DEPARTMENT OF ECONOMICS

EUI Working Papers

ECO 2012/07
DEPARTMENT OF ECONOMICS

A COMPARISON OF MIXED FREQUENCY APPROACHES FOR MODELLING EURO AREA MACROECONOMIC VARIABLES

Claudia Foroni and Massimiliano Marcellino

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
DEPARTMENT OF ECONOMICS

A Comparison of Mixed Frequency Approaches for Modelling Euro Area Macroeconomic Variables

CLAUDIA FORONI

and

MASSIMILIANO MARCELLINO

EUI Working Paper **ECO** 2012/07

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

A comparison of mixed frequency approaches for modelling Euro area macroeconomic variables*

Claudia Foroni
European University Institute

Massimiliano Marcellino
European University Institute,
Bocconi University and CEPR

20 February 2012

Abstract

Forecast models that take into account unbalanced datasets have recently attracted substantial attention. In this paper, we focus on different methods proposed so far in the literature to deal with mixed-frequency and ragged-edge datasets: bridge equations, mixed-data sampling (MIDAS), and mixed-frequency (MF) models. We discuss their performance on now- and forecasting the quarterly growth rate of Euro area GDP and its components, using a very large set of monthly indicators taken from Eurostat dataset of Principal European Economic Indicators (PEEI). We both investigate the behavior of single indicator models and combine first the forecasts within each class of models and then the information in the dataset by means of factor models, in a pseudo real-time framework. Anticipating some of the results, MIDAS without an AR component performs worse than the corresponding approach which incorporates it, and MF-VAR seems to outperform the MIDAS approach only at longer horizons. Bridge equations have overall a good performance. Pooling many indicators within each class of models is overall superior to most of the single indicator models. Pooling information with the use of factor models gives even better results, at least at short horizons. A battery of robustness checks highlights the importance of monthly information during the crisis more than in stable periods. Extending the analysis to a real-time context highlights that revisions do not influence substantially the results.

J.E.L. Classification: E37, C53

Keywords: mixed-frequency data, mixed-frequency VAR, MIDAS, factor models, nowcasting, forecasting

*We would like to thank Eric Ghysels, Helmut Luetkepohl, Gian Luigi Mazzi, Mike McCracken and Christian Schumacher, as well as seminar participants at the European University Institute and at the 8th IIF Workshop at Banque de France, for very helpful comments on a previous draft. The usual disclaimers apply.

1 Introduction

In recent times, forecast models that take into account the information in unbalanced datasets have attracted substantial attention. Policy-makers, in particular, need to assess the current state of the economy in real-time, when only incomplete information is available.

In real-time, the unbalancedness of datasets arises mainly due to two features: the different sampling frequency with which the indicators are available and the so-called "ragged-edge" problem, namely publication delays of indicators cause missing values of some of the variables at the end of the sample, see Wallis (1986). As an example, one of the key indicators of macroeconomic activity, the Gross Domestic Product (GDP), is released quarterly and with a considerable publication lag, while a range of leading and coincident indicators is available more timely and at a monthly or even higher frequency.

In this paper, we focus on different classes of models which deal with an unbalanced dataset. In particular, we concentrate on three main streams in the literature: the bridge models, the state-space approach and the MIDAS approach. Bridge equations are one of the most used techniques, since they link monthly to quarterly variables, choosing the regressors because of their timely information content (see e.g. Baffigi et al. (2004)). State-space approaches aim at capturing the joint dynamics of indicators at different frequency. Two main models are developed within this framework: the mixed-frequency VAR and the factor models (see e.g. Zadrozny (1988) and Mariano and Murasawa (2003)). In both cases, the use of the Kalman filter allows obtaining a monthly estimate of the quarterly series. The third approach, the MIDAS method, is based on a univariate reduced form regression, which uses highly parsimonious lag polynomials to exploit the content in the higher-frequency explanatory variable and provide a high-frequency update of the quarterly frequency variable (see e.g. Ghysels et al (2004) for financial applications and Clements and Galvao (2008) for macroeconomic applications). Recently, these Factor and MIDAS approaches have been merged in the Factor-MIDAS model, augmenting the MIDAS regressions with the factors extracted from a large dataset in high frequency (see Marcellino and Schumacher (2010)).

All these approaches tackle data at different frequency and with publication delays, but at the same time they display different characteristics, and this makes it difficult to rank them a priori only on the basis of theoretical considerations. Therefore, we compare them in an extensive and detailed empirical application. Specifically, in this paper we extend the analysis in Kuzin, Marcellino and Schumacher (2011, 2012), and focus on now- and forecasting the quarterly growth rate of Euro Area GDP, using a very large set of monthly indicators (around 150 monthly series), with a wide number of forecasting methods. In particular, the main distinctive features of this analysis with respect to the one conducted in Kuzin et al. (2011, 2012) are a different sample, including the financial crisis, and the different set of indicators included. Moreover, more approaches are assessed in this analysis, including the bridge models, and small models are compared with large ones, with the introduction of factors.

In addition, to compare the different approaches, we both investigate the behavior of single indicator models and combine the forecasts within each class of models. We conduct the analysis recursively in a pseudo real-time framework, taking into account the ragged-edge structure of the dataset, and we assess the now- and forecasting performance of the models by comparing the resulting mean-squared errors (MSE). Specifically, we first investigate the performance of a large number of single indicator models. Then, since all the approaches can be subject to misspecification issues, related e.g. to indicator selection, the number of lags, etc., we propose forecast pooling as a way of dealing with this model uncertainty (see Timmermann (2006) for an extensive review of forecast pooling approaches). We consider three simple weighting schemes: the average, the median and a performance-based weighting scheme (where the weights are equal to the inverse of the past MSE performance), and discuss whether they provide robust results against misspecification and parameter instability. We also consider the use of factor models, as a way of pooling information instead of pooling forecasts. Even in the case of factor models, we take into account the mixed-frequency nature of the dataset and compare different techniques. As additional robustness checks for our findings, we extend the forecast horizon up to four quarters ahead and compare the results obtained from recursive and rolling estimation methods. In the case of MIDAS models, we also extend the analysis to higher than monthly frequency data, incorporating in the analysis financial variables and interest rates and spreads at weekly frequency[1].

Moreover, in order to replicate the real-time situation in which policy makers and institutions need to assess the state of the economy, we investigate the behavior of a small group of indicators in a genuine real-time context, using data as they were at that moment, which allows taking into account data revisions.

Finally, we extend our analysis to the now- and forecasting of the economic activity components, disaggregating GDP from the output side in six branches following the NACE classification, and from the expenditure side, distinguishing among consumption, gross fixed capital formation and external balance.

We anticipate here some of the results concerning now- and forecasting of the aggregate GDP. First, there is no clear evidence of which approach to handling mixed-frequency data with ragged-edges is the best. A general finding is that looking at single indicator models, MF-VAR does not show particular improvement in terms of MSE, while generally MIDAS and bridge equations appear good methods to obtain better forecasts. Moreover, MIDAS without an AR component performs worse than the corresponding approach which incorporates it. Second, pooling forecasts is an helpful device in improving forecasting performance. Looking at the pooling results throughout different approaches, evidence seems to be in favour of a better performance of the MIDAS approach at every now-

[1] Andreou, Ghysels and Kourtellos (2009) also looked at large cross-section of daily financial data for forecasting quarterly real GDP growth, finding that MIDAS regression models provide substantial forecast gains against various benchmark forecasts.

and forecast horizon. Even better results are obtained by pooling information with the use of factor models instead of pooling forecasts[2]. Even the standard quarterly factor models can outperform the AR benchmark. The same general findings hold for those GDP components for which it is easy to find monthly indicators. However, checking the performance in different subsamples does not conduct to uniform results over time. In particular, in periods before the current crisis neither one of these more sophisticated models displays a clear superiority against the benchmark. However, during the first quarters of the crisis, at the end of 2008 and beginning of 2009, the models which take into account monthly information are clearly better in outperforming the benchmark. Finally, conducting the same analysis in a genuine real-time framework confirms the results found in a pseudo real-time context, allowing for the conclusion that data revisions, even though quite big in size, do not discard the reliability of the results obtained with final-vintage data.

The paper proceeds as follows. Section 2 describes the data. Section 3 discusses model specification. Section 4 presents the results on now- and forecasting quarterly Euro Area GDP, using single indicator models. Section 5 focuses on forecast pooling of the individual models, with different combination schemes. Section 6 summarizes the robustness analyses, including extending the forecast horizon, conducting a rolling evaluation and splitting the evaluation sample to assess stability of the findings over time, introducing several frequencies in the explanatory variables and conducting the analysis in a genuine real-time framework. Appendix A provides full details. Section 7 looks at the results from factor models, which pool information from a large number of time series, differently from what shown in Section 5 where the forecasts of individual models are pooled. Section 8 summarizes the results for GDP components, with more details provided in Appendix B. Section 9 concludes.

2 Data

The dataset contains Euro area quarterly and monthly series taken from the Eurostat dataset of Principal European Economic Indicators (PEEI). We consider a fixed country composition of the Euro area, as it is in 2009 at the end of the sample, with 16 countries. We collect quarterly GDP from 1996Q1 until 2009Q2, both at aggregate level and disaggregated into branches of activity and expenditure components. We also dispose of around 150 macroeconomic monthly indicators from January 1996 to August 2009, including con-

[2]Andreou, Ghysels and Kourtellos (2009) extract information from a large daily dataset and use the factors in a MIDAS framework. Differently from what we do, they first extract a number of daily factors, then estimate MIDAS models with each one of the factors and finally pool the forecasts obtained from the different single-factor models. What they find is that the use of daily factors improves the forecasting perfomance compared to their random-walk benchmark. Moreover, comparing the performance of the forecasts obtained pooling factors with those obtained pooling single indicators, they obtain quite similar results, but with a slightly better performance of factor models. These results are similar to what we obtain in Section 7.

sumer and producer price index by sector, industrial production and (deflated) turnover indexes by sector, car registrations, new orders received index, business and consumers surveys with their components, sentiment indicators, unemployment indices, monetary aggregates, interest and exchange rates. All series are seasonally adjusted. Details about transformations and stock/flow nature of the data can be found in the Appendix C.

The dataset is a final dataset. However, we take into account one of the specific characteristics of the macroeconomic data in real time, the ragged-edge structure of the dataset due to different publication lags of the series. The timing of data releases is more or less the same every month, and this allows us to replicate the same pattern of missing values at the end of each recursive sample. To have an idea of the ragged-edge structure of the dataset, we show in Table 1 the lags of the main series in the dataset.

As outlined in the recent literature, the use of pseudo real-time datasets, which replicate the differences in the availability of data, lead to significant differences in results with respect to the use of artificially balanced datasets, see Giannone, Reichlin and Small (2008) and Breitung and Schumacher (2008) among the others. Therefore, in our paper, we replicate the ragged-edge structure of the dataset we observed at the downloaded date (31st August 2009) in each of the recursive subsamples: for each series we observe the number of missing values at the end, and we impose the same number of missing observations at each recursion, so to mimic the availability of data in real-time. As a clarifying example, at the end of August 2009, we have data on the CPI index and financial variables available until July 2009, but data on unemployment and industrial production only until June 2009, i.e. while the former variables have a delay of one month, the latter become available with a delay of two months. Therefore, when in our recursive exercise we use a subsample from January 1996 to March 2009, we impose to use the CPI index until February 2009 and unemployment until January 2009, to replicate the same data availability we would have in real-time. Moreover, we impose a similar structure of publication delays also on quarterly variables, namely GDP and its components on the supply and expenditure side. To do this, we take into account that in the Euro area, GDP and its breakdown into components are available in the third month after the end of the quarter of interest, for example the GDP figure for 2009Q1 becomes available in June 2009.

3 Model Specification

The aim of the experiment is to evaluate the performance of different methods available in the literature which deal with unbalanced mixed-frequency datasets, when the number of indicators is very large.

We first recursively estimate and then now- and forecast Euro area GDP growth rate, with the first evaluation quarter fixed at 2003Q1 and the last at 2009Q1, for a total of 25 recursive evaluation samples. For each quarter we compute nowcasts and forecasts, based on different information sets. In each recursion we want to predict the current GDP growth and the values of one and two quarters ahead. Therefore, for every single quarter in the evaluation sample we have three nowcasts and six forecasts. As an example, for 2005Q3, we have a nowcast computed in September 2005, one in August 2005 and one in July 2005; moreover, we have one-quarter ahead forecasts computed in June, May and April 2005 and two-quarters ahead forecasts computed in March, February and January 2005. Each of the nine projections we have for every realization of the GDP growth rate in the evaluation sample is based on different information, available at the point in time in which the projection is computed. Therefore we exploit the ragged-edge structure of the dataset and consider only the information available at that moment. As we have already stated in describing the data, the GDP in Euro Area is available with a delay of three months after the end of the quarter of interest.

In terms of notation, we denote GDP growth as y_{t_q}, where $t_q = 1, 2, 3, ..., T_q^y$ is a quarterly time index and T_q^y is the final quarter for which GDP is available. GDP growth can be expressed also as a monthly variable with missing values, by setting $t_m = t_q, \forall t_m = 3t_q$, where t_m is the monthly time index. GDP growth is observable only in $t_m = 3, 6, 9, ..., T_m^y$ where $T_m^y = 3T_q^y$. Therefore, what we want to obtain is the nowcast or forecast of the economic activity h_q quarters ahead or, equivalently, $h_m = 3h_q$ months ahead. We exploit monthly stationary indicators x_{t_m}, with $t_m = 1, 2, 3, ..., T_m^x$, where T_m^x is the final month for which the indicator is available. Usually monthly indicators are available earlier during the quarter than the GDP release, so generally we condition the forecast on the information available up to month T_m^x, which includes GDP information up to T_q^y and indicator observations up to T_m^x with $T_m^x \geq T_m^y = 3T_q^y$. The GDP growth forecast is indicated as $y_{T_m^y+h_m|T_m^x}$.

3.1 The Bridge Model approach

One of the early econometric approaches in the presence of mixed-frequency data relies on the use of bridge equations, see e.g. Baffigi, Golinelli, Parigi (2004) and Diron (2008). Bridge equations are linear regressions that link ("bridge") high frequency variables, such as industrial production or retail sales, to low frequency ones, e.g. the quarterly real GDP growth, providing some estimates of current and short-term developments in advance of the release. The "Bridge model" technique allows computing early estimates of the low-frequency variables by using high frequency indicators. They are not standard macroeconometric models, since the inclusion of specific indicators is not based on causal relations, but on the statistical fact that they contain timely updated information.

In our exercise, since the monthly indicators are usually only partially available over the projection period, the predictions of quarterly GDP growth are obtained in two steps. First, monthly indicators are forecasted over the remainder of the quarter, on the basis of univariate time series models, and then aggregated to obtain their quarterly correspondent values. Second, the aggregated values are used as regressors in the bridge equation which allows to obtain forecasts of GDP growth.

Therefore, the bridge model to be estimated is:

$$y_{t_q} = \alpha + \sum_{i=1}^{j} \beta_i (L)\, x_{it_q} + u_{t_q} \tag{1}$$

where $\beta_i (L)$ is a lag polynomial of length k, and x_{it_q} are the selected monthly indicators aggregated at quarterly frequency.

In order to forecast the missing observations of the monthly indicators which are then aggregated to obtain a quarterly value of x_{it_q} , it is common practice to use autoregressive models, where the lag length is based on information criteria.

In our exercise, we use autoregressive models, where the lag length is chosen according to the BIC criterion, with the maximum lag fixed to 12. The data are then aggregated with standard methods, according to the stock/flow nature of the variables, specifically averaging over one lower-frequency period for stock variables and summing over the high-frequency indicators for flow variables. Once the data are aggregated, the number of lags of the indicators to include in the bridge model is chosen according to the BIC criterion, with a maximum lag equal to 4.

3.2 The MF-VAR approach

One of the most compelling approaches in the literature to deal with mixed-frequency time series at the moment is the one proposed by Zadrozny (1988) for directly estimating a VARMA model sampled at different frequencies. The approach treats all the series as generated at the highest frequency, but some of them are not observed. Those variables that are observed only at the low frequency are therefore considered as periodically missing.

Following the notation of Mariano and Murasawa (2004), we consider the state-space representation of a VAR model, treating quarterly series as monthly series with missing observations. The disaggregation of the quarterly GDP growth, y_{t_m}, into the unobserved month-on-month GDP growth, $y^*_{t_m}$, is based on the following aggregation equation:

$$\begin{aligned} y_{t_m} &= \frac{1}{3}\left(y^*_{t_m} + y^*_{t_m-1} + y^*_{t_m-2}\right) + \frac{1}{3}\left(y^*_{t_m-1} + y^*_{t_m-2} + y^*_{t_m-3}\right) + \\ &\quad + \frac{1}{3}\left(y^*_{t_m-2} + y^*_{t_m-3} + y^*_{t_m-4}\right) \\ &= \frac{1}{3}y^*_{t_m} + \frac{2}{3}y^*_{t_m-1} + y^*_{t_m-2} + \frac{2}{3}y^*_{t_m-3} + \frac{1}{3}y^*_{t_m-4} \end{aligned} \tag{2}$$

where $t_m = 3, 6, 9, ..., T_m$, since GDP growth, y_{t_m}, is observed the last month of each quarter, while $y^*_{t_m}$ is never observed.

This aggregation equation comes from the assumption that the quarterly GDP series (in log levels), Y_{t_m}, is the geometric mean of the latent monthly random sequence

$Y^*_{t_m}, Y^*_{t_m-1}, Y^*_{t_m-2}$. Taking the three-period differences and defining $y_{t_m} = \Delta_3 Y_{t_m}$ and $y^*_{t_m} = \Delta Y^*_{t_m}$, we obtain eq. (2).

Let for all t_m the latent month-on-month GDP growth $y^*_{t_m}$ and the corresponding monthly indicator x_{t_m} follow a bivariate VAR(p) process

$$\phi(L_m)\begin{pmatrix} y^*_{t_m} - \mu^*_y \\ x_{t_m} - \mu_x \end{pmatrix} = u_{t_m}, \tag{3}$$

where $u_{t_m} \sim N(0, \Sigma)$.

State-space representation

In our exercise we determine the number of lags, p, according to the Bayesian Information Criterion (BIC), with a maximum lag order of $p = 4$ months.

With $p \leq 4$, and defining s_{t_m} and z_{t_m} as

$$s_{t_m} = \begin{pmatrix} z_{t_m} \\ \vdots \\ z_{t_m-4} \end{pmatrix}, \quad z_{t_m} = \begin{pmatrix} y^*_{t_m} - \mu^*_y \\ x_{t_m} - \mu_x \end{pmatrix},$$

a state-space representation of the MF-VAR is

$$s_{t_m} = F s_{t_m-1} + G v_{t_m} \tag{4}$$

$$\begin{pmatrix} y_{t_m} - \mu_y \\ x_{t_m} - \mu_x \end{pmatrix} = H s_{t_m} \tag{5}$$

with $\mu_y = 3\mu^*_y$ that holds, and $v_{t_m} \sim N(0, I_2)$.

In the notation,

$$F = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}; \quad F_1 = \begin{bmatrix} \phi_1 & \ldots & \phi_p & 0_{2\times 2(5-p)} \end{bmatrix}; \quad F_2 = \begin{bmatrix} I_8 & 0_{8\times 2} \end{bmatrix}, \tag{6}$$

$$G = \begin{bmatrix} \Sigma^{1/2} \\ 0_{8\times 2} \end{bmatrix}; \quad H = \begin{bmatrix} H_0 & \ldots & H_4 \end{bmatrix} \tag{7}$$

where H contains the lag polynomial

$$H(L_m) = \begin{bmatrix} 1/3 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2/3 & 0 \\ 0 & 0 \end{bmatrix} L_m + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} L_m^2 + \begin{bmatrix} 2/3 & 0 \\ 0 & 0 \end{bmatrix} L_m^3 + \begin{bmatrix} 1/3 & 0 \\ 0 & 0 \end{bmatrix} L_m^4 \tag{8}$$

Estimation and forecasting

The state-space representation of the mixed-frequency VAR model, described by equations (4) and (5), can be estimated by maximum-likelihood even in the presence of missing observations due to publication lags and the low-frequency nature of GDP. However, when the number of parameters is large, the ML method can fail to converge.

Therefore, we implement the EM algorithm modified to allow for missing observations. As in Mariano and Murasawa (2004), we consider the missing values as realizations of some iid standard normal random variables, i.e.

$$y_{t_m}^{+} = \begin{cases} y_{t_m} & \text{if } y_{t_m} \text{ is observable} \\ \zeta_{t_m} & \text{otherwise} \end{cases} \tag{9}$$

where ζ_{t_m} is a draw from a standard normal distribution independent of the model parameters.

The measurement equation is modified accordingly in the first two months of each quarter, where the upper row of H is set to zero and a standard normal error term is added, so that the Kalman filter skips the random numbers. Since the realizations of the random numbers do not matter in practice, we replace the missing values with zeros.

We use the Kalman smoother to obtain forecasts of the economic activity. Although GDP growth for a particular month is not available, the smoother considers the monthly indicators available for the same quarter, so that nowcasting is also possible. For the months in which no observations are available also for the monthly indicators, the Kalman smoother acts exactly as the Kalman filter. What we obtain are iterative multistep forecasts and an estimate of the expected value of GDP growth in each month.

3.3 The MIDAS approach

MIDAS regressions are essentially tightly parameterized, reduced form regressions that involve processes sampled at different frequencies. The response to the higher-frequency explanatory variable is modelled using highly parsimonious distributed lag polynomials, to prevent the proliferation of parameters that might otherwise result, as well as the issues related to lag-order selection (see Ghysels et al. (2006), Andreou et al. (2010)).

3.3.1 The basic MIDAS model

MIDAS models are a direct forecasting tool, which directly rely on current and lagged indicators to estimate current and future GDP. This yields different models for different forecasting horizons. The forecast model for horizon $h_q = h_m/3$ is:

$$y_{t_q+h_q} = y_{t_m+h_m} = \beta_0 + \beta_1 b\left(L_m, \theta\right) x_{t_m+w}^{(3)} + \varepsilon_{t_m+h_m}, \tag{10}$$

where y_{t_m} and x_{t_m} are respectively the GDP growth and the monthly indicator, $x_{t_m}^{(3)}$ is the corresponding skip sampled monthly indicator, $w = T_m^x - T_m^y$ and $b\left(L_m, \theta\right)$ is the exponential Almon lag,

$$b\left(L_m,\theta\right)=\sum_{k=0}^{K}c\left(k,\theta\right)L_m^k,\quad c\left(k,\theta\right)=\frac{\exp\left(\theta_1 k+\theta_2 k^2\right)}{\sum\limits_{k=0}^{K}\exp\left(\theta_1 k+\theta_2 k^2\right)}. \tag{11}$$

We estimate the MIDAS model using nonlinear least squares (NLS) in a regression of y_{t_m} on $x_{t_m-k}^{(3)}$, yielding coefficients $\widehat{\theta}_1,\widehat{\theta}_2,\widehat{\beta}_0$ and $\widehat{\beta}_1$. Since the model is $h-$dependent, we reestimate it for multi-step forecasts and when new information becomes available. The forecast is given by:

$$\widehat{y}_{T_m^y+h_m|T^x}=\widehat{\beta}_0+\widehat{\beta}_1 B\left(L^{1/m};\widehat{\theta}\right)x_{T_m^x}^{(3)}. \tag{12}$$

As far as the specification is concerned, we use a large variety of initial parameter specifications, compute the residual sum of squares from equation (10) and choose the parameter set which gives the smallest RSS as initial values for the NLS estimation. K in the exponential Almon lag function is fixed at 12, whether the parameters are restricted to $\theta_1<5$ and $\theta_2<0$.

3.3.2 An extension: the AR-MIDAS model

A natural extension of the basic MIDAS model is the introduction of an autoregressive term. Including the AR dynamics is desirable but not straightforward. Ghysels, Santa-Clara and Valkanov (2004) show that the introduction of lagged dependent variables creates efficiency losses. Moreover, it would result in the creation of seasonal patterns in the explanatory variables.

Therefore, we follow Clements and Galvao (2008) and introduce the AR dynamics as a common factor to rule out seasonal patterns. We estimate the AR-MIDAS, defined as:

$$y_{t_m+h_m}=\beta_0+\lambda y_{t_m}+\beta_1 b\left(L_m,\theta\right)\left(1-\lambda L_m^3\right)x_{t_m+w}^{(3)}+\varepsilon_{t_m+h_m}, \tag{13}$$

where the λ coefficient can be estimated together with the other coefficients by NLS. Even in this case, we follow the procedure described for the MIDAS approach: first compute the RSS from (13), choose the parameters that minimize it, and use them as initial values for the NLS estimation.

4 Results for Euro area GDP

We first show the results of individual models for different now- and forecast horizons. We consider MSE as measure to compare the performance of the different models. As a benchmark, we recursively estimate an AR model of GDP growth, where the lag length is specified according to the BIC criterion. In our exercise we also considered the recursively estimated in-sample mean as a benchmark, but since the resulting MSE is greater than the MSE of the AR process, we preferred to adopt the AR in our analysis.

Table 2 provides evidence on the average performance of the different classes of mixed-frequency models, in order to investigate their properties and capture their differences and similarities over the full set of individual indicators. We report the average relative MSE performance for now- and forecasting quarterly GDP growth at different horizons for different classes of models, against the AR benchmark[3]. First, we estimate every individual model and compute the relative MSE with respect to the benchmark, i.e. we calculate the MSE of every single indicator model relative to the MSE of the AR model. Then, we take the average all across the indicators of the relative MSE within a model class (Bridge, MIDAS, AR-MIDAS and MF-VAR).

Bridge models perform well and generally outperform the benchmark, though the gains are small. For most of the horizons, MIDAS cannot clearly perform better than the benchmark (the relative MSE is very close or greater than 1), whereas when we introduce the AR component, the AR-MIDAS model beats the benchmark at all the horizons up to $h_m = 7$, behaving particularly well at short horizons. MF-VAR provides an average relative MSE larger than one, and equal to one only for larger horizons, showing therefore no particular gains in terms of forecasting performance.

A special comment on the results for $h_m = 1$ is needed: all the models except the AR-MIDAS perform very badly. This is due to the specific publication lag of the GDP in the Euro area. Since, as mentioned before, the GDP is released at the beginning of the third month of the next quarter, the one-month ahead nowcast is the only one computed with the GDP figure of the previous quarter already available. Looking at the results, therefore, this means that the information contained in the first lag for GDP matters a lot. When the performance in the previous quarter is available, it is very hard to reach a better nowcast with only the information contained in monthly indicators. On the contrary, the releases of the monthly series can improve the performance of a simple AR process, when the monthly information is added to the autoregressive component, as it is clear from the results of the AR-MIDAS.

As another way to compare the alternative mixed frequency models, we compute the relative MSE of the (AR-)MIDAS and MF-VAR to the MSE of bridge equations for each single indicator, and then average over all the relative MSEs we obtain. The results are shown in Table 3 (where we also report the median over all the relative MSEs).

The only model which is able to beat the bridge equations approach is the AR-MIDAS, for all the horizons up to $h_m = 7$. The inclusion of the lagged GDP therefore makes the difference in the performance of the different models. The same approach without including the autoregressive component shows worse results at every horizon. As already

[3]Since all the results are expressed in ratios with respect to the AR benchmark, in Table 2 we report also the absolute value of the MSE of the benchmark. Moreover, for the sake of completeness, we report also the variance of the GDP growth. As we can observe from these numbers, the AR benchmark outperform the naive variance of GDP growth, at least up to $h_m = 4$, that is at least for each nowcasting horizon.

mentioned before, the MF-VAR performs slightly better than the (AR-)MIDAS only for relatively long horizon.

We can conclude that bridge equations and AR-MIDAS outperform MF-VAR for short now- and forecast horizons with the latter better than the former. The MF-VAR improves its performance only for $h_m = 8, 9$, but the differences with the other approaches are very limited for these longer horizons.

So far, we looked at the average performance of the models over all indicators. We can have some insights also looking at the best individual indicators. We find few monthly variables that outperform the AR benchmark up to at least horizon $h_m = 6$ and with different estimation methods. More in detail, there are three variables - the economic sentiment indicator, the production expectations for the months ahead and the unemployment rate under 25 years - which outperform the benchmark with all the four methods described above (bridge models, MIDAS, AR-MIDAS and MF-VAR), and two others - the manufacturing new orders received index and the general economic situation over the next 12 months - which outperform the benchmark with three methods out of four. Three of the best performing indicators are business survey components: this confirms the evidence found in the literature about the importance of the survey data as a source of timely information about the current economic situation. The two other best performing indicators are instead "hard data", that is variables on actual production and demand, which have usually more relevance in forecasting economic activity but on the other hand they are less timely. In Table 4, we show the performance of these best performing individual models. As it is evident from the results in the table, the gains for some horizons, especially the very short ones, are quite relevant.

5 Forecast pooling

The availability of many indicators leads to many forecasts of the same variable, This suggests to exploit information in the individual forecasts and combine them. Implementing forecast combination allows us to overcome misspecification bias, parameter instability and measurement errors in the datasets which may be present in the individual forecasts (see Timmermann (2006) for a detailed overview on forecast combination).

Estimating combination weights is hard since a large data sample relative to the number of the models is required to obtain appropriate estimates of the weights. Hence, here we provide results from forecast combinations within the same class of models, based on simpler combination schemes. More precisely, we consider three different combination schemes: the mean - the most exploited method in the literature, the median - the simplest rank-based weighting scheme, and a weighted mean that lets the combination weights be inversely proportional to the MSE of the previous four-quarter performance of the model (see Stock and Watson (2001))[4].

[4] Aiolfi and Timmermann (2006) propose more sophisticated combination schemes which produce

In Table 5, we provide the relative MSE performance of model pooling within a given class of models against the benchmark. The steps we conducted are the following: first, forecasts from single indicator models are computed, and means, medians and weighted means of all the forecasts within a single model are obtained. Second, the MSE of these three different forecast combinations is calculated and divided by the MSE of the benchmark.

The results show that AR-MIDAS pooling performs pretty well: it outperforms the benchmark at each of the nine horizons with all the three weighting combination schemes. The weighted scheme works best, though the differences with respect to the other combination schemes are small. Pooling within MIDAS model also performs well for some horizons, but the gains are smaller than for the AR-MIDAS. Bridge models confirm their good performance across horizons, especially when looking at the mean and weighted mean aggregation schemes. Pooling is instead less useful for the MF-VARs, where the gains with respect to the benchmark are small or non existent. However, also in this case the weighted combination performs best. Very small gains in using pooling within this class show up only for long horizons ($h = 8, 9$). The last part of Table 5 contains the results of now- and forecast combinations of all the models under consideration. The weighted average is once again the combination scheme that provides the best results. However, the relative performance of pooling all the models together does not behave better than all the results obtained from pooling within single classes of models: it outperforms the pooling within the two weakest classes of models (MIDAS and MF-VAR), but does not beat the average relative performance of the two best approaches, the AR-MIDAS and the bridge models.

As in the case of single indicator models, in Table 6 we provide the relative MSE of pooled (AR-)MIDAS and MF-VAR against pooled bridge models. First the single forecasts are computed and aggregated with the different weighting schemes, then the MSE of the combination is computed. The benchmark in this case is represented by the bridge equations approach.

Contrary to the individual indicator models, pooled MIDAS shows a better performance with respect to the bridge equations, when mean or median combinations are used. Except for small horizons, where evidence is mixed, for the other now- and forecasting horizons the relative MSE is below one. AR-MIDAS displays an almost uniform superior performance with respect to the bridge equations, as in the case of individual indicator models, and MF-VAR keeps underperforming compared to bridge models, even at longer horizons.

In summary, pooling mixed frequency models based on a large set of alternative indicators is promising, and MSE weighted combinations of AR-MIDAS models overall produces

improved forecasts, based on information on past forecasting performance. However, since in our case the number of forecasts to be combined is by far larger than the number of observations in our sample, we decided to consider only the easiest combination schemes, one of which is also based on the same principle of weighting the different forecasts based on the past performance of the models.

the best results.

6 Robustness analysis

We have assessed the robustness of the reported findings to a variety of modifications in the experiment design, including a longer forecast horizon (up to four quarters ahead), the role of recursive and rolling estimation, subsample analysis, several and possibly higher frequencies (adding weekly financial data to the best performing monthly indicators), and using real time data. We now summarize the main findings, with more details provided in Appendix A.

First, when extending the forecast horizon to three and four quarters ahead, the AR benchmark on average performs best. Hence, as expected, the relevance of the higher frequency indicators decreases with the forecast horizon. In terms of mixed frequency methods, the MF-VAR is slightly better than MIDAS (with and without the inclusion of the AR component), confirming the results in Section 4. As for pooling, for horizons beyond $h_m = 9$ combining the different MIDAS forecasts outperforms the AR benchmark, while pooling the MF-VAR forecasts is rather ineffective, in line with the findings in Section 5 for shorter horizons. Also in line with those results, pooling across all the individual models with a performance based weighting scheme is useful.

Second, overall in our application rolling estimation, commonly considered as a way to robustify the results in the presence of parameter instability, is not better than recursive estimation for mixed frequency models based on single indicators. Similarly, pooling forecasts obtained from recursive estimation is better than combining rolling estimation based forecasts. These findings could be due to the rather short sample size, which forces the rolling windows to be even shorter (seven years).

Third, and as another way to check for temporal stability, we have split the evaluation sample into a pre-crisis and a crisis period, where the two subsamples cover, respectively, 2003Q1 to 2006Q4 and 2007Q1 to 2009Q1. In line with other studies based on single frequency data, we find that single indicator mixed frequency models cannot on average outperform the AR benchmark prior to the crisis, at any horizon and for any class of models. However, pooling forecasts within each class still allows to obtain a better now- and forecast forecasting performance than the benchmark at least at short horizons, up to one quarter ahead. After the crisis, we find results in line to what described in Sections 4 and 5 for the entire evaluation sample. Specifically, the models that exploit the timely high frequency information have a better forecasting performance than the AR benchmark, both if we look at the average performance of the single indicator models and if we pool the forecasts within each class. Part of the better performance during the crisis is related to a major deterioration in the AR forecasts, due to the large changes in GDP growth over these quarters. In terms of good indicators both before and during the crisis, we can list the total number of unemployed people and the total unemployment rate, some business survey components (services confidence indicators and orders placed with suppliers), financial indicators (two- and five- years interest rates) and the turnover indexes. A final interesting issue related to temporal stability is to assess whether the recovery period following the business cycle trough of 2009q1 was more similar to the 2000-2006 sample or to the 2007q1-2009q1 sample. We have therefore updated the time series for GDP and a few of the best performing indicators with data up to 2010q2, and compared the performance of a few mixed frequency models with that of the benchmark AR. We find that the best single indicators up to 2009q1 remain quite good also over 2009q2-2010q2, actually their performance generally further improves, but a large part of the additional gains are due to the worsening of the AR forecasts. It should be however remembered that, while interesting, evaluations based on such short samples are subject to substantial uncertainty.

Fourth, the relative performance of the models over the whole sample, and particularly over the recessionary and expansionary phases, could be driven by few large errors, whose relevance becomes even larger when squared for the computation of the MSE. To assess whether this is the case, we have repeated the analysis for the best indicators using the mean absolute error (MAE) as an evaluation criterion instead of the MSE. It turns out that the forecasting performance of the AR model still deteriorates over the recessionary phase, and even more over 2009q2-2010q2, but the extent of the deterioration is smaller than when measured in terms of MSE. However, the best single indicators in terms of MSE yield gains also in terms of MAE with respect to the AR , though their extent is reduced. Hence, the choice of the loss function does matter, but overall there seem to remain gains from exploiting higher frequency information in mixed frequency models.

Fifth, the MIDAS approach is flexible enough to allow for the inclusion of multiple explanatory variables at different frequencies, since each indicator is modelled with its own polynomial parameterization. The other approaches could be also modified to allow for regressors at different high frequencies but the computational costs are much higher. Hence, we have combined each of the five overall best performing monthly variables (general economic situation over the next 12 months, production expectations for the months ahead, the economic sentiment indicator, the manufacturing new orders received index and the unemployment under 25 years), with a weekly financial indicator: the three-months German interest rate, the ten-years Bund and the spread between the two. Evidence about the use of weekly data turns out to be quite mixed in our application. In general there is no clear signal that the inclusion of data at higher frequency improves the forecasting performance, not even at very short horizons. However, the weekly spread between the three-months and the ten-years interest rate, the best of the three series considered, often reduces the MSE with respect to a model based on monthly and quarterly information only.

To conclude, when we repeat the evaluation using real time data for hard indicators, with both monthly and quarterly data revised, we obtain results similar and of the same magnitude to the ones obtained with pseudo real-time datasets, which do not take into account data revisions, confirming previous studies in the literature (see e.g. Diron (2008) and Schumacher and Breitung (2008)). Despite consistent data revisions, especially for the economic activity, the forecast results obtained with pseudo real-time datasets are reliable. As a general observation, the MIDAS approach seems to be the most sensible to data revisions, while the mixed-frequency VAR produces similar results with or without data revisions. The results do not change much when we look at the business surveys, which are generally not revised.

7 Large scale models

In our empirical analysis, we use a very large dataset, with many series whose aim is to capture the movements in the euro area economy (see Section 2). The information included in these time series can be summarized in few factors that represent the key economic driving element. Therefore, factor models, which have a long tradition in econometrics, are appealing from an economic point of view.

So far, we tried to combine the information coming from the different indicators by averaging forecasts based on different individual equations which contain only one indicator. Now, with the use of factor models, instead of pooling forecasts we pool the information contained in the dataset and summarize it into a few factors.

In what follows, we compare the results obtained from a standard quarterly factor model (see Stock and Watson (2002)) with the ones obtained from the recent approach proposed in the literature by Marcellino and Schumacher (2010), the Factor-MIDAS, which merges factor models based on large datasets with the forecast methods based on MIDAS. In the final subsection we evaluate alternative mixed frequency factor models.

7.1 Quarterly factor model

We employ the standard factor model proposed by Stock and Watson (2002). The h_q-step ahead forecast model is:

$$y_{t+h_q} = \beta_0 + \beta\left(L_q\right)\widehat{f}_{t_q} + \lambda\left(L_q\right)y_{t_q} + \varepsilon_{t_q+h_q}. \tag{14}$$

where $\beta\left(L_q\right)$ is an unrestricted lag polynomial of lag order P and $\lambda\left(L_q\right)$ is of order R. The estimation is conducted with a two-step procedure. First, the quarterly dataset, obtained by aggregating the monthly indicators over time, is used to estimate the factors by principal component analysis (PCA). Second, the estimators $\widehat{\beta}_0$, $\widehat{\beta}\left(L_q\right)$ and $\widehat{\lambda}\left(L_q\right)$ are obtained regressing y_{t+h_q} onto a constant, $\widehat{f}_{t_q}$ and y_{t_q} and lags. The forecast then is formed as $\widehat{y}_{t+h_q} = \widehat{\beta}_0 + \widehat{\beta}\left(L_q\right)\widehat{f}_{t_q} + \widehat{\lambda}\left(L_q\right)y_{t_q}$. In our application, we choose a quarterly model with a fixed number of factors (one) and the number of lags chosen by BIC.

7.2 Factor-MIDAS models

It is possible to augment the MIDAS regressions with the factors extracted from a large dataset to obtain a richer family of models that exploit a large high-frequency dataset to predict a low-frequency variable. While the basic MIDAS framework consists of a regression of a low-frequency variable on a set of high-frequency indicators, the Factor-MIDAS approach exploits estimated factors rather than single or small groups of economic indicators as regressors.

The Factor-MIDAS model for forecast horizon h_q quarters with $h_q = h_m/3$ is

$$y_{t_q+h_q} = y_{t_m+h_m} = \beta_0 + \beta_1 b\left(L_m;\theta\right)\widehat{f}^{(3)}_{t_m+w} + \varepsilon_{t_m+h_m}, \tag{15}$$

where $b\left(L_m;\theta\right) = \sum_{k=0}^{K} c\left(k;\theta\right) L_m^k$ and $c\left(k;\theta\right) = \frac{\exp\left(\theta_1 k+\theta_2 k^2\right)}{\sum_{k=0}^{K}\exp(\theta_1 k+\theta_2 k^2)}$. As described above in the MIDAS models, the exponential lag function provides a parsimonious way to consider monthly lags of the factors.

The model can be estimated using nonlinear least squares in a regression of y_{t_m} onto the factors $\widehat{f}^{(3)}_{t_m+w-h}$. The forecast is given by

$$y_{T_m+h_m|T_m+w} = \widehat{\beta}_0 + \widehat{\beta}_1 b\left(L_m;\widehat{\theta}\right)\widehat{f}^{(3)}_{T_m+w}. \tag{16}$$

The projection is based on the final values of estimated factors.

MIDAS regression can be extended with the addition of autoregressive dynamics as

$$y_{t_q+h_q} = y_{t_m+h_m} = \beta_0 + +\lambda y_{t_m} + \beta_1 b\left(L_m;\theta\right)\widehat{f}^{(3)}_{t_m+w} + \varepsilon_{t_m+h_m}. \tag{17}$$

The same two-step procedure described for quarterly factor models can be used also in case of mixed-frequency data. To handle the ragged-edge structure of the dataset, we follow the procedure outlined by Stock and Watson (2002), which combines the EM algorithm with PCA. Since not all observations are available, due to publication lags, the authors write the relation between observed and not fully observed variables as

$$X_i^{obs} = A_i X_i, \tag{18}$$

where X_i^{obs} contains the observations available for variable i, as a subset of X_i, and A_i is the matrix that tackles missing values. Taking this relation into account, the EM algorithm provides an estimate of the missing values (for more details, see Stock and Watson (2002) and Marcellino and Schumacher (2010)).

In the next section we provide the results only for the case of models with $r = q = 1$, where r and q are respectively the number of static and dynamic factors. The EM algorithm is used to interpolate the missing values, but to avoid convergence problems the pairwise covariances are computed over the periods when both series are available.

Marcellino and Schumacher (2010) compare this case with larger values of r and q in a similar application for forecasting German GDP growth, and find only small changes in the results.

7.3 Results

Nowcast and forecast results for the different kinds of factor models described in the Sections 7.1 and 7.2 are presented in Table 7. The numbers in the table show the relative MSE of each model to the benchmark, which once again is an AR process where the lag length is selected accordingly to the BIC criterion.

As a general result, there is evidence that the nowcasting and forecasting performance benefits a lot from the use of a large information set, summarized by factors. Factor models perform quite well up to 2 quarters ahead. They behave particularly well in nowcasting, relatively to the individual models we saw in Section 4. The standard quarterly factor model performs better than the benchmark for horizons up to one quarter, while there are no improvements for longer horizons. Factor-MIDAS models outperform the benchmark model at every horizon, and they also show a better performance compared to the quarterly factor model. This confirms the importance of taking into account the ragged-edge and mixed frequency structure of the dataset in terms of forecasting performance.

Since an AR process is supposed to be an appropriate benchmark, the inclusion of autoregressive dynamics in the MIDAS equation can further enhance the now- and forecasting performance. The results in Table 7 confirm that adding an AR term in the forecasting equation is a good option, especially at very short horizons, while the gains are very small for longer ones. This confirms what already detected in Section 4.

Comparing the results in Table 7 with those on forecast pooling shown in Table 5, it turns out that generally pooling information into a factor model provides better results than pooling forecasts from different individual models, in particular up to one quarter ahead.

This evidence is confirmed even if we consider only the last part of the evaluation sample (2007Q1 - 2009Q1), as done in Section 6.3 for the individual models. There is a clear indication of a better performance of the factor models compared to the AR benchmark, and also a better performance than forecast pooling. Contrary to that, in the first part of the sample (2003Q1 - 2006Q4) it is difficult to outperform the benchmark, as in the case of individual models. Therefore, forecast pooling performs better in this first part of the sample.

7.4 Alternative mixed frequency factor models

There are alternative factor estimation methods developed in the literature to take into account unbalanced datasets. One convenient way is proposed by Altissimo et al. (2010) for estimating the New Eurocoin indicator, when each time series is realigned in order to obtain a balanced dataset and then dynamic principal components are applied to estimate the factors (vertical alignment - principal components method, VA-PCA). Another approach is the one proposed by Doz et al. (2011), based on a complete representation of the large factor model in a state-space form. The model consists of a factor representation of the monthly time series and a VAR structure which rules the behavior of the factors (Kalman filter - principal components method, KF-PCA) .

As shown in Table 8, the differences between the three factor estimation methods are relatively small overall, confirming the evidence found for Germany by Marcellino and Schumacher (2010). Comparing the performances, the factors estimated with the EM algorithm and PCA (EM-PCA method) behave relatively better for most of the forecast horizons, therefore we provide the results only for this method. Only for nowcasting one and two months ahead the method proposed by Altissimo et al. (2010) has a better performance than the method proposed by Stock and Watson (2002)[5].

8 Results for euro area GDP components

From a policy perspective, it is generally important to have timely now- and forecasts not only of GDP but also of its components from the demand and supply sides. Hence, we have repeated the analysis of the forecasting performance of the mixed frequency models for the euro area GDP components, and we now summarize the main findings, with additional details provided in Appendix B.

Considering a decomposition from the output side, we follow the NACE classification of the GDP and obtain six branches of activity: agriculture, hunting, forestry and fishing; industry, excluding construction; construction; trade, hotels and restaurants, transport and communication services; financial services and business activities; other services. From the expenditure side, instead, we obtain five components: household final consumption; government final consumption; gross fixed capital formation, imports and exports.

Perhaps not surprisingly given the heterogeneity in the time series properties of the GDP components, the evidence on the relevance of the mixed frequency models is quite mixed. More specifically, starting with the GDP disaggregation from the output side, the mixed-frequency approaches outperform the benchmark for those components for which many monthly indicators are available, as in the case of the industry sector, trade and financial services. For agriculture, availability of monthly indicators is critical. The same holds for the last branch that includes a variety of economic activities (public administration and defence, compulsory social security; education; health and social work; other community, social and personal service activities; private households with employed persons) for which it is not easy to find reliable and timely monthly indicators of value added.

[5] For the details on the specification of these alternative factor models we refer to Marcellino and Schumacher (2010).

Looking now at the GDP components from the expenditure side, for the government final consumption it is very hard to find monthly indicators, so that it is very difficult to beat the benchmark. However, for all the other components, bridge equations and AR-MIDAS have (on average across indicators) a better performance than the benchmark, with the MF-VAR approach ranked third, since since it often performs better than the benchmark, especially for longer horizons.

Now- and forecast pooling is helpful also for the components, but only for those with a sufficiently large set of indicators available, namely total industry and trade and financial services from the output side, and household final consumption, gross fixed capital formation and external balance from the expenditure side. Forecast combinations of AR-MIDAS models perform pretty well, outperforming the benchmark at several horizons. Also combinations of simple MIDAS and bridge equations allow for gains at some horizons.

To conclude, about the behavior of large scale factor models we can reach similar conclusions as for the aggregate GDP growth. Evidence is in favour of the use of factor models to predict the quarterly growth of each component for which the dataset contains enough useful information. There are significant gains especially at very short horizons: generally, exploiting the unbalanced structure of the dataset improves the performance, and the inclusion of an AR component reduces the MSE, even though not systematically. As for the case of GDP growth discussed in Section 7, also for the components of GDP the factor models seem to better perform relative to the benchmark than forecast pooling. This is true especially for nowcasts and short-term forecasts. However, when the forecast horizon increases, the outperformance of the factor models is no longer evident, and in many cases forecast pooling is better. Finally, for these long horizons, both methods of summarizing information (factors and forecast pooling) generally fail in beating the AR benchmark.

9 Conclusions

This paper extends the analysis presented in Kuzin, Marcellino, Schumacher (2011), considering a dataset of more than 150 monthly indicators to now- and forecast quarterly Euro Area GDP growth, and comparing different approaches which take into account the mixed frequency and ragged edge structure of the dataset. To start with, we compared the bridge model, the MIDAS model, with its extension incorporating an AR component, and the MF-VAR. The three approaches display some marked differences: while the bridge equations approach is a pure statistical model, where regressors are chosen by their timeliness more than by any specific economic reason, the other models presented in this paper are more sophisticated and exploit different ways to deal with an unbalanced dataset. Just as an example, while MIDAS is a single-equation approach and a direct multi-step forecast tool, MF-VAR jointly explains the indicator and GDP growth, and it is a recursive instrument to produce multi-step forecasts. Moreover, while with bridge equations and MIDAS models we obtain a monthly update of the quarterly GDP growth, with the state-space approach we can have an estimate of the monthly missing values of the GDP.

These approaches are therefore too different to have a ranking based only on theory. Hence it is preferable to compare them in empirical applications. The main results we obtained from our exercise hint at a better performance of MIDAS models, especially for the short horizons and when incorporating an AR component in the MIDAS model. Bridge models,which are less sophisticated than the other approaches, have overall a good performance. Finally, overall MF-VAR is the least promising mixed-frequency approach, at least for very short horizons.

Pooling within each class of models results to be a good strategy to improve the performance: the MSE of the forecast combinations is smaller than the MSE of most of the individual models at every horizon. Comparing the different performance of forecast combination within each class, AR-MIDAS appears to be the best strategy.

An even better performance is obtained with the factor models, confirming that pooling information from a large number of series is useful in short-term forecasting and reduces the MSE. In particular, we looked at the factors estimated with the EM algorithm and PCA, which behave relatively better than the other models proposed in the literature, and we included these factors in a MIDAS framework. The factor-MIDAS with the inclusion of the AR component is the best in terms of relative MSE.

We assessed the robustness of the findings about Euro area GDP growth extending the forecast horizons up to four quarters ahead. While individual models hardly beat the benchmark, pooling the forecasts still allows for some gains, especially in the case of AR-MIDAS approach. We also compared recursive and rolling estimation, checking for temporal stability, but the results for individual models are not satisfactory when estimated with a rolling technique, possibly because the size of the rolling window and of our sample is still too small. Splitting the sample onto 2003-2007 and 2008-2009 evidences the difficulties in beating the benchmark before the crisis, while the mixed-frequency approaches improve their performance during quarters of dramatic drop in GDP growth. Moreover, exploiting weekly financial data in the MIDAS approach does not appear to contribute at improving the performance of the model significantly, though the spread is sometimes useful.

Since the analysis was conducted on a pseudo real-time dataset, we repeated the evaluation for a small number of best performing indicators in a genuine real-time context to check for the role of data revisions. We obtained similar ranking of the methods, and same magnitude of gains, as with pseudo real-time data. Despite consistent data revisions, especially for the economic activity measure, the findings obtained with pseudo real-time datasets are therefore reliable.

As a final contribution, we extended the analysis to the single components of the GDP, from the output side and from the expenditure side. The findings are in line with those obtained for the aggregate measure of economic activity, at least for those components for which timely monthly indicators are available.

References

[1] Aiolfi, M. and A. Timmermann (2006): "Persistence in Forecasting Performance and conditional combination strategies", Journal of Econometrics, 135, 31-53.

[2] Altissimo, F. R. Cristadoro, M. Forni, M. Lippi, G. Veronese (2010): "New Eurocoin: Tracking Economic Growth in Real Time," The Review of Economics and Statistics, 92(4), 1024-1034.

[3] Baffigi, A., R. Golinelli, and G. Parigi (2004): "Bridge models to forecast the Euro area GDP", International Journal of Forecasting, 20(3), 447-460.

[4] Breitung, J. and C. Schumacher (2008): "Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data", International Journal of Forecasting, 24(3), 386-398.

[5] Clements, M. P., and A. B. Galvao (2008): "Macroeconomic forecasting with mixed-frequency data: Forecasting US output growth", Journal of Business and Economic Statistics, Journal of Business and Economic Statistics 26, 546-554.

[6] Clements, M. P., and A. B. Galvao (2012): "Real-time Forecasting of Inflation and Output Growth with Autoregressive Models in the Presence of Data Revisions", Journal of Applied Econometrics, forthcoming.

[7] Diron, M. (2008): "Short-term forecasts of Euro Area real GDP growth. An assessment of real-time performance based on vintage data", Journal of Forecasting, 27(5), 371-390.

[8] Doz, C., D. Giannone, and L. Reichlin (2011): "A Two-step estimator for large approximate dynamic factor models based on Kalman filtering", Journal of Econometrics, 164(1), 188-205.

[9] Ghysels, E., P. Santa-Clara, and R. Valkanov (2004): "The MIDAS touch: MIxed DAta Sampling regression models", mimeo, Chapel Hill, N.C.

[10] Ghysels, E., P. Santa-Clara, and R. Valkanov (2006): "Predicting volatility: Getting the most out of return data sampled at different frequencies", Journal of Econometrics, 131(1-2), 59-95.

[11] Giannone, D., L. Reichlin, and D. Small (2008): "Nowcasting GDP and inflation: The real-time informational content of macroeconomic data releases", Journal of Monetary Economics, 55, 665-676.

[12]

[13] Kuzin, V., M. Marcellino, and C. Schumacher (2011): "MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro Area", International Journal of Forecasting, 27, 529-542.

[14] Kuzin, V., M. Marcellino and C. Schumacher (2012): "Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries", Journal of Applied Econometrics, forthcoming.

[15] Marcellino, M., and C. Schumacher (2010): "Factor-MIDAS for now- and forecasting with ragged-edge data: A model comparison for German GDP", Oxford Bulletin of Economics and Statistics, 72, 518-550.

[16] Mariano, R., and Y. Murasawa (2003): "A new coincident index of business cycles based on monthly and quarterly series", Journal of Applied Econometrics, 18(4), 427-443.

[17] Mariano, R., and Y. Murasawa (2004): "Constructing a coincident index of business cycles without assuming a one-factor model", Discussion Paper 2004-6, College of Economics, Osaka Prefecture University.

[18] Stock, J., and M. Watson (2001): "A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series", in R.F. Engle and H. White, eds., Festschrift in Honour of Clive Granger. Cambridge University Press, 1-44.

[19] Stock, J., and M. Watson (2002): "Macroeconomic forecasting using diffusion indexes", Journal of Business and Economic Statistics, 20, 147-162.

[20] Wallis, K. (1986): "Forecasting with an econometric model: The 'ragged-edge' problem", Journal of Forecasting, 5, 1-13.

[21] Timmermann, A. (2006): "Forecast combinations", Handbook of Economic Forecasting, 2006.

[22] Zadrozny, P. A. (1988): "Gaussian-Likelihood of countinuous-time ARMAX models when data are stocks and flows at different frequencies", Econometric Theory, 4(1), 108-124.

Table 1: Main publication lags

Main releases	Publishing lag	Frequency
HICP	1 month	monthly
PPI	2 months	monthly
Industrial production	2 months	monthly
Industrial new orders	2 months	monthly
Turnover index	2 months	monthly
Hours worked	2 months	monthly
Car registrations	2 months	monthly
Retail trade	2 months	monthly
Construction output	2 months	monthly
Business survey	current month	monthly
Business climate indicator	current month	monthly
Consumer survey	current month	monthly
Money supply	1 month	monthly
Exchange rates (average)	1 month	monthly
Interest rates (average)	1 month	monthly
Stock exhange indexes (average)	1 month	monthly
Unemployment	2 months	monthly
GDP: disaggregation of sectorial value added	1 quarter	quarterly
GDP: disaggregation from expenditure side	1 quarter	quarterly

Notes: The publishing lags correspond to the number of missing observations at the end of the sample at the downloaded date.

Table 2: Average relative MSE performance of different classes of mixed-frequency models against AR benchmark

	horizon (h_m)								
model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
bridge	1.73	0.95	0.99	1.03	0.96	0.98	1.01	**0.99**	**0.98**
midas	1.69	0.96	1.03	1.06	0.96	0.99	0.99	0.99	1.00
ar-midas	**0.88**	**0.82**	**0.86**	**0.88**	**0.94**	**0.95**	**0.96**	1.02	1.02
mf-var	1.37	0.99	1.04	1.08	1.01	1.02	1.03	1.00	1.00
absolute values									
MSE bm	0.28	0.62	0.62	0.62	0.71	0.71	0.71	0.74	0.74
variance GDP	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63

Notes: The entries in the table are obtained as follows: first, estimate recursively every individual model and compute the relative MSE with respect to the benchmark; then, take the average all across the indicators of the relative MSE within a model class. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the best relative MSE performance for each horizon. For completeness, at the bottom of the table we report the absolute value of the MSE of the benchmark, and, as a term of comparison, the variance of the GDP growth.

Table 3: Average relative MSE performance of (AR-)MIDAS and MF-VAR against bridge

		horizon (h_m)								
	model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
mean	midas	1.02	1.05	1.07	1.05	1.01	1.01	**0.99**	1.00	1.02
	ar-midas	**0.57**	**0.90**	**0.91**	**0.88**	**0.99**	**0.97**	**0.95**	1.03	1.04
	mf-var	**0.86**	1.07	1.07	1.06	1.06	1.04	1.02	1.02	1.02
median	midas	**0.98**	1.01	1.02	1.01	**0.99**	1.01	**0.99**	1.00	1.01
	ar-midas	**0.54**	**0.85**	**0.87**	**0.83**	**0.96**	**0.96**	**0.95**	1.01	1.03
	mf-var	**0.80**	1.01	1.02	1.01	1.02	1.01	1.00	1.00	1.00

Notes: The entries in the table are obtained as follows: first, estimate recursively every individual model and compute the relative MSE with respect to the benchmark; then, take the average and the median all across the indicators of the relative MSE within a model class. The benchmark is the corresponding bridge model. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the classes of models which outperform the bridge models for each horizon.

Table 4: Relative MSE performance of the best mixed-frequency models with different indicators against AR benchmark

		horizon (h_m)								
monthly indic.	**model**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
General economic situation over the next 12 months	midas	0.70	0.65	0.67	0.67	0.86	0.88	0.97	0.96	0.99
	ar-midas	0.93	0.71	0.75	0.83	0.88	0.87	0.94	0.94	0.94
	mf-var	0.98	0.75	0.71	0.80	0.87	0.89	1.00	1.00	0.95
Production expectations for the months ahead	bridge	0.63	0.37	0.45	0.61	0.63	0.83	0.97	0.96	0.98
	midas	0.51	0.44	0.55	0.61	0.82	0.93	0.92	1.07	1.11
	ar-midas	0.66	0.45	0.54	0.66	0.80	0.90	0.97	0.97	0.98
	mf-var	0.57	0.44	0.59	0.71	0.81	0.94	1.00	1.07	1.06
Economic sentiment indicator	bridge	0.67	0.40	0.38	0.46	0.58	0.63	0.92	0.92	0.89
	midas	0.59	0.42	0.47	0.57	0.82	0.86	0.93	0.93	0.94
	ar-midas	0.60	0.54	0.60	0.70	0.81	0.88	0.98	0.97	1.06
	mf-var	0.77	0.57	0.63	0.77	0.81	0.88	1.00	1.02	1.03
New orders received index - Manufacturing	bridge	0.45	0.32	0.49	0.72	0.84	0.92	1.01	1.01	0.99
	midas	0.44	0.42	0.64	0.60	0.84	0.92	0.94	1.02	0.99
	ar-midas	0.49	0.41	0.58	0.66	0.84	0.96	0.97	1.00	1.00
Unemployment - Under 25 years	bridge	0.69	0.72	0.81	0.89	0.92	0.95	0.99	0.98	0.98
	midas	0.81	0.74	0.80	0.84	0.76	0.76	0.96	0.90	0.89
	ar-midas	0.94	0.74	0.78	0.84	0.75	0.76	0.93	0.91	0.90
	mf-var	0.77	0.79	0.87	0.95	0.95	0.98	1.01	1.00	1.00

Notes: The entries in the table are obtained as follows: first, estimate recursively every individual model, then compute the relative MSE with respect to the benchmark. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1.

Table 5: Relative MSE performance of model pooling within a given class of models against AR benchmark

		horizon (h_m)								
model		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
bridge	mean	1.53	0.89	0.94	0.99	0.94	0.97	0.99	0.98	0.97
	weighted mean	1.09	**0.73**	**0.82**	0.89	**0.88**	0.92	0.95	0.93	0.93
	median	1.67	0.98	1.03	1.07	1.00	1.03	1.03	1.01	1.00
midas	mean	1.46	0.90	0.98	1.02	0.93	0.96	0.96	0.95	0.95
	weighted mean	1.27	0.83	0.96	0.98	0.90	0.92	0.91	**0.92**	**0.92**
	median	1.63	0.94	1.04	1.07	0.94	1.00	1.00	0.98	0.98
ar-midas	mean	0.80	0.78	0.84	0.86	0.91	0.92	0.92	0.95	0.96
	weighted mean	**0.79**	0.76	0.83	**0.85**	**0.88**	**0.89**	**0.88**	**0.92**	0.93
	median	0.86	0.81	0.86	0.88	0.92	0.95	0.95	0.99	0.99
mf-var	mean	1.19	0.94	1.00	1.05	1.00	1.01	1.02	0.99	1.00
	weighted mean	1.09	0.89	0.98	1.04	0.99	1.00	1.02	0.99	0.99
	median	1.23	1.03	1.09	1.12	1.03	1.03	1.04	1.01	1.01
all	mean	1.20	0.87	0.94	0.98	0.94	0.96	0.97	0.97	0.97
	weighted mean	1.08	0.82	0.90	0.95	0.92	0.93	0.94	0.94	0.95
	median	1.19	0.91	0.95	1.00	0.97	1.01	1.01	1.00	1.00

Notes: The entries in the first part of the table are obtained as follows: first, forecasts from single indicator models are computed, and means, medians and weighted means of all the forecasts within a single model are obtained; second, the MSE of these three different forecast combinations is calculated and divided by the MSE of the benchmark. In the second part the entries are obtained in the same way, but the different combinations are obtained across all individual models of different classes. The estimation is conducted recursively. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the best relative MSE performance for each horizon.

Table 6: Relative MSE performance of model pooling within (AR-)MIDAS and MF-VAR models against bridge

		horizon (h_m)								
model		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
midas	mean	**0.95**	1.01	1.04	1.03	**0.99**	**0.99**	**0.97**	**0.97**	**0.98**
	weighted mean	1.16	1.13	1.16	1.10	1.02	1.00	**0.96**	**0.98**	**0.99**
	median	**0.97**	**0.96**	1.01	1.00	**0.94**	**0.97**	**0.97**	**0.97**	**0.98**
ar-midas	mean	**0.52**	**0.87**	**0.89**	**0.87**	**0.97**	**0.95**	**0.93**	**0.97**	**0.99**
	weighted mean	**0.72**	1.04	1.01	**0.96**	1.00	**0.96**	**0.93**	**0.98**	1.00
	median	**0.51**	**0.83**	**0.84**	**0.83**	**0.92**	**0.93**	**0.92**	**0.98**	**0.99**
mf-var	mean	**0.78**	1.05	1.06	1.06	1.06	1.04	1.03	1.01	1.03
	weighted mean	1.00	1.22	1.19	1.17	1.12	1.08	1.08	1.06	1.07
	median	**0.74**	1.05	1.06	1.05	1.02	1.00	1.01	1.00	1.01

Notes: The entries in the table are obtained as follows: first, forecasts from single indicator models are computed, and means, medians and weighted means of all the forecasts within a single model are obtained; second, the MSE of these three different forecast combinations is calculated and divided by the MSE of the benchmark. The estimation is conducted recursively. The benchmark is the bridge equations approach. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the classes of models which outperform the bridge models for each horizon.

Table 7: Relative MSE performance of different classes of factor models against AR benchmark

	horizon (h_m)								
model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Factor-MIDAS (basic)	0.79	0.61	0.64	0.76	0.87	0.88	0.95	0.98	0.97
Factor-MIDAS (ar)	**0.68**	**0.57**	**0.59**	**0.66**	**0.85**	**0.87**	**0.93**	**0.97**	**0.96**
Quarterly factor model	0.95	0.83	0.83	0.83	0.93	0.93	**0.93**	1.03	1.03

Notes: The entries in the table are obtained as follows: first, estimate recursively every factor model, then compute the relative MSE with respect to the benchmark. The factors are estimated with the EM algorithm together with PCA as in Stock and Watson (2002). The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1.

Table 8: Relative MSE performance of different classes of factor models against AR benchmark, comparison of nowcast and forecast results for different factor estimation methods for r = 1

		horizon (h_m)								
model	**estimation method**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Factor-MIDAS (basic)	EM-PCA	0.79	0.61	**0.64**	**0.76**	0.87	**0.88**	**0.95**	**0.98**	**0.97**
	VA-DPCA	**0.57**	**0.46**	0.78	0.88	**0.83**	0.97	1.00	1.07	1.15
	KF-PCA	0.78	0.67	0.78	0.91	0.94	0.97	1.03	1.04	1.04
Factor-MIDAS (ar)	EM-PCA	0.68	0.57	**0.59**	**0.66**	0.85	**0.87**	**0.93**	**0.97**	**0.96**
	VA-DPCA	**0.38**	**0.44**	0.74	0.75	**0.83**	0.96	0.97	1.11	1.24
	KF-PCA	0.62	0.61	0.68	0.75	0.91	0.94	0.98	1.04	1.07

Notes: In the estimation method column, EM-PCA refers to the EM algorithm together with PCA as in Stock and Watson (2002), VA-DPCA is the vertical realignment and dynamic PCA used in Altissimo et al. (2006), and KFS-PCA is the Kalman smoother of state-space factors according to Doz et al. (2006). The entries in the table are obtained as follows: first, estimate recursively every factor model, then compute the relative MSE with respect to the benchmark. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the best relative MSE performance for each horizon.

10 Appendix

A Robustness analysis

This Appendix assesses the robustness of the reported findings to a variety of modifications in the experiment design. In the first subsection, we extend the forecast horizon up to four quarters ahead. In the second subsection, we compare recursive and rolling estimation, checking for temporal stability. In the third subsection, we conduct a subsample analysis and look at the results for different periods. In the fourth subsection, we consider MIDAS models with several frequencies, adding weekly financial data to the best performing monthly indicators. In the final subsection, we repeat the analysis for the best indicators in a genuine real-time framework, which takes into account data revisions.

A.1 Extending the forecast horizon

In Sections 4 and 5, we presented the results for nowcasts and forecasts of the GDP growth up to two quarters ahead. Here, we extend the forecast horizon and look at the results for the forecasts three and four quarters ahead. We report here only the main results, and discuss the general findings.

Looking at individual models, the AR benchmark results to be the best model for horizons $h_m = 10$ up to $h_m = 15$. Table 9 reports the average relative MSE performance for now- and forecasting quarterly GDP growth at different horizons for different classes of models, against the AR benchmark. All the approaches, bridge, MIDAS and MF-VAR, show an average performance worse than the benchmark for three and four quarters ahead forecasts. Focusing on the relative performance of the different methods, it seems that the MF-VAR is performing slightly better than the MIDAS approach, with and without the inclusion of the AR component, confirming that the MF-VAR works better than MIDAS at longer horizons, as already highlighted in Section 4.

As for the horizons up to $h_m = 9$, we also consider three different forecast combination schemes: the mean, the median and a weighted mean that lets the combination weights be inversely proportional to the MSE of the previous four-quarter performance of the model.

In the first part of Table 10, we provide the relative MSE performance of model pooling within a given class of models against the benchmark.

Combining the different forecasts, MIDAS models outperform the benchmark. In this case, the basic model without the autoregressive component seems to work better than the correspondent model augmented by the AR dynamics. MF-VAR approach does not show any ability to beat the benchmark. Evidence is more mixed in the case of bridge models, with no clear evidence of these models outperforming the benchmark. Confirming the findings in Section 5 for shorter horizons in the case of forecast pooling, MIDAS shows a better performance than MF-VAR. In the second part of Table 10 we look at the results of pooling across all the individual models we have: these results are quite satisfactory, especially when the weighted scheme is used, as we already noticed in Section 5 for shorter horizons.

A.2 Temporal stability

So far, we first recursively estimated and then now- and forecasted Euro area GDP growth rate. Now we compare these results with those obtained by rolling estimation, which can produce better results in the presence of parameter instability. Rolling estimation is more robust but it may increase the variance of the parameters estimates and therefore the mean square forecast error, in particular in a rather short sample as ours.

In Table 11, we show the average relative MSE performance for now- and forecasting quarterly GDP growth against the AR benchmark, comparing recursive and rolling estimation. Since both the individual models and the benchmark can be estimated with rolling and recursive method, we have four types of ratios: the relative MSE performance when both individual models and the benchmark are estimated recursively (these is what is presented in Section 4, so we do not report the results again), when both individual models and the benchmark are estimated with a rolling method, when the individual models are estimated with a rolling method and the benchmark recursively and viceversa. The size of the rolling window is equivalent to seven years.

If we look at panel A and B in Table 11, we see that when we do a rolling estimation of the mixed-frequency models, independently on how the benchmark is estimated, we see fewer gains compared to the results in Section 5. More specifically, with a rolling estimation the MIDAS class does not outperform the benchmark anymore, while the MF-VAR slightly improves its performance at longer horizons. However, when the individual models are estimated recursively, they beat not only the AR process estimated recursively but also the same process estimated with a rolling technique.

In Table 12, we provide the relative MSE performance of model pooling within a class against the benchmark. As in the case of individual models we combine recursive and rolling estimations of the mixed-frequency models and the benchmark. We report only the results obtained with the median, to save space. Here, we see benefits in pooling compared to the AR benchmark, either in the case of pooling obtained after recursive or rolling estimation. However the gains are bigger when the mixed-frequency models are estimated recursively. Generally, the results still hint a better performance of the AR-MIDAS compared to the simple MIDAS without autoregressive component, and good results from pooling bridge equations. Moreover, differently from the case of recursive estimation, with a rolling estimation good results are also reached in the case of MF-VAR, especially for longer horizons. Pooling across all the models we have does not show significant gains independently on the recursive or rolling estimation method used.

To sum up, we looked at recursive and at rolling estimation of the benchmark and the mixed-frequency models. We can restrict our attention to the recursive AR benchmark because it gives a smaller MSE than the corresponding one obtained with rolling estimation (for all forecast horizons except for $h_m = 1$). Whenever the individual models are estimated recursively, we find bigger gains in terms of relative MSE than in the case of rolling estimation, especially within the MIDAS class. This can be due to the fact that our sample is still too short and we cannot have a reasonable size for the rolling window.

A.3 Subsample analysis

Another way to check for temporal stability is to split the evaluation sample and look at the results for different periods. In our analysis we consider two periods, the first from 2003Q1 to 2006Q4 and the second from 2007Q1 to 2009Q1. Here we summarize the main results.

As highlighted in different papers, many sophisticated forecast models cannot outperform a naive benchmark in periods going from the beginning of the 2000s up to the beginning of the financial crisis in 2007. This is exactly what we find with our data: the single indicator models cannot on average outperform the AR benchmark, at any horizon and for any class of models. However, pooling forecasts within each class still allows to obtain a better now- and forecast forecasting performance than the benchmark at least at short horizons, up to one quarter ahead.

Looking now at the results for the second subsample, we find results in line to what described in Sections 4 and 5 for the entire evaluation sample. The models that exploit the timely high frequency information have a better forecasting performance than the AR benchmark, both if we look at the average performance of the single indicator models and if we pool the forecasts within each class. This highlights how exploiting monthly information is useful especially in periods of crisis, as in the quarters 2008Q4 and 2009Q1 where the mixed-frequency models performed much better than an AR process. Part of the better performance is related to a major deterioration in the AR forecasts, due to the large changes in GDP growth over these quarters.

Moreover, if we look at the best performing indicators, we see that in the first subsample fewer variables contain enough information to outperform the benchmark. However, the indicators which have a MSE smaller than the benchmark up to $h_m = 9$ in the first part of the sample usually have a good performance also in the second part. Among these outperforming indicators, we find the total number of unemployed people and the total unemployment rate, some business survey components (services confidence indicators and orders placed with suppliers), financial indicators (two- and five- years interest rates) and the turnover indexes. As a remark, those indicators who are among the best in the first subsample up to $h_m = 9$ stay among the outperforming indicators even in the second period but only for shorter horizons, generally up to $h_m = 6$.

Another interesting issue to be considered is whether the recovery period following the business cycle trough of 2009q1 was more similar to the 2000-2006 sample or to the 2007q1-2009q1 sample. We have therefore updated the time series for GDP and a few of the best performing indicators with data up to 2010q2, and compared the performance of a few mixed frequency models with that of the benchmark AR. The results are reported in Table 13 in terms of relative MSE of each model/indicator versus that of the AR, while the last row of the table shows the MSE of the AR. For simplicity we focus on 1- to 3-step ahead forecasts only, results for longer horizons are available upon request.

Four main findings are worth noting. First, the results for the samples 2003q1-2009q1, 2003q1-2006q4 and 2007q1-2009q1 are very similar to those obtained with the 2009 data vintage, suggesting that recent data revisions were not so relevant for these indicators, more on this topic in subsection 6.5. Second, the extent of the mentioned deterioration in the AR performance during the crisis period is evident from the table. The AR MSE increases from about 0.06 during 2003-2006 to 0.64 for h=1 and 1.56 for h=2,3 during 2007-2009. The performance further deteriorates at the beginning of the recovery phase, with values reaching 1.38 for h=1 and 2.40 for h=2,3. Third, the best single indicators up to 2009q1 remain quite good also over 2009q2-2010q2, actually their performance generally further improves, but a large part of the additional gains are due to the worsening of the AR forecasts. Fourth, the ranking of the alternative mixed frequency models remains unclear, but in general bridge and mixed frequency VARs appear to behave better than over the previous periods. It should be however remembered that, while interesting, evaluations based on such short samples are subject to substantial uncertainty.

Finally, the relative performance of the models over the whole sample, and particularly over the recessionary and expansionary phases, could be driven by few large errors, whose relevance becomes even larger when squared for the computation of the MSE. To assess whether this is the case, in Table 14 we have repeated the analysis presented in the Table 13 but using the mean absolute error (MAE) as an evaluation criterion instead of the MSE. It turns out that the forecasting performance of the AR model still deteriorates over the recessionary phase, and even more over 2009q2-2010q2, but the extent of the deterioration is smaller than when measured in terms of MSE. Moreover, the best single indicators in terms of MSE yield gains also in terms of MAE with respect to the AR , though their extent is reduced. Hence, the choice of the loss function does matter, but overall there seem to remain gains from exploiting higher frequency information in mixed frequency models.

A.4 Several frequencies

Financial variables, as interest rates and term spreads, are available even at frequencies higher than monthly. The MIDAS approach is flexible enough to allow for the inclusion of multiple explanatory variables at different frequencies, since each indicator is modelled with its own polynomial parameterization. The other approaches could be also modified to allow for regressors at different high frequencies but the computational costs are much higher.

In the case of monthly and weekly data, the MIDAS framework is extended as follows:

$$y_{t_{m_1}} = \beta_0 + \beta_1 b\left(L_{m_1};\theta_1\right) x^{(m_1)}_{1,t_{m_1}+w-h_{m_1}} + \beta_2 b\left(L_{m_2};\theta_2\right) x^{(m_2)}_{2,t_{m_2}+w-h_{m_2}} + \varepsilon_{t_{m_1}}, \tag{19}$$

where m_1 represents the monthly frequency and m_2 the weekly frequency.

As a robustness check, we take the five best performing monthly variables: three business survey components, general economic situation over the next 12 months, production expectations for the months ahead and the economic sentiment indicator, the manufacturing new orders received index and the unemployment under 25 years. These five variables are the only ones which beat the AR benchmark up to at least horizon $h_m = 6$, with both MIDAS and AR-MIDAS approaches (as discussed in Section 4). To obtain a model with monthly and weekly data, we add a weekly financial variable to each of these 5 monthly variables; we consider three weekly series: the three-months German interest rate, the ten-years Bund and the spread between the two. We end up therefore with 15 different models, each of them with a monthly and a weekly variable.

In order to compare these models with several frequencies with the ones with only monthly explanatory variables, we compute the forecasts only 1 to 9 months ahead, even though it is theoretically possible to compute a new forecast every week. Moreover, MIDAS models consider a fixed ratio between the releases of high-frequency and low-frequency data. While this is evident in the case of monthly and quarterly data, where we always have 3 months in a quarter, it is not obvious when we consider also weekly data, since the number of weeks varies during months and quarters. In our application, we consider 12 weeks per quarter, and 4 weeks per months, skipping the first week of each month in which there are five of them.

In Table 15, we show the ratio of the MSE of the model with monthly and weekly variable relative to the MSE of the model with only the correspondent monthly indicator. A ratio smaller than one indicates that the introduction of weekly data improves the forecasting performance. All the models are recursively estimated.

Evidence about the use of weekly data is quite mixed and there is no clear signal that the inclusion of data at higher frequency improves the forecasting performance, not even at very short horizons. More in details, weekly data appear to have some information content more in the MIDAS models than in the AR-MIDAS, where the introduction of the autoregressive component improves the forecasting performance of the models. Moreover, looking at the different weekly data considered, the spread between the three-months and the ten-years interest rate is the best of the three series considered, and it is able to reduce the MSE in almost every case. The performance of the interest rates, on the other hand, is not clear and it depends on which monthly indicator is used.

A.5 The role of real-time data

In the analysis conducted so far, we used a pseudo real-time dataset, taking into account the pattern of missing values due to publication lags, but using only the final vintage of data available at the downloaded date. In this way, we do not consider data revisions, which can be substantial in case of real activity. In what follows, we consider only the best performing indicators, as selected in Section 4. Among them, we face two types of data: the business surveys, which are not revised (or revised very few times and only marginally), and the "hard data", unemployment rate and industrial production (durable consumer goods components), which can be substantially revised. Therefore, in the following analysis we will show the results of these two group separately. In fact, while in considering unemployment rate and industrial production we have data revisions in both the explanatory variables and in the GDP growth, in considering business survey components we have revisions only in the dependent variable.

It is not straightforward to construct a real-time dataset for industrial production components, since a new version of the European standard classification of production activities has been introduced by Eurostat in January 2009, so it is hard to match the data before and after the reclassification. Since the manufacturing new orders received index is available only after January 2009, we consider another component of industrial production, the durable consumer goods production, in order to have vintages for the entire evaluation sample. Moreover, in the course of 2005 and 2006 national accounts data displayed major revisions, due to the introduction of chain-linking of real activity series. Therefore the vintages of GDP are at constant prices up to the end of 2005 and chain-linked afterward. Finally, the same concept of Euro area has evolved over time, changing the country coverage. Two different concepts of Euro area country composition are employed in general: the fixed composition, which uses the same group of countries throughout all periods, and the changing composition, which follows the evolution of the euro area composition through time. In practice little differences are expected due to the small size in terms of GDP of the new member states. In order to conduct an analysis in real-time and to have the largest available number of vintages, we prefer to use the changing composition of Euro area, since for the fixed composition we have fewer available data: vintages of Euro area at 16 countries are available only after the enlargement to Slovakia in 2009, while for the other fixed compositions, e.g. Euro area at 12 countries, we do not have the most recent vintages after the reclassification of the industry activity.

We can have an idea of the magnitude of the revisions looking at Table 16, which summarizes the main statistics on the revisions.

The statistics reported in Table 16 refer to revisions of the series compared to the last vintage we have available at the downloaded date. Looking at the minimum and maximum, we see that the revisions can be substantial, especially in the monthly series (change in the unemployment rate and industrial production growth). On the contrary, the revisions do not appear to be so wide in the GDP growth rate. However, on average the revisions are zero for both the GDP growth, and for the two monthly series considered.

Some comments are needed also for the business survey series. Even though, as mentioned earlier, these series are not revised (or are very mildly revised due to the changing country composition of the Euro area), we can observe a change in the timing of the releases of these series: while at the beginning of our evaluation sample business surveys were released at the beginning of the month after the month of interest, recently the releases of the business surveys are available already in the last week of the month they refer to, therefore we can observe a change in the publication lags.

In our empirical analysis, we replicate the same forecasting exercise described in Section 3, with the four mixed-frequency methods (bridge models, MIDAS with and without AR component, MF-VAR) plus the quarterly autoregressive model, using real-time data. The forecasts are evaluated by comparing them to the final vintage of GDP available[6].

When we look at hard data, with both monthly and quarterly data revised, we obtain results similar and of the same magnitude to the ones obtained with pseudo real-time datasets, which do not take into account data revisions, confirming previous studies in the literature (see e.g. Diron (2008) and Schumacher and Breitung (2008)). Despite consistent data revisions, especially for the economic activity, the forecast results obtained with pseudo real-time datasets are reliable. There is no clear evidence that the use of real-time data reduces the MSE. In Table 17 we compare the relative MSE obtained with a real-time dataset and the relative MSE obtained with a pseudo real-time dataset. Data revisions have no clear impact on the forecasting accuracy of the different models. The relative MSEs calculated on the two different datasets are similar in most of the cases, but the results are not uniform across the models and the horizons. As a general observation, the MIDAS approach seems to be the most sensible to data revisions, while the mixed-frequency VAR produces similar results with or without data revisions.

The results do not change much when we look at the business surveys, which are generally not revised. However, we note that the performance of the individual mixed-frequency models relative to the AR benchmark is slightly better when we use a pseudo real-time dataset than when we conduct the analysis in real-time. This can be due to the fact that the publication lag changed during the evaluation sample, as mentioned earlier in the paragraph. In constructing the pseudo real-time dataset, in fact, we impose the publication lags we see at the downloaded date in each recursion. Therefore, in the first part of the evaluation sample the pseudo real-time dataset accounts for one observation more than what we find in the genuine real-time dataset. This is reflected in the results provided in Table 18, where the relative MSE is slightly bigger, especially for very short

[6] In this exercise we use the data from the latest-available vintage. While in their paper, Clements and Galvao (2012) argue that the use of lightly-revised data helps improving the forecasting performance of the models, we prefer to use the latest-available vintages of data for two reasons. First, as explained in this section, there are various complications in contructing a real-time dataset for Euro area. Second, we compare the forecasts to the final vintage of GDP available, therefore the forecasting target is not first-released data. Even in the evidence found by Clements and Galvao (2012), the use of latest-available or lightly-revised data does not influence the results when forecasting post-revision data.

horizons. However, since all in all the results are not so different depending on the dataset used, GDP growth revisions have no clear impact on forecasting performance, and the best performing indicators keep outperforming the benchmark.

Table 9: Average relative MSE performance of different classes of mixed-frequency models against AR benchmark

	horizon (h_m)					
model	**10**	**11**	**12**	**13**	**14**	**15**
bridge	0.99	**1.00**	**1.00**	**1.00**	**1.00**	**1.01**
midas	**0.98**	**1.00**	1.02	1.04	1.03	1.07
ar-midas	1.02	1.02	1.05	1.07	1.15	1.20
mf-var	1.01	1.01	1.01	1.02	1.01	**1.01**

Notes: The entries in the table are obtained as follows: first, estimate recursively every individual model and compute the relative MSE with respect to the benchmark; then, take the average all across the indicators of the relative MSE within a model class. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the best relative MSE performance for each horizon.

Table 10: Relative MSE performance of model pooling within a given class of models against AR benchmark

		horizon (h_m)					
model		**10**	**11**	**12**	**13**	**14**	**15**
bridge	mean	0.98	0.99	0.99	1.00	0.99	1.00
	weighted mean	0.95	0.95	0.94	0.96	0.96	0.98
	median	1.01	1.01	1.01	1.01	1.01	1.01
midas	mean	0.95	0.95	0.95	0.97	0.95	0.97
	weighted mean	**0.92**	**0.93**	**0.93**	**0.95**	**0.94**	**0.96**
	median	0.98	0.98	0.98	0.99	0.98	0.99
ar-midas	mean	0.96	0.97	0.96	0.98	1.06	1.09
	weighted mean	0.94	0.94	0.94	0.96	1.04	1.08
	median	1.00	1.01	1.00	1.02	1.13	1.14
mf-var	mean	1.00	1.01	1.01	1.01	1.01	1.01
	weighted mean	1.00	1.00	1.01	1.01	1.01	1.01
	median	1.01	1.01	1.01	1.01	1.01	1.01
all	mean	0.97	0.98	0.98	0.99	0.99	1.01
	weighted mean	0.95	0.96	0.96	0.97	0.98	1.00
	median	1.00	1.00	1.00	1.01	1.01	1.01

Notes: The entries in the first part of the table are obtained as follows: first, forecasts from single indicator models are computed, and means, medians and weighted means of all the forecasts within a single model are obtained; second, the MSE of these three different forecast combinations is calculated and divided by the MSE of the benchmark. In the second part the entries are obtained in the same way, but the different combinations are obtained across all individual models of different classes. The estimation is conducted recursively. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the best relative MSE performance for each horizon.

Table 11: Average relative MSE performance of different classes of mixed-frequency models against AR benchmark - Rolling/Recursive comparison

	horizon (h_m)								
model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
A. Rolling/Rolling									
bridge	1.89	**0.90**	**0.94**	**0.95**	**0.96**	0.97	0.98	1.01	1.00
midas	1.74	0.95	1.01	1.09	1.03	1.03	1.06	1.11	1.11
ar-midas	**1.01**	0.91	0.95	1.05	1.04	1.02	1.06	1.19	1.16
mf-var	1.53	0.94	0.96	0.99	0.98	**0.96**	**0.97**	**0.98**	**0.96**
B. Rolling/Recursive									
bridge	1.66	**0.97**	**1.02**	**1.02**	**1.01**	1.02	1.03	1.03	1.03
midas	1.53	1.02	1.09	1.17	1.08	1.08	1.11	1.13	1.13
ar-midas	**0.89**	0.98	**1.02**	1.13	1.10	1.07	1.11	1.22	1.19
mf-var	1.34	1.01	1.03	1.07	1.03	**1.00**	**1.01**	**1.00**	**0.98**
C. Recursive/Rolling									
bridge	2.00	0.90	0.92	0.97	0.92	0.94	0.97	**0.97**	**0.97**
midas	1.94	0.90	0.96	0.99	0.92	0.95	0.95	**0.97**	0.98
ar-midas	**1.00**	**0.76**	**0.80**	**0.82**	**0.90**	**0.91**	**0.92**	1.00	1.00
mf-var	1.55	0.93	0.96	1.00	0.96	0.97	0.98	0.98	0.98

Notes: The entries in the table are obtained as follows: first, estimate every individual model and compute the relative MSE with respect to the benchmark; then, take the average all across the indicators of the relative MSE within a model class. The benchmark is the estimate of an AR model with lag length specified accordingly to the BIC criterion. In each panel we compare estimations obtained recursively or rolling as indicated in the header of the panel: in Panel A, the individual models and the benchmark are both estimated rolling; in Panel B, the individual models are estimated rolling and the benchmark recursively; in Panel C the individual models are estimated resursively and the benchmark rolling. In each panel we also provide results of pooling across different classes of models. The rolling window has a size of 7 years. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the best relative MSE performance for each horizon in each panel.

Table 12: Relative MSE performance of model pooling within a given class of models against AR benchmark - Rolling/Recursive comparison (combination scheme: median)

	horizon (h_m)								
model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
A. Rolling/Rolling									
bridge	1.68	0.90	0.98	1.00	**0.97**	0.98	**0.97**	**0.98**	0.98
midas	1.38	0.89	1.00	1.06	1.00	1.01	1.02	1.04	1.05
ar-midas	**0.87**	**0.86**	**0.89**	**0.92**	0.99	0.99	1.00	1.10	1.08
mf-var	1.24	0.97	1.02	1.06	0.98	**0.97**	**0.97**	**0.98**	**0.96**
all	1.17	0.89	0.97	1.03	0.98	0.98	0.98	0.99	0.99
B. Rolling/Recursive									
bridge	1.48	0.96	1.05	1.08	**1.02**	**1.02**	**1.01**	**1.00**	1.00
midas	1.22	0.95	1.08	1.14	1.05	1.06	1.07	1.07	1.07
ar-midas	**0.77**	**0.93**	**0.96**	**0.99**	1.04	1.04	1.05	1.12	1.10
mf-var	1.09	1.04	1.10	1.14	1.03	**1.02**	1.02	**1.00**	**0.98**
all	1.03	0.96	1.05	1.11	1.03	1.03	1.03	1.02	1.01
C. Recursive/Rolling									
bridge	1.90	0.91	0.96	0.99	0.96	0.98	0.98	0.98	0.98
midas	1.84	0.87	0.96	0.99	0.90	0.96	0.95	**0.96**	**0.96**
ar-midas	**0.98**	**0.76**	**0.80**	**0.81**	**0.88**	**0.90**	**0.91**	0.97	0.97
mf-var	1.39	0.95	1.01	1.04	0.98	0.99	0.99	0.98	0.98
all	1.19	0.91	0.95	1.00	0.97	1.01	1.01	1.00	1.00

Notes: The entries in the table are obtained as follows: first, forecasts from single indicator models are computed, and the median of all the forecasts within a single model is obtained; second, the MSE of these the forecast combination is calculated and divided by the MSE of the benchmark. The benchmark is the estimate of an AR model with lag length specified accordingly to the BIC criterion. In each panel we compare estimations obtained recursively or rolling as indicated in the header of the panel: in Panel A, the individual models and the benchmark are both estimated rolling; in Panel B, the individual models are estimated rolling and the benchmark recursively; in Panel C the individual models are estimated resursively and the benchmark rolling. the rolling window has a size of 7 years. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the best relative MSE performance for each horizon in each panel.

Table 13: Relative MSE performance of the best mixed-frequency models with different indicators against AR benchmark

		sample: 2003q1 - 2009q1			sample: 2003q1 - 2006q4			sample: 2007q1 - 2009q1			sample: 2009q2 - 2010q2		
		horizon (h_m)			horizon (h_m)			horizon (h_m)			horizon (h_m)		
monthly indic.	**model**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
General economic situation over the next 12 months	bridge	1.26	0.73	0.74	1.08	0.82	0.79	1.29	0.73	0.74	0.22	0.15	0.08
	midas	1.03	0.78	0.73	1.07	0.94	1.03	1.03	0.76	0.71	0.19	0.12	0.10
	ar-midas	0.71	0.65	0.67	1.16	0.94	1.02	0.64	0.62	0.63	0.33	0.35	0.34
	mf-var	0.95	0.72	0.75	0.99	0.74	0.75	0.95	0.72	0.76	0.37	0.23	0.12
Production expectations for the months ahead	bridge	0.65	0.38	0.46	0.88	0.71	0.73	0.61	0.35	0.44	0.13	0.05	0.07
	midas	0.59	0.46	0.61	0.90	0.72	0.77	0.54	0.43	0.60	0.24	0.16	0.26
	ar-midas	0.52	0.42	0.54	0.72	0.65	0.79	0.49	0.40	0.52	0.41	0.26	0.50
	mf-var	0.79	0.41	0.52	0.93	0.72	0.72	0.77	0.38	0.50	0.27	0.10	0.13
Economic sentiment indicator	bridge	0.66	0.39	0.40	0.69	0.54	0.72	0.66	0.38	0.37	0.08	0.04	0.04
	midas	0.81	0.59	0.66	0.77	0.65	0.82	0.81	0.59	0.65	0.34	0.21	0.28
	ar-midas	0.60	0.54	0.60	0.78	0.72	0.88	0.57	0.52	0.58	0.46	0.52	0.65
	mf-var	0.58	0.41	0.48	0.67	0.60	0.75	0.56	0.39	0.45	0.14	0.06	0.06
New orders received index - Manufacturing	bridge	0.78	0.46	0.64	1.15	0.87	1.14	0.72	0.42	0.59	0.34	0.19	0.29
	midas	0.67	0.41	0.68	1.34	1.02	1.35	0.57	0.35	0.62	0.25	0.31	0.56
	ar-midas	0.55	0.39	0.63	0.89	1.04	1.04	0.49	0.33	0.59	0.23	0.34	0.80
	mf-var	0.71	0.48	0.63	1.46	0.89	1.03	0.60	0.44	0.60	0.11	0.08	0.12
Unemployment - Under 25 years	bridge	0.73	0.49	0.79	1.25	0.76	0.95	0.64	0.46	0.78	0.36	0.21	0.17
	midas	0.78	0.52	0.75	1.21	1.15	1.53	0.71	0.46	0.68	0.38	0.23	0.29
	ar-midas	0.62	0.50	0.72	1.17	1.18	1.58	0.53	0.44	0.65	0.70	0.31	1.22
	mf-var	0.64	0.57	0.79	0.98	0.66	0.95	0.59	0.56	0.77	0.30	0.16	0.14
AR: MSE		**0.27**	**0.61**	**0.61**	**0.06**	**0.08**	**0.08**	**0.64**	**1.56**	**1.56**	**1.38**	**2.40**	**2.40**

Notes: The entries in the table are obtained as follows: first, estimate recursively every individual model, then compute the relative MSE with respect to the benchmark. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is indicated in the top row. The absolute value of the AR MSE (benchmark) is displayed in the bottom row of the table.

Table 14: Relative MAE performance of the best mixed-frequency models with different indicators against AR benchmark

		sample: 2003q1 - 2009q1			sample: 2003q1 - 2006q4			sample: 2007q1 - 2009q1			sample: 2009q2 - 2010q2		
		horizon (h_m)			horizon (h_m)			horizon (h_m)			horizon (h_m)		
monthly indic.	**model**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
General economic situation over the next 12 months	bridge	1.08	0.90	0.91	1.01	0.96	0.94	1.12	0.88	0.90	0.56	0.41	0.36
	midas	1.01	0.95	0.95	1.00	1.04	1.09	1.02	0.91	0.89	0.52	0.43	0.43
	ar-midas	0.91	0.88	0.89	1.00	0.96	0.95	0.85	0.84	0.86	0.67	0.74	0.70
	mf-var	0.98	0.89	0.90	0.93	0.90	0.89	1.00	0.88	0.90	0.83	0.55	0.48
Production expectations for the months ahead	bridge	0.84	0.72	0.77	0.87	0.84	0.90	0.82	0.66	0.71	0.53	0.32	0.39
	midas	0.87	0.78	0.86	0.98	0.86	0.96	0.80	0.73	0.82	0.62	0.47	0.60
	ar-midas	0.79	0.76	0.84	0.85	0.85	0.98	0.75	0.72	0.78	0.78	0.56	0.76
	mf-var	0.92	0.73	0.81	0.93	0.82	0.88	0.91	0.69	0.77	0.68	0.45	0.53
Economic sentiment indicator	bridge	0.80	0.67	0.73	0.80	0.71	0.92	0.81	0.65	0.64	0.39	0.23	0.25
	midas	0.92	0.85	0.90	0.91	0.90	1.00	0.94	0.82	0.85	0.76	0.63	0.69
	ar-midas	0.82	0.83	0.88	0.85	0.91	0.99	0.80	0.79	0.82	0.78	0.91	0.97
	mf-var	0.78	0.70	0.79	0.75	0.76	0.94	0.80	0.68	0.72	0.51	0.31	0.31
New orders received index - Manufacturing	bridge	0.89	0.76	0.90	1.08	1.03	1.14	0.77	0.64	0.80	0.78	0.56	0.66
	midas	0.92	0.78	0.95	1.16	1.09	1.16	0.77	0.63	0.85	0.67	0.66	0.90
	ar-midas	0.79	0.79	0.91	0.89	1.13	1.07	0.72	0.63	0.83	0.61	0.68	1.02
	mf-var	0.83	0.76	0.85	1.10	0.92	0.98	0.67	0.68	0.78	0.45	0.35	0.45
Unemployment - Under 25 years	bridge	0.88	0.78	0.94	1.03	0.94	1.03	0.79	0.71	0.90	0.86	0.57	0.57
	midas	0.98	0.84	0.98	1.16	1.11	1.22	0.86	0.71	0.87	0.81	0.57	0.72
	ar-midas	0.90	0.83	0.98	1.10	1.10	1.24	0.78	0.70	0.85	1.01	0.65	1.23
	mf-var	0.89	0.82	0.97	1.00	0.89	1.07	0.81	0.79	0.92	0.79	0.48	0.50
AR: MAE		**0.34**	**0.45**	**0.45**	**0.20**	**0.23**	**0.23**	**0.59**	**0.86**	**0.86**	**0.76**	**1.02**	**1.02**

Notes: The entries in the table are obtained as follows: first, estimate recursively every individual model, then compute the relative MAE with respect to the benchmark. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is indicated in the top row. The absolute value of the AR MAE (benchmark) is displayed in the bottom row of the table.

Table 15: Relative MSE performance of models with monthly and weekly data against the correspondent model with only the monthly series

			horizon (h_m)								
model	**monthly indic.**	**weekly indic.**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
MIDAS	General economic situation over the next 12 months	3m Fibor	**0.84**	**0.87**	**0.89**	1.11	1.02	**0.69**	1.16	1.07	1.02
		10y Bund	1.07	**0.94**	1.09	1.01	**0.97**	**0.92**	**0.99**	1.01	**0.99**
		Spread	1.00	**0.88**	**0.88**	**0.78**	**0.74**	**0.73**	**0.82**	**0.89**	**0.86**
	Production expectations for the months ahead	3m Fibor	**0.92**	1.02	**0.95**	1.08	1.09	1.01	1.34	1.04	1.07
		10y Bund	1.02	**0.90**	1.03	**0.90**	**0.94**	**0.94**	**0.97**	**0.99**	1.02
		Spread	1.04	**0.90**	**0.80**	**0.75**	**0.87**	**0.85**	**0.85**	**0.97**	**0.94**
	Economic sentiment indicator	3m Fibor	**0.97**	1.10	**0.97**	1.12	1.03	**0.71**	1.17	1.04	**0.98**
		10y Bund	1.03	**0.91**	1.03	**0.97**	**0.94**	**0.94**	**0.98**	**0.99**	**0.97**
		Spread	**0.94**	**0.77**	**0.70**	**0.65**	**0.73**	**0.69**	**0.75**	**0.79**	**0.85**
	New orders received index - Manufacturing	3m Fibor	**0.94**	**0.87**	**0.90**	1.04	**0.90**	**0.69**	1.20	1.02	1.04
		10y Bund	1.06	1.15	1.08	1.03	1.02	**0.91**	**0.97**	**0.99**	1.00
		Spread	1.04	1.00	**0.95**	**0.90**	**0.96**	**0.94**	**0.94**	**0.99**	**0.95**
	Unemployment - Under 25 years	3m Fibor	1.01	**0.77**	**0.92**	1.04	**0.91**	1.06	1.14	1.02	1.08
		10y Bund	1.04	**0.92**	1.06	**0.98**	**0.98**	**0.92**	1.00	1.07	1.03
		Spread	**0.94**	**0.67**	**0.74**	**0.76**	**0.78**	**0.87**	**0.77**	**0.84**	**0.94**
AR-MIDAS	General economic situation over the next 12 months	3m Fibor	1.27	**0.94**	1.19	1.12	**0.95**	**0.71**	1.25	1.09	**0.96**
		10y Bund	1.09	1.06	1.02	1.08	**0.98**	**0.77**	1.03	**0.92**	**0.94**
		Spread	1.23	**0.98**	**0.92**	**0.88**	**0.77**	**0.81**	**0.92**	1.02	**0.88**
	Production expectations for the months ahead	3m Fibor	**0.99**	1.01	1.67	1.10	**0.97**	1.10	1.42	1.08	1.05
		10y Bund	1.09	1.01	1.04	1.04	**0.94**	**0.93**	1.07	1.02	1.00
		Spread	1.11	**0.97**	**0.85**	**0.80**	**0.86**	**0.85**	**0.92**	1.05	**0.94**
	Economic sentiment indicator	3m Fibor	1.31	**0.99**	1.19	1.14	**0.93**	**0.76**	1.42	1.05	**0.94**
		10y Bund	**0.96**	**0.98**	1.02	1.04	**0.96**	**0.89**	1.00	**0.94**	**0.96**
		Spread	1.12	**0.82**	**0.73**	**0.70**	**0.73**	**0.73**	**0.84**	**0.91**	**0.89**
	New orders received index - Manufacturing	3m Fibor	1.09	1.00	1.12	1.13	**0.83**	**0.74**	1.16	1.09	1.03
		10y Bund	1.21	1.07	1.09	1.13	**0.97**	**0.95**	1.02	**0.97**	**0.99**
		Spread	1.03	1.10	**0.95**	**0.93**	**0.89**	**0.88**	**0.87**	1.03	**0.96**
	Unemployment - Under 25 years	3m Fibor	1.12	**0.78**	**0.94**	1.07	**0.94**	1.01	1.38	1.00	1.08
		10y Bund	1.21	**0.90**	1.02	**0.94**	1.00	**0.89**	1.13	**0.96**	1.04
		Spread	1.21	**0.76**	**0.78**	**0.72**	**0.83**	**0.91**	**0.74**	**0.91**	1.04

Notes: The entries in the table are obtained as follows: estimate recursively every individual model with monthly and weekly data and compute the relative MSE with respect to the model with only the corresponding monthly indicator. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the cases in which the introduction of weekly data improves the performance.

Table 16: Revisions from January 2002 to March 2009: main statistics

	variable		**revisions**		
	average	**sd**	**average**	**min**	**max**
Unemployment - under 25 years	-0.1	0.2	0.0	-0.7	1.3
Industrial production -durable consumer	0.0	1.9	0.0	-3.9	4.9
GDP - market prices	0.6	0.5	0.0	-0.3	0.6

Notes: The statistics refer to the series and to their revisions compared to the last available vintage (August 2009).

Table 17: Relative MSE performance of individual models within a given class of models against AR benchmark, obtained with a pseudo real-time dataset and a real-time dataset. Both indicators and GDP revised

		horizon (h_m)								
model	**monthly indicators**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
A. Pseudo real-time dataset										
midas	Industrial production - durable consumer goods	0.84	0.38	0.83	0.85	1.01	0.96	1.12	1.15	1.04
	Unemployment - under 25 years	0.75	0.68	0.76	0.81	0.89	0.94	0.89	0.91	0.98
ar-midas	Industrial production - durable consumer goods	0.79	0.38	0.85	0.89	0.95	0.93	1.08	1.16	1.05
	Unemployment - under 25 years	0.79	0.71	0.77	0.89	0.90	0.93	0.93	0.93	1.00
mf-var	Industrial production - durable consumer goods	0.78	0.73	0.97	1.02	1.01	0.99	1.00	1.00	1.00
	Unemployment - under 25 years	0.71	0.69	0.84	0.85	0.96	0.97	0.98	0.97	0.98
bridge	Industrial production - durable consumer goods	0.52	0.43	0.76	0.88	1.01	0.97	1.00	1.00	0.98
	Unemployment - under 25 years	0.68	0.65	0.72	0.75	0.85	0.89	0.89	0.95	0.91
B. Real-time dataset										
midas	Industrial production - durable consumer goods	0.78	0.50	0.86	0.90	1.03	1.01	1.12	1.10	1.13
	Unemployment - under 25 years	0.65	0.69	0.76	0.83	0.90	0.97	0.92	0.94	0.97
ar-midas	Industrial production - durable consumer goods	0.71	0.51	0.87	0.99	0.96	1.01	1.11	1.13	1.16
	Unemployment - under 25 years	0.75	0.71	0.74	0.88	0.91	0.94	1.00	0.97	0.95
mf-var	Industrial production - durable consumer goods	0.83	0.77	0.98	1.01	1.01	0.99	1.00	1.00	1.00
	Unemployment - under 25 years	0.69	0.70	0.85	0.87	0.97	0.98	0.99	0.98	0.98
bridge	Industrial production - durable consumer goods	0.63	0.55	0.88	0.97	1.04	0.98	1.00	1.01	1.00
	Unemployment - under 25 years	0.69	0.74	0.77	0.80	0.87	0.90	0.90	0.96	0.89

Notes: Panel A reports the results obtained with a pseudo real-time dataset, Panel B reports the results obtained with a genuine real-time dataset. Both the indicators and the GDP growth are subject to revisions. Forecasts are evaluated against the final vintage of GDP available at the downloaded date. The entries in the table are obtained as follows: first, estimate recursively every individual model and compute the relative MSE with respect to the benchmark; then, take the average all across the indicators of the relative MSE within a model class. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1.

Table 18: Relative MSE performance of individual models within a given class of models against AR benchmark, obtained with a pseudo real-time dataset and a real-time dataset. Only GDP revised

		horizon (h_m)								
model	**monthly indicators**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
A. Pseudo real-time dataset										
	General economic situation over the next 12 months	0.45	0.55	0.52	0.57	0.74	0.75	0.93	0.90	0.86
midas	Production expectations for the months ahead	0.32	0.34	0.40	0.50	0.67	0.81	0.90	0.94	1.02
	Economic sentiment indicator	0.32	0.38	0.43	0.55	0.67	0.73	0.86	0.87	0.87
	General economic situation over the next 12 months	0.45	0.51	0.53	0.59	0.73	0.73	0.97	0.90	0.90
ar-midas	Production expectations for the months ahead	0.33	0.36	0.43	0.53	0.68	0.81	0.89	0.95	0.99
	Economic sentiment indicator	0.34	0.39	0.46	0.59	0.66	0.72	0.86	0.85	0.96
	General economic situation over the next 12 months	0.59	0.54	0.57	0.63	0.74	0.74	0.83	0.85	0.86
mf-var	Production expectations for the months ahead	0.38	0.38	0.42	0.54	0.67	0.78	0.89	0.93	0.95
	Economic sentiment indicator	0.48	0.34	0.35	0.50	0.67	0.73	0.83	0.86	0.89
	General economic situation over the next 12 months	0.61	0.55	0.56	0.61	0.71	0.68	0.82	0.85	0.84
bridge	Production expectations for the months ahead	0.39	0.38	0.43	0.52	0.64	0.78	0.92	0.96	0.98
	Economic sentiment indicator	0.34	0.39	0.40	0.44	0.59	0.61	0.82	0.85	0.81
B. Real-time dataset										
	General economic situation over the next 12 months	0.62	0.58	0.59	0.72	0.76	0.93	0.89	0.91	0.96
midas	Production expectations for the months ahead	0.35	0.34	0.50	0.64	0.68	0.90	0.91	0.94	1.08
	Economic sentiment indicator	0.42	0.40	0.54	0.66	0.69	0.89	0.87	0.89	0.97
	General economic situation over the next 12 months	0.57	0.54	0.60	0.72	0.77	0.96	0.91	0.89	0.99
ar-midas	Production expectations for the months ahead	0.36	0.36	0.51	0.65	0.68	0.90	0.91	0.95	1.14
	Economic sentiment indicator	0.43	0.40	0.56	0.68	0.67	0.88	0.86	0.88	0.97
	General economic situation over the next 12 months	0.61	0.56	0.65	0.71	0.75	0.82	0.82	0.86	0.91
mf-var	Production expectations for the months ahead	0.41	0.40	0.57	0.61	0.68	0.89	0.90	0.92	0.98
	Economic sentiment indicator	0.56	0.49	0.69	0.58	0.70	0.85	0.84	0.87	0.93
	General economic situation over the next 12 months	0.62	0.57	0.63	0.69	0.71	0.82	0.81	0.85	0.91
bridge	Production expectations for the months ahead	0.40	0.38	0.54	0.61	0.65	0.91	0.93	0.94	1.00
	Economic sentiment indicator	0.45	0.44	0.49	0.59	0.61	0.83	0.83	0.86	0.94

Notes: Panel A reports the results obtained with a pseudo real-time dataset, Panel B reports the results obtained with a genuine real-time dataset. Only GDP growth is subject to revisions. Forecasts are evaluated against the final vintage of GDP available at the downloaded date. The entries in the table are obtained as follows: first, estimate recursively every individual model and compute the relative MSE with respect to the benchmark; then, take the average all across the indicators of the relative MSE within a model class. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1.

B Results for euro area GDP components

In this Appendix, instead of focusing on the economic activity at aggregate level, we look at the disaggregate GDP components. Considering a decomposition from the output side, we follow the NACE classification of the GDP and obtain six branches of activity: agriculture, hunting, forestry and fishing; industry, excluding construction; construction; trade, hotels and restaurants, transport and communication services; financial services and business activities; other services. From the expenditure side, instead, we obtain five components: household final consumption; government final consumption; gross fixed capital formation, imports and exports.

In the Tables 19 and 20, we compare the performance of the different approaches with respect to the benchmark, which is an AR model for each singular component, where the lag length is specified accordingly to the BIC criterion. What we show is the average relative MSE performance of the different classes of mixed-frequency models against the AR benchmark. We consider nowcasts and forecasts up to two quarters ahead. In Table 19 we present the results for the components from the output side, while in Table 20 we present the results for the components from the expenditure side.

The evidence is quite mixed, depending on which component we are focusing on. More in detail, looking at the GDP disaggregation from the output side, the mixed-frequency approaches outperform the benchmark for those components for which many monthly indicators are available, as in the case of the industry sector, trade and financial services. For agriculture, availability of monthly indicators is critical. The same holds for the last branch that includes a variety of economic activities (public administration and defence, compulsory social security; education; health and social work; other community, social and personal service activities; private households with employed persons) for which it is not easy to find reliable and timely monthly indicators of value added.

As a general remark, the AR-MIDAS outperform the correspondent basic MIDAS. Moreover, this kind of models seems to work particularly well for short horizons. The evidence in favour of MF-VAR is less strong, and, as for the results on the aggregate measure of economic activity discussed in Section 4, this approach provides better results for longer horizons. Bridge models perform well, usually as well as MIDAS but they are outperformed by AR-MIDAS especially at very short horizons.

Looking now at the components from the expenditure side, we can reach the same kind of conclusions. Since for the government final consumption it is very hard to find monthly indicators, it is very difficult to beat the benchmark (no gains from using monthly indicators appear at any horizon). For all the other components, bridge equations and AR-MIDAS have a better performance than the benchmark (the latter showing slightly better results than the former even in this case). The MF-VAR approach has a less clear cut performance also in this case, but it performs better than the benchmark especially for longer horizons, beating even the MIDAS (with and without AR dynamics) for $h_m = 8, 9$ forecast horizons for some components.

We now move to analyze the gains in case of forecast pooling, to assess whether there is any benefit from combining forecasts from alternative models with different explanatory variables. We calculate the forecasts with the usual three different combination schemes, but we report only the results obtained with the median, to save space.

Even with pooling, improvements are only obtained for those components for which there is a variety of indicators available, namely total industry, trade and financial services from the output side (see Table 21), and household final consumption, gross fixed capital formation and external balance from the expenditure side (see Table 22).

Generally, forecast combinations of AR-MIDAS models perform pretty well, outperforming the benchmark at several horizons. Also combinations within simple MIDAS and bridge equations allow for gains at some horizons. Pooling within the MF-VAR class also beats the benchmark but not very often, and even when the performance is better, the gains are not so big, confirming the results found for the aggregate GDP. Pooling across all the methods gives good results for some components, and generally only for very short horizons.

To conclude, as in Section 6, we analyze the behavior of large scale factor models in now- and forecasting each single component of the GDP, from the expenditure (Table 23) and supply side (Table 24). We can reach the same conclusions as for the aggregate GDP growth. Evidence is in favour of the use of factor models to predict the quarterly growth of each component for which the dataset contains useful information. There are significant gains especially at very short horizons: generally, exploiting the unbalanced structure of the dataset improves the performance, and the inclusion of an AR component reduces the MSE, even though not systematically. As for the case of the GDP growth discussed in Section 7, also for the components of GDP the factor models seem to better perform relative to the benchmark than forecast pooling. This is true especially for nowcasts and short-term forecasts. However, when the forecast horizon increases, the outperformance of the factor models is no longer evident, and in many cases forecast pooling is better. Finally, for these long horizons, both methods of summarizing information (factors and forecast pooling) generally fail in beating the AR benchmark.

Table 19: Average relative MSE performance of different classes of mixed-frequency models against AR benchmark - GDP components (supply side)

		horizon (h_m)								
component	**model**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Agriculture, hunting, forestry and fishing	bridge	**0.70**	**0.98**	**0.97**	**0.97**	1.10	1.09	1.09	**1.13**	**1.13**
	midas	0.72	1.04	1.03	1.05	1.19	1.19	1.18	1.22	1.27
	ar-midas	0.81	1.10	1.09	1.11	1.32	1.34	1.31	1.07	1.12
	mf-var	0.76	1.04	1.01	1.00	**1.09**	**1.08**	**1.08**	1.14	1.14
Total industry, excluding construction	bridge	2.06	0.94	0.98	1.01	**0.93**	**0.95**	0.96	**0.94**	**0.94**
	midas	2.01	0.95	1.00	1.04	0.94	0.96	**0.95**	0.94	0.95
	ar-midas	**0.92**	**0.89**	**0.94**	**0.97**	0.95	0.96	0.96	0.95	0.96
	mf-var	1.45	0.94	0.98	1.02	0.95	0.96	0.97	0.95	0.95
Construction	bridge	0.95	1.10	1.07	1.08	1.08	1.06	1.05	**1.07**	1.06
	midas	**0.91**	**1.05**	**1.03**	**1.04**	**1.05**	1.03	1.04	**1.07**	1.11
	ar-midas	0.98	1.09	1.08	1.08	1.08	1.06	1.06	1.13	1.22
	mf-var	1.02	1.95	1.07	1.05	2.72	**1.01**	**1.01**	3.26	**1.01**
Trade, hotels and restaurants, transport and communication services	bridge	1.18	1.12	1.13	1.16	0.94	0.96	0.98	0.98	**0.98**
	midas	1.11	1.11	1.14	1.18	**0.91**	**0.95**	**0.95**	**0.96**	0.99
	ar-midas	**1.05**	**0.94**	**0.95**	**0.98**	0.94	0.97	0.97	0.99	0.99
	mf-var	1.22	1.12	1.15	1.20	1.00	1.01	1.01	0.99	0.99
Financial services and business activities	bridge	0.87	0.86	0.90	0.94	0.98	0.98	1.00	1.01	1.01
	midas	0.87	0.82	0.86	0.90	0.96	0.98	0.99	1.04	1.03
	ar-midas	**0.83**	**0.78**	**0.82**	**0.83**	**0.90**	**0.91**	**0.95**	1.04	1.08
	mf-var	0.86	0.85	0.91	0.94	1.00	0.98	1.00	**0.99**	**0.99**
Other services	bridge	1.07	**1.07**	**1.05**	**1.04**	**1.00**	**0.99**	**0.99**	**1.02**	**1.02**
	midas	**1.02**	1.15	1.10	1.10	1.07	1.06	1.10	1.18	1.27
	ar-midas	1.09	1.17	1.12	1.12	1.13	1.13	1.16	1.40	1.36
	mf-var	1.15	1.14	1.11	1.08	1.03	1.02	1.02	1.03	1.03

Notes: The statistics refer to the series and to their revisions compared to the last available vintage (August 2009).

Table 20: Average relative MSE performance of different classes of mixed-frequency models against AR benchmark - GDP components (expenditure side)

		horizon (h_m)								
component	**model**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Final consumption - households	bridge	1.08	0.98	0.95	0.96	0.95	0.96	0.98	**0.98**	0.97
	midas	1.10	1.03	1.00	1.02	0.98	1.07	1.06	1.05	1.06
	ar-midas	**1.02**	**0.86**	**0.83**	**0.85**	**0.82**	**0.86**	**0.88**	1.03	1.06
	mf-var	1.10	0.98	0.95	0.96	0.97	0.97	0.98	0.96	**0.96**
Final consumption - Government	bridge	**1.10**	**1.03**	**1.02**	**1.02**	**1.03**	**1.03**	**1.03**	**1.04**	**1.04**
	midas	1.15	1.10	1.03	1.05	1.11	1.10	1.18	1.15	1.15
	ar-midas	**1.10**	1.12	1.07	1.07	1.23	1.24	1.29	1.27	1.30
	mf-var	1.41	1.63	1.56	1.45	1.79	1.66	1.53	1.85	1.74
Gross fixed capital formation	bridge	1.35	1.06	1.06	1.11	1.00	1.01	1.03	**1.00**	**0.99**
	midas	1.32	1.10	1.15	1.14	1.03	1.06	1.06	1.02	1.05
	ar-midas	**0.93**	**0.91**	**0.95**	**0.95**	**0.98**	**0.96**	**0.95**	1.05	1.06
	mf-var	1.38	1.14	1.18	1.21	1.07	1.07	1.08	1.04	1.03
Imports	bridge	1.07	0.88	0.92	0.97	0.89	0.90	0.92	0.82	**0.82**
	midas	1.07	0.90	0.97	0.96	0.88	0.88	0.87	**0.80**	0.84
	ar-midas	**0.82**	**0.81**	**0.86**	**0.87**	**0.86**	**0.85**	**0.85**	0.84	0.87
	mf-var	1.03	0.91	0.96	1.01	0.93	0.93	0.94	0.83	0.83
Exports	bridge	1.41	0.91	0.96	1.00	**0.96**	0.97	0.99	0.98	**0.97**
	midas	1.38	0.91	0.97	0.99	0.97	**0.95**	**0.95**	**0.94**	0.98
	ar-midas	**0.91**	**0.86**	**0.93**	**0.95**	0.97	0.96	0.96	0.96	0.98
	mf-var	1.16	0.92	0.98	1.01	0.98	0.99	0.99	0.98	0.98

Notes: The entries in the table are obtained as follows: first, estimate recursively every individual model and compute the relative MSE with respect to the benchmark; then, take the average all across the indicators of the relative MSE within a model class. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the best relative MSE performance for each horizon and each component.

Table 21: Relative MSE performance of model pooling within a given class of models against AR benchmark (combination scheme: median)

		horizon (h_m)								
component	**model**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Agriculture, hunting, forestry and fishing	bridge	**0.68**	**0.95**	**0.94**	**0.94**	**1.06**	**1.06**	**1.06**	1.10	1.10
	midas	0.69	0.96	0.96	0.97	1.09	1.09	1.08	1.12	1.13
	ar-midas	0.75	1.01	1.01	1.02	1.18	1.18	1.18	**0.98**	**0.99**
	mf-var	0.69	0.96	0.95	**0.94**	1.07	1.07	**1.06**	1.10	1.10
	all	0.69	0.96	0.96	0.96	1.08	1.07	1.07	1.09	1.10
Total industry, excluding construction	bridge	2.06	0.99	1.02	1.06	0.96	0.97	0.97	0.95	0.95
	midas	2.02	0.97	1.02	1.05	**0.93**	0.96	**0.96**	**0.94**	**0.93**
	ar-midas	**0.90**	**0.90**	**0.94**	**0.96**	**0.93**	**0.95**	**0.96**	**0.94**	0.94
	mf-var	1.19	0.97	1.01	1.06	0.96	0.97	0.97	0.95	0.95
	all	1.42	0.94	0.98	1.02	0.95	0.96	0.97	0.95	0.95
Construction	bridge	**0.83**	**0.96**	**0.96**	**0.98**	**0.99**	**0.97**	**0.96**	**0.98**	**0.98**
	midas	0.87	1.00	1.00	1.00	1.01	1.00	1.01	1.01	1.01
	ar-midas	0.93	1.03	1.03	1.04	1.02	1.01	1.02	1.05	1.05
	mf-var	0.88	0.99	1.00	1.00	1.00	1.00	1.00	1.01	1.01
	all	0.85	1.00	1.00	1.01	1.00	0.99	0.99	1.01	1.01
Trade, hotels and restaurants, transport and communication services	bridge	1.17	1.22	1.24	1.27	1.00	1.00	1.00	0.99	0.99
	midas	1.04	1.11	1.15	1.17	**0.88**	**0.94**	**0.94**	**0.96**	0.96
	ar-midas	**1.01**	**0.94**	**0.95**	**0.97**	0.91	0.95	0.96	**0.96**	**0.95**
	mf-var	1.17	1.14	1.17	1.22	1.00	1.01	1.01	0.99	0.99
	all	1.08	1.04	1.08	1.12	0.96	0.98	0.99	0.98	0.98
Financial services and business activities	bridge	0.76	0.82	0.90	0.95	1.00	0.99	1.00	1.00	1.00
	midas	0.78	**0.70**	0.80	0.83	0.89	0.95	0.96	**0.96**	**0.96**
	ar-midas	0.73	**0.70**	**0.78**	**0.76**	**0.84**	**0.86**	**0.90**	0.97	0.98
	mf-var	**0.71**	0.79	0.92	0.96	1.00	0.99	1.00	1.00	1.00
	all	0.74	0.75	0.85	0.87	0.94	0.96	0.98	0.99	0.99
Other services	bridge	0.98	0.99	**1.00**	**1.00**	**0.98**	**0.98**	**0.97**	**1.00**	**1.00**
	midas	0.91	**0.98**	1.01	**1.00**	0.99	1.01	1.02	1.06	1.06
	ar-midas	**0.95**	1.00	1.01	1.02	1.03	1.06	1.06	1.29	1.24
	mf-var	0.99	1.02	1.03	1.03	1.00	1.01	1.01	1.02	1.02
	all	0.94	0.99	1.02	1.01	1.00	1.01	1.01	1.02	1.02

Notes: The entries in the table are obtained as follows: first, forecasts from single indicator models are computed, and the median of all the forecasts within a single model is obtained; second, the MSE of the forecast combination is calculated and divided by the MSE of the benchmark. Only the last row of each component cosiders the median of all the forecasts across methods. The estimation is conducted recursively. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the best relative MSE performance for each horizon and each indicator.

Table 22: Relative MSE performance of model pooling within a given class of models against AR benchmark (combination scheme: median)

		horizon (h_m)								
component	**model**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Final consumption - households	bridge	1.04	0.99	1.00	1.01	0.99	0.99	0.99	0.98	0.98
	midas	1.03	0.93	0.99	0.99	0.92	1.01	0.98	0.98	0.99
	ar-midas	**0.92**	**0.75**	**0.77**	**0.79**	**0.73**	**0.75**	**0.75**	**0.95**	**0.96**
	mf-var	0.95	0.93	0.96	0.98	0.98	0.99	0.99	0.97	0.97
	all	0.95	0.87	0.91	0.93	0.91	0.95	0.95	0.97	0.97
Final consumption - Government	bridge	1.04	**0.98**	0.98	**0.98**	**1.00**	**1.00**	**1.00**	1.01	1.01
	midas	1.06	0.99	**0.95**	**0.98**	**1.00**	1.02	1.04	1.05	1.05
	ar-midas	**1.01**	0.99	0.97	0.99	1.08	1.12	1.12	1.13	1.12
	mf-var	1.11	1.04	1.03	1.03	**1.00**	1.01	1.01	**1.00**	**1.00**
	all	1.04	0.98	0.98	0.98	**1.00**	1.01	1.01	**1.00**	**1.00**
Gross fixed capital formation	bridge	1.31	1.07	1.10	1.16	1.03	1.05	1.05	1.01	1.01
	midas	1.28	1.06	1.17	1.15	1.01	1.04	1.03	**0.99**	**1.00**
	ar-midas	**0.88**	**0.90**	**0.94**	**0.94**	**0.95**	**0.94**	**0.94**	1.01	1.02
	mf-var	1.27	1.12	1.20	1.24	1.07	1.08	1.08	1.03	1.03
	all	1.07	1.00	1.08	1.11	1.01	1.04	1.04	1.01	1.01
Imports	bridge	0.98	0.87	0.95	1.01	0.94	0.94	0.94	0.83	0.83
	midas	0.99	0.90	0.97	0.95	0.87	0.88	0.86	**0.78**	**0.77**
	ar-midas	**0.81**	**0.79**	**0.82**	**0.86**	**0.85**	**0.84**	**0.84**	0.79	0.79
	mf-var	0.96	0.89	1.00	1.05	0.93	0.94	0.94	0.83	0.83
	all	0.88	0.84	0.91	0.94	0.89	0.90	0.91	0.82	0.82
Exports	bridge	1.41	0.95	1.00	1.05	1.00	1.00	1.00	0.98	0.98
	midas	1.32	0.92	0.97	0.99	0.95	0.94	**0.94**	**0.93**	**0.93**
	ar-midas	**0.90**	**0.85**	**0.91**	**0.94**	**0.93**	**0.93**	**0.94**	**0.93**	**0.93**
	mf-var	1.07	0.93	1.01	1.04	0.99	1.00	1.00	0.98	0.98
	all	1.05	0.90	0.96	0.99	0.97	0.97	0.98	0.97	0.97

Notes: The entries in the table are obtained as follows: first, forecasts from single indicator models are computed, and the median of all the forecasts within a single model is obtained; second, the MSE of the forecast combination is calculated and divided by the MSE of the benchmark. Only the last row of each component cosiders the median of all the forecasts across methods. The estimation is conducted recursively. The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1. The numbers in bold show the best relative MSE performance for each horizon and each indicator.

Table 23: Relative MSE performance of different classes of factor models against AR benchmark - GDP components (supply side)

		horizon (h_m)								
component	**model**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Agriculture, hunting, forestry and fishing	Factor-MIDAS (basic)	0.71	1.04	1.02	1.06	1.17	1.22	1.32	1.38	1.27
	Factor-MIDAS (ar)	0.81	1.12	1.08	1.10	1.22	1.43	1.34	1.19	1.15
	Quarterly factor model	0.69	0.98	0.98	0.98	1.10	1.10	1.10	0.96	0.96
Total industry, excluding construction	Factor-MIDAS (basic)	1.05	0.69	0.73	0.82	0.88	0.90	0.92	0.93	0.92
	Factor-MIDAS (ar)	0.57	0.69	0.71	0.79	0.89	0.91	0.94	0.93	0.93
	Quarterly factor model	0.90	0.98	0.98	0.98	0.98	0.98	0.98	0.94	0.94
Construction	Factor-MIDAS (basic)	0.76	0.96	0.90	0.99	1.00	0.97	1.04	1.06	1.07
	Factor-MIDAS (ar)	0.83	0.99	0.92	1.01	0.99	0.97	1.05	1.10	1.11
	Quarterly factor model	0.98	1.10	1.10	1.10	1.05	1.05	1.05	1.03	1.03
Trade, hotels and restaurants, transport and	Factor-MIDAS (basic)	0.55	0.66	0.69	0.84	0.81	0.84	0.89	0.93	0.91
	Factor-MIDAS (ar)	0.56	0.58	0.61	0.76	0.82	0.86	0.93	0.90	0.89
	Quarterly factor model	0.93	0.89	0.89	0.89	1.01	1.01	1.01	1.01	1.01
Financial services and business activities	Factor-MIDAS (basic)	0.42	0.43	0.46	0.55	0.74	0.71	0.81	0.97	1.03
	Factor-MIDAS (ar)	0.40	0.44	0.46	0.56	0.71	0.66	0.78	1.11	1.18
	Quarterly factor model	0.64	0.92	0.92	0.92	1.00	1.00	1.00	1.15	1.15
Other services	Factor-MIDAS (basic)	0.87	0.87	0.96	0.96	1.00	0.96	1.13	1.09	1.13
	Factor-MIDAS (ar)	0.96	0.91	0.96	0.96	1.13	1.00	1.22	1.13	1.09
	Quarterly factor model	0.87	1.00	1.00	1.00	0.96	0.96	0.96	1.13	1.13

Notes: The entries in the table are obtained as follows: first, estimate recursively every factor model, then compute the relative MSE with respect to the benchmark. The factors are estimated with the EM algorithm together with PCA as in Stock and Watson (2002). The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1.

Table 24: Relative MSE performance of different classes of factor models against AR benchmark - GDP components (expenditure side)

		horizon (h_m)								
component	**model**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Final consumption - households	Factor-MIDAS (basic)	0.71	0.72	0.72	0.78	0.97	0.93	0.97	1.21	1.06
	Factor-MIDAS (ar)	0.69	0.57	0.58	0.67	0.72	0.70	0.68	1.17	0.92
	Quarterly factor model	1.07	0.69	0.69	0.69	0.83	0.83	0.83	0.79	0.79
Final consumption - Government	Factor-MIDAS (basic)	1.03	1.06	0.96	1.00	0.93	0.91	1.00	1.01	1.04
	Factor-MIDAS (ar)	0.97	1.06	1.03	1.01	1.04	1.00	1.12	1.09	1.23
	Quarterly factor model	0.96	0.94	0.94	0.94	1.22	1.22	1.22	0.93	0.93
Gross fixed capital formation	Factor-MIDAS (basic)	0.68	0.78	0.79	0.92	0.99	0.97	1.02	1.05	1.03
	Factor-MIDAS (ar)	0.63	0.68	0.71	0.78	0.95	0.94	0.98	1.09	1.06
	Quarterly factor model	1.11	1.09	1.09	1.09	0.90	0.90	0.90	1.06	1.06
Imports	Factor-MIDAS (basic)	0.70	0.66	0.74	0.83	0.86	0.86	0.88	0.90	0.84
	Factor-MIDAS (ar)	0.60	0.59	0.66	0.59	0.82	0.84	0.75	0.83	0.86
	Quarterly factor model	0.50	0.99	0.99	0.99	0.83	0.83	0.83	0.87	0.87
Exports	Factor-MIDAS (basic)	0.72	0.64	0.72	0.79	0.87	0.88	0.96	0.93	0.94
	Factor-MIDAS (ar)	0.62	0.64	0.72	0.75	0.87	0.88	0.96	0.93	0.94
	Quarterly factor model	0.75	0.91	0.91	0.91	1.00	1.00	1.00	1.04	1.04

Notes: The entries in the table are obtained as follows: first, estimate recursively every factor model, then compute the relative MSE with respect to the benchmark. The factors are estimated with the EM algorithm together with PCA as in Stock and Watson (2002). The benchmark is the recursive estimate of an AR model with lag length specified accordingly to the BIC criterion. The evaluation sample is 2003Q1-2009Q1.

C Dataset description

MONTHLY DATA	log/diff
HICP - All items excluding energy and unprocessed food	5
HICP - All items excluding energy, food, alcohol and tobacco	5
HICP - All items excluding energy and seasonal food	5
HICP - All items excluding energy	5
HICP - All items excluding tobacco	5
HICP - All items (HICP=Harmonized Index of Consumer Prices)	5
HICP - Food and non alcoholic beverages	5
HICP - Alcoholic beverages and tobacco	5
HICP - Clothing and footwear	5
HICP - Housing, water, electricity,gas and other fuels	5
HICP - Furnishings, household equipment and maintenance	5
HICP - Health	5
HICP - Transport	5
HICP - Communication	5
HICP - Recreation and culture	5
HICP - Education	5
HICP - Hotels, cafes and restaurants	5
HICP - Miscellaneous goods and services	5
HICP - Energy	5
HICP - Food	5
Producer price index - Electricity, gas, steam and air conditioning supply	5
Producer price index - Industry (except construction), sewerage, waste management and remediation activities	5
Producer price index - Mining and quarrying; manufacturing; electricity, gas, steam and air conditioning supply	5
Producer price index - Mining and quarrying	5
Producer price index - B_TO_E36	5
Producer price index - Manufacturing	5
Producer price index - Manufacturing, for new orders	5
Producer price index - Electricity, gas, steam and air conditioning supply	5
Producer price index - Water collection, treatment and supply	5
Producer price index - Capital goods	5
Producer price index - Consumer goods	5
Producer price index - Durable consumer goods	5
Producer price index - Intermediate goods	5
Producer price index - Non-durable consumer goods	5
Producer price index - Energy	5
Business surveys - Construction - Construction confidence indicator	1
Business surveys - Construction - Employment expectations for the months ahead	1
Business surveys - Construction - Assessment of order books	1
Business surveys - Construction - Price expectations for the months ahead	1
Business surveys - Construction - Trend of activity compared with preceding months	1
Business Climate Indicator	1
Consumer confidence indicator	1
Consumer surveys - Financial situation over the last 12 months	1
Consumer surveys - Financial situation over the next 12 months	1
Consumer surveys - General economic situation over the last 12 months	1
Consumer surveys - General economic situation over the next 12 months	1
Consumer surveys - Major purchases over the next 12 months	1
Consumer surveys - Major purchases at present	1
Consumer surveys - Price trends over the last 12 months	1
Consumer surveys - Price trends over the next 12 months	1
Consumer surveys - Statement on financial situation of household	1
Consumer surveys - Savings over the next 12 months	1
Consumer surveys - Savings at present	1

Consumer surveys - Unemployment expectations over the next 12 months	1
Business surveys - Industry - Industrial confidence indicator	1
Business surveys - Industry - Employment expectations for the months ahead	1
Business surveys - Industry - Assessment of export order-book levels	1
Business surveys - Industry - Assessment of order-book levels	1
Business surveys - Industry - Expectations for the months ahead	1
Business surveys - Industry - Production trend observed in recent months	1
Business surveys - Industry - Assessment of stocks of finished products	1
Business surveys - Industry - Selling price expectations for the months ahead	1
Business surveys - Retail - Assessment of stocks	1
Business surveys - Retail - Retail confidence indicator	1
Business surveys - Retail - Expected business situation	1
Business surveys - Retail - Employment	1
Business surveys - Retail - Orders placed with suppliers	1
Business surveys - Retail - Present business situation	1
Business survey - Services - Assessment of business climate	1
Business survey - Services - Evolution of demand expected in the months ahead	1
Business survey - Services - Evolution of demand in recent months	1
Business survey - Services - Services Confidence Indicator	1
Business survey - Services - Evolution of employment in recent months	1
Construction confidence indicator	1
Consumer confidence indicator	1
Economic sentiment indicator	1
Industrial confidence indicator	1
Retail confidence indicator	1
Services Confidence Indicator	1
Unemployment according to ILO definition - Over 25 years - Total	2
Unemployment according to ILO definition - Under 25 years - Total	2
Unemployment according to ILO definition - Total	2
Unemployment according to ILO definition - Over 25 years - Total	1
Unemployment according to ILO definition - Under 25 years - Total	1
Unemployment according to ILO definition - Total	1
Production index	5
Production index - Buildings	5
Production index - Civil engineering works	5
Production index - Construction	5
Production index - Mining and quarrying; manufacturing; electricity, gas, steam and air conditioning supply	5
Production index - Mining and quarrying; manufacturing	5
Production index - Mining and quarrying	5
Production index - Manufacturing	5
Production index - Manufacturing, for new orders	5
Production index - Electricity, gas, steam and air conditioning supply	5
Production index - Capital goods	5
Production index - Consumer goods	5
Production index - Durable consumer goods	5
Production index - Intermediate goods	5
Production index - Non-durable consumer goods	5
Turnover index - domestic market - Mining and quarrying; manufacturing	5
Turnover index - non-domestic market - Mining and quarrying; manufacturing	5
Turnover index - total - Mining and quarrying; manufacturing	5
Turnover index - domestic market - Manufacturing	5
Turnover index - non-domestic market - Manufacturing	5
Turnover index - total - Manufacturing	5
Turnover index - domestic market - Manufacturing, for new orders	5
Turnover index - non-domestic market - Manufacturing, for new orders	5
Turnover index - total - Manufacturing, for new orders	5
Turnover index - domestic market - Capital goods	5

Turnover index - non-domestic market - Capital goods	5
Turnover index - total - Capital goods	5
Turnover index - domestic market - Consumer goods	5
Turnover index - non-domestic market - Consumer goods	5
Turnover index - total - Consumer goods	5
Turnover index - domestic market - Durable consumer goods	5
Turnover index - non-domestic market - Durable consumer goods	5
Turnover index - total - Durable consumer goods	5
Turnover index - domestic market - Intermediate goods	5
Turnover index - non-domestic market - Intermediate goods	5
Turnover index - total - Intermediate goods	5
Turnover index - domestic market - Non-durable consumer goods	5
Turnover index - total - Non-durable consumer goods	5
Hours worked index - Construction	5
New orders received index - Manufacturing, for new orders	5
New orders received index - Manufacturing, for new orders (except heavy transport equipment)	5
Number of new car registrations	5
Deflated turnover index - Retail sale of food, beverages and tobacco	5
Deflated turnover index - Retail trade, except of motor vehicles and motorcycles	5
Deflated turnover index - Retail sale of non-food products (including fuel)	5
Deflated turnover index - Retail sale of non-food products (except fuel)	5
Deflated turnover index - Retail trade, except of motor vehicles, motorcyles and fuel	5
MF-M1-SA Money supply M1 - SA	5
MF-M2-SA Money supply M2 - SA	5
MF-M3-SA Money supply M3 - SA	5
MF-3MI-RT 3-month interest rates (average)	2
MF-LTGBY-RT Long term government bond yields - Maastricht definition (average)	2
Exchange rates US Dollar against the ECU/euro (average)	2
Exchange rates Yen against the ECU/euro (average)	2
Exchange rates Pound Sterling against the ECU/euro (average)	2
DAX share price index	5
DJ EURO STOXX 50, price index	5
EM government bond yield - 2 year	2
EM government bond yield - 3 year	2
EM government bond yield - 5 year	2
EM government bond yield - 7 year	2
EM government bond yield - 10 year	2
Germany interbank 12 month - offered rate	2
Germany interbank 3 month - offered rate	2
Germany interbank 6 month - offered rate	2
German yields on fully taxed bonds outstanding - all bank bonds	2
German yields on fully taxed bonds outstanding - corporate bonds	2

Notes: log/diff: 1: unchanged; 2: first differencing, no logs; 3: second differencing, no logs; 4: only logs; 5: first differencing, logs; 6: second differencing, logs.

QUARTERLY DATA	log/diff	time aggreg
Gross value added at constant prices (mio euro) - Agriculture, hunting, forestry and fishing	5	4
Gross value added at constant prices (mio euro) - Total industry (excluding construction)	5	4
Gross value added at constant prices (mio euro) - Construction	5	4
Gross value added at constant prices (mio euro) - Wholesale and retail trade, repair of motor vehicles, motorcycles and personal and household goods; hotels and restaurants; transport, storage and communication	5	4
Gross value added at constant prices (mio euro) - Financial intermediation; real estate, renting and business activities	5	4
Gross value added at constant prices (mio euro) - Public administration and defence, compulsory social security; education; health and social work; other community, social and personal service activities; private houselholds with employed persons	5	4
Gross value added at constant prices (mio euro) - All NACE branches - Total	5	4
Gross domestic product at market prices - CLV2000	5	4
Gross value added - CLV2000	5	4
Final consumption expenditure: household and NPISH - CLV2000	5	4
Final consumption expenditure: general government - CLV2000	5	4
Gross fixed capital formation - total - CLV2000	5	4
Exports - total - CLV2000	5	4
Imports - total - CLV2000	5	4

Notes: log/diff: 1: unchanged; 2: first differencing, no logs; 3: second differencing, no logs; 4: only logs; 5: first differencing, logs; 6: second differencing, logs. Time aggregation: 1: *I(0)* stocks; 2: *I(0)* flows; 3: *I(1)* stocks; 4: *I(1)* flows.

European
University
Institute
European University Press