

Leading indicators for Euro-area Inflation and GDP growth ^{*}

Anindya Banerjee[¶]
Department of Economics
European University Institute
Via della Piazzuola, 43
50133 Firenze (FI), Italy
anindya.banerjee@iue.it

Massimiliano Marcellino
IEP-Bocconi University, IGIER
and CEPR
Via Salasco, 5
20136, Milano, Italy
massimiliano.marcellino@uni-bocconi.it

Igor Masten
Faculty of Economics
University of Ljubljana
Kadeljeva pl. 17
1101 Ljubljana, Slovenia
igor.masten@ef.uni-lj.si

September 2005

Abstract

In this paper we evaluate the role of a set of variables as leading indicators for Euro-area inflation and GDP growth. Our leading indicators are taken from the variables in the ECB Euro-area-wide model database, plus a set of similar variables for the US. We compare the forecasting performance of each indicator ex-post with that of purely autoregressive models. We also analyze three different approaches to combining the information from several indicators. First, ex-post, we discuss the use as indicators of the estimated factors from a dynamic factor model for all the indicators. Second, within an ex-ante framework, an automated model selection procedure is applied to models with a large set of indicators. No future information is used, future values of the regressors are forecast, and the choice of the indicators is based on their past forecasting records. Finally, we consider the forecasting performance of groups of indicators and factors and methods of pooling the ex-ante single-indicator or factor-based forecasts. Some sensitivity analyses are also undertaken for different forecasting horizons and weighting schemes of forecasts to assess the robustness of the results.

JEL Classification: C53, E37, C50

Keywords: leading indicator, factor model, model selection, GDP growth, inflation

[¶] Corresponding author: Tel: 00 39 055 468 5956, FAX: 00 39 055 468 5902, e-mail: anindya.banerjee@iue.it.

^{*}While not implicating them for any infelicities that remain, it is our pleasure to thank Antoni Espasa, Antonio Montañés, the Editors of the Special Issue, and two anonymous referees for their helpful comments on our paper. The financial generosity of the EUI under research grants EFN 402 and I 216 and of the Italian MIUR is gratefully acknowledged. Early versions of this paper were presented by the first-named author to seminars in the ECB, the Department of Statistics and Econometrics, University Carlos III, Madrid, and to the Aragonese Foundation hosted under the auspices of the Department of Economics in the University of Zaragoza.

1. Introduction

Inflation and GDP growth are probably the two most important macroeconomic variables, as they drive monetary and fiscal policy. Methods for forecasting these two variables have been the subject of much intensive research in econometrics, and several studies are by now available also for the Euro area. For example, Marcellino (2004a) evaluates the performance of a large set of univariate forecasting methods, finding that simple autoregressive models perform well although for some series using nonlinear methods produces forecasting gains. Marcellino, Stock and Watson (2003) adopt factor models for forecasting industrial production, inflation and unemployment both for the Euro area as a whole and for its member countries, and find some gains with respect to autoregressions, in particular for nominal variables. Camba-Mendez, Kapetanios, Smith and Weale (2001) develop automated procedures using systems of leading indicators. Fagan, Henry and Mestre (2001) and Dreger and Marcellino (2003) construct a medium scale macroeconomic model for the Euro-area variables, forecasts from which, in general, outperform those derived from time-series models.

The aim of our analysis is to conduct a detailed evaluation of the properties of a large set of leading indicators for Euro-area inflation and GDP growth, using not only Euro-area series but also US macroeconomic variables. Received wisdom suggests, for example, that the links between the US and the Euro area could be important, with Euro-area growth depending upon US growth, and the European Central Bank's (ECB) decisions following in part the policies of the Federal Reserve Board (Fed). Marcellino *et al.* (2003) find few gains from using US industrial production and inflation for forecasting their Euro-area counterparts. Here we conduct a more detailed analysis by evaluating a larger set of US indicators.

Following Banerjee and Marcellino (2005), we first compare the performance of single indicator models with pure autoregressions. The comparison is conducted using an ex-post approach. In the ex-post evaluation, future values of the exogenous regressors are assumed to be known in order to evaluate the informational content of the indicators independently of their forecastability. This provides the maximum advantage against the autoregressive models but as we shall see, in many cases, it is not enough to outperform them.

We next exploit the joint information set in three ways. First, ex-post, we model all the indicators by means of a dynamic factor model and use the estimated factors as leading indicators. This method performed well in terms of average mean square and mean absolute forecast error for the US (Stock and Watson, 2002a, 2002b), the Euro area (Marcellino *et al.*,

2003), the UK (Artis, Banerjee and Marcellino, 2005), and the Acceding countries (Banerjee, Marcellino and Masten, 2005).

Second, within an ex-ante framework, we jointly consider groups of indicators and an automated model selection procedure to obtain a parsimonious forecasting model. Previous in-sample results on the performance of this procedure were quite encouraging (Hendry and Krolzig, 1999 and Krolzig and Hendry, 2001).

Third, we evaluate the efficacy of pooling procedures based on combining the forecasts derived from single indicators and groups of indicators derived above. This evaluation is conducted ex-ante to reflect the exigencies of the real world. Since the pioneering work of Bates and Granger (1969), forecast combination has often been found to improve the forecasting performance (see e.g. Stock and Watson, 2003 and Marcellino, 2004b, for recent evidence on, respectively, the US and the Euro area).

We evaluate all these competing forecasting approaches using a particular criterion that is most relevant for policy making, where the same model is adopted for forecasting at several horizons, it is periodically evaluated (and possibly re-specified and/or re-estimated), and the goal is to obtain robust forecasts that perform well on a year-by-year basis and not only on average over a long period of time. This criterion is denoted by $RMSE-h$ and described in detail below.

In the current paper we focus solely on forecasting the growth rate of GDP and inflation, although turning point forecasts could also be of interest. The formulation of turning-point forecasts requires either the adoption of models for binary variables such as logit or probit, or simulation methods using models for continuous variables (see e.g. Marcellino, 2004c, for an overview). However, the former approach requires the specification of a different model for each forecast horizon, and the econometric theory for the use of large information sets in this context is as yet underdeveloped. The latter is very computationally intensive in our context where many indicators are compared. Therefore, we leave the evaluation of alternative methods and indicators for turning point forecasts for future research.

The paper is organized as follows. In Section 2 we briefly review the forecasting methods under analysis. In Section 3 we discuss the Euro-area and the US data sets. In Sections 4 and 5 we present the results of the forecasting exercise for, respectively, inflation and GDP growth. In Section 6 we summarize the main results and conclude.

2. The methodology

In this section we describe the forecasting methodologies we use and the evaluation criterion we adopt to rank the competing methods. We deal in turn with single-indicator forecasts, factor forecasts, automated model selection based forecasts, forecast combination procedures and, finally, the forecast evaluation criterion.

2.1 Single indicator forecasts

The estimated model takes the form

$$Y_t = \alpha + \sum_{i=1}^m \beta_i Y_{t-i} + \sum_{i=1}^k \delta_i IND_{t-i} + \varepsilon_t, \quad (1)$$

where Y_t is the variable of interest and IND_{t-i} is the i^{th} lag of the particular indicator variable chosen. The values of m and k are determined by the automated model selection procedure described below, since as reported in Banerjee and Marcellino (2005) there are in general gains from using this approach to determine the number of lags instead of a fixed lag length. Pure autoregressions are a special case of (1) when the indicator variables are excluded.

The model (1) is used to produce one- up to h -step-ahead forecasts of the Y variable, and compared with the forecasts arising from an autoregression (where the number of lags in the pure autoregression are also always chosen using an automatic model selection algorithm). Stock and Watson (1999, 2002a, 2002b) and Marcellino (2004a) use dynamic estimation for forecasting, but since this method requires us to specify and estimate a different model for each forecast horizon, it is computationally cumbersome in our context with several horizons and many indicators. Moreover, Marcellino, Stock and Watson (2005) provide empirical evidence against dynamic estimation for forecasting purposes, although in general the differences with respect to iterated forecasts are minor.

Whenever out-of-sample values of the Y variable are required to generate forecasts, the forecast value is used. In the ex-post evaluation, unknown values of the leading indicator variable are replaced by the actual values. This framework biases the analysis in favour of the indicator model versus the pure autoregressive model and is adopted to evaluate the information content of the indicator, which could otherwise be hidden by its poor forecastability. We will also consider below a pseudo ex-ante context where unknown values of the leading indicator variable are replaced by forecasts from autoregressive models.

2.2 Factor-based forecasts

Dynamic factor-models provide an efficient tool for extracting information from a large database, so that instead of a single indicator variable we can use the estimated factors from a set of indicators to forecast the variable of interest. This forecasting technique has recently been successfully applied to forecasting US, UK and Euro-area macroeconomic variables (Stock and Watson, 2002a, 2002b, Artis *et al.*, 2005 and Marcellino *et al.*, 2003 respectively). There are some country-specific differences in the types of variables for which the forecasting gains are larger. Typically factor forecasts are better for real variables for the US and nominal variables for Europe. Here we briefly introduce the representation and estimation theory for the dynamic factor model, while Geweke (1977), Sargent and Sims (1977), Forni, Hallin, Lippi, and Reichlin (2000) and, in particular, Stock and Watson (2002a, 2002b) may be consulted for details.

The N -macroeconomic variables to be modelled, grouped in the vector X_t , admit an approximate linear dynamic factor representation with \bar{r} common factors, f_t , if:

$$X_{it} = \lambda_i(L)f_t + e_{it} \quad (2)$$

for $i=1, \dots, N$, where e_{it} is an idiosyncratic disturbance with limited cross-sectional and temporal dependence, and $\lambda_i(L)$ are lag polynomials in non-negative powers of L . If $\lambda_i(L)$ have finite orders of at most q , the model in (2) can be rewritten as,

$$X_t = \Lambda F_t + e_t \quad (3)$$

where $F_t = (f_t', \dots, f_{t-q}')'$ is $r \times 1$, where $r \leq (q+1)\bar{r}$, and the i -th row of Λ in (2) is $(\lambda_{i0}, \dots, \lambda_{iq})$.

Stock and Watson (2002a, 2002b) prove that, under some technical assumptions (restrictions on moments and stationarity conditions), the column space spanned by the dynamic factors f_t can be estimated consistently by the principal components of the $T \times T$ covariance matrix of the X 's.¹ A condition that is worth mentioning for the latter result to hold is that the number of factors included in the estimated model has to be equal or larger than the true number.

The empirical analyses mentioned above have shown that two or three factors are sufficient to explain a large proportion of the variability of a large set of time series. We allow for estimation of up to six factors in what follows, and use them for forecasting the variables of interest in two ways. First, ex-post, in Sections 4.1 and 5.1 we use factors one by one in the

¹ Notice that the fact that the column space rather than the factors themselves can be estimated is not problematic for forecasting since the column space provides a summary of the information contained in the data set.

same way as we use single indicators in model (1). Second, in Sections 4.2 and 5.2, factors are used jointly as part of the ex-ante analysis. In such a case, the starting specification of the model for Y is

$$Y_t = \alpha + \sum_{i=1}^m \beta_i Y_{t-i} + \sum_{j=1}^6 \sum_{i=1}^k \delta_{j,i} f_{j,t-i} + \varepsilon_t \quad (4),$$

with the factor-choice at time t (in order to forecast say $t+4$) being determined by the best-performing (group of) factors at time $t-1$. An important result by Stock and Watson (2002a, 2002b), further refined by Bai and Ng (2003), is that the use of the estimated factors instead of the true factors in equation (4) above does not affect the mean square forecast error asymptotically.

2.3 Automated model selection

PcGets is a computer-automated algorithm for general to specific reductions of models developed by Hendry and Krolzig (1999) and Krolzig and Hendry (2001), see also Hoover and Perez (1999). The starting point for the algorithm is the specification of a general unrestricted model (GUM) containing all variables likely (or specified) to be relevant, including the maximum lag length of the independent and dependent variables. The algorithm starts from a ‘pre-search’ simplification by applying tests for variable deletion, following which the GUM is simplified. This step uses a loose significance level such as 10%, to delete highly non-significant regressors. The procedure is refined at the second stage, where many alternative further reductions of the GUM are considered, using both t and F tests and information criteria as reduction (or deletion of variables) criteria. Diagnostic tests ensure that the models chosen as valid simplifications/reductions are congruent representations of the data. The third stage is the encompassing step (see *e.g.* Mizon and Richard, 1986) where all valid reduced models from the second step are collected, and encompassing tests are used to evaluate the relative merits of these competing models. Only models that are not encompassed are retained. If more than one model survives the third stage, their union forms the new general model and the algorithm recommences. This process continues until the set of non-encompassed models reduces to one or the union is repeated.

We use *PcGets* in three ways: first to select the lag length of the Y variable and of the indicator in model (1), starting with a maximum of 6 lags; second, to determine what factors and how many lags should be used in the model (4), starting with 6 factors and 6 lags. Both these uses appear in Sections 4.1 and 5.1. Our third use of *PcGets* is to simplify a general model where several indicators are jointly considered as explanatory variables for Y , i.e.,

$$Y_t = \alpha + \sum_{i=1}^m \beta_i Y_{t-i} + \sum_{j=1}^q \sum_{i=1}^k \delta_{j,i} IND_{j,t-i} + \varepsilon_t \cdot \quad (5)$$

We select the indicators to be included in the GUM (5) based on their forecasting performance as single indicators, as described in more detail in Section 4.2 below.

2.4 Forecast combination

The factor approach and the model simplification method described above are alternative, possibly complementary, procedures to summarize large information sets into a relatively small explanatory model for the Y variable, which is then used for forecasting. As an alternative, the information in the large set of indicators could be exploited by combining the single indicator forecasts.

Bates and Granger (1969) advocated the use of combination of forecasts as a tool to reduce the root-mean-square forecast error (RMSE). Since then several studies have found this method useful including Stock and Watson (1999, 2003), Clements and Hendry (2004) and Marcellino (2004b). The weights should in principle depend upon the entire covariance matrix of the forecasts to minimize the RMSE. Since this is too complicated in our framework with many forecasts, we will consider two simpler procedures that have performed well in similar analyses such as Stock and Watson (1999, 2003), namely a simple average of a number of best performing single indicator forecasts and the median of these forecasts. The latter could be more robust since we will see that some indicators produce forecasts with high RMSEs in some periods. The results are contained in Tables 6 and 12 and are also discussed in Sections 4.2 and 5.2 below within the context of the ex-ante evaluation exercise.

2.5 Forecast evaluation

The models in (1), (4) and (5) are estimated with a starting sample of 10 years, and are then re-estimated adding one year each time. For each estimation period, one- up to four- step-ahead forecasts are computed, and the loss function is constructed as the square root of the average squared forecast errors one- up to four-steps ahead, denoted by RMSE- h . For example, in the case of inflation, the models are estimated first from 1975:1 to 1984:4 to provide forecasts for the four quarters up to 1985:4. The estimation sample is next augmented by one year (*i.e.* until 1985:4) and the models are re-estimated to forecast inflation for 1986.

This exercise of augmentation is continued recursively until the estimation sample extends to 1999:4 and forecasts are provided for the four quarters in 2000.²

This procedure, adopted for example by Cecchetti, Chu and Steindel (2000), differs from the standard practice of taking averages over the whole forecasting period of the forecast errors computed for a fixed horizon. Its main advantage is that it is closer to the practice of forecast evaluation by policy makers and practitioners, where the same model is used to forecast at different horizons and the interest is in the periodic evaluation of the model (and possibly in its periodic re-specification). This is done since we wish to keep track of forecasting performance over the estimation and forecasting samples recursively, instead of simply comparing average forecasting accuracy. Another important benefit is that the evaluation is more robust to structural changes over the forecast samples, which are quite frequent; see *e.g.* Stock and Watson (2001).

We also argue that the use of the standard average of the fixed horizon root-mean-square errors over a reasonably long period of time can be misleading, by hiding many interesting characteristics of the indicators. In particular, some indicators can outperform autoregressive models on average but forecast very poorly in some periods. This has serious consequences if the forecasts are used in a policy-making environment. The fact that the indicators should be changed from period to period, depending on the likelihood of particular economic shocks over the forecasting period, does not emerge under the choice of a loss function, which averages over time.

The main drawback of our approach is that since the series of the computed $RMSE-h$ statistics is short and its elements are highly correlated, it is not possible to provide a reliable statistical test for a significant difference in forecasting performance. Nevertheless, we believe that the point $RMSE-h$ estimates provide a clear ranking of the competing forecasting models.

We have also considered two modifications of our evaluation criterion. First, we consider averaging one- up to eight-step-ahead forecasts, in order to cover a horizon of up to two years, which is usually of interest for policy making. Our results are robust to this modification. Moreover, since long horizon forecasts for stationary variables converge to the unconditional mean of the variable, the five up to eight-step-ahead forecasts can be hardly used to differentiate the indicators or the forecasting methods.

² In principle, the factors used in model (4) should also be recomputed for each sub-sample. In the paper we only use factors estimated for the whole period because this simplifies the computation of the results presented considerably and because the weights used to compute principal components are fairly stable over time.

Second, since shorter forecast horizons can be of greater interest than longer horizons, we have considered weighting the squared forecasts errors at different horizons either linearly or exponentially, where the weights decline with the horizon rather than being fixed. However, the findings are again robust to the use of these alternative loss functions. The detailed results of both modification exercises are available from us upon request.

3. The Data

The construction of quarterly Euro-area data for a long enough time span is complicated. A partial list of the problems includes the choice of the aggregation method (fixed versus time-varying weights, role of exchange rates, choice of the proper weighting variable, etc.); the presence of major redefinitions and institutional changes (modifications in the national accounting systems, the German reunification, etc.); the presence of missing observations and the need to disaggregate annual figures for some countries into quarterly values; and the choice of the seasonal and working day adjustment methods (for example whether to seasonally adjust the national series individually or directly the aggregated variable).

To focus on the topics of the paper, we have decided to use as a starting point a pre-existing and widely used data set, the one constructed by Fagan, Henry and Mestre (2001) for the ECB Euro-area-wide model. The data set includes quarterly data for the period 1970:1-2000:4, for several macroeconomic variables, mainly constructed using the so-called “Index Method”. For example, the logarithm of the Euro-area GDP is a weighted sum of the logarithms of the country specific GDPs, with constant weights based on the 1995 real GDP shares.³ We have then added some other indicators to our data set to reflect the choice of variables in Stock and Watson (1999, 2001), even though many fewer series are available for the Euro area as a whole than for its member countries or for the US.

We consider 46 Euro-area variables as indicators, listed in the Data Appendix (which also describes the type of data transformation adopted). They include output variables such as GDP, industrial production and some of their components; employment and productivity indicators; wages; exchange rates; interest rates and spreads; monetary aggregates; price indexes; and some other miscellaneous variables. Industrial production, some monetary aggregates and producer prices come from Eurostat and are available for a shorter time span.

³ More sophisticated aggregation methods have been suggested, *e.g.*, by Beyer, Doornik and Hendry (2001), but they have produced a very limited number of (monthly) Euro area series.

To the list above we add 16 US variables that broadly reflect the groupings of European indicators, to evaluate their role in forecasting Euro-area series. For the US we also consider as indicators the factors extracted from the larger data set in Banerjee and Marcellino (2005), that covers the sample 1975:1-2001:4 and includes about 50 indicators, listed in their Data Appendix.

4. Forecasting Inflation

4.1 Euro-area indicators, US variables and factors: Ex-post analysis

We have 46 indicators, fewer in the early periods, whose performance in the 16 evaluation periods is summarized in Table 1. Four main comments are in order, all reflective of the good performance of autoregressive models under ex-post evaluation.

First, looking at the second column of Table 1, the autoregression (AR) yields a lower RMSE- h than at least 50% of the models with an indicator in 12 out of the 16 periods. Second, among the indicators which out-perform the AR, only a subset yields gains of at least 10% in terms of reduction of the loss function (compare the third column of Table 1 with the second column). Thus, only in 3 out of 16 evaluation periods do more than 50% of the models with single indicators perform at least 10% better than the AR. Third, the performance of the AR model is rather unstable over time (consider the fourth column of Table 1), which provides evidence in favour of the adoption of our evaluation criterion instead of averaging mean squared errors over a long period over time for a fixed horizon. Finally, while the best performing indicator is always better than a pure AR (compare the fifth column of Table 1 with the fourth), the best indicator changes over time, likely to reflect the different shocks that hit the Euro area over the period under analysis, no indicator is best more than twice, and the loss involved is also rather volatile over time.

More detailed and disaggregate information on the forecasting performance of the single indicators is provided in the two panels of Table 2 and in Table 3. Panel (a) of Table 2 shows that 4 out of the 46 indicators do better than the autoregression in more than 11 out of the 16 evaluation periods. In this context it is worth noting that two labour market variables, namely wage growth and growth of compensation to employees, and the growth of the consumption deflator beat the AR in about 75% of the evaluation periods. Nevertheless when the evaluation criterion is tightened, it is clear from Panel (b) of Table 2 that all three variables are now placed in the second category (*i.e.* out-performing the AR by at least 10%

between 7 to 10 times out of 16 evaluation periods). However, it may also be noted from Panel (b) that in addition to these three variables there is a further set of indicators that is at least 10% better than the AR in at least 7 out of the 16 evaluation periods (i.e. close to 50% or more of the periods). These same indicators also show up in Table 3 where the top 5 indicators are listed period by period and are used in the ex-ante analysis in the next subsection.

By considering Tables 2 and 3 together, the best-performing single indicators for inflation can be grouped into four economic categories, namely: (i) labour market variables, including the unemployment rate, share of employment in total labour force and the growth of wages and labour productivity; (ii) particular prices, mainly the growth of the commodity prices and the private consumption deflator; (iii) fiscal variables, mainly expenditure and receipts (both in growth rates and as shares in GDP) and (iv) real variables such as the growth rate of GDP, total factor productivity growth, growth in gross investment at constant prices and private consumption growth. Trade-related and financial variables, such as the current account surplus, changes in the exchange rate, as well as interest rates spreads are also important but to a lesser degree. Rather surprisingly, among the worst performing indicators are, for example, the growth of M3, measures of industrial production (both in levels and in growth rates), producer prices and consumer sentiment indices.

Table 4 repeats the evaluation exercise undertaken in Table 1 for data taken from the US variables data set. In this case the ‘best’ leading indicator may be seen to be the growth of the real effective exchange rate for the US which outperforms the AR in 10 out of 16 cases (10% better on 7 out of these 10 periods), and also produces the best forecast on three occasions. The US inflation rate is ranked next, also outperforming the AR in 10 out of 16 cases (10% better on 6 occasions) but not providing the best forecast for any of these periods. The other good indicators are also reasonable from an economic point of view, including the capacity utilization rate, the growth in hourly earnings and a 3-month interest rate.

Table 5 summarises the forecasting performance of Euro-area and US factors. As mentioned previously, the US factors are extracted from the data set in Banerjee and Marcellino (2005) that includes a larger selection of variables, similar to the one we use here for the Euro area. The European factors are instead extracted from the full data set in Fagan *et al.* (2001). In both cases we consider the first six factors that explain a fraction larger than 50% of the variance of all the indicators.

One Euro-area factor (EU-F6) and two US factors (US-F4 and US-F3) do better than the AR in 9, 10, and 8 periods respectively out of 16, and in fact perform at least 10% better

than the AR in 7, 8 and 5 out of these periods. The next ranked set of factors (EU-F2 and EU-F4, and US-F6) do better than the AR in 7 out of 16 periods, with the second EU factor (EU-F2) doing better than the AR in all seven of these periods by more than 10%. The corresponding numbers for EU-F4 and US-F6 are 4 and 6 respectively. It may also be helpful to note that except for EU-F1 no factor model produces the best forecast more than once and the best single indicator usually beats each of them.

In summary, we may state that there is some role for indicators in helping to forecast inflation. For example, we have seen that labour market, price, output and government expenditure variables, as well as US inflation, are useful. Factors are efficacious, especially some of the US factors but it is interesting to note that they are not those with the highest explanatory power for the indicators. A related fact is that US-F1 is systematically deleted by *PcGets* from the forecasting model, and of course EU-F1 (while being highly effective in some cases) is dominated by EU-F2 and EU-F4 in terms of overall performance.

We now consider generalising the forecasting exercise in two different ways – first to allow for groups of variables, or factors or in combination to be considered, but also to allow the selection of variables (constituting these groups) to be made *ex-ante* (i.e. based on past forecasting performance.) We then ask whether the conclusions on the efficacy of leading indicators, based on *ex-post* evaluation are modified when using an *ex-ante* evaluation procedure.

4.2 Factor-based forecasts, groups of indicators and pooling: Ex-ante analysis

We have assumed so far that future values of the indicators are known when forecasting. This scenario provides the most favourable environment for the use of indicators, in the sense that if they do not perform well here they can be expected not to do so in real time. In fact, we have found that single indicators (or factors) can outperform the autoregression but the choice of the indicator has to be updated continuously. Moreover, since we consider a large set of indicators, the finding that some of them can beat the autoregression *ex post* could be due to data-mining. Therefore in this section we evaluate whether the autoregression can also be beaten in a pseudo-real-time framework.

Our method of *ex-ante* evaluation can best be described by an example. Say that we are in the last quarter of 1992 and want to produce forecasts for 1993:1 until 1993:4. Then we can use 1990:4 for estimation and produce forecasts for 1991:1 until 1992:4 and compute the $RMSE-h$ for each indicator (at this stage we still use future values of the indicators to produce

the forecasts since they are in the available information set). The indicator that provides the lowest RMSE- h for 1991:1 until 1992:4 is then used to forecast from 1993:1 until 1993:4, where the estimation sample is extended until the last available observation, *i.e.*, 1992:4. Moreover, since values of the indicator variable over the period 1993:1 until 1993:4 are not known in 1992:4, autoregressive models are used to forecast them.

The procedure above is repeated for each year, starting in 1990:4 in order to have enough observations in the first evaluation period. This framework can be extended to consider the forecasting performance of groups of indicators or factors or in combination where the make-up of the group at time t is determined by a selection criterion based on forecasting performance until time $t-1$. The ex-ante or real-time forecasting methods we consider in period t are:

- (a) AR;
- (b) best single Euro indicator in period $t-1$ (with *PcGets* model selection for lag length);
- (c) best 5 Euro indicators in period $t-1$ (with *PcGets* model selection);
- (d) best 5 Euro indicators + best 2 US indicators in period $t-1$ (with *PcGets* model selection);
- (e) best 3 Euro factors in period $t-1$ (with *PcGets* model selection);
- (f) best 5 Euro indicators + best 2 US factors in period $t-1$ (with *PcGets* model selection);
- (g) mean of the forecasts from each of the 7 single indicators in (d);
- (h) median of the forecasts from each of the 7 single indicators in (d).

For methods (a) to (f), the exact specification of the forecasting model (5) is determined by *PcGets*. The variables to be included in the groups (methods b to f) at t are determined on the basis of their forecasting performance as indicators in period $t-1$. The last two methods (g and h) consider the impact of forecast pooling.

The results are reported in Table 6. The best single feasible Euro indicator (column labelled (b)) out-performs the AR in 7 out of the 10 evaluation periods considered, and on each occasion the reduction in the RMSE- h is in excess of 10%. It provides the best indicator in 5 out of these 8 periods. Ex-ante, the best performing groups overall are given by (f) (the best 5 Euro indicators combined with the best 2 US factors at $t-1$) followed by (c) (composed of the best 5 Euro indicators at $t-1$). The group given by (f) out-performs the AR in 7 of the 10 evaluation periods (with above 10% gains in 4 out of these 7 periods) and provides the lowest RMSE on 2 occasions. Group (c), which is a subset of group (f) does better than the AR in 6 of the 10 evaluation periods (with above 10% gains in 3 out of these 6 periods) and provides

the lowest RMSE on 1 occasion. The AR provides the best indicator in only 1 out of the 10 evaluation periods.

The good performance of feasible single indicators and of the groups of indicators considered in forecasting inflation is an encouraging finding. In particular, the result that the group including US factors does well, confirms a finding in Marcellino *et al.* (2003) which considered only Euro-area factors. It is however a finding that is at odds with the outcome for the US where the factors perform better for real variables (see Stock and Watson, 2002a, 2002b and Banerjee and Marcellino, 2005). Also worth noting is that single-indicator forecasts are beaten by the groups of indicators (or indicators and factors) 5 out of 10 times.

The last two columns of Table 6 report the results on the forecasting behaviour of pooled forecasts, where the pooling occurs over the best five single EU indicators *ex-ante*. The median outperforms the mean forecasts, in 7 out of the 10 evaluation periods possibly due to some high RMSE-*h* indicator-based forecasts. The mean and median do better than the AR in 1 and 2 out of 10 periods respectively, and the median provides the best forecast in only one period. Thus, in marked contrast with the encouraging performance of feasible single indicators and groups of indicators, the performance of forecast pooling within this context is rather disappointing. This is an interesting result since it contrasts with what is typically found when using the average of the mean square forecast error over a long number of periods as the evaluation criterion, where the latter smoothes out large errors in some periods. These are captured here instead by our RMSE-*h* criterion.

4.3 Sensitivity Analysis

In this subsection we summarise the results of a set of sensitivity analyses we performed to evaluate the robustness of the previous findings. Detailed tables are available upon request.

First, since the dynamic properties, particularly the persistence of the inflation series (whether $I(0)$ or $I(1)$), are open to doubt, we considered forecasting the first difference of inflation (with the right-hand-side or leading indicator variables for nominal or price variables correspondingly differenced twice). The good performance of the autoregressive method is even more evident from the results when compared to the case where inflation is treated as $I(0)$.

Second, we evaluated the role of the estimation method, using a rolling window of 10 years rather than a recursively extended sample (see *e.g.* Pesaran and Timmermann, 1995). Overall, recursive estimation appears to be preferable.

Finally, as mentioned, we considered a longer forecast horizon, focusing on up to two year ahead forecasts, *i.e.* $h=8$ rather than $h=4$, and also experimented with different weights for the elements of the loss function in order to give greater importance to the short horizons. When compared with the $h=4$ analysis with fixed weights in the loss function, the best indicator selected in each evaluation period can be different, but the list of the best-performing indicators is basically unaltered, which makes our conclusion fairly robust to these modifications.

5. Forecasting GDP growth

We now evaluate the ability of the 46 indicators under analysis for forecasting GDP growth, paralleling the analysis of inflation.

5.1 Euro-area indicators, US variables and factors: Ex-post analysis

First, looking at the second column of Table 7, the autoregression (AR) yields a lower RMSE - h than at least 50% of the models with an indicator in 13 out of the 16 periods. Second, among the indicators which out-perform the AR, only a subset yields gains of at least 10% in terms of reduction of the loss function (compare the third column of Table 1 with the second column). Thus, in no evaluation period do more than 50% of the models with single indicators perform at least 10% better than the AR. Third, in common with the results for inflation, the performance of the AR model is rather unstable over time (consider the fourth column of Table 7). While the best performing indicator is always better than a pure AR, (compare the fifth column of Table 7 with the fourth), the best indicator changes over time. The loss involved is also rather volatile over time. Finally, in contrast with our results for inflation, more indicators are deleted by *PcGets* because of their statistical insignificance in the forecasting regressions.

To turn to the more detailed and disaggregate information on the forecasting performance of the single indicators we need to consider Tables 8 and 9. Panel (a) of Table 9 shows that 12 out of the 46 indicators do better than the autoregression in more than 7 out of the 16 evaluation periods, but in contrast with our results for inflation, there are no entries in the column labelled 11-16. When the evaluation criterion is tightened further, it may be seen from Panel (b) of Table 8 that only 1 of the 12 variables in the 7 – 10 category remains, with the remaining 11 now being placed in the third category (*i.e.* out-performing the AR by at

least 10% between 3 to 6 times out of 16 evaluation periods). Overall these results are reflective of a poorer performance of single indicators in forecasting inflation, when compared with the corresponding results for inflation, conditional upon the forecasting exercise being undertaken ex-post.

Tables 8 and 9 together provide information on the best-performing single indicators for GDP growth. These include indices of world GDP and world demand, measures of short and long-term interest rates and the interest rate spread, labour market variables such as the unemployment rate and the ratio of employment to the total labour force, and government consumption variables. Among the worst performing indicators appear to be measures of real and nominal exchange rates and (in common with inflation) the consumer sentiment index for the Euro area.

Table 10 repeats the evaluation exercise undertaken in Table 7 for data taken from the US variables data set. In this case the 'best' leading indicator may be seen to be a measure of weekly hours worked in the US which outperforms the AR in 10 out of 16 cases (better than 10% on 3 occasions) but does not provide the best forecast in any period. The unemployment rate variable is ranked next outperforming the AR in 8 out of 16 periods (better than 10% on 3 occasions) but not providing the best forecast for any of these periods. The other good indicators are a measure of capital utilisation (in concordance with the corresponding result for inflation) and a consumer sentiment indicator for the US (which is in stark contrast with the poor performance of the corresponding measure for the Euro area).

The forecasting performance of US and Euro-area factors is summarized in Table 11. Three US factors (US-F1, US-F4 and US-F6) do better than the AR in 8 periods out of 16 periods, and perform at least 10% better than the AR in 3, 2 and 1 period(s) respectively. US-F2 is ranked next, doing better than the AR on 7 occasions, 2 times at better than 10%. The EU factors perform relatively poorly in comparison. EU-F3, EU-F5 and EU-F6 are ranked third, on par roughly with US-F3 and US-F5. This is a finding in line with Marcellino *et al.* (2003) who show that factors work better for nominal than for real variables, in contrast with the US (see also Stock and Watson, 2002a, 2002b, Banerjee and Marcellino, 2005).

Within the framework of this ex-post analysis we may conclude that the role of single indicators in forecasting GDP growth is markedly diminished in comparison with the results for inflation reported previously. US factors are efficacious but not to the same extent as for forecasting inflation and Euro-area factors display poor performance. It is possible to note, for example, that EU-F2 produces the worst forecast in 3 out of 16 periods.

5.2 Factor based forecasts, groups of indicators and pooling: Ex-ante analysis

Table 12 reports results for the ex-ante analysis, where the methodology and the forecasting methods under analysis are as for inflation. The best single feasible Euro indicator (column labelled (b)) now out-performs the AR in only 4 of the 10 evaluation periods (in contrast with inflation where the corresponding number is 7 out of 10 periods) and the reduction in the RMSE- h is in excess of 10% in 2 of these periods. It provides the best indicator in 3 out of these 4 periods. Ex-ante the best performing group overall is given by (e) (the best 3 Euro indicators combined at $t-1$). This group out-performs the AR in 6 of the 10 evaluation periods (with above 10% gains in 4 out of these 6 periods) and provides the best indicator on 2 occasions. Groups (c) (best 5 EU indicators) is ranked next, out-performing the AR in 5 periods, while groups (d) and (f) perform roughly equivalently, each doing better than the AR in 4 of the 10 evaluation periods. The AR does not provide the best indicator in any of the 10 evaluation periods.

Columns (g) and (h) of Table 12 provide the report the results on the forecasting behaviour of pooled forecasts, where the pooling occurs over the best five single EU indicators ex-ante. Two main comments can be made. First, as for inflation, the median always outperforms the average forecasts. Second, the median forecasts are at least as good as the autoregression in only 2 out of 16 evaluation periods for $h=4$. In comparison with the results reported in Section 4.2 for inflation, the even weaker performance of the pooled forecasts for GDP growth may be justified by the poorer than average performance of the single indicators for forecasting GDP growth.

5.3 Sensitivity Analysis

The relative ranking of rolling and recursive estimation is less clear cut than in the case of inflation, but in most cases the gains from the use of a rolling window are minor.

As for inflation, the performance of the methods for $h=8$ or with different weightings of the components of the loss function is rather similar to what we reported. In particular, the list of the overall best performing indicators is not affected. The detailed results are available upon request.

6. Conclusions

This paper presents a thorough analysis of leading indicators for Euro-area inflation and GDP growth. We consider many single (European and US) indicators, factors extracted from the set of indicators, groups of indicators or factors (with the final specification determined by an automated model selection procedure), and pooled forecasts. The comparison of the forecasting models is conducted with respect to an autoregressive model, both *ex post* and in a pseudo-real-time context, based on their performance for up to one-year and up to two-year forecasts. We argue that such a loss function, the RMSE- h criterion, is particularly relevant in a policy-making context.

A major observation that emerges is that both for inflation and GDP growth, *ex-post*, autoregressions are systematically beaten by univariate leading indicator models, but the best indicator changes over time. This reflects the fact that in forecasting practice, although some groups of indicators perform robustly, it is advisable to update continuously the choice of indicators or groups of indicators. Moreover, the dynamics of the variable forecast will be driven by different shocks at different points of time, so that different indicators will assume different relevance over different time periods. Thus, even in the best-case scenario – *i.e.* under *ex-post* evaluation – autoregressions (sufficiently differenced for stationarity), which are simple models, are hard to beat consistently with a single forecasting model and overall provide a robust forecasting tool.

Among the particular points about forecasting Euro-area inflation, seven are worth mentioning. First, some labour market variables, prices, fiscal series and the GDP growth rate on average outperform the autoregression. Second, some US indicators are also useful, in particular the growth in the real exchange rate, the inflation rate, the capacity utilization rate, the growth of M2, a 3-month interest rate and the growth in hourly earnings. Factors are also efficacious, especially some of the US factors. Third, in a pseudo *ex-ante* context – the closest to real-time forecasting – the indicators can beat the autoregressions quite often up to four quarters ahead. Fourth, an additional encouraging finding for forecasting inflation in the *ex-ante* framework is that grouping either the best performing single indicators or the US or Euro-area factors, complemented by the automatic model selection procedure implemented with *PcGets*, is often better than the autoregression. Fifth, forecast pooling produces rather poor results. Sixth, it is also confirmed within the pseudo *ex-ante* framework that for a one-year forecasting horizon, continuous updating of leading indicators is of major importance. Finally, recursive estimation appears to be better than rolling estimation, and, as expected, the

indicators appear to perform better for a one-year ($h=4$) than for a two-year ($h=8$) horizon, with the deterioration evident in both the ex-ante and ex-post contexts. The $h=8$ results are available upon request.

Six similar comments can be made for Euro-area GDP growth, in addition to the general observation that the indicators used in this paper on average appear to perform better for inflation. First, ex post, univariate leading indicator models can always beat the autoregression, but the best indicator changes over time. Second, the best indicators on average are world GDP and demand growth, interest rates, public expenditure and labour market variables. Third, the set of good US indicators includes labour market variables such as hours worked and unemployment, capacity utilization rate and the consumer confidence indicator, while rather surprisingly US GDP and industrial production growth are outperformed by the autoregression in most cases. Fourth, in line with the generally weaker performance of indicators in forecasting GDP growth relative to inflation ex post, the results for forecasting in real-time framework indicate a weaker performance of both single indicators, groups of indicators and pooled forecasts. It is only for a group of indicators (consisting of the three best individual indicators) that we can say that it on average beats a simple AR model in a pseudo ex-ante context. Fifth, sensitivity results for a longer forecast horizon or different weighting of the components of the loss function indicate a high degree of robustness of our findings. Finally, there is no clear ranking of recursive and rolling estimation as was the case for inflation.

Further work in this area would include evaluating the use of leading indicators under alternative specifications for the loss function. With a more standard loss function we could, for example, undertake density forecast comparisons or use Diebold-Mariano- type statistics. We could also pursue the potential benefits from looking at the possible use of more detailed survey data (on consumer or business confidence, for example) or more disaggregate data (such as the components of a price index) in constructing forecasts. Most importantly, we have considered the use of these methods to data from accession countries, where questions of inflation and GDP growth are of paramount importance in the years ahead and pose a considerable challenge to economists and forecasters alike.

References

- Artis, M., Banerjee, A. and Marcellino, M. (2005), "Factor Forecasts for the UK", *Journal of Forecasting*, 24, 279-298.
- Bai, J. and Ng, S. (2003), "Confidence Intervals for Diffusion Index Forecasts with a Large Number of Predictors", *mimeo*, University of Michigan.
- Banerjee, A. and Marcellino, M. (2005), "Are there any reliable leading indicators for US inflation and GDP growth?", *International Journal of Forecasting* (forthcoming).
- Banerjee, A., Marcellino, M. and Masten, I. (2005), "Forecasting Macroeconomic Variables for the Acceding Countries", ECB Working paper No. 482.
- Bates, J.M. and Granger C.W.J. (1969), "The Combination of Forecasts." *Operations Research Quarterly*, 20, 415-468.
- Beyer, A., Doornik, J.A., and Hendry, D.F. (2001), "Constructing historical Euro-zone data", *Economic Journal*, 111, 102-21.
- Camba-Mendez, G., Kapetanios, G., Smith, R.J. and Weale, M.R. (2001), "An Automatic Leading Indicator of Economic Activity: Forecasting GDP Growth for European Countries", *Econometrics Journal*, 4, 56-90.
- Cecchetti, S. G., Chu, R.S. and Steindel, C. (2000), "The Unreliability of Inflation Indicators", Current Issues in Economics and Finance 4/6, Federal Reserve Bank of New York.
- Clements, M.P. and Hendry, D.F. (1999), *Forecasting Non-stationary Economic Time Series* Cambridge, Massachusetts: MIT Press.
- Clements, M.P. and Hendry, D.F. (2004), "Pooling of forecasts", *Econometrics Journal*, 7, 1-31.
- Dreger, C. and Marcellino, M. (2003), "A macroeconomic model for the euro economy", *mimeo*, Bocconi University.
- Fagan, G., Henry, J. and Mestre, R. (2001), "An area-wide model for the Euro area", ECB Working Paper No. 42, European Central Bank.

- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000), "The Generalized Dynamic Factor Model: Identification and Estimation" *Review of Economics and Statistics*, 11, 369-379.
- Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series" In *Latent Variables in Socio-Economic Models*, edited by Denis J. Aigner and Arthur S. Goldberger, Ch. 19. Amsterdam: North-Holland.
- Hendry, D.F. and Krolzig, H-M. (1999), "Improving on `Data mining reconsidered by K.D. Hoover and S.J. Perez", *Econometrics Journal*, 2, 202-219.
- Hoover, K. and Perez, S.J. (1999), "Data mining reconsidered: encompassing and the general-to-specific approach to specification search", *Econometrics Journal*, 2, 167-191.
- Krolzig, H-M. and Hendry, D.F. (2001), "Computer automation of general-to-specific model selection procedures", *Journal of Economic Dynamics and Control*, 25, 831-866.
- Marcellino, M. (2004a), "Forecasting EMU macroeconomic variables", *International Journal of Forecasting*, 20, 359-72.
- Marcellino, M. (2004b), "Forecast pooling for short time series of macroeconomic variables", *Oxford Bulletin of Economics and Statistics*, 66, 91-112.
- Marcellino, M. (2004c), "Leading indicators: What have we learned?" mimeo, Bocconi University (in preparation for Elliot, G., Granger, C.W.J. and A. Timmermann (eds), *Handbook of Economic Forecasting*, Elsevier-North Holland.)
- Marcellino, M., Stock, J.H. and Watson, M.W. (2003), "Macroeconomic Forecasting in the Euro Area: Country Specific Versus Euro Wide Information", *European Economic Review*, 47, 1-18.
- Marcellino, M., Stock, J.H. and Watson, M.W. (2005), "A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Series h-Steps Ahead", *Journal of Econometrics* (forthcoming).
- Mizon, G.E. and Richard, J-F. (1986), "The encompassing principle and its application to non-nested hypothesis tests", *Econometrica*, 54, 657-678.
- Pesaran, H. and Timmermann, A. (1995), "Predictability of Stock Returns: Robustness and Economic Significance", *Journal of Finance*, 50, 1201-1228.

Sargent, T.J. and Sims, C.A.. (1997), “Business Cycle Modelling without Pretending to have Too Much a-priori Economic Theory” In *New Methods in Business Cycle Research*, edited by Christopher A. Sims, Minneapolis: Federal Reserve Bank of Minneapolis.

Stock, J.H. and Watson, M.W. (1999), “Forecasting Inflation”, *Journal of Monetary Economics*, 44, 293-335.

Stock, J.H. and Watson, M.W. (2001), “Forecasting Output and Inflation: The Role of Asset Prices”, NBER Working Paper No. 8180.

Stock, J. H. and M. W. Watson (2002a), “Macroeconomic Forecasting Using Diffusion Indexes”, *Journal of Business and Economic Statistics*, 20, 147-62.

Stock, J. H. and Watson, M. W. (2002b), “Forecasting Using Principal Components from a Large Number of Predictors”, *Journal of the American Statistical Association*, 97, 1167-1179.

Stock, J.H. and Watson, M.W. (2003), “Combination forecasts of output growth and the 2001 US recession”, *mimeo*, Harvard University.

List of variables and transformations used

Variable	Transformations used	Description
Output variables		
GDP	DLV	Gross Domestic Product
IndProd	DLV	Industrial production – total, series starts in 1978q1
IndProd_man	DLV	Industrial production – manufacturing, series starts in 1980q1
OutGap	LV	Output gap
TotDemand	DLV	Total demand
Cons_real	DLV	GDP – private consumption at constant prices
Cons_nom	DLV	GDP – private consumption at current prices
DisInc	DLV	Household's disposable income
GovC	DLV	GDP – government consumption at constant prices
GovC_nom	DLV	GDP – government consumption at current prices
GovExp	DLV	Government expenditure
Inv_real	DLV	Gross investment, constant prices
Inv_nom	DLV	Gross investment, current prices
GDPworld	DLV	World GDP
WorldDemand	DLV	World Demand Composite Indicator
Employment and productivity		
Empl	DLV	Total Employment
Empl/L	LV	Ratio Total Employment/Labour Force
LabProd	LV, DLV	Labour Productivity
UnemRate	LV	Unemployment Rate
TFP	DLV	Trend Total Factor Productivity
Exchange rates		
RealXR	LV, DLV	Real effective exchange rate
NomXR	LV, DLV	Nominal effective exchange rate
Interest rates and spreads		
Long_Intrate	LV	Long-term interest rate (% p.a.)
IntRate_Short	LV	Short-term interest rate (% p.a.)
IntSpread	LV	Long_Intrate – IntRate_Short
Monetary aggregates		
M1	DLV	Monetary aggregate M1, series starts in 1980q1
M3	DLV	Monetary aggregate M3, series starts in 1980q1
Price indexes		
HICP	DLV	Harmonized Index of Consumer Prices (1996=100)
ConsDefl	DLV	Private consumption deflator
ProdPr	DLV	Producer prices – total industry, series starts in 1980q1
ProdPr_man	DLV	Producer prices – manufacturing, series starts in 1985q1
ComPrice	DLV	Commodity Prices
Wages		
LabInc	DLV	Compensation to employees
Wage	DLV	Wage rate
UnitLabCost	DLV	Unit labor costs

Miscellaneous

Debt/GDP	LV	Ratio Public Debt/GDP
GovExp/GDP	LV	Ratio Government Expenditure/GDP
Surpl/GDP	LV	Ratio Government Primary Surplus/GDP
GovRev/GDP	LV	Ratio Government Revenue/GDP
CurrAcc	LV	Current account balance, current prices
TrBal	LV	Trade balance
Imp	DLV	Imports of Goods and Services, constant prices
Exp	DLV	Exports of Goods and Services, constant prices
Confind	LV	Industrial confidence indicator, series starts in 1985q1
Ecsent	LV	Economic sentiment indicator, series starts in 1985q1

US variables

GDP	DLV	GDP – Total (BN \$, 1996 prices, S.A.)
IndProd	DLV	Industrial production – total (1995=100, S.A.)
CapUtil	LV	Capacity utilization rate (% , S.A.)
WorkHours	LV	Weekly hours worked – manufacturing (hours, S.A.)
UnemRate	LV	Unemployment rate (% of civilian labor force, S.A.)
SharePrice	DLV	NYSE share prices (1995=100)
IntRate-10ygov	LV	Government composite bonds (>10 years, % p.a.)
IntRate-cod	LV	Certificates of deposits (3 month, % p.a.)
Spread-3m	LV	fcod – Federal Funds rate
Spread-10y	LV	fy10gov – Federal Funds rate
RealXR	DLV	US real effective exchange rate (1995=100)
M2	DLV	Monetary aggregate M2 (BN\$, S.A.)
Infl	LV	Growth rate of CPI index (1995=100, S.A.)
Wage	DLV	Hourly earnings – total private (1995=100, S.A.)
UnitLabC	DLV	Unit labor cost – manufacturing (1995=100, S.A.)
Conf	LV	Consumer sentiment (1995=100, S.A.)

Transformations used: LV – levels, DLV – annual growth rate. Variables transformed into growth rates have the ending “g” in the tables below.

All data for Euro Area have been seasonally adjusted at source (Eurostat) or using the SABL method (Fagan *et al.*, 2001). The base year for all series is 1990 if not indicated otherwise. See Fagan *et al.*, 2001 for details on data sources and data construction.

US data have been collected from OECD Main Economic Indicators database. S.A. indicates that the data have been seasonally adjusted at source.

INFLATION INDICATORS FOR EURO AREA

Table 1: Performance of indicators in forecasting Euro-area Inflation (up to four quarters ahead)

Estimation period	No. of Indicators That Relative to AR Performed		RMSE- <i>h</i>		
	Better	At least 10% Better	AR	Best Indicator	Worst Indicator
75:1 84:4	32	29	4.31	1.81 Debt/GDP	4.76 LabProd
75:1 85:4	8	8	1.42	0.60 ComPriceg	5.10 Empl/L
75:1 86:4	22	17	1.65	0.46 TFPg	6.40 IntSpread
75:1 87:4	16	14	0.80	0.25 Empl/L	3.51 NomXR
75:1 88:4	19	9	1.36	0.88 IntSpread	7.78 Surpl/GDP
75:1 89:4	22	18	1.24	0.82 NomXR	3.31 TFPg
75:1 90:4	14	9	0.85	0.38 TrBal	2.55 CurrAcc
75:1 91:4	11	10	1.68	0.51 Debt/GDP	3.69 M1g
75:1 92:4	14	4	2.06	1.55 GovC_nomg	6.86 Impg
75:1 93:4	9	3	0.99	0.54 GovC_nomg	5.10 Debt/GDP
75:1 94:4	15	4	1.11	0.59 M1g	2.29 Debt/GDP
75:1 95:4	29	14	1.26	0.96 IntSpread	1.94 ProdPr_mang
75:1 96:4	19	18	1.15	0.57 DisIncg	3.41 Empl/L
75:1 97:4	4	0	0.76	0.69 ProdPr_mang	1.91 IndProd_mang
75:1 98:4	35	26	1.52	0.72 GovExp/GDP	49.66 M1g
75:1 99:4	30	26	1.41	0.67 OutGap	2.08 UnitLabCostg

Note: Variable ending “g” indicates growth rate of original variable.

Table 2: Ranking the Euro-area Inflation Indicators (up to four quarters ahead)

Number of Times the Indicator Outperforms the Autoregression (out of 16)			
0-2	3-6	7-10	11-16
IndProd_mang***, Confind*, ProdPrg**	NomXR, RealXR, RealXRg, TotDemandg, Debt/GDP, LabProd, Long_Intrate, IntSpread, IntRate_Short, UnitLabCostg, M3g, GovRev/GDP, IndProdg***, Ecsent*, ProdPr_mang*, TrBal, Impg, GDPworldg, M1g	ComPriceg, NomXRg, GovC_nomg, GovCg, GovExp, GovExp/GDP, Empl/L, LabProdg, Cons_nomg, Cons_realg, UnemRate, OutGap, CurrAcc, Surpl/GDP, Inv_realg, DisIncg, Exp, WorldDemandg, GDPg, TFPg	ConsDeflg, Wageg, Inv_realg, LabIncg
Number of Times the Indicator Outperforms the Autoregression by at least 10% (out of 16)			
0-2	3-6	7-10	11-16
RealXR, LabProd, M3g, IndProdg***, IndProd_mang***, Confind*, ProdPr_mang*, ProdPrg**, M1g	ComPriceg, NomXR, NomXRg, RealXRg TotDemandg, GovCg, Debt/GDP, GovExp, GovExp/GDP, Empl/L, LabProdg, Long_Intrate, IntSpread, Cons_Nomg, IntRate_Short, UnitLabCostg, OutGap, CurrAcc, Surpl/GDP, GovRev/GDP, Ecsent*, TrBal, Impg, Exp, GDPworldg, WorldDemandg, TFPg	GovC_nomg, ConsDeflg, Cons_realg, UnemRate, Wageg, Inv_realg, DisIncg, LabIncg, GDPg	

Notes: Variable ending “g” indicates growth rate of original variable. Not all variables in the data set start in 1975. For this reason it is impossible to test the forecasting performance of all indicators for all 16 periods.

Stars denote the following: * – out of 5 periods, ** – out of 10 periods, *** – out of 12 periods.

Table 3: Top five Euro-area Inflation Indicators (up to four quarters ahead)

75:1 84:4	Debt/GDP, NomXR, ComPriceg, GovC_nomg, TrBal
75:1 85:4	ComPriceg, ConsDeflg, NomXRg, TFPg, Cons_realg
75:1 86:4	TFPg, NomXRg, GovCg, ConsDeflg, RealXRg
75:1 87:4	Empl/L, UnemRate, Inv_realg, GovExp, GovRev/GDP
75:1 88:4	IntSpread, Impg, Inv_realg, Inv_realg, GDPg
75:1 89:4	NomXR, Impg, Cons_realg, TotDemandg, Empl/L
75:1 90:4	TrBal, Cons_realg, OutGap, Surpl/GDP, Inv_realg
75:1 91:4	Debt/GDP, Cons_nomg, ConsDeflg, GDPg, RealXRg
75:1 92:4	GovC_nomg, Wageg, LabIncg, M3g, UnitLabCostg
75:1 93:4	GovC_nomg, ConsDeflg, GovExp/GDP, Cons_realg, Wageg
75:1 94:4	M1g, ComPriceg, ConsDeflg, OutGap, Inv_realg
75:1 95:4	IntSpread, GovCg, GovC_nomg, Cons_nomg, Impg
75:1 96:4	DisIncg, CurrAcc, GovRev/GDP, TFPg, WorldDemandg
75:1 97:4	ProdPr_mang, Surpl/GDP, Exp, ProdPrg, M3g
75:1 98:4	GovExp/GDP, Wageg, LabIncg, CurrAcc, UnemRate
75:1 99:4	OutGap, GDPg, UnemRate, Empl/L, Long_Intrate

Table 4: Forecasting performance of selected US variables for Euro-area Inflation (up to four quarters ahead)

	Number of Times the Indicator			
	Outperforms Autoregression	Outperforms AR by at least 10%	Produces Best Forecast	Produces Worst Forecast
GDPg	2	2	-	-
IndProdg	4	3	-	-
CapUtil	9	6	-	1
WorkHours	4	4	-	-
UnemRate	7	6	-	-
SharePriceg	7	5	-	-
IntRate-10ygov	7	3	-	-
IntRate-cod	8	7	-	-
Spread-10y	5	3	-	1
Spread-3m	6	2	-	-
RealXR	4	3	1	1
RealXRg	10	7	3	-
M2g	7	3	-	-
Infl	10	6	-	-
Wageg	9	5	-	1
UnitLabCg	5	3	1	-
Conf	7	6	-	-

Note: Variable ending “g” indicates growth rate of original variable.

Table 5: Forecasting performance of US and Euro-area factors for Euro-area Inflation (up to four quarters ahead)

	Number of Times the Indicator			
	Outperforms Autoregression	Outperforms AR by at least 10%	Produces Best Forecast	Produces Worst Forecast
US-F1	4	2	-	-
US-F2	2	2	-	-
US-F3	8	5	-	1
US-F4	10	8	-	-
US-F5	4	2	-	-
US-F6	7	6	-	-
EU-F1	6	5	2	-
EU-F2	7	7	-	-
EU-F3	5	2	1	-
EU-F4	7	4	1	-
EU-F5	4	3	-	-
EU-F6	9	7	-	1

Table 6: Performance of feasible indicators (ex-ante) for forecasting Euro-area Inflation (up to four quarters ahead)

Point in time	AR	Best indicator	Best 5 EU indicators	Best 5 EU + best 2 US indicators	Best 3 EU factors	Best 5 EU indicators + best 2 US factors	Forecast pooling of each single indicators from column 5	
							Mean	Median
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
90:4	0.85	2.93	1.36	1.54	0.90	1.99	0.91	0.77
91:4	1.68	0.48	1.46	2.00	2.05	0.91	2.25	2.36
92:4	2.06	2.17	2.02	2.54	2.34	2.02	2.74	2.38
93:4	0.99	0.69	1.60	1.29	1.13	1.25	1.40	1.25
94:4	1.11	0.70	1.19	1.99	1.23	0.89	1.19	1.21
95:4	1.26	0.81	1.15	1.05	1.20	1.19	1.33	1.34
96:4	1.15	0.91	0.56	1.01	0.47	0.88	0.92	0.88
97:4	0.76	1.77	1.12	1.17	0.85	0.97	1.13	0.92
98:4	1.52	1.11	0.97	1.65	1.48	0.91	1.73	1.59
99:4	1.41	0.72	1.38	1.38	0.88	1.38	1.23	1.00

Note: BOLD indicates that the corresponding RMSE- h is smaller than the RMSE- h of the pure AR model.

GDP GROWTH INDICATORS FOR EURO AREA

Table 7: Performance of indicators in forecasting Euro-area GDP growth (up to four quarters ahead)

Estimation period	No. of Indicators That Relative to AR Performed		RMSE- <i>h</i>		
	Better	At least 10% better	AR	Best Indicator	Worst Indicator
75:1 84:4	6	4	1.46	1.19 Long_Intrate	17.43 OutGap
75:1 85:4	11	1	3.40	3.05 GovExp/GDP	4.68 OutGap
75:1 86:4	3	1	4.05	3.53 GovCg	7.65 OutGap
75:1 87:4	23	19	2.01	0.92 IntRate_Short	30.11 OutGap
75:1 88:4	26	10	2.26	1.50 LabProd	6.28 TotDemandg
75:1 89:4	28	22	1.91	1.09 Impg	5.65 OutGap
75:1 90:4	12	7	0.90	0.47 IntSpread	4.46 M1g
75:1 91:4	22	10	3.17	1.75 TFPg	5.64 TotDemandg
75:1 92:4	14	11	3.32	2.00 TFPg	5.68 Inv_realg
75:1 93:4	13	12	1.15	0.37 IndProd_mang	4.47 LabProd
75:1 94:4	10	4	1.15	0.35 OutGap	4.03 IntSpread
75:1 95:4	10	5	0.73	0.56 Emplg	2.79 IntRate_Short
75:1 96:4	12	5	1.73	1.38 OutGap	2.99 CurrAcc
75:1 97:4	11	1	1.17	1.00 TFPg	3.72 LabProd
75:1 98:4	26	17	1.55	0.72 Long_Intrate	508.95 TFPg
75:1 99:4	8	2	0.78	0.57 Inv_realg	2.05 IndProd_mang

Note: Variable ending “g” indicates growth rate of original variable.

Table 8: Ranking the Euro-area GDP growth indicators (up to four quarters ahead)

Number of Times the Indicator Outperforms the Autoregression (out of 16)			
0-2	3-6	7-10	11-16
Wageg, Ecsent*, Confind*, ProdPr_mang*	NomXR, NomXRg, RealXR, RealXRg, TotDemandg, GovC_nomg, GovCg, GovExp, Surpl/GDP, Empl/L, LabProd, LabProdg, Long_Intrate, IntSpread, Cons_nomg, Cons_realg, Impg, Expg, TrBal, UnitLabCostg, OutGap, M3g, GovRev/GDP, Inv_realg, Inv_realg, IndProdg***, IndProd_mang***, ProdPrg**, LabIncg, M1g	Infl, CurrAcc, Debt/GDP, GovExp/GDP, Emplg, IntRate_Short, TFPg, UnemRate, GDPworldg, WorldDemandg, DisIncg, ComPriceg, IndProdg	
Number of Times the Indicator Outperforms the Autoregression by at least 10% (out of 16)			
0-2	3-6	7-10	11-16
NomXR, NomXRg, RealXRg, TotDemandg, GovC_nomg, GovCg, GovExp, Surpl/GDP, LabProd, Cons_nomg, Expg, TrBal, UnitLabCostg, Wageg, M3g, Inv_realg, DisIncg, Ecsent*, Confind*, ProdPr_mang*, LabIncg	Infl, CurrAcc, RealXR, Debt/GDP, GovExp/GDP, Emplg, Empl/L, LabProdg, Long_Intrate, IntSpread, Cons_realg, IntRate_Short, Impg, TFPg, UnemRate, OutGap, GDPworldg, GovRev/GDP, Inv_realg, ComPriceg, IndProdg***, IndProd_mang***, ProdPrg**, M1g	WorldDemandg	

Notes: Variable ending “g” indicates growth rate of original variable. Not all variables in the data set start in 1975. For this reason it is impossible to test the forecasting performance of all indicators for all 16 subperiods.

Stars denote the following: * – out of 5 subperiods, ** – out of 10 subperiods, *** – out of 12 subperiods.

Table 9: Top five Euro-area GDP growth indicators (up to four quarters ahead)

75:1 84:4	Long_Intrate, Empl/L, UnemRate, WorldDemandg, IntRate_Short
75:1 85:4	GovExp/GDP, Empl/L, IntRate_Short, HICPg, NomXRg
75:1 86:4	GovCg, DisIncg, GDPworldg, Surpl/GDP, Impg
75:1 87:4	IntRate_Short, Empl/L, UnemRate, GovRev/GDP, HICPg
75:1 88:4	LabProd, DisIncg, Inv_realg, Debt/GDP, CurrAcc
75:1 89:4	Impg, WorldDemandg, Inv_realg, LabProdg, GovC_nomg
75:1 90:4	IntSpread, IndProdg, ComPriceg, Cons_realg, NomXR
75:1 91:4	TFPg, UnemRate, M1g, IntSpread, Empl/L
75:1 92:4	TFPg, IndProdg, IntSpread, IndProd_mang, Emplg
75:1 93:4	IndProd_mang, Emplg, Impg, OutGap, IndProdg
75:1 94:4	OutGap, TFPg, GovExp/GDP, Cons_realg, Long_Intrate
75:1 95:4	Emplg, GovRev/GDP, ProdPrg, DisIncg, WorldDemandg
75:1 96:4	OutGap, HICPg, Cons_nomg, IndProd_mang, IndProdg
75:1 97:4	TFPg, Surpl/GDP, NomXRg, RealXR, RealXRg
75:1 98:4	Long_Intrate, IntRate_Short, M1g, Debt/GDP, LabProdg
75:1 99:4	Inv_realg, ProdPrg, GovRev/GDP, Long_Intrate, TotDemandg

Table 10: Forecasting performance of selected US variables for Euro-area GDP growth (up to four quarters ahead)

	Number of Times the Indicator			
	Outperforms Autoregression	Outperforms AR by at least 10%	Produces Best Forecast	Produces Worst Forecast
GDPg	4	1	-	-
IndProdg	6	1	-	-
CapUtil	8	1	-	-
WorkHours	10	3	-	-
UnemRate	8	3	-	-
SharePriceg	5	3	-	-
IntRate-10ygov	4	1	-	-
IntRate-cod	5	2	1	-
Spread-10y	2	1	-	-
Spread-3m	2	2	-	-
RealXR	4	1	-	-
RealXRg	3	1	-	-
M2g	4	1	-	-
Infl	6	4	-	-
Wageg	2	1	-	-
UnitLabCg	5	3	-	-
Conf	7	3	-	-

Note: Variable ending “g” indicates growth rate of original variable.

Table 11: Forecasting performance of US and Euro-area factors for Euro-area GDP growth (up to four quarters ahead)

	Number of Times the Indicator			
	Outperforms Autoregression	Outperforms AR by at least 10%	Produces Best Forecast	Produces Worst Forecast
US-F1	8	3	-	-
US-F2	7	2	-	-
US-F3	6	2	-	-
US-F4	8	2	-	-
US-F5	6	1	-	-
US-F6	8	1	-	-
EU-F1	4	2	-	-
EU-F2	4	3	-	3
EU-F3	6	2	1	-
EU-F4	5	2	1	-
EU-F5	6	3	-	-
EU-F6	6	1	-	-

Table 12: Performance of feasible indicators (ex-ante) for forecasting Euro-area GDP growth (up to four quarters ahead)

Point in time	AR	Best indicator	Best 5 EU indicators	Best 5 EU + best 2 US indicators	Best 3 EU factors	Best 5 EU indicators + best 2 US factors	Forecast pooling of each single indicators from column 5	
							Mean	Median
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
90:4	0.90	2.57	0.90	1.27	0.68	0.96	1.00	0.93
91:4	3.17	3.33	2.34	3.64	3.18	2.53	3.14	3.13
92:4	3.31	2.50	3.30	3.26	3.06	3.40	3.66	3.61
93:4	1.15	4.70	1.85	2.48	1.12	2.00	1.72	1.43
94:4	1.15	10.3	2.39	1.77	1.25	1.77	1.19	1.20
95:4	0.73	0.64	1.70	0.73	0.66	1.04	1.12	0.81
96:4	1.73	1.71	1.86	2.00	1.78	1.86	1.83	1.80
97:4	1.17	1.14	1.09	0.93	1.16	0.93	1.11	1.09
98:4	1.55	2.19	1.13	1.00	1.06	1.30	1.69	1.63
99:4	0.78	1.80	0.75	0.75	0.89	0.75	0.90	0.80

Note: BOLD indicates that the corresponding RMSE- h is smaller than the RMSE- h of the pure AR model.