BANCA D'ITALIA

Temi di discussione

del Servizio Studi

Forecasting Industrial Production in the Euro Area

by G. Bodo, R. Golinelli and G. Parigi

Number 370 - March 2000

The purpose of the "Temi di discussione" series is to promote the circulation of working papers prepared within the Bank of Italy or presented in Bank seminars by outside economists with the aim of stimulating comments and suggestions.

The views expressed in the articles are those of the authors and do not involve the responsibility of the Bank.

SINTESI

Il contenuto di questo lavoro esprime esclusivamente le opinioni degli autori: pertanto non rappresenta la posizione ufficiale della Banca d'Italia.

Nel lavoro viene presentata una metodologia per la previsione di breve termine (1-2 mesi) dell'indice della produzione industriale per il complesso dell'area dell'euro. La metodologia rende disponibili stime tempestive della dinamica di un importante indicatore congiunturale, arricchendo così l'insieme di informazioni su cui si basano le decisioni di politica monetaria per l'area dell'euro. In particolare, con questo tipo di strumenti previsivi è possibile disporre di stime con un solo mese di ritardo rispetto a quello di riferimento (quindi con un mese di anticipo rispetto alla pubblicazione dell'indice provvisorio da parte dell'Eurostat).

Nell'attuale versione del lavoro si sono considerate diverse procedure alternative per la previsione dell'indice aggregato per l'area dell'euro:

a) modelli univariati di serie storiche sia a livello di singolo paese, sia a livello aggregato;

b) modelli VAR (vettoriali autoregressivi) multivariati con riferimento ai quattro principali paesi dell'area dell'euro (Germania, Francia, Italia e Spagna);

c) modelli VAR, con riferimento alle due aree aggregate: euro e Stati Uniti, con e senza indicatori congiunturali tratti dalle indagini qualitative della Commissione europea;

d) utilizzo di previsioni per singoli paesi, derivate da modelli basati su indicatori congiunturali, aggregate per ottenere l'indice dell'intera area dell'euro.

L'analisi riguarda il periodo 1985.1-1999.6 (1987.1-1997.12 per la stima). In una prima fase si presentano i risultati della stima di un modello univariato (*arima*) per il complesso dell'area dell'euro. Quest'analisi, che genera previsioni a breve termine (1-2 mesi) piuttosto affidabili, rappresenta un termine di confronto per le altre metodologie esplorate. Il ricorso a formulazioni di tipo VAR più complesse, ristrette ai soli quattro principali paesi dell'area (l'aggregato di questi paesi assomma a poco più dell'80 per cento della produzione dell'intera area dell'euro) non ha prodotto miglioramenti apprezzabili, soprattutto a causa del ridotto numero di gradi di libertà che queste procedure comportano. Se, tuttavia, si limita l'attenzione a un VAR bivariato, basato sull'indice di produzione degli USA e su quello dell'aggregato euro, si ottengono risultati nettamente migliori rispetto alla semplice formulazione *arima*. Quando anche l'indice del clima di fiducia delle imprese europee viene incluso tra le variabili a cui la previsione risulta condizionata, il miglioramento è ancora più netto.

Infine, in una terza fase si presentano le stime di modelli “nazionali”, dove gli indici di produzione dei vari paesi sono messi in relazione con svariati indicatori, fra cui variabili desunte dalle indagini qualitative della Commissione europea, indici di produzione di paesi esterni all’area dell’euro (Regno Unito e Stati Uniti) e altri indicatori specifici, come il consumo di energia elettrica (per la sola Italia). I risultati sono più soddisfacenti in alcuni casi – Italia e Francia - meno in altri – Germania e, soprattutto, Spagna.

Il confronto tra l’aggregazione delle previsioni “nazionali” e quelle calcolate secondo le procedure descritte in precedenza mette in luce per il periodo 1998.1-1999.6 una migliore capacità previsiva delle seconde, caratterizzate da un errore medio sensibilmente più basso.

FORECASTING INDUSTRIAL PRODUCTION IN THE EURO AREA

By Giorgio Bodo*, Roberto Golinelli** and Giuseppe Parigi***

Abstract

The creation of the Euro area has increased the importance of obtaining timely information about short-term changes in the area's real activity. In this paper we propose a number of alternative short-term forecasting models, ranging from simple ARIMA models to more complex cointegrated VAR and conditional models, to forecast the index of industrial production in the euro area. A conditional error-correction model in which the aggregate index of industrial production for the area is explained by the US industrial production index and the business confidence index from the European Commission harmonised survey on manufacturing firms achieves the best score in terms of forecasting capacity.

Contents

*Fiat, **University of Bologna, ***Bank of Italy, Research Department.

1. Introduction[1]

The creation of the Euro area has certainly posed a number of important economic issues: the extent of the convergence process, from both the real and nominal points of view, the effects deriving from the Stability Pact, the role played by the European Central Bank and by the national central banks, etc. At the same time one of the key aspects has become that of obtaining timely information concerning the Euro area economic cycle and, if possible, its future developments[2]. Obviously, these problems have already been examined at a national level but so far we do not know of any recent contribution at the Euro aggregate level. We therefore decided to try to develop some simple time series models able to forecast short-term changes in real activity in the Euro area. Because the economic cycle is influenced by the dynamics of the industrial sector, and because the information is provided without much delay, we focused our research on forecasting models for the industrial production index. Such information should be very helpful for economic policy decisions.

We began our analysis from simple univariate Arima models for the Euro area as a whole and by country (Germany, France, Italy and Spain; paragraph 2) and then moved to a VAR system in which we consider the industrial production indices for these four euro countries (paragraph 3). At this stage we introduced new variables which might significantly affect the European industrial production index; i.e. the United States production index and the level of confidence in the business sector computed by the European Commission. The US index is supposed to approximate the evolution of demand outside the Euro area. Apart from the fact that the US index produces better results than some alternatives (in particular the index of the UK), this choice may also be motivated by the reliability of the US index and the very short delay with which it is published[3].

[1] We are grateful to G. Bacchilega, L. Picci, M. Marcellino, F. Signorini, M. Magnani, P. Zaffaroni and the participants in a seminar at the research department of the Bank of Italy for useful comments. The usual caveats apply. The views contained here are those of the authors only and do not necessarily reflect those of the institutions for which they work.

[2] We use the terms Europe, EMU and Euro to refer to the 11 member countries of the Euro area.

[3] For the usefulness of the US index for forecasting see also Bodo *et al*. (1997).

In paragraph 4 we tried to evaluate the possibility of improving our forecast for the area by building country specific models with indicators and then by aggregating them. Finally, in paragraph 5 we present a comparison of the different proposed forecasting approaches and then conclusions are drawn.

2. Statistical analysis of the indices of industrial production

In a recent forecasting exercise for 215 US monthly macroeconomic time series with alternative methods, Stock and Watson (1998) find that the best overall performance of a single method is achieved by autoregressions with unit root pretest: «If a macroeconomic forecaster is restricted to use a single method, then for the family of loss functions considered here she would be well advised to use an autoregression with a unit root pretest and data-dependent lag-length selection.» (p. 21).

Though this paragraph is devoted to building univariate models for benchmark forecasts, after preliminary data inspection (section 2.1) we followed the Stock and Watson's suggestion to test for seasonal and non seasonal unit roots in the variables under scrutiny (section 2.2). Finally, the retained univariate ARIMA models for the Euro area as a whole, Germany, France, Italy and Spain are presented (section 2.3).

2.1 Preliminary data analysis

Analysis of the levels and the first differences of the manufacturing output indices for Germany, France, Italy, Spain, the Euro area and the United States (Fig. 1; see the Appendix Data for more details on the statistical sources) suggests that seasonality and (to a less extent) non-stationarity are the main features of the variables over the period 1985-1999. In particular, seasonality seems to follow a regular path in all European countries; for Germany, the seasonal fluctuations of August are smaller than in the other countries and give the (false) impression that seasonality is not important. The cyclical fluctuations of the log-levels around the trend line are about the same for all European countries: the peaks in 1991-1992, the 1993 recession and the subsequent expansion are common to all series (in Italy this evolution is blurred by the largest seasonal fluctuations). The US seasonal pattern is less evident and, since 1991, cyclical fluctuations are absent.

Figure 1

INDUSTRIAL PRODUCTION, LOG-LEVELS AND FIRST DIFFERENCES

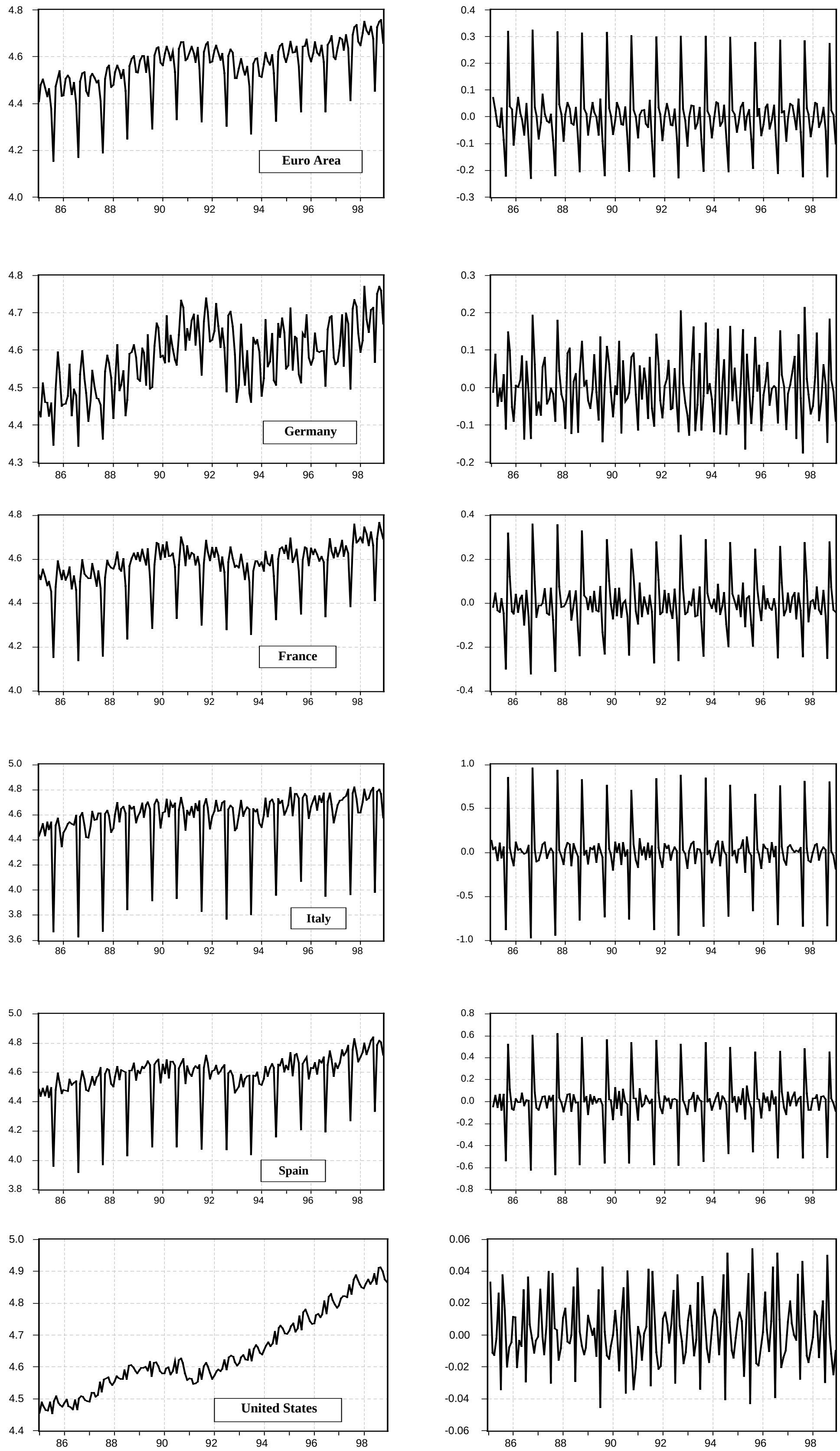


Table 1 shows three descriptive statistics for the first differences of the series reported in Figure 1. From the first row of Table 1 we see that the euro area industrial production index grew on average at about 2 per cent per year (adjusted for the effect of working days), while the US growth rate was by far more sustained.

Table 1

DESCRIPTIVE STATISTICS FOR FIRST DIFFERENCED DATA (1)
(1987.1-1997.12)

	Euro area	Germany	France	Italy	Spain	United States
Mean (on annual basis)	0.0192	0.0146	0.0145	0.0177	0.0176	0.0324
Standard deviation	0.1148	0.0851	0.1215	0.3403	0.2281	0.0220
R^2	0.9863	0.6619	0.9405	0.9771	0.9635	0.9177

(1) All variables are expressed in logarithms.

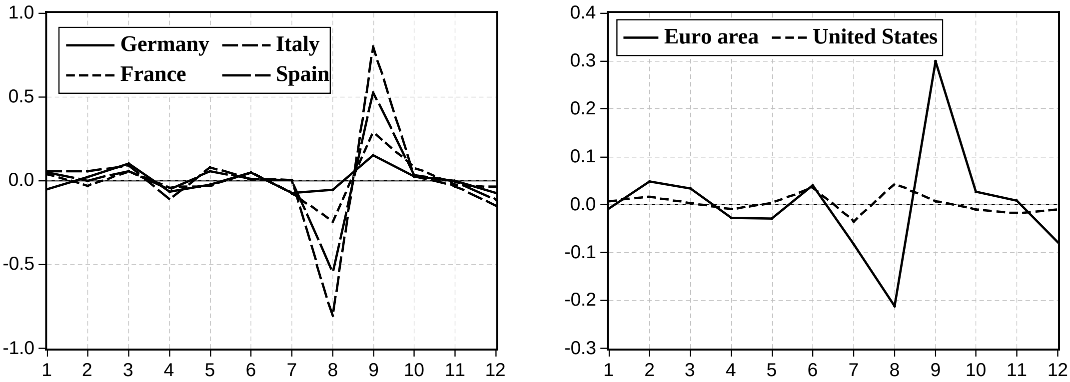
Sample standard deviations show greater volatility for Italy and Spain; overall the Euro area variability is about five times that of the US. Output growth volatility seems to be related to the seasonal pattern, as summarised by the R^2 (obtained from the regression of the first difference of the industrial production on twelve seasonal dummies). With the exception of Germany, more than 90 per cent of the output variability is explained by seasonality, even though a high R^2 can arise in the presence of seasonal unit roots and hence does not necessarily imply deterministic seasonal patterns (see also Osborn *et al.*, 1999, p. 31).

Figure 2 reports the estimates of the seasonal parameters of the previous «R^2 regressions». In Italy, Spain and France seasonality is particularly strong in August and September, while in Germany the August effect is partly anticipated in July. From the graph on the right it is clear that the euro area and the US seasonal paths are very different, both in timing and amplitude.[4] It is also worth noting that the regressions depicted in Figure 2 were characterised by fairly stable parameter estimates, as suggested by their recursive least squares estimation (results not reported).

[4] These findings are similar to those reported, at industry level, in Osborn *et al.* (1999, p. 33) and Miron (1996, Table 3.3).

Figure 2

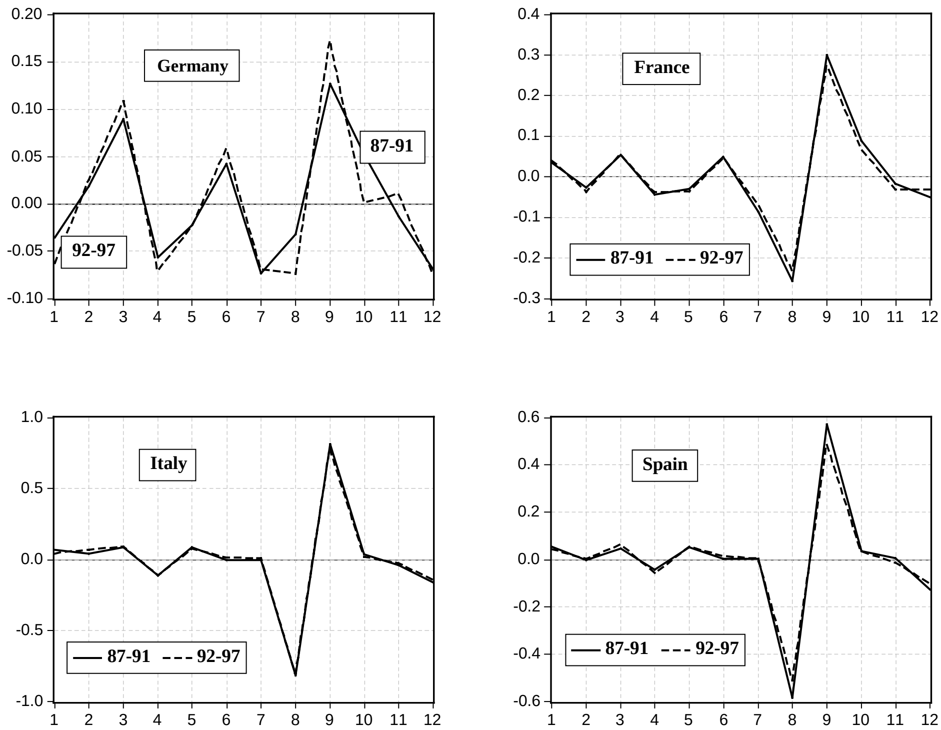
SEASONAL ESTIMATES FOR FIRST DIFFERENCES (1987.1-1997.12)



In addition, the above regressions with the sample split into two sub-periods (1987.1-1991.12 and 1992.1-1997.12) are stable, as depicted in Figures 3 (at country level) and 4 (at macro area level).

Figure 3

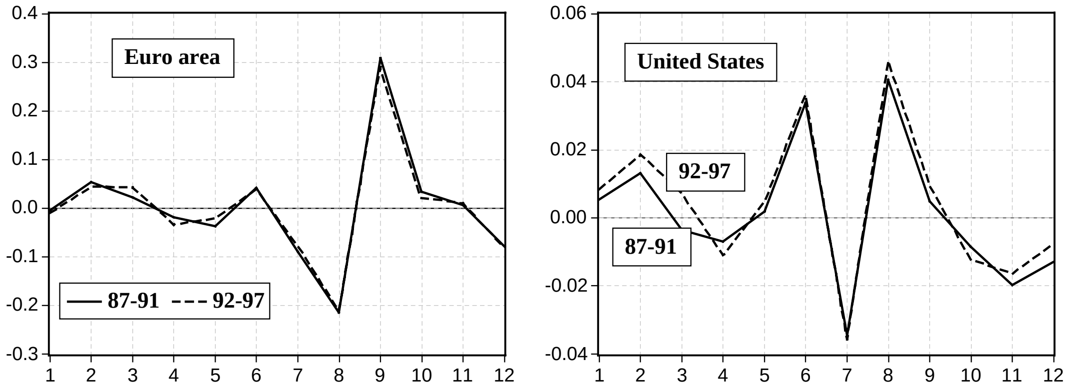
SEASONAL STABILITY IN FOUR EUROPEAN COUNTRIES



The descriptive nature of our analysis and the brevity of the time span prevent us from examining thoroughly the nature (deterministic or stochastic) of the seasonality of economic variables; this may help to explain the difference between our results and those reported in Hylleberg *et al.* (1993, p. 329).

Figure 4

SEASONAL STABILITY IN EURO AREA AND UNITED STATES



The main findings of this section can be summarised as follows. As far as the four European economies are concerned, we have found evidence that: seasonality matters and seems to be stable over the period under scrutiny; output growth over the period 1987-1997 was substantial, even though graphically hidden by seasonality. The US index grew more rapidly than the European one, with considerably lower volatility and a different seasonal pattern; nevertheless, seasonality is also a relevant component of the US output fluctuations as well.

2.2 Unit root tests and optimal differencing filters

The traditional approach to time series modelling proposed by Box and Jenkins (1970) sometimes leads to an incorrect filter selection and, consequently, to biased forecasts (see Franses, 1991). As an alternative, a particular differencing filter may be chosen by assuming an appropriate number of seasonal and non-seasonal unit roots in a time series, following a formal testing procedure based on the seminal Dickey and Fuller (1979) approach.

According to Dickey and Pantula (1987) we start from the highest level of differencing and then test down to lower levels. More specifically, we first test for both seasonal and non-seasonal unit roots, using the monthly version of the Osborn-Chui-Smith-Birchenhall approach (OCSB, see *e.g.* Franses and Hobijn, 1997 and Osborn, *et al.*, 1999).

The OCSB test results are summarised in Table 2; they broadly confirm the outcomes of the descriptive analysis in which we found no changing seasonal patterns, as typical of data generated by seasonal unit roots. Thanks to this result, we avoided further seasonality analyses based on the suggestions of Hylleberg et al. (1990) and extended to monthly data in Franses (1998); for a more detailed discussion of these topics see Franses (1996).

Table 2

OCSB SEASONAL AND NON-SEASONAL UNIT ROOT TEST (1)
(1987.1-1997.12)

Country	$t(\pi_1)$	$t(\pi_2)$	$F(\pi_1, \pi_2)$	p	Q stat (12)	Q stat (24)	JB
Euro area	-1.81*	-5.87**	23.99***	8	6.8 [0.87]	20.2 [0.68]	1.46 [0.48]
Germany	-2.18**	-5.75**	20.75**	10	16.0 [0.19]	33.8 [0.09]	0.51 [0.77]
France	-2.29**	-5.69**	21.39**	12	13.6 [0.33]	23.2 [0.51]	1.04 [0.59]
Italy	-2.30**	-7.11***	32.44***	12	11.6 [0.48]	26.3 [0.34]	1.62 [0.44]
Spain	-2.23**	-4.60	15.42	7	13.8 [0.32]	31.4 [0.14]	19.9 [0.00]
US	1.01	-10.18***	54.99***	9	16.2 [0.18]	27.5 [0.28]	7.63 [0.02]

(*) 10%, (**) 5% and (***) 1% significant (critical values are from Franses and Hobijn (1997, Table 12)).
Q and JB are the Lijung-Box and Jarque-Bera test statistics (p-values in squared brackets).

(1) The OCSB test is based on the regression:

$$A(L)\, \Delta_1\Delta_{12} y_t = \Sigma_s\, \delta_s\, D_{st} + \pi_1\, \Delta_{12} y_{t-1} + \pi_2\, \Delta_1 y_{t-12} + \varepsilon_t$$

where $A(L)$ is a polynomial of order p in the lag operator L such that the estimated residuals are white noise; D_{st} is a seasonal dummy equal to 1 in the month s (s = 1, 2, …, 12) and zero otherwise; δ_s, π_1 and π_2 are parameters. Interaction terms between seasonality and trend are not included, given the preliminary data analysis results. The linear trend is excluded because it is never significant. We test the null hypotheses of the appropriateness of: a) the $\Delta_1\Delta_{12}$ filter against that of Δ_{12} or Δ_1 or no filter at all; b) the Δ_{12} filter against the alternatives Δ_1 or no filter. The first test is accomplished by a joint $F(\pi_1, \pi_2)$ statistic, the second by a $t(\pi_2)$ statistic (notice that since Δ_{12} filter contains the Δ_1 component, the test for $\pi_1 = 0$ is not *per se* decisive).

However, since the presence of unit roots at the zero frequency cannot be excluded, we evaluated a number of non-seasonal stationarity tests. The first two columns of Table 3 show the augmented Dickey-Fuller (ADF) test results for both levels and first differences of the output indices: output levels are never at least 10% stationary, while first differences often are, with the exception of Germany, Italy and the US. The results for Italy and Germany are contradicted by the Kwiatkowski-Phillips-Schmidt-Shin test statistics (KPSS, see Kwiatkowski *et al.*, 1992) shown in the third column of Table 3. As is known, while the ADF is a unit root test, the KPSS is a stationarity test, since it concentrates on partial sums of the residuals from an auxiliary regression of the output levels on a constant, a trend and seasonals. As this test needs an estimate of the long run variance of the residuals;

we chose a window size equal to 12, the square root of the total number of observations, as this value should yield the most favourable results in terms of rejection frequencies (see Franses, 1998). The KPSS outcomes are the same for all European countries: output levels are I(1), while first differences are always stationary.

Table 3

UNIT ROOT TESTS AT ZERO FREQUENCY
(1987.1-1997.12)

Variable (1)	ADF	k (2)	KPSS	PB	k (2)	break in:
Lipue11	-1.89	18	0.378***	-4.03	20	
Dlipue11	-2.74*	17	0.077			
Lipde	-2.36	18	0.479***	-3.74	18	
Dlipde	-1.92	17	0.013	-7.06***	4	1993.1
Lipfr	-1.50	22	0.346***	-3.02	22	
Dlipfr	-2.97**	21	0.020			
Lipit	-2.48	23	0.246***	-5.02*	23	1992.5
Dlipit	-2.16	22	0.008			
Lipsp	-1.52	11	0.405***	-4.75	15	
Dlipsp	-3.69***	10	0.019			
Lipus	-1.67	18	0.809***	-5.04*	24	1990.12
Dlipus	-1.21	17	0.495**			

(***) 1%, (**) 5% and (*) 10% significant.
(1) A letter *L* at the beginning of the name stands for «logarithm», a *D* stands for «first difference». The variable labels are composed of a fixed part, *ip*, for «industrial production index» followed by the country indicator: *ue11* (Euro area), *de* (Germany), *fr* (France), *it* (Italy), *sp* (Spain) and *us* (United States). For example, *Dlipit* is the first difference of the log of the Italian industrial production index.-(2) The lag length was chosen by starting from a large number of lags (*k*=24, to consider additional seasonal effects) and dropping the lag parameters not 10% significant (see Campbell and Perron, 1991). For the levels a constant, a trend and seasonals are used; for the first differences the trend variable is omitted.

The ADF and KPSS tests are based on the assumption of stable deterministic components. In order to assess the robustness of this hypothesis, the last three columns in Table 3 report the outcomes of the Perron (1997) test (PB) with breaks in both the intercept and the trend slope at an unknown point in time (Perron defines it as model 2). On the basis of preliminary analysis, we still assume constancy of seasonal dummies over the whole sample. The PB test confirms the ADF results for the Euro area, France and Spain, while it finds possible breaks every time the ADF test detected two possible roots. For Germany, it explains one of the two roots with a break in the growth rate after January

1993; Italy and the US output series seem to be trend stationary with a 10% significant trend slope break (respectively, in May 1992 and December 1990).

2.3 Univariate benchmark forecasting models

The somewhat confused results of the zero frequency unit root tests led us to search for the univariate benchmark models by setting up a number of alternative specifications (levels, first differences, structure of lags, etc.). Both the diagnostic tests and the explanatory power of alternative models (standard errors of the regression) guided our choice of the «best» models over the estimation period (1987.1-1997.12). As far as the out-of-sample (1998.1-1999.6) forecasting performance is concerned, the results are comparatively analysed in paragraph 5. Table 4 presents the preferred models for each country and the Euro area aggregate.

Table 4

BENCHMARK MODELS
(1987.1-1997.12)

	Euro area		Germany		France		Italy		Spain	
Specification ([1])	Estimate	t	Estimate	t	Estimate	t	Estimate	t	Estimate	t
AR(1)	-0.371	-4.3	-1.012	-14.0	-0.842	-10.6	-0.704	-7.6	-0.846	-10.2
AR(2)			-0.613	-8.5	-0.527	-6.6	-0.492	-4.8	-0.439	-5.2
AR(3)							-0.196	-2.1		
SAR(12)	0.337	3.8			0.374	4.2	0.352	4.0	0.242	2.7
SAR(24)							-0.355	-4.0		
SAR(36)							-0.337	-3.7		
Diagnostic checks ([2])	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Adjusted R^2	0.988		0.859		0.970		0.993		0.980	
S.E. of regression (%)	1.27		3.20		2.09		2.94		3.23	
Durbin-Watson	2.10		2.01		1.88		2.02		2.02	
LM(12)	1.25	0.26	1.21	0.28	1.23	0.27	1.18	0.31	1.03	0.43
LM(24)	1.33	0.17	1.23	0.24	1.64	0.05	1.28	0.20	0.93	0.57
ARCH(12)	0.69	0.76	1.03	0.43	0.85	0.60	0.92	0.53	0.83	0.62
WHITE	0.67	0.76	1.35	0.21	0.94	0.51	1.83	0.06	1.90	0.05
JB	1.56	0.46	0.63	0.73	0.46	0.79	0.07	0.97	10.73	0.01
RESET(2)	1.69	0.19	1.02	0.36	0.36	0.70	1.63	0.20	0.51	0.60
CHOW	1.02	0.44	0.92	0.53	0.29	0.99	1.27	0.25	1.12	0.35

([1]) The dependent variable is the Δlog of the industrial production index. Coefficients estimates of seasonal dummy variables and linear trend are not shown.-([2]) AR(p) p-th autoregressive component. SAR(p): p-th multiplicative autoregressive seasonal component. LM(p) p-th residuals autocorrelation test; ARCH(p) p-th autoregressive conditional heteroschedasticity test; WHITE heteroschedasticity test; JB normality test; RESET nonlinearity test up to the third power; CHOW predictive failure test over the period 1998.1-1999.6.

The main characteristics are the following:

a) Seasonality is modelled with dummies plus seasonal autoregressive multiplicative components (deterministic seasonal estimates, in common with other deterministic components, are not reported in Table 4). Estimates never include impulse dummy variables, even though we sometimes checked for the robustness of our results to the presence of outliers (as for the case of Spain, see below).

b) Stochastic components are specified in first differences and autoregressive terms, *i.e.* ARIMA(p,1,0); hence we never detected relevant moving average components. Whenever ARMA trend stationary models appeared as valid representations of output levels, their estimation results did not differ significantly (after reparametrisation) from the corresponding, preferred, ARIMA model, and showed reduced forecasting ability. An explanation is reported in Franses and Kleibergen (1996): ARIMA models adapt more rapidly to structural parameter changes (if any occur along the forecasting horizon) than ARMA trend stationary specifications. Moreover, ARIMA models are more parsimonious, since they impose unit root restrictions.

c) Preferred models are never affected by relevant specification problems, as evidenced by the performance of the diagnostics tests reported in Table 4 (for Spain, the results of the heteroschedasticity and normality tests are affected by the presence of two outliers).

d) Overall, the sample performance of the models is satisfactory: regression standard errors exceed 3 per cent in only two cases, though the marked seasonality pattern of the variables help to increase the explanatory power of models.

3. Multivariate forecasting models

Sometimes the forecasting performance may be improved by modelling the interrelationships among countries or areas through vector autoregressive models (VAR), though we are conscious that modelling levels is often a delicate task because of possible structural breaks occurring in the forecasting horizon (see Clements and Hendry, 1998). Given the univariate evidence, two cases are possible: (a) the integrated variables are cointegrated, in this case levels matter; (b) the variables are not cointegrated but the multivariate model in first differences captures useful causal links.

Paragraph 3.1 analyses two VAR models for the outputs of Germany, France, Italy and Spain (with and without the US); given the poor results obtained, in paragraph 3.2 we try to improve the forecasting ability by using more parsimonious models, with only the euro area and the US output indices. Finally, the importance of augmenting the previous bivariate VAR with the European business confidence index is assessed. Overall, the results show that the gains from extending the information set to single country data are smaller than those that can be obtained from a more articulated lag structure for aggregate areas.

3.1 A vector autoregressive model for the Euro area countries and the United States

To avoid specification problems, the minimum lag starting point is a VAR model for the industrial output levels of Germany, France, Italy and Spain with 13 lags and a standard deterministic nucleus with constant, seasonal dummies and linear trend. In this model we tested for cointegration according to the Johansen (1995) procedure (section (a), Table 5).

Table 5

COINTEGRATION ANALYSIS OF INDUSTRIAL PRODUCTION LEVELS
(1987.1-1997.12)

Ho: rank = p	Max eigenv. Test	c.v. (1)	Trace test	c.v. (1)
	(a) without the US			
$p = 0$	24.86	31.5	53.91	63.0
$p \leq 1$	16.22	25.5	29.05	42.4
$p \leq 2$	7.55	19.0	12.83	25.3
$p \leq 3$	5.28	12.3	5.28	12.3
	(b) with the US			
$p = 0$	30.36	37.5	67.97	87.3
$p \leq 1$	19.85	31.5	37.61	63.0
$p \leq 2$	9.97	25.5	17.76	42.4
$p \leq 3$	4.99	19.0	7.79	25.3
$p \leq 4$	2.79	12.3	2.79	12.3

(1) Osterwald-Lenum (1992) 5% critical values.

The cointegration rank of the VAR(13) was found to be zero, so that no linear combination among output levels (each assumed to be I(1)) is stationary.[5] Similar results

[5] We tested for the cointegration rank of the VAR model in differences, with and without US output: the null of reduced rank was always found to be significant, suggesting that output levels can be considered I(1).

are obtained if a US output equation is included in the model (section (b), Table 5). It must be stressed that, given the short sample and the huge number of parameters to be estimated in the VAR, the absence of cointegration cannot imply economic consequences regarding convergence or compatibility of national patterns. In the absence of cointegration, we simply state that levels are not useful to improve ARIMA model specifications.

Table 6

RESTRICTED ΔVAR FOR GERMANY, FRANCE, ITALY AND SPAIN
(1987.1-1997.12)

	Germany		France		Italy		Spain	
Specification (1)	Estimate	t	Estimate	t	Estimate	t	Estimate	t
DLIPDE_1	-1.074	-18.2	-0.181	-3.6	-0.372	-4.9	-0.170	-2.1
DLIPDE_2	-0.579	-8.9	-0.171	-3.5	-0.355	-4.4	-0.209	-2.5
DLIPDE_6					-0.207	-3.7		
DLIPDE_7					-0.240	-3.4		
DLIPDE_10					-0.180	-4.1	-0.264	-6.9
DLIPDE_12			-0.257	-7.5			-0.275	-3.9
DLIPFR_1			-0.440	-6.3				
DLIPFR_2			-0.224	-4.1				
DLIPFR_3					0.547	5.0		
DLIPFR_7					0.212	3.3		
DLIPFR_11	0.243	3.6					-0.313	-4.7
DLIPFR_12	0.725	5.2	0.597	8.7	0.619	7.0	0.572	5.2
DLIPIT_1			-0.082	-2.5	-0.547	-10.6		
DLIPIT_2	-0.356	-5.8			-0.275	-7.9		
DLIPIT_3	-0.490	-5.3			-0.478	-8.7		
DLIPIT_4	-0.246	-3.3	-0.043	-3.0	-0.253	-6.0		
DLIPIT_5							0.111	7.6
DLIPSP_1							-0.665	-9.5
DLIPSP_2	0.285	3.2					-0.287	-4.3
DLIPSP_3	0.487	4.5						
DLIPSP_4	0.254	2.9			0.165	2.8		
DLIPSP_10	0.152	3.7			0.066	3.0	0.165	4.3
DLIPSP_12	-0.323	-4.1						
Diagnostic checks (2)	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
S.E. of reg. (%)	2.81		1.97		3.20		2.86	
LM(12)	1.39	0.19	1.82	0.06	1.76	0.07	2.18	0.02
LM(24)	1.21	0.26	2.05	0.01	2.73	0.01	1.64	0.06
ARCH(12)	0.81	0.64	0.49	0.92	0.45	0.93	0.56	0.87
WHITE	0.37	0.99	0.52	0.99	0.74	0.85	0.69	0.90
JB	1.02	0.60	1.29	0.53	1.81	0.40	9.27	0.01

(1) FIML estimation. Coefficient estimates of seasonals and linear trend are not shown. Label names as in Table 3, with lag _*p*.-(2) Diagnostic tests as in Table 4. System diagnostics test statistics [p-values]: autocorrelation up to 12th order 0.90 [0.77], vector normality 19.3 [0.01], heteroschedasticity 0.79 [0.99]; forecasting ability (1998.1-1999.6) 1.44 [0.06].

We therefore specified a VAR model in first differences and applied a specification search in order to reduce the huge number of parameters (240, i.e.: (4 countries × 12 lags + 12 deterministic variables) × 4 equations). Since the overparameterised VAR was data coherent (on the basis of not reported misspecification tests), we first dropped a number of not significant system regressors (country-lags) in all the equations (100 parameters were restricted to zero; the corresponding test $F(100,276) = 0.84$ has a p-value of 0.84). A further set of restrictions was then imposed by equation (61 parameters set to zero; the likelihood ratio test $\chi^2(61) = 50.5$ has a p-value of 0.83).

As shown in Table 6, all parameter estimates are significant and the standard errors of the regressions are similar to those of the corresponding equations listed in Table 4. Although the number of parameter estimates in the VAR is considerably higher than that in the ARIMA models, the (surprising) similarity between their explanatory powers may depend on the poor specification of the seasonal stochastic components in the VAR due to the low number of degrees of freedom. A signal of possible, relevant seasonal factor omission is shown by 24th order LM residual autocorrelation tests in Table 6, sometimes 5 per cent significant.

Table 7

GRANGER CAUSALITY TESTS
IN A VAR WITH US (1)

Causing country:	Statistic
Germany	177.1
France	108.3
Italy	140.8
Spain	111.0
United States	107.0

(1) H_0: the coefficients of the lagged values of the country listed in a row are zero in all other equations (countries). The test statistic is distributed as a χ^2 with 48 degrees of freedom (4 equations $\times$ 12 lags). All tests are largely 1% significant.

The importance of taking into account the relationships between the US and the Euro-area industrial production indices may be analysed through Granger causality tests. As the levels are not cointegrated, we used the first differences VAR(12) model. From the results in Table 7, it appears that each block of countries Granger-causes the output of the other four countries.

Notwithstanding the results of Table 7 the inclusion of the US output does not improve the forecasting ability of the multi-country VAR (results not reported). Again, the main problem seems to lie in the high number of parameters to be estimated (360 parameters, including deterministic explanatory variables). Thus benchmark ARIMA models have some forecasting advantages when compared with the VAR approach because they are more parsimonious in specifying seasonal effects and avoid the problem of overparameterisation. On the other hand, Granger causality tests show that dynamic interrelations would matter, if we were able to handle them in a more parsimonious parametric space.

3.2 Modelling the Euro area, the United States and the business confidence index

In order to forecast the industrial production index of the Euro area as a whole, a parsimonious way to give account of the dynamic interrelations with US output is a bivariate VAR framework that ignores single European countries. 24 lags are used to explain seasonal dynamics and to proxy for omitted short-term cyclical indicators.

Table 8

MISSPECIFICATION AND COINTEGRATION TESTS
FOR THE BIVARIATE VAR (EURO AREA AND UNITED STATES)
(1987.1-1997.12)

	Euro Area		United States	
UVAR diagnostic checks (1)	Statistic	p-value	Statistic	p-value
Adjusted R-squared	0.988		0.995	
S.E. of regression (%)	1.06		0.62	
Durbin-Watson	2.07		2.00	
LM(12)	2.38	0.01	1.79	0.07
LM(24)	1.74	0.05	1.84	0.04
ARCH(12)	0.46	0.93	0.34	0.98
WHITE	0.03	0.86	2.56	0.11
JB	4.84	0.09	6.54	0.04
RESET(1)	0.16	0.69	3.64	0.06
Cointegration H_0: rank = p	Max eigenv. Test	c.v. (2)	Trace test	c.v. (2)
p == 0	23.41	19.0	30.46	25.3
p <= 1	7.05	12.3	7.05	12.3

(1) Diagnostic tests as in Table 4. UVAR parameter constancy forecast test: F(24,71) is equal to 1.34 [p-value 0.17] in the more restrictive case, see Doornik and Hendry (1997).-(2) Osterwald-Lenum (1992) 5% critical values.

To compute the Johansen cointegration tests between the Euro area and the US indices we consider a VAR(24) model with intercept, seasonals and a trend constrained to lie in the cointegration space. The first section of Table 8 is devoted to analysis of the data congruency of the statistical model (UVAR, unrestricted VAR). Overall, the diagnostic tests of both equations hardly ever reject the null at 5 per cent significance (and never at 1 per cent; the poor result for the LM(12) in the euro area equation seems to depend on the specific sample used: in fact it disappears by simply rolling the sample one year ahead). The standard error of the euro area equation is lower than the best ARIMA model in Table 4, suggesting that useful information has been included in the bivariate VAR(24). All the roots of the companion matrix lying inside the unit circle ensure the stability of the system.

Both Johansen cointegration tests (maximum eigenvalue and trace), displayed in Table 8, reject the null of cointegration rank being zero, and fail to reject the reduced rank hypothesis. The cointegration results do not change if we model trend in alternative ways: VARs with unrestricted trend or without trend are still cointegrated.

The cointegration vector parameters β_i (i = US and UE11, the Euro area) and the loading parameters α_i estimates are reported in the first row of Table 9 case *(1)*, along with some additional results related to restricted cases of interest for further interpretation and assessment. More specifically, the results of Table 9 are separated into different blocks. In block *(2)* the hypotheses that one in two variables is trend stationary, resulting in a system with rank one but without cointegration between the euro area and the US, are both strongly rejected on the basis of the test statistics. In block *(3)*, tests of weak exogeneity for each variable are reported: in particular, the result *(3b)* suggests that the US is weakly exogenous. This result still holds even when the model does not contain a trend (case *4c*).

If the US production is strongly exogenous, the euro area can be forecast by using only the correspondent equation of the bivariate VAR. The strong exogeneity hypothesis for the US is tested by imposing p-1=23 zero restrictions to (all) the differenced Euro area output lag parameters in the US output equation of the cointegrated VAR - specified in error correction form - and a zero restriction to the (single) loading parameter of the US output equation. In this way, under the null of strong exogeneity, the US output equation is restricted to an ARIMA(23,1) univariate model.

Table 9

PARAMETER ESTIMATES AND TEST OF ALTERNATIVE HYPOTHESES ON THE PARAMETRIC SPACE (1)

Alternative cases (2)	β_{UE11}	β_{US}	trend	α_{UE11}	α_{US}	χ^2	d
1. Only normalisation choice	-1 (-)	0.848 (0.251)	-0.0006 (0.0005)	0.242 (0.065)	0.068 (0.039)	[-] [-]	0
2a. UE11 output is trend stationary (*)	-1 (-)	0 (-)	0.0014 (0.0002)	0.122 (0.078)	0.096 (0.043)	13.9 [0.0]	1
2b. US output is trend stationary (*)	0 (-)	-1 (-)	0.0025 (0.0002)	-0.118 (0.053)	0.010 (0.031)	11.0 [0.1]	1
3a Weak exogeneity of UE11 output (*)	-1 (-)	-0.118 (0.532)	0.0019 (0.0012)	0 (-)	0.061 (0.030)	16.2 [0.0]	1
3b Weak exogeneity of US output	-1 (-)	1.155 (0.354)	-0.0014 (.0008)	0.157 (0.046)	0 (-)	3.8 [0.05]	1
4a. No trend in cointegrating vector	-1 (-)	0.589 (0.046)	0 (-)	0.267 (0.076)	0.096 (0.045)	1.5 [0.23]	1
4b = 4a + 3a. No trend and UE11 weak exogeneity (*)	-1 (-)	0.756 (0.162)	0 (-)	0 (-)	0.043 (0.031)	20.2 [0.0]	2
4c = 4a + 3b. No trend and US weak exogeneity	-1 (-)	0.552 (0.058)	0 (-)	0.203 (0.068)	0 (-)	8.8 [0.01]	2
5 = 4c + US not Granger causes UE11 = Strong exogeneity of US output	-1 (-)	0.552 (0.058)	0 (-)	0.203 (0.063)	0 (-)	34.4 [0.08]	24

(1) UE11 (Euro area), US (United States). To facilitate interpretation, we inverted signs to parameter estimates (standard errors in parentheses). d is the number of overidentifying restrictions (degrees of freedom), zero means exact identification [p-values in squared brackets].-(2) Under the null. An asterisk shows that the null is 1% significant.

The result of this test, reported in Table 9, case *(5)*, shows that the strong exogeneity hypothesis is not rejected at 5 per cent significance level; the overall p-value (7.8 per cent) is composed of 1.2 per cent p-value (weak exogeneity hypothesis), and 21.5 per cent p-value (non-Granger causality hypothesis). The pretest bias that might affect the latter inferences is avoided by testing non-Granger causality in a VAR framework without cointegration restrictions (see Toda and Yamamoto, 1995). In this context, the hypothesis that the US index does not cause in the Granger sense that of the euro area is strongly rejected ($\chi^2(24) = 81.2$); the opposite case is not rejected at the 5 per cent significance level ($\chi^2(24) = 36.8$).

Table 10

CONDITIONAL ERROR CORRECTION MODEL
FOR EURO AREA INDUSTRIAL PRODUCTION
(1987.1-1997.12)

Specification (1)	Estimate	t		Estimate	t
DLIPUE11_1	-0.213	-2.5	DLIPUS	0.695	4.4
DLIPUE11_12	0.298	3.9	DLIPUS_1	-0.540	-3.0
DLIPUE11_13	0.181	2.3	ECM_1	-0.073	-2.6
Diagnostic checks (2)	Statistic	p-value		Statistic	p-value
Adjusted R^2	0.991		S.E. of regress. (%)	1.10	
Durbin-Watson	2.12				
LM(12)	0.68	0.77	LM(24)	1.08	0.39
ARCH(12)	0.57	0.86	WHITE	0.92	0.57
JB	2.66	0.26	RESET(2)	1.90	0.16
CHOW	0.80	0.65			

(1) Label names as in Table 6. $ECM = LIPUE11 - 0.552 \times LIPUS$. Coefficient estimates of the seasonal dummy variables and the linear trend are not shown.-(2) Diagnostic tests as in Table 4.

Table 10 reports the retained forecasting model of the Euro area production index. Given the previous results, we started from a uniequational (overparameterised) error correction mechanism (both autoregressive and distributed lags of the 24th order) conditional on the US industrial production index. After assessing the data congruency of the general model using mis-specification tests, the number of parameters to be estimated was progressively reduced from 60 to 18. The test of the zero restrictions imposed on the starting model is not rejected at 1% significance level: F(42,72) = 1.62 [0.04].

The short-run fluctuations of the euro area output index are explained by the same dynamics as in the ARIMA model in Table 4, by simultaneous and lagged effects of the US output in differences. An important extension of the ARIMA model is the inclusion of the variables in levels through the error correction term. Globally, the conditional model presents three additional (and highly significant) parameters with respect to the ARIMA benchmark.

The estimated long-run relationship between the levels of the euro area and US output (0.552) is about the same as the sample elasticity from Table 1 (0.593). We then extend the previous information set by adding the European Commission business confidence index, BCI (details about the definition of this variable are given in the

Appendix Data). As expected, the inclusion in the VAR of a cyclical variable makes it possible to reduce the number of lags in the model: the parameter restrictions from 24 to 14 lags are not rejected by the statistic F(90,135) = 1.15 (p-value: 0.23). The resulting VAR(14), with the same deterministic nucleus as in the previous bivariate model, is data coherent, as supported by a number of diagnostic tests not reported. The Johansen cointegration tests corroborate the presence of a single long-run relationship which, after a number of restrictions, becomes (standard errors in brackets):

$$LIPUE11 = LIPUS + \underset{(0.005)}{0.0204} * BCI. \tag{1}$$

Thanks to the positive contribution of the business confidence index, the long-run elasticity of the euro area output index to that of the US increases and is not significantly different from one. In equilibrium, the BCI is constant and exerts no influence on the euro-US output long-run interrelations.

Both the US output and the BCI are weakly exogenous to the previous cointegrating relationship; the rank and parameter restrictions are not rejected ($\chi^2(4) = 8.97$, p-value: 0.06), and we can model the euro area output with the uniequational error correction mechanism reported in Table 11. Similar results are obtained by specifying the BCI in logs after suitable transformation. In this case the four overidentifying restrictions are even more acceptable (p-value: 0.23).

With respect to the model without BCI in Table 10, the estimates of short-run effects of the US output on the euro area are about the same; the speed of adjustment to long-run is also similar, but much more significant. The importance of the inclusion of the BCI variable is evidenced by improved explanatory power (the standard error of the regression is lower than 1 per cent). The non-Granger causality of the US output and the BCI in relation to the euro area output is 5 per cent significant according to both the restricted cointegrated VAR and the Toda and Yamamoto approach (results not reported). The inclusion of the BCI in the information set determines a change in the Granger-causality characteristics of the system since both the euro and US production performances evidently influence it (Lutkepohl, 1982, shows that Granger causality in a bivariate system may be due to the omission of relevant variables). In one-step-ahead forecasting exercises, both models (with and without BCI) can be used safely, while only the model without the BCI is appropriate for multi-step forecasts.

Table 11

CONDITIONAL ERROR CORRECTION MODEL FOR THE EURO AREA WITH THE BCI
(1987.1-1997.12)

Specification (1)	Estimate	t		Estimate	t
LIPUE11_1 –			DLIPUS	0.612	4.3
LIPUE11_12	-0.442	-7.3	DLIPUS_1	-0.529	-3.7
BCI_1 – BCI_3	0.0018	4.9	ECM_1	-0.064	-7.2
Diagnostic checks (2)	Statistic	p-value		Statistic	p-value
Adjusted R^2	0.993		S.E. of regress. (%)	0.97	
Durbin-Watson	2.19				
LM(12)	0.57	0.86	LM(24)	1.12	0.34
ARCH(12)	1.20	0.30	WHITE	0.91	0.58
JB	1.66	0.44	RESET(2)	0.05	0.95
CHOW	0.89	0.56			

(1) Label names as in Table 6. *ECM = LIPUE11 – LIPUS – 0.020375* $\times$ *BCI.* Coefficient estimates of the seasonal dummy variables and the linear trend are not shown.-(2) Diagnostic tests as in Table 4.

4. A disaggregated approach

4.1 Models for the individual Euro area countries

As shown in paragraph 3, the multivariate analysis of individual European countries with the US cannot be exploited, given the insufficient number of degrees of freedom. This leaves open the question of the validity of a more disaggregated approach to forecasting the euro index. The basic hypothesis is that by exploiting country-specific indicators a better forecast of the aggregate index may be obtained. In this paragraph we present forecasting models for Germany, France, Italy and Spain covering up to 80 per cent of total euro output. For Italy we employ the model estimated by Marchetti and Parigi (1999), based on the relationship between manufacturing activity and the consumption of electricity. For Germany, France and Spain general indicators of this kind are not available and we use the US index and some variables derived from the EU Commission's harmonised qualitative survey. In particular, we consider not only the BCI but also its individual components as well, i.e. the order book level, the stock of finished products and the trend in production. Moreover, for one specific country the evolution of manufacturing activity in the other countries of the euro-area has been tried as a regressor (for instance the indices of France, Spain and Italy are used as explicative variables for the German model). The econometric methodology is based on the general to specific approach, according to which all variables with their lags, including the dependent one, are initially used along with a trend and seasonals:

$$(2)\quad y_{k,t} = \alpha + \Theta(L) y_{k,t} + \sum_{j=1}^{4} \mathrm{B}_j(L) lip_{j,t} + \sum_{i=1}^{m} \Phi_i(L) qual_{i,t} + \sum_{i=1}^{3} \delta_i trend^i + seasonals + u_t$$

y_k is the logarithm of the k-th country's industrial production index (k=1 … 3 for Germany, France and Spain); lip_j (j = 1,…,4) represents the logarithm of the US, Italian and two other indices of European countries used as regressors (i.e. if y_k is the German index, lip_j is given by the indices of France, Spain, Italy and the US); *qual* represents the i-th indicators from the EU Commission's survey for the k-th country (actually, in some specific cases the indicators for other countries are used as well; see the model for Spain below); *trend* and *seasonals* are the trend (initially specified as a third degree polynomial) and dummy variables for seasonal effects; $\Theta(L)$, $B_j(L)$ and $\Phi_i(L)$ are polynomials in the lag operator L (we started from a maximum lag of 12 for the lagged dependent variable and for the US index; for the other regressors the maximum lag was lower in order to preserve an adequate number of degrees of freedom). The estimation period for equation (2) is 1987.1-1997.12 (1988.1-1997.12 for Spain because the results of the EU Commission's survey are available only from 1986); 18 months are left to check for the out-of-sample forecasting performance of the models.

a) Germany

The performance of the model for Germany is good, albeit not entirely satisfactory (Table 12): the standard error of the equation is above 2 per cent (2.31%), albeit lower than both the ARIMA model in Table 4 and the VAR in Table 6.

Table 12

FORECASTING MODEL FOR GERMANY
ESTIMATION AND DIAGNOSTICS
(1987.1-1997.12)

Specification (1)	Estimate	t
y_{t-1}	0.843	13.731
ADL(1,3) Δy_{t-1}		
Δy_{t-1}	-0.967	-11.804
Δy_{t-2}	-0.532	-8.126
Δy_{t-3}	-0.097	-1.475
Σ	-1.596	-8.126
$lipus_{t-1}$	0.083	2.251
$M(\Delta lipus,12)_{t-1}$	-0.453	-3.520
$M(\Delta dbci,2)_{t-2}$	0.184	2.592
Δ(lipit,2)	0.327	7.426
Δ(lipfr,3)	0.350	3.966

$\overline{R}^2$ = 0.92
S.E.(%) = 2.31

Misspecification tests
(p-value in parentheses) (2)

Autocorrelation			Unit root test on residuals			Heteroschedasticity		
DW	1.99		ADF	-9.17		$ARCH_{1-12}$	21.79	(0.04)
LM_{1-12}	1.40	(0.18)						
LB_{12}	12.73	(0.39)						
			General specification			Predictive power		
			RESET	8.34	(0.00)	CHOW	0.633	(0.87)

(1) The regression includes seasonals. M(x,z) is the z-periods moving average for x. ADL(n,m) stands for Almon polynomial distributed lag of degree n and length m. S.E., regression standard error.-(2) DW, Durbin-Watson statistic; LM_{1-12}, Lagrange multiplier test for residual autocorrelation of order 1 through 12, F(12,100); LB, Lijung-Box test for residual autocorrelation up to the 12th lag, $\chi^2(12)$; CHOW, Chow test of predictive power over the period 1998.1-1999.6, F(18,130); RESET, test of functional form F(2,110); $ARCH_{1-12}$, autoregressive conditional heteroschedasticity test up to the 12th lag, $\chi^2(12)$; ADF, augmented Dickey-Fuller test (1% critical value: -5.8).

The German confidence index (*dbci*) and the production indices of France and Italy (*lipfr* and *lipit*) play a significant role in the specification. This is a fairly common result to models of countries other than Italy, where the consumption of electricity may obscure the role of confidence and other indices. The complex dynamics are presumably a consequence of the lack of a more suitable set of indicators: were an indicator such as the consumption of electricity available for Germany - as it is for Italy - better results could be obtained with a more parsimonious specification. However, the set of diagnostic checks confirms the validity of the estimated model.

b) France

The results for France are definitely better than those for Germany. The standard error is substantially lower (equal to 1.74%, see Table 13) than corresponding statistical models. As for Germany, production conditions in Italy and Germany play a specific role for French activity.

Table 13

FORECASTING MODEL FOR FRANCE
ESTIMATION AND DIAGNOSTICS
(1987.1-1997.12)

Specification (1)	Estimate	t
Y_{t-3}	0.387	5.014
Y_{t-12}	0.423	5.565
$Lipus_{t-1}$	0.097	3.246
$M(\Delta lipus,10)_{t-1}$	-0.277	-2.339
Δobf_{t-6}	0.121	2.465
$Fbci_{t-1}$	0.724	4.229
$M(\Delta fbci,2)_{t-3}$	0.128	2.990
$\Delta(ibci+dbci)_{t-8}$	0.154	2.824
$\Delta(lipes)_{t-2}$	0.142	3.594

$\overline{R}^2$ = 0.97
S.E.(%) = 1.74

Misspecification tests
(p-value in parentheses) (2)

Autocorrelation			Unit root test on residuals			Heteroschedasticity		
DW	1.92		ADF	-8.8		$ARCH_{1-12}$	11.69	(0.47)
LM_{1-12}	0.93	(0.52)						
LB_{12}	11.62	(0.48)						
			General specification			Predictive power		
			RESET	1.24	(0.29)	CHOW	0.64	(0.87)

(1) The regression includes seasonals. M(x,z) is the z-periods moving average for x. ADL(n,m) stands for Almon polynomial distributed lag of degree n and length m. S.E., regression standard error.-(2) DW, Durbin-Watson statistic; LM_{1-12}, Lagrange multiplier test for residual autocorrelation of order 1 through 12, F(12,100); LB, Lijung-Box test for residual autocorrelation up to the 12th lag, $\chi^2(12)$; CHOW, Chow test of predictive power over the period 1998.1-1999.6, F(18,130); RESET, test of functional form F(2,110); $ARCH_{1-12}$, autoregressive conditional heteroschedasticity test up to the 12th lag, $\chi^2(12)$; ADF, augmented Dickey-Fuller test (1% critical value: -5.8).

In the French case another qualitative variable, the order book level (*obf*), enters the specification, meaning that the confidence index is unlikely to be sufficient to capture the attitude of firms. Misspecification tests are generally good, the ARCH test for heteroschedasticity being influenced by some outliers. The model is stable, as shown by the Chow test on the predictive power of the equation.

c) Italy

The model for Italy is based on a richer set of information than those for other countries.

Table 14

FORECASTING MODEL FOR ITALY
ESTIMATION AND DIAGNOSTICS
(1987.1-1997.12)

Specification (1)	Estimate	t
$M(y,4)_{t-1}$	0.416	8.622
$Elco_t$	0.008	14.987
$Temp_t$	1.290	7.489
$Temp^2_t$	-0.051	-7.987
Trend	31.50	-4.462
$Trend^2$	0.183	2.294
$Trend^3$	-0.001	-1.528

$\overline{R}^2$ = 0.99
S.E.(%) = 1.11

Misspecification tests
(p-value in parentheses) (2)

Autocorrelation			Unit root test on residuals			Heteroschedasticity		
DW	1.87		ADF	-7.30		$ARCH_{1-12}$	14.36	(0.28)
LM_{1-12}	0.96	(0.50)						
LB_{12}	13.54	(0.33)						
			General specification			Predictive power		
			RESET	3.76	(0.03)	CHOW	1.12	(0.34)

(1) The regression includes seasonals. M(x,z) is the z-periods moving average for x. ADL(n,m) stands for Almon polynomial distributed lag of degree n and length m. S.E., regression standard error.-(2) DW, Durbin-Watson statistic; LM_{1-12}, Lagrange multiplier test for residual autocorrelation of order 1 through 12, F(12,100); LB, Lijung-Box test for residual autocorrelation up to the 12th lag, $\chi^2(12)$; CHOW, Chow test of predictive power over the period 1998.1-1999.6, F(18,130); RESET, test of functional form F(2,110); $ARCH_{1-12}$, autoregressive conditional heteroschedasticity test up to the 12th lag, $\chi^2(12)$; ADF, augmented Dickey-Fuller test (1% critical value: -5.8).

A complete description of the model, with its statistical properties, can be found in Marchetti and Parigi (2000). In brief, the main driving force of the industrial production index, adjusted for working days, is the consumption of electricity, with meteorological data to take account of non-industrial use of electricity. Other indicators, such as those from qualitative surveys and the US index, play no relevant role in the specification.

d) Spain

The model for Spain is similar to those of Germany and France, with the exception that along with the confidence index for Germany, France and Italy, the aggregate production index of these three countries (*ipue3*, see paragraph 4.2 below for more details regarding its computation) also plays a significant role in the model. The performance of the equation however is not completely satisfying, given the fairly high value of the standard error (2.54%). The p-value of the Chow test is lower than for the other countries, and it signals possible parameter instability.

Table 15

FORECASTING MODEL FOR SPAIN
ESTIMATION AND DIAGNOSTICS
(1988.1-1997.12)

Specification ([1])	Estimate	t
$M(y,5)_{t-2}$	0.600	5.775
y_{t-12}	0.230	2.979
$Lipus_{t-1}$	0.388	3.469
$M(\Delta lipus,12)_{t-1}$	-0.530	-3.011
$M(\Delta ebci,3)_{t-1}$	0.309	5.063
$M(\Delta(ibci+fbci+dbci),4)_{t-6}$	0.084	3.920
$\Delta lipue3_{t-3}$	0.231	4.054
Trend	-0.057	-2.818

$\overline{R}^2$ = 0.97
S.E.(%) = 2.54

Misspecification tests
(p-value in parentheses) ([2])

Autocorrelation			Unit root test on residuals			Heteroschedasticity		
DW	2.32		ADF	-8.4		$ARCH_{1-12}$	10.4	(0.59)
LM_{1-12}	0.82	(0.63)						
LB_{12}	8.39	(0.75)						
			General specification			Predictive power		
			RESET	2.35	(0.10)	CHOW	1.270	(0.22)

([1]) The regression includes seasonals. M(x,z) is the z-periods moving average for x. ADL(n,m) stands for Almon polynomial distributed lag of degree n and length m. S.E., regression standard error.-([2]) DW, Durbin-Watson statistic; LM_{1-12}, Lagrange multiplier test for residual autocorrelation of order 1 through 12, F(12,100); LB, Lijung-Box test for residual autocorrelation up to the 12^{th} lag, $\chi^2(12)$; CHOW, Chow test of predictive power over the period 1998.1-1999.6, F(18,130); RESET, test of functional form F(2,110); $ARCH_{1-12}$, autoregressive conditional heteroschedasticity test up to the 12th lag, $\chi^2(12)$; ADF, augmented Dickey-Fuller test (1% critical value: -5.8).

4.2 Forecasting performance

To assess the forecasting capacity of the models for Germany, France, Italy and Spain we computed one-step-ahead *ex ante* forecasts by estimating rolling regressions for each month in the period 1997.1-1999.6, with a fixed window of 8 years (see Table 16; the choice of considering the data for 1997 is motivated by the need to check for the forecasting ability over a longer period of time). For comparison purposes this same technique was also applied to the ARIMA models.

Tab. 16

ANALYSIS OF ONE-STEP-AHEAD FORECASTING PERFORMANCE ([1])
(1997.1-1999.6)

Models	RMSE	ME	MAE	HAE	Fraction of RMSE due to: bias	slope ≠ 1	Res.variance
Individual Euro area countries							
Germany ARIMA, Tab. 4	3.52	0.6	2.95	9.37	0.02	0.02	0.96
Germany model, Table 12	2.91	-0.1	2.61	6.72	0.00	0.01	0.99
France ARIMA, Table 4	2.00	0.5	1.71	6.93	0.06	0.06	0.88
France model, Table 13	1.79	-0.3	1.58	4.14	0.02	0.11	0.87
Italy ARIMA, Table 4	4.02	-0.2	2.76	7.81	0.00	0.31	0.69
Italy model, Table 14	1.18	0.1	1.03	2.65	0.00	0.00	1.00
Spain ARIMA, Table 4	3.51	0.6	2.79	14.05	0.03	0.18	0.79
Spain model, Table 15	3.50	-1.0	3.08	6.89	0.06	0.14	0.80
Euro area							
ARIMA, Table 4	1.29	0.2	1.13	3.75	0.03	0.01	0.96
Indic (Table 11)	0.93	0.1	0.8	2.29	0.00	0.00	1.00
Forecasts based on the production indices of Germany, France and Italy ([2])							
Sub3	1.44	0.4	1.29	3.06	0.07	0.02	0.91
Csmod3	1.46	0.3	1.29	3.36	0.04	0.03	0.93
Forecasts based on the production indices of Germany, France, Italy and Spain							
Sub4	1.41	0.3	1.26	3.09	0.05	0.02	0.93
Csmod4	1.48	-0.1	1.26	3.22	0.00	0.04	0.96

([1]) RMSE root mean square error; ME mean error; MAE mean absolute error; HAE highest absolute error. All statistics are in % values.-([2]) *Sub3* and *Sub4* stand for the forecasts obtained by aggregating the forecasts of individual countries; *Csmod3* and *Csmod4* stand for the forecasts obtained by using single-country forecasts separately.

Although the predicted values track fairly closely the actual ones in some countries, in others the results are not completely satisfactory. While for Italy and France the performance of the models is good, the same cannot be said for Germany or, above all, for Spain. In the first two countries the RMSE is below 2 per cent, while for Germany and Spain it is about 3 per cent. In all cases, there are no signs of systematic bias and the size of the errors is not large (with the possible exception of Spain). Overall, the analysis of the forecasting performance suggests that the indicators play an important role: according to the test statistics presented in Table 16, all the models clearly outperform their corresponding ARIMA benchmarks (again, with the exception of Spain).

The next step in our analysis is the forecast of the euro index. In this context, there are two alternative strategies: a) to aggregate the single country forecasts, using the same weights as for the euro index, so as to obtain a sort of sub-aggregate indicator; b) to link the forecasts of the single countries to the aggregate index in a sort of "bridge model" (see Parigi and Schlitzer, 1995, for a description of bridge models). In both cases we may consider only three countries, Germany, France and Italy, discarding Spain, given its lower weight in the aggregate index and the somewhat disappointing results shown above.

According to the first strategy, two sub-aggregate indicators may be computed with the weights of the value added at factor costs expressed in ecu 1995, the same data used by Eurostat to compute the aggregate production series (more specifically, the weights are 0.38 for Germany, 0.19 for France, and 0.16 for Italy and 0.08 for Spain): *ipue3* for the first three countries, covering up to about 73 per cent of the output of the whole area; *ipue4* for the four countries (81 per cent of the total index). These variables, *ipue3* or *ipue4*, are then related to the euro index through a simple, general, regression of the form:

$$(3) \qquad y_t = \alpha + \Theta(L) + \mathrm{B}(L) ipue(i)_t + seasonals, \ (\mathrm{i} = 3,4)$$

which has been simplified according to the usual general-to-specific procedure giving the following two specifications for the three- and four-country cases:

(4)
$$y_t = -0.10 + 0.741 * y_{t-1} + 0.268 * ipue3 + seasonals + u_t$$
$$\quad (-0.99) \; (19.09) \qquad\quad (6.81)$$

and

(5)
$$y_t = -0.11 + 0.719 * y_{t-1} + 0.292 * ipue4 + seasonals + u_t .$$
$$\quad (-1.09) \; (18.39) \qquad\quad (7.34)$$

The two estimates are very similar: in both cases the regression standard errors are close to 1.2 per cent, and the usual misspecification tests do not signal any problem. The joint restriction that the sum of the coefficients of the lagged endogenous variable and of *ipue(i)* be equal to 1 and that the intercept be zero is easily accepted but we have decided not to impose it, given our main interest on the forecasting performance of the models.

The one-step-ahead forecasts are computed by aggregating the one-step-ahead forecasts for the national industrial production indices and using models (4) and (5). Overall, the results are good (see Table 16): the RMSE is close to 1.4 per cent (direct use of the actual values for the sub-aggregate indices only slightly improves the RMSE to around 1.2; this may be due to the fact that the index for the aggregate area is adjusted for the effect of working days, while this is not true for the single-country indices).

The second forecasting strategy consists of using the national indices directly to predict the euro index - i. e. without computing the sub-aggregate indicators. In this case the «bridge» model between the aggregate and the country-specific indices is a simple regression along with the usual dummy variables and some dynamics (estimates are not reported and are available from the authors). In terms of forecasting performance the results are similar to those of the previous case, meaning that the use of sub-aggregate or country-specific indices does not change things significantly (according to the disaggregated approach, when using the actual values of the variables matters clearly improve; however, this procedure is precluded by the fact that the delay with which the single countries and the euro aggregate indices are released is basically the same).

In general, the two strategies provide similar results, with little difference between the three- and four-country cases. This means that in order to forecast the aggregate index the data for all countries are not necessary; in fact, the numbers in the Table show that a reasonably good forecast may be obtained with a very small set of countries. Clearly, should these same countries succeed in publishing their indices more timely, the forecasting performance could be improved by using their actual values.

5. Comparison of forecasts

The aim of this paragraph is to compare the various series of predicted values in order to establish a sort of hierarchy among the models. Initially, the equality between the RMSE's of the different models is tested with the test statistics proposed by Diebold and Mariano (1995; DM hereafter), adapted to small samples by Harvey et al. (1997). As this check is insufficient to establish the improved validity of one forecast over another we also compute a set of encompassing forecasting tests, following the procedures proposed by Harvey et al. (1998; EDM hereafter) and by Fair and Shiller (1990; FS hereafter).

Table 17 shows the results of the DM test for the forecasts produced by the model with indicators, (Indic, see paragraph 3); the ARIMA model (see paragraph 2), the models based on sub-aggregate indicators (Sub3 and Sub4 for the three- and four-country cases, respectively) and the models based on country-specific indices (Csmod3 and Csmod4).

Overall, it appears that the RMSE of the Indic model is the lowest and that it is significantly different from the RMSE of all the other models. The superiority of the Indic model is reinforced by the encompassing analysis according to both the EDM and FS tests (when we consider the forecasts obtained using the actual values of the country-specific indices the picture does not change substantially). Another interesting result is the very good performance of the aggregate ARIMA model, which outperforms the Sub3, Sub4, Csmod3 and Csmod4 models (this result disappears when the aggregate forecasts are computed using the actual values of the country indices).

The results of the comparison show that for one-step-ahead forecasts the aggregate model based on indicators is clearly the best, followed surprisingly by the ARIMA model[6].

---

[6] A forecasting model for the aggregate euro area has been recently estimated by Amisano *et al*. (1999).

Table 17

COMPARISONS OF ONE-STEP-AHEAD FORECASTING PERFORMANCE
(1997.1-1999.6)

	ARIMA	Indic	Sub3	Csmod3	Sub4	Csmod4
			Modified Diebold-Mariano test ([1])			
ARIMA	-	-1.638	2.763	2.386	2.202	1.187
Indic		-	-2.422	-2.698	-2.400	-3.048
Sub3			-	-0.605	1.532	-0.330
Csmod3				-	1.306	-0.186
Sub4					-	-0.534
Csmod4						-
			Modified Diebold-Mariano encompassing test ([1])			
ARIMA	-	2.484	-0.572	-0.085	-0.180	0.267
Indic	-0.197	-	-1.007	-0.802	-0.974	-1.129
Sub3	3.857	2.948	-	1.050	1.728	1.804
Csmod3	4.403	3.518	1.631	-	2.344	0.716
Sub4	3.708	2.917	-1.333	0.165	-	1.188
Csmod4	2.071	3.390	1.134	0.859	1.198	-
			Fair-Shiller test ([2])			
ARIMA v Indic	-0.328	4.421				
ARIMA v Sub3	1.773		0.172			
ARIMA v Csmod3	2.594			0.155		
ARIMA v Sub4	1.698				0.286	
ARIMA v Csmod4	3.430					-0.064
Indic v Sub3		6.784	-1.342			
Indic v Csmod3		6.449		-1.002		
Indic v Sub4		6.703			-1.265	
Indic v Csmod4		6.068				-1.175
Sub3 v Csmod3			2.032	-0.169		
Sub3 v Sub4			-0.491		0.746	
Sub3 v Csmod4			2.728			-0.275
Csmod3 v Sub4				-0.372	2.085	
Csmod3 v Csmod4				1.659		-0.406
Sub4 v Csmod4					2.809	-0.563

([1]) In small samples the test statistics are distributed as a t with n-1 degrees of freedom, where n is the number of forecasts (see Harvey *et al.*, 1997 and 1998). For the DM test, the null hypothesis is the equality of the RMSE's of the two models; for the EDM test, the null hypothesis is that the forecasts of the model on the row encompass the forecasts of the model on the column. -([2]) The table shows White-consistent t-values of the β_1 and β_2 estimates in: $\frac{(y_t - y_{t-12})}{y_{t-12}} = \alpha + \beta_1 * \frac{(\hat{y}_{1,t} - y_{t-12})}{y_{t-12}} + \beta_2 * \frac{(\hat{y}_{2,t} - y_{t-12})}{y_{t-12}}$ where $\hat{y}_{1,t}$ and $\hat{y}_{2,t}$ are the forecasts of the two models being compared (see Fair and Shiller, 1990).

A possible explanation lies in the fact that the aggregation operation somehow smoothes the dynamics with respect to the evolution of the single components, especially when they are characterised by asymmetric cycles. However, even if the single-country models provide disappointing forecasting performances, they should not be discarded too hastily. In effect, one specific advantage of the disaggregated approach is the possibility of decomposing the forecast for the aggregate area on a geographical basis. In this context, further research should be devoted to finding better specifications for the single-country models, in particular by enriching the information set with a more complete list of indicators[7].

6 Conclusions

In this paper we presented a methodology for forecasting the industrial production index for the euro area as a whole; there is no need to emphasise the importance of having reliable estimates for economic policy reasons. We considered a number of alternative solutions:

1) simple univariate time series models at aggregate or disaggregated level (where aggregation refers to geographic coverage);

2) vector autoregressive models referring to the four largest Euro area countries (France, Italy, Germany and Spain);

3) a two-country VAR (US and Euro area) with and without the inclusion of data on business conditions in Europe as measured by the harmonised European Commission survey;

4) A disaggregated procedure based on national forecasts obtained from single-country models and aggregated to provide the euro index.

The main results of this thorough and complete analysis are that a simple Euro area ARIMA model is able to generate a reliable short-term forecast (1–2 month lead) also

---

[7] A possible improvement of the country-specific models may come from the use of national (rather than harmonised EU) survey results. In this case, raw numbers may be directly employed, with a more articulated set of questions. See Bodo *et al*. (1997) and Gudin and Rauh (1999) for some examples of this approach.

exploiting the fact that the industrial production series are particularly influenced by seasonal factors. The attempt to improve on these results using a more complex and articulated VAR model (restricted to the four largest Euro area countries) failed because of overparameterization problems. However, if we include the US production index in the analysis and proceed to use a two country VAR model (US and Euro area), we obtain a significant improvement over the Arima estimate. The effort to use national models in order to provide a reliable forecast for the area did not lead to satisfactory results, showing that it is probably more effective to concentrate on the aggregate as a whole if this is the focus of the forecasting exercise. A reason for the latter finding is that aggregation solves a number of specification problems in the single-country equations.

Over the sample period 1987–1999, there is some evidence of Granger causality from the US to the Euro area (in a bivariate VAR), but not *vice versa*. However, this result disappears when the business confidence index of European manufacturing firms is considered.

Our findings may seem somewhat implausible if the US and the Euro area are to be considered as two closed areas in terms of trade flows. However it should be recalled that over the period of our analysis we compare two somewhat different situations: one "large" country, the US, and a set of "small" countries (Euro area). This may imply that the weight of the American economy may be high for the individual European economies. Table 18 shows the weights of foreign trade in the real effective exchange rate of the euro area.

The importance of the US for the euro area is evident: notwithstanding the decrease since 1985, in 1995 the global weight of the US (given by the average of the export and import shares) was still fairly high (22.2 per cent)[8]. Actually, this weight should be considered as an underestimate because it does not take into account the trade links between the American economy and other leading partners of the euro area, especially the UK and Japan (which together account for almost 40 per cent of the total euro trade flows).

[8] According to the weights published by the ECB, the value of the US over the period 1995-97 is close to 25 per cent, probably reflecting the economic downturn in Japan and the Southeast Asian region.

Table 18

FOREIGN TRADE WEIGHTS OF THE MAIN PARTNERS OF THE EURO AREA ([1])

Countries	1985			1995		
	Exports	Imports	Global	Exports	Imports	Global
United Kingdom	19.37	26.57	21.87	17.28	27.90	23.21
United States	29.77	23.12	27.46	25.55	19.48	22.17
Japan	18.41	16.12	17.61	18.12	14.85	16.29
Switzerland	8.11	12.21	9.53	7.20	11.17	9.41
Hong Kong SAR ([2])	3.47	2.60	3.17	7.85	5.62	6.61
Sweden	5.04	8.62	6.28	4.01	7.51	5.97
South Korea	2.64	2.19	2.49	5.56	3.03	4.14
Singapore	1.95	0.87	1.57	4.55	2.55	3.44
Denmark	3.07	2.91	3.02	2.86	3.67	3.31
Norway	2.16	2.07	2.13	1.47	1.79	1.65
Canada	2.84	1.20	2.27	2.04	1.07	1.50
Greece	1.53	1.29	1.45	1.50	0.92	1.18
Australia	1.64	0.23	1.15	2.01	0.44	1.12
	100.00	100.00	100.00	100.00	100.00	100.00

([1]) Source: Bank of Italy (1998). The methodology is similar to that employed by the ECB (see ECB, 1999). The main difference concerns the kind of weighting scheme, mobile for the Bank of Italy, fixed for the ECB. The export and import weights are the respective shares in foreign trade. For the export share the method of double weighting is adopted in order to capture third market effects.-([2]) Special Administrative Region.

The process of integration among the countries of the monetary union may have important consequences. As the aggregate values for the euro area are computed as an average of the corresponding country values, asymmetric, national cycles may “cancel out”, resulting in a smooth evolution at the aggregate level which can help the task of the forecaster. However, with the advent of the monetary union the integration process may increase the degree of symmetry among countries and alter the dynamic characteristics of the aggregate cycle, thus eroding the forecasting ability of models based on the past. Moreover, the relationship between the US and the Euro area may weaken as the reduction of the US weight over the period 1985-1995, characterised by the start of the single European market, seems to suggest (see tab. 18). Further research is needed on the nature and possible evolution of the relationship between the US (and other areas of the world) and the Euro area.

The process towards more complete integration should nonetheless be slow and cyclical differences among single countries persist for quite a long time. In this context, the high-frequency forecasting model described in this paper should be only slightly affected, at least in the near future. Clearly, this implies that the statistical properties of the specification should always be monitored and that recursive estimation techniques should be adopted.

Appendix: Data

Industrial production data. These data are released as index numbers of the quantity of output produced in one specific month in the industrial sector, net of the construction component (for the US the year base is 1992=100; for Spain and France, 1990=100; for Germany, Italy and the euro area, 1995=100). For all countries the series are not seasonally adjusted. We have taken into account the reunification of Germany by aggregating data for the former East Germany with that for the former West Germany. The data are released by the official statistical bureau for Germany, France, Italy and Spain and by the Federal Reserve for the USA.

Eurostat computes the aggregate index for the euro area, as a weighted average of the 11 national indices of industrial production adjusted for working days since for some countries the version of the index adjusted for working days is the only one available. The weights are derived from the country shares of the 1995 value added at factor costs expressed in ecu for the manufacturing sector, excluding construction.

The publication lag is very low for the US (3 weeks) and about 2 months for the European countries (more specifically: 8 weeks for the euro area, France and Spain; 6 weeks for Germany and Italy).

European Commission qualitative data. These data are derived from the harmonised survey of the industrial sector carried out by the European Commission on a monthly basis. The data are qualitative and are quantified by taking the balance between the percentage of positive and negative replies to the questions in the survey. The confidence index is an arithmetic average of the replies to the questions concerning the order levels, the expected trend in production and the stocks of finished products. All series are available only on a seasonally adjusted basis. The series for the euro area are obtained by the same aggregation procedure as for that used the industrial production index.

References

Amisano, G., C. Giannini, P. Guida, E. Lizzoli, M. Serati and L. Stanca (1999), "Un modello previsivo trimestrale per l'EU-11", Research Department, Unicredito, November, mimeo.

Bank of Italy (1998), *Economic Bulletin*, February, No. 30.

Bodo, G. E. Giovannini and M. G. Piras (1997), "La previsione a breve termine dell'indice della produzione industriale nei principali paesi europei: un modello con legami internazionali, LUISS, Quaderni di Ricerca, No. 84.

Box, G. E. P. and Jenkins G. M. (1970), *Time Series Analysis: Forecasting and Control*, Holden-Day.

Campbell, J. Y. and P. Perron (1991), "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots", in O. J. Blanchard and S. Fischer (eds.), *NBER Economics Annual 1991*, MIT Press, pp. 141-201.

Clements, M. P. and D. F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge University Press.

Dickey, D. A. and W. A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of American Statistical Association*, Vol. 74, pp. 427-31.

Dickey, D. A. and S. G. Pantula (1987), "Determining the Order of Differencing in Autoregressive processes", *Journal of Business and Economic Statistics*, Vol. 5, pp. 455-61.

Diebold, F. X. and R. S: Mariano (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, Vol. 13, pp. 253-63.

Doornik, J. A. and D. F. Hendry (1997), *Modelling Dynamic Systems Using PcFiml 9.0 for Windows*, International Thomson Business Press.

European Central Bank, (1999), *Monthly Bulletin*, October, pp. 29-31.

Fair, R. C. and R. B. Shiller (1990), "Comparing Information in Forecasts from Econometric Models", *American Economic Review*, Vol. 80, pp. 255-68.

Franses, P. H. (1991), "Seasonality, non-Seasonality and the Forecasting of Monthly Time Series", *International Journal of Forecasting*, Vol. 7, pp. 199-208.

Franses, P. H. (1996), *Periodicity and Stochastic Trends in Economic Time Series,* Oxford University Press.

Franses, P. H. (1998), *Time Series Models for Business and Economic Forecasting*, Cambridge University Press.

Franses, P. H. and B. Hobijn (1997), "Critical Values for Unit Root Tests in Seasonal Time Series", *Journal of Applies Statistics*, Vol. 24, No. 1, pp. 25-47.

Franses, P. H. and F. Kleibergen (1996), "Unit Roots in the Nelson-Plosser Data: Do They Matter for Forecasting?", *International Journal of Forecasting*, Vol. 12, pp. 283-88.

Franses, P. H. and A. B. Koeller (1998), "A Model Selection Strategy for Time Series with Increasing Seasonal Variation", *International Journal of Forecasting*, Vol. 14, pp. 405-14.

Gudin, P. and J. Rauh (1999), "The Industrial Recovery Is Well Under Way", *European Weekly Analyst*, Goldman Sachs, No. 99/37, pp.4-6.

Harvey, D. I., S. J. Leybourne and P. Newbold (1997), "Testing the Equality of Prediction Mean Squared Errors", *International Journal of Forecasting*, Vol. 13, pp. 281-91.

Harvey, D. I., S. J. Leybourne and P. Newbold (1998), "Tests for Forecast Encompassing", *Journal of Business and Economic Statistics*, Vol. 16, pp. 254-59.

Hylleberg, S., R. F. Engle, C. W. J. Granger and B. S. Yoo (1990), "Seasonal Integration and Cointegration", *Journal of Econometrics*, Vol. 44, No. 1-2, pp. 215-38.

Hylleberg, S., C. Jorgensen and N. K. Sorensen (1993), "Seasonality in Macroeconomic Time Series", *Empirical Economics*, Vol. 18, pp. 321-35.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root", *Journal of Econometrics*, Vol. 54, pp. 159-78.

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.

Lutkepohl, H. (1982), "Non Causality Due to Omitted Variables", *Journal of Econometrics*, Vol. 19, pp. 367-78.

Marchetti, D. and G. Parigi (2000), "Energy Consumption, Survey Data and the Prediction of Industrial Production in Italy: a Comparison and Combination of Different Models", *Journal of Forecasting*, Vol. 19, No. 3.

Miron, J. A. (1996), *The Economics of Seasonal Cycles*, MIT Press.

Osborn, D. R., A. P. L. Chui, J. P. Smith and C. R. Birchenhall (1988), "Seasonality and the Order of Integration for Consumption", *Oxford Bulletin of Economics and Statistics*, Vol. 50, No. 4, pp. 361-77.

Osborn, D. R., S. Heravi and C. R. Birchenhall (1999), "Seasonal Unit Roots and Forecasts of Two-Digit European Industrial Production", *International Journal of Forecasting*, Vol. 15, pp. 27-47.

Osterwald-Lenum, M. (1992), "A Note with Quantiles of the Asymptotic Distribution of the ML Cointegration Rank Test Statistics", *Oxford Bulletin of Economics and Statistics*, Vol. 54, pp. 461-72.

Parigi, G. and G. Schlitzer (1995), "Quarterly Forecasts of the Italian Business Cycle by Means of Monthly Economic Indicator", *Journal of Forecasting*, Vol. 14, pp. 117-41.

Perron, P. (1997), "Further Evidence on Breaking Trend Functions in Macroeconomic Variables", *Journal of Econometrics*, Vol. 80, pp. 355-85.

Stock, J. H. and M. W. Watson (1998), "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series", NBER, *Working Paper*, No. 6607, June.

Toda, H. Y. and T. Yamamoto (1995), "Statistical Inference in Vector Autoregressions with Possibly Integrated Processes", *Journal of Econometrics*, Vol. 66, pp. 225-50.

RECENTLY PUBLISHED "TEMI" (*)

(*) Requests for copies should be sent to:
Banca d'Italia - Servizio Studi - Divisione Biblioteca e pubblicazioni - Via Nazionale, 91 - 00184 Rome (fax 0039 06 47922059). They are available on the Internet at www.bancaditalia.it