ELSEVIER European Economic Review 47 (2003) 1–18

EUROPEAN ECONOMIC REVIEW

www.elsevier.com/locate/econbase

Macroeconomic forecasting in the Euro area: Country specific versus area-wide information

Massimiliano Marcellino[a,d], James H. Stock[b,e], Mark W. Watson[c,e,*]

[a]*Istituto di Economia Politica, Universita Bocconi and IGIER, Via Salasco 3/5, 20136 Milan, Italy*
[b]*Kennedy School of Government, Harvard University, Cambridge, MA 02138, USA*
[c]*Department of Economics and Woodrow Wilson School, Princeton University, Princeton, NJ 08544, USA*
[d]*CEPR, London, UK*
[e]*NBER, Cambridge, MA, USA*

Received 1 July 2000; accepted 12 February 2002

Abstract

This paper compares several time series methods for short-run forecasting of Euro-wide inflation and real activity using data from 1982 to 1997. Forecasts are constructed from univariate autoregressions, vector autoregressions, single equation models that include Euro-wide and US aggregates, and large-model methods in which forecasts are based on estimates of common dynamic factors. Aggregate Euro-wide forecasts are constructed from models that utilize only aggregate Euro-wide variables and by aggregating country-specific models. The results suggest that forecasts constructed by aggregating the country-specific models are more accurate than forecasts constructed using the aggregate data. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: C32; C53; F47

Keywords: Forecasting; Aggregation; Factor models

1. Introduction

The challenge of forecasting aggregate European economic performance is gaining increasing importance. Euro-area inflation forecasts are needed to effectively implement the European Central Bank's targets for inflation of the Euro. European integration also means that political and business decisions increasingly depend on aggregate European

* Corresponding author. Tel.: +1-609-258-4811; fax: +1-609-258-5533.
E-mail address: mwatson@princeton.edu (M.W. Watson).

PII: S0014-2921(02)00206-4

real economic activity, so it is also of increasing interest to consider the problem of forecasting real activity measures for the Euro area as a whole. Forecasting Euro-area aggregates is largely a new area and there is considerable uncertainty about the best econometric procedures to approach this task.

This paper investigates several time series methods for forecasting four Euro-area wide aggregate variables: real GDP, industrial production, price inflation, and the unemployment rate. We consider two empirical questions arising from this problem. First, is it better to build aggregate Euro-area wide forecasting models for these variables, or are there gains from aggregating country-specific forecasts for the component country variables? Second, are there gains from using information from additional predictors beyond simple univariate time series forecasts, and if so, how large are these gains and how are they best achieved?

These questions are attacked by applying an array of forecasting models to data on the eleven countries originally in the EMU, over the period 1982–1997, at both the monthly and quarterly level (the data set is discussed in detail in Section 2). We focus on short-run forecasts at the one-, two- and four-quarter horizons. Using these data, we consider forecasts constructed at the country level and, alternatively, at the aggregate EU level. Comparable forecasting models are constructed for each country, and the resulting forecasts are aggregated to the Euro-area level. This permits a comparison of models and forecasts based on different information sets and constructed at different levels of aggregation.

Five sets of forecasting models are considered: autoregressions; vector autoregressions (VARs); a model in which the Euro-area aggregate is used at the country level as a predictor for the country-specific variable; one in which the comparable aggregate for the United States is used as a predictor; and a large-model forecasting framework in which forecasts are based on estimates of common dynamic factors. The first four of these methods are conventional.

The fifth approach, forecasting using estimated common factors, is motivated by postulating that there are common sources for co-movements across the Euro area and that these co-movements are useful for forecasting both at the country and at the aggregate level. One way to model these co-movements is as arising from one or more common dynamic factors (Geweke, 1977; Sargent and Sims, 1977). Recent advances in the theory of dynamic factor models have shown that, under suitable technical conditions, it is possible to estimate the dynamic factors consistently in an approximate dynamic factor model when the time series and cross-sectional dimensions are large (Forni et al., 2000; Stock and Watson, 1998). If the data are generated by an approximate dynamic factor model, then factors estimated by principal components can also be used for efficient forecasting under quite general conditions (Stock and Watson, 1998). Recent empirical applications of these and related methods include forecasting US inflation (Stock and Watson, 1999) and constructing coincident and leading indexes for the Euro area (Forni et al., 1999. (Additional references to this growing literature are contained in Stock and Watson, 1998 and Forni et al., 1999, 2000.)

Although the main focus of this paper is on comparing forecasting models, our findings might be of interest to macroeconomists more generally. We have two principal conclusions. First, there typically are gains from forecasting these series at the country level, then pooling the forecasts, relative to forecasting at the aggregate level; the coefficient restrictions that would permit direct modeling of the aggregates are strongly rejected, and the pooled forecasts are more accurate than forecasts constructed using the aggregate series. This suggests that structural macroeconometric modeling of the Euro area is appropriately done at the country-specific level, rather than directly at the aggregate level. For example, this apparent failure of the aggregation restrictions complicates the interpretation of models of Euro-wide inflation dynamics estimated at the aggregate level, cf. Gali et al. (2001).

Second, our simulated out-of-sample forecast experiment provides little evidence that forecasts from multivariate models are more accurate than forecasts from univariate models. Looking across variables and forecast horizons, the most accurate forecasts are produced by pooling country-specific univariate autoregressions, a method called “autoregressive components” by Fair and Shiller (1990). If we restrict attention to multivariate models, the forecasts based on estimated factors appear to be somewhat more accurate than the other methods. There are several possible reasons for the comparative success of the autoregressive components forecasts. One is that our sample covers a period of great economic change in Europe, and consequent instability of the multivariate relations could make simple autoregressions more reliable. Another is that our short sample (short because of data limitations) favors forecasting models with very few parameters. Our findings do not imply that there are not important multivariate relations in these data, just that it is difficult to exploit them profitably for real time forecasting.

The remainder of the paper is organized as follows. The data are discussed in Section 2, and the estimators and forecasting models are described in Section 3. Results are presented in Sections 4 and 5 concludes.

2. The Euro-area data

2.1. Country level data

The data are monthly national macroeconomic series for the 11 Euro-area countries. (Greece, which joined the EMU in 2001, is not included.) The data are taken from the OECD main economic indicators, monthly, for the period 1982:1–1997:8. There are approximately 50 variables for each country. These 50 variables typically include industrial production and sales (disaggregated by main sectors); new orders in the manufacturing sector; employment, unemployment, hours worked and unit labor costs; consumer, producer, and wholesale prices (disaggregated by type of goods); several monetary aggregates (M1, M2, M3), savings and credit to the economy; short-term and long-term interest rates, and a share price index; the effective exchange rate and the exchange rate with the US dollar; several components of the balance of payments; and some miscellaneous additional series.

Not all variables are available for the full span for all countries using the same series definitions and construction methods. An alternative approach would have been to adopt a completely harmonized data set with a balanced panel. However the coverage and span of a balanced panel of harmonized series is small, so instead we will use a richer and significantly longer data set. One cost of using this expanded data set is handling missing observations, an issue returned to below.

The data were preprocessed in four stages prior to use in the forecasting exercise. First, the data were inspected visually for indications of major series redefinitions or other inconsistencies. Discrepancies that could not be reconciled resulted in the series being dropped.

Second, the series were transformed to account for stochastic or deterministic trends. In general, logarithms were taken of all nonnegative series that were not already in rates or percentage units. Generally the same transformation, including degree of differencing, was applied to a group of series, and the series transformation was the same across countries. For example, real quantity variables (industrial production, sales, etc.) were generally transformed to first log differences (growth rates). Two sets of transformations were considered for prices, wages, and interest rates. In the first, prices and wages were transformed to rates of inflation (first differences of logs) and interest rates were left in levels. In the second, prices and wages were transformed to changes in rates of inflation (second differences of logs) and interest rates appear as changes. The first of these transformations will be referred to as the I(1) case for prices, and the second will be referred to as the I(2) case for prices. The I(2) case provided the most accurate forecasts of price inflation and this will serve as our baseline model. Results for the I(1) case will be summarized as a robustness check on our main results.

Third, the series were seasonally adjusted. Some of the OECD series are seasonally adjusted and some are not. The seasonal adjustment is sporadic and some of the series that are reported as seasonally adjusted in fact have pronounced and statistically significant seasonal components. We therefore passed all the series through a two-step seasonal adjustment procedure, whether or not they were reported as being seasonally adjusted. First, the transformed monthly series (as discussed in the preceding paragraph) was regressed against 11 monthly indicator variables. Second, if the HAC F-test on these 11 coefficients was significant at the 10% level, the transformed series was seasonally adjusted using Wallis (1974) linear approximation to X-11 ARIMA.

Fourth, the transformed seasonally adjusted series were screened for large outliers (outliers exceeding 6 times the interquartile range). These were presumed to be coding errors or to result from an anomaly, and each outlying observation was recoded as missing data.

Forecasts are reported below both from the models estimated using monthly data and from the models estimated using quarterly data. Since GDP is only available quarterly, monthly forecasts are not presented for this variable. The quarterly data were constructed as the average of the monthly data in the relevant quarter, after the first three of the four stages outlined above. For example, the quarterly growth rate of IP was constructed as the growth rate of the average of the monthly seasonally adjusted values of IP for 3 months in the quarter. Subsequent to this quarterly transformation, the quarterly series were passed through the fourth preprocessing screen and outliers exceeding four interquartile ranges were replaced by missing values.

This resulted in a total of 580 monthly series, and 562 quarterly series, being available for at least some of the sample; this will be referred to as the nonbalanced panel. Of these, 253 monthly series, and 401 quarterly series, were available for the full sample with no missing data, and these constitute the balanced panel subset of the data. The balanced panel is larger for quarterly than monthly data because fewer of the quarterly series contained large outliers. Specifics, including data definitions and transformation codes, are listed in Appendix A. [1]

2.2. Euro-area aggregate data

This forecasting exercise also uses data on Euro-area aggregates. Unfortunately, the official harmonized series for the Euro area, e.g. the HICP, are available only since 1993 and thus have too short a span to be useful for forecasting experiments. We therefore constructed our own Euro-area aggregate series for GDP, industrial production, consumer price inflation, and the unemployment rate. Quarterly data for GDP over the entire sample period were not available for Belgium, Ireland, Luxembourg and Portugal so our GDP aggregate excludes these countries.

The Euro-area aggregates were constructed as the weighted average of the (transformed) country level data for all 11 countries. Various weighting schemes were tried, including weighting by the relative share of each country's GDP in the Euro-area aggregate, quarter by quarter; a version of this with monthly interpolation; and a fixed-weight scheme using each country's share of nominal GDP for 1997, measured in DM at exchange rates averaged over 1997. Preliminary analysis showed that all the three methods yielded similar aggregate series, and forecast model comparisons were insensitive to the aggregation method used. For simplicity, the results reported here use the fixed weight method based on the 1997 GDP shares.

The aggregates were computed as the share-weighted average of the transformed, seasonally adjusted series. That is, aggregate IP growth is the weighted average of individual country IP growth, aggregate inflation is the weighted average of country inflation, and similarly for GDP growth. The aggregate unemployment rate is the weighted average of the country unemployment rates. For IP, GDP and prices this corresponds to aggregation by geometric weighted averaging.

The resulting series on GDP, IP growth rates, inflation and unemployment changes are quite similar to the official harmonized series for the short span over which both series are available. For the purpose of the forecasting exercise reported here, there are three related advantages to using our constructed series. First, our aggregated series is consistently constructed over the entire sample period, so that any instabilities uncovered in our analysis reflect something other than changes in aggregation method. Instability associated with aggregation could be particularly severe for GDP because of the missing data mentioned above. Second, our aggregation method yields aggregates that are time-invariant linear combinations of the country-specific variables (after transformation). This linear aggregation simplifies the econometric testing problem for comparing alternative forecasting methods. Third, linear aggregation makes that it is straightforward to pool the country-specific forecasts to form a forecast of the aggregate. Such exact aggregation is more difficult using the official harmonized series, as construction of these series evidently includes some degree of temporal smoothing.

[1] This is available from the authors at www.wws.princeton.edu/∼mwatson.

3. Forecasting methods

3.1. Forecasting framework and horizons

For each series to be forecasted, several models are constructed. Each model has a similar structure. Forecasts are made at forecast horizons of one, two, and four quarters for the quarterly data and at 3, 6 and 12 months for the monthly data. All models are specified and estimated as a linear projection of a h-step ahead variable, y^h_{t+h} onto t-dated predictors, which at a minimum include lagged transformed values y_t of the variable of interest. Specifically, the forecasting models all have the form

$$y^h_{t+h} = \mu + \alpha(L)y_t + \beta(L)'Z_t + \varepsilon^h_{t+h}, \tag{1}$$

where $\alpha(L)$ is a scalar lag polynomial, $\beta(L)$ is a vector lag polynomial, μ is a constant, and Z_t is a vector of predictor variables. (For notational ease, we suppress the dependence of μ, $\alpha(L)$, and $\beta(L)$ on the forecast horizon h.)

The "h-step ahead projection" approach reflected in (1) contrasts with the textbook approach of estimating a one-step ahead model, then iterating that model forward to obtain h-step ahead predictions. There are two main advantages of the h-step ahead projection approach. First, it eliminates the need for estimating additional equations for simultaneously forecasting Z_t, e.g. by a VAR. Second, it can reduce the potential impact of specification error in the one-step ahead model (including the equations for Z_t) by using the same horizon for estimation as for forecasting.

Implementation of (1) requires making a decision about how to model the order of integration of the dependent variable. In the base case, the logarithm of IP, the logarithm of GDP, and the unemployment rate were all treated as I(1), so that y_t, respectively, is the growth rate of IP, the rate of growth of GDP, and the first difference of the unemployment rate. The base case modeled the logarithm of prices as I(2), so that y_t denotes the first difference of price inflation. As a sensitivity check, the analysis was repeated with a transformation in which log prices were modeled as I(1) (the "I(1) prices" case).

The dependent variable in (1), y^h_{t+h} was chosen to focus on forecasting problems that we take to be of particular interest. Specifically, in the case of quarterly inflation, at the one quarter horizon we are interested in forecasting the quarter-upon-quarter rate of inflation. At the two quarter horizon, we are interested in forecasting the average inflation rate over the next two quarters, that is, the percentage increase in prices over two quarters. For GDP, we are similarly interested in forecasting either the one-, two- or four-quarter growth of the quarterly index. For the unemployment rate, our interest focuses on the future level of the unemployment rate.

The particulars of the construction of y^h_{t+h} depend on whether the series is modeled as I(1) or I(2) and whether the model is estimated with quarterly or monthly data. First consider the case of quarterly data, so t denotes quarters. In the I(1) case (i.e. y_t is quarterly GDP or IP growth, the quarterly rate of inflation, or the quarterly change in the rate of unemployment), $y^h_{t+h} = \sum_{s=t+1}^{t+h} y_s$. Thus, y^h_{t+h} represents growth in the series between time periods t and $t+h$. In the I(2) price case, $y^h_{t+h} = \sum_{t+1}^{t+h} \Delta p_s - h\Delta p_t$ and

$y_t = \Delta^2 p_t$ where p_t is the logarithm of the price index. For the monthly models (so t denotes months), y^h_{t+h} is constructed to correspond to its quarterly counterpart. For example, consider IP forecasts at the 6 month horizon, and let $\overline{\mathrm{IP}} = (\mathrm{IP}_t + \mathrm{IP}_{t-1} + \mathrm{IP}_{t-2})/3$, where IP is the monthly seasonally adjusted index of industrial production. Then $y^6_{t+6} = \ln(\overline{\mathrm{IP}}_{t+6}/\overline{\mathrm{IP}}_t)$, the rate of change in the quarterly series between months t and $t+6$.

The various forecasting models considered differ in their choice of Z_t. All the methods entail some model selection choices, in particular the number of autoregressive lags and the number of lags of predictor variables Z_t to include in (1). A standard approach to this problem is to employ data-dependent lag selection using an information criterion. However the sample size here is short and this would entail estimating some models with quite low degrees of freedom, yielding results that would be difficult to interpret. For the results reported here, we therefore fix the number of autoregressive lags and the number of predictor variables a priori and do not use data-dependent model selection. In all quarterly models, two lags of y_t were used (that is, lags 0 and 1 of y_t appeared in the right-hand side of (1)). In all monthly models, three lags are used (that is, lags 0–2 of y_t appeared in the right-hand side of (1)).

3.2. Forecasting methods: Country-specific forecasts

Autoregressive benchmark forecast: The autoregressive forecast is a univariate forecast based on (3.1) excluding Z_t.

VAR forecasts: VAR forecasts were constructed using three-variable VARs. Forecasts for industrial production, price inflation and the unemployment rate were constructed using IP–CPI–Unemployment rate VARs. When GDP data were available, GDP forecasts were constructed from a GDP–CPI–Unemployment rate VAR. In all cases the variables were transformed as described in the last section. The multistep forecasts differ from the usual VAR procedure, in that they are computed by the multistep projection method (1) in which the coefficients of (1) are estimated directly by OLS; this contrasts with the usual method in which the coefficients in (1) would be computed from the coefficients estimated in a one-step ahead VAR and iterating the VAR forward h periods. The quarterly VARs included two lags and the monthly VARs used three lags.

AR with EU aggregates: For these forecasts, Z_t is the value of the Euro-area aggregate of the specific series being forecast. For example, for forecast of IP growth, Z_t is the Euro-area aggregate IP growth series (two lags of Z_t were used for the quarterly forecasts and three lags for the monthly forecasts).

AR with US variables: For these forecasts, Z_t is the value of the US aggregate variable that corresponds to the variable being forecast. For example, for forecasts of European IP growth, Z_t is US IP growth; for forecasts of European CPI inflation, Z_t is US CPI inflation. In the I(1) specifications, US inflation is treated as I(1), in the I(2) specifications, as I(2). Two lags of Z_t were used for the quarterly forecasts and three lags were used for the monthly forecasts.

Principal components forecasts: These forecasts are based on setting Z_t to be the principal components from a large number of I(0) candidate predictor time series.

Stock and Watson (1998) show that, if these data can be described by an approximate dynamic factor model, then under certain conditions the space spanned by the latent factors can be estimated consistently (as the cross section and time series dimension increase) by the principal components of the covariance matrix of the predictor time series. Accordingly, we will refer to the principal components as the estimated factors. Stock and Watson (1998) also provide conditions under which these estimated factors can be used to construct asymptotically efficient forecasts by a second stage forecasting regression in which the estimated factors are the predictors. The dynamic factor model interpretation of the principal components forecasts is briefly reviewed in Appendix A.

Here, two sets of principal components are used to construct two alternative sets of forecasts. The first set consists of country-specific principal components of the standardized country-specific series; these will be referred to as the country-specific estimated factors. For example, the principal components for Italy were constructed using the collection of time series pertaining to the Italian economy, each transformed to have mean zero and unit variance. The second set of principal components was computed for the Euro area as a whole by stacking the standardized data for each individual country; these will be referred to as the Euro-wide estimated factors. In all cases, missing data in the nonbalanced panel were handled by computing the principal components using the EM algorithm, as detailed in Stock and Watson (1998).

These two sets of principal components were used to create three forecasting models for each country: one includes the country specific factors, one included the Euro factors, and one included both sets of factors. For the country-specific forecasting model, Z_t consists a set of country-specific factors and for the Euro-area PC forecasting model, Z_t consists of a set of factors from the pooled Euro-area data. For the combined CS/Euro PC forecasting model, Z_t consists of both country-specific and Euro-area factors. In all cases, two lags of the variable being forecast was included in the regression for the quarterly specifications and three lags were used for the monthly models. Two factors were used in the benchmark quarterly I(2) price specification and four factors were included in the monthly model. More discussion of these choices is provided below.

3.3. Forecasting methods: Forecasts of Euro-area aggregates

The Euro-area aggregates were forecast in two ways: by pooling country-specific forecasts, and by directly forecasting the aggregate variables using other aggregate variables.

The pooled country specific forecasts were computed as a weighted average of the individual country-specific forecasts, using the same fixed 1997 real GDP share weights as were used to construct the Euro-wide aggregate series. For example, the pooled AR forecast of two-quarter IP growth is the weighted average of the 11 individual country-specific AR forecasts of two-quarter IP growth.

The other approach is to forecast directly the aggregate series at the Euro-wide level. This entails applying the forecasting methods described in the previous section to the Euro aggregates, with the following modifications and clarifications. The AR forecasts of the aggregate were computed using two lags of the aggregate being forecast. The VAR forecasts were computed using three-variable VARs identical to the country-specific specifications, but with the Euro aggregates replacing the country-specific series. The aggregate PC forecasts were computed using Euro-wide factors; two lags of the aggregate being forecast was included in the quarterly models and three lags were included in the monthly models. The number of Euro-wide factors was determined by Bai and Ng (2002) selection criteria. Two factors were used in the benchmark quarterly I(2) price model, three factors in the quarterly I(1) price model and four factors in the monthly I(2) price model. [2]

From a theoretical perspective, pooling the country-specific forecasts should produce lower mean squared forecast errors than directly forecasting the aggregates, provided that the country-specific models are time invariant, that they are correctly specified, that the model parameters differ across countries, that there are no data irregularities, and that there are enough observations (the theoretical results being asymptotic); see e.g. Lutkepohl (1987). Whether these assumptions are useful approximations in practice, and thus whether pooled country forecasts or direct forecasts of the aggregates actually work better, is an empirical question.

3.4. Model comparison methods

Full sample comparisons: The models are compared over the full sample using several conventional test statistics. These include the F-test of the restriction that Z_t does not enter (1). Also reported are F-tests of the restriction that the coefficients of the country-specific models are equal; under this restriction, there is no information loss by aggregation (the linear models aggregate to an aggregate model with the same coefficients) and the pooled forecasts would simply be an overparameterized version of the forecasts of the aggregates using Euro-wide data. For these comparisons, all models were estimated over the full dataset (adjusted for initial conditions).

Simulated out of sample comparisons: A simulated out of sample forecasting experiment was also performed. For this exercise, all statistical calculations were done using a fully recursive, or simulated out of sample, methodology. This includes all model estimation, standardization of the data, calculation of the estimated factors, etc. At each date in the simulated out of sample period, the factors were recomputed, models re-estimated, etc. The simulated out of sample forecast periods are 1993:I–1997:II (quarterly) and 1993:1–1997:8 (monthly). All forecast-based statistics, and all discussion of forecast performance below refer to these simulated out of sample forecasts over the simulated out of sample period.

The forecasting performance of the various models were examined by comparing the simulated out of sample mean squared forecast error of a candidate forecast, relative to the mean squared forecast error of the benchmark AR forecast.

[2] Bai and Ng (2002) propose several criteria that provide consistent estimators of the number of factors. We computed their ICP_1, ICP_2 and ICP_3 criteria using the pooled balanced panel for each of the dates in our out-of-sample period. The number of factors (2 for quarterly I(2), 3 for quarterly I(1) and 4 for quarterly I(2)) correspond to the modal values of the estimated number of factors over the out-of-sample period.

4. Empirical results

4.1. Quarterly models with I(2) prices

Results for the models of the Euro-wide aggregates for the base case (quarterly data, prices modeled as I(2), factors computed using the nonbalanced panel) are given in Table 1. Table 2 provides a summary of the relative performance of the various methods in the simulated out-of-sample forecast period, and results for the individual countries are given in an unpublished appendix.[3]

Panel A of Table 1 presents results based on the full-sample regressions. The top section of the table shows p-values for exclusion tests for Z_t in the various specifications, and the bottom section shows p-values for equality of coefficients in the country-specific models. Panel B of Table 1 presents detailed results for the simulated out-of-sample forecasting experiment for each of the three forecast horizons. The top section of each panel presents the root mean square forecast error (RMSFE) of the univariate autoregression over the out-of-sample period, and the implied RMSFE computed from the full-sample autoregression. The next section presents the mean square forecast error of each of the models over the out-of-sample period relative to the mean square forecast error for the univariate autoregression

Panel A of Table 2 summarizes the results in Table 1, by presenting binary comparisons of the RMSFE of each of the forecasting methods over the out-of-sample period. For each of the 12 forecasts (four variables, three horizons), the panel shows the fraction of forecasts in which the method corresponding to a given row was more accurate than the method corresponding to a given column. Panel B contains a similar summary, but now for the 120 country-specific forecasts (seven countries with four variables, four countries with three variables, all with three horizons).

These tables suggest seven conclusions.

First, looking across all variables and horizons, none of the multivariate methods consistently outperform the univariate autoregression. This result holds for the forecasts of the Euro aggregates and the country-specific variables. For example, from panel B of Table 2, the VAR provided more accurate forecasts than the univariate AR model in only 33% of the country-specific forecasts. Similar results hold for the other multivariate forecasts.

Second, aggregating the country-specific forecasts provides more accurate forecasts than the forecasts based on models for the Euro aggregates. From panel A of Table 2, in 83% of the cases, the aggregated AR models were more accurate that the AR model constructed for the Euro aggregate. Similar results hold for the VAR, the factor model (PC-Euro) and the model that incorporates the US variable (AR + US). These results are consistent with the in-sample tests for coefficient equality shown in the bottom of panel A of Table 1, which shows that the aggregation restriction is strongly rejected for all of the models except one.

Third, the multivariate models provided more accurate forecasts for the unemployment rate. From panel B of Table 1, the best performing multivariate models are the

[3] This is available from the authors at www.wws.princeton.edu/∼mwatson.

aggregated factor models and the aggregated VAR. For the unemployment rate, these models dominate the univariate autoregression and the other models for each of forecast horizons considered.

Fourth, a general feature of these results is that the full-sample statistics and simulated out-of-sample forecasting comparisons often give conflicting conclusions. This is true both for the forecasts of the aggregate and for the individual country forecasts. One possible explanation for this discrepancy is that the coefficients in the predictive relationships change over the course of the sample; indeed, this would be expected given the move towards European integration over this period. Another explanation is that the simulated out of sample forecast period simply is too short to make sharp inferences because of sampling uncertainty, as is suggested by the large standard errors on the relative MSEs.

Fifth, there is no compelling evidence that using the counterpart US series helps to forecast the European variables. At the Euro-aggregate level, the US series is significant in the in-sample regressions only for IP growth Table 2 indicates that it provides more accurate forecasts than a univariate autoregression in far fewer than half the cases.

Sixth, panel B of Table 1 shows that the multivariate models perform particularly poorly for predicting GDP growth rates. The first two rows of the table suggest part of the explanation: evidently GDP was much less volatile in the out-of-sample period that in the earlier part of the sample. The multivariate models do not perform poorly relative to their historical performance; instead, the univariate autoregression performs much better than it had in past. The out-of-sample RMSFE of univariate autoregression is roughly one half of its in-sample value.

Seventh, looking across all series and horizons the best performing Euro-aggregate forecast method is the aggregated (or pooled) univariate autoregression. Evidently these pooled forecasts provide a rigorous benchmark for the other forecasting methods.

Finally, all of these conclusions must be tempered by the paucity of information in the simulated out-of-sample experiment. For each series, there are only 17 one-quarter ahead forecasts, eight nonoverlapping two-quarter ahead forecasts and only four nonoverlapping four-quarter ahead forecasts. This means that there is considerable uncertainty in the relative forecasting ability of the different methods.

4.2. Results from alternative models

Because it is unclear a priori whether log prices should be treated as I(1) or I(2), as a sensitivity check, the analysis was repeated treating log prices as I(1) and interest rates as I(0). The results are summarized in Table 3. This change in specification changes the estimated factors, both country specific and Euro wide, and also changes the base autoregressive specification for inflation. The table indicates that the major conclusions from the benchmark model continue to hold for this specification: the aggregated (pooled) forecasts are generally more accurate than the forecasts constructed from models using only the Euro aggregates and the aggregated AR model continues to offer the best overall performance.

Results for monthly I(2) data are summarized in Table 4. Here too, the major results of the quarterly data remain evident. However, in the monthly data the multivariate

Table 1
Results for quarterly Euro-wide aggregates

A. Full sample regression results

Alternative model	CPI inflation	Unemployment rate	Industrial production	Real GDP
P-Values for F-tests of univariate AR models versus alternative models				
AR + US	0.41	0.18	0.01	0.53
VAR	0.04	0.00	0.02	0.00
PC-Euro	0.52	0.01	0.73	0.00
P-Values for tests of equality of country-specific models				
AR + US	0.00	0.00	0.00	0.00
VAR	0.30	0.00	0.00	0.00
PC-Euro	0.00	0.00	0.00	0.00

B. Results for simulated out of sample forecasts
B.1. One-quarter forecast horizon

	CPI inflation	Unemployment rate	Industrial production	Real GDP
In-sample AR RMSFE	0.0023	0.0785	0.0091	0.0070
Out-sample AR RMSFE	0.0016	0.0929	0.0083	0.0033
MSFE relative to AR model				
AR	1.00	1.00	1.00	1.00
AR + US	1.24	1.15	1.07	1.93
VAR	1.26	0.86	0.92	2.27
PC-Euro	1.95	0.79	1.06	1.38
Agg − AR	0.92	0.81	0.97	0.94
Agg − VAR	1.10	0.55	1.03	2.21
Agg − AR + Eu	1.03	0.97	0.98	1.05
Agg − AR + US	0.96	0.96	1.02	1.99
Agg − PC-CS	1.04	0.52	1.03	1.58
Agg − PC-Eu	1.64	0.53	1.04	1.23
Agg − PC-C&E	1.58	0.48	1.09	2.27

B.2. Two-quarter forecast horizon

	CPI inflation	Unemployment rate	Industrial production	Real GDP
In-sample AR RMSFE	0.0047	0.1786	0.0141	0.0102
Out-sample AR RMSFE	0.0025	0.2185	0.0164	0.0053
MSFE relative to AR model				
AR	1.00	1.00	1.00	1.00
AR + US	1.33	1.33	1.00	2.50
VAR	1.99	0.93	0.93	4.25
PC-Euro	2.43	0.94	1.09	1.39
Agg − AR	1.03	0.77	0.85	0.91
Agg − VAR	0.72	0.53	0.80	3.96
Agg − AR + Eu	1.28	1.02	0.96	0.99
Agg − AR + US	1.13	0.92	0.84	2.48
Agg − PC-CS	0.94	0.53	0.89	3.54
Agg − PC-Eu	1.85	0.62	1.02	1.30
Agg − PC-C&E	1.57	0.56	1.00	3.21

Table 1 (*Continued*)
B.3. Four-quarter forecast horizon

	CPI inflation	Unemployment rate	Industrial production	Real GDP
In-sample AR RMSFE	0.0098	0.4343	0.0266	0.0153
Out-sample AR RMSFE	0.0052	0.6016	0.0371	0.0091
MSFE relative to AR model				
AR	1.00	1.00	1.00	1.00
AR + US	1.43	1.78	0.97	1.79
VAR	4.26	1.18	1.07	6.16
PC-Euro	1.76	1.11	1.13	2.91
Agg − AR	0.90	0.60	0.70	1.09
Agg − VAR	1.13	0.51	0.62	4.48
Agg − AR + Eu	1.37	0.97	1.06	1.72
Agg − AR + US	0.99	0.76	0.73	1.74
Agg − PC-CS	0.57	0.47	0.82	3.98
Agg − PC-Eu	1.20	0.70	1.01	2.78
Agg − PC-C&E	0.93	0.63	1.19	4.71

Note: The models labeled AR, AR + US, VAR and PC-Euro were constructed with the EU-Aggregates. Models prefaced with “Agg−” are the pooled country-specific models. AR is the univariate autoregression, AR + US includes the US aggregate, VAR is vector autoregression, AR + EU includes the EU aggregate, PC-Euro includes factors estimated from the pooled EU dataset, PC-CS includes factors estimated from the country specific datasets, and PC-C&E includes both sets of factors.

models appear to perform somewhat better. In particular the aggregated VAR model performs better than the aggregated AR model in more that half the cases, and the AR + US model performs better than the AR model in country-specific forecasts.

Finally, as an additional check we constructed the factor models for the EU-aggregates using the balanced panel of data series (401 series versus the 562 series in the unbalanced panel). The resulting forecasts were very close to the nonbalanced forecasts, and the forecast summaries are very similar to the results reported in Tables 1 and 2.

5. Discussion and conclusions

Macroeconometric analysis of time series data as short as these, with as many missing observations as these, and with series definitions which vary from country to country poses special challenges. Still, some interesting conclusions emerge from this analysis.

As was discussed in the introduction, some of these findings have relevance for empirical macroeconometrics beyond the specific European forecasting problems considered here. The finding that, in many cases, pooling country-specific forecasts outperforms directly modeling the Euro-area aggregates suggests that, even if interest is in aggregate measures of Euro-area economic performance, country-specific details matter. This is consistent with the very different political and economic situations of these 11 countries over this period, and over time this might change as the economies of these countries become more closely integrated.

Table 2
Comparison of simulated out-of-sample forecasting results quarterly data, I(2) prices

A. Euro-aggregate forecasts. Fraction of series/horizons in which row-method beat column-method

	AR	AR + US	VAR	PC-EU	Agg − AR	Agg − VAR	Agg− AR + EU	Agg− AR + US	Agg− PC-CS	Agg− PC-Eu	Agg− PC-C&E
AR	—	0.83	0.67	0.83	0.17	0.50	0.58	0.42	0.42	0.75	0.58
AR + US	0.17	—	0.58	0.50	0.00	0.25	0.08	0.08	0.17	0.42	0.67
VAR	0.33	0.42	—	0.50	0.08	0.08	0.33	0.17	0.08	0.25	0.42
PC-Euro	0.17	0.50	0.50	—	0.08	0.25	0.17	0.25	0.25	0.00	0.42
Agg − AR	0.83	1.00	0.92	0.92	—	0.50	1.00	0.92	0.58	0.83	0.83
Agg − VAR	0.50	0.75	0.92	0.75	0.50	—	0.58	0.50	0.33	0.67	0.75
Agg − AR + Eu	0.42	0.92	0.67	0.83	0.00	0.42	—	0.33	0.42	0.58	0.67
Agg − AR + US	0.58	0.92	0.83	0.75	0.08	0.50	0.67	—	0.50	0.58	0.67
Agg − PC-CS	0.58	0.83	0.92	0.75	0.42	0.67	0.58	0.50	—	0.75	0.83
Agg − PC-Eu	0.25	0.58	0.75	1.00	0.17	0.33	0.42	0.42	0.25	—	0.42
Agg − PC-C&E	0.42	0.33	0.58	0.58	0.17	0.25	0.33	0.33	0.17	0.58	—

B. Country forecasts. Fraction of series/horizons/countries in which row-method beat column-method

	AR	AR + EU	AR + US	VAR	PC-CS	PC-EU	PC-EU&CS
AR	—	0.63	0.61	0.68	0.61	0.61	0.61
AR + Eu	0.38	—	0.47	0.60	0.46	0.52	0.50
AR + US	0.39	0.53	—	0.68	0.46	0.57	0.58
VAR	0.33	0.40	0.32	—	0.33	0.41	0.40
PC-CS	0.39	0.54	0.54	0.68	—	0.63	0.70
PC-EU	0.39	0.48	0.42	0.59	0.38	—	0.63
PC-EU&CS	0.39	0.50	0.42	0.60	0.30	0.37	—

Note: Each entry shows the fraction of times that the forecast corresponding to the rows of the table had a lower MSFE than the forecast corresponding to the column in the simulated out-of-sample period. See the notes to Table 1 for a definition of the models.

Table 3
Comparison of simulated out-of-sample forecasting results quarterly data, I(1) prices

A. Euro-aggregate forecasts. Fraction of series/horizons in which row-method beat column-method

	AR	AR + US	VAR	PC-EU	Agg − AR	Agg − VAR	Agg− AR + EU	Agg− AR + US	Agg− PC-CS	Agg− PC-Eu	Agg− PC-C&E
AR	—	0.75	0.58	0.67	0.08	0.42	0.58	0.33	0.50	0.50	0.58
AR + US	0.25	—	0.58	0.50	0.00	0.25	0.25	0.08	0.42	0.33	0.58
VAR	0.42	0.42	—	0.50	0.08	0.00	0.50	0.17	0.25	0.25	0.33
PC-Euro	0.33	0.50	0.50	—	0.17	0.42	0.33	0.42	0.25	0.25	0.58
Agg − AR	0.92	1.00	0.92	0.83	—	0.67	1.00	0.75	0.58	0.67	0.67
Agg − VAR	0.58	0.75	1.00	0.58	0.33	—	0.67	0.33	0.42	0.50	0.58
Agg − AR + Eu	0.42	0.75	0.50	0.67	0.00	0.33	—	0.33	0.50	0.50	0.58
Agg − AR + US	0.67	0.92	0.83	0.58	0.25	0.67	0.67	—	0.42	0.58	0.67
Agg − PC-CS	0.50	0.58	0.75	0.75	0.42	0.58	0.50	0.58	—	0.83	0.83
Agg − PC-Eu	0.50	0.67	0.75	0.75	0.33	0.50	0.50	0.42	0.17	—	0.67
Agg − PC-C&E	0.42	0.42	0.67	0.42	0.33	0.42	0.42	0.33	0.17	0.33	—

B. Country forecasts. Fraction of series/horizons/countries in which row-method beat column-method

	AR	AR + EU	AR + US	VAR	PC-CS	PC-EU	PC-EU&CS
AR	—	0.57	0.52	0.71	0.62	0.64	0.71
AR + Eu	0.43	—	0.49	0.63	0.57	0.57	0.66
AR + US	0.48	0.51	—	0.72	0.57	0.59	0.70
VAR	0.29	0.38	0.28	—	0.38	0.46	0.63
PC-CS	0.38	0.42	0.42	0.63	—	0.58	0.81
PC-EU	0.36	0.43	0.41	0.54	0.42	—	0.71
PC-EU&CS	0.29	0.34	0.30	0.38	0.19	0.29	—

Note: Each entry shows the fraction of times that the forecast corresponding to the rows of the table had a lower MSFE than the forecast corresponding to the column in the simulated out-of-sample period. See the notes to Table 1 for a definition of the models.

Table 4
Comparison of simulated out-of-sample forecasting results monthly data, I(2) prices

A. Euro-aggregate forecasts. Fraction of series/horizons in which row-method beat column-method

	AR	AR + US	VAR	PC-EU	Agg − AR	Agg − VAR	Agg− AR + EU	Agg− AR + US	Agg− PC-CS	Agg− PC-Eu	Agg− PC-C&E
AR	—	0.56	0.56	0.56	0.11	0.33	0.33	0.22	0.44	0.56	0.56
AR + US	0.44	—	0.56	0.56	0.33	0.33	0.33	0.00	0.56	0.56	0.56
VAR	0.44	0.44	—	0.56	0.33	0.11	0.44	0.33	0.22	0.33	0.33
PC-Euro	0.44	0.44	0.44	—	0.22	0.33	0.00	0.22	0.33	0.56	0.56
Agg − AR	0.89	0.67	0.67	0.78	—	0.33	0.56	0.56	0.67	0.67	0.67
Agg − VAR	0.67	0.67	0.89	0.67	0.67	—	0.67	0.67	0.67	0.67	0.78
Agg − AR + Eu	0.67	0.67	0.56	1.00	0.44	0.33	—	0.33	0.44	0.56	0.67
Agg − AR + US	0.78	1.00	0.67	0.78	0.44	0.33	0.67	—	0.67	0.56	0.78
Agg − PC-CS	0.56	0.44	0.78	0.67	0.33	0.33	0.56	0.33	—	0.33	0.67
Agg − PC-Eu	0.44	0.44	0.67	0.44	0.33	0.33	0.44	0.44	0.67	—	0.67
Agg − PC-C&E	0.44	0.44	0.67	0.44	0.33	0.22	0.33	0.22	0.33	0.33	—

B. Country forecasts. Fraction of series/horizons/countries in which row-method beat column-method

	AR	AR + EU	AR + US	VAR	PC-CS	PC-EU	PC-EU&CS
AR	—	0.60	0.41	0.66	0.66	0.59	0.62
AR + Eu	0.40	—	0.43	0.68	0.59	0.69	0.73
AR + US	0.59	0.57	—	0.67	0.72	0.72	0.77
VAR	0.34	0.32	0.33	—	0.40	0.46	0.52
PC-CS	0.34	0.41	0.28	0.60	—	0.57	0.62
PC-EU	0.41	0.31	0.28	0.54	0.43	—	0.69
PC-EU&CS	0.38	0.27	0.23	0.48	0.38	0.31	—

Note: Each entry shows the fraction of times that the forecast corresponding to the rows of the table had a lower MSFE than the forecast corresponding to the column in the simulated out-of-sample period. See the notes to Table 1 for a definition of the models.

Looking across series and forecast horizons, no multivariate model beat the pooled univariate autoregressions. The pooled univariate autoregression therefore provides a powerful benchmark for future forecast comparisons.

Within the multivariate methods, factor models, either based on country specific or Euro-wide factors, regularly outperform VARs at the country level based both on full-sample F statistics and on simulated out of sample forecast comparisons. This suggests that conventional small-scale macroeconometric VAR models, and associated VAR policy analysis exercises, could miss important information contained in a large number of variables excluded from the VAR.

Acknowledgements

We would like to thank Fabio Canova, Vitor Gaspar, Jordi Gali, Jerome Henry, Ricardo Mestre, Lucrezia Reichlin, two referees, and seminar participants at the European Central Bank, the Bank of England, the Bank of Spain, and at the CEPR-ESSIM Conference (May 2000) for helpful comments on an earlier draft of this paper. This research was supported in part by National Science Foundation grant no. SBR-9730489 and by the Center for European Studies at Harvard University.

Appendix A. Dynamic factor models and principal components forecasts

This appendix briefly reviews the link between the principal components forecasts and dynamic factor models. This material draws on Stock and Watson (1998).

Let X_t be a N-dimensional multiple time series of variable, observed for $t=1,\ldots,T$. Suppose that X_t has an approximate linear dynamic factor representation with $\bar{r}$ common dynamic factors f_t:

$$X_{it} = \lambda_i(L) f_t + e_{it} \tag{A.1}$$

for $i=1,\ldots,N$, where $e_t=(e_{1t},\ldots,e_{Nt})'$ is a $N \times 1$ vector of idiosyncratic disturbances with limited cross-sectional and temporal dependence, and $\lambda_i(L)$ are lag polynomials in nonnegative powers of L; see for example Geweke (1977), Sargent and Sims (1977), and Forni et al. (1999, 2000). If $\lambda_i(L)$ have finite orders of at most q, (A.1) can be rewritten as

$$X_t = \Lambda F_t + e_t, \tag{A.2}$$

where $F_t = (f_t', \ldots, f_{t-q}')'$ is $r \times 1$, where $r \leqslant (1+q)\bar{r}$, the ith row of Λ in (A.2) is $(\lambda_{i0}, \ldots, \lambda_{iq})$.

Stock and Watson (1998) show, under this finite lag assumption and some additional technical assumptions (restrictions on moments and nonstationarity), 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. The principal component estimator is computationally convenient, even for very large N. Importantly, it can be generalized to handle data irregularities such as missing observations using the EM algorithm.

The consistency of the estimated factors implies that they can be used to construct efficient forecasts for a single time series variable. Specifically, suppose one is interested in forecasting the scalar time series y_{t+1} using the predictors in X_t and suppose that y_{t+1} has the factor structure

$$y_{t+1} = \beta(L)f_t + \gamma(L)y_t + \varepsilon_{t+1}, \tag{A.3}$$

where $\mathrm{E}(\varepsilon_{t+1} \mid \{f_\tau, X_\tau, y_\tau\}_{\tau=-\infty}^{t}) = 0$ (The different time subscripts used for y and X emphasize the forecasting relationship.) If $\{f_t\}$, $\beta(L)$, and $\gamma(L)$ were known, the minimum mean square error forecast of y_{T+1} would be $\beta(L)f_T + \gamma(L)y_T$. It is shown in Stock and Watson (1998) that, if $\beta(L)$ and $\gamma(L)$ have finite orders, then forecasts that are asymptotically efficient to first order can be obtained from OLS estimation of (A.3) with the estimated factors replacing the true factors.

References

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

Fair, R.C., Shiller, R.J., 1990. Comparing information in forecasts from econometric models. American Economic Review 80, 375–389.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 1999. Reference cycles: The NBER methodology revisited. manuscript.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic factor model: Identification and estimation. Review of Economics and Statistics 82, 540–554.

Gali, J., Gerter, M., López-Salido, J.D., 2001. European inflation dynamics. European Economic Review 45, 1237–1270.

Geweke, J., 1977. The dynamic factor analysis of economic time series. In: Aigner, D.J., Goldberger, A.S. (Eds.), Latent Variables in Socio-Economic Models. North-Holland, Amsterdam (Chapter 19).

Lutkepohl, H., 1987. Forecasting Aggregated Vector ARMA Processes. Springer, Berlin.

Sargent, T.J., Sims, C.A., 1977. Business cycle modeling without pretending to have too much a priori economic theory. In: C. Sims et al. (Eds.), New Methods in Business Cycle Research, Federal Reserve Bank of Minneapolis, Minneapolis, MN.

Stock, J.H., Watson, M.W., 1998. Diffusion indexes. NBER working paper #6702.

Stock, J.H., Watson, M.W., 1999. Forecasting inflation. Journal of Monetary Economics 44, 293–335.

Wallis, Kenneth F., 1974. Seasonal adjustment and relations between variables. Journal of the American Statistical Association 69 (345), 18–31.