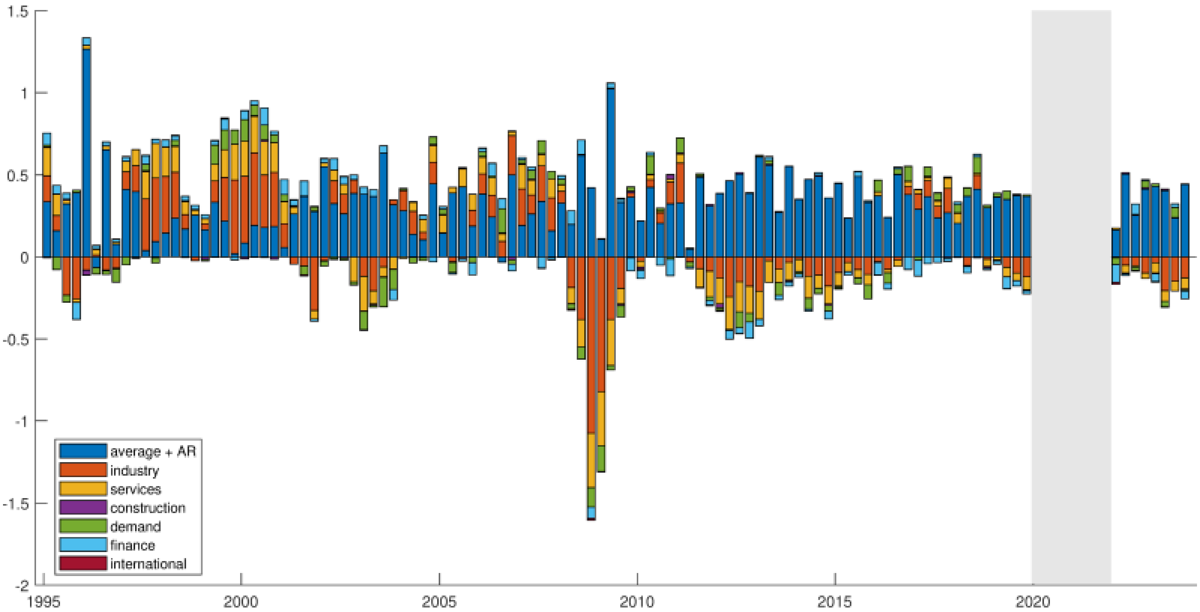
Notes: In the top panel, the figures show the results of the CUSUM test for the forecasting equation in months 1, 2 and 3 over the sample period (without the covid period). In the bottom panel, the figures depict the number of autoregressive lags and factors in months 1, 2 and 3 in recursive estimations from 1995Q1 to 2010Q1, 1995Q1 to 2010Q2,..., 1995Q1 to 2023Q4. The shaded area corresponds to the covid period.

Figure 4: Monthly factor and business climate for France



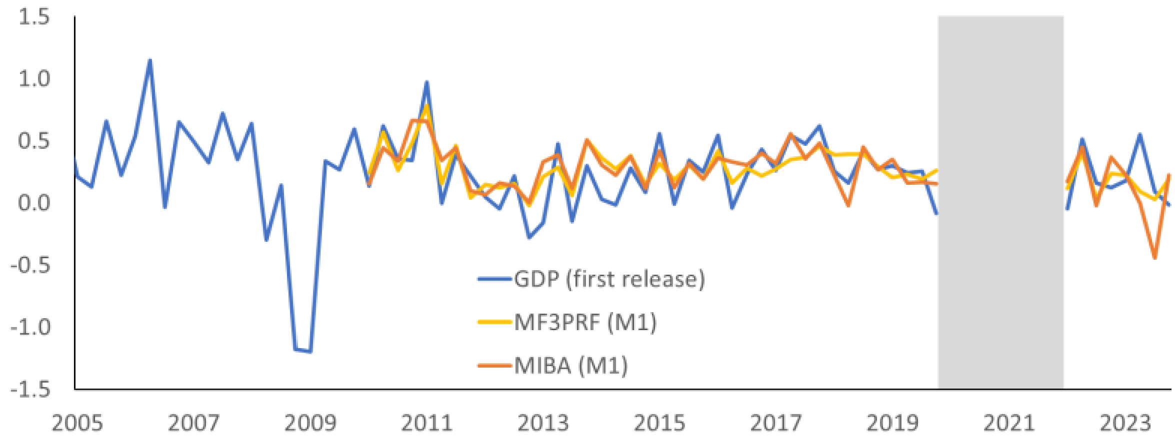
Notes: This figure shows the monthly factor (estimated in month 3) and the business climate for France provided by Insee. The shaded area corresponds to the covid period.

Figure 5: Contributions of the predictors to the estimated GDP

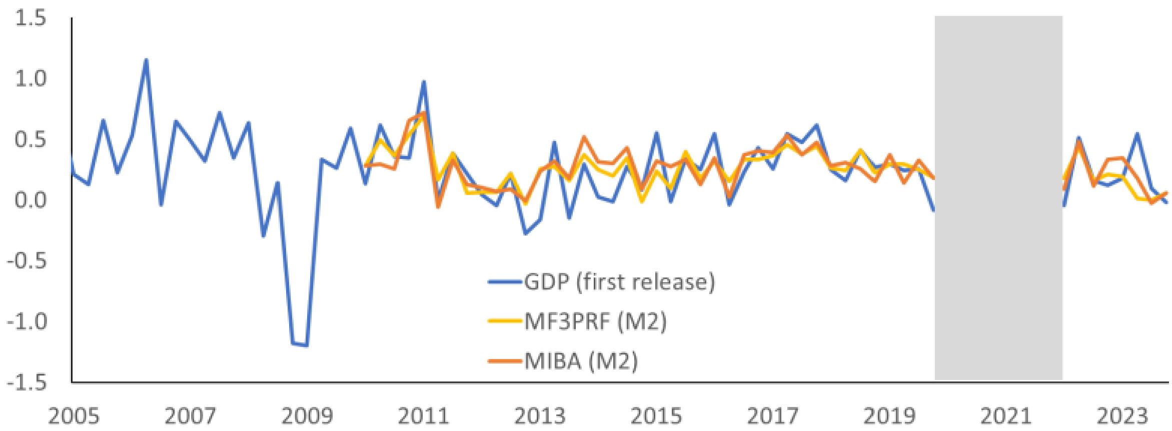


Notes: This figure shows the contributions of the variables to the estimated GDP from 1995Q1 to 2023Q4. The set of variables is divided into six blocks: industry, services, construction, demand, financial and international variables. The shaded area corresponds to the covid period.

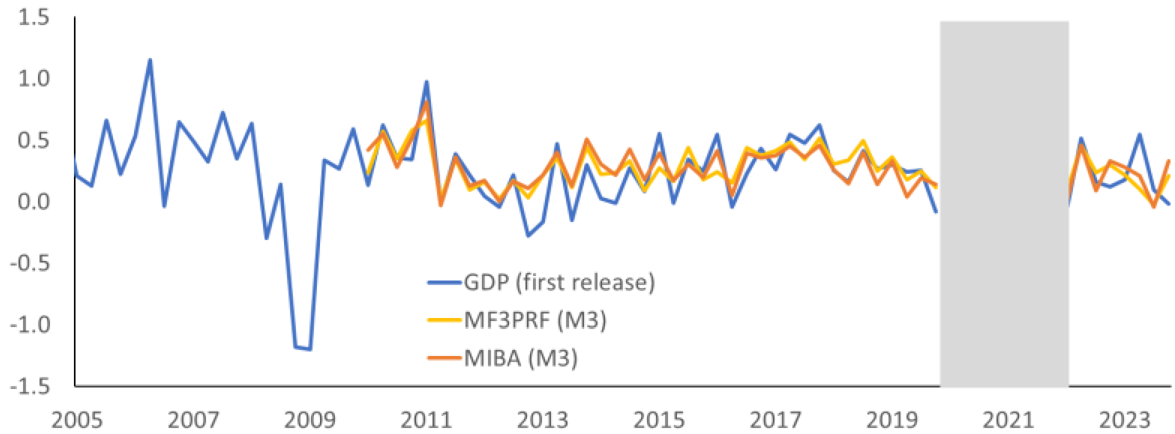
Figure 6: Pseudo real-time forecast of GDP growth



(a) month 1



(b) month 2



(c) month 3

Notes: This figure displays the MF-3PRF and the MIBA forecasts generated in the out-of-sample design for the period 2010Q1 to 2023Q4. The results are shown for M1, M2 and M3. The blue line represents the first release of GDP growth, while the shaded area corresponds to the pandemic period.

Table 2: Out-of-sample evaluation of the models - reference case and variants

(a) Treatment of the missing values at the end of the sample

Reference specification																		
	2010Q1-2023Q4						2010Q1-2019Q4						2022Q1-2023Q4					
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
	RMSE			MAE			RMSE			MAE			RMSE			MAE		
AR(p)	0.308	0.308	0.308	0.234	0.234	0.234	0.260	0.260	0.260	0.214	0.214	0.214	0.480	0.480	0.480	0.335	0.335	0.335
MIBA	0.213	0.184	0.169	0.163	0.153	0.137	0.186	0.186	0.163	0.143	0.155	0.132	0.316	**0.177**	0.200	0.261	0.145	0.164
MF3PRF	**0.179**	0.174	0.172	**0.149**	**0.131***	0.138	0.174	**0.165***	0.165	0.147	**0.130***	0.133	**0.203***	0.215	0.202	**0.161***	**0.140**	0.159
COMB	0.186*	**0.173***	**0.166**	0.150*	0.140**	**0.134**	**0.170***	0.169**	**0.159**	**0.138**	0.139**	**0.131**	0.250*	0.191	**0.196**	0.211*	0.142	**0.153**
Treatment of the missing values: naive forecast (zero)																		
	2010Q1-2023Q4						2010Q1-2019Q4						2022Q1-2023Q4					
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
MF3PRF	1.02	1.06	1.05	1.01	1.09	1.04	1.03	1.07	1.06	1.01	1.08	1.05	**0.98**	1.01	1.01	1.03	1.11	1.00
COMB	1.02	1.03	1.03	1.01	1.04	1.02	1.02	1.04	1.04	1.00	1.05	1.02	**0.99**	1.01	1.01	1.01	1.03	1.02
Treatment of the missing values: realignment																		
	2010Q1-2023Q4						2010Q1-2019Q4						2022Q1-2023Q4					
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
MF3PRF	1.17	1.18	1.11	1.14	1.22	1.14	1.18	1.16	1.11	1.10	1.16	1.16	1.12	1.22	1.11	1.35	1.49	1.09
COMB	1.07	1.08	1.02	1.06	1.10	1.05	1.08	1.06	1.02	1.04	1.07	1.05	1.05	1.13	1.03	1.13	1.21	1.05
Treatment of the missing values: EM algorithm																		
	2010Q1-2023Q4						2010Q1-2019Q4						2022Q1-2023Q4					
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
MF3PRF	1.00	1.06	1.15	**0.98**	1.09	1.08	1.00	1.07	1.01	**0.97**	1.09	1.00	**0.99**	1.02	1.53	1.04	1.11	1.43
COMB	1.00	1.03	1.05	**0.99**	1.04	1.04	1.00	1.03	1.01	**0.98**	1.04	1.00	1.00	1.02	1.17	1.01	1.03	1.23

Notes: This table reports the RMSE and MAE criteria of the nowcasts in the first (M1), second (M2) and third (M3) month of the quarter. The results are given for three out-of-sample windows: 2010Q1-2023Q4, 2010Q1-2019Q4 and 2022Q1-2023Q4. The top part of the table reports the RMSE and MAE for the AR, MIBA, the MF-3PRF model in the reference configuration (1 factor, AR forecast of the missing values at the end of the sample) and the combination of MIBA and MF-3PRF forecasts (COMB). The best model for a given horizon is shown in bold. ** (*) significantly better than MIBA at 5% (10%) according to the one-sided Coroneo and Iacone (2020) test. The bottom panel reports the ratio of the criteria for MF-3PRF and COMB relative to their values in the reference case (first block of the table) for several alternative treatments of the missing values at the end of the sample: filling the missing values with 0, the realignment strategy, or an EM algorithm. A ratio less than 1 (in bold) indicates a gain relative to the reference case.

(b) Other variants

	Aggregation scheme in step 1																	
	2010Q1-2023Q4						2010Q1-2019Q4						2022Q1-2023Q4					
	RMSE			MAE			RMSE			MAE			RMSE			MAE		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
MF3PRF	1.04	1.03	1.07	1.02	1.05	1.05	1.04	1.03	1.09	1.01	1.05	1.08	1.04	1.01	1.01	1.06	1.05	**0.93**
COMB	1.02	1.01	1.02	1.01	1.01	1.02	1.02	1.01	1.03	1.00	1.01	1.02	1.02	1.01	**0.99**	1.02	1.01	1.00
	Pre-selection of variables with LARS-EN																	
	2010Q1-2023Q4						2010Q1-2019Q4						2022Q1-2023Q4					
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
MF3PRF	1.10	1.13	1.06	1.12	1.13	1.04	1.12	1.20	1.07	1.13	1.16	1.06	1.02	**0.89**	1.02	1.07	**0.98**	**0.99**
COMB	1.03	1.05	1.03	1.06	1.03	1.02	1.05	1.07	1.04	1.07	1.04	1.02	**0.99**	**0.93**	1.02	1.02	**0.99**	1.03
	Low frequency (Kelly-Pruitt, 2015)																	
	2010Q1-2023Q4						2010Q1-2019Q4						2022Q1-2023Q4					
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
MF3PRF	**0.99**	1.07	1.12	**0.98**	1.15	1.13	**0.97**	1.09	1.15	**0.97**	1.13	1.15	1.05	1.00	1.04	1.02	1.23	1.03
COMB	**0.99**	1.03	1.04	**0.99**	1.05	1.05	**0.98**	1.03	1.05	**0.99**	1.04	1.05	1.03	1.01	1.00	1.01	1.12	1.03
	2 factors																	
	2010Q1-2023Q4						2010Q1-2019Q4						2022Q1-2023Q4					
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
MF3PRF	1.11	1.09	1.02	1.07	1.10	1.01	1.12	1.08	**0.97**	1.06	1.10	**0.97**	1.06	1.14	1.18	1.10	1.12	1.15
COMB	1.04	1.04	1.00	1.02	1.01	**0.98**	1.05	1.03	**0.98**	1.01	1.01	**0.96**	1.03	1.07	1.08	1.04	1.04	1.09
	Blocking approach																	
	2010Q1-2023Q4						2010Q1-2019Q4						2022Q1-2023Q4					
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
MF3PRF	1.04	1.03	1.06	1.02	1.05	1.06	1.04	1.03	1.07	1.01	1.05	1.07	1.04	1.01	1.03	1.06	1.05	**0.99**
COMB	1.02	1.01	1.02	1.01	1.01	1.02	1.02	1.01	1.02	1.00	1.01	1.02	1.02	1.01	1.00	1.02	1.01	1.01
	Bridge approach																	
	2010Q1-2023Q4						2010Q1-2019Q4						2022Q1-2023Q4					
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
MF3PRF	1.06	1.05	1.00	1.04	1.08	1.00	1.09	1.07	1.00	1.05	1.08	1.00	**0.96**	1.00	1.00	**0.97**	1.07	1.00
COMB	1.03	1.02	1.00	1.02	1.02	1.00	1.05	1.03	1.00	1.02	1.03	1.00	**0.99**	1.01	1.00	**0.99**	1.01	1.00

Notes: In this table, we consider several variants and reports the ratio of the RMSE and MAE criteria of MF-3PRF and COMB relative to their values in the reference case (first block in Table 2a). First, we change the aggregation scheme in step 1 (taking the month 1 in M1, the month 2 in M2 and averaging the 3 months in M3). Second, we preselect the predictors with a LARS-EN algorithm before the estimation of the factor. Third, we use the 3PRF approach of Kelly and Pruitt (2015). The next variants are the two-factor model and the blocking approach. Last, we forecast the monthly predictors up to the end of the quarter before the estimation of the factor. A ratio below 1 (in bold) indicates a gain relative to the reference case.

APPENDIX 1 - Calculation of the contributions

Let $y_{T/3}$ be the target variable ($(T/3,1)$ vector), X_T a (T,N) matrix of predictors, Z a $(T/3,L)$ matrix of proxies and F_T a (T,L) matrix of factors. For simplicity, we consider $L=1$ factor.

In the *first step*, we run a (time-series) regression at a quarterly frequency:

$$X_{T/3} = \iota\Phi_0 + Z\Phi' + \varepsilon$$

or equivalently without the intercept $J_{T/3}X_{T/3} = J_{T/3}Z\Phi' + \tilde{\varepsilon}$ with $X_{T/3}$ the explanatory variables converted to the quarterly frequency (e.g. by average or by taking a specific month of each quarter) and $J_{T/3} = I_{T/3} - \frac{1}{T/3}\iota_{T/3}\iota'_{T/3}$ a matrix with the following properties: $J_{T/3}J'_{T/3} = J_{T/3}$ and $J'_{T/3} = J_{T/3}$ with and ι_T a T-vector of ones. The OLS estimator for Φ is $\hat{\Phi}' = \left(Z'J_{T/3}Z\right)^{-1}\left(Z'J_{T/3}X_{T/3}\right)$.

In the *second step*, we run a (cross-sectional) regression for each month:

$$X'_T = \iota\tilde{\Phi_0} + \hat{\Phi}F'_T + \eta$$

or equivalently without the intercept:

$$J_N X'_T = J_N\hat{\Phi}F'_T + \tilde{\eta}$$

with $J_N = I_N - \frac{1}{N}\iota_N\iota'_N$. The OLS estimator for F_T writes as follows:

$$\hat{F_T}' = \left(\hat{\Phi}'J_N\hat{\Phi}\right)^{-1}\left(\hat{\Phi}'J_N X'_T\right) \tag{1}$$

that is when replacing $\hat{\Phi}$ by its expression in the first step:

$$\hat{F_T}' = \underbrace{\left(Z'J_{T/3}Z\right)}_{S_{ZZ}}\left(\underbrace{Z'J_{T/3}X_{T/3}J_N}_{W'_{XZ}}\underbrace{X'_{T/3}J_{T/3}Z}_{S_{XZ}}\right)^{-1}\underbrace{Z'J_{T/3}X_{T/3}J_N}_{W'_{XZ}}X'_T \tag{2}$$

In the *third step*, we estimate the forecasting equation. Consider for example the case where we want to include the last three months of the estimated factor in the forecast equation:

$$y_{T/3} = c + Q^{(1)}_{T/3,T}F_T\beta_1 + Q^{(2)}_{T/3,T}F_T\beta_2 + Q^{(3)}_{T/3,T}F_T\beta_3 + \eta_{T/3}$$

with $Q^{(1)}_{T/3,T}F_T$ the vector of the contemporaneous values of the factor in month 1, $Q^{(2)}_{T/3,T}F_T$ in month 2 and $Q^{(3)}_{T/3,T}F_T$ in month 3. The matrices of temporal aggregation are defined as:

$$Q^{(1)}_{T/3,T} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

$$Q^{(2)}_{T/3,T} = \ . \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 &) & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

$$Q^{(3)}_{T/3,T} = \ . \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

The forecast equation can be rewritten in a more synthetic way as follows:

$$\underbrace{y_{T/3}}_{(T/3,1)} = c + \underbrace{Q_{T/3,3T}\tilde{F}_{3T}}_{(T/3,3)}\beta + \nu_t = c + F_{T/3}\beta + \nu_t$$

avec $\underbrace{Q_{T/3,3T}}_{(T/3,3T)} = \begin{bmatrix} Q^{(1)}_{T/3,T} & Q^{(2)}_{T/3,T} & Q^{(3)}_{T/3,T} \end{bmatrix}$, $\underbrace{\tilde{F}_{3T}}_{(3T,3)} = \begin{pmatrix} F_T & 0 & 0 \\ 0 & F_T & 0 \\ 0 & 0 & F_T \end{pmatrix}$ and $\underbrace{\beta}_{(3,1)} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$.

Since $F_T = \underbrace{X_T W_{XZ}}_{(T,1)} \underbrace{\left(S_{XZ}{}' W_{XZ}\right)^{-1} S_{ZZ}}_{(1,1)}$ (second step estimate), $\tilde{F}_{3T}$ rewrites as follows:

$$\underset{(3T,3)}{\tilde{F}_{3T}} = \begin{pmatrix} X_T W_{XZ} & 0 & 0 \\ 0 & X_T W_{XZ} & 0 \\ 0 & 0 & X_T W_{XZ} \end{pmatrix} \underbrace{\left(S_{XZ}{}' W_{XZ}\right)^{-1} S_{ZZ}}_{(1,1)}$$

As previously, we use the matrix $J_{T/3}$ to remove the intercept in the forecast equation:

$$\underset{(T/3,T/3)}{J_{T/3}} \underbrace{y_{T/3}}_{(T/3,1)} = \underset{(T/3,T/3)}{J_{T/3}} \underbrace{F_{T/3}}_{(T/3,3)} \beta + \widetilde{\nu}_t$$

and the forecast writes as follows:

$$J_{T/3}\hat{y} = \underset{(T/3,T/3)}{J_{T/3}} \underset{(T/3,3)}{Q_{T/3,3T}\tilde{F}_{3T}} \underset{(3,1)}{\widehat{\beta}} = \underset{(T/3,T/3)}{J_{T/3}} \underset{(T/3,3)}{F_{T/3}} \underset{(3,1)}{\widehat{\beta}}$$

with $\widehat{\beta}_{(3,1)} = \left(\underbrace{F'_{T/3} J_{T/3} F_{T/3}}_{(3,3)} \right)^{-1} \underbrace{F'_{T/3} J_{T/3} y_{T/3}}_{(3,1)}$

Finally, we have:

$$
\begin{aligned}
J_{T/3}\hat{y} = J_{T/3}Q_{T/3,3T}\tilde{F}_{3T}\widehat{\beta} = J_{T/3} \left[\begin{array}{ccc} Q^{(1)}_{T/3,T} & Q^{(2)}_{T/3,T} & Q^{(3)}_{T/3,T} \end{array}\right] \\
\begin{pmatrix} X_T W_{XZ} & 0 & 0 \\ 0 & X_T W_{XZ} & 0 \\ 0 & 0 & X_T W_{XZ} \end{pmatrix} \left(S_{XZ}'W_{XZ}\right)^{-1} S_{ZZ}\ \widehat{\beta}
\end{aligned}
$$

$$
= \underbrace{\left[\begin{array}{ccc} J_{T/3}Q^{(1)}_{T/3,T}X_T W_{XZ} & J_{T/3}Q^{(2)}_{T/3,T}X_T W_{XZ} & J_{T/3}Q^{(3)}_{T/3,T}X_T W_{XZ} \end{array}\right]}_{(T/3,3)} \underbrace{\left(S_{XZ}'W_{XZ}\right)^{-1} S_{ZZ}}_{(1,1)}\ \underbrace{\widehat{\beta}}_{(3,1)}
$$

or equivalently:

$$
\begin{aligned}
\hat{y} = \iota\overline{y} + J_{T/3}Q^{(1)}_{T/3,T}X_T \underbrace{W_{XZ}\left(S_{XZ}'W_{XZ}\right)^{-1} S_{ZZ}\widehat{\beta}_1}_{\widehat{\alpha_1}} + J_{T/3}Q^{(2)}_{T/3,T}X_T \underbrace{W_{XZ}\left(S_{XZ}'W_{XZ}\right)^{-1} S_{ZZ}\widehat{\beta}_2}_{\widehat{\alpha}_2} \\
+ J_{T/3}Q^{(3)}_{T/3,T}X_T \underbrace{W_{XZ}\left(S_{XZ}'W_{XZ}\right)^{-1} S_{ZZ}\widehat{\beta}_3}_{\widehat{\alpha}_3}
\end{aligned}
$$

with $\widehat{\beta}_i$ the i-th element of $\widehat{\beta}$. The contribution of variable i is given by:

$$
\left[\begin{array}{ccc} J_{T/3}Q^{(1)}_{T/3,T}X^i_T W_{XZ} & J_{T/3}Q^{(2)}_{T/3,T}X^i_T W_{XZ} & J_{T/3}Q^{(3)}_{T/3,T}X^i_T W_{XZ} \end{array}\right] \times \left(S_{XZ}'W_{XZ}\right)^{-1} S_{ZZ}\ \widehat{\beta}
$$

with X^i_T the matrix X_T whose columns are replaced by columns of zeros except column i.

APPENDIX 2 - Monthly database

Area	Type	Variable	Code	Start date	delay	Transf	Source
Industry	Survey	Change in output, compared with previous month	EVPRO	Jan-76	M+0	-	BdF
Industry	Survey	Change in deliveries, compared with previous month	EVLIV	Jan-81	M+0	-	BdF
Industry	Survey	Change in overall level of new orders, compared with previous month	EVCOM	Jan-81	M+0	-	BdF
Industry	Survey	Change in foreign orders, compared with previous month	EVCOME	Jan-81	M+0	-	BdF
Industry	Survey	Change in inventories of final goods, compared with previous month	EVSTPF	Jan-76	M+0	-	BdF
Industry	Survey	Current order books	ETCC	Jan-76	M+0	-	BdF
Industry	Survey	Current position in inventories of final goods	STPF	Jan-76	M+0	-	BdF
Industry	Survey	Average capacity utilisation rate (TUC)	TUC	Jan-81	M+0	-	BdF
Industry	Survey	Expected production for the coming month	PREVPRO	Jan-76	M+0	-	BdF
Industry	Real	IPI - Food products and beverages (C1)	IPI_C1	Jan-90	M+1	Δ log	Insee
Industry	Real	IPI - Coke and refined petroleum products (C2)	IPI_C2	Jan-90	M+1	Δ log	Insee
Industry	Real	IPI - Electrical and electronic equipment and machinery (C3)	IPI_C3	Jan-90	M+1	Δ log	Insee
Industry	Real	IPI - Transport equipment (C4)	IPI_C4	Jan-90	M+1	Δ log	Insee
Industry	Real	IPI - Textiles, wearing apparel, leather and related products (CB)	IPI_CB	Jan-90	M+1	Δ log	Insee
Industry	Real	IPI - Wood and paper products (CC)	IPI_CC	Jan-90	M+1	Δ log	Insee
Industry	Real	IPI - Chemicals and chemical products (CE)	IPI_CE	Jan-90	M+1	Δ log	Insee
Industry	Real	IPI - Pharmaceutical products (CF)	IPI_CF	Jan-90	M+1	Δ log	Insee
Industry	Real	IPI - Rubber and plastics products (CG)	IPI_CG	Jan-90	M+1	Δ log	Insee
Industry	Real	IPI - Basic metals and fabricated metal products (CH)	IPI_CH	Jan-90	M+1	Δ log	Insee
Industry	Real	IPI - Other manufacturing (CM)	IPI_CM	Jan-90	M+1	Δ log	Insee
Industry	Real	IPI - Mining and quarrying; energy, water supply, waste (DE)	IPI_DE	Jan-90	M+1	Δ log	Insee
Services	Survey	Change in activity over month, compared with previous month	EVACT	Oct-02	M+0	-	BdF
Services	Survey	Expected overall activity over next month	PREVACT	Oct-02	M+0	-	BdF
Services	Survey	Change in aggregate demand, compared with previous month	DETEM	Oct-02	M+0	-	BdF
Services	Survey	Cash positions at end of month	NIVTRES	Oct-02	M+0	Δ	BdF
Services	Real	Index of services production (IPS)	IPS_tot	March-05	M+2	Δ log	Insee
Construction	Real	Housing starts (total)	houstar	Jan-00	M+1	Δ log	MEEDDM
Construction	Real	Housing authorized (total)	housaut	Jan-00	M+1	Δ log	MEEDDM
Construction	Real	IPI - Construction (FZ)	IPI_FZ	Jan-90	M+1	Δ log	Insee
Demand	Real	Consumption - Food products	CONS1	Jan-80	M+1	Δ log	Insee
Demand	Real	Consumption - Transport equipment	CONS2	Jan-80	M+1	Δ log	Insee
Demand	Real	Consumption - Household durables	CONS3	Jan-80	M+1	Δ log	Insee
Demand	Real	Consumption - Other durables	CONS4	Jan-80	M+1	Δ log	Insee
Demand	Real	Consumption - Textile-leather	CONS5	Jan-80	M+1	Δ log	Insee
Demand	Real	Consumption - Other engineered goods	CONS6	Jan-80	M+1	Δ log	Insee
Demand	Real	Consumption - Energy	CONS7	Jan-80	M+1	Δ log	Insee
Demand	Real	New private car registrations	immat	Jan-58	M+0	log	OECD
Demand	Real	Consumption - electricity	CELEC	Jan-96	M+0	Δ log	RTE
Demand	Real	Sales volume in wholesale and retail trade	sales	Jan-05	M+2	Δ log	Insee
Demand	Real	Total unemployment	unemp	Jan-83	M+2	Δ	Eurostat
Demand	Real	Unemployment - Less than 25 years	unempy	Jan-83	M+2	Δ	Eurostat
Finance	Text	EPU - France	EPU	Jan-87	M+0	-	Link
Finance	Stock	CAC 40	CAC	March-90	M+0	Δ log	Yahoo
Finance	Stock	DAX	DAX	Jan-88	M+0	Δ log	Yahoo
Finance	Stock	SP500	SP	Jan-85	M+0	Δ log	Yahoo
Finance	Stock	Nikkei	NIKKEI	Jan-85	M+0	Δ log	Yahoo
Finance	Money	M1	M1	Dec-77	M+1	Δ log	BdF
Finance	Money	M2	M2	Jan-80	M+1	Δ log	BdF
Finance	Money	M3	M3	Jan-70	M+1	Δ log	BdF
Finance	Interest rate	Overnight interbank rate - France	TIOV	Jan-90	M+0	-	OECD
Finance	Interest rate	3 month interbank rate - France	TI3M	Jan-90	M+0	-	OECD
Finance	Interest rate	Long-term interest rate - France	TILT	Jan-90	M+0	-	OECD
Finance	Price	Consumer Price Index	CPI	jan-70	M+1	Δ log	Insee
Finance	Price	Oil price	Brent	Jan-60	M+1	Δ log	Insee
Finance	Price	Agricultural Raw Material Index	PXAGR	Jan-90	M+1	Δ log	FMI
Finance	Price	Gold price	Gold	Jan-55	M+1	Δ log	FMI
International	Survey	ESI Germany	ESI_DE	Jan-80	M+0	Δ	Eurostat
International	Real	Manufacturing production - Germany	IPI_DE	Jan-91	M+1	Δ log	Eurostat
International	Survey	ESI Euro Area	ESI_EA	Jan-80	M+0	Δ	Eurostat
International	Real	Manufacturing production - Euro Area	IPI_EA	Jan-91	M+2	Δ log	Eurostat

Notes: This table presents the characteristics of the 60 monthly indicators, including the area (used to separate the contributions), their type, code, starting date, publication lags, transformation and source. The publication lags are calculated with respect to the release of the Banque de France's business surveys.

APPENDIX 3 - Benchmark models

Estimation results of the MIBA equations (1995Q1-2023Q4)

M1			M2			M3		
Variable	coef	t-stat	Variable	coef	t-stat	Variable	coef	t-stat
Intercept	0.11	2.89	Intercept	0.04	1.09	Intercept	0.05	1.36
Dum09Q1	-0.97	-3.54	Dum09Q1	-0.72	-2.87	Dum09Q1	-0.80	-3.24
AR(1)	-0.42	-5.42	AR(1)	-0.38	-5.50	AR(1)	-0.40	-5.87
$EVLIV_{t-2/3}$	0.02	4.41	$EVLIV_{t-1/3}$	0.02	7.08	$EVLIV_{t}$	0.02	4.75
$PREVPRO_{t-2/3}$	0.04	7.12	$PREVPRO_{t-1/3}$	0.03	4.38	$EVLIV_{t-1/3}$	0.0218	7.2979
			$EVLIV_{t-2/3}$	0.0202	4.6927	$EVLIV_{t-2/3}$	0.0245	6.389
Adj-R2	0.60		Adj-R2	0.67		Adj-R2	0.68	
$\hat{\sigma}_e$	0.25		$\hat{\sigma}_e$	0.22		$\hat{\sigma}_e$	0.22	
BIC	-2.62		BIC	-2.78		BIC	-2.81	
AR(4)	0.87		AR(4)	0.42		AR(4)	0.65	
Het	0.89		Het	0.78		Het	0.42	
JB	0.03		JB	0.50		JB	0.50	

Notes: This table reports the estimation results of the MIBA model in months 1, 2 and 3 for the period 1995Q1 to 2023Q4 (excluding the observations from 2020Q1 to 2021Q4). The table shows the parameter estimates and the corresponding t-statistics. The p-values of the residual tests are reported at the bottom of the table. AR(4) denotes the Breusch-Godfrey test for residual serial correlation up to order four, Het the Breusch-Pagan-Godfrey test for heteroskedasticity, and JB the Jarque Bera test for normality.