



ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

**The Role of Indicator Selection in
Nowcasting Euro Area GDP in
Pseudo Real Time**

Alessandro Girardi
Roberto Golinelli
Carmine Pappalardo

Quaderni - Working Paper DSE N°919



The Role of Indicator Selection in Nowcasting Euro Area GDP in Pseudo Real Time

Alessandro Girardi^{*}, Roberto Golinelli^{**} and Carmine Pappalardo^{*}

Abstract

Building on the literature on regularization and dimension reduction methods, we have developed a quarterly forecasting model for euro area GDP. This method consists in bridging quarterly national accounts data using factors extracted from a large panel of monthly and quarterly series including business surveys and financial indicators. The pseudo real-time nature of the information set is accounted for as the pattern of publication lags is considered. Forecast evaluation exercises show that predictions obtained through various dimension reduction methods outperform both the benchmark AR and the diffusion index model without pre-selected indicators. Moreover, forecast combination significantly reduces forecast error.

JEL Classification: C53, C22, E37, F47

Keywords: Euro Area GDP forecasts, Bridge and Factor Models, Indicators' selection and pre-screening, Forecasting ability.

This draft: 13/01/2014 16:17

^{*} Division of Economic Studies and Econometric Forecasts, Italian Institute of Statistics, ISTAT, Rome.

^{**} Department of Economics, University of Bologna, Italy; corresponding author; roberto.golinelli@unibo.it

1. - Introduction.....	2
2. - The state of the art in short run modelling for GDP forecasting.....	5
3. - Methodological issues: indicator selection, diffusion index modelling and aggregation.....	9
3.1. Indicators' selection.....	9
3.2. Modelling the modeller.....	12
3.3. Aggregation vs disaggregation.....	13
4. - A model taxonomy based on indicator selection.....	15
4.1. AR, BM and FM specifications.....	17
4.2. The PFM setup.....	17
5. - Data and experimental design of the forecasting exercise.....	20
5.1. The dataset.....	20
5.2. Managing the publication lag.....	22
6. - Empirical results.....	26
6.1. A look at the selection procedure: the role of shrinkage.....	27
6.2. Forecasts from baseline specifications.....	28
6.3. Forecast averaging and combination.....	31
7. - Extensions to the baseline results.....	33
7.1. GDP forecasting from the demand side.....	34
7.2. Alternative FM settings.....	36
8. - Conclusions.....	38
References.....	42
Tables and figures.....	47
Appendix – The indicators dataset.....	55

1. - Introduction¹

The process of European integration made it clear right from the beginning that political and economic decisions would increasingly be based on aggregate European figures. Decision makers in various sectors of the economy (business, government, the central bank, financial markets, etc.) base their choices on an early understanding of the state of economic activity, usually measured by GDP, which is generally considered the best variable to capture aggregate economic conditions. Together with the establishment of the European Monetary Union (EMU), it is not therefore

¹ Paper presented at the 33rd International Symposium on Forecasting of the International Institute of Forecasters, Seoul, June 2013; at Experts' Group "Economic Forecasts" of the European Commission, Bruxelles, October 2013; and at the 14th IWH-CIREQ Macroeconometric Workshop on "Forecasting and Big Data", Halle (Saale), December 2013. We are grateful, without involvement, for comments and feedback to Valentina Corradi, Katja Drechsel, Domenico Giannone, Michael Graff, Hyun Hak Kim, Esther Ruiz, Norm Swanson and meetings' participants. The opinions it expresses are the authors' only, and do not necessarily reflect those of the institutions they work for. We would like to thank PRIN for funding (R. Golinelli).

surprising that a number of papers have been devoted to the issue of forecasting developments in euro area economic activity in the short term (see e.g. Baffigi et al. (2004), Bodo et al. (2000), Forni et al. (2001), Grassmann and Keereman (2001), Marcellino et al. (2003), and Rünstler and Sédillot (2003)).

However, quarterly GDP data are available only with a delay, which weakens their role in short-term policy decision making. For this reason, market participants and policymakers pay high attention to short-term information from indicators, which are promptly available at monthly frequencies. These indicators can be used to assess the state of the economy either in a purely qualitative manner or by incorporating their information into econometric models to draw an early picture of the evolution of current economic activity. Since the seminal papers of Trehan (1989), Klein (1990), and Rathjens and Robins (1993), a number of models have been introduced to predict GDP by using monthly indicators. Since then, results in the literature have unambiguously concluded that the use of indicators, available for some/all months of the quarter to be forecast, provides considerable improvements in forecasts of the not yet available current-quarter GDP growth. Furthermore, these improvements grow with the flexibility of the modelling approach in embodying the newest information as soon as it is released (for a recent example, see Kitchen and Kitchen, 2013).

The aim of this paper is to look in greater depth at the results of the literature on euro area GDP nowcasting, by introducing a procedure to extract the most valuable indicators from a flow of data issued every month. Thus, this work mainly belongs to the strand of research devoted to the selection of short-run indicators in pseudo real-time (because in this way we are also closer to the existing literature), and to the issue of forecast combination. More precisely, we are contributing to the literature on the nowcasting of euro area GDP in three ways.

Firstly, targeting predictors (by pre-selecting information through pre-screening) is an effective way to improve forecast performance, confirming the findings of Bai and Ng (2008),

Bulligan et al. (2012) and Kim and Swanson (2013). However, we should add here the results of the sensitiveness analysis to the number of targeted indicators which is driven, in the training sample of the pre-screening procedure, by the quantiles of the empirical probability distribution of each indicator to be picked by a set of six data reduction (one hard- and five soft-thresholding) rules. Secondly, we provide new evidence of the potential improvements in euro area GDP forecasting ability by averaging and combining forecasts from alternative approaches of pre-screening the indicators, see e.g. Hendry and Clements (2002), Stock and Watson (2004), Costantini and Pappalardo (2010), Kisinbay (2010), and Costantini and Kunst (2011). Thirdly, we assess the effectiveness of forecasting GDP from the supply and from the demand side, i.e. from a (possibly pre-screened) factor model approach which combines indicator information *versus* another in which forecasts of GDP expenditure items are combined, in line with the work of Huang and Lee (2010), and Hendry and Hubrich (2011).

This paper is set out as follows. Section 2 surveys the state of the art regarding GDP forecasting/nowcasting literature, with particular attention to euro area GDP. Section 3 introduces and discusses the main methodological issues of the paper - indicator selection, automated modelling, and aggregation - in the unifying context of the diffusion index models. Section 4 lists the taxonomy of the models used in the forecasting exercise on the basis of the underlying process of indicator selection and amount of disaggregate information used. Section 5 describes the data and design of the pseudo out-of-sample forecasting exercise. Section 6 reports the baseline empirical results in terms of the euro area GDP forecasting ability of classical bridge and factor models together with models that exploit alternative amounts of indicator information through pre-screening rules. In Section 7, we compare the forecasting ability of single aggregate models combining information with the combination of forecasts obtained from different models for GDP demand components, and assess the solidity of our baseline findings to some alterations of its modelling settings. Section 8 concludes.

2. - The state of the art in short run modelling for GDP forecasting

From the stream of results available in the GDP forecasting/nowcasting literature applied to virtually all the countries and areas of the world, three basic questions clearly emerge: (i) how to optimally select the indicators? (ii) what is the impact on GDP predictability of the real time nature of the forecasting activity? (iii) what is the level of aggregation at which it is better to model the relationships between forecast targets and indicators?

Regarding the indicator selection issue (i), since Klein and Sojo (1989) the extraction of reliable signals from noisy high frequency indicators has been carried out via two main routes: empirical indicators and bridge models. The empirical indicators approach lead to the development of factor-based models (FM; for a survey see Stock and Watson, 2006) which summarise all the available information into the extraction of some common factors from the full set of indicators. Bridge models (BM) link forecast targets to "suitable" indicators, selected *a priori* on the basis of researcher experience and statistical inference (see Golinelli and Parigi, 2007).²

As far as euro area GDP is concerned, Angelini et al. (2011), Banbura and Rünstler (2011), Marcellino et al. (2003), and Rünstler et al. (2009) assess the predictive performance of alternative FM nowcasts, while Baffigi et al. (2004), Diron (2008) Grassmann and Keereman (2001), and Rünstler and Sédillot (2003) follow the BM approach. In all these papers, the ability of both BM and FM approaches in nowcasting euro area GDP is assessed against simple benchmark time series models.

In spite of their widespread usage, both BM and FM approaches are subject to criticism. BM may appear excessively *ad hoc* because of the "incredible" exclusion restrictions underlying the list of the pre-selected indicators; FM may be biased by unbalanced sources of information (see Boivin

² These two routes were first compared in their ability to forecast the US economy in the short run by Klein and Ozmucur (2008). Then, other studies extended the FM-BM comparison to other countries/areas: Bulligan et al., (2010) for Italy, Antipa et al. (2012) for Germany, and Brunhes-Lesage and Darné (2012) for France

and Ng, 2006). In fact, the main requirement emerging from the asymptotic properties of the FM approach (such as factor estimators, structure and convergence to optimal forecasts) is that the sources of common dynamics remain limited as the number of cross-sections increases to infinity. Boivin and Ng (2006) question exactly this point, and argue that, in many practical applications, the (growing) size and (unbalanced) composition of the indicators can negatively affect factor estimates.

However, some papers have recently shown that the use of factors extracted from fewer but informative indicators can yield better forecasts than those obtained using large indicator datasets. In the literature, alternative techniques and methods have been introduced to perform this information reduction: targeting indicators with thresholding rules (Bai and Ng, 2008, Schumacher, 2010, Bulligan et al., 2012, and Kim and Swanson, 2013), estimating weighted principal components and preselecting indicators with rules that eliminate irrelevant information (Boivin and Ng, 2006, and Caggiano et al., 2011), estimating factors under a sparse prior (e.g. Kaufmann and Schumacher, 2013), and selecting one "representative" indicator of each category in which the large panel can be classified (Alvarez et al., 2012).

As for the real time nature of the data and of the forecasting activity, issue (ii), it impacts on the way forecasting ability experiments should be conducted. Specifically, two main aspects are addressed: (ii.a) experiments should mimic the actual situation faced by forecasters in terms of the schedule of data releases and thereby the availability of monthly indicators; (ii.b) experiments should use data actually available at the time the forecast was made, i.e. they cannot exploit the latest available (revised) series of both GDP and monthly indicators (see Croushore and Stark, 2001, and Pesaran and Timmermann, 2005).

Considering aspect (ii.a), almost all the papers of the euro area literature carefully account for the timing of the data releases (see e.g. Giannone et al., 2008, for FM, and Diron, 2008, for BM). Furthermore, some studies show that the depth in accounting for the timing of data releases

plays a crucial role in forecast evaluation, especially to assess the role of survey indicators (see Angelini et al., 2011, Banbura and Rünstler, 2011, Giannone et al., 2009, and Rünstler et al., 2009). A comprehensive study on the marginal information embodied in each monthly indicator update can be found in Drechsel and Maurin (2011). A related aspect to data issues is that GDP and indicator data are mixed-frequency - usually, quarterly and monthly - and asynchronous, i.e. released with different publication lags. The majority of studies simply convert all the data at the lower available frequency by taking quarterly averages of monthly indicators, and the ragged-(or jagged-) edge nature of the data requires that missing monthly observations for the quarter to be forecast are predicted usually with univariate autoregressive models; on this, see McGuckin et al. (2007).³ Camacho and Perez-Quiros (2010), Camacho et al. (2012), Ferrara et al. (2010), Giannone et al. (2009), Kuzin et al. (2009) are notable exceptions, as they respectively use approximate Kalman filter models, Markov-switching dynamic factors, non parametric methods, mixed-frequency VARs, and MIDAS regressions of Clements and Galvão (2008).

As far as aspect *(ii.b)* is concerned, a large part of the evidence about forecasting ability is based on the latest available data because of the dearth of data vintages. Exceptions are few and are always conducted in data parsimonious environments, such as the BM approach of Diron (2008) and Ferrara et al. (2010), and the FM approach of Camacho and Perez-Quiros (2010). Therefore, the literature mainly disregards the implications of data revisions although, since Diebold and Rudebush (1991), it is well known that the use of the latest available data can significantly overstate the forecasting performance of models based on preliminary and unrevised data. They also note that the more dangerous insurgence of spurious forecasting abilities emerges in contexts where leading

³ Another strategy with FM would be the use of more specialised techniques are rooted in the FM approach, see e.g. the shifting operator (by which all indicators with missing observations for the latest month are shifted in time so as to have a balanced panel), and the expectation-maximization algorithm of Stock and Watson (2002a). Additional references and descriptions can be found in Bulligan et al. (2010).

and coincident indexes are used as indicators (Heij et al., 2011, reassessed this finding). However, if the aim of the pseudo real time exercise is to compare the relative forecasting ability of alternative approaches (rather than to measure absolute forecasting ability), then their ranking should not be greatly affected by neglecting data revisions, as shown for both BM and FM contexts in Bernanke and Boivin (2003), Golinelli and Parigi (2008), Schumacher and Breitung (2008), Bulligan et al. (2010).

The choice of the level of aggregation, issue (iii), is about inspecting whether it is better to model aggregate or disaggregate information. Disaggregation can be either about functional GDP components, as in BM by sector (Hahn and Skudelny, 2008, and Burriel, 2012) and BM by expenditure (Baffigi et al. 2004), or about modelling single countries belonging to the euro area in both BM and FM (see Marcellino et al., 2003). When the aggregate variable is the target of the forecast, one way of increasing the information set is to disaggregate the target in its components. If this enlargement is relevant, we would expect an improvement in forecasting ability from the disaggregation. If disaggregate data are modelled, aggregate euro-wide forecasts may be obtained by aggregating disaggregate FM and BM forecasts. Therefore, as first argued in Baffigi et al. (2004), and Hubrich (2005), and generalised in Clements and Hendry (2011), the choice of the "optimal" aggregation level cannot be stated *a priori*, but depends on the properties of the disaggregated specifications and might also vary over the forecast horizon. Results in Marcellino et al. (2003), and Baffigi et al. (2004) suggest that forecasts obtained by aggregating forecasts from country-specific models are more accurate than forecasts from aggregate (area-wide) models, while functional disaggregation seems to be less useful in improving area wide GDP forecasts (the latter point is also confirmed by the results in Hubrich, 2005 regarding HICP forecasts).⁴

⁴ The heterogeneous statistical information available for the pre-euro periods, structural breaks in area-wide GDP/indicator relationships and short samples of available data are the main candidates to explain early results in favour of the use of simple models by country. However, the FM advantage of using disaggregate information from

3. - Methodological issues: indicator selection, diffusion index modelling and aggregation

In this Section we will deepen the three main methodological aspects emerging from the previous survey: the selection of suitable indicators, the procedures to emulate the modeller in pseudo real time and the amount of information exploited by modelling the target at aggregate/disaggregate level.

3.1. Indicators' selection

Equation (1) is the ARDL (p, q_j) representation of the dynamic relationships between the logs of the quarterly target variable y_t (in our case the euro area GDP) and N quarterly averages of monthly indicators x_{jt} ($j = 1, 2, \dots, N$).

$$\alpha(L)y_t = \sum_{j=0}^N \beta_j(L)x_{jt} + \varepsilon_t \quad (1)$$

where $\alpha(L)$ and $\beta_j(L)$ denote lag polynomials of order p and q_j , respectively, and ε_t is the random error shock, assumed to be iid. Given that indicators x_{jt} are published in advance of the release of y_t , equation (1) can be used to predict the latter from indicator data for the same quarter (GDP nowcast).

The number N of indicators x_{jt} entering the list of explanatory variables of equation (1) is crucial for the choice of its empirical implementation (for a recent and extensive survey of variable selection methods see Ng, 2013). In forecasting practice, the availability of a very large number of timely monthly indicators (the Appendix lists the 259 indicators used in this study) leads to a curse of dimensionality which prevents the direct estimation of α and β_j parameters, unless just a few (k) of these are pre-selected with experience and "art" by the researcher, as implied in the BM

large datasets could improve the performance of aggregate GDP models by exploiting factors that embody disaggregate information (see Hendry and Hubrich, 2011).

approach. Although unavoidably arbitrary, this extraction of $k \ll N$ indicators has proved to be quite effective in forecasting euro area GDP in the short-run with BM, see e.g. Rünstler and Sédillot (2003), Baffigi et al. (2004), and Diron (2008).

An alternative approach, based on statistical procedures, and not on *ex ante* selection, could be that of Bayesian model averaging and selection (see Koop and Potter, 2004, for an application in the field of macroeconomic forecasting with factor-based models). In this context, however, De Mol et al. (2008) show that Bayesian shrinkage techniques deliver predictions that are highly correlated to principal component forecasts, FM. Therefore, in this paper we will concentrate on BM and its FM polar approach.

With FM, the (arbitrary) role of the researcher with BM is displaced by the statistical extraction of few (again, for simplicity, k) factors from a large-panel of N (standardized) indicators (either stationary or transformed to achieve stationarity), by assuming that the following factor structure is admissible: $x_{jt} = \lambda_j^{0'} F_t^0 + e_{jt}$, where F_t^0 is a $k \times 1$ vector of factors common to all variables, $\lambda_j^{0'}$ is the vector of factor loadings for x_{jt} , and e_{jt} is an idiosyncratic error.⁵

Although the FM approach does not require the explicit selection of k indicators, the researcher has to address issues which give rise to a wide-ranging set of FM with potentially different (implicit) indicator selection: (i) the estimation of factors with either static or dynamic principal components (see respectively Stock and Watson, 2002a and 2002b, henceforth SW, and Forni et al., 2005, FHLR); (ii) the forecast of the target variable, given the factor estimates, which can be either parametric (associated with the static approach of SW) or non-parametric (associated

⁵ Luciani (2014) assess the forecasting ability of models which also exploit the idiosyncratic dynamics. Results show that when forecasting aggregate variables (such as GDP), which respond primarily to pervasive (i.e. macroeconomic) shocks, accounting for non-pervasive (i.e. idiosyncratic) shocks is not useful. This fact supports our choice of focusing on only pervasive shocks, i.e. on F_t^0 only.

with the dynamic approach of FHLR); (iii) the determination of k to estimate the unknown number of factors r adopting optimality criteria (see e.g. Bai and Ng, 2002), or the share of the variance explained (in any case, all these approaches estimate k without reference to the target variable).⁶

The issue of determining k is related to one of the biggest issues with the large-panel FM approach that was raised in Boivin and Ng (2006): the FM problem of extracting factors "blindly", i.e. without taking into consideration the properties of the variables to be forecast with those factors. In fact, as N increases, it can be that a number of indicators will be not very highly correlated with the target variable and, after their inclusion in the panel, the average common component to explain the target will be smaller and/or the residual cross-correlation will be larger. As a result, the best factors driving the variable to be forecast can be dominated by other less useful factors in oversized panels. Therefore, the factor extraction from a subset n of "targeted indicators" (with $n < N$) tailored to a specific variable can bring gains in terms of forecasting accuracy. In fact, the pre-screening activity to retain the most valuable indicators brings into the FM approach additional information that enters all the modelling stages.

Pre-screening before factor extraction leads to a mixed approach between BM and FM, which we will refer to as pre-screened FM (PFM) from now on. As with the BM approach, PFM pre-selects several indicators. As compared to BM, the selection process is less arbitrary, because it can be based on, for e.g., statistical thresholding rules, and less extreme, as $k < n < N$. Like the FM approach, PFM extracts k factors from a panel of indicators, but the panel of targeted indicators is more likely to carry useful information with which to forecast the target than in FM, where a large panel of unbalanced sources of information is used. As with BM and FM, recent literature has

⁶ More details and discussion of these aspects are in Bulligan et al. (2010), and D'Agostino and Giannone (2012). These options, together with the choice of the balancing strategy discussed above (shifting, autoregressive and other specialised algorithms) brings to alternative FM of which much of the literature has supported the usefulness in forecasting euro area GDP; see e.g. Marcellino et al. (2003), Rünstler et al. (2009), Angelini et al. (2011).

documented the advantage over simple benchmarks of forecasting euro area GDP with the PFM approach (see Caggiano et al., 2011; Bulligan et al., 2012). After selection (with BM) or extraction (with FM and PFM) of a vector of predictors P_t , FM, PFM and BM can be all represented by the same dynamic quarterly relationship (of order p and q_j) between the dependent variable y (usually in log-differences, to prevent non-stationarity problems) and the vector of predictors P_t :

$$\Delta y_t = \sum_{i=0}^{q_j} \Gamma_i P_{t-i} + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \varepsilon_t \quad (2)$$

where Γ_i and α_i are parameters, and ε_t is the random error shock. Equation (2), belonging to the framework of the diffusion index (or factor augmented) models, explains short run fluctuations in y_t in part through its co-movements with P_t , in part through the idiosyncratic and unpredictable shocks ε_t and in part through dynamic propagation (leading) mechanisms and inertia, represented by lags of both P_t and y_t .

3.2. Modelling the modeller

Independently of the way the predictors P_t are obtained, the use of model (2) must cope with the problem of neutralizing the advantage of knowing how the data look ex-post (see Stark and Croushore, 2002). The modelling approach in which a new specification is chosen and estimated before each forecast round is called adaptive, while the non-adaptive alternative implies only the estimation of the parameters without changing the equation specification. Swanson and White (1997) have shown that adaptive models, estimated over rolling windows, perform better than fixed-specification models, since they may limit the effects of heterogeneity over time and structural breaks. In addition, the dimension of the window must balance the trade-off between the number of observations (the more they are, the more efficient the inferences) and the risk of

parameter structural breaks (increasing with the sample dimension), see Stock and Watson (1996) and Giacomini and White (2006).

In this context, automated modelling and inference is a viable option, because it is based on predetermined rules and guards against future information creeping into the model specification and the pseudo ex-ante forecasts. As in other analyses emulating the real-time behaviour of the researcher through the LSE general-to-specific modelling strategy (see Banerjee et al., 2005, Golinelli and Parigi, 2008, 2013; Barhoumi et al., 2011; Bulligan et al., 2012; Brunhes-Lesage and Darné, 2012), in this paper we started from a general dynamic equation (2), then reduced its complexity by eliminating statistically insignificant regressors and checked that the resulting model satisfies a number of misspecification and parameter constancy tests.⁷ Although the reduced model preserves the same features (in terms of information) of the initial unrestricted model, this way of “modelling the modeller” is not without costs: the application of the automatic procedure neglects the researcher’s skill and, for this, is bound to worsen the performance of the final model. In other words, it has to be taken as a sort of “lower bound” of the researcher’s modelling ability, which is one of the main ingredients of the “art” of forecasting.

3.3. Aggregation vs disaggregation

Regarding the choice of the amount of information to be exploited by modelling the target at aggregate/disaggregate level, it is clear that one possible way of increasing the information set is to disaggregate the target in its components. If the disaggregate information is relevant, we would expect an improvement in forecasting ability over the use of aggregate information. Symmetrically, aggregating the data amounts to throwing away information but, if the amount of noise in the

⁷ Namely, the Lagrange Multiplier test for serial correlation in the residuals up to 5 lags, normality tests, tests for quadratic heteroscedasticity between regressors, and Chow in-sample predictive failure test on 90% of the sample.

disaggregate data swamps any additional signal, parameter estimates are less reliable in the disaggregate context, and incremental uncertainty leads to less accurate forecasts.⁸

On the basis of the alternatives listed in Hendry and Hubrich (2011), the present paper will consider three ways to tackle the aggregation issue in forecasting euro area GDP: *(I)* aggregate models using aggregate information for direct aggregate GDP forecasts; *(II)* component models using disaggregate information for direct disaggregate forecasts (for e.g. consumption and investment) and their aggregation in indirect GDP forecasts; *(III)* aggregate models using disaggregate information for direct aggregate GDP forecasts.

If the predictors are pre-selected (as in the BM approach), equation (2) for euro area GDP target represents case *(I)*: some indicators (such as industrial production) can be seen as the aggregate information that the researcher feels is needed to directly forecast GDP. On the other hand, if, for e.g., two equations (2) are devoted to explaining consumption and investment targets with pre-selected retail sales (for consumption) and business confidence (for investment), the resulting disaggregate BM forecasts - case *(II)* - need an aggregator equation to estimate the weight of consumption and investment and convert their forecasts into GDP forecasts.⁹ In the euro area literature, Rünstler and Sédillot (2003) and Diron (2008) report examples of case *(I)*, i.e. aggregate "supply-side" BM for forecasting GDP through aggregate indicators. Examples of case *(II)* are in Baffigi et al. (2004), which document a disaggregate "demand-side" BM, and in Hahn and Skudelny (2008) for a "sectoral/supply-side" BM.

⁸ In the words of Grunfeld and Griliches (1960) « [...] the aggregate equation may explain the aggregate data better than all micro equations combined if our micro equations are not "perfect". Since perfection is unlikely, aggregation may result in a "net gain"» (p. 10).

⁹ See e.g. Baffigi et al. (2004), and Golinelli and Parigi (2007). Regarding the performance of the aggregator function, it is worth noting that the change in GDP measurement by statistical agencies from fixed to chained bases implies that the component weights (to be estimated) also depend on relative prices of the demand components and this fact potentially requires further extensions of the disaggregate information set.

If the predictors are extracted from a large panel of indicators (as in the FM approach), equation (2) for the euro area GDP target might also be seen as an example of case *(III)*. Factors extracted through FM combine disaggregate information about the target components that are likely to be included in the indicator data-set and bring this disaggregate information into the aggregate equation (2) through predictors measured by such factors (see Hendry and Hubrich (2011)). For this reason, it is not useful to make disaggregate FM for GDP components given that, as seen above, factor extraction does not explicitly account for the model's target variables.

Given that the PFM approach shares features from both BM and FM, it can represent all the cases listed above. Different PFM interpretations emerge depending on how many of the N indicators are excluded from the targeted subset of n indicators in the pre-screening phase. When n is closer to k than to N (i.e. many indicators are discarded) PFM can be seen as an example of case *(I)*, as the pre-screened selection of few indicators brings PFM features closer to supply-side BM, while when the larger and closer n is to N , the more PFM can be seen as an example of case *(III)*, as its features are closer to those of FM. Finally, if in the pre-screening phase different "disaggregate" target indicators are pre-screened for each GDP component, the resulting system of PFM equations (2) (together with GDP aggregator equation) is an example of case *(II)*.

4. - A model taxonomy based on indicator selection

Although equation (2) is not the only way to forecast with FM (as this is the so called parametric approach), its use in forecasting GDP allows for the comparability of alternative forecasts, as the modelled link between predictor and target is always the same, independently of the different ways the predictors (pre-selection or factors) are obtained. In order to further improve comparability (also

with the existing results in literature), in this paper (a) we will estimate factors (in both the FM and PFM approaches) by following the SW method only and (b) we will use latest available data.¹⁰

The former choice was made because we are not interested in comparing the forecasting ability of static (SW) and dynamic (FHLR) methods to estimate factors, as this has already been accomplished in D'Agostino and Giannone (2012) but rather our aim is to explain how different selection approaches (from BM to FM through PFM) impact on forecasting performance.¹¹ The second choice is because all the literature about forecasting with factors is based on the last vintage of revised data. As D'Agostino and Giannone (2012) also acknowledged, though we are aware that by neglecting data revisions we run the risk of overestimating the forecasting accuracy of each method, this should have a limited effect on the assessment of the relative performance of BM and FM. Given that data revision to indicators in the Appendix and to GDP can be seen as either innovation shocks or location shifts to all modelling approaches, they are expected to affect both absolute and relative forecast accuracy measures in a very similar way (Hendry and Hubrich, 2011).

¹⁰ For this, we devote particular care to excluding composite leading and coincident indexes (CLI) from our large indicator dataset (see the Appendix), as CLI are affected by a spurious forecasting performance when using revised data. In fact, CLI revisions are not only due to statistical adjustments of their components as soon as more complete historical data are available, but components are often added and eliminated *ex post* to improve their performance retrospectively (Diebold and Rudebusch, 1991). In addition, we account for the source of revision related to the indicators statistical treatment before their use (check for outliers, forecast to balance the panel, quarterly aggregation, and seasonal adjustment). This procedure, described below, exactly mimics that adopted in the genuine out-of-sample forecasting practice.

¹¹ D'Agostino and Giannone (2012) found that SW and FHLR methods perform similarly and produce highly collinear forecasts, corroborating our choice of focusing on the SW method alone as the "representative method" of the FM extraction of factors in order to simplify result reporting, while, in the light of the results in Boivin and Ng (2005), the SW method (i.e. our choice here) seems to perform systematically better when more complicated but realistic error structures are considered. In addition, Alvarez et al. (2012) have shown that the FHLR method to estimate factors results in similar problems as SW when the number of indicators in the data-set is large.

4.1. AR, BM and FM specifications

In the light of equation (2), the univariate AR model is the most natural benchmark model for our study. AR forecasts are based on an assumption of zero indicator usefulness because it does not use them at all. The AR benchmark is nested in equation (2) under the assumption that $\Gamma_i = 0, \forall i$ or equivalently $k = 0$. In this study, the AR lag length p (measuring GDP growth inertia through α_i parameters) has been selected on the basis of the Schwarz criterion.

With reference to the three cases regarding the use of aggregate/disaggregate information, the benchmark AR can be interpreted as the simplest example of case (I): a model which uses only past aggregate GDP information to produce aggregate GDP forecasts. In this context, the inclusion of *ad hoc* pre-selected indicators in the benchmark AR leads to the classical BM, where $\Gamma_i \neq 0$ for predictors P_t . In this paper, BM pre-selection involves the same euro area quantitative indicators ($k = 4$) used in the BM of Rünstler and Sédillot (2003): Industrial production, Production in construction, the Retail sales indicator and Car registrations. In this way, we use a pre-selection dated prior to the beginning of the out-of-sample forecast exercise in 2008 (see next section).

The FM approach is at the other extreme of the selection process. Following SW, predictors in equation (2) are obtained as the first $k = 3$ principal components extracted from a large-panel of $N = 259$ indicators (described in the next section and in the Appendix; technical details are reported in next section too).¹²

4.2. The PFM setup

The PFM approach in this paper pre-screens indicators to extract factors from a reduced dataset which only includes specific (targeted) indicators, i.e. selected in a way that they are the most tailored to the target variable. PFM is between BM and FM as it is less arbitrary than BM in the process of indicator selection and, contrary to the FM approach, factors are – if they are targeted –

¹² In Section 7.2 we assess the robustness of the results to alternative approaches to determine the number of factors k .

likely to provide useful information with which to forecast the variable of interest. Regarding PFM use of aggregate/disaggregate information, the pre-screen for the best indicators to explain the target puts PFM between case (I) and case (III), depending on the nature of the indicators screened.

We have implemented PFM using six data reduction methods: one hard- and five soft-thresholding rules. By applying the hard-thresholding rule, an indicator is selected if the significance of its correlation coefficient with the target variable is below the 5% threshold, its main drawback being the tendency to select indicators highly collinear with the target and disregard the information content of the other indicators. The soft-thresholding rules consist of criteria of ordering and selecting indicators on the basis of a minimization problem of the form: $\min_{\beta} [\Phi (RSS) + \lambda\Psi(\beta_1, \dots, \beta_j, \dots, \beta_n)]$, where RSS is the Residual Sum of Squares of a regression of the target variable on the retained indicators, λ (the Lagrange multiplier) is the shrinkage parameter (the higher the λ , the higher is the penalty for having extra regressors in the model), Φ and Ψ are functions of RSS and the regression coefficients ($\beta_j, j=1, \dots, n$). With soft-thresholding rules, the cross-correlations among indicators are taken explicitly into account when minimizing the loss function, with the disadvantage that n cannot be too large in relation to the available time span. Depending on the functional form of Φ and Ψ , specific soft-thresholding rules are obtained. Here, we focus on five of them: Least angle regressions (*Lars*); Least absolute shrinkage selection operator (*Lasso*); Elastic net estimator (*Enet*); Forward stagewise linear regressions (*Fstage*); Ridge regression (*Ridge*).¹³

In order to implement an algorithm for the definition of targeted indicators using the one hard- and five soft-screening rules listed above, we have followed three steps.

Step 1. We have partitioned the full sample of T observations into a first portion T_1 used for model training, and another $(T-T_1)$ left to run the pseudo *ex ante* forecast exercise. In order to pre-

¹³ An overview of these methods can be found in Bai and Ng (2008) and Ng (2013), among others.

screen indicators also on the basis of the leading/coincident properties of the target, the initial dataset of N indicators has been expanded to include their lags up to order l ; therefore, the full panel of indicators is made up of $N \times (l+1)$ variables and T_l observations over the training-sample.

Step 2. For a given rolling window of size T_w , ($T_w < T_l$), each of the 6 screening rules listed above have been applied to the large initial dataset of $N \times (l+1)$ series to rank the targeted indicators.

Step 3. We have associated each of the $N \times (l+1)$ indicators and for each of the 6 screening rules a binary variable which takes value 1 if that indicator was ranked by that given method among the top 10 variables and assigned it a 0 otherwise. In addition, we have introduced a 7th screening rule ("union of all soft rules") that gives value 1 to the indicators that are ranked in the top 10 by at least one soft method.¹⁴

The algorithm is run over the first $T_l \square T_w+1$ rolling windows. As a result, for each indicator, 7 binary variables (one for each screening rule plus the "union of all soft rules") of $T_l \square T_w+1$ observations have been obtained. The sample means by indicator and binary variable, over all $T_l \square T_w+1$ runs, ranges between 0 (if the indicator was never selected by the specific screening rule) and 1 (if the indicator was always selected by the specific screening rule) and can be interpreted as the probability of each indicator of being selected. This probability is conditional on a given thresholding rule.

In order to assess the sensitiveness of GDP forecasting ability to the decreasing size of indicator data-sets, in this paper we have considered six quantiles of the empirical probability distribution of being picked by a thresholding rule: $Q = [\text{min}, 10^{\text{th}}, 25^{\text{th}}, 50^{\text{th}}, 75^{\text{th}}, 90^{\text{th}}]$ percentiles, where "min" means "indicator picked at least once by that rule in the $T_l \square T_w+1$ samples". Indicators belonging to the large panel also enter the targeted panel (i.e. are selected by one rule) if their

¹⁴ As we have used only one hard-screening rule, we have introduced this "union of all soft rules" rule (labelled in the following as "soft") as a sort of average of the five soft rules, in order to summarise the outcomes of all alternative pre-screening rules in two categories: hard and soft.

probability of being picked, conditional on that rule, is being larger than a given quantile Q . A sequence of six quantiles is set for each of the seven rules and, correspondingly, the number of selected indicators by rule decreases as the reference quantile increases.

Overall, for each quantile, we can identify seven different PFM whose factors are extracted from the corresponding pre-screening thresholding rules: *hard*, *lars*, *lasso*, *enet*, *fstage*, *ridge*, and *soft*. For example, if we set $Q = 90^{\text{th}}$ percentile (the highest in the six listed above), we select the lowest set of indicators (n) out of N and the seven models that exploit with factors these n targeted indicators will be respectively labeled: PFMhard90, PFMIars90, PFMIasso90, PFMenet90, PFMfstage90, PFMridge90 and PFMsoft90. Given that 6 quantiles times 7 pre-screening rules leads to 42 different PFM, in Section 5 we have summarised the outcomes by only reporting those of the 12 of them, i.e. PFMhard_{xx}, PFMsoft_{xx}, $xx = \text{min}, 10, 25, 50, 75, 90$) which respectively denote hard and soft (the union of all single soft rules) pre-screened FM.

The list of all the models used in this paper for nowcasting euro area GDP is reported in Table 1. Along the different rows, starting from the benchmark AR, models are listed by the growing number of indicators entering the forecast and by the decreasing degree of arbitrariness in the selection process.

Table 1 here

5. - Data and experimental design of the forecasting exercise

5.1. The dataset

The forecasting exercise has been performed using a large dataset of short-term indicators mainly concerning the euro area and the US economy.¹⁵ The complete set of indicators consists of 259 time

¹⁵ The inclusion of US figures aims at capturing the linkages related to the transmission of international business cycles, see e.g. Bodo et al. (2000).

series (a large part of which are at monthly intervals) ranging from January 1990 to December 2012.

The indicators reported in the Appendix can be allocated to the following four types of information:¹⁶

- (1) qualitative indicators from business tendency surveys (manufacturing, construction, services) and consumer surveys;
- (2) quantitative indicators from the supply and demand side;
- (3) financial variables: interest rate and stock market indices; nominal and effective real exchange rates;
- (4) price variables (consumer, energy and commodity prices).

The composition of the information set is reported in Table 2 in which indicator counts are allocated by type of information along the columns, and by geographic area, publication frequency and transformation along the rows.

Table 2 here

Qualitative indicators are about 60% of the total information set. A large number of them have been gathered from business and consumer surveys. As far as the former are concerned our dataset includes – as well as business confidence climates - the assessment on the current level of order-books (both overall and export), production, the stock of finished products, prices and employment. In addition, entrepreneurs' expectations on short-term trends in orders, production and prices have been considered in the main sectors of economic activity. Furthermore, firms' assessments of the main constraints on production activities are considered. As far as consumer

¹⁶ Most soft indicators, such as business and consumer surveys, are not subject to revision, while some of the quantitative variables (such as industrial production indices) are revised after the first release. For all these latter indicators, and for GDP and its demand components, we have used the latest available (revised) series for the reasons explained in Section 4.

surveys are concerned (for both the euro area and the US), the information set contains consumers' evaluation on the current and future economic situation of the general economy, their assessment of the financial situation of their own households (both during the past and the next 12 months), present and future saving intentions and planned durable goods purchases. Though the unemployment rate (the main labor market indicator) is a quantitative indicator, there are a number of qualitative indicators (on both employment and unemployment, such as unemployment expectations) that are drawn from the business and consumer surveys which belong to the group of qualitative indicators.

As for quantitative indicators, these are about 11% of the total information. Supply side quantitative indicators consist of industrial production indices (IP, both general and by main sub-sectors) and the unemployment rate. The retail sales index and car registrations are relevant demand-side quantitative indicators. Although they usually account for low shares in large-panels, quantitative indicators are BM favorite indicators.

Financial indicators make up about 24% of the total information. The financial data set includes several interest rates in the term structure both for the euro area as a whole and its most relevant economies (France, Germany, Italy, and Spain) and a set of stock price indices for the main sectors of economic activity. It also includes euro nominal exchange rates with the US dollar, the UK Pound and the Japanese Yen. Public finances have been accounted for through an indicator of public debt for euro area countries. Finally, price indicators are 5% of the total information. Here, we have considered the headline inflation rate and price indices for the main energy goods (crude oil, natural gas, gasoil) and commodities (such as wheat) which are traded internationally.

5.2. Managing the publication lag

In our forecasting exercise we have mirrored a predictor which updates quarterly forecasts for euro area GDP in each of the 63 months (the forecast rounds) over the time span 2008q1-2013q1. The real time dimension of the dataset is ensured by organizing data in such a way as to exactly mimic

the time schedule of actual data releases and thereby the availability of monthly/quarterly indicators along the same period (quarter). The ragged-edge issue due to the asynchronous release of indicators and GDP has been evaluated by assuming that each forecast round is carried out around the middle of each month when industrial production figures for the euro area are released.

According to the timeliness of the indicators, the 12 monthly forecast rounds of each year can be classified into three different vintages and, each forecast round is carried out using only one vintage, that corresponds to indicator availability. Table 3 details by vintage the share of indicators which are missing for at least one month in the quarter to be predicted in each vintage.

Table 3 here

First vintage ($v = 1$). For the March, June, September and December forecast rounds, only one month of quantitative indicators and 2-3 months of qualitative indicators are known for the quarter to be forecast (specific publication lags are in Table A.1 of the Appendix). This is the worst case information availability scenario. In fact, 51.7% of indicators are not available for all the months of the quarter to be predicted (see Table 3).

Second vintage ($v = 2$). For the January, April, July and October forecast rounds, two months of quantitative indicators and three months of soft ones are known for the quarter to be forecast. This is an intermediate case of information availability, as 12.7% of indicators are not available for all the months of the quarter to be predicted (see Table 3).

Third vintage ($v = 3$). For the February, May, August and November forecast rounds, all the indicators, both qualitative and quantitative, are available for all the months of the quarter to be forecast. This is the best case information availability scenario, often referred to as nowcast. In Table 3 it emerges that less than 1% of indicators are not available.

The growing information content of the three vintages (when v goes from 1 to 3, the share of the unavailable series drops from 50% to about zero) is one of the criteria in the assessment of the

role of the indicators in improving the GDP forecasting performance of models over the benchmark AR.¹⁷

The jagged-edge nature of the panel of indicators requires predictions of one/two missing monthly indicators in order to fill the quarter to be forecast.¹⁸ In the light of Table 3, financial variables are released daily and are always available for all the quarters to be forecast (i.e. they never need to be predicted).

Auxiliary AR models are used to project missing indicators by following a four-step procedure: data transformation, univariate autoregressive (AR) modelling, monthly extrapolation, quarterly averages and seasonal adjustment.

Step 1. The indicators are subject to transformations (such as logs and first-differences) to remove possible non-stationarity. Unit root tests have confirmed the appropriateness of the choices listed in Table 2. Then, outliers (defined as those points which are more/less than three standard deviations away from the mean of the series) are replaced by their sample average plus (minus) two standard errors of the remaining observations.

Step 2. All series from *Step 1* are modelled as parsimonious AR where the number of lags is chosen according to the Schwarz information criterion. We use adaptive models of rolling windows of 132 months (i.e. 11 years as in all the rolling windows of this paper, irrespective of data periodicity).

Step 3. The AR models estimated in *Step 2* are used to extrapolate the indicators over horizons that depend on both publication lags and the specific forecast round. As the objective of the paper is to perform one-step ahead GDP forecasts, the AR models are used to compute out-of-sample forecast usually at 1- and/or 2-steps ahead. As an example, business and consumer surveys

¹⁷ The sensitiveness of the results to alterations of indicator selection and pre-screening is the other one.

¹⁸ Although we refer to missing data extrapolation as if all indicators were at monthly intervals, the same steps are followed to predict quarterly indicators, when they are missing for the quarter to be forecast.

are never predicted through univariate models thanks to their prompt availability at the end of the reference month. By contrast, quantitative indicators are usually forecast using AR models to complete the monthly information for the quarter of interest.

Step 4. The balanced dataset of raw data is converted at a quarterly frequency by averaging monthly indicators. Time series are then seasonally adjusted using the X12 seasonal adjustment procedure.

This large dataset of quarterly data spans the period 1990q1-2013q1 and consists of 148 monthly rounds. As they are rolling over 11 years, each round spans 44 quarters and the first sample goes from 1990q1 to 2000q4 and the last from 2002q2 to 2013q1.¹⁹

We have split the 148 monthly rounds into two parts: the first 85 rounds (about 60% of our set of vintages) have been devoted to model training (the last training round #85th goes from 1997q1 to 2007q4) and the remaining 63 rounds focus on the assessment of pseudo real time forecasting ability over the horizon 2008q1-2013q1.

In order to train hard- and soft-thresholding rules to pick the targeted indicators for the alternative PFM listed in Table 1, the algorithm described in Section 4 runs on the first 85 rolling windows. As a result, we have obtained estimates of the empirical probability distribution of each indicator to be picked conditional on a given thresholding rule (overall, we used seven rules). The quantiles of these distributions have been used to pre-screen the targeted indicators and their number, i.e. by excluding many of them (if we pick indicators with a probability above the 90th percentile), or by excluding a few of them (if we pick indicators selected at least once in the training sample).

¹⁹ The total number of spans is obtained by accounting for the 3 different monthly vintages available for each quarter. Specifically, we have considered a full set of three vintages for each quarter of the 49 time spans (the former is from 1990q1 to 2000q4, the latter from 2002q1 to 2012q4) by 3 plus one additional vintage for the period 2002q2-2013q1.

The resulting panels of indicators, selected by alternative rules and quantiles, are the inputs of the SW method (to obtain the k factors/predictors in equation (2)) and denote the PFM listed in Table 1. Specifically, the predictors for both FM and PFM are the first three factors extracted by applying the SW principal components method and $q_j = 1$ (lags are used up to the first order). The restricted (targeted) datasets which emerge from the training phase are held fixed over the whole forecasting exercise. In the same way, if we extract factors from the large panel of indicators available at the end of Step 4 above, i.e. before the pre-screening procedure, we obtain the predictors for the (untargeted) FM listed in the last row of Table 1.²⁰

6. - Empirical results

The forecasting exercise consists in 63 one-step ahead forecasts of euro area GDP over the time span 2008q1-2013q1.²¹ To account for the different availability of indicators, three predictions for each quarter (indicated by vintage $\nu = 1, 2, 3$) have been performed. The evaluation of the forecasting ability refers to the models listed in Table 1 and discussed in Section 4, i.e. BM, FM and PFM. As anticipated in Section 4, for each of the 6 quantile thresholds, we report the outcomes only for the following two PFM: the hard thresholding rule (PFM*hard*) and the union of all the 5 soft thresholding rules (PFM*soft*).

²⁰ Given a sample of $N = 259$ indicators, we expanded it in order to account for the leading properties of indicators up to the fourth order. This choice delivers a panel of $259 \times 5 = 1295$ variables which both enter the factor extraction in FM and the pre-screening procedure in PFM.

²¹ It could be argued that the great recession of 2009 favours more those models which exploit indicators compared to the AR benchmark as during deep recession episodes indicators embody such news in the nowcast. Therefore, forecasting ability over the AR benchmark may be overstated. However, our aim was to rank the forecasting performance of BM, PFM and FM, i.e. of models that exploit (albeit to a different extent) indicator information.

6.1. A look at the selection procedure: the role of shrinkage

As a preliminary step we can investigate the outcome from the two thresholding rules. Table 4 presents the ten most selected indicators by hard- and soft-thresholding rule (along the columns), independently of the six quantile thresholds, Q .

Table 4 here

Four basic remarks ensue: (1) *PFMsoft* tend to select contemporaneous variables, whilst the column for *PFMhard* includes lagged variables up to the fourth order; (2) the selected targeted predictors are broadly the same for each vintage of the training period when selected according to the *PFMsoft* method: variables are included in the vast majority of the vintages of the training period, ranging from 52.9% (corresponding to 45 out of 85 cases) for the lagged value of consumers' intention of major purchases within next 12 months, to 100% (corresponding to 85 out of 85 possible cases) for production level in capital goods and production in manufacturing of basic metals; (3) opposedly, the *PFMhard* approach is not able to find similarly stable patterns of indicators' selection, and the range between the most and the less selected top regressors is between 15.3% and 41.2% (corresponding respectively to production level in buildings, and to retail business situation); (4) among the set of the most selected predictors, there are both euro area and US-related variables, giving support to our choice of including them in the information set.

A picture of indicators' counts coming from the hard and soft selection procedures by quantile threshold is in Table 5, where columns also detail the number of indicators by type.

Table 5 here

As expected, the selection related to the *PMFsoft* approach tends to include a very small number of regressors when compared to the *PMFhard* approach because - when screening each indicator - the selection process in the *PMFhard* setup disregards the information content of the other indicators. The selected predictors by type shows that both methods favour the inclusion of qualitative data in the set of regressors. In general, at the highest level of shrinking (that is at the

90% threshold), qualitative and quantitative indicators are preferred to financial and nominal ones, and price indicators are never selected by the soft thresholding rules.

6.2. Forecasts from baseline specifications

Forecast ability has been measured in terms of the root mean squared forecast errors (RMSE).²² Forecasting performance has been evaluated on the basis of each model's RMSE relative to that of the AR benchmark. The results have been presented in Table 6 and plotted in Figure 1. The latter depicts the three columns of Table 6 with a sequence of three bars of different colours (grey for $\nu=1$, white for $\nu=2$, and black for $\nu=3$, the nowcast) for each model progressively coded along the horizontal axis.

Table 6 and Figure 1 here

The forecasting models are listed in the first (*Code*) and second (*Model*) columns of Table 6. RMSE by vintage ν of each model relative to the benchmark are in the following three columns, together with the significance at 5 and 10% of the Diebold and Mariano (1995) test as adapted to small samples by Harvey et al. (1997).

The main results can be discussed as follows.

The forecast accuracy of each model significantly improves over the AR benchmark as indicator information increases, i.e. from the first to the third vintage. This result, not new in the literature, stresses the usefulness of accumulating indicator information through monthly updates of the data within the quarter to be forecast. However, this improvement from $\nu=1$ to $\nu=3$ does not show the same pattern for all the models.

²² We have also computed other measures of forecasting accuracy, i.e. mean errors (ME) and mean absolute errors (MAE) that qualitatively give us the same outcomes as those with RMSE (these unreported results are available upon request).

For example, when $v=1$, i.e. in the less favourable situation (because about half of the indicators are missing for the quarter to be forecast, see Table 3), the improvement of our models, in terms of relative RMSE over the AR benchmark (whose RMSE levels are in the first line of Table 6) is around 20-30%. BM performance is a remarkable exception: its improvement (about 10-15% over the AR RMSE) is below that of many PFM. This result, quite commonly found in the literature with different samples and specifications (see, for e.g. Golinelli and Parigi, 2007 and Bulligan et al. 2012) can be explained by considering that BM include pre-selected (hard) indicators which, when $v=1$, are not yet available for the largest portion of the quarter to be forecast. As soon as monthly quantitative information is released (i.e. when $v=2$ and $v=3$) BM performance improves considerably and its RMSE ratio over the AR benchmark drops from 0.87 (for $v=1$) to 0.45 (for the nowcast, $v=3$).

On the other hand, considering the nowcast, the RMSE ratio of BM (in terms of the AR benchmark) is better than FM (0.45 against 0.65). On the contrary, the latter outperforms BM when $v=1$. This outcome can be explained by the fact that FM include both qualitative and financial indicators whose availability in $v=1$ is much larger than that of quantitative ones (again, see Table 3).

These outcomes are shown in the two extremes of Figure 1: on the left (for BM coded #1), we have three bars that drop like a stone passing from grey ($v=1$) to black ($v=3$); on the right (for FM coded #26) we have three bars that do not improve as markedly as BM over vintages. But, considering the first vintage, FM relative forecast error is significantly lower compared to that of BM (bars in grey for $v=1$).

Moving from far left to right along the sequence of bars depicted in Figure 1 we have an increase in the number of PFM-exploited indicators, see Table 5 and the last column in Table 6. If the bars show monotonically increasing or decreasing patterns, useful insights for forecast practice can be learned.

A glance at Figure 1 shows two main patterns in RMSE ratios in terms of the AR benchmark. Firstly, the relative RMSE increases as the number of targeted indicators rises. Secondly, within each block, RMSE ratios seem to relate to quantile settings. Both patterns can be explained by the number of indicators entering the factor extraction and suggest that wider indicator information, if not properly selected, delivers more noise than signal to nowcast GDP. RMSE ratios for the PFM, computed for quantiles below the 25th percentile, turn out to be systematically above 0.4. PFM based on soft rules systematically outperform the PFM*hard* rule within all blocks. As far as the amount of exploited information is concerned, the hard thresholding rule retains much more - and less useful - information than the soft ones.

A main “message” coming from our results can be well summarised with the *maxim*: “less modelling and more shrinking” which constitutes an advance on the findings based on the hybrid combination “factor plus shrinkage” (see e.g. Kim and Swanson, 2013). In fact, higher-quantiles PFM (using fewer indicators) reduce the degree of freedom for the researcher and the unavoidable arbitrariness of subjective choices. In other terms, an interesting feature of employing soft-thresholding rules is that they combine both variable selection and parameter estimation and, in so doing, they relate to the variable to be predicted. Therefore, if this target-driven selection aims to retain very few variables (as in our 90th quantile case) the outcome of our hybrid approach is a sort of quasi-orthogonalized version of Bayesian regressions.²³ It coincides with BM in purpose, but exploits shrinkage techniques to reduce over-fitting problems in model specification (rather than the practitioner's art of forecasting necessarily constrained by working experience).

²³ Note also that orthogonal regressors improve the ability of our automatic general-to-specific modelling procedure of discovering the better diffusion index model specification.

6.3. Forecast averaging and combination

The large set of nowcasts obtained from pre-screened PFM with 5 soft-threshold rules (not reported but available upon request) suggests to compute averages and combinations of them in order to check whether is possible to further reduce the forecast error of the single soft-rules.

In this respect, Hendry and Clements (2002) have shown that the average of different model predictions is often associated with better forecasting performance than the single predictions mainly because averaging amends the effects of some forms of model misspecification. Their results are confirmed by Stock and Watson (2004) who also tried more complex techniques of forecast combination but without any significant improvement. Therefore, in Table 7 we have considered the simple average of the five forecasts, $PFM_{avg_{xx}}$, obtained from PFM based on the soft pre-screening rules (i.e. $PFM_{lars_{xx}}$, $PFM_{lasso_{xx}}$, $PFM_{enet_{xx}}$, $PFM_{fstage_{xx}}$ and $PFM_{ridge_{xx}}$ models).

With the aim of challenging - at least *ex post* - the average-view stated above, we have also applied the procedure proposed by Costantini and Pappalardo (2010) and Kisinbay (2010) which selects a subset of available forecasts for combination by exploiting the complementarity between RMSE and the encompassing test of Harvey, Leybourne and Newbold (1998) (hereafter HLN).²⁴ More in details, it provides a hierarchically ordered efficient selection of non-nested PFM forecasts. After this stage, several combination methods might be applied to get a hierarchical forecast combination (HFC).²⁵ The algorithm selects a given model with a greater forecasting accuracy *vis-à-vis* the rival models only if the former presents a greater information content. Using out-of-sample

²⁴ Given the available span of data, we cannot afford a second training sample to run *ex ante* the hierarchical forecast combination procedure (i.e. besides the first one needed to estimate the quantiles of the empirical probability distribution in the PFM approach) to run fully *ex ante* in the last part of the sample the procedure to assess the forecasting ability. Therefore, we must be aware that, below, the ability of the combination method can be overstated.

²⁵ In this paper, combination is obtained as the simple average of the selected models, as it has been shown to perform better than a set of several alternative approaches (Costantini and Pappalardo, 2010).

forecasts as inputs, the algorithm cross compares all forecasting models using the HLN test so as to eliminate the encompassed models and then averaging the remaining forecasts. The main steps of the hierarchical procedure can be summarised as follows:

- I) the RMSE of the out-of-sample forecast for each model are computed by using out-of-sample forecasts and realized values;
- II) the models are ranked according to their past performance based on RMSE so that the best forecasting model is identified as the one with the lowest RMSE;
- III) the HLN statistic is employed so as to sequentially test whether the best forecasting model encompasses other remaining models: If the best model encompasses the alternative model at some significance level τ , the alternative model is deleted from the list;²⁶
- IV) Step III) is then repeated using the second best model (having discarded the encompassed models according to the previous step) so as to include those which are not encompassed by the first best model;
- V) The procedure continues with the following best models until no encompassed model remains on the list.
- VI) The hierarchical forecast combination (HFC) is based on a simple average of the previously selected models.

The result of this *ex post* combination of soft rules is labeled $PFMcomb_{xx}$.

RMSE ratios reported in Table 7 depict a clear ranking: the performance of the $PFMsoft$ model is always improved by the two combination methods (i.e. $PFMavg$ and $PFMcomb$), and this outcome is related with the higher number of indicators exploited by the $PFMsoft$ approach.

Table 7 here

²⁶ In our context we set τ at the 10 per cent significance level.

As a result, moving from $Q = \text{"min"}$ to $Q = 90^{\text{th}}$ leads to monotonic improvements of the RMSE ratios over the AR benchmark of both the combination methods. This fact further supports the notion of using statistical procedures to appropriately pre-select the indicators rather than leaving the process to researcher judgment and beliefs. The best two cases of forecast combination (PFMavg and PFMcomb at $Q = 90^{\text{th}}$) outperform not only the AR benchmark, but also - and more interestingly - both the FM without pre-selection and the Rünstler and Sédillot (2003) BM, especially for $\nu=1$. Overall, our results indicate that the forecast combination can help in preventing bias in the simplest models. Although alternative very stringent soft rules lead to simple models that, as such, are prone to misspecification, forecast combination can mitigate bias emerging from the potential misspecification of individual models generating individual forecasts.²⁷

Comparing the best scorers combination methods, the hierarchical procedure (PFMcomb) very often shows lower RMSE ratios than the unweighted average of individual forecasts (PFMavg) but this evidence - against the view expressed in e.g. Hendry and Clements (2002) and Stock and Watson (2004) - is weakened by the not comparability of the two roads to combine forecasts, as only PFMavg is fully *ex ante*.

7. - Extensions to the baseline results

In this Section, we will assess the robustness of our main findings to alterations coming from three sources. Firstly, disaggregate modelling of GDP demand components (i.e. demand-side GDP forecasts); secondly, changes in some settings in implementing the baseline FM approach; thirdly, further indicator pre-selection techniques. In this respect, note that disaggregation affects all the models listed in Table 1, while alternative settings relating to both the number of factors in the diffusion index model and the dimension of the panel of indicators affect only the FM approach.

²⁷ The range of cases in which forecast averaging and combination proved to be useful can be extended by noting with Timmermann (2006) that, besides general misspecification, the presence of structural breaks in individual model parameters can also lead to biased individual forecasts.

7.1. GDP forecasting from the demand side

In the context of alternative uses of aggregate/disaggregate information, the euro area GDP forecast from the demand side corresponds to case (II) introduced in Section 3.3: the modelling of disaggregate information (i.e. the demand items) to forecast aggregate GDP by aggregating the corresponding disaggregate forecasts. In implementing case (II), we have explicitly modelled five GDP components: private consumption, government expenditure, investments (including changes in inventories), imports and exports. Each of these has been first modelled and forecast by following the same steps as those used for the forecast of aggregate (supply-side) GDP. Then, the five forecasts have been aggregated into GDP by using an aggregator equation (see, for e.g., Baffigi et al., 2004).²⁸

The list of the models used to forecast GDP from the demand side has been described in the first two columns of Table 8 (on the subject of codes and labels see Table 1).²⁹

Table 8 here

The nine-columns of results in Table 8 are grouped into three blocks of vintages, ($v = 1, 2$ and 3), depending on data availability. For each vintage, the first two columns report the RMSE ratios of each model to the corresponding benchmark. The first column reports the ratios for the baseline forecasting scenario,³⁰ the second reports the ratios for the demand side forecasts. Finally,

²⁸ For each forecast round, we first estimated the *quasi*-identity relating GDP growth rates to the growth rates of the five demand components listed above, then parameter estimates were used to aggregate individual forecasts to get the demand-side GDP forecast. Growth rates were proxied by first differences in log-levels.

²⁹ In order to avoid the risk of over-reporting obscuring the relevant evidence, in Table 8 we have not reported the performance of both averages and combinations of the soft thresholding rules. Consequently, we have emphasised a few basic cases and left all the other results to one side - which are in line with those reported here and available upon request.

³⁰ Although the supply-side (baseline) results are in Table 4, they are also reported in Table 5 in order to ease the comparison of supply- and demand-side forecasting ability.

the third column reports the relative RMSE of the demand side approach compared to the supply side one for each vintage and model within each tern.

If we consider the building of the demand-side benchmark model, each demand component has first been AR modelled and, then forecasts have been used to predict GDP forecasts through the aggregator equation, see e.g. Golinelli and Parigi (2013). Both benchmark models do not exploit any indicator information.

In the context of demand-side modelling, FM extract the same factors for all the GDP components: the right-hand side specification of each of the five diffusion index models (the general unrestricted model, see equation 2) includes the same P regressors. In other words, as they differ each other only in the *target* variable (i.e. the five demand components), FM use disaggregate information only at the stage of equation (2) parameter estimation.

By contrast, the PFM approach treats the available information in a completely different way than the FM one. Pre-screening activities apply to the same large panel of indicators, but the thresholding rules target different indicator subsets which are specific for each demand component.

Finally, the demand-side BM preselect specific indicators by component on the basis of the choices in Baffigi et al. (2004).³¹

³¹ In the specification of the BM models for private consumption (C), government consumption (G), investments (I), imports (M) and exports (X) we consider current values of the regressors and their first lag. The following variables are used: consumer confidence, total retail turnover, expectations on unemployment (C); economic situation on the overall economy, expectations on public deficit, expectations on short-term interest rates and on the trade balance over the next 6 months (G); level of total order book, expected production levels, building activity with respect to the previous month; industrial production excluding buildings (I); stock of finished manufactured goods, euro/US dollar exchange rate, industrial production excluding buildings, expectations on import volumes over the next six months (M); level of foreign order book, industrial production excluding buildings, expectations on export volumes over the next six months, US expectations on import volumes over the next six months (X).

Overall, the main results relating to the forecasting ability of euro area GDP from the demand side can be summarised in the following four findings. Firstly, models using indicators tend to improve over the benchmark as soon as monthly indicator information is released. Secondly, PFM with soft thresholding rules generally show better performance than hard ones. Thirdly, PFM with soft thresholding rules show improving performance as the number of exploited indicators decreases. Fourthly, the demand-side approach always underperforms the supply-side approach.

The first three outcomes are in line with the findings of the baseline (i.e. supply-side) case discussed above and denote the robustness of the notion of using few indicators pre-screened by the soft thresholding rules.

The fourth outcome confirms the findings in Baffigi et al. (2004), Hubrich (2005), and Barhoumi et al. (2011) in our more extended modelling context. Functional disaggregation does not improve the forecasting ability over the aggregate approach. In fact, the third columns of each tern in Table 8 (labelled as "S/D") report ratios lower than one, suggesting that it is better to forecast GDP with aggregate rather than disaggregated approaches. Although the number of indicators shrank with both BM and soft rules PFM with $Q = 90^{\text{th}}$, the worsening in their forecasting ability is due to the relative large number of predictors. Any demand component is bound to select its specific indicators and, on average, the noise introduced by them prevails over the signal. The RMSE increase, often more than 30-40% of the supply-side method, is common to all demand-side modelling approaches and reaches its "best case" (of about a 20% RMSFE increase) in correspondence with the benchmark and FM models.

7.2. Alternative FM settings

The robustness of our results to alterations in the way the FM approach is implemented in the baseline forecast has been assessed along three directions in order to disentangle the effects on FM forecasting ability of:

- > changing the number (l) of lags in expanding the panel of indicators from which principal components are extracted (in the baseline $l = 4$);
- > partitioning in b blocs (namely, qualitative indicators, quantitative indicators, financial variables and prices) the input panel of indicators and extracting k "representative" indicators by bloc (in the baseline $b=1$, as the input panel is considered as a single block);³²
- > using an alternative number (k) of principal components which are the predictors of the diffusion index model translating indicator information in terms of GDP (in the baseline $k = 3$).

The results in terms of RMSE of alternative FM embodying the alterations listed above have been reported in the rows of Table 9, while the results in the columns have been ordered by vintage (i.e. $v = 1, 2$ and 3). More specifically, the different models have been labelled as $FM(l, b, k)$ in the rows where l is the number of lags in expanding the panel of indicators prior to principal component extraction; b is the number of blocks in which the input panel is structured and k is the number of principal components extracted with SW. Finally, the row in bold shows the RMSFE of the baseline $FM(4, 1, 3)$.

Table 9 here

The minimum points by vintage have been achieved by models with five factors extracted from the un-extended input panel: $FM(1, 1, 5)$ for $v = 1$, and $F(0, 1, 5)$ for $v = 2$ and 3 . This is not surprising, considering that FM with 5 principal components (as suggested by Bai and Ng, 2002), extracted from an input panel of 259 simultaneous indicators is the model suggested by the literature implementing SW forecasting approach in the present context. However, it should be

³² Note that block-partitioning is an alternative way of injecting external information into the dataset, instead of pre-selecting with BM or pre-screening with PFM as we did in the baseline forecast, see Moench et al. (2009) and Alvarez et al. (2012).

noted that, even adopting these improved outcomes as the FM of our baseline forecast, the findings emerged in the previous section are qualitatively the same. In other words, the FM baseline features resist to alterations in their basic settings.

Finally, the last two rows of Table 9 RMSE worsen by more than 15-20% as compared to the RMSE levels scored by the baseline FM. At least in the present context, the extraction of representative factors from an input panel structured in blocks of indicators is not as effective as pre-selection applied to BM or pre-screening to PFM in improving GDP forecasting ability.

8. - Conclusions

The focus of the paper is to compare the performance of alternative methods to nowcast euro area GDP on the basis of several ways to select/exploit indicators' information. Building on the literature on regularization and dimension reduction methods, our approach consists in bridging quarterly national accounts data using factors extracted from a pre-screened panel of monthly and quarterly series. We find that targeting predictors is an effective way to improve forecast performance, thus confirming the results of Bai and Ng (2008), Bulligan et al. (2012) and Kim and Swanson (2013).

Indicators' selection is the core contribution of this paper. Pre-screening before factor extraction leads to a mixed approach between bridge models (BM) and factor models (FM), which we will refer to as pre-screened factor models (PFM). We argue that the factor extraction from a subset of targeted indicators to a specific variable, retains the most valuable piece of information to be passed into the FM approach and can deliver substantial gains in terms of forecasting accuracy.

Forecasting ability is assessed according two main specific features: the pseudo real-time nature of the information set, and the empirical probability of each indicator to be picked conditional to a screening rule (one hard and 5 soft rules). For each rule, a threshold Q is defined and is based on the quantiles of the empirical probability distributions: $Q = [\text{min}, 10^{\text{th}}, 25^{\text{th}}, 50^{\text{th}}, 75^{\text{th}}, 90^{\text{th}}]$.

The main results can be summarized as follows. *First*, forecast accuracy of each model (BM, FM, PFM) significantly improves as indicator information increases. Overall, when quantitative information (such as the manufacturing production indexes) is lacking, modeller preferences do not pay. On the contrary, in the nowcast, the available quantitative information works better than other qualitative information sources (such as survey indicators data), which are more up-to-date but also much noisier. Therefore, when virtually all the indicators are available for all the months of the quarter to be forecast, the contribution of qualitative indicators becomes considerably less useful. *Second*, by exploiting the moments of the empirical distribution of selected indicators, forecasting ability monotonically increases with the progressive reduction in the number of indicators from the large-panel pool. As a result, wider information, if not properly selected, might deliver more noise than signal to nowcast GDP. *Third*, PFM based on soft rules (PFM_{soft}) systematically outperform the PFM based on hard-thresholding rule (PFM_{hard}) within all blocks, as the latter retains much more, and less useful, information than the soft ones. *Fourth*, PFM_{soft} at the 90th quantile (PFM_{soft90}) is the best selected model over all vintages. It tends to retain very few variables and coincides with BM in purpose, but PFM_{soft90} significantly outperform BM nowcasts since it also exploits shrinkage techniques, which allow to additionally reducing over-fitting problems in model specification.

In addition to the above findings, the paper provides evidences that euro area GDP forecasting ability may be further improved by applying forecasts' combination techniques. In particular, two techniques are adopted in the paper: the un-weighted average of single soft forecast and a hierarchical procedure (Costantini and Pappalardo, 2010; Kisinbay, 2010) that aims to pre-select models before forecasts' combination by accounting for both RMSE and encompassing properties. We document that moving from $Q = \text{"min"}$ to $Q = 90^{\text{th}}$ leads to monotonic improvements of both combination methods; moreover, the hierarchical method at $Q = 90^{\text{th}}$ provides the best forecast, as it can mitigate the bias emerging from the potential misspecification of models

generating individual forecasts. Finally, the *ex post* dominance of hierarchical procedure compared to the unweighted average combination vanishes as the thresholding rule becomes less stringent: the larger the bulk of information, the less effective the combination approach is in lowering RMSE ratios compared to the unweighted average approach.

Several sensitivity analyses are performed. Firstly, the robustness of the supply-side predictions of euro area GDP is assessed with respect to the forecasting from the demand side, which is obtained by combining forecast of GDP expenditure components (Huang and Lee, 2010; and Hendry and Hubrich, 2011). According to our results, the demand-side approach systematically underperforms the supply-side one. We conclude that functional disaggregation does not improve the forecasting ability over the aggregate approach (see Baffigi et al., 2004; Hubrich, 2005; Barhoumi et al., 2011). Secondly, this exercise confirms the findings of the baseline (i.e. supply-side) GDP forecast: PFM*soft* perform better than its hard counterparts, with lowering forecast errors as the number of exploited indicators decreases. Thirdly, baseline results are also robust to a number of alternative specifications of the FM model with respect to *i*) changes in the lags of the panel of indicators ($l=4$ in the baseline); *ii*) the partition of the panel of indicators in blocks (qualitative, quantitative indicators, financial and price indicators) and extracting k factors by block (rather than a set of k factors from the whole panel); *iii*) the number of principal components exploited as predictors of the diffusion index model ($k = 3$ in the baseline).

Overall, the PFM approach to nowcast the euro area GDP looks promising. It allows for a more effective use of the available short-run statistical information, thus increasing the reliability of the forecasting exercises through a significant reduction of the prediction error with respect to the one obtained from more traditional tools (such as *ad hoc* BM and standard FM). We argue that the PFM empirical methodology should be applied to forecast other short-run indicators which can be relevant for decision making and public policy. This would primarily benefit policymakers and

market participants who base their choices on the early understanding of the state of the economic activity.

References

- Alvarez, R., M. Camacho and G. Pérez-Quirós (2012), "Finite sample performance of small versus large scale dynamic factor models", *CEPR Discussion Paper*, No. 8867.
- Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin and G. Runstler (2011), "Short-term forecasts of euro area GDP growth", *Econometrics Journal*, Vol. 14, pp. 25-44.
- Antipa, P., K. Barhoumi, V. Brunhes-Lesage and O. Darné (2012), "Nowcasting German GDP: a comparison of bridge and factor models", *Journal of Policy Modeling*, Vol. 34, pp. 864-878.
- Baffigi, A., R. Golinelli, and G. Parigi (2004), "Bridge models to forecast the euro area GDP", *International Journal of Forecasting*, Vol. 20, No 3, pp. 447-460.
- Bai, J. and S. Ng (2002), "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, Vol. 70, pp. 191-221.
- Bai, J. and S. Ng (2008), "Forecasting economic time series using targeted predictors", *Journal of Econometrics*, Vol. 146, pp. 304-317.
- Banbura, M. and G. Runstler (2011), "A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP", *International Journal of Forecasting*, Vol. 27, pp. 333-346.
- Banerjee, A., M. Marcellino and I. Masten (2005), "Leading indicators for euro-area inflation and GDP growth", *Oxford Bulletin of Economics and Statistics*, Vol. 67, pp. 785-813.
- Barhoumi, K., O. Darné, L. Ferrara and B. Pluyaud (2011), "Monthly GDP forecasting using bridge models: application for the French economy", *Bulletin of Economic Research*, Vol. 64, No. 1, pp. 53-70.
- Bernanke, B. S. and J. Boivin (2003), "Monetary policy in a data-rich environment", *Journal of Monetary Economics*, Vol. 50, pp. 525-546.
- Bodo, G., R. Golinelli and G. Parigi (2000), "Forecasting industrial production in the euro area", *Empirical Economics*, Vol. 25, pp. 541-61.
- Boivin J. and S. Ng (2005), "Understanding and comparing factor-based forecasts", *International Journal of Central Banking*, Vol. 1, No. 3, pp. 117-151.
- Boivin, J. and S. Ng (2006), "Are more data always better for factor analysis?", *Journal of Econometrics*, Vol. 132, pp. 169-194.
- Brunhes-Lesage, V. and O. Darné (2012), "Nowcasting the French index of industrial production: a comparison from bridge and factor models", *Economic Modelling*, Vol. 29, pp. 2174-2182.
- Bulligan, G., R. Golinelli and G. Parigi (2010), "Forecasting monthly industrial production in real-time: from single equations to factor-based models", *Empirical Economics*, Vol. 39, No. 2, pp 303-336.
- Bulligan, G., M. Marcellino and F. Venditti (2012), "Forecasting economic activity with higher frequency targeted predictors", *Temi di Discussione*, Bank of Italy, No. 847.
- Burriel, P. (2012), "A real-time disaggregated forecasting model for the Euro area GDP", *Economic Bulletin*, Bank of Spain, April, pp. 93-103.

- Caggiano, G., G. Kapetanios and V. Labhard (2011), "Are more data always better for factor analysis? Results for the Euro area, the six largest Euro area countries and the UK", *Journal of Forecasting*, Vol. 30, pp. 736-752.
- Camacho, M., and G. Perez-Quiros (2010), "Introducing the euro-sting: short-term indicator of euro area growth", *Journal of Applied Econometrics*, Vol. 24, pp. 663-694.
- Camacho, M., G. Perez-Quiros and P. Poncela (2012), "Markov-switching dynamic factor models in real time", *CEPR Discussion Paper*, No. 8866.
- Clements, M. P. and A. B. Galvao (2008), "Macroeconomic forecasting with mixed frequency data: forecasting US output growth", *Journal of Business and Economic Statistics*, Vol. 26, No. 4, pp. 546-554.
- Clements, M. P. and D. F. Hendry (2011), "Forecasting from mis-specified models in the presence of unanticipated location shift", in M. P. Clements and D. F. Hendry (eds.), *The Oxford Handbook of Economic Forecasting*, Oxford University Press, Ch. 10, pp. 271-314.
- Constantini, M. and C. Pappalardo (2010), "A hierarchical procedure for the combination of forecasts", *International Journal of Forecasting*, Vol. 26, pp. 725-743.
- Constantini, M. and R. M. Kunst (2011), "Combining forecasts based on multiple encompassing tests in a macroeconomic core system", *Journal of Forecasting*, Vol. 30, No. 6, pp. 579-596.
- Croushore, D. and T. Stark (2001), "A Real-Time Data Set for Macroeconomists", *Journal of Econometrics*, Vol. 105, pp. 111-130.
- D'Agostino, A. and D. Giannone (2012), "Comparing alternative predictors based on large-panel factor models", *Oxford Bulletin of Economics and Statistics*, Vol. 74, No. 2, pp. 306-326.
- De Mol, C., D. Giannone and L. Reichlin (2008), "Forecasting using a large number of predictors: is Bayesian shrinkage a valid alternative to principal components?", *Journal of Econometrics*, Vol. 146, pp. 318-328.
- Diebold, F.X. and R. S. Mariano (1995), "Comparing predictive accuracy", *Journal of Business and Economic Statistics*, Vol. 13, pp. 253-63.
- Diebold, F. X. and G. D. Rudebusch (1991), "Forecasting output with the composite leading index: a real time analysis", *Journal of the American Statistical Association*, Vol. 86, pp. 603-610.
- 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*, Vol. 27, pp. 371-390.
- Drechsel, K. and L. Maurin (2011), "Flow of conjunctural information and forecast of euro area economic activity", *Journal of Forecasting*, Vol. 30, pp. 336-354.
- Ferrara, L., D. Guegan and P. Rakotomaroahy (2010), "GDP nowcasting with ragged-edge data: a semi-parametric modeling", *Journal of Forecasting*, Vol. 29, pp. 18-199.
- Forni, M., M. Hallin, M. Lippi and L. Reichlin (2001), "Coincident and leading indicators for the euro area", *The Economic Journal*, Vol. 111, May, pp. 62-85.
- Forni, M., M. Hallin, M. Lippi and L. Reichlin (2005), "The generalized factor model: one sided estimation and forecasting", *Journal of the American Statistical Association*, Vol. 100, pp. 830-840.
- Giacomini, R. and H. White (2006), "Tests of conditional predictive ability", *Econometrica*, Vol. 74, No. 6, pp. 1545-1578.

- Giannone, D., L. Reichlin and D. Small (2008), "Nowcasting: the real time informational content of macroeconomic data", *Journal of Monetary Economics*, Vol. 55, pp. 665-676.
- Giannone, D., L. Reichlin and S. Simonelli (2009), "Nowcasting euro area economic activity in real time: the role of confidence indicators", *National Institute Economic Review*, No. 210, October, pp. 90-97.
- Golinelli, R. and G. Parigi (2007), "The use of monthly indicators to forecast quarterly GDP in the short run: an application to the G7 countries", *Journal of Forecasting*, Vol. 26, No. 2, pp. 77-94.
- Golinelli, R. and G. Parigi (2008), "Real-time squared: a real-time data-set for real-time GDP forecasting", *International Journal of Forecasting*, Vol. 24, pp. 368-385.
- Golinelli, R. and G. Parigi (2013), "Tracking world trade and GDP in real time", *Temi di Discussioni*, Bank of Italy, No. 920.
- Grassman, P. and F. Keereman (2001), "An indicator-based short-term forecast for quarterly GDP in the euro area", *EC Economic Paper*, No. 154, June.
- Grunfeld, Y. and Z. Griliches (1960), "Is aggregation necessarily bad?", *The Review of Economics and Statistics*, Vol. 42, No. 1, pp. 1-13.
- Hahn, E. and F. Skudelny (2008), "Early estimates of euro area real GDP growth: A bottom up approach from the production side", *ECB Working Paper*, No. 975, December.
- Harvey, D. I., S. J. Leybourne and P. Newbold (1997), "Testing the equality of prediction mean squared errors", *International Journal of Forecasting*, Vol. 13, pp. 281-91.
- Harvey, D. I., S. Leybourne and P. Newbold, (1998), "Tests for forecast encompassing", *Journal of Business and Economic Statistics*, Vol. 16, pp. 254-259.
- Heij, C., D. van Dijk and P. J. F. Groenen (2011), "Real-time macroeconomic forecasting with leading indicators: an empirical comparison", *International Journal of Forecasting*, Vol. 27, pp. 466-481.
- Hendry, D. F. and M. P. Clements (2002), "Pooling of forecasts", *Econometrics Journal*, Vol. 5, pp. 1-26.
- Hendry, D. F. and K. Hubrich, K. (2011), "Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate", *Journal of Business and Economic Statistics*, Vol. 29, No. 2, pp. 216-227.
- Huang, H. and T.-H. Lee (2010), "To combine forecasts or to combine information?", *Econometric Reviews*, Vol. 29, No. 5-6, pp. 534-570.
- Hubrich, K. (2005), "Forecasting euro area inflation; does aggregating forecasts by HICP component improve forecasting accuracy?", *International Journal of Forecasting*, Vol. 21, pp. 119-136.
- Kaufmann, S. and C. Schumacher (2013), "Bayesian estimation of sparse dynamic factor models with order-independent identification", *Study Center Gerzensee Working Paper*, No. 13.04.
- Kim, H. H. and N. R. Swanson (2013), "Forecasting financial and macroeconomic variables using data reduction methods: new empirical evidence", *Journal of Econometrics*, forthcoming.
- Kisinbay, T. (2010), "The use of encompassing tests for forecast combinations", *Journal of Forecasting*, Vol. 29, No. 8, pp. 715-727.

- Kitchen, J. and J. Kitchen (2013), "Real-time forecasting revisited: letting the data decide", *Business Economics*, Vol. 48, No. 1, pp. 8-28.
- Klein, L. R. (1990), "Cyclical indicators in econometric models", in P. A. Klein (ed.), *Analysing Modern Business Cycles*, M.E. Sharpe Inc., pp. 97-106.
- Klein, L. R. and S. Ozmucur (2008), "The University of Pennsylvania models for high frequency macroeconomic modeling", in R. S. Mariano and Y. Tse, *Econometric Forecasting and High-Frequency Data Analysis*, Lecture Notes Series, National University of Singapore, Vol. 13, World Scientific, pp. 53-91.
- Klein, L. R., and E. Sojo (1989), "Combinations of high and low frequency data in macroeconometric models", in L. R. Klein and J. Marquez (eds.), *Economics in Theory and Practice: An Eclectic Approach*, Kluwer Academic Publishers, pp. 3-16.
- Koop, G. and S. Potter (2004), "Forecasting in dynamic factor models using Bayesian model averaging", *Econometrics Journal*, Vol. 7, pp. 550-565.
- Kuzin, V., M. Marcellino and C. Schumacher (2009), "MIDAS versus mixed-frequency VAR: nowcasting GDP in the euro area", *Discussion Paper*, Deutsche Bundesbank, No. 7.
- Luciani, M. (2014), "Forecasting with approximate dynamic factor models: the role of non-pervasive shocks", *International Journal of Forecasting*, Vol. 30, pp. 20-29.
- Marcellino, M., J. H. Stock and M. W. Watson (2003), "Macroeconomic forecasting in the euro area: country specific versus area wide information", *European Economic Review*, Vol. 47, pp. 1-18.
- McGuckin, R. H., A. Ozyildirim and V. Zarnowitz (2007), "A more timely and useful index of leading indicators", *Journal of Business and Economic Statistics*, Vol. 25, No. 1, pp. 110-120.
- Moench, E, S. Ng and S. Potter (2009), "Dynamic hierarchical factor models," *Staff Reports*, Federal Reserve Bank of New York, No. 412.
- Ng, S. (2013), "Variable selection in predictive regressions", in G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Vol. 2, Chapter 14, pp. 752-789.
- Pesaran, M. H. and A. Timmermann (2005), "Real-time econometrics", *Econometric Theory*, Vol. 21, pp. 212-231.
- Rathjens, P. and R. P. Robins (1993), "Forecasting quarterly data using monthly information", *Journal of Forecasting*, Vol. 12, pp. 321-330.
- Runstler, G., K. Barhoumi, S. Benk, R. Cristadoro, A. den Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth and C. van Nieuwenhuyze (2009), "Short-term forecasting of GDP using large datasets: a pseudo real time forecast evaluation exercise", *Journal of Forecasting*, Vol. 28, pp. 595-611.
- Runstler, G. and F. Sedillot (2003), "Short-term estimates of euro area real GDP by means of monthly data", *ECB Working Paper*, No. 276, September.
- Schumacher, C. (2010), "Factor forecasting using international targeted predictors: the case of German GDP", *Economics Letters*, Vol. 107, pp. 95-98.
- Schumacher, C. and J. Breitung (2008), "Real-time forecasting of German GDP based on large factor model with monthly and quarterly data", *International Journal of Forecasting*, Vol. 24, pp. 386-398.

- Stark, T. and Croushore, D. (2002), "Forecasting with a real-time data set for macroeconomists", *Journal of Macroeconomics*, Vol. 24, No. 4, pp. 507–568.
- Stock, J. H. and M. W. Watson (1996), "Evidence on structural instability in macroeconomic time series relations", *Journal of Business and Economic Statistics*, Vol. 14, pp. 11-30.
- Stock, J. H. and M. W. Watson (2002a), "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, Vol. 20, pp. 147-162.
- Stock, J. H. and M. W. Watson (2002b), "Forecasting Using Principal Components from a Large Number of Predictors ", *Journal of the American Statistical Association*, Vol. 97, pp. 1167-1179.
- Stock, J. H. and M. W. Watson (2004), "Combination forecasts of output growth in a seven-country data set", *Journal of Forecasting*, Vol. 23, pp. 405-430.
- Stock, J. H. and M. W. Watson (2006), "Forecasting with Many Predictors", in G. Elliott, C. W. J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Vol. 1, Chapter 10, pp. 515-554.
- Swanson, N. R. and H. White (1997), "A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks", *Review of Economics and Statistics*, Vol. 79, No. 4, pp. 540–550.
- Timmermann, A. G. (2006), "Forecast combinations", in G. Elliott, C. W. J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Vol. 1, Chapter 4, pp. 135-196.
- Trehan, B. (1989), "Forecasting growth in current quarter real GNP", *Economic Review*, Winter, 39-52.

Tables and figures

Tab. 1 - Summary of our forecasting models ^a

Code ^b	Label	Description ^c
		<i>Benchmark Model</i>
	AR	AR model, lags selected with Schwarz criterion
		<i>Bridge Model</i>
1	BM	few pre-selected (fixed) indicators
		<i>Pre-screened Factor Models, using:^d</i>
2	PFMhard90	hard-thresh. rule, probability to be picked > 90 th pct.
3	PFMsoft90	union of 5 soft-thresh. rules, probability to be picked > 90 th pct.
4	PFMavg90	average of 5 soft-thresh. rules, probability to be picked > 90 th pct.
5	PFMcomb90	optimal combination of 5 soft-thresh. rules, probability to be picked > 90 th pct.
6	PFMhard75	hard-thresh. rule, probability to be picked > 75 th pct.
7	PFMsoft75	union of 5 soft-thresh. rules, probability to be picked > 75 th pct.
8	PFMavg75	average of 5 soft-thresh. rules, probability to be picked > 75 th pct.
9	PFMcomb75	optimal combination of 5 soft-thresh. rules, probability to be picked > 75 th pct.
10	PFMhard50	hard-thresh. rule, probability to be picked > 50 th pct.
11	PFMsoft50	union of 5 soft-thresh. rules, probability to be picked > 50 th pct.
12	PFMavg50	average of 5 soft-thresh. rules, probability to be picked > 50 th pct.
13	PFMcomb50	optimal combination of 5 soft-thresh. rules, probability to be picked > 50 th pct.
14	PFMhard25	hard-thresh. rule, probability to be picked > 25 th pct.
15	PFMsoft25	union of 5 soft-thresh. rules, probability to be picked > 25 th pct.
16	PFMavg25	average of 5 soft-thresh. rules, probability to be picked > 25 th pct.
17	PFMcomb25	optimal combination of 5 soft-thresh. rules, probability to be picked > 25 th pct.
18	PFMhard10	hard-thresh. rule, probability to be picked > 10 th pct.
19	PFMsoft10	union of 5 soft-thresh. rules, probability to be picked > 10 th pct.
20	PFMavg10	average of 5 soft-thresh. rules, probability to be picked > 10 th pct.
21	PFMcomb10	optimal combination of 5 soft-thresh. rules, probability to be picked > 10 th pct.
22	PFMhardmin	hard-thresh. rule, picked at least once
23	PFMsoftmin	union of 5 soft-thresh. rules, picked at least once
24	PFMavgmin	average of 5 soft-thresh. rules, picked at least once
25	PFMcombmin	optimal combination of 5 soft-thresh. rules, picked at least once
		<i>Factor Model</i>
26	FM	All indicators, static components, see Stock-Watson method

^(a) From high to low, models are listed by growing number of indicators and falling degree of arbitrariness. Models in the lower part of the table are bound to use more disaggregate information.

^(b) With *code* the alternative models are numbered progressively.

^(c) For a detailed description see Section 4.

^(d) Over repeated samples, hard- and soft-thresholding rules estimate the probability of each indicator to be picked conditional on a rule. The PFM pre-screening of targeted indicators selects those indicators with a probability larger than alternative percentiles of that distribution; the larger the percentile is, the lower the number of picked indicators.

Tab. 2 - The composition of the panel of indicators ^a

	Soft	Hard	Financial	Prices	Total	% shares
<i>Reference area</i>						
- Euro area	91	29	48	13	181	69.9%
- USA	65	0	0	0	65	25.1%
- World	0	0	13	0	13	5.0%
<i>Publication frequency</i>						
- Daily	0	0	61	0	61	23.6%
- Monthly	73	29	0	13	115	44.4%
- Quarterly	83	0	0	0	83	32.0%
<i>Data transformation</i>						
- No transformation	72	0	0	0	72	27.8%
- First differences, Δ	84	2	1	0	87	33.6%
- Δ of logs	0	27	60	13	100	38.6%
Total	156	29	61	13	259	
% shares	60.2%	11.2%	23.6%	5.0%		

(^a) Frequency of counts and % shares on N = 259 total indicators in the data-set.

Tab. 3 – Share of not available indicators by vintage ^a

	<i>First vintage</i>	<i>Second vintage</i>	<i>Third vintage</i>
<i>Indicators by type</i>			
Soft	35.5	1.5	0
Hard	11.2	10.4	0.8
Financial ^b	0	0	0
Prices	5.0	0.8	0
<i>Total</i>	51.7	12.7	0.8

(^a) Percent on the total number of indicators.

(^b) Financial variables for the first vintage are computed as the mean of the available daily outturns up the publication of the first release of the National Accounts.

Tab. 4 – The ten most selected indicators by the rules ^a

<i>Variable</i>	<i>Lag</i>	<i>PFMhard</i>	<i>PFMsoft</i>
EA IPI CPTL. GDS.	0	35.3%	100.0%
EA IPI INTERMEDIATE GDS.	0		64.7%
EA IPI EXCL. BUILD.	0		96.5%
EA IPI MFG. BAS MTLs.	0	22.4%	100.0%
EA CPTL. SHORTAGE	0		85.9%
US ECON. SIT. OVERALL ECON.	0		85.9%
US ECON. SIT. PRIV. CNSPT.	0	38.8%	70.6%
EA CONS. MAJOR PURCH. NEXT 12M	1		52.9%
EA IPI MFG. REPAIR MACH.	1	31.8%	
US CPTL. SHORTAGE.	1		61.2%
US INSUFF. DEMAND.	1		68.2%
EA RET. BUS. SIT.	1	41.2%	
US CHI PMI DELIVERIES	2	17.6%	
TR. EW. CCI LIVESTOCK	2	28.2%	
EA CPI NON NRG. IND. GDS. (DUR.)	3	18.8%	
US CONS. CONF. 6M INC. SAME	1	30.6%	
EA IPI BUILD.	4	15.3%	

^(a) Percent on the total number of possible cases (i.e. the 85 training rounds).

Tab. 5 – Number of selected indicators by hard/soft rule, quantile threshold and type

<i>Model</i>	<i>Threshold</i>	<i>Indicator type</i>				<i>Total</i>
		<i>Qualitative</i>	<i>Quantitative</i>	<i>Financial</i>	<i>Price</i>	
<i>PFMhard</i>	90	16	9	7	1	33
	75	58	22	27	2	109
	50	92	32	35	4	163
	25	126	42	58	11	237
	10	126	42	58	11	237
	min	126	42	58	11	237
<i>PFMsoft</i>	90	4	4	0	0	8
	75	9	6	4	0	19
	50	21	9	7	0	37
	25	38	13	13	0	64
	10	38	13	13	0	64
	min	41	13	17	0	71

Tab. 6 - GDP forecasting ability: RMSE ratios to AR benchmark

Code ^a	Model ^{b, c}	1 st vintage	sig ^d	2 nd vintage	sig ^d	3 rd vintage (nowcast)	sig ^d	# of indicators ^e
	<i>AR^f</i>	0.0075		0.0073		0.0073		0
1	BM	0.8689 *		0.4668 **		0.4499 **		4
2	PFM <i>hard</i> 90	0.8407		0.6692 *		0.6755 *		33
3	PFM <i>soft</i> 90	0.6932 **		0.4253 **		0.3161 **		8
6	PFM <i>hard</i> 75	0.8179		0.7040		0.6819		109
7	PFM <i>soft</i> 75	0.7729 *		0.5595 *		0.4978 *		19
10	PFM <i>hard</i> 50	0.7573		0.6516		0.6324		163
11	PFM <i>soft</i> 50	0.7489 *		0.5281 *		0.4862 *		37
14	PFM <i>hard</i> 25	0.8269		0.7748		0.7676		237
15	PFM <i>soft</i> 25	0.7005 *		0.6094 *		0.5736 *		64
18	PFM <i>hard</i> 10	0.8269		0.7748		0.7676		237
19	PFM <i>soft</i> 10	0.7005 *		0.6094 *		0.5736 *		64
22	PFM <i>hard</i> min	0.8269		0.7748		0.7676		237
23	PFM <i>soft</i> min	0.6689 *		0.5169 *		0.4914 *		71
26	FM	0.7134 *		0.6666 *		0.6571 *		1,295

(^a) With *code* the alternative models are numbered progressively. From high to low, models are listed by growing number of indicators and falling degree of arbitrariness. Models in the lower part of the table use more indicators.

(^b) For a short description see Table 1; further details are given in Section 4.

(^c) With reference to only PFM, the numbers at the end of each label indicates percentiles. Over repeated samples, hard- and soft-thresholding rules estimate the probability of each indicator to be picked conditional on a rule. The PFM pre-screening of targeted indicators selects those indicators with a probability larger than alternative percentiles of that distribution; the larger the percentile is, the lower the number of picked indicators.

(^d) *, and ** means 10 and 5% significant Harvey et al. (1997) test for equal conditional predictive ability. Under the null, the RMSE ratio to AR is one, i.e. the RMSE of the model in each row is not significantly lower than that of the benchmark.

(^e) Total number of indicators whose information is used by each model; the detail by indicator type is given in Table 5.

(^f) This row reports the RMSE levels of the benchmark AR, rather than ratios that would be 1 by definition. Note that 1st vintage RMSE is slightly different to that of the 2nd and 3rd vintages because of a small difference in the number of forecast errors.

Tab. 7 - GDP forecasting ability: RMSE ratios to each PFMsoft

Code ^a	Model ^{b, c}	1 st vintage	sig ^d	2 nd vintage	sig ^d	3 rd vintage (nowcast)	sig ^d	# of indicators ^e
3 ^f	PFMsoft90	0.0052		0.0031		0.0023		8
4	PFMavg90	0.8719		0.9944		0.9611		5
5	PFMcomb90	0.8686		0.9052		0.8285		5
7 ^f	PFMsoft75	0.0058		0.0041		0.0036		19
8	PFMavg75	0.7683 *		0.7498		0.6828		13
9	PFMcomb75	0.7619		0.7734		0.6016		13
11 ^f	PFMsoft50	0.0056		0.0039		0.0035		37
12	PFMavg50	0.8737 **		0.7830 **		0.7585 **		27
13	PFMcomb50	0.8224		0.7355 **		0.7182 **		27
15 ^f	PFMsoft25	0.0053		0.0044		0.0042		64
16	PFMavg25	0.9352 *		0.8175 **		0.7962 **		42
17	PFMcomb25	0.8848 ***		0.7629 ***		0.7456 ***		42
19 ^f	PFMsoft10	0.0053		0.0044		0.0042		64
20	PFMavg10	0.9657 *		0.8740 **		0.8532 **		45
21	PFMcomb10	0.8848 ***		0.7629 ***		0.7300 ***		45
23 ^f	PFMsoftmin	0.0050		0.0038		0.0036		71
24	PFMavgmin	1.0341		1.0443		1.0077		49
25	PFMcombmin	0.9792		0.8994		1.0065		49

(^a) With *code* the alternative models are numbered progressively. From high to low, models are listed by growing number of indicators and falling degree of arbitrariness. Models in the lower part of the table use more indicators.

(^b) For a description see Table 1.

(^c) With reference to only PFMsoft, the numbers at the end of each label indicates percentiles.

(^d) *, and ** means 10 and 5% significant Harvey et al. (1997) test for equal conditional predictive ability. Under the null, the RMSE ratio to PFMsoft by percentile is one, i.e. the RMSE of the model in each row is not significantly lower than that of the PFMsoft benchmark of its percentile.

(^e) Total number of indicators whose information is used by each model.

(^f) The row reports the RMSE of the benchmark by percentile, rather than ratios that would be 1 by definition.

Tab. 8 - GDP forecasting ability: comparing supply- and demand-side models

Vintages of available indicators information:		1 st vintage			2 nd vintage			3 rd vintage		
Code ^a	Models ^a	S ^b	D ^b	S/D ^c	S ^b	D ^b	S/D ^c	S ^b	D ^b	S/D ^c
	<i>Benchmark</i> ^d	0.0075	0.009	0.8333	0.0073	0.0088	0.8295	0.0073	0.0088	0.8295
1	BM	0.8689	0.6667	1.3033	0.4668	0.6023	0.7750	0.4499	0.5682	0.7918
2	PFM <i>hard</i> 90	0.8407	0.9667	0.8697	0.6692	0.8977	0.7455	0.6755	0.8864	0.7621
3	PFM <i>soft</i> 90	0.6932	0.6444	1.0757	0.4253	0.4886	0.8704	0.3161	0.4091	0.7727
6	PFM <i>hard</i> 75	0.8179	0.9000	0.9088	0.704	0.9318	0.7555	0.6819	0.9432	0.7230
7	PFM <i>soft</i> 75	0.7729	0.8556	0.9033	0.5595	0.6023	0.9289	0.4978	0.625	0.7965
10	PFM <i>hard</i> 50	0.7573	0.8111	0.9337	0.6516	0.7727	0.8433	0.6324	0.75	0.8432
11	PFM <i>soft</i> 50	0.7489	0.7889	0.9493	0.5281	0.8523	0.6196	0.4862	0.8636	0.5630
14	PFM <i>hard</i> 25	0.8269	0.900	0.9188	0.7748	0.8636	0.8972	0.7676	0.8295	0.9254
15	PFM <i>soft</i> 25	0.7005	0.7333	0.9553	0.6094	0.8182	0.7448	0.5736	0.8409	0.6821
18	PFM <i>hard</i> 10	0.8269	0.9111	0.9076	0.7748	0.875	0.8855	0.7676	0.8409	0.9128
19	PFM <i>soft</i> 10	0.7005	1.0556	0.6636	0.6094	1.0000	0.6094	0.5736	0.9205	0.6231
22	PFM <i>hard</i> min	0.8269	0.9111	0.9076	0.7748	0.875	0.8855	0.7676	0.8409	0.9128
23	PFM <i>soft</i> min	0.6689	1.1111	0.6020	0.5169	1.0568	0.4891	0.4914	1.0568	0.4650
26	FM	0.7134	0.7000	1.0191	0.6666	0.6705	0.9942	0.6571	0.6364	1.0325

^(a) With *code* the alternative models are numbered as in Table 1, where models' description is also given (further models' details are in Section 3). From high to low, models are listed by growing number of indicators and falling degree of arbitrariness. Models in the lower part of the table use more indicators.

^(b) S = supply-side GDP forecasts' RMSE ratios over the AR (the same as those reported in Table 4), D = demand-side GDP forecasts' RMSE ratios over the AC. For S and D columns, the AR-AC row reports RMSE levels, rather than ratios that would be 1 by definition (see also note *f* in Table 4).

^(c) RMSE ratios by model of demand-side over the corresponding supply-side model. Figures larger than one denote the better performance of supply side forecasts.

^(d) Depending on supply (S) or demand (D) side approaches, benchmarks are different: in S columns the benchmark is the univariate AR model for GDP, while in D columns the benchmark is the aggregation of the AR GDP components, see Golinelli and Parigi (2013).

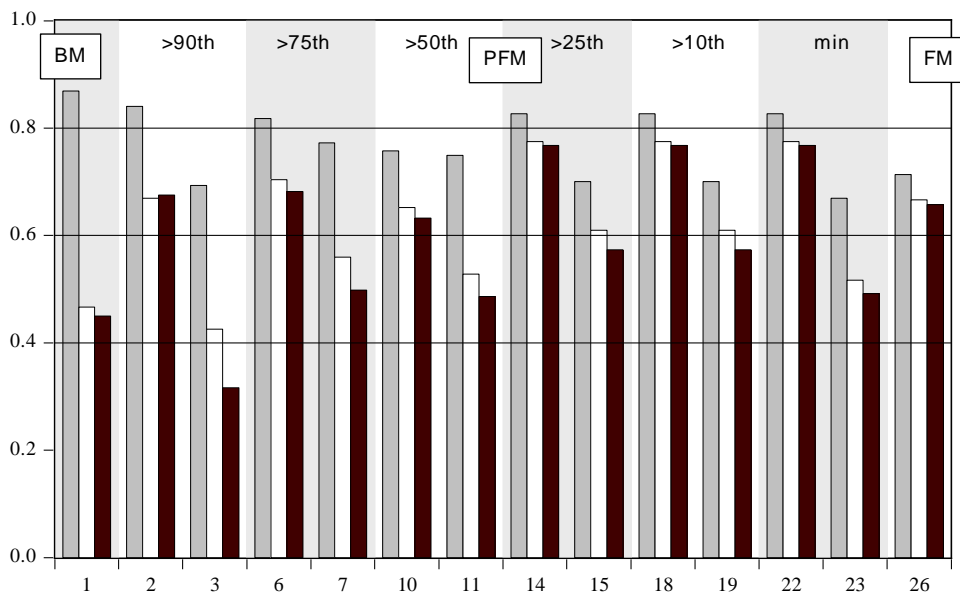
Tab. 9 - GDP forecasting ability: RMSE ratios of alternative FM ^a

FM(<i>l</i> , <i>b</i> , <i>k</i>) models	1 st vintage	2 nd vintage	3 rd vintage
FM(0, 1, 1)	1.0185	1.0205	1.0209
FM(0, 1, 2)	1.0556	0.9796	0.9791
FM(0, 1, 3)	1.0926	1.0000	1.0000
FM(0, 1, 5)	1.1667	0.9184	0.8125
FM(1, 1, 1)	1.0185	1.0409	1.0416
FM(1, 1, 2)	0.9260	0.9388	0.9374
FM(1, 1, 3)	1.0185	1.0817	1.0833
FM(1, 1, 5)	0.9630	1.0000	1.0000
FM(2, 1, 1)	1.7408	1.8572	1.8958
FM(2, 1, 2)	1.0740	1.0817	1.0625
FM(2, 1, 3)	1.0371	1.0409	1.0416
FM(2, 1, 5)	1.0740	1.0817	1.0833
FM(4, 1, 1)	1.2778	1.3470	1.3542
FM(4, 1, 2)	1.0185	0.9796	0.9791
<i>FM(4, 1, 3)</i> ^b	<i>0.0054</i>	<i>0.0049</i>	<i>0.0048</i>
FM(4, 1, 5)	0.9815	1.0205	1.0416
FM(4, 4, 1)	1.2593	1.2245	1.2291
FM(4, 4, 2)	1.2963	1.3470	1.1667

^(a) Along the rows alternative RMSE of FM(*l*, *b*, *k*) ratios with respect to the baseline forecast FM(4, 1, 2) are reported; where *l* = number of lags in expanding the panel of indicators prior to principal components extraction; *b* = number of blocks in which the input panel is structured; *k* = number of principal components extracted with SW. The row in italics reports the RMSE of the baseline FM(4, 1, 3).

^(b) RMSE levels.

Fig. 1 - RMSE ratios to AR benchmark ^{a, b}



(^a) Codes represented along the horizontal axis indicate alternative models: 1=BM; 2=PFM*hard*90; 3=PFM*soft*90; 6=PFM*hard*75; 7=PFM*soft*75; 10=PFM*hard*50; 11=PFM*soft*50; 14=PFM*hard*25; 15=PFM*soft*25; 18=PFM*hard*10; 19=PFM*soft*10; 22=PFM*hard*min; 23=PFM*soft*min; 26=FM. For a short description of the models see Table 1; more details are in Section 4.

(^b) For each code, there are three bars representing: 1st vintage (in grey), 2nd vintage (in white) and 3rd vintage (nowcast, in black) RMSE ratios.

Appendix – The indicators dataset

Tab. A.1. - The list of the indicators used in PFM and FM approaches ^a

Variable	T	F	O	C	A	S	Variable	T	F	O	C	A	S
EA IND. CONF.	1	m	ea	s	1	sa	BD TOTAL 1-3Y GOVT.	2	d	ea	f	0	-
EA RET. CONF.	1	m	ea	s	1	sa	ES BMK. 10Y GOVT.	2	d	ea	f	0	-
EA BUILD. EMPL. EXP.	1	m	ea	s	1	sa	ES BMK. 5Y GOVT.	2	d	ea	f	0	-
EA BUILD. ORDER B. POSITION	1	m	ea	s	1	sa	ES TOTAL 1-3Y GOVT.	2	d	ea	f	0	-
EA BUILD. PRICE EXP.	1	m	ea	s	1	sa	EA CPI ALL ITEMS (HARMONISED)	2	m	ea	p	1	sa
EA IND. EMPL. EXP.	1	m	ea	s	1	sa	EA CPI GDS. EXCL. SVS.	2	m	ea	p	1	sa
EA IND. EXP. ORDER B. LEV.	1	m	ea	s	1	sa	EA CPI NON NRG. IND. GDS. (DUR.)	2	m	ea	p	1	sa
EA IND. ORDER B. LEV.	1	m	ea	s	1	sa	EA CPI NON NRG. IND. GDS.	2	m	ea	p	1	sa
EA IND. PROD. EXP.	1	m	ea	s	1	sa	EA CPI NON NRG. IND. GDS. (NON DUR.)	2	m	ea	p	1	sa
EA IND. PROD. TREND	1	m	ea	s	1	sa	EA CPI NON NRG. IND. GDS. (SEMI DUR.)	2	m	ea	p	1	sa
EA RET. EMPL. EXP.	1	m	ea	s	1	sa	EA CPI EXCL. NRG. FOOD	2	m	ea	p	1	sa
EA RET. EXP. BUS. SIT.	1	m	ea	s	1	sa	EA CPI ELEC.	2	m	ea	p	1	sa
EA RET. INTNS. PLACING ORDS.	1	m	ea	s	1	sa	EA CPI NRG.	2	m	ea	p	1	sa
EA RET. BUS. SIT.	1	m	ea	s	1	sa	EA CPI SVS. EXCL. GDS.	2	m	ea	p	1	sa
EA RET. VOL. OF STOCKS	1	m	ea	s	1	sa	EA CPI IND. GDS.	2	m	ea	p	1	sa
EA CONS. CONF.	1	m	ea	s	1	sa	EA PPI (EXCL. BUILD.)	2	m	ea	p	2	sa
EA IND. MFG. SELLING PRICE EXP.	1	m	ea	s	1	sa	EA PPI EXCL. BUILD. & NRG.	2	m	ea	p	2	sa
EA IND. MFG. STOCKS OF FIN. PROD.	1	m	ea	s	1	sa	EA IPI CONS. NON DUR.S	2	m	ea	h	2	sa
EA BUILD. CONF.	1	m	ea	s	1	sa	EA IPI CPTL. GDS.	2	m	ea	h	2	sa
EA BUILD. ACT. WRT PREV. MO	1	m	ea	s	1	sa	EA IPI INTERMEDIATE GDS.	2	m	ea	h	2	sa
EA CONS. ECON. SIT. LAST 12M	1	m	ea	s	1	sa	EA IPI CONS. DUR.S	2	m	ea	h	2	sa
EA CONS. ECON. SIT. NEXT 12M	1	m	ea	s	1	sa	EA IPI CONS. GDS.	2	m	ea	h	2	sa
EA CONS. MAJOR PURCH.	1	m	ea	s	1	sa	EA IPI MFG. ELECT. PRDS	2	m	ea	h	2	sa
EA CONS. MAJOR PURCH. NEXT 12M	1	m	ea	s	1	sa	EA IPI MFG. CHEMICAL PRDS.	2	m	ea	h	2	sa
EA CONS. PRICE TRENDS LAST 12M	1	m	ea	s	1	sa	EA IPI MFG. REF. PETR. PROD.	2	m	ea	h	2	sa
EA CONS. PRICE TRENDS NEXT 12M	1	m	ea	s	1	sa	EA IPI MFG. FOOD PRDS.	2	m	ea	h	2	sa
EA CONS. SAV.	1	m	ea	s	1	sa	EA IPI MFG. OTH. ELEC.	2	m	ea	h	2	sa
EA CONS. SAV. NEXT 12M	1	m	ea	s	1	sa	EA IPI EXCL. BUILD.	2	m	ea	h	2	sa
EA CONS. UNEMPL. NEXT 12M	1	m	ea	s	1	sa	EA IPI BUILD.	2	m	ea	h	3	sa
EA CONS. HSLD. FIN. SIT. LAST 12M	1	m	ea	s	1	sa	EA IPI NRG.	2	m	ea	h	2	sa
EA CONS. HSLD. FIN. SIT. NEXT 12M	1	m	ea	s	1	sa	EA IPI MFG. ELEC. DISTRB.	2	m	ea	h	2	sa
EA CONS. HSLD. FIN. SIT.	1	m	ea	s	1	sa	EA IPI MFG. PLASTIC MACH.	2	m	ea	h	2	sa
EA CONS. CONF.	1	m	ea	s	1	sa	EA IPI MFG. BAS MTL.S.	2	m	ea	h	2	sa
EA CONS. ECON. SIT. LAST 12M	1	m	ea	s	1	sa	EA IPI MFG. TEXT.	2	m	ea	h	3	sa
EA CONS. ECON. SIT. NEXT 12M	1	m	ea	s	1	sa	EA IPI MFG. WOOD	2	m	ea	h	2	sa
EA CONS. HSLD. FIN. SIT. LAST12M	1	m	ea	s	1	sa	EA IPI MFG. PHARMA. PRDS.	2	m	ea	h	2	sa
EA CONS. HSLD. FIN. SIT. NEXT12M	1	m	ea	s	1	sa	EA IPI ELEC. SUPPLY	2	m	ea	h	2	sa
EA CONS. HSLD. FINANCIAL SIT.	1	m	ea	s	1	sa	EA IPI MFG. REPAIR MACH.	2	m	ea	h	2	sa
EA CONS. MAJOR PURCH.	1	m	ea	s	1	sa	EA RET. SALES T/OVER. TOTAL	2	m	ea	h	2	sa
EA CONS. MAJOR PURCH. NEXT 12M	1	m	ea	s	1	sa	EA RET. T/OVER. EXCL. MV	2	m	ea	h	2	sa
EA CONS. PRICE TRENDS LAST 12M	1	m	ea	s	1	sa	EA RET. T/OVER. NON FOOD PRODS.	2	m	ea	h	2	sa
EA CONS. PRICE TRENDS NEXT 12M	1	m	ea	s	1	sa	EA RET. T/OVER. EXCL. MV & FUEL	2	m	ea	h	2	sa
US CONS. CONF.	1	m	us	s	1	sa	EA RET. T/OVER. FOOD & BEV.	2	m	ea	h	2	sa
US CONS. CONF. EXP.	1	m	us	s	1	sa	EA CAR REG.	2	m	ea	h	2	sa
US CONS. CONF. SIT.	1	m	us	s	1	sa	EA IND. MV TRAILERS ORDER B.	1	m	ea	h	1	sa
US CONS. CONF. IN 6M BUY H.	1	m	us	s	1	sa	EA IND. MV TRAILERS EXP. ORDER B.	1	m	ea	h	1	sa
US CONS. CONF. CURR. JOBS HTF.	1	m	us	s	1	sa	EA CONS. CAR PURCHASE NEXT 12M	1	q	ea	s	0	sa
US CONS. CONF. CURR. JOBS PLENT.	1	m	us	s	1	sa	EA CONS. H. IMPTS. NEXT 12M	1	q	ea	s	0	sa
US CONS. CONF. IN 6M JOBS MORE	1	m	us	s	1	sa	EA CONS. H. PURCHASE NEXT 12M	1	q	ea	s	0	sa
US CONS. CONF. IN 6M JOBS SAME	1	m	us	s	1	sa	EA IND. CURR. PROD. CAP.	1	q	ea	s	0	sa
US CONS. CONF. BUS. CNDT. BAD	1	m	us	s	1	sa	EA IND. LIMITS TO PROD.	1	q	ea	s	0	sa
US CONS. CONF. BUS. CNDT. GOOD	1	m	us	s	1	sa	EA IND. LIMITS TO PROD. DEM.	1	q	ea	s	0	sa
US CONS. CONF. BUS. CNDT. NORM.	1	m	us	s	1	sa	EA IND. LIMITS TO PROD. EQPT.	1	q	ea	s	0	sa
US CONS. CONF. JOBS NSP.	1	m	us	s	1	sa	EA IND. LIMITS TO PROD. FNCL.	1	q	ea	s	0	sa

US CONS. CONF. 6M EC. CNDT. BETTER	1	m	us	s	1	sa	EA IND. LIMITS TO PROD. LAB.	1	q	ea	s	0	sa
US CONS. CONF. 6M EC. CNDT. SAME	1	m	us	s	1	sa	EA IND. LIMITS TO PROD. OTH.	1	q	ea	s	0	sa
US CONS. CONF. 6M EC. CNDT. WORSE	1	m	us	s	1	sa	EA IND. MFG. CAP. UT.	1	q	ea	s	0	sa
US CONS. CONF. 6M INC. DECREASED	1	m	us	s	1	sa	EA CPTL. SHORTAGE	0	q	ea	s	0	sa
US CONS. CONF. 6M INC. INCREASED	1	m	us	s	1	sa	EA USD WRT TO GBP	0	q	ea	s	0	su
US CONS. CONF. 6M INC. SAME	1	m	us	s	1	sa	EA USD WRT TO US\$	0	q	ea	s	0	su
US CAP. UT. RATE. MFG.	1	m	us	s	2	sa	EA USD WRT TO YEN	0	q	ea	s	0	su
US AVG CONS. EXP. FOR BUS. CNDT.	1	m	us	s	2	sa	EA ECON. CLIM.	0	q	ea	s	0	su
US CAP. UT. RATE. IND.	1	m	us	s	2	sa	EA ECON. EXP.	0	q	ea	s	0	su
US VEND. PERF. SLOWER DEL. DIFF.	1	m	us	s	2	sa	EA ECON. SIT. CPTL. EXPEND. S	0	q	ea	s	0	su
US CHI PMI BUS. BAROMETER (SA)	1	m	us	s	1	sa	EA ECON. SIT. OVERALL ECON.	0	q	ea	s	0	su
US CHI PMI BACKLOG	1	m	us	s	1	sa	EA ECON. SIT. PRIV. CMPT.	0	q	ea	s	0	su
US CHI PMI DELIVERIES	1	m	us	s	1	sa	EA ECON. SIT. NEXT 6M CPTL. EXPND.	0	q	ea	s	0	su
US CHI PMI EMPLMT.	1	m	us	s	1	sa	EA ECON. SIT. NEXT 6M OVERALL EC.	0	q	ea	s	0	su
US CHI PMI INVENTORIES	1	m	us	s	1	sa	EA ECON. SIT. NEXT 6M PRIV. CMPT.	0	q	ea	s	0	su
US CHI PMI NEW ORDERS	1	m	us	s	1	sa	EA ECON. SIT. CPTL. EXPND.	0	q	ea	s	0	su
US CHI PMI PRICES PAID	1	m	us	s	1	sa	EA ECON. SIT.	0	q	ea	s	0	su
US CHI PMI PROD.	1	m	us	s	1	sa	EA ECON. SIT. OVERALL ECON.	0	q	ea	s	0	su
US CHI PMI PROD. AV. DAYS	1	m	us	s	1	sa	EA ECON. SIT. PRIV. CNSPT.	0	q	ea	s	0	su
WTI CRUDE OIL	2	d	w	f	0	-	EA EXP. INFL. R. (% AVG. OFY)	0	q	ea	s	0	su
NAT. GAS	2	d	w	f	0	-	EA EXP. VOL. NEXT 6M	0	q	ea	s	0	su
NORTH SEA CRUDE OIL	2	d	w	f	0	-	EA FGN. INV. NEXT 6M ADMIN. RCTNS.	0	q	ea	s	0	su
WHEAT US	2	d	w	f	0	-	EA FGN. INV. NEXT 6M POL. STAB.	0	q	ea	s	0	su
GAS OIL	2	d	w	f	0	-	EA FGN. INV. ADMIN. RCTNS.	0	q	ea	s	0	su
TR. EW. CCI ENGY.	2	d	w	f	0	-	EA FGN. INV. POL. STAB.	0	q	ea	s	0	su
TR. EW. CCI IND.	2	d	w	f	0	-	EA FRGN. DEBTS	0	q	ea	s	1	su
TR. EW. CCI INT.	2	d	w	f	0	-	EA FRGN. INV. CLIM.	0	q	ea	s	0	su
TR. EW. CCI LIVESTOCK	2	d	w	f	0	-	EA FRGN. INV. CLIM. NEXT 6M	0	q	ea	s	0	su
TR. EW. CCI PRECIOUS MTLs.	2	d	w	f	0	-	EA IMP. VOL. NEXT 6M	0	q	ea	s	0	su
TR. EW. CCI SOFTS	2	d	w	f	0	-	EA INFL.	0	q	ea	s	1	su
TR. EW. CCI GRAINS	2	d	w	f	0	-	EA INFL. RATE NEXT 6M	0	q	ea	s	0	su
TR. EW. CCI	2	d	w	f	0	-	EA INSUFF. DEMAND	0	q	ea	s	1	su
EURO STOXX BANKS	2	d	ea	f	0	-	EA LACK OF CONF. GOVT. ECON. POL.	0	q	ea	s	1	su
EURO STOXX INSURANCE	2	d	ea	f	0	-	EA LACK OF INTL. COMP.	0	q	ea	s	1	su
EURO STOXX CHEMICALS	2	d	ea	f	0	-	EA LACK OF SKILLED LABOUR	0	q	ea	s	1	su
EURO STOXX TELECOM	2	d	ea	f	0	-	EA PUBL. DEF.	0	q	ea	s	1	su
EURO STOXX UTILITIES	2	d	ea	f	0	-	EA SHORT TERM INT. R. NEXT 6M	0	q	ea	s	0	su
EURO STOXX AUTO & PARTS	2	d	ea	f	0	-	EA TRD. BALANCE NEXT 6M	0	q	ea	s	0	su
EURO STOXX BASIC MATS	2	d	ea	f	0	-	EA TRD. BARRIERS TO EXP.	0	q	ea	s	1	su
EURO STOXX BASIC RESOURCE	2	d	ea	f	0	-	EA UNEMPL.	0	q	ea	s	1	su
EURO STOXX CON & MAT	2	d	ea	f	0	-	EA VALUE OF US\$ NEXT 6M	0	q	ea	s	0	su
EURO STOXX FINANCIAL SVS	2	d	ea	f	0	-	US CPTL. SHORTAGE.	0	q	us	s	1	su
EURO STOXX FINANCIALS	2	d	ea	f	0	-	US USD WRT TO EURO	0	q	us	s	0	su
EURO STOXX FOOD & BEV	2	d	ea	f	0	-	US USD WRT TO GBP	0	q	us	s	0	su
EURO STOXX HEALTH	2	d	ea	f	0	-	US USD WRT TO YEN	0	q	us	s	0	su
EURO STOXX HEALTH CARE	2	d	ea	f	0	-	US ECON. SIT. LASTY. CPTL. EXPND.	0	q	us	s	0	su
EURO STOXX INDS GDS & SVS	2	d	ea	f	0	-	US ECON. SIT. LASTY. OVERALL EC.	0	q	us	s	0	su
EURO STOXX INDUSTRIALS	2	d	ea	f	0	-	US ECON. SIT. LASTY. PRIV. CMPT.	0	q	us	s	0	su
EURO STOXX MEDIA	2	d	ea	f	0	-	US ECON. SIT. NEXT 6M CPTL. EXP.	0	q	us	s	0	su
EURO STOXX OIL & GAS	2	d	ea	f	0	-	US ECON. SIT. NEXT 6M OVERALL EC.	0	q	us	s	0	su
EURO STOXX OIL & GAS	2	d	ea	f	0	-	US ECON. SIT. NEXT 6M PRIV. CNSPT.	0	q	us	s	0	su
EURO STOXX TECH.	2	d	ea	f	0	-	US ECON. SIT. CPTL. EXPND.	0	q	us	s	0	su
EURO STOXX TECH.	2	d	ea	f	0	-	US ECON. SIT. OVERALL ECON.	0	q	us	s	0	su
EURO STOXX TELECOM	2	d	ea	f	0	-	US ECON. SIT. PRIV. CNSPT.	0	q	us	s	0	su
EURO STOXX UTILITIES	2	d	ea	f	0	-	US EXP. INFL. RATE(% AVG. OFY)	0	q	us	s	0	su
EURO STOXX 50	2	d	ea	f	0	-	US EXP. VOL. NEXT 6M	0	q	us	s	0	su
STOXX EUROPE 600	2	d	ea	f	0	-	US FGN. INV. NEXT 6M POL. STAB.	0	q	us	s	0	su
EURO STOXX.	2	d	ea	f	0	-	US FGN. INV. ADMIN. RCTNS.	0	q	us	s	0	su
STOXX EUROPE 50	2	d	ea	f	0	-	US FGN. INV. POL. STAB.	0	q	us	s	0	su
STOXX EUROPE SMALL 200	2	d	ea	f	0	-	US FGN. INV. NEXT 6M ADMIN. RCTNS.	0	q	us	s	0	su
STOXX EUROPE LARGE 200	2	d	ea	f	0	-	US FRGN. DEBTS.	0	q	us	s	1	su

STOXX EUROPE MID 200	2	d	ea	f	0	-	US FRGN. INV. CLIM.	0	q	us	s	0	su
GBP TO EURO EXCH. R.	2	d	ea	f	0	-	US FRGN. INV. CLIM. NEXT 6M	0	q	us	s	0	su
US \$ TO EURO EXCH. R.	2	d	ea	f	0	-	US IMP. VOL. NEXT 6M	0	q	us	s	0	su
SWISS FRANC TO EURO EXCH. R.	2	d	ea	f	0	-	US INFL.	0	q	us	s	1	su
CAN \$ TO EURO EXCH. R.	2	d	ea	f	0	-	US INFL. RATE NEXT 6M	0	q	us	s	0	su
JAP YEN TO EURO EXCH. R.	2	d	ea	f	0	-	US INSUFF. DEMAND.	0	q	us	s	1	su
EURO BMK. BOND RED. YLD.	1	d	ea	f	0	-	US LACK OF CONF. GMNT. ECON. POL.	0	q	us	s	1	su
IT BMK. 10Y GOVT.	2	d	ea	f	0	-	US LACK OF INTL. COMP.	0	q	us	s	1	su
IT BMK. 5Y GOVT.	2	d	ea	f	0	-	US LACK OF SKILLED LABOUR	0	q	us	s	1	su
IT TOTAL 1-3Y GOVT.	2	d	ea	f	0	-	US PUBL. DEF.	0	q	us	s	1	su
FR BMK. 10Y GOVT.	2	d	ea	f	0	-	US SHORT TERM INT. RATES NEXT 6M	0	q	us	s	0	su
FR BMK. 5Y GOVT.	2	d	ea	f	0	-	US TRD. BALANCE NEXT 6M	0	q	us	s	0	su
FR TOTAL 1-3Y GOVT.	2	d	ea	f	0	-	US TRD. BARRIERS TO EXP.	0	q	us	s	1	su
BD BMK. 10Y GOVT.	2	d	ea	f	0	-	US UNEMPL.	0	q	us	s	1	su
BD BMK. 5Y GOVT.	2	d	ea	f	0	-							

(^a) Legenda of the columns:

T is type of transformation (0 = no transformation, 1 = first differences of levels, 2 = first differences of log-levels);

F is the frequency (d = daily, m = monthly, q = quarterly);

C is the geographical area (ea = Euro area, us = United States, w = international/world);

B is the block (s = qualitative indicators, h = quantitative indicators, f = financial variables, p = prices);

A is the availability, i.e. the maximum number of periods to be forecast to complete the quarter (from 0 to 3);

S indicates seasonality treatment (sa = seasonally adjusted, su = seasonally unadjusted, - not seasonal).



Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
<http://www.dse.unibo.it>