DISCUSSION PAPER SERIES

No. 3108

# EUROCOIN: A REAL TIME COINCIDENT INDICATOR OF THE EURO AREA BUSINESS CYCLE

Filippo Altissimo, Antonio Bassanetti,
Riccardo Cristadoro and Mario Forni,
Marco Lippi, Lucrezia Reichlin and
Giovanni Veronese

*INTERNATIONAL MACROECONOMICS*

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP3108.asp

*ISSN 0265-8003*

# EUROCOIN: A REAL TIME COINCIDENT INDICATOR OF THE EURO AREA BUSINESS CYCLE

**Filippo Altissimo,** Banca d'Italia and Harvard University and CEPR
**Antonio Bassanetti,** Banca d'Italia
**Riccardo Cristadoro,** Banca d'Italia
**Mario Forni,** Università di Modena and CEPR
**Marco Lippi,** Università di Roma La Sapienza
**Marc Hallin,** ECARES, Université Libre de Bruxelles
**Lucrezia Reichlin,** ECARES, Université Libre de Bruxelles and CEPR
**Giovanni Veronese,** CENSIS

Discussion Paper No. 3108
December 2001

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 3108

December 2001

# ABSTRACT

## EuroCOIN: A Real Time Coincident Indicator of the Euro Area Business Cycle*

This Paper is the result of the Bank of Italy-CEPR project to construct a monthly coincident indicator of the business cycle of the euro area. The index is estimated on the basis of a harmonized data set of monthly statistics of the euro area (951 series) which we constructed from a variety of sources. We use the information of this large panel to obtain an indicator which has three characteristics: (i) it provides real time information on monthly coincident activity since it is updated as new information becomes available in a non-synchronous way; (ii) it is cleaned from noise originated from measurement error and idiosyncratic national and sectoral dynamics; (iii) it is cleaned from seasonal and short-run dynamics through a filter that requires very little revision at the end of the sample. Unlike other methods used in the literature, the procedure takes into consideration the cross-country as well as the within-country correlation structure and exploits all information on dynamic cross-correlations. As a by-product of our analysis, we provide a characterization of the commonality and dynamic relations of the series in the data set with respect to the coincident indicator and a dating of the euro area cycle.

JEL Classification: C51, E32 and O30
Keywords: business cycle and dynamic factor model

Filippo Altissimo
Kennedy School of Government
Harvard University
79 JFK Street
Cambridge, MA 02138
USA
Email: filippo_altissimo@ksg.harvard.edu

Antonio Bassanetti
Banca D'Italia
Via Nazionale
91-00184 Roma
ITALY
Email: bassanetti.antonio@insedia.interbusiness.it

For further Discussion Papers by this author see:
www.cepr.org/pubs/new~dps/dplist.asp?authorid=135843

For further Discussion Papers by this author see:
www.cepr.org/pubs/new~dps/dplist.asp?authorid=154606

Riccardo Cristadoro
Servizio Studi
Banca d'Italia
Via Nazionale
91-00184 Roma
ITALY
Tel: (39 06) 4792 3341
Fax: (39 06) 4792 3720
Email: cristadoro.riccardo@insedia.interbusiness.it

For further Discussion Papers by this author see:
www.cepr.org/pubs/new~dps/dplist.asp?authorid=135845

Mario Forni
Dipartimento Di Economia Politica
Università di Modena e Reggio Emilia
Via Berengario 51
41100 MODENA
ITALY
Tel: (39 59) 417 852
Fax: (39 59) 417 948
Email: forni@unimo.it

For further Discussion Papers by this author see:
www.cepr.org/pubs/new~dps/dplist.asp?authorid=128014

Marco Lippi
Dipartimento di Scienze Economiche
Università di Roma La Sapienza
Via Cesalpino 12
00161 Roma
ITALY
Tel: (39 06) 4428 4202
Fax: (39 06) 4404 572
Email: marco.lippi@uniroma1.it

For further Discussion Papers by this author see:
www.cepr.org/pubs/new~dps/dplist.asp?authorid=101408

Lucrezia Reichlin
ECARES
Université Libre de Bruxelles
50, avenue Roosevelt CP 114
B-1050 Brussels
BELGIUM
Tel: (32 2) 650 4221
Fax: (32 2) 650 4475
Email: lreichli@ulb.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new~dps/dplist.asp?authorid=109257

Giovanni Veronese
Banca D'Italia
Via Nazionale
91-00184 Roma
ITALY
Tel: (49 621) 189 1886
Fax: (49 621) 189 1884
Email: veronese.giovannifurio@insedia.interbusiness.it

For further Discussion Papers by this author see:
www.cepr.org/pubs/new~dps/dplist.asp?authorid=154605

Marc Hallin
ECARES
Université Libre de Bruxelles
50, avenue Roosevelt CP 114
B-1050 Brussels
BELGIUM
Tel: (32 2) 650 5886
Fax: (32 2) 650 5899
Email: mhallin@ulb.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new~dps/dplist.asp?authorid= 127894

*The views expressed in this Paper are those of the authors and do not necessarily reflect those of the Bank of Italy, or any other institutions with which the authors are affiliated.

Submitted 12 November 2001

# 1 Introduction

The creation of the European Monetary Union (EMU) and the implementation of a common monetary policy call for the development of new analytical and empirical tools. The monetary policy of the European Central Bank (ECB) is based on *euro wide* economic developments and therefore rests upon the monitoring of a large number of national and area-wide statistics to obtain a reliable picture of the current and future economic situation.

A euro area business cycle indicator can be a valuable instrument for policy makers, since it would synthesize information coming from different sources and provide a clear signal as to the current business situation. Ideally, this indicator should exploit information on the correlation structure of many macroeconomic variables within and between the countries of the union. It should be free from idiosyncratic national dynamics and other measurement errors affecting national statistics. Finally, the ideal index should be cleaned from seasonal components and high frequency volatility of no concern for business cycle analysis.

Several problems need to be tackled to obtain this indicator. First, the lack of data availability on a comparable basis and for a reasonably long time span. On this point, there is still no agreement on which are the most reliable data sources for analyzing the euro economy and there is no standard data set containing cross-nationally comparable time series on all relevant macroeconomic variables. Second, GDP is not recorded on a monthly basis, so that information other than GDP should be used to obtain a monthly index. Third, data become available in a non-synchronous way, so that, to construct a real time index of economic activity, observations on those variables which become known with a lag, should be forecast. Fourth, to extract the cyclical component of the index, data must be filtered without inducing a phase shift of the index and, at the same time, so that little revision is necessary as new data become available.

This paper, which is the result of joint research between the Bank of Italy and the CEPR, is a developement of previous research on the same topics (see Altissimo et al. 2001). The aim of this paper is to solve the afore mentioned problems and produce a monthly real time coincident indicator of the euro business cycle which will be made public and regularly updated.

The first step of the project has been the construction of a data set of monthly time series covering a wide range of economic phenomena for the major euro–area countries. These series were collected from many different sources with the aim of obtaining a euro–area databank with a coverage comparable to those available for the US. We selected the statistics according to criteria of 'minimum harmonization' to allow cross–country comparability. In particular, wherever possible, we maintained a common sectoral breakdown across countries. The final product is a panel of 951 time series starting from 1987:1 up to mid-2001. The second objective has been to develop a new method for exploiting the information contained in this data set for the construction of the index. This method builds on previous research by Forni, Hallin, Lippi and Reichlin (2000 and 2001) (FHLR henceforth), but, we will argue, improves in several directions.

The model in FHLR can be summarized in the following way. Given a panel with a large number of variables $x_{it}$, it is assumed that the comovement of the $x$'s can be described by a *small* number of *common shocks*, so that each $x_{it}$ is the sum of a *common component*, which dynamically depends on the common shocks, and a residual *idiosyncratic component*. In FHLR (2001) a non-parametric estimator is constructed which is consistent and performs quite well when the number of variables in the panel is large compared to the number of time observations. In the present application, the method just outlined is employed in order to: (i) extract the common component of all series in the panel and, in particular, from monthly GDP, to be used as the reference variable; (ii) use all cross-sectional information for forecasting variables which are published with a lag, thus obtaining a timely indicator with only small revisions as data become available; (iii) use cross-sectional information for filtering out high frequency dynamics.

Regarding the last point, the idea is that cross sectional information can make up for a long lead and lag structure as in standard band pass filtering. Indeed a good cleaning from the high frequency component is obtained with a very small time window, thus causing an almost negligible end-of-sample unbalance. This seems a promising line of research and a possible improvement on recent work on band-pass filtering of macro economic time series.

The coincident indicator will be defined as the common component of the euro area GDP, filtered so as to eliminate high frequency variation (14 months and less).

As a by-product of our analysis we establish a dating of the euro cycle, characterize the degree of synchronization of national business cycles and the leading-lagging properties of various economic sectors in the main European countries. Together those results provide a useful basis for the assessment and understanding of the current economic situation in the area as well as of its likely future developments.

Some comments on the relation of the present paper to previous literature are in order. We use the generalized dynamic factor model studied in FHLR to construct a coincident indicator of the euro area. This allows for the use information contained in a large panel of time series to extract a signal which is cleaned from idiosyncratic noise. This general idea is also behind Stock and Watson's forecasting method applied to the US economy (Stock and Watson, 1998) and to the euro economy (Marcellino, Stock and Watson, 2000). As in FHLR (2000) we exploit information on the dynamic structure of the panel. However, we introduce a second step proposed in FHLR (2001), to obtain a re-estimation of the common component as well as of its optimal forecast which is a linear combination of past observations only. The second step is exploited to update the index and to solve the end-of-sample problem created by the two-sided filter used in FHLR (2000).

The paper is organized as follows. In Section 2 we outline the theoretical framework on which the definition and estimation of the coincident indicator is based. In Section 3 we describe our data set. From 3000 available monthly time series we initially selected, mainly for homogeneity reasons, about 1000 series. In Section 4 we show that a dynamic factor model with four common shocks gives an adequate representation of the series contained in the panel. In Section 5 we provide a first estimation, based on the previously selected panel, of the coincident indicator. The latter is then taken as a reference cycle to assess, for each relevant group of variables, pro- and anti-cyclicality, and the phase shift at business cycle frequencies, so that a classification of the main variables into leading, coincident and lagging is obtained. However, with the 1000-variable panel problems arise both for computational and timely availability reasons. In Section 6 we provide criteria, based on the analysis of Section 5, to substantially reduce the number of variables employed (the final set contains 246 series) and to construct our final coincident indicator. In Section 7 the euro area business cycle indicator is finally presented and discussed. Section 8 concludes. Technical details related to the theoretical part or the data treatment are confined to the Appendix.

# 2 Theoretical foundations

## 2.1 The model

We assume that our $j$-th time series, suitably transformed, is a realizations from a zero-mean, wide-sense stationary process $x_{jt}$. Each process in the panel is thought of as an element from an infinite sequence, so that $j = 1, \dots, \infty$. Moreover, all of the $x$'s are co-stationary, i.e. stationarity holds for the $n$-dimensional vector process $(x_{1t}, \dots, x_{nt})'$, for any $n$.

As in the traditional dynamic factor model, each variable is represented as the sum of two mutually orthogonal unobservable components: the 'common component' and the 'idiosyncratic component'. The common component is driven by a small number, say $q$, of 'factors' or 'shocks', common to all of the variables in the system, but possibly loaded with different lag structures. By contrast, the idiosyncratic component is driven by variable-specific shocks. In the traditional factor model, such component is orthogonal to all of the other idiosyncratic components in the cross-section, while here a limited amount of correlation is allowed for.

More formally, we assume

$$x_{jt} = \chi_{jt} + \xi_{jt} = \mathbf{b}_j(L)\mathbf{u}_t + \xi_{jt} = \sum_{h=1}^{q} b_{jh}(L)u_{ht} + \xi_{jt} \qquad (1)$$

where $\chi_{jt}$ is the common component, $\mathbf{u}_t = (u_{1t}, \dots, u_{qt})'$ is the vector of the common shocks, i.e. a (covariance stationary) $q$-vector process, whose spectral density matrix can be assumed to be the identity matrix with no loss of generality ($\mathbf{u}_t$ is an orthonormal white noise), $\mathbf{b}_j(L) =$

$(b_{j1}(L), \ldots, b_{jq}(L))$ is a row vector of $s$-order polynomials in the lag operator $L$, and the idiosyncratic component $\xi_{jt}$ is orthogonal to $u_{t-k}$ for any $k$ and $j$.

For both identification and estimation purposes we need the additional assumptions listed in FHLR (2001). In particular, in order to distinguish the common from the idiosyncratic components we need an assumption on the joint covariance structure of the $\xi$'s which does not rule out cross-sectional correlation, but puts limits on it. On the other hand, we need an assumption ensuring a minimum amount of correlation between the common components. For the technical details we refer to the paper above. Here let us simply say that correlations between the idiosyncratic components are such that the simple cross-sectional average vanishes in variance as $n \to \infty$, just as if they were pairwise orthogonal, while this property does not hold for the common components.

### 2.2 The indicator

Our proposed indicator is the common component of the european GDP growth at cyclical frequencies. As we show below our indicator fulfills three standard criteria to be an important synthesis of information on the euro area business cycle.

**Criterion 1: cross-sectional smoothing.**

Let us first observe that the idiosyncratic component captures both variable-specific shocks, such as shocks affecting, say, the production index of a particular industrial sector, and local-specific shocks, such as, for instance, a natural disaster, having possibly large but geographically concentrated effects. We do not expect these local or sectoral shocks to explain a large fraction of the European GDP, because of an obvious consequence of aggregation. However, as we shall see, though small, they are non-negligible.

As stated in the Introduction, the idea behind this paper is that eliminating the idiosyncratic component will produce a better signal for policy makers. Shocks originating from a local or sectoral source generate dynamics that should be monitored by local or sectoral policy makers, even if they were large enough to affect the European GDP to a certain extent. By contrast, common, Europe-wide policy should monitor the dynamics generated by common shocks. Hence, an index of the European business cycle—reference indicator for common European policy—should be cleaned by the idiosyncratic component.

Another important reason for cleaning the GDP from idiosyncratic components is that the latter should include most of measurement errors, as they are likely to be cross-sectionally poorly correlated. In evaluating the size of measurement errors, we should consider that the European GDP is obtained by aggregating data provided by heterogeneous sources, not all equally reliable, and that monthly estimates are obtained by interpolating data which are observed only quarterly.

**Criterion 2: intertemporal smoothing.**

The idiosyncratic component is not the only undesirable noise affecting the variables. In particular, GDP growth will be affected both by cyclical and by shorter-run movements, including seasonal and very short-run, high-frequency changes. For constructing a cyclical indicator such temporary changes should be washed out, in order to unveil the underlying medium- and long-run tendency of the economy.

As is well known, the common components $\chi_{jt}$, just like any other stationary variable, can be decomposed into the sum of waves of different periodicity (the so-called "spectral decomposition")[1] More specifically, we can disentangle a cyclical, medium- and long-run, component, say $\chi_{jt}^C$ and a non-cyclical, short-run, component, say $\chi_{jt}^{NC}$, by aggregating respectively waves of periodicity larger than, or smaller than, a given critical period $\tau$. This is done by applying to the series the theoretical band-pass filter discussed in Sargent (1987) and Baxter and King (1999), i.e.:

$$\chi_{jt} = \chi_{jt}^C + \chi_{jt}^{NC} = d^C(L)\chi_{jt} + d^{NC}(L)\chi_{jt}, \tag{2}$$

where $d^{NC}(L) = 1 - d^C(L)$ and $d^C(L)$ is a two-sided, symmetric, infinite-order, square-summable filter whose $k$-th coefficient is:

$$d_k^C = \frac{1}{\pi k} \sin(k \cdot \tau). \tag{3}$$

Hence, assuming that the European GDP is the first variable in the panel, our cyclical indicator is $\chi_{1t}^C$.

Since such indicator is not observed, it has to be estimated. We shall deal with estimation in a moment. Let us only anticipate here that, having an estimate of $\chi_{1t}$, obvious estimates of our indicator $\chi_{1t}^C$ could be obtained by applying the truncation of the filter $d_j^C(L)$ proposed by Baxter and King (1999) or the data-dependent approximation suggested by Christiano and Fitzgerald (2001). Such univariate filtering, however, would not exploit the superior information embedded in the cross-sectional dimension of the model. As we shall see in detail, our procedure is more efficient in that it can be regarded as a multivariate version of Christiano and Fitzgerald's. This enables us to obtain a good temporal smoothing with a very short filter, a fact which greatly reduces the typical end-of-sample distortions of two-sided filtered series. An efficient cleaning of short-run noise is a second important advantage of our factor model approach.

**Criterion 3: updating**

As already observed, a good indicator of the business cycle should be up-to-date. This is the case with our indicator. Precisely, each month $t$ we will be able to produce an estimate of the indicator for the previous month $t-1$ (and the previous quarter). Moreover, the estimation procedure involves estimation of the common factors at time $t$. Since the data will not be

[1] See e.g. Priestley (1989).

available either for time $t$ or for time $t-1$, our estimation will in fact be a prediction and will be subject to revision for a (short) period. This has the advantage that we describe what is happening now, not three or four months ago. But clearly we have a prediction error. Reducing the prediction error is the reason why the role of the information coming from the cross-section, and particularly from the leading variables, is crucial. Moreover, the leading variables will obviously play a crucial role in predicting the indicator (and/or the GDP itself) at time $t+k$, $k \geq 0$.

Identifying the leading-lagging relations of the variables and predicting efficiently is a third important motivation for the use of the generalized dynamic factor model in this context. For an extensive analysis of the performances of the model in prediction we refer to FHLR (2001). Here we simply give a simple intuition of the role played by the cross-section in forecasting.

Let us go back to model (1) and precisely to the impulse-response functions $b_{j1}(L), \ldots, b_{jq}(L)$ to the shocks $u_{ht}$. We assumed for simplicity that these functions are finite-order polynomials, but, apart for this, we did not place any further restriction. Hence the model is quite flexible, in that the reaction of each variable to each common shock may be small or large, negative or positive, immediate or delayed. This can accommodate pro-cyclical and counter-cyclical as well as leading and lagging or even more complicated behaviors.

For instance, assuming just a single shock $u_t$, the four dynamic loadings 1, $L$, $-1$ and $-L$ would characterize pro-cyclical and leading, pro-cyclical and lagging, counter-cyclical and leading, counter-cyclical and lagging variables. Notice that in this example, the leading variables are completely unpredictable given information at time $t$. The lagging variables, which are unpredictable by means of univariate modeling, can be predicted perfectly by using the leading ones.

In practice, common shocks are more than one and the dynamic responses are not so simple, so that we shall need specific criteria in order to classify the variables as counter-cyclical or leading (these criteria will be explained in detail in Section 5). Correspondingly, the relation between forecasting ability and "leadership" is less obvious. Nevertheless, the example provides a good intuition of the reasons why the model can perform well in forecasting.

### 2.3 The estimation procedure

Estimation of the coincident index is in three steps. In the first one we estimate the covariance structure of the common and the idiosyncratic components. More precisely, we estimate the spectral density matrix of the common and the idiosyncratic components by means of a dynamic principal component procedure. The theoretical basis of such procedure is found in FHLR (2000) and is summarized in Appendix B. Consistency results for the entries of this matrix as both $n$ and $T$ go to infinity can easily be obtained from the results in that paper.

From these estimated spectral-density matrices we can obtain the auto-covariances and cross-covariances for common and idiosyncratic components at all leads and lags by applying the inverse Fourier transform. Notice that we can easily get also covariances for the cyclical and the non-cyclical components $\chi_{jt}^{C}$ and $\chi_{jt}^{NC}$ simply by applying such transformation to the relevant band of the estimated spectra and cross-spectra. The details are reported in Appendix B.

In the second step, we compute an estimate of the static factors, following FHLR (2001). With the term "static factors" we mean the $q(s+1)$ variables $u_{jt-k}$ appearing in representation (1), so that, say, $u_{1t}$ and $u_{1t-1}$ are different static factors. To be precise, the static factors are not identified in the model unless we introduce additional assumptions, so that we shall in fact estimate a vector of linear combinations of such factors, say $\mathbf{v}_t$, spanning the same information space. Such estimates, say $\hat{\mathbf{v}}_t$, are obtained as the *generalized principal components* of the $x$'s, a construction which involves the (contemporaneous) variance-covariance matrices of the common and the idiosyncratic components estimated in the first step (see Appendix B). Such generalized principal components have an important "efficiency" property: they are the contemporaneous linear combinations of the $x$'s with the smaller idiosyncratic-common variance ratio. They can consistently approximate any point in the common-factor space, including the common components $\chi_{jt}$'s, as $n, T \to \infty$ in a proper way. Similarly, we can get forecasts of the common components (and the factors themselves) simply by projecting $\chi_{jt+k}$ (or the $k$-th lead factor) on $\hat{\mathbf{v}}_t$. This forecast approximates consistently the theoretical projection.

In the third and final step we use the present, past and future of the static factors to get our estimate of the cyclical component of the GDP $\chi_{1t}^{C}$. Precisely, we project $\chi_{1t}^{C}$ on $\mathbf{v}_{t-m}, \ldots \mathbf{v}_{t+m}$.[2] The lag-window size $m$ should increase with the sample size $T$, but at a slower rate. Consistency of such estimator is ensured, for appropriate relative rates of $m$, $T$ and $n$, by the fact that (a) the projection of $\chi_{1t}^{C}$ on the first $m$ leads and lags of $\chi_{1t}$ is consistent because of consistency of $\hat{\chi}_{1t}$ and the estimated covariances involved; (b) $\chi_{1t}$ is a linear combination of the factors in $v_t$, so that projecting on the factors cannot be worse than projecting on the common component itself (see Appendix B for further details).

Notice that here we have something like a multivariate version of the procedure by Christiano and Fitzgerald (2001) to approximate the band-pass filter. Exploiting the superior information embedded in the cross-sectional dimension enables us to obtain a good smoothing by using a very small window ($m = 1$). This is very important in that we get readily a reliable end-of-sample estimation and are not forced to revise our estimates for a long time (say 12 months or more) after the first release, as with the univariate procedure. To get an intuition of the reason why we get good results with a narrow window, consider the extreme case $m = 0$. Clearly with univariate prediction we cannot get any smoothing at all. By contrast, the static factors

[2] We do not estimate OLS, but use the projection coefficients derived by the covariance matrices of the cyclical components estimated in the first step.

will include in general both the present and the past of the common shocks and can therefore produce smooth linear combinations.

### 2.4 End-of-sample unbalance

Finally let us explain shortly how we treat the problem of end-of-sample unbalance (further details can be found in Appendix B). Typically data referring to period $T$ become available some periods later and different variables have in general a different delay. Hence if we want to estimate the model as it stands, we are forced to wait until the latest observation arrives. Clearly we can reduce the problem by eliminating from the data set series whose delay is larger than a given limit, as we explain in Section 6. But even so we are necessarily left with three or four months, at the end of the sample, for which some observations are available and some others are not.

Our procedure to handle this problem is the following. Let $T$ be the last date for which we have all the data set. Until $T$ we estimate the static factors as explained above, i.e. by taking the generalized principal components of the vector $\mathbf{x}_T = (x_{1T}, ..., x_{nT})$. From $T$ on, we use the generalized principal components of a modified $n$-dimensional vector $\mathbf{x}_T^*$ which includes, for each process in the data set, only the last observed variable, in such a way to exploit, for each process, the most recent information. Clearly computation will involve the estimated covariance matrices of the common and the idiosyncratic component of $\mathbf{x}_T^*$ in place of those of $\mathbf{x}_T$. Having an estimate of these factors, call them $w_T$, we estimate $\chi_{1T+k}^C$, $k > 0$, by projecting it on $w_{T-m}, \dots, w_T$.

## 3 A unified euro area database

Unlike the U.S. case, where analysts can easily access well established and large databases,[3] nothing of this sort yet exists in Europe. We had therefore to consult and evaluate many different sources: among others, national statistical institutes, the OECD and the Eurostat statistics; from these we collected and examined a large number of series, organizing them in a detailed dataset.

The final database—whose richness of properly organized and monthly updated information could make it a particularly useful tool for further research—has been organized into the following eleven homogeneous blocks, corresponding to different major sectors: industrial production; producer prices; consumer prices; money aggregates; interest rates; financial variables and exchange rates; European Commission surveys; national Institues surveys; trade; labor market series and a miscellanea of other variables. On the whole, they should provide an almost

[3] See, as an example, the DRI-McGraw Hill Basic Economics database, formerly known as 'Citibase'.

exhaustive description of the European economy.

Each block contains time series for Germany, France, Italy, Spain, The Netherlands, Belgium and, when available, for the euro area as a whole. Indeed, since a business cycle indicator for the euro area reflects economy wide fluctuations *common across countries*, one should collect data covering a wide variety of sectors for *all* European economies. Unfortunately, data limitations forced us to restrict the focus on the six largest countries that, nonetheless, accounted for more than 90% of the euro aggregate GDP in 2000.[4] Some key macroeconomic series not directly referred to the euro area were also included to capture phenomena that might be relevant to explain fluctuations across Europe; some examples are oil and raw material prices and some indicators of the business cycle in other large economies (UK, US and Japan).[5]

Altogether the database contains about 3000 time series. Among these we selected only those variables that satisfy two crucial requirements: one concerning the length of the series and the other their homogeneity over time and across countries. As regards the first one, the largest common sample for the dataset spans the period January 1987 - March 2001.[6] Although many time series are available for a longer period, the decision to set the starting date in 1987 is the result of a trade off between obtaining richer time series information and maintaining a large cross-sectional dimension for the dataset.

As for the second requirement (i.e. homogeneity over time and across countries) we selected variables from each of the eleven blocks trying to maintain, for each of them and wherever possible, a common breakdown for all countries. In some cases to obtain series of sufficient length we had to join together statistics covering shorter time spans (for example HICP and CPI or Pan-German with West German data), trying to match definitions and disaggregations as closely as possible (see Appendix A for details). Whenever we had to intervene with some kind of manipulation to obtain time series of the desired quality, the strategy adopted for data reconstruction was the following. For the most recent years series were collected from Eurostat or the ECB, since these institutions coordinate national sources in the process of statistical harmonization. Then other international institutions (like the OECD) or national sources (e.g. Insee or Istat) were used to obtain series of sufficient length or to cover important economic phenomena with the desired detail.

As a last step, in order to avoid overweighting a single country or a particular economic sector, in selecting time series we tried to maintain a satisfactory balancing in terms of numerosity across countries and blocks. Nevertheless, closely pursuing this criterion with the available statistics

[4] The few series relative to the euro area as a whole included in the dataset should contribute to counterbalance the drawback due to lack of variables from the smallest countries.

[5] These time series account for a very small portion of the whole dataset.

[6] This constitutes a difference with respect to other studies that focused on a single source and a shorter time span. See for example Marcellino, Stock and Watson (2000), where only the OECD Main Economic Indicators database for the period 1982:1-1998:12 is exploited.

would have forced us to work with a minimal common set of indicators, thereby forgoing much information. This is the reason why, to meet the requirement of a large database, we preferred to relax the condition on *perfect balancing*.

Applying this selection criteria we ended up with 951 monthly time series (see Table A1 for details on data sources and Table A2 for numerosity of time series in each block). Further details on the data and a description of the procedures applied to treat outliers, non-stationarity and seasonality can be found in Appendix A.

# 4 Is there a euro area business cycle?

As we have argued in the Introduction, applying a dynamic factor model to the construction of a euro area coincident indicator of the business cycle requires that economic time series of different countries and sectors strongly co-move at the business cycle frequencies. In turn, comovement means that a very small number of common shocks are able to explain a considerable fraction of the variability of the series in the panel. A growing body of empirical literature has addressed this question using various methods[7] and has recently received a further impulse by the creation of the EMU. Most studies find evidence of a rising degree of integration and synchronization among European economies, while some differences in the cyclical behavior across countries still persist. In this paper we will not explicitly address the question of synchronization of business cycles across euro area countries, even though our results throw some light also on this issue. Rather we will attempt to describe the dynamic behavior of the series included in our panel and show that comovements at business cycle frequencies are relevant across countries and sectors and are captured by a limited number of common factors.

We can firstly investigate the existence of business cycle co-movements by looking at the *average spectral shape* of the series in each main block (sector) of our cross-section as well as in the whole data set. The simple arithmetic average of the spectral density functions of all the variables in the data set tells us how on average the overall volatility is distributed across different periodicities and exhibits relevant dynamics at low frequencies (see Figure 1). The same property, i.e. a large portion of the variance concentrated at business cycle frequencies (shaded area in Figure 1), holds for the majority of sectoral blocks. Low frequency fluctuations account for large part of the variance of the producer price indices (PPI), the consumer price indices (HICP) and the labor market variables, while in the case of the industrial production and the survey data high-frequency noise downplays the business cycle component. Overall we can conclude that monthly series have, on average, a clearly detectable cyclical behavior, responsible for a sizeable part of their variation.

[7] McDermott and Scott (2000), Lumsdaine and Prasad (1999), Cheung and Westermann (2000), Dickerson et al. (1998), Artis et al. (1999).

Given this evidence, it is natural to ask whether the movements at business cycle frequencies are common across Europe as well as across different types of economic activities. The question can be answered by principal component analysis, as extended by Brillinger (1981) to take into account the dynamic relationships between the series in the panel.[8] Let us briefly recall that the first dynamic principal component of the variables $x_{it}$, $i = 1, \ldots, n$, call it $Z_t$, is the linear combination of lags and leads of the $x$'s, such that the variance of $x_{it}$ explained by $Z_t$, summed for $i = 1, \ldots, n$, is maximum among all linear combinations. The second dynamic principal component has the same definition, but for the constraint that it must be orthogonal to $Z_t$ at any lead and lag, and so on. If the variables $x_{it}$ are not correlated at any lead and lag then the variance explained by $Z_t$ is, roughly speaking, one $n$-th of the total variance of the $x$'s. On the contrary, a high fraction of the variance of the $x$'s explained by $Z_t$, or by the first $q$ dynamic principal components, with $q$ very small as compared to $n$, reveals the presence of a strong comovement of the $x$'s. As an extreme example, if the $x$'s are all driven by just one white noise shock, i.e. $x_{it} = b_i(L)u_t$, then the first dynamic principal component would explain 100% of the total variance of the $x$'s.

We denote by $\lambda_j(\theta)$, $j = 1, \ldots, n$, the $j$-th eigenvalue, at frequency $\theta$ and in decreasing order, of the spectral density matrix of the $x$'s. It may be shown that the contribution of the $i$-th dynamic principal component to the total variance at frequency $\theta$ is then given by:

$$\frac{\lambda_i(\theta)}{\sum_{j=1}^{n} \lambda_j(\theta)} \tag{4}$$

Figure 2 exhibits the first eight "normalized" dynamic eigenvalues on the interval $[0; \pi]$. The average contribution to total variability of the first dynamic principal component is around 20% and increases to 28% in the cyclical interval $[0; \frac{1}{7}\pi]$.[9] The second dynamic principal component accounts, on average, for 14% of the total variance, which increases to 16% in the interval $[0; \frac{1}{7}\pi]$. The third and fourth dynamic principal components explain 11 and 10%, respectively; each of the remaining principal components accounts for less than 7%. Overall the first four dynamic principal components explain more than 55% of the total variance of the 951 series; which increases to 65% when focusing on the interval $[0; \frac{1}{7}\pi]$. In Figure 3 the overall explained share of variance, cumulated for the first four dynamic principal components, is shown as a function of the frequency.

We conclude that not only our data exhibit on average large variability at low frequencies, but also that there are strong co-movements across series and that this commonality is particularly significant at business cycle periodicities.

[8] In addition to Brillinger (1981), see FHLR (2000).

[9] The cyclical band comprises oscillations with periodicity between 14 and 120 months, i.e. $\theta \in [0.05; \frac{1}{7}\pi]$. However in the empirical application, for reasons that will be explained in Section 5, we will consider a slightly larger band which includes also the zero frequency.

As argued in FHLR (2000), a reasonable criterion to select the number of common factors in the dynamic factor model, which is preliminary to the estimation step, consists in fixing $q$ as the number of dynamic principal components of the $x$'s that individually explain more than a conventional percentage. Here we set such percentage at 10% and therefore take in the sequel $q = 4$.

In Tables 1 and 2 the share of explained variance (indicated by $\mathrm{var}(\chi)/\mathrm{var}(x)$; see Appendix B), is detailed by country and sector (the individual contribution of the first four dynamic principal components is also shown). Looking at Table 1, we observe shares going from a low of around 50% of the surveys, with a considerable part of their variance concentrated at high frequencies, to a peak of around 60% for producer prices, consumer prices and interest rates, with a large part of their variability concentrated at low frequencies.

The results in the tables allow the conclusion that none of the selected factors can be associated to a specific country and/or economic activity.

| SECTOR | $\frac{var(\chi)}{var(x)}$ | D.F. I | D.F. II | D.F. III | D.F. IV |
|---|---|---|---|---|---|
| IP | 0.55 | 0.18 | 0.15 | 0.11 | 0.10 |
| PPI | 0.58 | 0.22 | 0.15 | 0.12 | 0.09 |
| HICP | 0.57 | 0.23 | 0.16 | 0.10 | 0.09 |
| SURVEY | 0.50 | 0.18 | 0.12 | 0.11 | 0.10 |
| MONEY | 0.53 | 0.19 | 0.16 | 0.10 | 0.07 |
| INTEREST RATES | 0.64 | 0.26 | 0.15 | 0.12 | 0.11 |
| FINANCIAL | 0.57 | 0.22 | 0.15 | 0.11 | 0.09 |
| LABOR | 0.55 | 0.20 | 0.14 | 0.12 | 0.09 |
| OTHER VAR. | 0.60 | 0.27 | 0.14 | 0.09 | 0.08 |
| TRADE | 0.58 | 0.24 | 0.13 | 0.11 | 0.08 |
| INTERNATIONAL | 0.53 | 0.19 | 0.16 | 0.11 | 0.08 |

Table 1 - Share of explained variance per blocks.

| COUNTRY | $\frac{var(\chi)}{var(x)}$ | D.F. I | D.F. II | D.F. III | D.F. IV |
|---|---|---|---|---|---|
| GERMANY | 0.58 | 0.24 | 0.14 | 0.10 | 0.09 |
| FRANCE | 0.55 | 0.20 | 0.14 | 0.10 | 0.10 |
| ITALY | 0.54 | 0.19 | 0.14 | 0.11 | 0.09 |
| SPAIN | 0.53 | 0.19 | 0.13 | 0.11 | 0.09 |
| HOLLAND | 0.56 | 0.21 | 0.13 | 0.11 | 0.10 |
| BELGIUM | 0.55 | 0.19 | 0.14 | 0.10 | 0.09 |
| EURO | 0.56 | 0.24 | 0.14 | 0.08 | 0.09 |

Table 2 - Share of explained variance per country.

# 5 Cyclical behavior of the variables

Having set to four the number of common shocks in the preliminary analysis of Section 4, we can now use the method introduced in FHLR (2000) to estimate the spectral density matrix of the common components $\chi_{it}$. The well known integral formulas then allow the computation of the implied covariance matrices (see Appendix B). If the integrals are computed over a given frequency band, the resulting covariances correspond to the outcome of band-passing the $\chi$'s over the given frequency interval. More precisely, the cyclical band was extended to include the zero frequency. This extention, on one hand, does not make an important difference with respect to taking the strict cyclical interval. On the other hand computing the mean lag, as we do below, requires the spectral density at zero. It is important to point out that for the moment we are only interested in studying pro- and anti-cyclicality and lead and lag relationships between the common components and the reference cycle, defined as the common component of aggregate European GDP. This requires only spectral densities and covariances, not actual estimation of the $\chi$'s. As a consequence, the issue of end-of-sample unbalance, due to two-sidedness of band-pass filters, does not arise here. In Section 7 instead we will be concerned with actual estimation of band-passed common components. In that case, the two-step procedure briefly outlined in Section 2 will find full application.

The GDP is an overall measure of economic activity and it has clear advantage with respect to a more limited measure such as industrial production; the drawback of using GDP as the basis of our reference variable is that it is measured only every three months. However it can be regarded as the outcome of an unobserved monthly process; the linear interpolation of the quarterly figures is therefore a proxy for the unobserved GDP. Since we are interested in the *common component* of this variable, this assumption should ensure that the we can obtain a consistent estimate if the approximation error is not correlated with the dynamic factors driving the cross section. Indeed this condition does not seem too demanding given that this particular type of measurement error affects only the GDP variables in our cross section.

Prior to studying the correlation of the common components with the reference cycle, they have been classified as pro- or counter-cyclical according to the phase angle with respect to the reference series evaluated at zero frequency (or, equivalently, according to their mean lag, i.e. the average correlation with the reference cycle across all lags). When the phase is zero (positive mean lag) a variable is classified as pro-cyclical, when it is equal to $\pi$ (negative mean lag) as counter cyclical.

Having split the sample into two groups we further distinguished variables into three categories: lagging, coincident and leading. For procyclical variables we looked at the time displacement of the maximal *positive* correlation, classifying the series as leading when the correlation was maximal at a time displacement smaller than $-2$, lagging when greater than 2 and coincident otherwise (where for a given common component $\chi$ and reference cycle $r$, the correlation at displacement $k$ is given by $\rho_k = E\left(r_t \cdot \chi_{t+k}\right))$. The same criterion was applied for countercyclical variables, the displacement considered being in this case that of the minimal *negative* correlation.

Overall there are 802 procyclical and 149 countercyclical variables, 258 leading, 404 coincident and 289 lagging (see Tables A4 and A5).

We study the correlation structure of our panel across countries and across economic sectors, both dimensions being interesting for a full characterization of the business cycle fluctuations in Europe.

A first feature worth noting is that while the explanatory power of the common factors is uniform across countries and sectors, as illustrated in Section 4, more distinctive characteristics emerge when one considers the lead and lag relationships among variables.

Considering countries first, Belgium and The Netherlands on average lead the euro area cycle thus confirming a widely held view (see table A4), while Spain and Italy are lagging. However our inference about countries might be influenced by difference in the data collected for each national economy; a better understanding of the country properties should follow by the investigation of the national GDP behavior in the last section. The analysis by sector is what we consider now.

### 5.1 Industrial Production

The 176 series of industrial production revealed a widespread pro-cyclical behavior being, on average, coincident with the European business cycle: the median lead for this block is one month and the average variance explained by the common shocks at cyclical frequencies is 40% (see Table A5).

Only 14 variables displayed a countercyclical behavior, namely tobacco, extraction of coke and other minerals and manufacturing of radio TV and communication equipment for France and Spain. Most of the industrial production series (more than 50%) appear to be coincident, but almost one third of the series show fairly good leading properties of the European business cycle (see Figure 4).

Not surprisingly the euro area total industrial production (excluding construction) is almost coincident with the euro area cycle (leading of one month) and it has maximal correlation with the reference variable equal to 0.83; these properties are also shared by the country indeces with the noticeable exception of The Netherlands whose total industrial production leads the euro area cycle by five months and it is poorly correlated with the euro area (see Figure 5).

Some common interesting features emerged across economies. The production of chemicals and of basic metals share the same leading properties area wide, both sectors on average anticipate the euro area cycle by 4 months; while the production of intermediate goods is leading by three months across the euro area with the exception of The Netherlands where the lead increases to seven months.[10]

### 5.2 Prices and Wages

*Price variables* display a strong comovement within the cross section: more than 60% of their variation at cyclical periodicity being explained by the first four factors. They are procyclical and tend to lag the fluctuations of our reference variable. This result concerning the growth rates of prices is consistent with Stock and Watson's finding with US data.[11]

*Consumer prices* in almost all countries appear to be lagging with an average displacement of about 4 months. In Belgium and in The Netherlands the overall indices of consumer prices and the core components (goods excluding energy and food, and services prices) lead the European cycle by a few months; in contrast the same items turn out to be slightly lagging in the other economies considered. A noteworthy feature is that prices of energy products tend to be countercyclical and leading of around two years in all countries.

*Producer prices* are in phase with the reference cycle with an average time displacement of about 2 months, hence their movements anticipate the corresponding cycle in consumer prices, consistently with the commonly entertained transmission mechanism. Similarly to what was found for consumer prices, in Belgium they revealed to be leading with respect to our reference variable.

*Nominal wages* are procyclical and on average strongly correlated with the cycle, thus displaying a behavior similar to prices. On the other hand, they result generally lagging by more

[10] The sectors involved in the manufacturing of packing materials - like pulp, paper and paper products - present a strong leading property both in Italy and Spain.

[11] See Stock and Watson (1999) p. 42.

than two quarters, thus following prices, with the exception of wages in The Netherlands that are countercyclical and leading by two quarters.

### 5.3 Employment Statistics

Vacancies are pro-cyclical and leading in Belgium, lagging in Germany. In accordance with previous empirical findings, unemployment is always lagging and countercyclical with a high correlation with the cycle, the only exceptions being Belgium and Spain where it appears coincident. Statistics on hours worked and on temporary layoffs are not generally available at a monthly frequency for euro area countries. In Italy temporary layoffs are countercyclical, as could be expected, but their correlation with the business cycle is rather weak.

### 5.4 Survey Data

The surveys, both the ones coming from European Commission and those from the national institutes, commonly used by short term analysts to assess current and perspective economic situation, contain indeed relevant information for business cycle analysis, in particular more than 40% of the series included in this group are classified as leading. Nonetheless caution has to be paid in interpreting this evidence since the variance explained by common factors for this group is among the lowest, meaning that the signals released are quite noisy.

In the manufacturing sector survey, the questions concerning the level of orders, production expectations and overall business situation are leading across countries and comove positively and strongly with the euro area business cycle; in the three major countries the average lead is of a quarter and it increases to five months for the Belgian economy. Among the questions in the survey, the one pertaining to the short term production expectations has the largest average time lead across different countries, whereas assessment of stocks and unemployment expectations are countercyclical for all countries.

The construction sector shares the same anticipating properties of the manufacturing sector but with a shorter lead, while the retail trade sector seems less correlated with the cycle and present mixed signals cross counties, being strongly leading in Belgium and Italy while lagging by more than two quarters in Germany.

Among the business confidence indicators, the common component of the IFO indicator leads the euro cycle by about one quarter and it has maximal correlation of 0.64 with the reference variable; similar lead and correlation properties are shared by the manufacturing confidence indicators for France and Italy while the indicator for Belgium presents a longer lead.

The consumer confidence indicator has weaker leading properties than the business indicator, reflecting the mixed evidence coming from different questions; it is homogeneously leading of about 2 months in the various countries. Beside the good time lead characterizing the expectations on the general economic situation of the country and the intentions of carrying out major purchases, consumers' evaluations on price trends appear to generally lag the business cycle.

### 5.5 Monetary aggregates

The relationship between money and real activity is one of the most debated issues in macroeconomics and no clear agreement has emerged so far in the literature about the existence and the direction of a causal link. However a growing body of theoretical research has argued that nominal rigidities should play an important role in the description of the mechanics of business cycle. Without taking a position in this debate, we included in our panel the most commonly used measures of money, M1, M2 and M3 (for each country and for the euro area), both in nominal and in real terms to avoid ruling out a possible important source of fluctuations. A further reason why we thought it important to include money lies in the prominent role assigned to money (and in particular to M3) in the conduct of monetary policy by the ECB.

The cyclical component of nominal monetary aggregates growth rates explains a sizable part of their variability and it is linked with the business cycle as shown by their quite high correlations with the indicator (see table A5). On the other hand the commonly recognized difficulties in obtaining a reliable measure of money are probably the explanation for cyclical patterns that differ across countries. In most cases monetary aggregates are procyclical (see in particular Italy, France and Spain), but no clearcut leading properties emerge for these measures. The euro area M1 and M2 lead by 2 quarters the cycle, but, surprisingly, M3 lags it by 18 months. A possible explanation of the result is that given the quite short duration of business cycles in Europe, a variable that leads the cycle by a large number of months can easily be classified as lagging being closer to the preceding cycle than to the next.

The same pattern emerges for real money aggregates (where nominal variables have been divided by the current price level as measured by the HICP). The euro area M1 and M2 are leading, while M3 is still lagging by 18 months. Overall real money aggregates seem to have a better predictive content than the nominal ones. Germany again is countercyclical.

### 5.6 Interest Rates

We included in our panel various nominal interests rates (on t-bills and t-bonds as well as on banking loans), we also constructed time series of real *ex post* interest rates and of spreads between long term and short term rates.

Interest rate spreads are procyclical and leading, with an average lead of more than one quarter in accordance with the commonly held view that spreads have a good predictive content of future real activity (the only exceptions being Belgium and Italy). The association with the reference cycle—measured by the mean absolute correlation—is rather low (0.43) compared with that of other variables. This would suggest that spreads provide a noisy signal and not always give reliable news concerning the future business outlook.

As for real rates, 8 out of 12 are leading and countercyclical as should be expected. Their average lead is 6 months. However, even though most of their variance is explained by the common cycle, the maximal correlation remains low, as in the case of spreads, leading to the same caveat.

All other nominal rates are procyclical and either coincident or lagging and both highly common and correlated with the cycle.

### 5.7 Other Financial Variables

This group includes share prices and exchange rates.

Stock exchange indexes and share prices are procyclical and leading, with an average lead of 5 months even though their variance is only weakly linked with business cycle variability. It is worth recalling that here we are concerned not with the tic-to-tic or the daily variation of those indexes, that would clearly be hardly linked with business cycle movements, but with the filtered variation of the index that by construction is free from high frequency variability[12].

No clear common pattern seems to characterize the effective exchange rates (both nominal and real), some are countercyclical and some are not, the same ambiguity emerging for their lead-lag relation with the cycle.

### 5.8 Trade flows

Trade sector includes almost 100 series. This rather large sample includes total exports and imports of each euro area country together with some disaggregation by commodity, but also bilateral exchanges, allowing us to distinguish, in particular, between intra UE and extra UE trading.

On average trade variables are less related to the cycle than series included in other groups, the variability explained by the common cyclical components being in fact only 0.48. Furthermore—as shown by Figure A5—they are generally lagging.

Imports are more closely linked to domestic demand than exports so it is not surprising that their correlation with the cycle is higher than that of exports whether they come from the european union or from extra UE countries. They also result on average more lagging than exports.

Trade in some commodities tends to lead the cycle, in particular import and exports of raw materials, crude oil and food and beverages anticipate by one or two quarters the reference cycle.

[12] For a similar point see Stock and Watson (1999) p. 43.

Interestingly enough exports toward non UE countries and towards the UK are negatively related to the European cycle. Being able to break trade by country is actually very useful to explain some results that would otherwise be puzzling. For example, the italian merchandise export is countercyclical, but distinguishing between extra and intra UE leads to the conclusion that while the latter is procyclical and positively (even though weakly) related to the cycle, the first is responsible for the countercyclicality and negative relationship of total exports. Also exports toward the UK in Germany and in Italy have a negative relationship with the cycle, pointing to a closer link between UK and the US than between UK and the rest of Europe.

### 5.9 Other Variables

Variables in this block were chosen to capture particular phenomena that could help to forecast economic activity. They are, therefore, very heterogeneous and a small part of their variability is captured by the first 4 dynamic factors. Leading features were displayed by new orders of construction and residential buildings, wherever available, and by some specific indicators of industrial activity: in particular car registrations in Italy and Spain. In the Italian case, electricity consumption - whose properties have been well documented[13] - is found to have a 2 months time lead on European GDP.

## 6 The reduction of the data set and the real time feasible indicator

As we have already observed in Section 3, the dataset employed up to now, containing 951 time series, has two serious drawbacks. Firstly, computer elaborations are rather heavy. Secondly, and more importantly, too many variables in the panel are published with a long delay with respect to the reference period. This motivates our decision to substantially reduce the dataset. Three criteria were applied.

The first criterion was aimed at reducing the delay-in-publication problem. Typically, financial variables and monetary aggregates are updated with the shortest delay; consumer prices, firms and consumer surveys take slightly more time and are released one month after the reference period. Industrial productions and producer prices are usually published with a delay of about two months. Labor market and, above all, trade variables are the ones released with the largest time displacement with respect to the reference period. Eventually we decided to keep the variables whose most recent observation was not older than 2001.03 (i.e. three periods before the most updated series).

[13] See Marchetti and Parigi (2000).

The second criterion is based on the share of cyclical variability explained by the four common shocks. We kept the variables with a high degree of commonality. In particular, among the leading and the lagging variables, only those with more than 60% of the variance explained by dynamic factors were held, this share rising to 70% for the coincident. Eliminating time series with a lower percentage share led to an equally proportioned reduction of the three groups of variables.

The third criterion requires some explanation. As we have seen in Section 2, the final step of our estimation procedure consists in a contemporaneous aggregation of the $x$'s. On the other hand, we have assumed that the common components $\chi_{it}$ load the shocks $u_{jt}$, $j = 1, 2, 3, 4$, with lags from 0 to $s$. Now, the most 'extreme' static factors, $u_{jt}$ and $u_{jt-s}$ in particular, are loaded with significant coefficients only by 'very leading' and 'very lagging' variables respectively. Since such extreme variables are rare in the panel, the factors $u_{jt}$ and $u_{jt-k}$ are likely to be poorly estimated. Therefore we decided to drop too leading or to lagging variables. Precisely, the leading and lagging variables kept for the reduced dataset are those with a maximum time displacement not larger than seven months.

The three criteria proved to work remarkably well: the dimension of the cross-section decreased to 246 time series. All major European economies remain adequately represented; only Germany can count on a larger number of variables than other countries but this feature already characterized the full data set. Sectors are heterogeneously included: this does not seem to be a problem since we already had to relax the perfect balancing condition for the full database. Within the reduced cross-section, 118 variables were classified as being coincident, 69 as leading and 59 as lagging. These numbers broadly reflect the proportions in the full database; only the lagging variables are slightly under-represented, but this is a feature we do not dislike.

# 7 The Euro Area Business Cycle Indicator

The reduced dataset is used to implement the four-step procedure described in Section 2.

1. In the first step we estimate the spectral density matrix and the autocovariance matrices of the common and idiosyncratic components.

2. In the second step these covariance matrices are employed to estimate the space spanned by the shocks $u_{jt-k}$, $j = 1, 2, 3, 4$, $k = 1, \dots, s$, call $S_t$ such a space. Then, defining $\chi_{jt}^C = d^C(L)\chi_{jt}$, where $d^C(L)$ is the band-pass filter corresponding to the cyclical frequency-interval, we compute the projection of $\chi_{jt}^C$ on the space spanned by

$$S_{t-m},\ S_{t-m+1}, \dots, S_t, \dots, S_{t+m-1},\ S_{t+m}.$$

The number of lags $s$ has been chosen on the basis of the selection criterion proposed by Bai and Ng (2001), in our case $s = 17$. Inspection of the spectral density of the estimated common component suggested setting $m$ equal to one. As already pointed out, such a small $m$ ensures a minimum end-of-sample unbalance and revision.

3. Our final indicator will be the *three-month* change of the common component of the euro area GDP.

4. The end of sample adjustment of the indicator in order to couple with the different time release of the data.

The coincident indicator for the euro area is estimated over the period 1987.02 to 2001.02 on the reduced data set and is shown in Figure 6. The solid line is the indicator constructed by means of the common components calculated with the two step procedure method on the reduced data set; while the dotted line is the indicator constructed on the complete data set. The two indicators are very similar, showing that there has been no substantial loss of information in the data reduction process.

Being defined in terms of the three-months output growth rates, the index is consistent with a *growth cycle* definition of the business cycle, i.e. it can be interpreted as deviations of economic activity from its long-term trend, identified by the zero line in the figure. Positive values of the indicator signal periods of growth above the long run average, and the reverse for values below zero. Hence the peaks (troughs) have to be interpreted as periods of maximal (minimal) growth, that are followed by a deceleration (acceleration) in overall activity. While this kind of definition already existed in the traditional literature, it should be stressed that the procedures embodied in the original NBER methodology are based on the 'classical cycle' concept, which focuses on fluctuations in the *absolute level* of economic activity.

### 7.1 Data irregularity at the end of sample

The procedure used to construct the indicator requires time series defined on a common time range. As described in Section 3, the 951 time series span the common time range 1987.02-2001.02; nonetheless observations for the most recent months are available for many of them.

In the reduction process we eliminated the variables with the greatest delay -i.e. those ending in 2001.2. However, without a procedure to handle the end of sample unbalance problem, the coincident indicator could be computed at most until February 2001 thus loosing the information coming from the series available with the greatest timeliness. This is not a negligible issue: among the 246 variables of the reduced dataset, only 47 are updated up to March 2001 (among which are the GDPs), whilst 50 have observations until April (industrial productions and some surveys data); 102 until May (in particular producer prices, firms and consumers surveys) and 46 until June (financial variables and exchange rates).

The method to handle the end of sample unbalance, proposed in Section 2, allows us to update our coincident indicator to June 2001, i.e. the most recent observations in the panel. The idea underlying the method is to compute the static factors on re-aligned data, in which each single variable is recorded with its most recent observation (see Appendix B). This allows to make use also of the most updated information, providing a "real time" cyclical coincident indicator.

Information concerning the GDPs of the second quarter of the year will be released either late in the summer or even at the beginning of autumn. Waiting for their release to compute the indicator up to June 2001 would imply little contribution of it to the analysis of the current economic situation. Instead of this, the estimate of June 2001, according to the solution described, enables to have timely signals to interpret the current cyclical phase.

### 7.2 The dating

The dating of the Euro area business cycle will rest on the behavior of the indicator just constructed. The visual inspection of the coincident indicator shows that the euro area from the end of the eighties to the year 2000 experienced four complete cycles (from peak to peak): 1989.03-1992.11, 1993.01-1994.10, 1994.11-1997.10 and 1997.11-1999.12 and now. Applying the Bry-Boschan dating scheme to our coincident indicator confirmed this dating for the european business cycle, (see Figure 7). It is worth noting that the dating of the business cycle is an ex-post assessment of the past dynamic and it is not performed in real time.

The average duration of expansion episodes is roughly 12 months, while the recessions lasted on average 13 months. The first episode at the end of the eighties concludes the long expansion of this decade, which ends at the last months of 1988. The use of the cross sectional information casts an interesting light on this downturn episode: the decline in the coincident index appears in contrast with the dynamics of the *original* GDP variables, steadily growing up to the mid 1990s, while the *common component* of other series in the panel signaled a downturn. The recession ends with the short expansion between 1991 and early 1992, mainly related to the German unification. The 1993.01-1994.10 episode includes in particular the 1992 currency crisis which led to strong devaluations of the Italian Lira and the British Sterling. Afterwards the euro area cycle experienced two expansionary phases (1993.01-1994.10 and 1995.12-1997.11) lasting around two years each and two recessions (1994.11-1995.11 and 1997.12-1998.10) of short duration, an year each. The most recent peak occurred at the end of 1999.

Unlike the US experience, which register only a short recession at the beginning of the nineties and a continuous growth subsequently, in the same period the euro area economy experienced four complete phases of acceleration and deceleration of the economic activity.

### 7.3 Assessing the index

As a first check on the quality of our reference variable we compared the coincident indicator with the quarter on quarter changes of Euro Area GDP (see Figure 8), after properly rescaling the two variables. The index closely tracks the GDP while being monthly and having no delay in the release; interestingly the two variables resemble each other more in the second part of the sample, pointing to a higher degree of commonality across euro area economies in the second half of the nineties.

This finding is confirmed by investigating the behavior of the national components of the coincident index that make up the overall index (see Figure 9, where three month changes of the common component of the GDP for each countries are reported). Those indicators represent the part of national cycle that is common across the European countries, and therefore may be different from the actual country specific one.

Few general comments are in order. First, the three largest countries seem to share a similar cyclical pattern, whereas the Netherlands and Belgium experienced more peculiar dynamics. Belgium and the Netherlands seem to anticipate the last three turning points, supporting the view that those economies tend to lead the euro area dynamics. This leading property however was not present in the early nineties. Spain seems to have over-performed the other economies in the area across the period, while France is the most closely related to the average behavior. Finally, after 1992 and the German unification there is some evidence of a stronger synchronization among the six euro area economies.

The information provided by the indicator is compared with other two indicators, routinely utilized in the short-run assessment of the economy of the euro area, namely the IFO German business climate and the European Commission euro area industrial climate. The three month changes of those indexes are compared with the indicator in Figure 10 and 11. As expected those two survey indicators are leading the proposed index. However the time lead seems not to be constant at different turning points and, as noted in the previous section, there is a high level of noise-to-signal ratio, making difficult the real time interpretation of those data. This fact is particularly true for IFO, while the European Commission data, though being characterized by lower volatility, presents large oscillation at turning points.

Looking at the most recent dynamic of the indicator, if the information underlying the projection of the euro area coincident index will be confirmed in the near future, there are positive signals that the phase of deceleration of the euro area economic activity, started in the early 2000, might have reached a stop in the second quarter of 2001 (see Figure 12).

# 8 Conclusions

This paper is the result of a joint research project between the Bank of Italy and the CEPR, for the construction of a coincident indicator for the European business cycle. The paper uses a newly constructed monthly data bases of more than 900 time series for the six major economies of the area and proposes a new practical method for the construction of an index which provides "real time" information on the state of the economy. The index is estimated on the basis of an econometric method that allows to exploite all the information potentially available to policy makers and to promptly update results as new data are released.

On the basis of our index we establish cyclical dating and assess historical characteristics of expansions and recessions. We conclude that from the end of the eighties to date we experienced four complete cycles with the last peak occurring at the end of 1999. The most recent developments suggest that the period of deceleration of the euro economy has come to a softening. A comparison of our index with the IFO and the European Commission business climate indicator suggests that there are advantages in the pooling of information implied by our method. Our index is less volatile, in particular around turning points: this should indicate a higher signal to noise ratio.

The method we propose for the construction of the index allows, as a by-product, to study the synchronization and covariance of the elements of the panel in order to understand the sectoral and national structure of the aggregate cycle. These results highlight the performance of key indicators such as survey data and industrial production, which are widely used in short term analysis as leading indicators. We show that survey data, although mostly leading, are noisy signals of economic activity; as a consequence, new releases have to be read with caution. Financial variables are also mostly leading, but are often poorly correlated with the aggregate cycle.

# A Data and Data Treatment

## A.1 Data

This appendix describes the main guidelines followed setting up the database and, in particular, each of the blocks into which it has been split. As already noticed in section 3, the general strategy adopted was to collect data for most recent years from Eurostat and the European Commission, whenever they were available: these sources should grant a proper statistical harmonization across countries for the information released. Nevertheless, many other international sources and national institutions were consulted in order to construct a dataset that gives a comprehensive account of the economic phenomena emerging from the largest European countries (see Table A1 for details). In these cases attention was paid to gather data of homogeneous quality. Finally the database has been organized in a way that allows monthly updates of all the time series therein: this is obviously a fundamental requirement in view of monthly releases of the cyclical indicators built upon it.

The trading days and seasonally adjusted series on Industrial Production were extracted from the Eurostat database, organized according to the NACE Rev. 1 classification method and generally covering a sufficiently long time span. Nevertheless in some cases earlier data were collected from the OECD database, according to the ISIC classification; the Eurostat time series were then linked backward trying to match definitions and disaggregations as closely as possible. In spite of this, most industrial production time series for The Netherlands and for Belgium start only in the nineties and therefore cannot be used to perform the dynamic factor model estimation.

For producer prices we replicated the sectoral breakdown used for industrial production (NACE Rev.1); in doing so we resorted to the Eurostat database on PPIs and on some national sources, such as ISTAT for Italy and INSEE for France. Consumer price series are the result of a link between the most recent HICP data available from Eurostat, starting in 1995, and a combination of earlier data from either the main economic indicators database of the OECD, or national statistical institutes (ISTAT, INSEE) and Datastream.

The monetary block includes various definitions of money aggregates (M1, M2 and M3) for the largest European economies; besides this, a wide variety of interest rates was gathered covering both short and long term government bonds, bank deposits and bank loans. When available, some spreads between interest rates were included too, especially for the Italian economy. Effective exchange rates were also collected for all of the countries considered, both in real and in nominal terms. The main sources consulted for the variables belonging to this block are the BIS (Bank of International Settlements), the ESCB and some national institutions.

Harmonizing the data collected by national sources, the European Commission monthly provides seasonally adjusted business and households survey results, both for the Euro Area and for each member country. Constructions, retail trade and manufacturing sectors are investigated and the economic sentiment indicator is obtained to synthesise the overall business climate. Time series reporting balances of the answers start in the mid eighties and are regularly updated; some of them regard questions addressed quarterly to economic agents and are therefore not exploited in the present work. National institutions (e.g. IFO for Germany, INSEE for France, ISAE for Italy etc.) survey datasets cover longer time spans and a deeper disaggregation level of economic activities; for these reasons they were included in our database too, in addition to those provided by the European Commission.

Relevant business cycle information can be extracted from data that are not classifiable among the previously described sets of time series. A further group was consequently formed, containing a miscellanea collection of variables concerning many different economic phenomena, such as passenger car and other vehicles registrations, new companies formation, declarations of bankruptcy, share-price indexes, orders, turnovers, construction permits, rail transportations of passengers and goods and many others. Due to the particular nature of these variables, it was not always possible to collect them for each country; as a consequence, this set of series is not perfectly balanced but, nonetheless, proved to be useful.

It has been particularly difficult to obtain labour-market variables satisfying the requirements listed in section 3.1 and needed for the estimation of the model. OECD and BIS databases were consulted, obtaining sufficiently exhaustive information concerning the unemployment in all European countries. Although with a lesser detail, time series on wages and unit labour costs were found, whilst very few information about vacancies are available.

Finally, exports and imports time series - especially regarding consumer, intermediate and capital goods - were extracted from BIS and OECD datasets to constitute the Trade block. More interestingly, this block of data includes time series on the trade volume between each euro area countries and its main commercial partners.

## A.2 Data treatment

The dynamic behavior of the series collected are remarkably non-homogeneous. Most of them are raw series, others have been adjusted to take into account working days effects and, finally, a few are available only in a seasonally adjusted version. Preliminary inspection reveals that our series are not affected by the same kind of non-stationarity. Given the large number of series in the panel, careful individual treatment of non-stationarity was not feasible. Rather, we followed an automatic procedure treating in the same way all the series of a given economic class (e.g. industrial production, consumer prices and so on). Then we checked whether this resulted in an improper treatment of the data, such as over-differencing, incomplete removal of outliers or inadequate seasonal adjustment. When this was the case, if the problem could not be fixed with some *ad hoc* adjustment, the variable was discarded from the dataset.

Our data treament procedure can be detailed in four steps as follows:

1. We detected and removed outliers from each series using Tramo, a procedure developed by Gomez and Maravall;[14] in particular we focused on transitory changes, level shifts and additive outliers. The same procedure allowed to adjust for working days effects, whenever requested. Once these deterministic factors were removed, each series was seasonally adjusted using seasonal dummy variables.[15] To take into account the possibility of changes in the seasonal pattern over time, the dummies were also coupled with a linear time trend. We did not resort to other more sophisticated procedures (e.g. Seats or X12) to avoid the use of bilateral filters, which would imply large revisions in the seasonally adjusted series and therefore in the cyclical indicator.[16]

2. The adjusted data were further inspected to make sure that the procedure described above successfully removed all major irregularities. In a few cases we had to drop time series that even after the first step displayed major breaks or other inconsistencies that could not be accounted for and that were therefore attributed to the poor quality of the data.

3. In order to estimate the cross spectral density matrix the series need to be covariance stationary. The transformation inducing stationarity was applied to each *outliers free* and *seasonally adjusted* series. The first log difference was taken for industrial productions, trade variables, financial series, monetary aggregates and labor market variables; exceptions were made for some series (e.g. unemployment rates) for which first differences of natural values were taken. First difference of natural values was also applied to business and household survey responses and to the vast majority of interest rates. *Real* interest rates and the spreads between long and short term nominal interest rates did not need any transformation.[17] The order of integration of price variables is controversial since the choice between I(1) and I(2) models is not always clear-cut. Given that in the majority of cases an I(1) classification seems appropriate, we decided in favor of considering all price indexes as I(1).

[14] See Gomez, Maravall (1999)

[15] These seasonal dummies were defined in such a way that they sum to zero over each full year.

[16] For the same reason, in the selection of the 951 time series finally used to estimate the cyclical indicator, we preferred to collect raw data to be regressed on seasonal dummies instead of series released by the original sources as already seasonal adjusted. Anyway some variables are not available in a raw version; in these cases series were treated as the raw ones in order to remove any residual seasonality. Although we did not use them, to complete our database we also seasonally adjusted time series applying Tramo-Seats.

[17] In general the stationarity inducing transformation exploited was coherent with the model identified by applying Tramo-Seats to the dataset.

4. Finally the series were normalized subtracting their mean and then dividing for their standard deviation. This standardization is necessary to avoid overweighting series with large variance when estimating the spectral density.

# B Technical details

## B.1 Estimating the covariances of the common components

In the first step of our procedure, we estimate the spectral-density matrix and the covariances of the common components. We start by estimating the spectral-density matrix of $\mathbf{x}_t = \begin{pmatrix} x_{1t} & \cdots & x_{nt} \end{pmatrix}'$. Let us denote the theoretical matrix by $\mathbf{\Sigma}(\theta)$ and its estimate by $\hat{\mathbf{\Sigma}}(\theta)$. The estimation is accomplished by using a Bartlett lag-window of size $M = 18$, i.e. by computing the sample auto-covariance matrices $\hat{\mathbf{\Gamma}}_k$, multiplying them by the weights $w_k = 1 - \frac{|k|}{M+1}$ and applying the discrete Fourier transform:

$$\hat{\mathbf{\Sigma}}_x(\theta) = \frac{1}{2\pi} \sum_{k=-M}^{M} w_k \cdot \hat{\mathbf{\Gamma}}_k \cdot e^{-i\theta k}.$$

The spectra were evaluated at 101 equally spaced frequencies in the interval $[-\pi, \pi]$, i.e. at the frequencies $\theta_h = \frac{2\pi h}{100}$, $h = -50, \ldots, 50$.

Then we performed the dynamic principal component decomposition (see Brillinger, 1981). For each frequency of the grid, we computed the eigenvalues and eigenvectors of $\hat{\mathbf{\Sigma}}(\theta)$. By ordering the eigenvalues in descending order for each frequency and collecting values corresponding to different frequencies, the eigenvalue and eigenvector functions $\lambda_j(\theta)$ and $U_j(\theta)$, $j = 1, \ldots, n$, are obtained. The function $\lambda_j(\theta)$ can be interpreted as the (sample) spectral density of the $j$-th principal component series and, in analogy with the standard static principal component analysis, the ratio

$$p_j = \int_{-\pi}^{\pi} \lambda_j(\theta) d\theta / \sum_{j=1}^{n} \int_{-\pi}^{\pi} \lambda_j(\theta) d\theta$$

represents the contribution of the $j$-th principal component series to the total variance in the system. Letting $\mathbf{\Lambda}_q(\theta)$ be the diagonal matrix having on the diagonal $\lambda_1(\theta), \ldots, \lambda_q(\theta)$ and $\boldsymbol{U}_q(\theta)$ be the $(n \times q)$ matrix $\begin{pmatrix} U_1(\theta) & \cdots & U_q(\theta) \end{pmatrix}$ our estimate of the spectral density matrix of the vector of the common components $\boldsymbol{\chi}_t = \begin{pmatrix} \chi_{1t} & \cdots & \chi_{nt} \end{pmatrix}'$ is given by

$$\hat{\mathbf{\Sigma}}_\chi(\theta) = \boldsymbol{U}(\theta)\mathbf{\Lambda}(\theta)\tilde{\mathbf{U}}(\theta), \tag{5}$$

where the tilde denotes conjugation and transposition. Given the correct choice of $q$, consistency results for the entries of this matrix as both $n$ and $T$ go to infinity can easily be obtained from Forni, Hallin, Lippi and Reichlin (2000).

Following Forni, Hallin, Lippi and Reichlin (2000), we identified the number of common factors $q$ by requiring a minimum amount of explained variance. Precisely, we required $p_q > 0.1$ and $p_{q+1} < 0.1$. We found $q = 4$.

An estimate of the spectral density matrix of the vector of the idiosyncratic components $\boldsymbol{\xi}_t = \begin{pmatrix} \xi_{1t} & \cdots & \xi_{nt} \end{pmatrix}'$ can be obtained as the difference $\hat{\boldsymbol{\Sigma}}_\xi(\theta) = \hat{\boldsymbol{\Sigma}}(\theta) - \hat{\boldsymbol{\Sigma}}_\chi(\theta)$.

Starting from the estimated spectral-density matrix we obtain estimates of the covariance matrices of $\boldsymbol{\chi}_t$ at different leads and lags by using the inverse discrete Fourier transform, i.e.

$$\hat{\boldsymbol{\Gamma}}_{\chi k} = \frac{2\pi}{101} \sum_{h=-50}^{50} \hat{\boldsymbol{\Sigma}}_\chi(\theta_h) e^{i\theta_h k}.$$

Moreover, we compute estimates of the covariance matrices of the cyclical componennt $\boldsymbol{\chi}_t^C = (\chi_{1t}^C, ..., \chi_{nt}^C)'$ by applying the inverse tranform to the frequency band of interest, i.e. $[-2\pi/\tau, 2\pi/\tau]$. Precisely, letting $\boldsymbol{\Gamma}_{\chi^C k} = \mathrm{E}\left(\boldsymbol{\chi}_t^C \boldsymbol{\chi}_{t-k}^{C\prime}\right)$, the corresponding estimate will be

$$\hat{\boldsymbol{\Gamma}}_{\chi^C k} = \frac{2\pi}{2H+1} \sum_{h=-H}^{H} \hat{\boldsymbol{\Sigma}}_\chi(\theta_h) e^{i\theta_h k},$$

where $H$ is defined by the conditions $H/101 > \tau$ and $(H+1)/101 < \tau$.

## B.2 Estimating the static factors

Starting from the covariances estimated in the first step, we estimate the static factors as linear combinations of (the present of) the observable variables $x_{jt}$, $j = 1, \ldots, n$. Indeed, as observed in the main text, the static factors appearing in representation (1), i.e. $u_{ht-k}$, $h = 1, \ldots, q$, $k = 1, \ldots, s$, are not identified without imposing additional assumptions and therefore cannot be estimated. This however is not a problem, since we need only a set of $r = q(s+1)$ variables forming a basis for the linear space spanned by the $u_{ht}$'s and their lags. We can then obtain $\hat{\chi}_{jt}$ by projecting $\chi_{jt}$ on such factors and $\hat{\chi}_{jt}^C$ by projecting $\chi_{jt}^C$ on the leads and the lags of such factors.

Our strategy is to take the first $r$ generalized principal components of $\hat{\boldsymbol{\Gamma}}_{\chi 0}$ with respect to the diagonal matrix having on the diagonal the variances of the idiosyncratic components $\xi_{jt}$, $j - 1, \ldots, n$, denoted by $\hat{\boldsymbol{\Gamma}}_{\xi 0}$. Precisely, we compute the generalized eigenvalues $\mu_j$, i.e. the $n$ complex numbers solving $\det(\boldsymbol{\Gamma}_{\chi 0}^T - z\hat{\boldsymbol{\Gamma}}_{\xi 0}) = 0$, along with the corresponding generalized eigenvectors $V_j$, $j = 1, \ldots, n$, i.e. the vectors satisfying

$$V_j \hat{\boldsymbol{\Gamma}}_{\chi 0} = \mu_j V_j \hat{\boldsymbol{\Gamma}}_{\xi 0},$$

and the normalizing condition

$$V_j \hat{\boldsymbol{\Gamma}}_{\xi 0} V_i' = \begin{cases} 0 & \text{for} \quad j \neq i, \\ 1 & \text{for} \quad j = i. \end{cases}$$

Then we order the eigenvalues in descending order and take the eigenvectors corresponding to the largest $r$ eigenvalues. Our estimated static factors are the generalized principal components

$$v_{jt} = V_j' \mathbf{x}_t, \quad j = 1, \ldots, r.$$

The motivation for this strategy is that, if $\hat{\mathbf{\Gamma}}_{\xi 0}$ is the variance-covariance matrix of the idiosyncratic components (i.e. the $\xi_{jt}$'s are mutually orthogonal), the generalized principal components are the linear combinations of the $x_{jt}$'s having the smallest idiosyncratic-common variance ratio (for a proof see Forni, Hallin, Lippi and Reichlin, 2001). We diagonalize the idiosyncratic variance-covariance matrix since, as shown in the paper cited above, this gives better results under simulation when $n$ is large with respect to $T$ as is the case here.

By using the generalized principal components and the covariances estimated in the first step we can estimate and forecast $\boldsymbol{\chi}_t$. Precisely, setting $\mathbf{V} = (V_1 \ \cdots \ V_r)$ and $\mathbf{v}_t = (v_{1t} \cdots v_{rt})' = \mathbf{V}'\mathbf{x}_t$, our estimate of $\boldsymbol{\chi}_{t+h}$, $h = 0, \ldots, s$, given the information available at time $t$, is

$$\begin{aligned} \hat{\boldsymbol{\chi}}_{t+h} &= \hat{\mathbf{\Gamma}}_{\chi h} \mathbf{V} \left( \mathbf{V}' \hat{\mathbf{\Gamma}}_0 \mathbf{V} \right)^{-1} \mathbf{v}_t \qquad (6) \\ &= \hat{\mathbf{\Gamma}}_{\chi h} \mathbf{V} \left( \mathbf{V}' \hat{\mathbf{\Gamma}}_0 \mathbf{V} \right)^{-1} \mathbf{V}' \mathbf{x}_t. \end{aligned}$$

In Forni, Hallin, Lippi and Reichlin (2001) it is shown that, as both $n$ and $T$ go to $\infty$ in a proper way, $\hat{\boldsymbol{\chi}}_t$ converges in probability, entry by entry, to $\boldsymbol{\chi}_t$, and $\hat{\boldsymbol{\chi}}_{t+h}$ converges to the theoretical projection of $\boldsymbol{\chi}_{t+h}$ on the present and the past of $u_{1t}, \ldots, u_{qt}$.

### B.3 Estimating the cyclical part of the common components

Finally we estimate the cyclical common components $\chi_{jt}^C$ by using the covariances estimated in the first step in order to project $\chi_{jt}^C$ on the present and $m$ leads and lags of the estimated static factors.

Set $\mathbf{V}_t = (\mathbf{v}_{t+m}' \cdots \mathbf{v}_t' \cdots \mathbf{v}_{t-m}')'$ and

$$\mathbf{W} = \left( \underbrace{\begin{array}{cccc} \mathbf{V} & 0_{n\times r} & \cdots & 0_{n\times r} \\ 0_{n\times r} & \mathbf{V} & \cdots & 0_{n\times r} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n\times r} & 0_{n\times r} & \cdots & \mathbf{V} \end{array}}_{2m+1 \text{ blocks}} \right).$$

Moreover, set $\mathbf{X}_t = (\mathbf{x}_{t+m}' \cdots \mathbf{x}_t' \cdots \mathbf{x}_{t-m}')'$, so that $\mathbf{V}_t = \mathbf{W}'\mathbf{X}_t$. The sample variance-covariance matrix of $\mathbf{X}_t$ is

$$\mathbf{M} = \begin{pmatrix} \hat{\mathbf{\Gamma}}_0 & \hat{\mathbf{\Gamma}}_1 & \cdots & \hat{\mathbf{\Gamma}}_{2m} \\ \hat{\mathbf{\Gamma}}_1' & \hat{\mathbf{\Gamma}}_0 & \cdots & \hat{\mathbf{\Gamma}}_{2m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{\Gamma}}_{2m}' & \hat{\mathbf{\Gamma}}_{2m-1}' & \cdots & \hat{\mathbf{\Gamma}}_0 \end{pmatrix},$$

while $\mathrm{E}\left(\boldsymbol{\chi}_t^C \boldsymbol{X}_t'\right)$ can be estimated by

$$\mathbf{R} = \begin{pmatrix} \hat{\mathbf{\Gamma}}_{\chi^C m}' & \cdots & \hat{\mathbf{\Gamma}}_{\chi^C 0}' & \cdots & \hat{\mathbf{\Gamma}}_{\chi^C m} \end{pmatrix}.$$

Our estimate of the common cyclical components is then

$$\hat{\chi}_t^C = \mathbf{RW}\left(\mathbf{W}'\mathbf{MW}\right)^{-1}\mathbf{W}'\mathbf{X}_t. \tag{7}$$

At the end of the sample, i.e. from $T-m$ onward, we have the problem that $\mathbf{x}_{T+h}$, $h > 0$, is not available. Our estimate is then obtained by substituting our forecast of the common components $\hat{\boldsymbol{\chi}}_{T+h}$, in place of $\mathbf{x}_{T+h}$ and applying formula (7).

## B.4 Treatment of the end-of-sample unbalance

Let us assume that $T$ is the last date for which we have observations for all of the variables in the data set and that there are some variables for which we have observations until dates $T+1, \ldots, T+w$. Without loss of generality we can then reorder the variables in such a way that

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{x}_t^{1\prime} & \mathbf{x}_t^{2\prime} & \cdots & \mathbf{x}_t^{w\prime} \end{pmatrix},$$

where $\mathbf{x}_{jt}$, $j = 1, ..., w$, is such that the last available observation reefers to $T+j-1$. Correspondingly, the sample covariance matrices $\hat{\mathbf{\Gamma}}_k$ are partitioned as follows

$$\hat{\mathbf{\Gamma}}_k = \begin{pmatrix} \hat{\mathbf{\Gamma}}_k^{11} & \hat{\mathbf{\Gamma}}_k^{12} & \cdots & \hat{\mathbf{\Gamma}}_k^{1w} \\ \hat{\mathbf{\Gamma}}_k^{21} & \hat{\mathbf{\Gamma}}_k^{22} & \cdots & \hat{\mathbf{\Gamma}}_k^{2w} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{\Gamma}}_k^{w1} & \hat{\mathbf{\Gamma}}_k^{w2} & \cdots & \hat{\mathbf{\Gamma}}_k^{ww} \end{pmatrix}.$$

A similar partition holds for $\hat{\mathbf{\Gamma}}_{\chi k}$.

Our idea is simply to shift the variables in such a way to retain, for each one of them, only the most updated observation, and compute the generalized principal components for the realigned vector. In such a way we are able to get information on the factors $u_{hT+j}$, $h = 1, \ldots, q$, $j = 1, \ldots, w$, and to exploit it in prediction.

Precisely, we set

$$\mathbf{x}_t^* = \begin{pmatrix} \mathbf{x}_t^{1\prime} & \mathbf{x}_{t+1}^{2\prime} & \cdots & \mathbf{x}_{t+w-1}^{w\prime} \end{pmatrix}.$$

Notice that the sample covariance matrices of $\mathbf{x}_t^*$ are then

$$\hat{\mathbf{\Gamma}}_k^* = \begin{pmatrix} \hat{\mathbf{\Gamma}}_k^{11} & \hat{\mathbf{\Gamma}}_{k-1}^{12} & \cdots & \hat{\mathbf{\Gamma}}_{k-w+1}^{1w} \\ \hat{\mathbf{\Gamma}}_{k+1}^{21} & \hat{\mathbf{\Gamma}}_k^{22} & \cdots & \hat{\mathbf{\Gamma}}_{k-w+2}^{2w} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{\Gamma}}_{k+w-1}^{w1} & \hat{\mathbf{\Gamma}}_{k+w-2}^{w2} & \cdots & \hat{\mathbf{\Gamma}}_k^{ww} \end{pmatrix}$$

and the matrices $\hat{\mathbf{\Gamma}}_{\chi k}^*$ are defined in the same way. Then we compute the matrix $\mathbf{V}^*$ of the generalized eigenvectors of $\hat{\mathbf{\Gamma}}_{\chi k}^*$ with respect to $\hat{\mathbf{\Gamma}}_{\xi k}$ (the latter matrix is diagonal and thefore is the same for $\mathbf{x}_t$ and $\mathbf{x}_t^*$) and obtain forecasts of $\boldsymbol{\chi}_{T+h}^*$ as in equation (7):

$$\hat{\boldsymbol{\chi}}_{T+h}^* = \hat{\mathbf{\Gamma}}_{\chi h}^* \mathbf{V}^* \left( \mathbf{V}^{*\prime} \hat{\mathbf{\Gamma}}_0^* \mathbf{V}^* \right)^{-1} \mathbf{V}^{*\prime} \mathbf{x}_T^*.$$

Finally we use the forecasts in $\hat{\boldsymbol{\chi}}_{T+h}^*$, $h = 1, \dots$ to replace missing data and to get the forecasts of $\boldsymbol{\chi}_{T+h}$, $h > w$, which are needed to apply (7).

| Data bank | series |
|---|---|
| Eurostat (various collections) | 98 |
| Main Economic Indicators (OECD) | 44 |
| Directions of Trade (OECD) | 70 |
| ECB Short Term Statistics and others | 297 |
| Bank of Italy database | 101 |
| ISAE | 51 |
| Bank of Iternational Settlements | 191 |
| Datastream | 99 |

Table A1 - Data sources

| Economic sector | Number of series | % |
|---|---|---|
| Industrial production | 176 | 18.5 |
| Production prices | 94 | 9.9 |
| Consumer prices | 39 | 4.1 |
| surveys | 237 | 24.9 |
| Monetary aggregates (nominal) | 25 | 2.6 |
| Monetary aggregates (real) | 21 | 2.2 |
| Real interest rates | 12 | 1.3 |
| Nominal interest rates | 78 | 8.2 |
| Spreads | 6 | 0.6 |
| Other financial variables | 44 | 4.6 |
| Labour market | 37 | 3.9 |
| of which: | | |
| wages | 19 | 2.0 |
| unemployment | 9 | 0.9 |
| Other indicators | 58 | 6.1 |
| Trade | 98 | 10.3 |
| Non Euro countries | 19 | 2.0 |
| GDP | 7 | 0.7 |
| All variables | 951 | 100 |

Table A2 - Data by economic sector

| Country | Number of series | % |
|---|---|---|
| Germany | 212 | 23.2 |
| France | 135 | 14.8 |
| Italy | 190 | 20.8 |
| Spain | 129 | 14.1 |
| Belgium | 121 | 12.2 |
| Netherlands | 111 | 13.3 |
| Euro-area | 14 | 1.5 |
| Total | 912 | 100 |

Table A3 - Data by country

| ***Country*** | ***Total*** | ***Leading*** | ***Coincident*** | ***Lagging*** | ***Total*** | ***Leading*** | ***Coincident*** | ***Lagging*** |
|---|---|---|---|---|---|---|---|---|
| ***GERMANY*** | 212 | 54 | 93 | 65 | 24% | 25% | 44% | 31% |
| ***FRANCE*** | 135 | 25 | 69 | 41 | 15% | 19% | 51% | 30% |
| ***ITALY*** | 190 | 55 | 70 | 65 | 21% | 29% | 37% | **34%** |
| ***SPAIN*** | 129 | 34 | 41 | 54 | 14% | 26% | 32% | **42%** |
| ***BELGIUM*** | 121 | 40 | 51 | 20 | 13% | **33%** | 42% | 17% |
| ***NETHERLAND*** | 111 | 40 | 53 | 28 | 12% | **36%** | 48% | 25% |
| ***TOTAL*** | **898** | **248** | **377** | **273** | **100** | **28%** | **42%** | **30%** |

Table A4 - Cyclical behavior breakdown by countries.

| SECTOR | N° of variables | N° of countercyclical variables | | Leading | | Coincident | | Lagging | | Explained variance | Mean abs. Correlation | Average time shift |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***INDUSTRIAL PRODUCTION*** | 176 | 14 | 8.0% | 51 | 29.0% | 93 | 52.8% | 32 | 18.2% | 0.407 | 0.573 | -0.6 |
| ***PRODUCTION PRICES*** | 94 | 17 | 18.1% | 14 | 14.9% | 41 | 43.6% | 39 | 41.5% | 0.698 | 0.414 | 2.8 |
| ***CONSUMPTION PRICES*** | 39 | 12 | 30.8% | 12 | 30.8% | 6 | 15.4% | 21 | 53.8% | 0.633 | 0.417 | 4.3 |
| ***E. C. SURVEYS*** | 237 | 28 | 11.8% | 97 | 40.9% | 106 | 44.7% | 34 | 14.3% | 0.517 | 0.539 | -0.9 |
| ***MONETARY AGGREGATES*** | 25 | 8 | 32.0% | 7 | 28.0% | 6 | 24.0% | 12 | 48.0% | 0.571 | 0.574 | 4.4 |
| ***MONETARY AGGREGATES - REAL*** | 21 | 7 | 33.3% | 9 | 42.9% | 3 | 14.3% | 9 | 42.9% | 0.553 | 0.538 | 0.7 |
| ***SPREAD*** | 6 | 2 | 33.3% | 3 | 50.0% | 1 | 16.7% | 2 | 33.3% | 0.661 | 0.435 | -3.8 |
| ***REAL INTEREST RATES*** | 12 | 8 | 66.7% | 7 | 58.3% | 1 | 8.3% | 4 | 33.3% | 0.771 | 0.421 | -0.4 |
| ***INTEREST RATES*** | 78 | 0 | 0.0% | 0 | 0.0% | 47 | 60.3% | 31 | 39.7% | 0.749 | 0.783 | 2.6 |
| ***OTHER FINANCIAL VARIABLES*** | 44 | 22 | 50.0% | 16 | 36.4% | 6 | 13.6% | 22 | 50.0% | 0.606 | 0.482 | 0.4 |
| ***LABOUR MARKET*** | 37 | 15 | 40.5% | 5 | 13.5% | 11 | 29.7% | 21 | 56.8% | 0.634 | 0.667 | 4.5 |
| ***of which:*** | | | | | | | | | | | | |
| ***WAGES*** | *19* | *11* | *57.9%* | *8* | *42.1%* | *5* | *26.3%* | *6* | *31.6%* | *0.568* | *0.643* | *-0.3* |
| ***UNEMPLOYMENT*** | *9* | *8* | *88.9%* | *2* | *22.2%* | *2* | *22.2%* | *5* | *55.6%* | *0.691* | *0.824* | *1.6* |
| ***OTHER INDICATORS*** | 58 | 6 | 10.3% | 18 | 31.0% | 24 | 41.4% | 16 | 27.6% | 0.381 | 0.486 | 1.1 |
| ***TRADE*** | 98 | 6 | 6.1% | 14 | 14.3% | 46 | 46.9% | 38 | 38.8% | 0.475 | 0.544 | 2.0 |
| ***UK, JAPAN and US*** | 19 | 4 | 21.1% | 5 | 26.3% | 6 | 31.6% | 8 | 42.1% | 0.600 | 0.577 | 1.4 |
| ***GDP*** | 7 | 0 | 0.0% | 0 | 0.0% | 7 | 100.0% | 0 | 0.0% | 0.693 | 0.896 | -0.4 |
| ***ALL VARIABLES*** | **951** | **149** | **15.7%** | **258** | **27.1%** | **404** | **42.5%** | **289** | **30.4%** | **0.544** | **0.577** | **0.9** |

Table A5 - Cyclical behaviour breakdown by sector

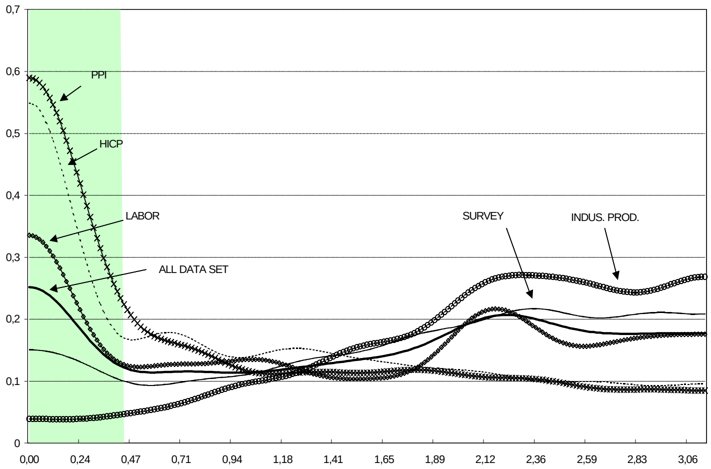


Figure 1 - Average spectral shape.

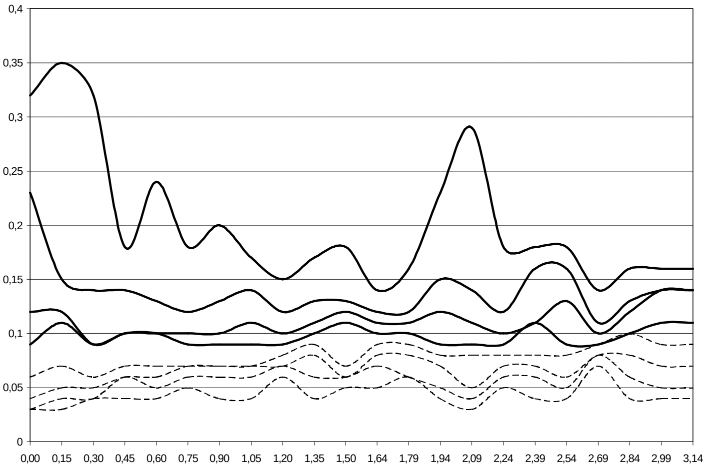


Figure 2 - First eight dynamic eigenvalues.

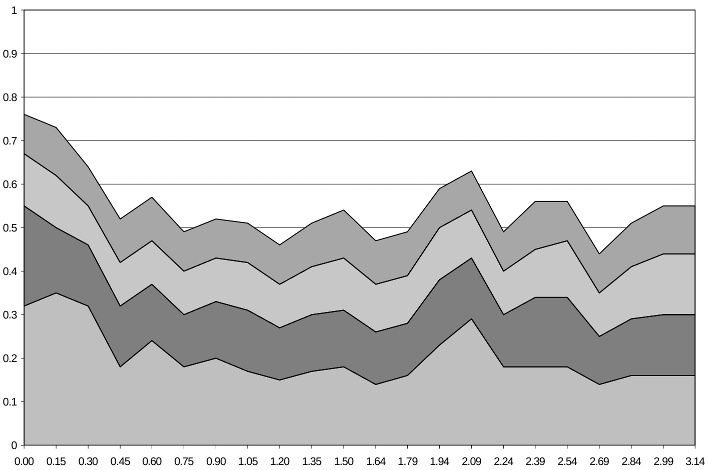


Figure 3 - Cumulated share of variance explained by the first four dynamic eigenvalues.

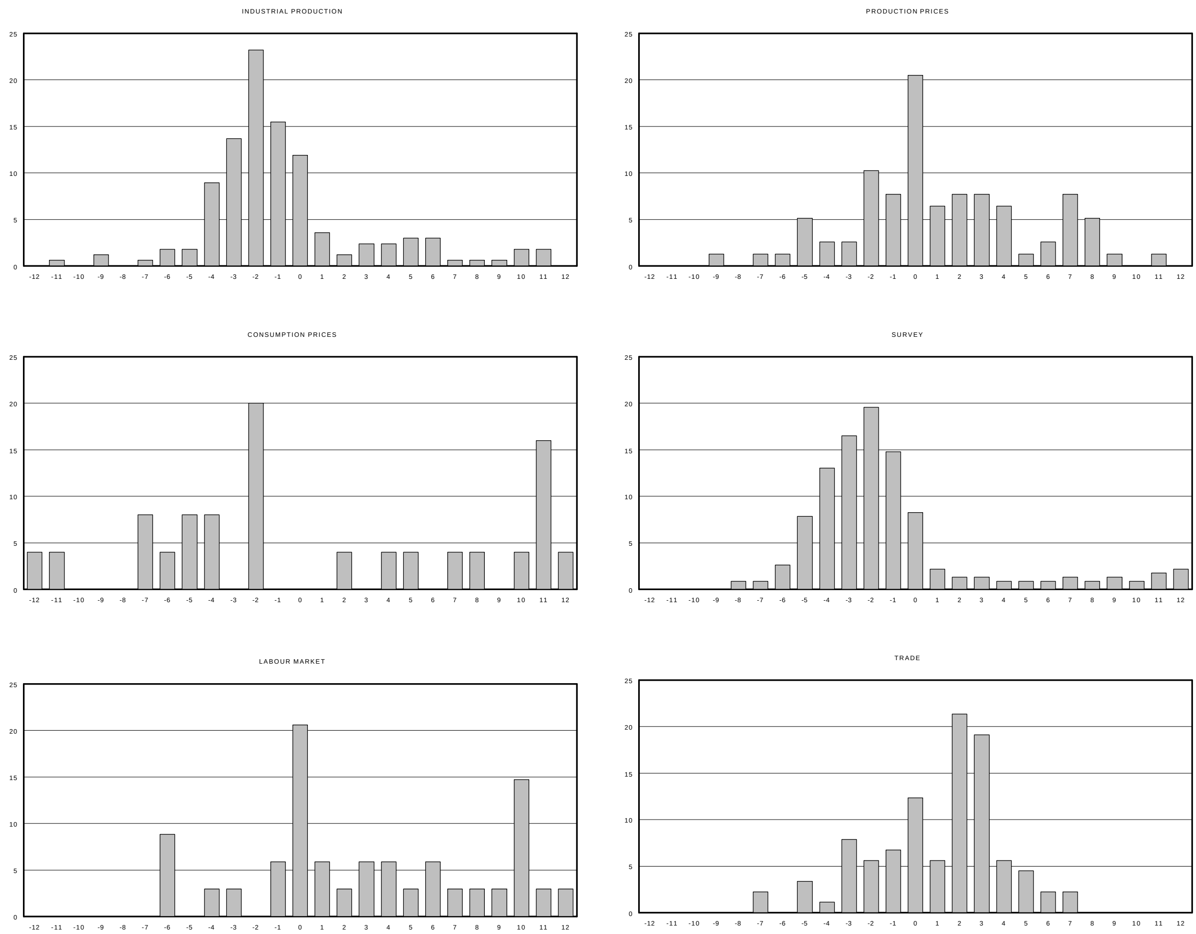


Figure 4 - Displacement Distribution.

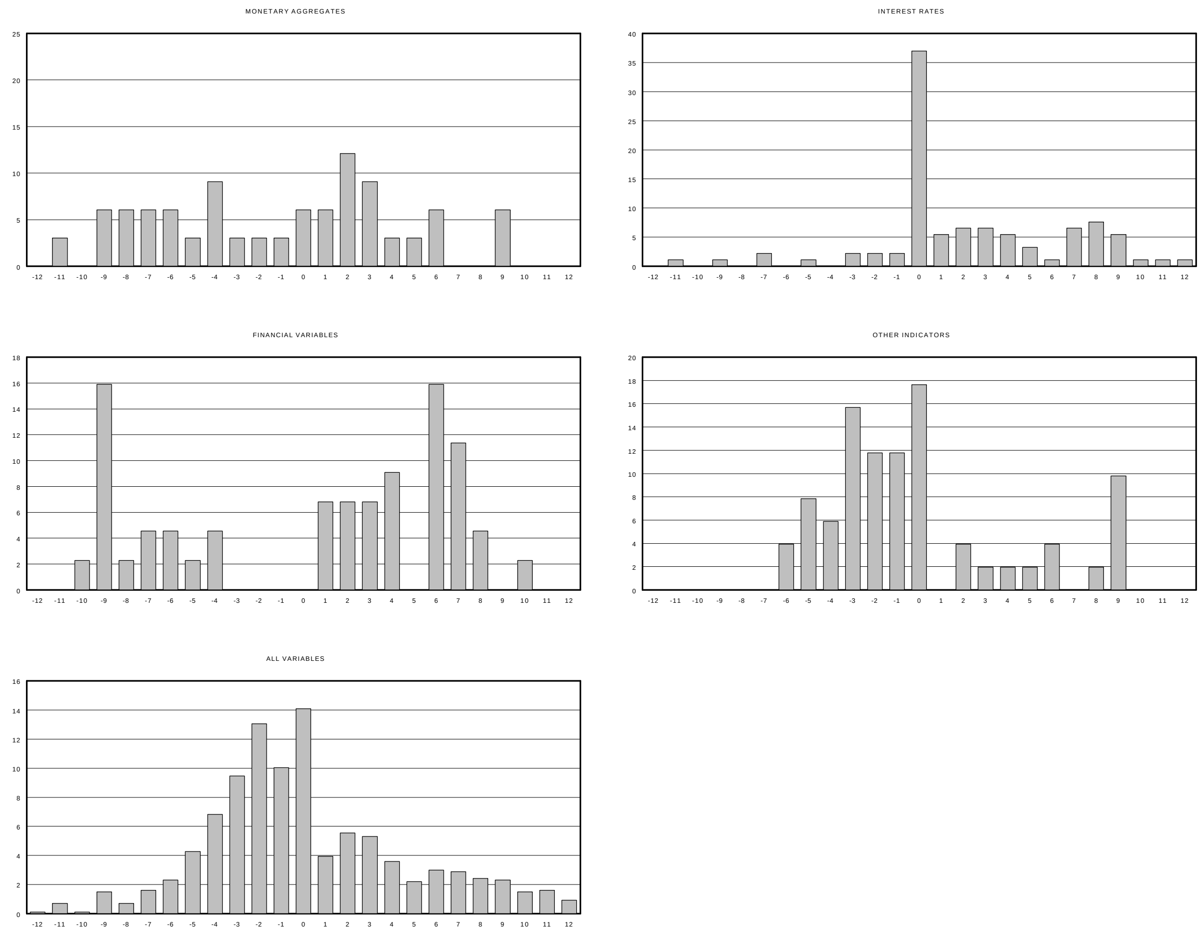


Figure 4 - Displacement Distribution.

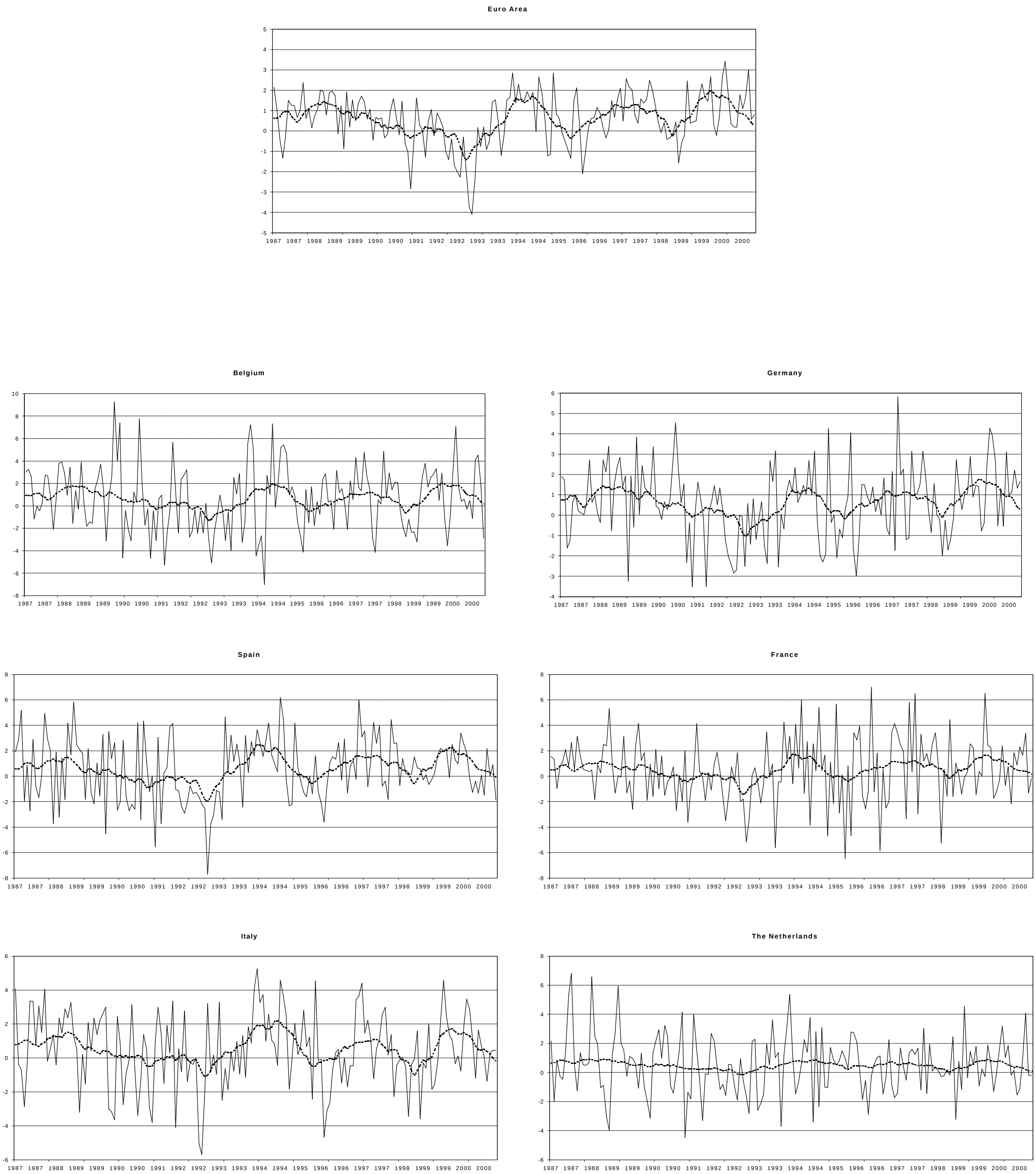


Figure 5 - Three month changes of the total industrial index of production (excluding construction) with the relative common component.


Figure 6 - Euro Area Coincident Indicator.

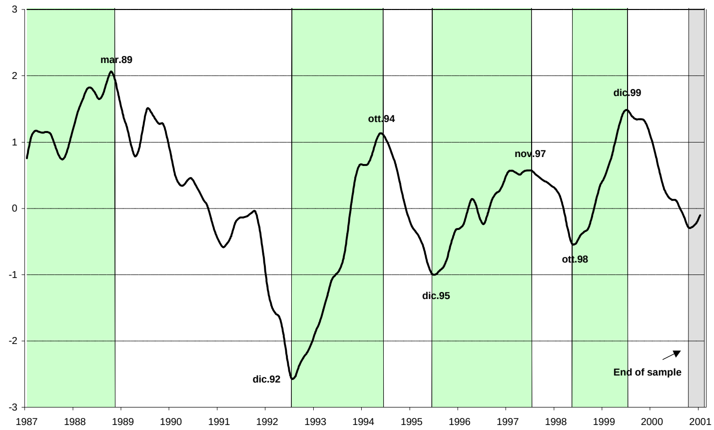


Figure 7 - Euro Area Coincident Indicator and business cycle dating.

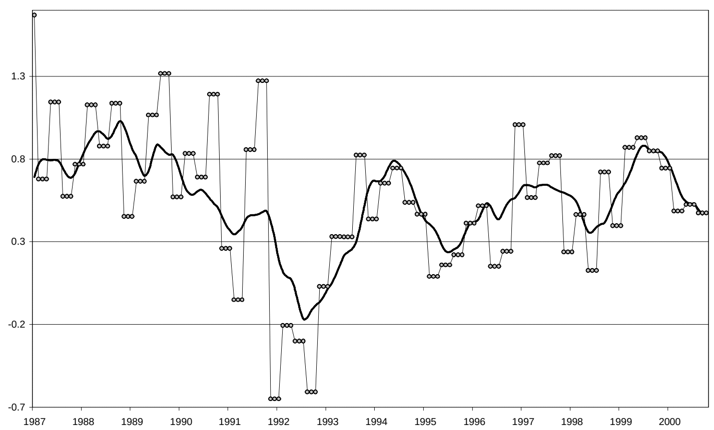


Figure 8 - Euro Area Coincident Indicator and the q-o-q changes of the Euro Area GDP.

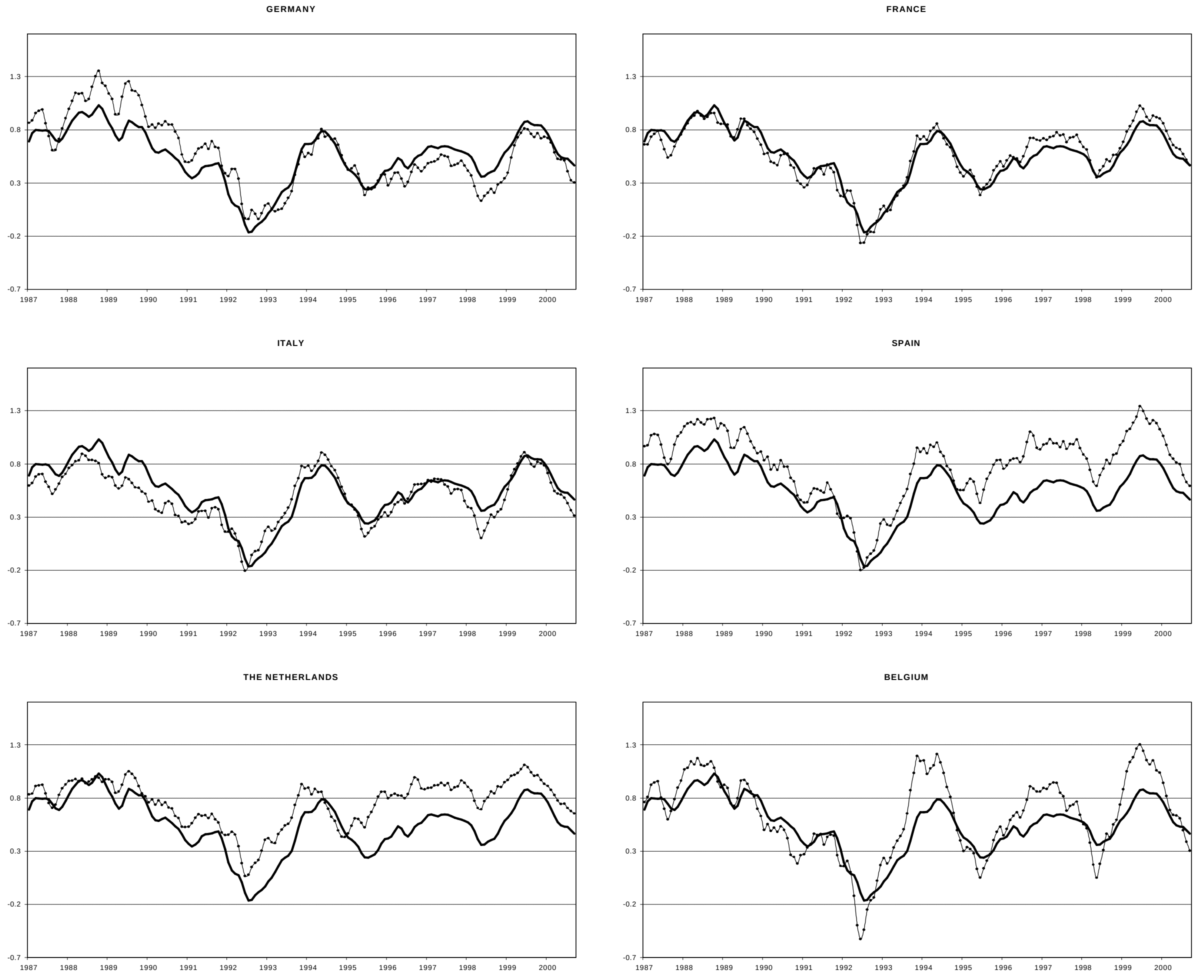


Figure 9 - Euro Area coincident indicator and common components of national GDPs.

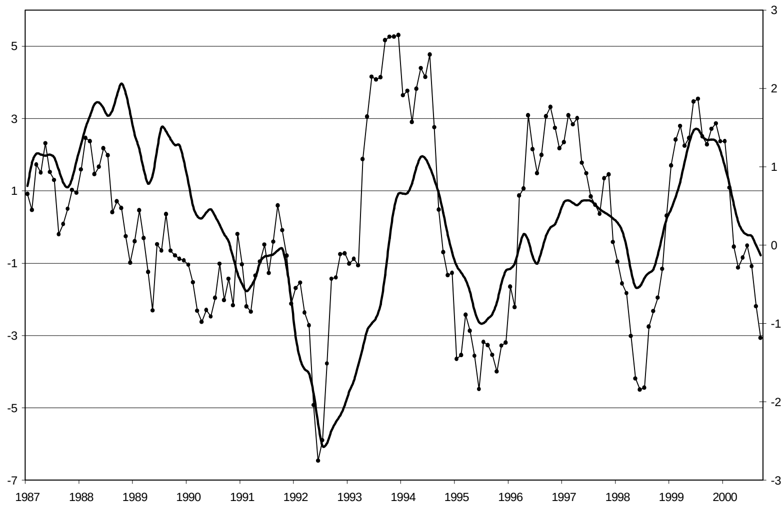


Figure 10 - Euro area coincident indicator and the euro area business climate survey.

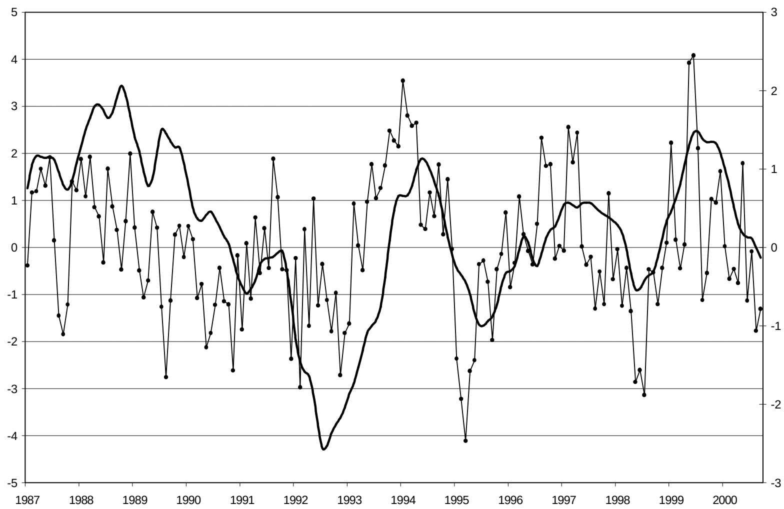


Figure 11 - Euro area coincident indicator and the IFO German business climate.

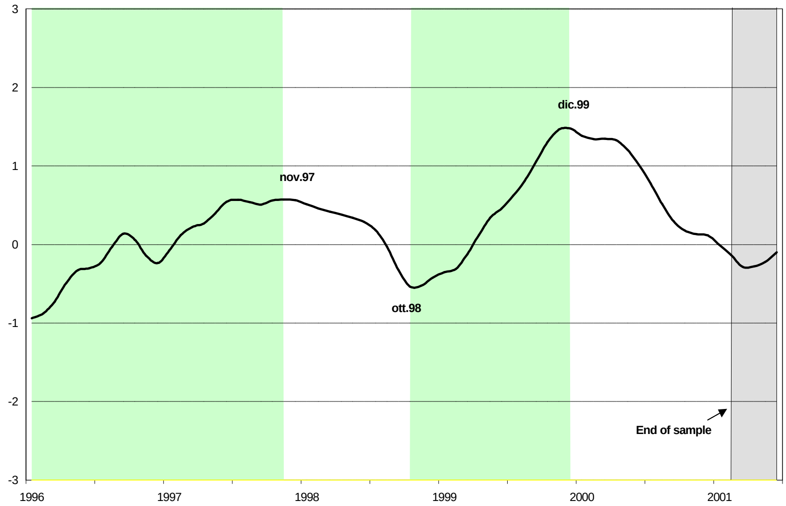


Figure 12 - Euro Area Coincident Indicator

## References

[1] Altissimo, F. Bassanetti, A. Cristadoro, R. Reichlin, L. and Veronese, G. (2001) "The construction of coincident and leading indicators for the euro area business cycle" in Banca d'Italia, Temi di discussione n. XXX.

[2] Artis, M. Krolzig, H.M., and Torr,. J., (1999), "The European Business Cycle", European University Institute Working Paper, N.24.

[3] Baxter, A. and King, (1999), "Measuring Business Cycles Approximate Band-Pass Filters for Economic Time Series", *International Economic Review.*

[4] Bowden, R.J. and V. Martin (1993) "Reference Cycles in the Time and Frequency Domain: Duality Aspects of the Business Cycle", in P.C.B. Phillips (ed.) Models, Methods and Applications of Econometrics. Essays in Honor of A.M. Bergstrom, Cambridge Mass.: Basil Blackwell.

[5] Brillinger, D. R. (1981), Time Series Data Analysis and Theory, New York: Holt, Rinehart and Winston.

[6] Brockwell, P.J. and R. A. Davis (1987), Time Series: Theory and Methods, New York: Springer-Verlag.

[7] Bry G. and Boschan C. (1971), "Cyclical Analysis of Times Series: Seelected Procedures and Computer Programs", NBER, Technical Paper n. 20.

[8] Burns, A.F., and W.C. Mitchell (1946), Measuring Business Cycles. New York: NBER.

[9] Cheung, Y., Westermann, F., (1999), "Output Dynamics of the G7 countries—Stochatic Trends and Cyclical Movements", CES-ifo Working Paper, N.220.

[10] Christiano, L.J. and T.J. Fitzgerald (1999), "The Band Pass Filter", *NBER Working Paper* N.7257.

[11] Dickerson, A.P., Gibson,H.D. and Tsakalotos, E., (1998), "Business Cycle Correspondence in the European Union", *Empirica* 25: 51-77, 1998.

[12] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000) "The Generalized Factor Model: Identification and Estimation", *The Review of Economic and Statistics*, 82(4), 540-554.

[13] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2001) "The Generalized Factor Model: One-Sided Estimation and Forecast", mimeo

[14] Forni, M. and Lippi, M., (2000) "The Generalized Factor Model: Representation Theory", *Econometric Theory*, forthcoming.

[15] Forni M. and Reichlin, L. (2001), "Federal policies and local economies: Europe and the US", *European Economic Review*, No. 45, pp. 109-134.

[16] Gomez, V., Maravall, A., (1996) "Programs TRAMO and SEATS, Instructions for the user (Beta version: September 1996)", Working Paper No. 9628 Bank of Spain.

[17] Lumsdaine, R.L., Prasad, E.S., (1999), "Identifying the Common Component in International Economic Fluctuations: A New Approach", IMF Working Paper, N.154.

[18] Marchetti, D.J., Parigi, G., (2000), "Energy Consumption, Survey Data and the Prediction of Industrial Production in Italy: A Comparison and Combination of Different Models", Journal of Forecasting, 19, 419-440.

[19] McDermott, C.J., Scott, A. (2000), "Concordance in Business Cycles", IMF Working Paper, N.37.

[20] Geweke, J. (1977) "The Dynamic Factor Analysis of Economic Time Series", in D.J. Aigner and A.S. Golberger (eds.) Latent Variables in Socio-Economic Models, Amsterdam, North-Holland, Ch. 19.

[21] Marcellino, M., J.H. Stock and M.W. Watson (2000), Macroeconomic forecasting in the Euro area: country specific versus area-wide information, unpublished manuscript.

[22] Sargent, T.J. (1987), Macroeconomic Theory, Second edition, Academic Press.

[23] Sargent, T.J. and Sims, C. A. (1977) "Business Cycle Modelling Without Pretending to Have Too Much A Priori Economic Theory" in Sims, C.A. (ed.) New Methods in Business Research, Minneapolis: Federal Reserve Bank of Minneapolis.

[24] Stock, J.H. and Watson, M.H. (1989) "New Indexes of Coincident and Leading Economic Indicators", NBER Macroeconomic Annual 1989, 351-94.

[25] Stock, J.H. and Watson, M.H. (1999) "Diffusion indexes", NBER Working Paper N.6702.

[26] Zarnowitz, V. (1992) Business Cycles. Theory, History, Indicators and Forecasting, Chicago: The University of Chicago Press.