BANCA D'ITALIA

Temi di discussione

del Servizio Studi

The construction of coincident and leading indicators for the euro area business cycle

by F. Altissimo, A. Bassanetti, R. Cristadoro, L. Reichlin and G. Veronese

Number 434 - December 2001

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

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

This paper is part of the research project "Monitoring the Euro Area Business Cycle". As the others which are part of the project, it was presented and discussed in an internal seminar and at an international conference, held at the Bank of Italy in Rome, 7-8 September 2001.

THE CONSTRUCTION OF COINCIDENT AND LEADING INDICATORS FOR THE EURO AREA BUSINESS CYCLE

by Filippo Altissimo*, Antonio Bassanetti**, Riccardo Cristadoro**, Lucrezia Reichlin*** and Giovanni Veronese**

Abstract

This paper investigates the business cycle properties of the euro area and computes a coincident and a leading indicator of economic activity. We accomplish this by applying the newly introduced generalized factor model to a properly constructed and harmonized data set of short term statistics of the euro area (794 monthly series). 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 byproduct 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, O30.

Keywords: business cycle, dynamic factor model.

---

* Bank of Italy, Economic Research Department and CEPR.

** Bank of Italy, Economic Research Department.

***Université Libre de Bruxelles, ECARES and CEPR.

Contents

1. Introduction[1]

The creation of the European Monetary Union (EMU) has given rise to new challenges for macroeconomic analysis, particularly in view of the policy problems created by the new economic and institutional environment. Also in short term analysis a growing attention has been devoted to euro-wide economic developments, mainly as a consequence of the presence of new actors that have to base their action on the European rather than national macroeconomic situation. The European Central Bank bases its monetary policy on euro wide economic developments and the official target of price stability is given a quantitative content in terms of the monetary union Harmonized Index of Consumer Prices (HICP), whose year on year growth rate should not exceed 2 per cent in the medium term. The two pillar strategy followed by the ECB implies that the Central Bank should monitor a large number of indicators with the aim of obtaining a reliable picture of the current economic situation within the area as well as of its future developments. Even if growth prospects are not a direct concern of the monetary authority they can influence policy decisions since they have an impact on prices (our analysis shows that both consumer and producer prices are strongly pro-cyclical). Furthermore the Treaty establishing the European Community (EC) states that, without prejudice for the price stability, the monetary policy should be oriented to favor other objectives of the EC.[2] In this context, a growing body of empirical literature has analysed the business cycle properties of euro area countries trying to discover if they share a similar behaviour and whether it is appropriate to study a European-wide cycle. The development of reliable synthetic indicators of the business situation in the euro area would be an important input in the policy decision making process.

---

[1] 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. Email: lreichli@ulb.ac.be; altissimo.filippo@insedia.interbusiness.it; cristadoro.riccardo@insedia.interbusiness.it.

[2] The article 105(1) of the Treaty say that "*The primary objective of the ESCB shall be to maintain price stability. Without prejudice to the objective of price stability, the ESCB shall support the general economic policies in the Community with a view to contributing to the objectives of the Community as laid down in Article 2*", where Article 2 refers to "*sustainable and non-inflationary growth*". For a general discussion of the monetary policy strategy within the European System of Central Banks (ESCB) see the ECB Monthly Bulletin of January 1999.

In this paper we assess the existence of a common cycle for the euro area and we characterize it building area wide coincident and leading business cycle indicators. We also investigate the leading 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. The construction of the coincident and leading indexes requires, on the one hand, a large dataset of macroeconomic time series covering a wide range of economic phenomena for a sufficient number of years; on the other, an appropriate methodology to synthesize efficiently the information contained in a large array of data. The two synthetic indexes are built on monthly statistics and apply a recently developed methodology by Forni, Hallin, Lippi and Reichlin (2000, 2001; FHLR here on) that gives a rigorous foundation to the heuristic NBER approach. The application of this new methodology to monthly data in the study of the euro area business cycle constitutes an absolute novelty.

Since our goal is to extract information from a very large dataset, we need to use a method that is appropriate when the number of data in the panel is possibly larger than the number of time observations. Here we use the *generalized dynamic factor model* that is based on a generalization of the static principal components approach. The basic insight of dynamic factor models is the following. Due to strong comovements between macroeconomic time series, the dynamics of each variable can be represented as the sum of a part explained by a small number of common components and an orthogonal idiosyncratic residual. The generalized dynamic factor model reconciles dynamic factor analysis with dynamic principal components and the estimator is constructed so as to take into account dynamic differences between time series by appropriately weighting leading and lagging variables. This feature allows to estimate the model on all available variables – leading, coincident and lagging – without needing to pre-classify them a priori. The method differs from the previous literature which defines the coincident indicator as a common factor extracted from an index model estimated on a small number of coincident variables which are identified, prior to estimation, by heuristic criteria (see Stock and Watson, 1989).

This methodology has been applied to a properly constructed 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 since up to now there is no euro-area databank with a coverage comparable to those available for the US. Furthermore we selected the statistics according to criteria of 'minimum harmonization' to enhance cross-country comparability. In particular, wherever possible, we maintained a common sectorial breakdown across countries.

The analysis conducted can be summarized in the following steps. First we estimated the unobserved common component of each series in the panel. This allowed us to "clean" variables from measurement errors as well as local and/or variable-specific components. Second, we constructed the coincident indicator for the EMU, defined as a weighted average of the common ("cleaned") components of countries' GDP's. This index is therefore constructed taking into account cross-correlations within and across countries and summarizes that part of the dynamics of GDP which, being the "most" correlated with the rest of the economy, is of interest for measuring the EMU-aggregate business cycle. On the basis of this coincident indicator we propose a dating of the euro area business cycle. A first distinguishing feature that emerges from this analysis is the high number of cyclical upturns and downturns that have characterized the European economy in the nineties as compared with the uninterrupted expansion experienced, over the same period, by the US.

The information on the leading-lagging relation of each variable in the panel and the coincident indicator can then be recovered from the estimates of contemporaneous and lagged cross-covariances between the common component of each variable and the coincident indicator. This enables us to identify the leading variables in all EMU economies. The latter then can be aggregated to construct a leading indicator. The leading indicator can finally be used to forecast the coincident indicator.

A by-product of the analysis is an evaluation of the degree of "commonality" of the dynamics of different countries and variables with respect to the coincident indicator. Moreover, we establish the basic facts on the leading-lagging structure or sectorial and national cycles.

The paper is organized as follows: in section 2 we describe the theoretical basis of the generalized factor model. In the third section we present the database used for the analysis of the business cycle in the euro area, stressing the wide range of indicators contained and the treatment of the data prior to the analysis with the Dynamic Principal Components (DPC) method. The fourth section investigates the presence of significant movements at cyclical periodicity in our monthly macroeconomic variables and whether these are common across countries and sectors, thus resulting in a euro-area business cycle. In the fifth section the business cycle coincident indicator for the euro area is constructed and on this basis a dating for the business cycle is proposed. The sixth section presents a detailed analysis of the cyclical behaviour of different sectors of the national economies, focusing in particular on the leading properties of some series, most notably those belonging to the European Commission surveys and commonly used in short term analysis. The next section, building on the findings of the previous one, proposes a leading index for the euro area business cycle. The eighth section shows the results of an in-sample exercise devoted to the assessment of the size of revision errors for the indicators estimated in real time. As a by-product of this, we can construct confidence bounds around our coincident and leading indexes, taking into account the uncertainty due to the bilateral nature of the filters. The last section concludes the paper highlighting the main results and pointing to directions for further research.

2. Large dataset in presence of a small number of common factors

The basic idea of our approach is that the movements observed in a large set of macroeconomic time series, like the one that we want to analyze, are generated by few common sources. The macroeconomic time series can therefore be thought of as being guided by a small number of common shocks plus an idiosyncratic shock for each variable, whose impact in the aggregate vanishes. When investigating movements that are common across the dataset we want to disregard the part of the dynamics of the series that is due to measurement errors and idiosyncratic shocks. Index models that formalize this idea have been introduced in the literature by Sargent and Sims (1977) and Geweke (1977). Here we briefly present a more general version of those models, known as generalized factor model due to Forni, Hallin, Lippi and Reichlin (2000) on which the empirical part of the paper is based, in doing this we follow closely their presentation.

2.1 *Cleaning the variables through dynamic principal components*

A crucial preliminary step before investigating the cyclical properties of the data consists in cleaning each observed series in the panel from the noise, i.e. from that part of its own dynamics which is poorly correlated with the rest of the cross section. To this purpose, we must, first, define few factors (aggregates) which capture most of the variance of the variables in the panel and, second, project each variable on the leads and lags of those aggregates. This allows us to decompose each time series into two orthogonal components, the first one capturing the part of individual dynamics which has 'strong correlation' with the rest of the panel and the second one being of no interest for our purposes. We do this by using as aggregates the first few dynamic principal components, which are the dynamic generalizations of the well-known static concept of principal component.

Let us formalize the problem in the following way. We assume that our macroeconomic time series, suitably transformed, are realizations from a zero mean, wide-sense stationary n-dimensional vector process $y_t = (y_{1,t}, ..., y_{n,t})'$. We wish to summarize what the processes $y_{j,t}$, $j = 1, \dots, n$, have in common by a small number q of 'aggregate indexes'. Precisely, we look for q processes $z_{h,t}$, $h = 1, \dots, q$, satisfying the following properties:

(a) $z_{h,t}$ is a linear combinations of the leads and lags of the variables in y_t, i.e.

$$z_{ht} = p_h(L)y_t, h = 1, \dots, q,$$

where L is the lag operator and $p_h(L)$ is a $(1 \times n)$ row vector of two-sided linear filters

(b) z_{ht} and z_{kt} are mutually orthogonal at any lead and lag for $h \neq k$ and the filters $p_h(L)$ are normalized in such a way that $p_h(L)p_k(F)' = 0$ for $h \neq k$ and $p_h(L)p_h(F)' = 1$, where prime denotes transposition and $F = L^{-1}$. Finally, let us focus on the decomposition

$$y_t = \gamma_t^q + \zeta_t^q = C^q(L)z_t^q + \zeta_t^q = K^q(L)y_t + \zeta_t^q, \tag{1}$$

where $\gamma_t^q = (\gamma_{1,t}, ..., \gamma_{n,t})'$ is the projection of y_t on the present, past and future of $z_t^q = (z_{1,t}, ..., z_{q,t})'$ and ζ_t^q is the residual vector.

(c) the filters $p_h(L)$ and the associated processes $z_{h,t}$, $h = 1, \dots, q$, are such that the sum of the explained variances (for given q):

$$\sum_{j=1}^{n} var(\boldsymbol{\gamma}_{jt}^q) \tag{2}$$

is maximized.

Processes $z_{1,t}, \ldots, z_{q,t}$ satisfying requirements (a), (b) and (c) for $q = 1, \ldots, n$ do exist under quite general conditions and are called 'dynamic principal components' of y_t and are a natural generalization of the well known concept of static principal component.[3] What we propose here as the first step of our procedure is precisely to 'clean' the vector y_t by replacing it with its projection γ_t^q on the present, past and future of the first q principal components series.

A comprehensive treatment of the principal component series can be found in Brillinger (1981). Here we need only to remark a few facts. A first observation is that dynamic principal components are related to the eigenvalues and the eigenvectors of the spectral-density matrix of y_t, just like the static principal components are related to the eigenvalues and the eigenvectors of the variance-covariance matrix. Precisely, let $\Sigma(\theta)$, $-\pi < \theta \leq \pi$, be the spectral-density matrix of x_t: the vector $p_h(e^{-i\theta})$ is the eigenvector corresponding to the h-th eigenvalue of $\mathbf{\Sigma}(\theta)$ in descending order. Moreover, denoting by $\lambda_h(\theta)$ this eigenvalue and setting $\lambda_h = \int_{-\pi}^{\pi} \lambda_h(\theta) d\theta$, the maximal explained variance (2) is $\lambda_1 + \cdots + \lambda_q$ and the percentage of explained variance is given by the ratio

$$\frac{\lambda_1 + \cdots + \lambda_q}{\lambda_1 + \cdots + \lambda_n}. \tag{3}$$

As we shall see below, the above ratio provides useful indications for the choice of q.

Second, we can get an explicit expression for the filters $C^q(L)$ and $K^q(L)$ appearing in (1). These filters are given by

$$\begin{aligned} C^q(L) &= (p_1(F)', \cdots, p_q(F)'), \\ K^q(L) &= C^q(L)C^q(F)' = p_1(F)'p_1(L) + \cdots + p_q(F)'p_q(L)', \end{aligned} \tag{4}$$

[3] It is worth noting that the projection γ_t^q solving the maximization problem is unique, whereas the principal components themselves are not. To see this, let us focus for simplicity on the first principal component and set $q = 1$. Now let us consider any invertible two-sided filter $a(L)$. Clearly, the linear space spanned by the leads and lags of $a(L)z_{1t}$ and that spanned by the leads and lags of z_{1t} coincide. Hence if z_{1t} solves the maximization problem, also $a(L)z_{1t}$ solves the problem, since the projection γ_t^1 is the same. The normalization $p_1(L)p_1(F)' = 1$, which is usually adopted, is not sufficient to imply uniqueness, since it simply imposes $a(L)a(F) = 1$, i.e. the amplitude of $a(L)$ must be 1 at all frequencies, but the phase can be chosen arbitrarily. For instance, we can get a different set of principal components simply by taking their lags, i.e. by multiplying $P_h(L)$, $h = 1, \ldots, q$, by $a(L) = L^k$.

in close analogy with the static case.

Finally, it is worth stressing that the definition of the filters $p_h(L)$ involves unknown quantities like the variances in (2) and therefore must be estimated from a finite realization of y of length T. The estimator we use here (which is denoted by $\chi^T_{n,t}$ for reasons which will be clear below) is described in detail in Appendix B. Here we give only a short hint. As a preliminary step, we estimate the spectral density matrix $\Sigma(\theta)$ at different frequencies. Then, for each frequency, we compute the first q eigenvalues and eigenvectors and use (4) to compute $K^q(e^{-i\theta})$. Lastly, we use the inverse Fourier transform to estimate the filter $K^q(L)$ and apply it to the data. The estimates of $\Sigma(\theta)$ and $K^q(e^{-i\theta})$ can be exploited to estimate the spectral density matrix of the common components, which is $K^q(e^{-i\theta})\Sigma(\theta)K^q(e^{i\theta})'$.

The 'cleaning' of the variables obtained through the filtering process just described is motivated by the fact that one wants to look only at that part of the movement of the series that is common across the panel. Since our aim is to construct a business cycle indicator for the Euro area, we are only interested in the common movements at business cycle periodicity. In terms of the derivation of the filters $C^q(L)$ and $K^q(L)$ this implies that we can further restrict the space on which we project our series to extract the common components. In particular we can compute the inverse Fourier transform of $K^q(e^{-i\theta})$ only for θ belonging to the sub-interval of $[-\pi;\pi]$ corresponding to business cycle frequencies. In this way the common component obtained through filtering will capture that part of the variability of a given series that is common across the panel *and* that is related to business cycle swings.

2.2 *Principal components and the general dynamic factor model*

Our cleaning procedure is based on the choice of a small number q of principal components, and seems therefore open to considerable arbitrariness. However, if we assume that the y's are generated by a factor model, then the procedure can be given a more sound justification and a criterion for the choice of q can be constructed. In the dynamic factor approach, the variables are represented as the sum of two unobservable components: the 'common components', which are driven by a small number of 'factors', common to all of the variables in the system (but possibly loaded with different lag structures) and the 'idiosyncratic components', which are uncorrelated with the common components and are specific to a particular variable. If we take this point of view, eliminating the idiosyncratic part and retaining the common one appears as a quite natural cleaning procedure.

To better understand the factor model we are dealing with, it will be convenient to think of the vector y_t as formed by the first n elements of the infinite sequence $y_{j,t}$, $j = 1, \ldots \infty$. To emphasize the dependence on n, we write y_{nt} in place of y_t. In our model,

$$y_{j,t} = \chi_{j,t} + \xi_{j,t} = b_j(L)u_t + \xi_{j,t} \tag{5}$$

where $\chi_{j,t}$ is the common component, $u_t = (u_{1,t}, ..., u_{n,t})'$ is the vector of the common shocks, i.e. a (covariance stationary) q-vector process with non-singular spectral density matrix, $b_j(L)$ is a row vector of possibly two-sided, square-summable filters, and the idiosyncratic component $\xi_{j,t}$ is orthogonal to u_{t-k} for any k and j. Hence, with obvious notation,

$$y_{n,t} = \chi_{n,t} + \xi_{n,t} = B_n(L)u_t + \xi_{n,t}. \tag{6}$$

Finally, we require the following properties. Let us denote by $\lambda^{\chi}_{h,n}(\theta)$, $h = 1, \ldots, n$, the h-th eigenvalue of the spectral density matrix of $\chi_{n,t}$, in descending order of magnitude. Similarly, $\lambda^{\xi}_{h,n}(\theta)$ is the h-th eigenvalue of the spectral matrix of $\xi_{n,t}$. We assume that:

(i) the eigenvalues of $\xi_{n,t}$ are bounded as $n \to \infty$; precisely, $\lambda^{\xi}_{h,n}(\theta) < \bar{\lambda}$ a.e. in $[-\pi, \pi]$, for any h and n;

(ii) the first q eigenvalues of $\chi_{n,t}$ diverge, precisely, $\lim_{n\to\infty} \lambda^{\chi}_{h,n}(\theta) = \infty$ for $h \leq q$, a.e. in $[-\pi, \pi]$.

Model (5) is the generalized dynamic factor model proposed by Forni, Hallin, Lippi and Reichlin (2000) and Forni and Lippi (2000). The basic difference with respect to the dynamic factor model of Sargent and Sims (1977) and Geweke (1977) is that here the idiosyncratic components are not assumed to be mutually uncorrelated. Instead of this rather restrictive assumption, we require conditions (i) and (ii), which impose a particular behavior to the common and the idiosyncratic eigenvalues as the cross-sectional dimension becomes larger and larger. Heuristically, we require that the amount of cross-correlation between the idiosyncratic components is limited in the sense that idiosyncratic causes of variation, although possibly shared by many (even all) units, have their effect concentrated on a finite number of units and tending to zero as n tends to infinity. On the other hand, we want a minimum amount of cross-correlation between the common components. With a slight oversimplification, we want each $u_{h,t}$ to be present in infinitely many cross-sectional units, with non-decreasing importance. These requirements define the notion of 'common' and idiosyncratic in an asymptotic sense and guarantee the uniqueness of the common and the idiosyncratic components (the uniqueness of the common shocks and the factor loading requires additional assumptions).

Now let us go back to equation (1) and rewrite it as

$$y_{n,t} = \gamma_{n,t} + \zeta_{n,t} = C_n(L)z_{n,t} + \zeta_{n,t}, \tag{7}$$

where for convenience, we have added the subscript n and dropped the superscript q, which is not useful in this context. Now let us add the following assumptions:

(iii) the non-zero eigenvalues of $\zeta_{n,t}$ (i.e. the last $n-q$ eigenvalues of $x_{n,t}$) are bounded as $n \to \infty$; precisely, $\lambda_{h,n}(\theta) < \Lambda$, $h = q+1, \ldots, n$, a.e. in $[-\pi, \pi]$, for any n;

(iv) the first q eigenvalues of $\chi_{n,t}$ (i.e. the first q eigenvalues of $y_{n,t}$) diverge; precisely, $\lim_{n\to\infty} \lambda_{h,n}(\theta) = \infty$ for $h \leq q$, a.e. in $[-\pi, \pi]$.

Assuming (iii) and (iv), the similarity between representations (7) and (6) is striking. The basic difference is that the sequence $\chi_{n,t}$, $n = 1, \ldots, \infty$ is nested, in the sense that the first $n-1$ entries of $\chi_{n,t}$ are the same as that of $\chi_{n-1,t}$. By contrast, the sequence $\gamma_{n,t}$ in non-nested in general, so that the two decompositions do not coincide.

However, there is a deep relation between them. Forni and Lippi (2000) show that if conditions (iii) and (iv) on the eigenvalues of the x's are satisfied, then the generalized dynamic factor representation (5) does exist and, conversely, if (5) holds, then (iii) and (iv) are satisfied.

This result is a dynamic generalization of a basic theorem in Chamberlain and Rothschild (1983). Moreover, the j-th entry of $\gamma_{n,t}$, call it $\gamma_{j,n,t}$, converge to $\chi_{j,t}$ in mean square as $n \to \infty$, for any j. Hence, for n large, $\gamma_{n,t}$ is a good approximation of $\chi_{n,t}$.

These results build a firm bridge linking principal component and factor analysis. The basic intuition behind them is that, by taking the principal components, we are taking an average of the x's. When n is large, we get a kind of Large Number result. The idiosyncratic components, which are poorly correlated, disappear, so that we are essentially left with linear combinations of (the leads and lags of) the common components. Such linear combinations span almost the same dynamic space as the common factors. Hence, by projecting $y_{j,t}$ on the former space, we approximate $\chi_{j,t}$, which is the projection of $y_{j,t}$ on the latter.

The above results also suggest a simple criterion for the choice of the number of principal components to be retained. If model (5) holds, the eigenvalues $\lambda_{h,n} = \int_{-\pi}^{\pi} \lambda_{h,n}(\theta) d\theta$ are bounded for $h > q$ and diverge for $h \leq q$ as $n \to \infty$. Hence, for large n, we expect that there is a 'jump' between $\lambda_{q,n}$ and $\lambda_{q+1,n}$. This suggests adding principal components until the increase in the explained variance is less than some predefined value. Precisely, denoting by $\lambda_{h,n}^T$, where T is the number of time observations, the estimate of $\lambda_{h,n}$, and given a number $\alpha \in (0,1)$ the criterion consists in selecting $q = q^*$ such that

$$\begin{aligned} \lambda_{q^* n}^T / \sum_{h=1}^{n} \lambda_{h,n}^T &> \alpha \quad and \\ \lambda_{q^*+1,n}^T / \sum_{h=1}^{n} \lambda_{h,n}^T &< \alpha. \end{aligned} \tag{8}$$

2.3 *The procedure*

Having clarified the basic theoretical background, we can now go on and present the various steps of procedure followed for the identification of the common business cycle behavior of the series in the panel and for the construction of the coincident and leading indicators of the euro area. The procedure consists of six steps:

A. Choice of the variables to include in the panel and their pre-treatment;
B. Identification of the number of common components and ”cleaning” of the variables for the non-business cycle components;
C. Construction of the coincident indicator;
D. Classification of the variables as pro and anti-cyclical and leading and lagging vs the coincident index;
E. Construction of the leading indicators as the best forecast five steps ahead of the coincident;
F. The end point adjustment and the forecast evaluations.

3. A unified euro area database

The key message from the theoretical analysis is that, in order to capture the factors underlying the comovements observed across the euro area, the cross sectional dimension of the dataset needs to be very large, with a number of series possibly larger than the number of observations. Furthermore prefiltering time series to achieve stationarity and to remove possible outliers is a preliminary step towards the correct estimation of the correlation structure in the dataset, as explained in paragraph 2.

3.1 *The data*

The construction of a business cycle indicator for the euro area that exploits the dynamic factor model approach demands a large amount of data to meet the dimensionality requirement of the cross-section. Unlike for the U.S. case, where analysts can promptly access well established databases,[4] in Europe nothing of this sort yet exists.[5] We therefore had to devote a big effort in consulting many different sources: among others, national statistical institutes, the OECD and the ESCB statistics (see Table 1 for details); from these we collected and examined a large number of series, organizing them in a detailed dataset, covering the vast majority of economic phenomena pertaining to the euro area.

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

---

[4] See, as an example, the DRI-McGraw Hill Basic Economics database, formerly known as ‘Citibase’.

[5] The so called Short Term Statistics (STS) Database under construction at Eurostat is still very lacking.

into the following eight homogeneous blocks, corresponding to different major topics: industrial production; producer prices; consumer prices; money, interest rates and exchange rates; European Commission surveys; trade; labour and other variables. On the whole, they should provide an almost exhaustive description of the European economy (see Table 2 for the numerosity of time series held in each block).

Each block contains time series for Germany, France, Italy, Spain and, when available, for Belgium and The Netherlands. Actually, since a business cycle indicator for the euro area reflects economy wide fluctuations common across countries, one should collect data covering an ample 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 per cent of the euro aggregate GDP in 2000. Some macroeconomic series not directly referrable to the euro area were also gathered to capture phenomena that, likewise, might be relevant to explain business cycle fluctuations across Europe; some examples are oil and raw materials prices and some indicators of the business cycle position in other large economies (US and Japan).[6]

In the construction of the dataset two crucial requirements were pursued to capture fluctuations at business cycle periodicity: the proper length of the series and their homogeneity over time and across countries. As regards the first one, the largest common sample for the dataset spans the period 1985.1-2000.9.[7] Although many time series are available for a longer sample, the decision to set the starting date in 1985 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 collected variables for each of the eight blocks maintaining within each of them, wherever possible, a common breakdown for all countries. In many cases we obtained satisfactory series only joining currently available statistics covering a short time range (for example HICP or Pan-German data) with older ones (CPI, West German data), trying to match definitions and

---

[6] These time series account for no more then 10 per cent of the whole dataset.

[7] 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.

disaggregations as closely as possible (see Appendix A for details). On this regard, whenever we had to intervene with some kind of 'manipulation' to obtain time series of the desired quality, the strategy adopted for data reconstrucion was the following. For the most recent years series were collected from Eurostat and the European Commission whenever available with the desired length and quality, since such institutions coordinate national sources in the process of statistical harmonization. As a second option other international institutions (like the OECD) and national sources (e.g. INSEE or IFO) were used to obtain a sufficient length in the series and to cover some important economic phenomena with the desired detail.

Finally, in order to avoid overweighting a single country or a particular economic sector, we tried to maintain a satisfactory balancing across blocks and countries. Nevertheless, the available statistics forced us to make a choice between working with a minimal common set of indicators, thereby greatly reducing the richness of information,[8] or working with an unbalanced dataset; this is the reason why, to meet the requirement of a large database, we relaxed the condition on *perfect balancing* (see Table 3 for the structure of the final dataset).

The effort dedicated to this stage of the work ended up with the collection of about 800 monthly time series satisfying the requisites listed above, drawn from a much ampler set of variables gathered from the consulted sources (see Table 2).[9]

3.2 *Data treatment*

The original data collected present very diverse characteristics for what concerns their dynamic behavior; some of them are raw data, others have been adjusted to take into account working day effects and some are available only in a seasonally adjusted version. Furthermore preliminary inspection reveals that they cannot be characterised by the same kind of non stationarity. It was not feasible to carefully analyze each series included in the panel; therefore we followed an 'automatic' procedure that ensures a homogeneous treatment to all series, to fit the typical behavior in each group. Particular attention was paid in checking whether this procedure resulted in an improper treatment of the data, such as over-differencing and removal

---

[8] In our case this option would have reduced the number of series eventually used in the analysis from 794 to approximately 250 (see, Table 2).

[9] The construction of a quarterly dataset and the development of a method to properly mix information obtained at different frequencies is however part of the work in progress at the Bank of Italy.

of too little seasonal variation. Unfortunately, sometimes the problem could not be fixed since it derived from the poor quality of the original data; in such instances we discarded the variable from the dataset. Our data treament procedure can be detailed in the following four stages.[10]

First, we seasonally adjusted each series using Tramo-Seats (TS), a model based procedure developed by Gomez and Maravall.[11] This package provides an ARIMA-based decomposition of a time series into three unobserved components (trend-cycle, seasonal and irregular). It also contains a routine to detect and remove several types of outliers; in particular we focused on additive outliers, transitory changes and level shifts. In the analysis we used the outliers free and seasonally adjusted version of each series resulting from the application of the TS procedure; series released from the original sources as seasonally adjusted were also put through this stage to remove any residual seasonality or outliers.

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

Third, in order to estimate the cross spectral density matrix the series need to be covariance stationary. The stationarity inducing transformation was applied to each *outliers free and seasonally adjusted series.* The first log difference was taken for groupings involving quantity variables such as money and industrial production; first differences were applied to interest rates, business and household survey responses; no transformation was needed for interest rate spreads. In general the stationarity inducing transformation exploited was coherent with the model identified by TS. The most controversial issue concerned the order of integration of some price variables where the choice between I(1) and I(2) models was borderline in some cases. After some further checks we decided to consider them as I(1).

Finally the series obtained from stage 3 were normalized, dividing them for their standard deviation and then subtracting their mean.

After this preliminary treatment, 794 variables spanning the period 1985.1-2000.5 were ready to be used in the estimation of the dynamic factor model (see Table 2).

---

10 A similar procedure was adopted by Marcellino, Stock and Watson (2000).

11 See Gomez, Maravall (1999).

4. Is there a euro area business cycle?

The construction of a business cycle composite indicator for the euro-area rests on the assumption that a sizeable degree of co-movement in macroeconomic variables across countries as well as across sectors does exist and can be properly exploited. A growing body of empirical literature has addressed this question using various methods,[12] and it 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 the rest of the 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 movements at business cycle frequencies are relevant across countries and sectors and are captured by a limited number of common factors.

One way to explore the existence of co-movements among the different types of macroeconomic variables included in our panel is by looking at the *typical spectral shape* of the series in the main blocks (sectors) of our panel. The inspection of simple arithmetic averages of the spectral density functions of the variables within each block (already transformed to achieve stationarity and properly standardized) reveals the presence of relevant dynamics at low frequencies (see Figure 1). The same behaviour emerges looking at the entire set of data. We can conclude that monthly series have, on average, a clearly detectable cyclical behaviour, responsible for the larger part of their variability.

Given this evidence, the natural question to ask is whether the movements at business cycle frequencies are common across Europe. Following the dynamic factor approach, the presence of a euro wide business cycle implies that a large part of the variance at business cycle periodicities can be explained by a small number of factors driving all the cross section. Recalling the notation of the methodological section, the spectral density matrix of the data can be interpreted as the decomposition of the total variance-covariance of the series over different periodicities (frequencies) and the exsistence of few driving factors, in turn, would

---

[12] Engle and Kozicki (1993), McDermott and Scott (2000), Lumsdaine and Prasad (1999), Cheung and Westermann (2000), Dickerson et al. (1998), Artis et al. (1999).

result in the presence of some dominating eigenvalues of the spectral matrix, concentrated at those frequencies.

At any given frequency θ, it is possible to write the following factorization of the spectral matrix, as in the static principal component analysis:

$$\Sigma(\theta) = U(\theta) \cdot \Lambda(\theta) \cdot U(\theta)^* \tag{9}$$

where $U(\theta)$ is the $(n \times n)$ matrix containing the row eigenvectors $p_j(\theta)$ (principal components)[13] and $\Lambda(\theta)$ is the $(n \times n)$ diagonal matrix containing the eigenvalues in decreasing order and generic element $\lambda_j(\theta)$. Similar to the static case it is possible to decomposes the spectral density at frequency θ into the contributions of the components:

$$\Sigma(\theta) = \sum_{j=1}^{n} p_j(\theta) \cdot p_j^*(\theta) \cdot \lambda_j(\theta), \tag{10}$$

and the contribution of the first principal component to the total variance at frequency θ is then given by:

$$\frac{\lambda_1(\theta)}{\sum_{j=1}^{n} \lambda_j(\theta)} \tag{11}$$

The plot of the first six normalized eigenvalues on the interval $[0;\pi]$ shows that there is commonality among the series and that this commonality is larger at business cycle frequencies (see Figure 2). The first dynamic principal component explains more than 25 per cent of the variability at business cycle frequencies, the first twos around 50 per cent and the first four more than 75 per cent. To give a banchmark to this result it is usefull to recall that if the 794 series in the panel were independent from each other, then each eigenvalue would account for $\frac{1}{794}$ of the total variance and there would not be a dominating one. We can conclude that not only our data exhibit large variability at low frequencies, but also that the co-movements across series are concentrated at business cycle periodicity. In the following, we choose to work with two factors, i.e. $q = 2$ in the notation of the previous section; this allows to account for half of the variability of the series at business cycle. The criterium of choice of the number of factors is described in the previous section and we retain only those components which explain more than 15 per cent of the total variability of the series at business cycle, i.e. $\alpha = 0.15$ in (8).

---

[13] The symbol * stands for conjugate transpose and the eigenvectors have been unitary normalized, i.e. $UU^* = I$.

Moreover since the presence of dominant eigenvalues is concentrated only at business cycle frequencies, common components χ_t - i.e. the projections of the series onto the space spanned by the common factors - can be obtained by computing the weights only with respect to the cyclical band. The cyclical band, comprising oscillations with periodicity between 14 and 120 months, i.e. $\theta \in [0.05, 0.45]$.[14]

This choice derives from the fact that at those frequencies there is a clear dominating pattern of the first two eigenvalues, but a further advantage one gets by limiting the inversion on to the cyclical frequencies is evidenced in Figure 2. In fact looking at the amount of total variation explained by the first common factor (i.e. looking at the plot over $[0; \pi]$ of the first eigenvalue) two further peaks become visible at higher frequencies. The first one, around a periodicity of one year, can be attributed to residual *common seasonality,* potentially filter-induced, since all data have been treated with a similar filtering process. The second one, at a quarterly periodicity, is presumably due to time disaggregation performed on data that are originally available only at a frequency lower than monthly.[15]

The ability to identify co-movements at various frequencies relevant to economic analysis is an interesting feature of the methodology, that opens the way to various applications, like seasonal adjustment. These aspects are not further pursued in the present work, since we want to isolate only that part of the commonality that is related to the business cycle, but it can be an element of future investigation in term both of analysis of more efficient multivatiate procedure of isolating seasonal effect and of the importance of revisions in the data induced by the seasonal adjustment procedures. On the other hand, the fact that some commonality exists also at other frequencies supports the decision to confine to the business cycle periodicity if the purpose is to construct indicators that reflect only medium run swings in economic activity.

Two objections could be raised at this point. First, one might wonder why we proceeded in the analysis of the series through all the steps required by our methodology just to end up with a filtering that resembles the one derived by standard *band pass filter*. In truth,

---

[14] For details on the actual procedure see section 2 and the techincal Appendix B.

[15] Many statistical institute are probably computing some of their monthly statistics partly relying on quarterly data. This data can be transformed into monthly statistics by temporal disaggregation procedures such as that proposed by Chow and Lin.

with respect to traditional band pass filtering,[16] the filter obtained with the dynamic principal components approach, exploits a much wider set of information and isolates movements that are not only of a certain periodicity but also common across all series. The second objection relates to the possibility, which we excluded at the outset, that seasonal and business cycles are related.[17] While the empirical literature on the subject has not reached clear conclusions, we still believe that the theoretical underpinnings of the approach are still rather weak. Furthermore, at the present stage, the factor analytical approach does not allow a clear cut identification of the common factors, attributing one factor to ”business cycles” and the other to ”seasonal cycles”, including commonality at frequencies other than business cycles would entail having common factors that are a mixture of seasonal and cyclical shocks.[18]

The decision to restrict the attention to the $[0.05; 0.45]$ interval implies that only two factors are sufficient to explain about 50 per cent of the business cycle variation of the series included in the panel. The other factors (again looking at Figure 2) do not seem more relevant to capture cyclical movement than they are to capture higher frequency dynamics. Hence we fixed the number of factors to 2. On average the variance of the common components $\chi_{j,t}$ thus obtained is about 40 per cent of the total variance of the series, rising to almost 60 per cent for labour market variables.

5. The coincident indicator

Having estabilished that the observed co-movements at business cycle frequencies can be imputed to a small set of common factors we obtained, for each variable i, the projection, χ_i, on the space spanned by these factors. While the χs variables represent the common component of each series with respect to the cross section, it is still ambiguous how a synthetic measure of the business cycle in the euro area can be defined.

In the business cycle literature there is a long standing tradition in the construction of cyclical indicators which goes back to the work of Burns and Mitchell (1946) and the NBER. In the NBER tradition the analysis involved a first stage in which a large number of candidate

---

16 See Baxter and King (1999).

17 See Beaulieu, Mason and Miron (1992).

18 This line of research is currently being pursued. Some results and a general discussion are in Forni and Reichlin (2001).

macroeconomic series were subject to detailed scrutiny. These were evaluated according to their leading-coincident properties with respect to a well defined set of 'business cyle dates'; in particular the assessment relied on turning point analysis and on the stability of the identified lead/lag relationship. While not based on a well defined probabilistic model, this procedure produced a reliable measure of economic activity. More recently, a different approach proposed cyclical indicators derived directly from a stochastic model, where the cycle is identified as the common unobservable component driving a small set of variables. In this framework state-space models have been employed to derive a measure of the unobservable cyclical component. This approach, introduced by Stock and Watson (1989, 1991) led to the construction of a new version of the coincident and leading indicators for the US that is currently published.

The dynamic factor approach allows us to retain the original NBER idea of exploiting the full information set available from European short term statistics, while relying on a well defined statistical model. Unlike in the US case, in the euro area context no official and reliable business cycle dating yet exists; hence, the construction of a coincident indicator first has to address the problem of defining a reference variable, with respect to which to evaluate lead/lag relationships. In the Italian experience Altissimo et al. (2000), consider a a set of candidate variables, including GDP and industrial production, to avoid the circularity flaw implicit in the older NBER tradition.

In our analysis we defined the euro area coincident index, coi_t, as the weighted average of the common component of the GDP of the six largest euro area countries, where the weights account for the relative size of the countries,[19] i.e.

$$COI_t = \sum w_k \cdot \chi_{GDP_k,\, t}$$

where w_k are the weights and $\chi_{GDP_k,\, t}$ is the common component of the GDP in the k_{th} country (Germany, France, Italy, Spain, The Netherlands and Belgium). Such indicator has to be interpreted as a measure of the area wide common fluctuations at business cycle perioditicity, rather than as an average of the (potentially different) cyclical positions of each country: the latter could be affected also by factors specific to a particular country. For example, if a country is heading towards a recession as a result of a pure country specific factor, this should not, in

19 Weights are calculated using PPPs in the base year.

principle, contribute to the area wide index. On the contrary if the same cyclical episode is shared by the majority of the European economies, this should be reflected also in the overall index. A full understanding of the different cyclical position of the euro area countries and the relevance of common versus country specific cyclical dynamics goes beyond the scope of this work. However, the adoption of a unique monetary policy, the introduction of the single market and the single currency, and the coming into place of various constraints for national fiscal policies (like the Stability and Growth Pact) envisages that the common area wide component of the business cycle fluctuation is becoming the most relevant one to consider also at the country level.

The GDP is an overall measure of economic activity and it has a 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. To formalize this point, let $\widetilde{y}$ be the approximating value of the unobservable monthly GDP y, as

$$\widetilde{y}_t^{GDP} = y_t^{GDP} + \varepsilon_t = \chi_{GDP_k,\,t} + \zeta_t^{GDP} + \varepsilon_t = C^{GDP}(L)z_t + \zeta_t^{GDP} + \varepsilon_t,$$

then to estimate consistently the common component of y, i.e. $C^{GDP}(L)z_t$, it is sufficient that ε_t and z_t are orthogonal at all lead and lags.

The resulting coincident indicator for the euro area is constructed over the period 1987.06 to 2000.09 and is reported in Figure 3. In order to recover the common component of the variables, and consequently of the index, for the observations at the end of the sample, auxiliary forecasts of the variables have been performed as described in the next section.

Being associated to the cyclical fluctuations of the output growth rates, the index is coherent with a *growth cycle* definition: the cycle is identified with the deviations of economic activity from its long-term trend, identified by the zero line in the figure. Positive value of the indicator signal periods of growth above the long run growth rate, and the reverse for periods 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 were based on the ”classical cycle” concept, which focuses on fluctuations in the *absolute level* of economic activity.

As a further check on the quality of our reference variable we also constructed a coincident indicator in the NBER fashion (see Figure 7), that is by simply averaging the common component of all the variables classified as coincident. Reassuringly, the resulting indicator displays properties similar to our reference variable, COI_t, signaling the same turning points in economic activity and amplitude of the business cycle episodes.

• *The dating*

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): 1988.10-1991.12, 1992.01-1994.09, 1994.10-1997.11 and 1997.12 to now. Applying the Bry-Boschan (1971) dating scheme to our coincident indicator obtained a dating for the European business cycles. The average duration of expansion and recession episodes is roughly similar 17 months the former, 16 the latter. 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 1992.01-1994.09 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 (1993.01-1994.09 and 1995.11-1997.11) lasting around two years each and two recessions (1994.10-1995.10 and 1997.12-1998.11) of short duration, one year each. The most recent peak occurred in the middle of the year 2000. However, the exact timing of this episode could be affected by some end of sample uncertainty. Differently from 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.

• *The countries*

In Figure 4, the coincident index is compared with its national component, namely the $\boldsymbol{\chi}_{GDP,\,t}$ of the single countries that make up the overall index. These 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. Two general comments can be easily derived from the visual inspection of the figure. First, the German component seems to lag at turning points the other countries indicators: this can be the result both of a lower importance of the euro area shock for the German economy and its sluggishness to adjust. Second, after 1992 and the German unification there is some evidence of a stronger synchronization among the six euro area economies.

To conclude this section in Figure 5, we compare our coincident indicator with the one recently developed by the European Commission (EC) for the euro area, which exploits the unobserved component methodology on business survey data. The figure reports also the three month moving average of the growth rate of European industrial production. Quite strikingly the EC index seems to lag both our indicator and the industrial production series.

6. Pro-counter cyclicality and lead-lag relationships

Having selected the common component of the euro area GDP as our reference index, we examine how its fluctuations at business cycle frequencies relate to the ones of other variables. In particular we examine the cross correlation between the coincident indicator and the common component of each series in the panel, filtered over the cyclical band, $\widetilde{\chi}_{it}$. We classify a variable as being pro-cyclical (counter-cyclical) if it displays a positive maximum (negative) correlation, $corr(\,\widetilde{\chi}_{i,t-h}, \widetilde{\chi}_{eugdp,t})$, with respect to the common component of GDP, $\widetilde{\chi}_{eugdp,t}$.The same cross-correlation analysis establishes a full set of lead-lag relationships between each variable and our reference series. In particular we used the following classification scheme. If the correlation is significantly different from zero and the displacement of the maximal correlation is negative (positive) and greater or equal in absolute value to three months, the variable is classified as leading (lagging); a displacement between -2 and +2 months instead characterizes coincident variables. All variables with a non significant correlation were classified as *uncertain*. An alternative procedure commonly used in the literature using frequency domain techniques (FLHR, AMO) consists in classifying variables according to their phase angle with respect to the reference series, evaluated at frequency zero.

We decided not to use this criteria to be able to handle also *smoothed* series such as the $\widetilde{\chi}_{it}$, i.e. series that by construction should display only very little variation at frequencies that fall outside the cyclical band, such as zero.

6.1 *Overall results*

On average the dataset resulted well balanced in terms of the number of leading, coincident and lagging variables: the median delay relative to the common component of the euro area GDP is -1 months. Around 286 variables (36 per cent of the dataset) are found to be leading with respect to the reference index, 286 are coincident and 222 are lagging.

About 120 series resulted to be countercylical with respect to the European GDP;[20] in particular it is interesting to note that the *unemployment variables* for all the countries considered (both in actual and in expected level) fall into this category. The same feature is shared by the firms' assessment of *stocks*: this is indeed conforting because it is line with the predictions of inventory business cycle models.

6.2 *Industrial production*

Industrial productions revealed a widespread pro-cyclical behaviour and, on average, fairly good leading properties of the European business cycle: the median delay for this block is -2 months, the variance explained by the common factors at cyclical frequencies is 27 per cent. Nevertheless some differences emerged across countries and sectors; in particular, the dynamics of the Spanish and French manufacturing activity appear to be in advance with respect to the overall economic fluctuations, whilst in Germany and in Italy most of the industrial productions show coincident characteristics. In spite of these diversities, some common features emerged across economies: one of these, not surprisingly, regards the sectors involved in the manufacturing of packing materials - like pulp, paper and paper products - which share the same leading properties area wide.

The analysis revealed that several other production activities are characterized by this common feature, although with slightly shorter time leads with respect to the previous one.

---

[20] The results remained substantially similar when we classified variables according to their phase angle with the reference index.

Among these, intermediate goods and consumer durable goods productions stand out, together with the manufacturing of motor vehicles, trailers and semi-trailers and the sectors involved in the production of basic metals.

On the other side, some industrial activities appear to be coincident or even to lag the overall fluctuations. Capital goods, machinery and equipment sectors form part of the first group; in the second one, wearing apparel and dressing manufactures are included.

6.3 *Producer and consumer prices*

Price variables display a strong comovement within the cross section: around 40 per cent of their variation at cyclical periodicity is explained by the first two factors. On average consumer prices in almost all countries appear to be procyclical and in phase with the fluctuations of our reference variable, their average displacement being of about 3 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 of 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 result procyclical and show an average lag of about 2 months with respect to our coincident indicator, that is a slightly inferior time displacement than consumer prices. Similarly to what was found for consumer prices, in Belgium they revealed to be leading with respect to our reference variable.

6.4 *Survey data*

As could be expected, the European Commission business and consumers surveys were confirmed to be relevant information sources for the business cycle analysis; nonetheless attention has to be paid in interpreting the results: among them the variance explained by common factors amounts to about 30 per cent, meaning that the signals they release are quite noisy. The construction industry survey have leading properties in almost all European countries, particularly in The Netherlands and in Belgium; with a lesser degree in France and in Germany. Italy constitutes an exception: the balances of the answers given by the building sector firms generally show coincident or even lagging characteristics: the sectorial confidence indicator lags 1 month the overall European economic fluctuations.[21]

The manufacturing industry survey has good leading properties too and, just like the case of the construction sector, the results coming from The Netherlands and Belgium are the strongest in this sense: their industrial confidence indicators move, respectively, with a 4 and 5 months time lead. The result for Belgium in particular are in line with the common wisdom regarding the leading properties of industrial activity in this country. Confirming the evidence emerged from the industrial production analysis, the German results are on the borderline between being either coincident or slightly leading.[22] Among the questions included in the manufacturing industry survey, the one pertaining to the short term production expectations have the largest average time lead across different countries, wheares, as already outlined, the assessment of stocks of finished products revealed a countercyclical behaviour.

The consumer confidence indicator has weaker leading properties than the manufacturing one, reflecting the mixed evidence deducible from different questions; it is homogeneously leading of about 3 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. The other questions reveal different properties according to the reference country.

The retail trade survey could be analyzed only for Germany and for The Netherlands, the time series of the other countries being not long enough to be used in the estimation of the dynamic factor model. The results are controversial: the German sector is clearly lagging, whilst on the whole the Dutch one appears to be leading.

6.5 *Money, interest rates and exchange rates*

The median delay of this block is of 3 months: 17 per cent of the variables resulted to lead the European reference cycle, while most of them (55 per cent) are lagging.

---

[21] It leads 6 and 5 months in The Netherlands and in Belgium, respectively, and 3 months both in Germany and in France.

[22] This feature is summarized by the confidence indicator which has a time lead of 2 months (4 months for France and Italy).

Monetary aggregates are generally classified as coincident; the major exceptions are M1 in Germany and in Italy, characterized by rather long leads and anticipating the European business cycle by 12 and 6 months respectively, and M2 in Belgium with a time lead of 10 months.

The spread between Italian long term and short term interest rates has good leading properties too, confirming the results found by Altissimo et al. (2000) about the Italian business cycle.

Finally the measures of competitiviness based on real exchange rates included in this block generally anticipate business fluctuations with leads ranging from 6 to 10 months.

6.6 *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 two dynamic factors: the explained variance is 0.24; nonetheless 45 per cent of them are leading.

Among these, noteworthy results regard the share-price indices that lead our coincident indicator by at least 3 months in all countries; a similar behaviour concern car and other vehicles registrations and, not surprisingly, indicators of rail transportation of goods, for countries where they are available.

The same leading features were displayed by some specific indicators of industrial activity: in particular the production of crude steel in Italy, Spain, Belgium and Germany revealed a strong and positive correlation with the European cycle. Similarly, in the Italian case, electricity consumption - whose properties have been well documented[23] - is found to have a 3 months time lead on European GDP.

Statistics on dwellings *started* for France, Belgium and Spain share leading features too, while construction permits *issued* in Germany and Belgium resulted to be either coincident or lagging. Commodity prices excluding food components displayed a strong pro-cyclical behaviour, appearing in phase with the reference variable.

---

[23] See Marchetti and Parigi (2000).

Some stylized facts emerge also for the labour market variables: the unemployment time series, wherever available, are clearly countercyclical and even the consumers' expectations on future unemployment trends - gathered by the European Commission surveys - show a similar behaviour in all countries.

Finally, it is interesting to point out that the series concerning the utilization of Wage Supplementation Fund by Italian industrial firms are countercyclical and lead the European GDP by about 2 months. This result confirms previous findings (see Altissimo et al., 2000) and is coherent with the fact that employers resort to this Fund only during recessions. On the other hand the percentage of overtime hours in large Italian industrial firms is pro-cyclical and coincident.

7. The leading indicator

In the above section, variables have been classified according to their degree of co-movement with the common factors and their lead/lag relation with respect to the coincident index. In particular a large number of them (more than 180) were found to have a time lead of at least four months. This enabled us to construct a leading indicator that exploits a vast and complex array of information on the European economies. Most importantly the dynamic factor model suggests that the leading variables in any given period contain information on the same common shocks that will hit the coincident series only later on and, by construction, the coicident index too. The leading variables therefore are the natural candidates to be used in the development of a forecast of the coincident indicator.

We decided to restrict the attention to those variables that were found to lead the reference index by at least four months: this choice is quite conservative because it excludes some potentially useful information coming from variables with a shorter lead. However it ensures that the resulting leading indicator does indeed anticipate our reference index and, most of all, its quarterly counterpart. The average time lead of the variables satisfying this criterion was around five months. This time interval for the forecast should be a reasonable approximation of a *real time* situation, when data is available only with a certain lag or in a provisional form, and policy makers are interested in assessing the current cyclical position of the economy.

To construct the five steps ahead prediction, we project the coincident index five months ahead on its current and past values and on simple averages of the common components of the leading variables. This means that the leading index has to be interpreted as the forecast of the coincident at $t+5$, given the information at time t. In Figure 4 the leading index is compared with the coincident one. It indeed performs extremely well in forecasting the coincident index and in anticipating turning points.

Since the calculation of the common components of the variables that make up the coincident and the leading indicators at time time t use information beyond t (because of the bilateral nature of the filters), the evaluation of the performance of our leading indicator reported in Figure 4 is only an ex post one, in the sense that it disregards the real time situation where part of the data is not yet available.

We takle the issue of end point adjustment, taking into account the bilaterality of the filters and putting our indicator under more demanding tests, in the next section.

8. The end point adjustment

The synthetic indicators COI_t and LEA_t obtained as linear combinations of the *common components* $\chi_{j,t}$ of each variable are themselves functions of the past, present and future observations of all the time series included in the panel. This derives from the fact that the $K(L)$ filters are bilateral. Hence, at the end of the sample these indicators cannot be constructed unless either forecasts of the variables are available or the filter is transformed from bilateral into unilateral. We resorted to the first method and constructed estimates for COI_t and LEA_t based on M steps ahead forecasts of their component series (in our case M is equal to 16). This procedure does indeed exploit the lead and lag relationships established in the construction of the business cycle indicators.

Let our sample of stationary and standardized variables be $\{ y_{j,t}\}_{t=1}^{T}$where $j = 1, 2, ..., n$. Having chosen a truncation lag equal to M the *common component* of the $j-th$ variable is given by

(12) $$\chi_{j,t} = \sum_{h=-M}^{M} K_{j,h} y_{t-h}$$

where y_{t-h} is the $(n \times 1)$ vector of variables at time $t-h$ and $K_{j,h}$ the $(1 \times n)$ weighting vector at lag (lead) h (see 4 in section 2). At the end of the sample (more precisely from time $T-M+1$ onwards) we no longer have the necessary observation on y_t to construct our indicators.

To build the forecasts of the $y's$ and $\chi's$, we first obtained two indicators c_t and l_t by taking a simple (unweighted) average of the variables previously classified as coincident (CO) or leading (LE): $c_t = \frac{1}{nc}\sum_{j\in CO} y_{t,j}$ and $l_t = \frac{1}{nl}\sum_{j\in LE} y_{t,j}$, where nc and nl are the number of coincident and leading variables, respectively. Forecasts for c_t and l_t (from $T+1$ to $T+16$) were then produced using the following VARMA model:

$$
\begin{aligned}
\begin{pmatrix} c_t \\ l_t \end{pmatrix} &= m + \sum_{k=1}^{5} B_k \begin{pmatrix} c_{t-k} \\ l_{t-k} \end{pmatrix} + \eta_t \\
\eta_t &= \rho\eta_{t-1} + u_t
\end{aligned} \tag{13}
$$

Next we run n (= 794) regressions to obtain 16 steps ahead forecasts of the original variables exploiting the projected c_t and l_t indexes:

$$
\begin{aligned}
y_{t,j} &= \mu_j + \sum_{k=1}^{5} \alpha_k y_{j,t-k} + \sum_{k=0}^{5} \beta_k c_{t-k} + \sum_{k=0}^{5} \gamma_k l_{t-k} + \upsilon_{j,t} \\
\upsilon_{j,t} &= \rho_j \upsilon_{j,t} + \varepsilon_{j,t}
\end{aligned} \tag{14}
$$

Once the forecast values for $\hat{y}_{j,T+1}, \hat{y}_{j,T+2}, ..., \hat{y}_{j,T+16}$ are available up to time $T+16$, one can construct the common component of each series (using the bilateral filters $\{K_h\}_{h=-16}^{16}$ estimated with data up to time T), getting $\hat{\chi}_{j,T-16+1}, \hat{\chi}_{j,T-16+2}, ..., \hat{\chi}_{j,T}$ and the coincident and leading indicators based on averages of these filtered components (see COI and LEA equations in sections 5 and 7).

This procedure can be seen as a multivariate generalization of the method usually adopted when applying bilateral filters to single time series (see for example the software TRAMO-SEATS or X12-REGARIMA for univariate seasonal adjustment). A direct consequence of this solution is that from time $T-M+1$ up to time T the indicators are subject to revisions. If revision errors are large and decrease only slowly as new data are added, then we cannot rely on our indicators for business cycle analysis since the provisional figures can experience large changes until sufficient new data are accumulated. Therefore what matters is how precise are the provisional estimates and when the signal about the cyclical situation can be considered reliable. Using information up to time T and the M-steps ahead projections - as was just described - one obtains M provisional estimates for the indicators. Let $I_{T/T}$ be the *concurrent* estimator (i.e. the estimator of index I at time T based on information up to time T, where by this we mean observations on y_t process) and $I_{T-k/T}$ the provisional estimate for time $T-k$.[24] When $k > M$ the estimator is *final* and can be simply written as I_{T-k} without reference to its information content. The concurrent estimator is based on sixteen forecasted values; as new observations are added, the provisional estimates converge to the final one. It is these process that we want to study.

To analyze the size of revision errors we run a simulation for the period 1992:12 - 1998:12 progressively increasing the sample (first we considered a sample ranging from 1985:6 to 1992:12, then from 1985:6 to 1993:1 and so on)[25]. At each step the VARMA was re-estimated and 16 steps ahead forecast values for c_t and l_t obtained. Next, on the basis of these values, 16 steps ahead forecasts for y_t were constructed from the ARX models. Finally, using the weights estimated on the whole sample, we built the filtered variables ($\chi_{j,t}$) and on the basis of the filtered values we constructed - for each step of the simulation - a new set of coincident a leading indicators. Hence at each step we can compute the revision error for the concurrent estimator $I_{t/t}$ and the 15 provisional ones $I_{t-k/t}$:

(15) $$r_{t-k/t} = I_{t-k} - I_{t-k/t}$$

with $k = 1, 2, ..., 15$. The simulations enabled us to evaluate the forecasting performance of the VARMA and of the equations used to project forward the $y's$ (see Table 4-5) and to construct confidence bounds around leading and coincident indicators, based on recorded revision errors (see Figure 4-7).

The statistics relative to the VARMA forecasts show Theil's U ranging from 0.8 (one step ahead) to 0.6 (16 steps ahead) and RMSE going from 0.15-0.20 to 0.33-0.24 for the coincident and leading index (c_t and l_t), respectively. The F-tests performed on the univariate ARX model used in the projections of the $y's$ reveal that the indicators c_t and l_t enter significantly at a 95 per cent confidence level in more than 450 cases (at a 90 per cent level the number increases to over 500). This notwithstanding the RMSE of forecasts are very close to those obtained with

---

24 Here $I_{T-k/T}$ stands for $coi_{T-k/T}$ and $lea_{T-k/T}$ for the coincident and the leading index respectively.

25 Since our sample has observations up to 2000:5 we stopped the simulation in 1998:12 to have a complete set of *final* estimates of our indicators to be compared with the provisional one.

simple AR models with the same lag truncation.[26] It is important to remark that this result have only a limited influence on the final revision errors, since the business cycle indicators are build by averaging filtered variables and the latter too imply averages of all variables. As it could be anticipated, this aggregation process greatly reduces the forecast errors: comparing final filtered versions of the variables (i.e. the common components $\chi_{j,t-k}$) with their provisional estimates $\hat{\chi}_{j,t-k/t}$ the RMSE are much lower (from 0 to 0.10, see Table 6).

The RMSE of the revision in the case of our reference coincident indicator is 0.5 for the concurrent estimator and narrows to just above 0.01 for the ”pre-final” estimate (see the first panel of Figure 7). For the leading and the ”NBER” coincident indicators the results are better since they entail the averaging of a larger set of variables (see second and third panel of Figure 7). As a final step we constructed 95 per cent confidence bands around our indicators for the period January 1999 to May 2000 (last available common date for the series included in the panel). We can conclude that the large errors at the very end of the estimation period for the reference coincident indicator should induce caution in interpreting the signal, on the other hand the confidence bands appear to be much narrower in the case of the leading and the ”NBER” coincident indicator (see Figure 4-7). Hence, reading the concurrent estimate of the reference indicator in conjunction with that of the leading and of the NBER indicators can significantly contribute to the detection of a reliable signal even at the end of the sample.

A further check on the ability of the indicators to capture at an early time the changes in the cyclical situation of the euro area is obtained through the analysis of their behaviour in simulation around turning points. Results in this crucial aspect of the performance of our indicators are encouraging. The ability to detect the turning points varies from case to case.[27] In January 1993 the *final* coincident indicator signals a trough. The turning point was detected correctly in January with a delay of only six months; the previous estimates of the indicator anticipated it in November 1992. This result is however reassuring since the false signal resulted in a slight anticipation of the upturn (see Figure 8). A much better result would have been obtained in November 1995, when the final coincident indicator shows another trough.

---

[26] This result might imply that there is room to improve our forecasting by considering a finer distinction of indicators and a better model to forecast them prior to their use in the ARX models. Tables with detailed results from the simulations are available upon request.

[27] We analysed the upturn in activity in early 1993, the slowdown in september 1994, the upturn in november 1995 and the slowdown at the end of 1997.

In this case since the very beginning, preliminary estimates correctly anticipated the upturn (see Figure 9). In general, at the other turning points episodes the provisional estimator of the coincident indicator had a performance that lies between the two extremes just discussed.

9. Conclusions

In the present paper we use a well defined statistical model (the Generalised Dynamic Factor Model) to analyze a large cross section of macroeconomic variables for the main European countries and to identify common cyclical movements in the euro area.

An important by-product of the present work is the construction of a large databank containing monthly series for a wide set of economic phenomena regarding the major euro area countries. These series were not only collected from many different sources (up to now no euro area databank exits with a coverage comparable to those available for the US) but also selected according to criteria of 'minimum harmonization' to guarantee cross-country comparability.

We constructed a coincident and a leading business cycle indicator for the euro area retaining the basic NBER idea of exploiting a large set of data while providing a rigorous foundation to the method. Our business cycle indicators are build on the basis of monthly data. In particular the euro area reference cycle (or coincident indicator) is based on the weighted average of the common components of GDPs of the main six euro area countries, which, though the GDP is observed quarterly, we can consistently estimate at monthly level by exploiting the cross sectional information. This is an absolute novelty with respect to the existing literature, that is mainly concerned with quarterly data. The comparative advantage connected to the use of higher frequency statistics is that they are available with a much shorter delay (for instance, monthly figures for industrial production and prices are generally available within two months, whereas national account data are usually disseminated with a delay of 3 to 5 months).

Our monthly business cycle coincident indicator (available for the period 1985-2000) reveals that four major cyclical episodes affected the euro area since the end of the eighties. This is a striking difference with respect to the US economy, that over the same period was affected only by a minor recession (in 1990-91) followed by 9 years of uninterrupted expansion.

We also show that, overall, our coincident index has a better performance with respect to other indicators available for the euro area, like the one published by the European Commission. In particular it provides a sharper signal and regularly anticipates the cyclical turning points with respect to the EC index allowing an earlier detection of the changes in growth prospects.

The methodology adopted allows us to study in a coherent way all the variables included in our panel (almost 800), isolating those showing leading properties with respect to the business cycle from those that are coincident or lagging. Some interesting patterns do show up. The industrial production seems to be coincident or marginally leading, while some particular sectors (pulp & paper and chemical) present systematic leading properties across Europe. As could be expected, the European Commission business and consumers surveys were confirmed to be relevant information sources for the business cycle analysis; nonetheless attention has to be paid in interpreting the results, given that the signals they release are quite noisy. Our finding confirms the common wisdom concerning leading properties of the belgian production surveys, although this property is not shared by the belgian industrial production. Production and consumer prices share a lot of commonality with the aggregate cycle and are coincident or lagging it.

A leading index was constructed as simple average of the common components of those variables having a lead of at least five months. Therefore a five steps ahead forecast of the coincident indicator can be derived from the leading one. The in-sample fit of these forecast is very good and well anticipates turning points. However the current (i.e. end of the sample) estimates of the indexes can be unreliable since they are partly based on forecast values. As new information becomes available, the indexes can be subject to large revisions. On the other hand it is exactly at the end of the sample that one mostly needs advice. Therefore we performed an in-sample exercise to derive a robustness check of the conclusions that can be reached on the basis of current (end of the sample) estimates of the indexes. This exercise proved that the error associated with the preliminary estimates of the indicators decreases rapidly as new information becomes available, furthermore especially for the leading index, it does not appear to affect the current estimates in a measure that would prevent their use at the end of the sample.

When the present exercise was started, data up to may 2000 were available. The indication derived from our coincident and leading indexes pointed to a slowdown in the economic activity in the euro area in the second half of 2000.

Appendix A: The dataset

This appendix describes the principal guidelines followed setting up the database and, in particular, each of the blocks into which it has been split. As already noticed in Section 3.1, 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 comprehensively account of the economic phenomena emerging from the largest European countries (see Table 1 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 held therein: this is obviously a foundamental 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, responding 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 early 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 this 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, an ample 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 surveys 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.) surveys 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 this 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, revealed to be useful.

It has been particularly difficult to obtain Labour market variables satisfying the requirements listed in Section 3.1 and needed to 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.

Appendix B: The construction of the indicators

This appendix presents the steps followed in the construction of the business cycle indicators, illustrating the practical solutions adopted in the implementation of the FHLR method.[28] The derivation of the indicators is based on the theoretical model presented in Section 2, whose notation is maintained also here; the derivation can be split in the following steps:

1. Preparation of the dataset and stationary transformation of the time series

The first step has been already detailed in the main text and in appendix 1, therefore here we only recall the fact that all series have been seasonally adjusted, corrected for the presence of outliers and transformed to induce stationarity when necessary. Finally each time series has been demeaned and divided by its standard deviation to prevent scale effects on the measurement of the influence of the series on the cycle. Thus one obtains an n-dimensional $(n = 794)$ jointly stationary vector stochastic process y_t, of which a sample of length $T = 180$ is observed.

2. Estimation of the spectral matrix of the cross section

The estimation of its $(n \times n)$ spectral matrix Σ_y is accomplished by computing first the covariance matrices Γ_k of the process y_t up to lag M and deriving then the matrices $\widehat{\Sigma}_y(\theta_j)$ for some $\theta_j \in [0; \pi]$ via Fourier transform:

$$\Gamma_k = \frac{1}{T-k} \sum_{t=k}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y}) \qquad \text{for } k = 0, 1, 2, ..., M \tag{B1}$$

[28] The actual computations have been performed with a FORTRAN routine that follows almost exactly steps 2 to 5 of this appendix, calling two programs from the NAG library, one for the eigenvalue problem of step 3 (NAG f02haf) the other for the integration required in the Fourier inverse transformation as explained in step 5 (NAG *d01gas*). Programming in FORTRAN proved to be a key improvement since the FORTRAN code dramatically reduced the CPU time and memory requirements with respect to the other programs previously developed (in MatLab and in Speakeasy), running more then 10 times faster. The time required to perform the entire exercise on 800 variables, with 16 lags in computing correlation and spectral matrices and 101 points in the $[-\pi; \pi]$ interval is about 1 hour on a 400Mhz PENTIUM with 240Mb of RAM.

where $\bar{y}=\frac{1}{T}\sum_{t=1}^{T}y_t$ and Γ_{-k} equals Γ_k'. The spectral matrix at frequency θ_j is:

$$\hat{\Sigma}_y\left(\theta_j\right)=\frac{1}{2\pi}\sum_{k=-M}^{M}w_k\cdot\Gamma_k\cdot e^{-i\theta_j k}\quad\text{where}\quad j=-ns,...,0,...,ns \tag{B2}$$

where $w_k=1-\frac{|k|}{M+1}$(Bartlett's window). Consistent estimation of $\mathbf{\Sigma}(\theta_s)$ is ensured, provided that $M(T)\rightarrow\infty$ and $M(T)/T\rightarrow 0$, as $T\rightarrow\infty$. $\hat{\Sigma}_y$ is evaluated at 101 equally spaced frequencies in the interval $[-\pi,\pi]$ i.e. $\theta_j=\frac{2\pi j}{100}$.
The choice of the truncation lag for the estimation of the covariances and therefore of the spectral matrix proved to be critical; we opted for a truncation lag, $M=16$. FHLR (2000) propose a simulation based rule of thumb for M, setting it equal to integer part of $\frac{\sqrt{T}}{4}$; their simulations involve simple AR and MA processes that are a reasonable benchmark for quarterly data but does not seem to the satisfactory in the case of monthly data.
Another discretionary element of the procedure is the number of points in the $[-\pi,\pi]$ interval in which the Σ_y is evaluted. In principle one can compute the spectral matrix for any number of points in the interval. Our choice was guided by the need for a sufficient basis to evaluate the spectra at business cycle frequencies (with 51 points in the interval $[0;\pi]$ there are 7 frequencies implying periodicity greater than 14 months, with quarterly data 20 points would suffice to have 7 frequencies at periodicity greater than 5 quarters).

3. Choice of the number of common factors and estimation of the corresponding dynamic principal components

The choice of the number of common factors (dynamic principal components) responsible for the comovement of the series is based on the amount of variance that they explain at cyclical frequencies and on their behaviour as n increases. For each frequency θ_j the spectral density matrix, being positive semi- definite, can be factorized as:

$$\hat{\Sigma}_y\left(\theta_j\right)=U\left(\theta_j\right)\cdot\Lambda\left(\theta_j\right)\cdot U^*\left(\theta_j\right) \tag{B3}$$

where $U\left(\theta_j\right)$ is the $(n\times n)$ matrix of eigenvectors (and $U^*\left(\theta_j\right)$ its conjugate transpose so that $U\left(\theta_j\right)\cdot U^*\left(\theta_j\right)=I$) and $\Lambda\left(\theta_j\right)$ is the diagonal matrix containing the n eigenvalues in descending order (i.e. $\lambda_1(\theta_j) \geq \lambda_2(\theta_j) \geq ... \geq \lambda_n(\theta_j)$).[29] The first q dynamic principal components are then obtained as combinations of the leads and lags of the variables y_t with weights based on the first q eigenvectors. Letting $p^1(\theta_j) = u_1^*(\theta_j)$ be the first eigenvector of dimension $n \times 1$, the weights $p^1(L)$ for the first dynamic principal component f_{1t} are computed through the inverse Fourier transform of $p^1(\theta_j)$:

$$f_{1t} = \underset{(1\times n)}{p^1(L)} \cdot y_t = \sum_{k=-K}^{K} \underset{(1\times n)}{p_k^1} \, y_{t-k} = \sum_{j=1}^{n} \sum_{k=-K}^{K} p_{j,k}^1 \, y_{j,t-k}, \tag{B4}$$

where the weight on the j-th variable at lag k is given by:

$$p_{j,k}^1 = \int_{-\pi}^{\pi} e^{i\theta k} p_j^1(\theta)\, d\theta. \tag{B5}$$

and the size of the bilateral filter is K. The integration in (16) has not been performed on the entire interval $[-\pi; \pi]$, but on the *sub-interval of frequencies corresponding to cyclical periodicity*, i.e. 14 to 120 months. The above numerical integration has been performed using a fourth order polynomial between a grid of points of the interval of integration and provides optimal treatment of the end-points.[30]

4. Cleaning the data (isolating the part of the data variation explained by the factors)

To isolate the total variation explained by the factors, the variables are projected on the past, present and future of the dynamic principal components. It is possible to prove that, if the variable is projected on the common factors, the polynomial term can be expressed in function of the eigenvectors and the common component $\chi_{j,t}$ of the j-th variable (being $q = 2$) is given by:

$$\begin{aligned} \chi_{j,t} &= p_j^1(F)\, f_{1,t} + p_j^2(F)\, f_{2,t} \\ &= p_j^1(F)\, p^1(L)\, y_t + p_j^2(F)\, p^2(L)\, y_t \\ &= K_j^2(L)\, y_t \end{aligned} \tag{B6}$$

---

[29] The eigenvalue problem at each frequency ω_j has been solved by calling the NAG subroutine *f02haf*.

[30] In our procedure this integration is performed via the NAG *d01gas*.

with polynomials in the lag operator obtained by Fourier inversion (on the cyclical band) of the first two eigenvectors of the spectral matrix. It is worth noting that the common components $\chi_{j,t}$ are a combination of past, present and future values of the factors $f_{1,t}$ and $f_{2,t}$ The weights in (16) and those in the last expression of (16) are linked by a reverse relation: the common component of the j-th variable, $\chi_{j,t}$, loads the k-th lag of the first common factor with weight $p^1_{j,-k}$ while the first common factors, $f_{1,t}$, loads the k-th lag of variable $y_{j,t}$ with weight $p^1_{j,k}$.

5. Construction of the coincident indicator - COI

The coincident indicator is the weighted average of the common componets of the GDP of the six major countires, where the weights (PPP at the base year) are 0.37, 0.23, 0.19, 0.09, 0.07 and 0.05 for Germany, Franch, Italy, Spain, The Netherland and Belgium.

6. Definition of the pro- or counter- cyclicality of each variable and classification of the variables according to the time lead/lag properties with respect to the coincident indicator.

The classification of the ’cleaned’ variables as pro- or counter- cyclical is based on the computation of their maximum correlation with the coincident indicator COI (i.e. the weighted average of the common components of the GDP’s). A positive sign is interpreted as procyclical behaviour, a negative sign as counter cyclical bahaviour. Next, the lead and lag relationships of the $\chi' s$ w.r.t. COI are established based on the displacment of the maximal correlation (the time lead or alg at which the correlation is maximized).

7. Construction of the indicators

Finally the leading indicator LEA_t is obtained as a simple average of the common components of the variables classified as leading. The indicator $LCOI_t$ is instead obtained by regressing the coincident indicator on LEA_t and on itself using the following equation:

$$LCOI_t = c + \sum_{j=5}^{8} a_j COI_{t-5-j} + \sum_{j=5}^{9} b_j LEA_{t-5-j} + \varepsilon_t$$

where the starting and truncation lags have been chosen on the basis of the goodness of fit of the equation.

Tables and Figures

Country \ Block	*Industrial Production*	*Producer Prices*	*Consumer Prices*	*Money Int. Rates*	*Surveys*	*Trade*	*Labor*	*Other Variables*
Belgium	Eurostat, OECD	Eurostat	Eurostat, OECD	ECB, BIS	European Commission, Datastream	BIS, OECD	BIS, OECD	BIS, OECD, Datastream
France	Eurostat, OECD	Eurostat	Eurostat, OECD	ECB, BIS	European Commission, Insee	BIS, OECD	BIS, OECD	BIS, OECD, Datastream
Germany	Eurostat, OECD	Eurostat	Eurostat, OECD	ECB, BIS	European Commission, Ifo	BIS, OECD	BIS, OECD	BIS, OECD, Datastream
Italy	Eurostat, Istat	Eurostat, Istat	Eurostat, Istat	ECB, BIS	European Commission, Isae	BIS, OECD	BIS, OECD	BIS, OECD, Datastream
The Netherlands	Eurostat, OECD	Eurostat	Eurostat, OECD	ECB, BIS	European Commission, Datastream	BIS, OECD	BIS, OECD	BIS, OECD, Datastream
Spain	Eurostat, OECD	Eurostat	Eurostat, OECD	ECB, BIS	European Commission, Datastream	BIS, OECD	BIS, OECD	BIS, OECD, Datastream
Other Countries	OECD	OECD	OECD	OECD	OECD	OECD	OECD	BIS, OECD, Datastream, IMF

Table 1 - Data sources

Block	*Industrial Production*	*Producer Prices*	*Consumer Prices*	*Money Int. Rates*	*Surveys*	*Trade*	*Labor*	*Other Variables*
Number of Original Variables	200	100	80	130	600	60	80	150
Final Dataset	144	95	36	147	170	35	45	122

Table 2 - Composition of the data set

Block / Country	All		Leading			Coincident			Lagging		
	Variance Ratio	*Delay*	*within block %*	*Variance Ratio*	*Delay*	*within block %*	*Variance Ratio*	*Delay*	*within block %*	*Variance Ratio*	*Delay*
Industrial Production	0.27	-2	49	0.32	-4	37	0.3	-1	14	0.13	7.5
Producer Prices	0.43	2	13	0.27	-8	42	0.46	0	45	0.45	7
Consumer Prices	0.41	3	27	0.27	-8	23	0.32	0	50	0.61	10
Money Int. Rates	0.4	3	17	0.38	-8	28	0.36	-1	55	0.42	6
Surveys	0.33	-3	53	0.37	-4	36	0.34	-2	11	0.19	6
Trade	0.18	-2	43	0.15	-3	20	0.31	-2	37	0.13	11
Labor	0.56	0	20	0.3	-5	60	0.58	0	20	0.43	4
Other Var.	0.24	-2	45	0.25	-5	34	0.23	1	22	0.24	9
All	0.34	-1.0	36	0.3	-5.7	36	0.39	-0.6	28	0.34	7.6

Table 3 - Lead/Lag Relationships and Explained Variance

Steps	RMSE	THEIL-U	RMSE	THEIL-U
1	0.161	0.823	0.195	0.842
2	0.161	0.843	0.199	0.856
3	0.189	0.837	0.192	0.792
4	0.208	0.826	0.192	0.787
5	0.210	0.728	0.195	0.746
6	0.217	0.712	0.204	0.757
7	0.232	0.665	0.203	0.701
8	0.251	0.667	0.208	0.731
9	0.265	0.663	0.211	0.691
10	0.279	0.648	0.215	0.686
11	0.290	0.633	0.226	0.691
12	0.303	0.636	0.235	0.689
13	0.316	0.642	0.231	0.686
14	0.335	0.655	0.242	0.636
15	0.329	0.650	0.239	0.699
16	0.325	0.638	0.238	0.694

Table 4 - Forecast performance of the VARMA model

STEPS	THEIL's U (overall) *min*	*ave*	*max*	THEIL's U (QUANTILES) 25%	50%	75%
1	0.38	0.73	1.39	0.60	0.75	0.86
2	0.40	0.79	1.52	0.72	0.77	0.83
3	0.54	0.79	1.64	0.72	0.77	0.83
4	0.44	0.77	1.65	0.70	0.75	0.82
5	0.46	0.77	1.58	0.71	0.74	0.80
6	0.59	0.77	1.52	0.71	0.74	0.81
7	0.45	0.76	1.44	0.69	0.73	0.79
8	0.45	0.77	1.39	0.70	0.74	0.80
9	0.49	0.77	1.39	0.69	0.74	0.81
10	0.41	0.75	1.41	0.67	0.72	0.79
11	0.46	0.77	1.39	0.69	0.74	0.80
12	0.48	0.72	1.52	0.64	0.69	0.75
13	0.50	0.75	1.45	0.68	0.72	0.78
14	0.50	0.74	1.39	0.67	0.71	0.77
15	0.49	0.74	1.43	0.67	0.71	0.76
16	0.51	0.74	1.46	0.66	0.71	0.77

Table 5 - ARX models

STEPS	RMSE for CLY *coinc.*	*lead.*
1	0.00	0.00
2	0.01	0.01
3	0.01	0.01
4	0.02	0.01
5	0.02	0.02
6	0.03	0.02
7	0.04	0.03
8	0.05	0.04
9	0.06	0.04
10	0.07	0.05
11	0.08	0.05
12	0.08	0.06
13	0.09	0.07
14	0.10	0.07
15	0.11	0.08
16	0.11	0.09

Table 6 - common components ($\chi' s$)

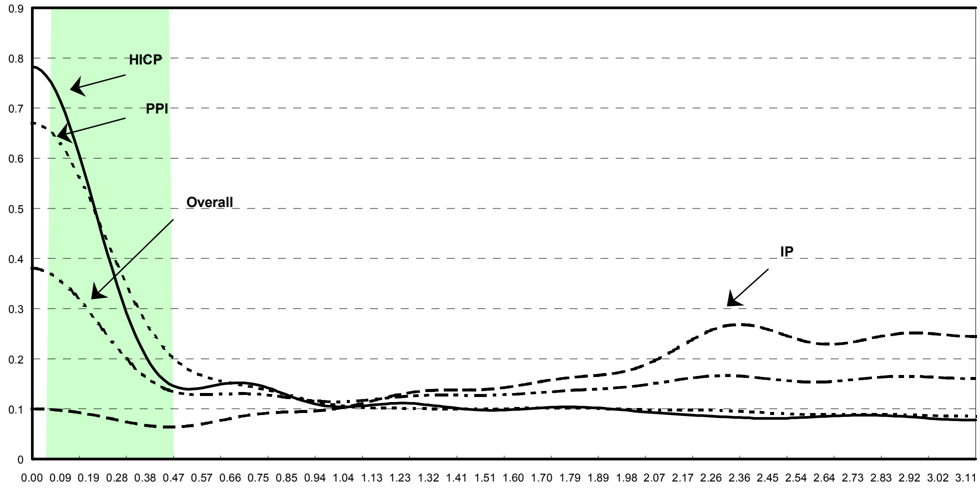


Figure 1- Typical spectral shape for the main sectors included in the panel (average spectra of the series)

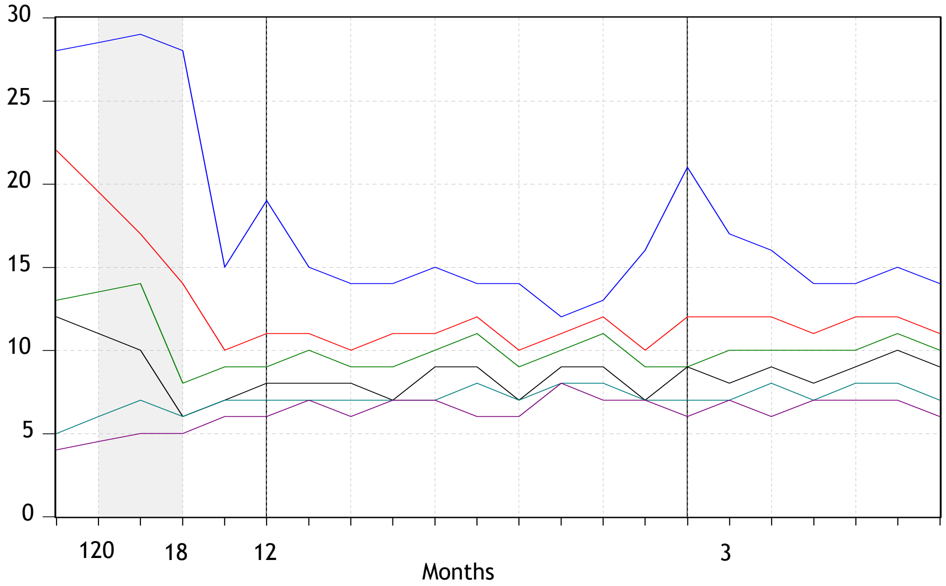


Figure 2 - Variance explained by the first six eigenvalues of the spectral matrix at various frequencies

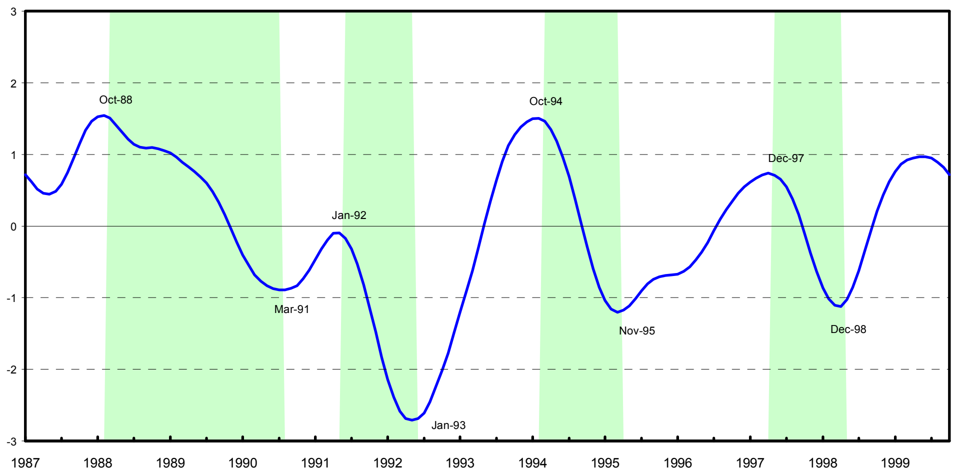


Figure 3 - Coincident indicator for the Euro area and business cycle dating

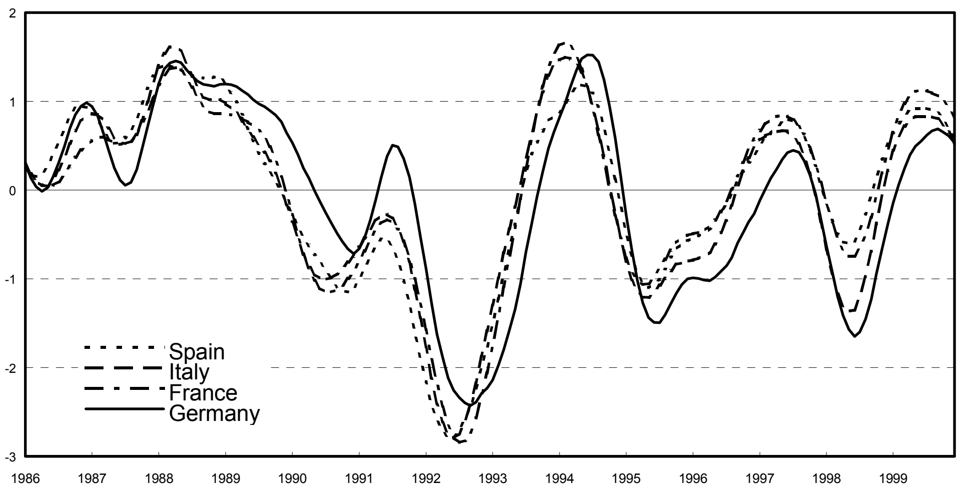


Figure 4 - Common components of the GDP of the main euro area countries

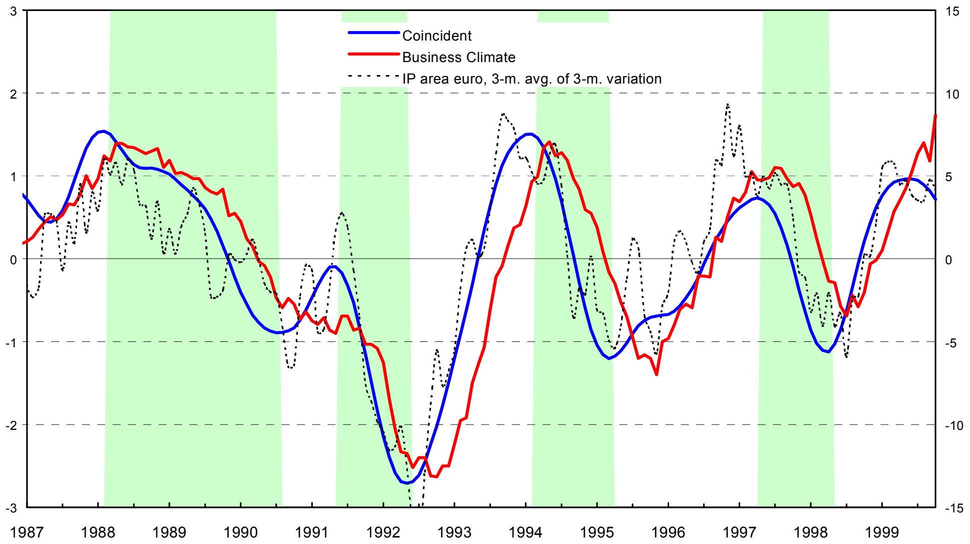


Figure 5 - Comparison with EC business climate indicator and euro area industrial production

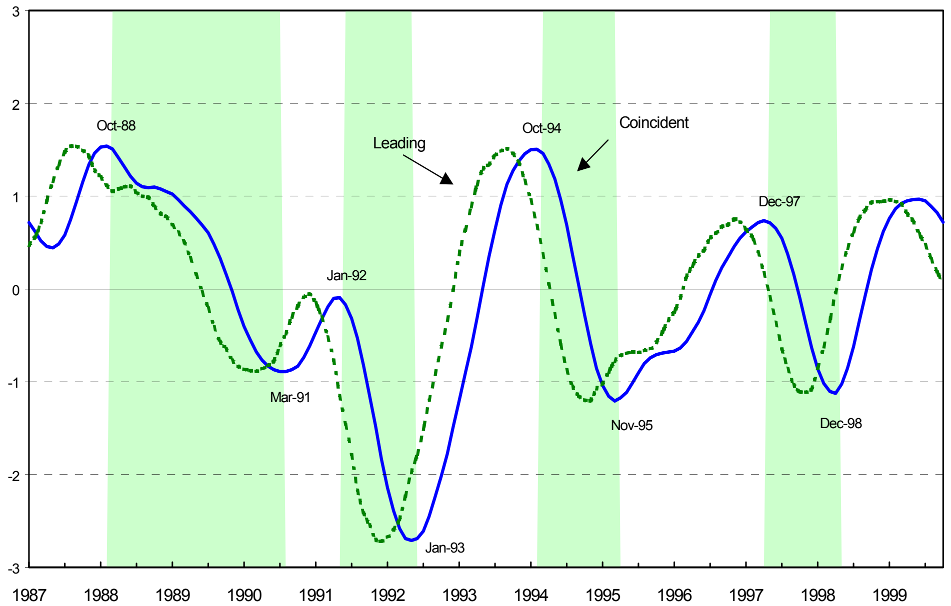
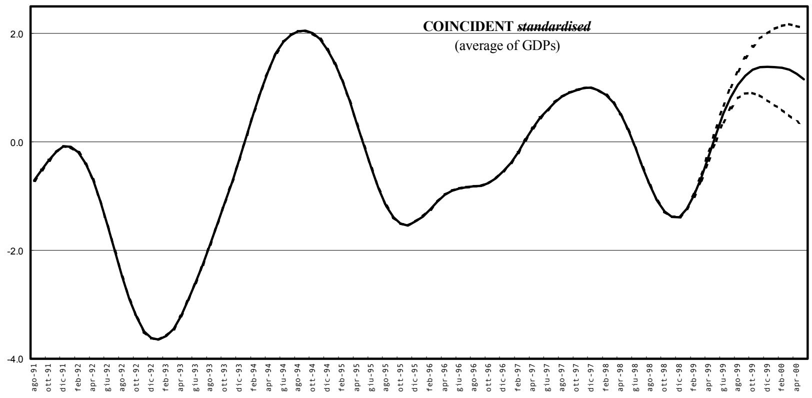
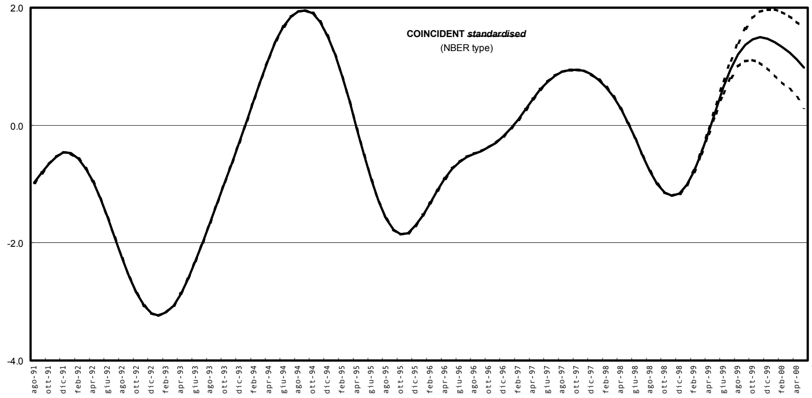


Figure 6 - Euro area coincident and leading indicator

95% Confidence bounds for the coincident indicator (GDP average)
(percentage variation)



95% Confidence bounds for the coincident indicator (NBER type)
(percentage variation)



95% Confidence bounds for the leading indicator 1
(percentage variation)

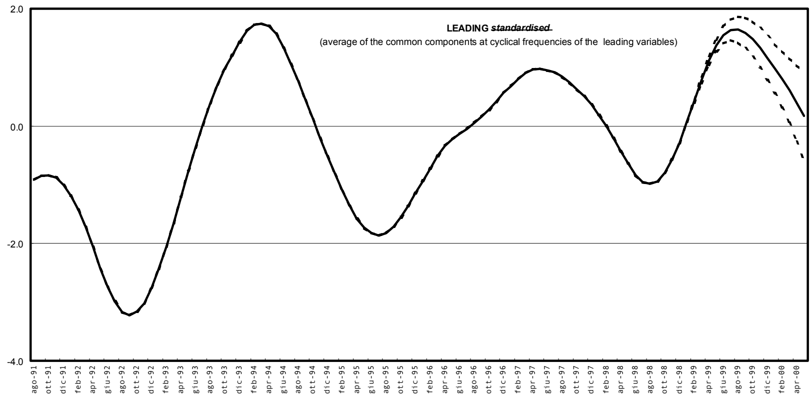


Figure 7 - 95% Confidence bounds

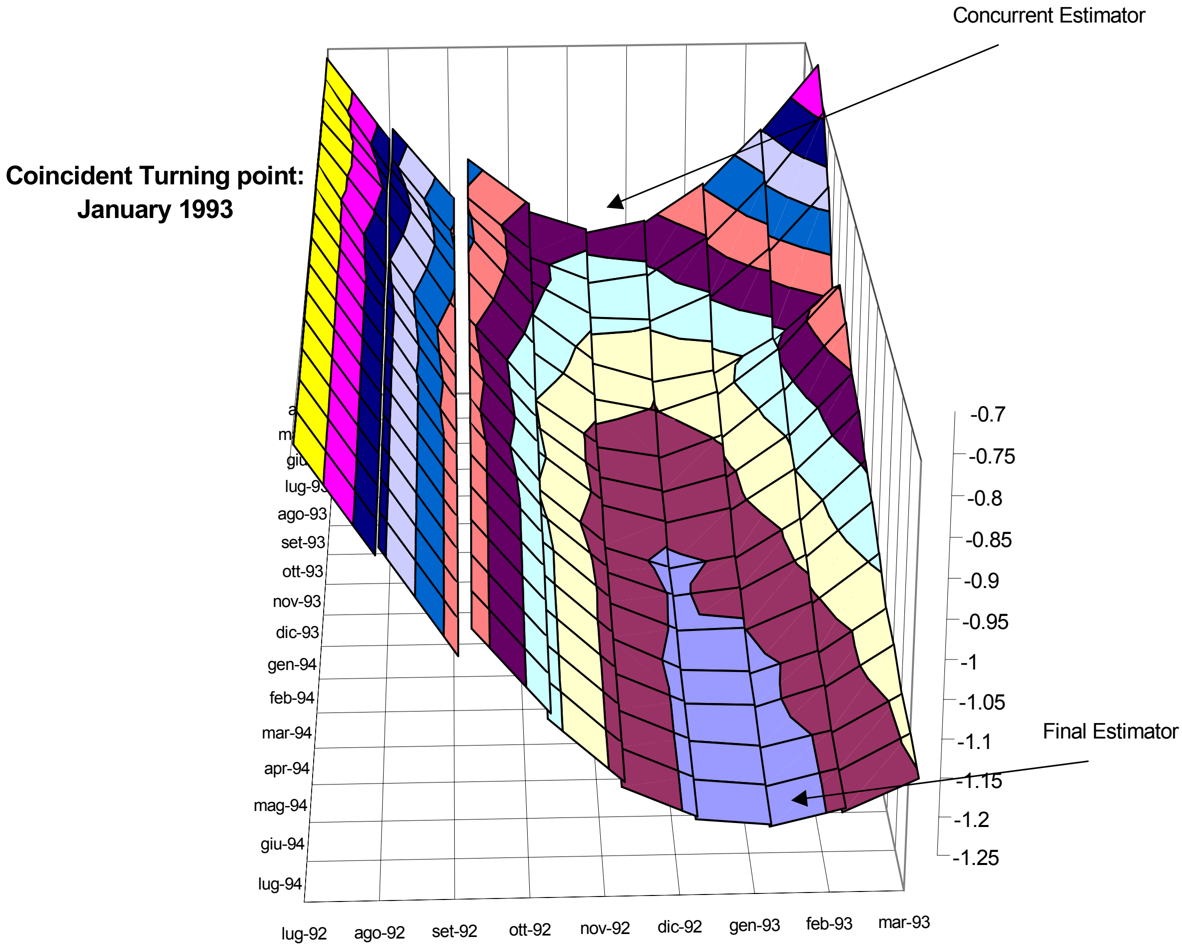


Figure 8 - Turning Points Detetction (Jan. 1993 episode)

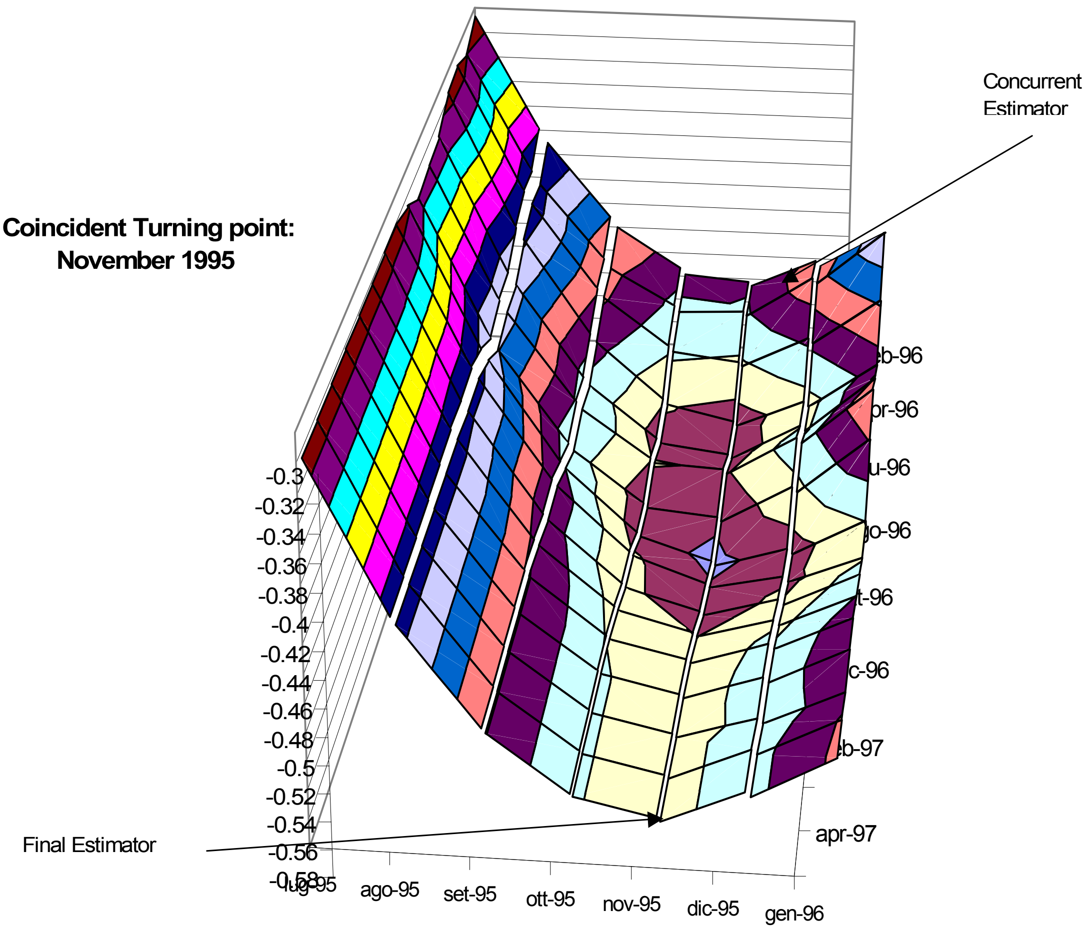


Figure 9 - Turning point detection (Nov. 1995 episode)

References

Altissimo, F., Marchetti, D.J. and Oneto,G.P., (2000), “The Italian Business Cycle: Coincident and Leading Indicators and Some Stylized Facts”, Bank of Italy, *Temi di Discussione*, N. 377.

Artis, M., Krolzig, H.M. and Torr, J., (1999), “The European Business Cycle”, European University Institute, Working Paper, No.24.

Baxter, M. and King, R.G., (1999), “Measuring Business Cycles: Approximate Band- Pass Filters for Economic Time Series”, *Review of Economics and Statistics*, Vol.81(4), 575-593.

Beaulieu, J.J., Makie-Mason, J.K. and Miron, J.A., (1992), “Why Do Countries and Industries With Large Seasonal Cycles Also Have Large Business Cycles?”, *Quarterly Journal of Economics*, No.2 , pp. 621-656.

Bowden, R.J. and Martin, V., (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.

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

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

Bry, G. and Boschan, C., (1971), “Cyclical Analysis of Times Series: Seelected Procedures and Computer Programs”, NBER, Technical Paper, No.20.

Burns, A.F. and Mitchell, W.C., (1946), " Measuring Business Cycles”, New York: NBER.

Cheung, Y. and Westermann, F., (1999), “Output Dynamics of the G7 countries—Stochatic Trends and Cyclical Movements”, CES-ifo Working Paper, No.220.

Chamberlain, G. and Rothschild, M., (1983), “Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets”, *Econometrica*, No.5, 1281-1304.

Dickerson, A.P., Gibson, H.D. and Tsakalotos, E., (1998), “Business Cycle Correspondence in the European Union”, *Empirica*, Vol.25, 51-77, 1998.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L., (2000), “The Generalized Factor Model: Identification and Estimation”, *The Review of Economic and Statistics*, Vol.82(4), 540-554, forthcoming.

Forni, M. and Lippi, M., (2000), “The Generalized Factor Model: Representation Theory”, *Econometric Theory*, forthcoming.

Forni, M. and Reichlin, L., (2001), "Federal Policies and Local Economies: Europe and the US", *European Economic Review*, No.45, 109-134.

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

Lumsdaine, R.L. and Prasad, E.S., (1999), "Identifying the Common Component in International Economic Fluctuations: A New Approach", IMF Working Paper, No.154.

McDermott, C.J. and Scott, A., (2000), "Concordance in Business Cycles", IMF Working Paper, No.37.

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.

Granger, C.W.J. and Hatanaka, M., (1964), *Spectral Analysis of Economic Time Series*, Princeton: Princeton University Press.

Marchetti, D.J. and 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*, Vol.19, 419-440.

Marcellino, M., Stock, J.H. and Watson, M.W., (2000), "Macroeconomic Forecasting in the Euro Area: Country Specific Versus Area-Wide Information", unpublished manuscript.

Sargent, T.J., (1987), *Macroeconomic Theory*, Second edition, Academic Press.

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.

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

Stock, J.H. and Watson, M.H., (1999), "Diffusion indexes", unpublished manustript.

Zarnowitz, V., (1992), *Business Cycles. Theory, History, Indicators and Forecasting*, Chicago, The University of Chicago Press.

RECENTLY PUBLISHED "TEMI" (*)

No. 410 — *International Transmission Via Trade Links: Theoretically Consistent Indicators of Interdependence for Latin America and South-East Asia*, by C. BENTIVOGLI and P. MONTI (June 2001).

No. 411 — *Why is the Business-Cycle Behavior of Fundamentals Alike Across Exchange-Rate Regimes?*, by L. DEDOLA and S. LEDUC (August 2001).

No. 412 — *Political Institutions and Policy Outcomes: What are the Stylized Facts?*, by T. PERSSON and G. TABELLINI (August 2001).

No. 413 — *Equilibrium Welfare and Government Policy with Quasi-Geometric Discounting,* by P. KRUSELL, B. KURUŞÇU and A. A. SMITH, Jr. (August 2001).

No. 414 — *Insurance within the Firm*, by L. GUISO, L. PISTAFERRI and F. SCHIVARDI (August 2001).

No. 415 — *Limited Financial Market Participation: A Transaction Cost-Based Explanation,* by M. PAIELLA (August 2001).

No. 416 — *Decentramento fiscale e perequazione regionale. Efficienza e redistribuzione nel nuovo sistema di finanziamento delle regioni a statuto ordinario,* by G. MESSINA (August 2001).

No. 417 — *Personal Saving and Social Security in Italy: Fresh Evidence from a Time Series Analysis,* by F. ZOLLINO (August 2001).

No. 418 — *Ingredients for the New Economy: How Much does Finance Matter?*, by M. BUGAMELLI, P. PAGANO, F. PATERNÒ, A.F. POZZOLO, S. ROSSI and F. SCHIVARDI (October 2001).

No. 419 — *ICT Accumulation and Productivity Growth in the United States: an Analysis Based on Industry Data,* by P. CASELLI and F. PATERNÒ (October 2001).

No. 420 — *Barriers to Investment in ICT,* by M. BUGAMELLI and P. PAGANO (October 2001).

No. 421 — *Struttura dell'offerta e divari territoriali nella filiera dell'information and communication technologies in Italia,* by G. IUZZOLINO (October 2001).

No. 422 — *Multifactor Productivity and Labour Quality in Italy, 1981-2000,* by A. BRANDOLINI and P. CIPOLLONE (October 2001).

No. 423 — *Tax reforms to influence corporate financial policy: the case of the Italian business tax reform of 1997-98,* by A. STADERINI (November 2001).

No. 424 — *Labor effort over the business cycle,* by D. J. MARCHETTI and F. NUCCI (November 2001).

No. 425 — *Assessing the effects of monetary and fiscal policy,* by S. NERI (November 2001).

No. 426 — *Consumption and fiscal policies: medium-run non-Keynesian effects,* by G. RODANO and E. SALTARI (November 2001).

No. 427 — *Earnings dispersion, low pay and household poverty in Italy, 1977-1998,* by A. BRANDOLINI, P. CIPOLLONE and P. SESTITO (November 2001).

No. 428 — *Nuove tecnologie e cambiamenti organizzativi: alcune implicazioni per le imprese italiane,* by S. TRENTO and M. WARGLIEN (December 2001).

No. 429 — *Does monetary policy have asymmetric effects? A look at the investment decisions of Italian firms,* by E. GAIOTTI and A. GENERALE (December 2001).

No. 430 — *Bank-specific characteristics and monetary policy transmission: the case of Italy,* by L. GAMBACORTA (December 2001).

No. 431 — *Firm investment and monetary transmission in the euro area,* by J. B. CHATELAIN, A. GENERALE, I. HERNANDO, U. VON KALCKREUTH and P. VERMEULEN (December 2001).

No. 432 — *Financial systems and the role of banks in monetary policy transmission in the euro area,* by M. EHRMANN, L. GAMBACORTA, J. MARTÍNEZ-PAGÉS, P. SEVESTRE and A. WORMS (December 2001).

No. 433 — *Monetary policy transmission in the euro area: what do aggregate and national structural models tell us?,* by P. VAN ELS, A. LOCARNO, J. MORGAN and J.P. VILLETELLE (December 2001).

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