No. 2400

REFERENCE CYCLES: THE NBER METHODOLOGY REVISITED

Mario Forni, Marc Hallin, Marco Lippi and Lucrezia Reichlin

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

ISSN 0265-8003

REFERENCE CYCLES: THE NBER METHODOLOGY REVISITED

Mario Forni, Universitá di Modena and CEPR
Marc Hallin, Université Libre de Bruxelles and ECARES
Marco Lippi, Universitá di Roma
Lucrezia Reichlin, Université Libres de Bruxelles, ECARES and CEPR

Discussion Paper No. 2400
March 2000

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

CEPR Discussion Paper No. 2400

March 2000

ABSTRACT

Reference Cycles: The NBER Methodology Revisited*

This Paper proposes a new way to compute a coincident and a leading index of economic activity. The method provides a unified approach for the selection of the coincident and the leading variables, for averaging them into coincident and leading indexes and for the identification of turning points. The statistical framework we propose reconciles dynamic principal component analysis with dynamic factor analysis. We use our procedure to estimate coincident and leading indexes for the EMU area as well as country-specific indexes. Unlike other methods used in the literature, the country indexes take into consideration the cross-country as well as the within-country correlation structure.

JEL Classification: C13, C33, C43
Keywords: coincident and leading indicators, dynamic factor models, dynamic principal components

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

Marc Hallin
Université Libre de Bruxelles
ECARES
39 ave. F.D. Roosevelt
B-1050 Bruxelles
BELGIUM
Tel: (32 2) 650 5886
Fax: (32 2) 650 5899
Email: mhallin@ulb.ac.be

Marco Lippi
Dipartimento di Scienze Economiche
Università di Roma
via Cesalpino 12
I-00161 Roma
ITALY
Tel: (39 06) 4428 4202
Fax: (39 06) 4404 572
Email: lippi@giannutri.caspur.it

Lucrezia Reichlin
Université Libre de Bruxelles
ECARES
39 ave. F.D. Roosevelt
B-1050 Bruxelles
BELGIUM
Tel: (32 2) 650 4221
Fax: (32 2) 650 4475
Email: lreichli@ulb.ac.be

*This Paper is produced as part of a CEPR research network on New Approaches to the Study of Economic Fluctuations, funded by the European Commission under the Training and Mobility of Researchers Programme (Contract No ERBFMRX-CT98-0213).

Submitted 1 February 2000

NON-TECHNICAL SUMMARY

The aim of this Paper is to provide a practical method for computing coincident and leading indicators of economic activity for the EMU as an aggregate and for EMU countries. The procedure we propose is inspired by the NBER tradition, but it is based on a fully developed statistical model which is a generalization of the dynamic factor model and has a great theoretical appeal in that it provides a reconciliation of factor analysis with principal component analysis (Forni, Hallin, Lippi and Reichlin 1999, Forni and Lippi, 1999).

We pass through all the steps of the traditional NBER procedure – selection of the coincident and the leading variables, identification of the turning points, averaging and cleaning from the noise – although in a different order. We first eliminate, from each time series in the panel, that part of the dynamics that is poorly correlated with the rest of the economy and can hence be considered as idiosyncratic. Then, in a second step, we select coincident and leading indicators by analysing the phase shifts between these `cleaned' time series. Finally, we aggregate coincident and leading variables into coincident and leading indexes and establish turning points. A major novelty of our methodology is that these steps are not conceptually disjointed operations, but are all consistently nested within a unified theoretical setting.

The advantages of our method over standard techniques can be summarized as follows. First, we do not need to identify the turning points and select the coincident variables on the basis of judgmental criteria before cleaning them from noise, measurement errors and other idiosyncratic disturbances. Second, by retaining leading and lagging variables in the data set we are able to exploit additional information for the estimation of the coincident index. Third, from the analysis of a panel of time series for different countries, we can construct both an aggregate index and country-specific indexes; the latter are estimated by taking into account cross-country as well as within-country correlation. This is particularly interesting for the construction of indexes for the EMU area where it would be inappropriate to analyse countries in isolation from each other.

1. Introduction[1]

There are basically two approaches to construct coincident and leading indicators of the business cycle. The best known is due to the NBER tradition and is based on two steps: first, the identification of turning points on the basis of some judgmental criterion and the classification of each single variable as leading or coincident; second, on the computation of averages of leading and coincident variables in order to construct the relative indexes (Burns and Mitchell, 1946 and, for a general review, Zarnowitz, 1992). Through averaging the analyst can not only produce a synthetic measure of economic activity, but also eliminate measurement error and, in general, specific characteristics of a single time series which are poorly correlated with the rest of the economy. This approach, although not founded on any well defined probabilistic model, has proved to be a useful way to summarize information on the macroeconomy, is built on an enormous amount of experience and has provided knowledge on US business cycle and reference for the macroeconomic profession for years.

A second approach is founded on statistical modeling, specifically on index (or factor) models. Index models are based on the idea that the dynamics of macroeconomic variables can be represented as the sum of a component which is common to all variables in the economy and an orthogonal idiosyncratic residual. The common component is of low dimension, due to the existence of comovements between macroeconomic variables (see Sargent and Sims, 1977 as a classic reference). A version of this model has been used by Stock and Watson (1989) to construct coincident and leading indicators which are regularly applied to the US economy and also published as NBER products. In this approach, the coincident index is the common factor estimated from a vector of few coincident variables and the classification of variables as coincident is performed a priori. The common factor estimate, in a sense, is the statistical counterpart of the heuristic averaging operation of the NBER methodology.

In this paper, we propose a new procedure, which retains basic features of both the approaches above. The procedure is based on a statistical model, which is a generalization of the dynamic factor model and has a great theoretical appeal in that it provides a reconciliation of factor analysis with principal component analysis (Forni, Hallin, Lippi and Reichlin 1999, Forni and Lippi, 1999). We pass through all the steps of the traditional NBER procedure—selection of the coincident and the leading variables, identification of the turning points, averaging and cleaning from the noise—though in a different order. We first eliminate, from each time series in

[1] This research has been supported by an A.R.C. contract of the Communauté française de Belgique and by the European Commission under the Training and Mobility of Researchers Programme (Contract no. ERBFMRXCT98-0213). We would like to thank Luca Sala for excellent research assistantship.

the panel, that part of the dynamics which is poorly correlated with the rest of the economy and can hence be considered as idiosyncratic. Then, in a second step, we select coincident and leading indicators by analyzing the phase shifts between these 'cleaned' time series. Finally, we aggregate coincident and leading variables into coincident and leading indexes and establish turning points. A major novelty of our methodology is that these steps are not conceptually disjoint operations, but are all consistently nested within a unified theoretical setting.

The advantages of our method over standard techniques can be summarized as follows. First, we do not need to identify the turning points and select the coincident variables on the basis of judgmental criteria before cleaning them from noise, measurement errors and other idiosyncratic disturbances. Second, by retaining leading and lagging variables in the data set we are able to exploit additional information for the estimation of the coincident index. Third, from the analysis of a panel of time series for different countries, we can construct both an aggregate index and country-specific indexes ; the latter are estimated by taking into account cross-country as well as within-country correlations. This is particularly interesting for the construction of indexes for the EMU area where it would be inappropriate to analyze countries in isolation from each other.

The paper is organized as follows. Section 2 presents our theoretical framework. The procedure for the construction of the coincident and the leading indexes is described in Section 3. A simple example illustrating the main features of our method is presented in Section 4. The indexes for the EMU area are shown in Section 5. Summary and conclusions follow.

2. Theory

'Cleaning' the variables through dynamic principal components

As anticipated above, a crucial preliminary step of our procedure is to clean 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 panel. Intuition suggests that, to this purpose, we must, first, define few aggregates which capture most of the variance of the variables in the panel and, second, project each variable on the leads and lags of these 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 $\mathbf{x}_t = (x_{1t} \quad \cdots \quad x_{nt})'$. We wish to summarize what the processes x_{it} have in common by a small number q of 'aggregate indexes'. Precisely, we look for q processes z_{ht}, $h = 1, \ldots, q$, satisfying the following properties. To begin with, (a) z_{ht} is a linear combinations of the leads and lags of the variables in $\mathbf{x}_t$, i.e.

$$z_{ht} = \mathbf{p}_h(L)\mathbf{x}_t, \qquad h = 1, \ldots, q,$$

where L is the lag operator and $\mathbf{p}_h(L)$ is a row vector of two-sided filters. Moreover, (b) z_{ht} and z_{kt} are mutually orthogonal at any lead and lag for $h \neq k$ and the filters $\mathbf{p}_h(L)$ are normalized in such a way that $\mathbf{p}_h(L)\mathbf{p}_k(F)' = 0$ for $h \neq k$ and $\mathbf{p}_h(L)\mathbf{p}_h(F)' = 1$, where prime denotes transposition and $F = L^{-1}$. Finally, let us focus on the decomposition

$$\mathbf{x}_t = \boldsymbol{\gamma}_t^q + \boldsymbol{\zeta}_t^q = \mathbf{C}^q(L)\mathbf{z}_t^q + \boldsymbol{\zeta}_t^q = \mathbf{K}^q(L)\mathbf{x}_t + \boldsymbol{\zeta}_t^q, \tag{1}$$

where $\boldsymbol{\gamma}_t^q = (\, \gamma_{1t}^q \quad \cdots \quad \gamma_{nt}^q \,)$ is the projection of $\mathbf{x}_t$ on the present, past and future of $\mathbf{z}_t^q = (\, z_{1t} \quad \cdots \quad z_{qt} \,)'$ and $\boldsymbol{\zeta}_t^q$ is the residual vector. We require that (c) the filters $\mathbf{p}_h(L)$ and the associated processes z_{ht}, $h = 1, \ldots, q$, are such that the sum of the explained variances

$$\sum_{j=1}^{n} \text{var}(\gamma_{jt}^q) \tag{2}$$

is maximized.

Processes $z_{1t}, \ldots, z_{nt}$ satisfying requirements (a), (b) and (c) for $q = 1, \ldots, n$ do exist under quite general conditions and are called 'principal component series' or 'dynamic principal components' of $\mathbf{x}_t$.[2] What we propose here as the first step of our procedure is precisely to 'clean' the vector $\mathbf{x}_t$ by replacing it with $\boldsymbol{\gamma}_t^q$, i.e. its projection 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 $\mathbf{x}_t$, just like the static principal components are related to the eigenvalues and the eigenvectors of the variance-covariance matrix. Precisely,

[2] It is worth noting that the projection $\boldsymbol{\gamma}_t^q$ 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 $\boldsymbol{\gamma}_t^1$ is the same. The normalization $\mathbf{p}_1(L)\mathbf{p}_1(F)' = 1$, 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 $\mathbf{P}_h(L)$, $h = 1, \ldots, q$, by $a(L) = L^k$.

let $\mathbf{\Sigma}(\theta)$, $-\pi < \theta \leq \pi$, be the spectral-density matrix of $\mathbf{x}_t$: the vector $\mathbf{p}_h(e^{-i\theta})$ 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 given by $\lambda_1 + \cdots + \lambda_q$ 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 $\mathbf{C}^q(L)$ and $\mathbf{K}^q(L)$ appearing in (1). These filters are given by

$$\begin{aligned} \mathbf{C}^q(L) &= \begin{pmatrix} \mathbf{p}_1(F)' & \cdots & \mathbf{p}_q(F)' \end{pmatrix} \\ \mathbf{K}^q(L) &= \mathbf{C}^q(L)\mathbf{C}^q(F)' = \mathbf{p}_1(F)'\mathbf{p}_1(L) + \cdots + \mathbf{p}_q(F)'\mathbf{p}_q(L) \end{aligned}, \tag{4}$$

with a close analogy with the static case.

Finally, it is worth stressing that the definition of the filters $\mathbf{p}_h(L)$ involves unknown quantities like the variances in (2) and therefore must be estimated from a finite realization of $\mathbf{x}$ of length T. The estimator we use here (which is denoted by $\boldsymbol{\chi}_{nt}^T$ for reasons which will be clear below) is described in detail in Appendix A.[3] Here we give only a short hint. As a preliminary step, we estimate the spectral density matrix $\mathbf{\Sigma}(\theta)$ at different frequencies. Then, for each frequency, we compute the first q eigenvalues and eigenvectors and use (4) to compute $\mathbf{K}^q(e^{-i\theta})$. Lastly, we use the inverse Fourier transform to estimate the filter $\mathbf{K}^q(L)$ and apply it to the data. The estimates of $\mathbf{\Sigma}(\theta)$ and $\mathbf{K}^q(e^{-i\theta})$ can be exploited to estimate the spectral density matrix of the common components, which is $\mathbf{K}^q(e^{-i\theta})\mathbf{\Sigma}(\theta)\mathbf{K}^q(e^{i\theta})'$.

Principal Components and the Generalized Dynamic Factor Model

Our cleaning procedure is based on the choice of the small number q, and seems therefore open to considerable arbitrariness. However, if we assume that the x'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,

[3] Consistency of this estimator is ensured by standard results (see Brillinger 1981, ch. 9).

eliminating the idiosyncratic part and retaining the common part 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 $\mathbf{x}_t$ as formed by the first n elements of the infinite sequence x_{jt}, $j = 1, \dots \infty$. To emphasize the dependence on n, we write $\mathbf{x}_{nt}$ in place of $\mathbf{x}_t$. In our model,

$$x_{jt} = \chi_{jt} + \xi_{jt} = \mathbf{b}_j(L)\mathbf{u}_t + \xi_{jt}, \tag{5}$$

where χ_{jt} is the common component, $\mathbf{u}_t = (\, u_{1t} \quad \cdots \quad u_{qt} \,)'$ is the vector of the common shocks, i.e. a (covariance stationary) q-vector process with non-singular spectral density matrix, $\mathbf{b}_j(L)$ is a row vector of possibly two-sided, square-summable filters, and the idiosyncratic component ξ_{jt} is orthogonal to $\mathbf{u}_{t-k}$ for any k and j. Hence, with obvious notation,

$$\mathbf{x}_{nt} = \boldsymbol{\chi}_{nt} + \boldsymbol{\xi}_{nt} = \mathbf{B}_n(L)\mathbf{u}_t + \boldsymbol{\xi}_{nt}. \tag{5$'$}$$

Finally, we require the following properties. Let us denote by $\lambda^{\chi}_{hn}(\theta)$, $h = 1, \dots, n$, the h-th eigenvalue of the spectral density matrix of $\boldsymbol{\chi}_{nt}$, in descending order of magnitude. Similarly, $\lambda^{\xi}_{hn}(\theta)$ is the h-th eigenvalue of the spectral matrix of $\boldsymbol{\xi}_{nt}$. We assume that (i) the eigenvalues of $\boldsymbol{\xi}_{nt}$ are bounded as $n \to \infty$; precisely, $\lambda^{\xi}_{hn}(\theta) < \Lambda$ a.e. in $[-\pi,\ \pi]$, for any h and n; (ii) the first q eigenvalues of $\boldsymbol{\chi}_{nt}$ diverge, precisely, $\lim_{n\to\infty} \lambda^{\chi}_{hn}(\theta) = \infty$ for $h \leq q$, a.e. in $[-\pi,\ \pi]$.

Model (5) is the generalized dynamic factor model proposed by Forni, Hallin, Lippi and Reichlin (1999) and Forni and Lippi (1999). 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 j 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_{ht} 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

$$\mathbf{x}_{nt} = \boldsymbol{\gamma}_{nt} + \boldsymbol{\zeta}_{nt} = \mathbf{C}_n(L)\mathbf{z}_{nt} + \boldsymbol{\zeta}_{nt}, \tag{1$'$}$$

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 $\boldsymbol{\zeta}_{nt}$ (i.e. the last $n-q$ eigenvalues of $\mathbf{x}_{nt}$) are bounded as $n \to \infty$; precisely, $\lambda_{hn}(\theta) < \Lambda$, $h = q+1, \ldots, n$, a.e. in $[-\pi,\ \pi]$, for any n; (iv) the first q eigenvalues of $\boldsymbol{\chi}_{nt}$ (i.e. the first q eigenvalues of $\mathbf{x}_{nt}$) diverge; precisely, $\lim_{n\to\infty} \lambda_{hn}(\theta) = \infty$ for $h \le q$, a.e. in $[-\pi,\ \pi]$.

Assuming (iii) and (iv), the similarity between representations ($1'$) and ($5'$) is striking. The basic difference is that the sequence $\boldsymbol{\chi}_{nt}$, $n = 1, \ldots, \infty$ is nested, in the sense that the first $n-1$ entries of $\boldsymbol{\chi}_{nt}$ are the same as that of $\boldsymbol{\chi}_{n-1,t}$. By contrast, the sequence $\boldsymbol{\gamma}_{nt}$ in non-nested in general, so that the two decompositions do not coincide.

However, there is a deep relation between them. Forni and Lippi (1999) 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 $\boldsymbol{\gamma}_{nt}$, call it γ_{jnt}, converge to χ_{jt} in mean square as $n \to \infty$, for any j. Hence, for n large, $\boldsymbol{\gamma}_{nt}$ is a good approximation of $\boldsymbol{\chi}_{nt}$. Finally, the latter result provides the basis for the consistency theorem proven in Forni, Hallin, Lippi and Reichlin (1999): the j-th entry of the estimator used here and described in Appendix A, call it χ^T_{jnt}, converges in probability to χ_{jt} as both n and T go to infinity at some appropriate rate.

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 x_{jt} on the former space, we approximate χ_{jt}, which is the projection of x_{jt} on the latter. We shall return on this point in Section 5, where we show a simple example illustrating our methodology.

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 $\lambda_{hn} = \int_{-\pi}^{\pi} \lambda_{hn}(\theta) d\theta$ are bounded for $h > q$ and diverge for $h \le q$ as $n \to \infty$. Hence, for large n, we expect that there is a 'jump' between λ_{qn} and $\lambda_{q+1,n}$. This suggests to add principal components until the increase in the explained variance is larger than some prespecified value. Precisely, denoting by λ^T_{hn}, where T is the number of time observations, the estimate of λ_{hn}, and given a number $\alpha \in (0,1)$ the criterion consists in selecting $q = q^*$ such that

$$\lambda^T_{q^* n} / \sum_{h=1}^{n} \lambda^T_{hn} > \alpha \quad \text{and} \quad \lambda^T_{q^*+1,n} / \sum_{h=1}^{n} \lambda^T_{hn} < \alpha. \tag{6}$$

Model (5) allows for possibly more than one shock with heterogenous impulse response functions across cross-sectional units. The model in Stock and Watson (1989) is obtained as the particular case in which, besides mutually orthogonal idiosyncratic components, there is only one common shock (so that $\mathbf{B}_n(L)$ is a vector), and all the entries of $\mathbf{B}_n(L)$ are proportional (i.e. there is a scalar filter $b(L)$ such that $\mathbf{B}_n(L) = b(L)\mathbf{B}_n$, $\mathbf{B}_n$ being a vector of constants). In this case we can unambiguously identify a unique 'common cycle' in a very strong sense, since the only common factor, defined as $b(L)u_t$, is loaded only contemporaneously by all of the variables in $\mathbf{x}_{nt}$. This is a rather demanding restriction. None of the x_{jt}'s can be leading or lagging, and, in addition, there must be only one source of common variation in the data set. Matching these requirements imposes a very accurate preliminary selection of the coincident variables, and, even so, the restrictions could be rejected by the data, leaving us without any theoretical foundation for the coincident index. Here we give up the idea that a common cycle in the strong sense above does exist. On the other hand, we are payed back with an enormous gain in flexibility, which is the source of the advantages of our methodology, as we shall see in detail in the following sections.

3. The procedure

Having clarified the basic theoretical background, we can now go on and present the whole procedure for the construction of the coincident and the leading indexes and the identification of turning points. To better fix ideas, we shall make reference to the panel of countries belonging to the European Monetary Union (EMU), but the procedure, possibly with minor modifications, has general validity. The procedure consists of five steps.

STEP 1: CHOICE OF THE VARIABLES TO INCLUDE IN THE PANEL

The first step is to decide what are the variables to include in the data set and to clean them by computing the common components. As we have seen, from a theoretical point of view, all available variables should in principle be included. In practice, however, it is not recommended to include variables which have small commonality and large idiosyncratic components, since the latter could survive aggregation and be wrongly interpreted as additional common factors. To select our data set we proceed as follows. We identify two sets of variables: a 'core' formed by variables which surely must be in, and a set of candidate variables. Then we transform all the variables so as to get stationarity and normalize them by subtracting the mean from each variable and dividing them by their standard deviation. We obtain stationarity by differencing (or differencing the logs). An alternative procedure would have been to apply a band-pass filter as, for example, in Stock and Watson (1998). The advantage of the latter strategy is that it eliminates high frequency fluctuations.

Its drawback, however, is that it implies a bilateral filter which poses problems of estimation at the end of the sample in addition to those that we will discuss below.

Using the core, we fix q^* according to criterion (6) and compute the corresponding 'degree of commonality,' as measured by the (estimated) variance ratio (3). Once this ratio, say μ, is fixed, we evaluate each candidate variable in turn.[4] For the evaluation, we consider the enlarged system formed by the core and the candidate variable, compute the quantity (3) with $q = q^*$ and decide that the candidate has passed the exam as soon as (3) is larger than μ. Successful candidates are not added to the core until all the candidates have been evaluated. The final data set is defined as including the core and the non-core successful variables.

At this stage, we estimate both the vector of the common components $\boldsymbol{\chi}_{nt}$ and its spectral density matrix $\boldsymbol{\Sigma}^{\chi}(\theta)$, with $q = q^*$, as explained in the previous Section and in Appendix A.

STEP 2: PRO-CYCLICAL AND ANTI-CYCLICAL VARIABLES

As the second step, we classify the common components χ_{jt} as being 'in phase' or in 'phase opposition' with respect to the common component of the GDP, which is taken as the reference point to establish what is 'coincident' and what is not. A variable in phase opposition is a counter-cyclical variable; an obvious example is the unemployment rate.

Precisely, we proceed as follows. Using the estimate of $\boldsymbol{\Sigma}^{\chi}(\theta)$, we compute the cross-spectral density of each common component with respect to the common component of the growth rate of the European GDP.[5] Then we compute the argument of these densities, which is the phase angle delay with respect to the European GDP, at frequency zero.[6] Let the phase angle shift for χ_{jt} be $\phi_j(\theta)$, $-\pi < \theta \leq \pi$. At frequency zero, the phase may be either 0 or π depending on whether long-run correlation is positive or negative. We interpret $\phi_j(0) = \pi$ as indicating that χ_{jt} is in 'phase opposition' and define the new series of interest as[7]

$$\omega_{jt} = \begin{cases} \chi_{jt} & \text{if } \phi_j(0) = 0, \\ -\chi_{jt} & \text{if } \phi_j(0) = \pi. \end{cases}$$

[4] Since we have many countries, when evaluating a variable we add to the core several time series at a time.

[5] The European growth rate is defined as the weighted average of the differences of the logs of the GDP's of the European countries, with weights proportional to the average levels across time. For a few countries we do not have the GDP; in these cases, we used energy consumption.

[6] We recall that the cross spectral density between two variables h and j can be expressed, in its 'polar form', as $S_{hj}(\theta) = A_{hj}(\theta)e^{-i\phi_{hj}(\theta)}$ where $A_{hj}(\theta)$ is the 'amplitude' and $\phi_{hj}(\theta)$ is the 'phase'. The phase $\phi_{hj}(\theta)$ measures the angular shift between the cosine waves of h and j at frequency θ, while $\phi_{hj}(\theta)/\theta$ measures the time shift.

[7] This procedure is suggested by Granger and Hatanaka 1964, ch. 12.

STEP 3: CLASSIFICATION OF VARIABLES AS COINCIDENT, LEADING AND LAGGING

The third step consists in classifying the resulting time series as being leading, coincident or lagging according to their phase delay with respect to the GDP.

We compute the phase angle shift of ω_{jt}, $j = 1, \dots, n$, with respect to the GDP, at a typical business cycle frequency, say $\theta^* > 0$, and, denoting such phase angle with $\psi_j(\theta^*)$, we classify ω_{jt} as coincident if $|\psi_j(\theta^*)|$ is smaller than a prespecified value τ, leading if $\psi_j(\theta^*) < -\tau$ and lagging if $\psi_j(\theta^*) > \tau$.[8]

STEP 4: THE COINCIDENT AND LEADING INDEXES

Now we are ready to compute the indexes. The coincident and the leading indexes for Europe are constructed as averages of the coincident and the leading common components, taken with the proper signs. More precisely, we assign to variable j the weight W_j, given by the average across time of the GDP level of the corresponding country. Then, denoting with $\mathcal{C}$ the set of the j's such that ω_{jt} is coincident and with $\mathcal{L}$ the set of the j's such that ω_{jt} is leading, the first differences of the coincident and the leading indexes, ΔC_t and ΔL_t, are defined as

$$
\begin{aligned}
\Delta C_t =& \frac{\sum_{j \in \mathcal{C}} \omega_{jt} W_j}{\sum_{j \in \mathcal{C}} W_j} \\
\Delta L_t =& \frac{\sum_{j \in \mathcal{L}} \omega_{jt} W_j}{\sum_{j \in \mathcal{L}} W_j}.
\end{aligned}
\tag{7}
$$

The indexes in levels, C_t and L_t, are defined as the cumulated sums, centered and divided by their standard deviations. A lagging index can be defined in a similar way.[9]

Notice that the estimated common components χ_{jnt}^T are obtained by applying to the data set a two-sided filter ($\mathbf{K}_n^T(L)$) with length $2M + 1$ (see Appendix A). Hence at the beginning and at the end of sample, i.e. at $t = 1, \dots, M$ and $t = T - M + 1, \dots, T$ the estimates are bad. For this reason we correct the estimates at these points by replacing ω_{jt} in equation (7) either with x_{jt} (if $\omega_{jt} = \chi_{jt}$), or with $-x_{jt}$ (if $\omega_{jt} = -\chi_{jt}$). In order to evaluate the adequacy of this correction, consider that if we have many coincident and leading variables, the idiosyncratic components are roughly eliminated by the averaging in (7), even when ω_{jt} is replaced by x_{jt}. In

[8] Of course, this is equivalent to computing the 'time delay' $\psi_j(\theta^*)/\theta^*$ and compare it with τ/θ^*. Notably, if θ^* is sufficiently close to 0, the estimate of the time delay $\psi_j(\theta^*)/\theta^*$ can be regarded as an estimate of the derivative of the phase angle at $\theta = 0$. This is interesting in that such derivative is equal to the 'mean lag', which is a well-known time-domain statistic measuring the 'delay' of a time series.

[9] As observed in Sargent (1987, ch. XI), a phase lead does not necessarily imply Granger causation. Hence, the ability of the leading index in predicting the coincident index has to be verified in practice.

our European data set, within sample, i.e. for $t = M+1, \ldots, T-M$, the mean-square error is 0.98 for the coincident index and 0.96 for the leading index.

STEP 5: TURNING POINTS

Having obtained the (corrected) coincident index C_t, the turning points are simply defined as the dates t^* in which C_t reaches local maxima and minima. To avoid the possibility that two maxima (or minima) are too close to each other, we can impose the further condition that, given a prespecified m, $t^* \geq t$ for any $t \in [t^* - m, t^* + m]$.

STEP 6: LOCAL INDEXES

Coincident and leading indexes for each European country can be constructed as follows. For country s we select the related series, i.e. the χ_{jt}'s such that $s_j = s$, and focus on the cross-spectra of each one of them with respect to the GDP of that country. Then we follow Step 2 in order to define the correct signs and the phase delays of each series. The coincident (leading) index is then defined as the simple average of the coincident (leading) series.

4. A stylized example

Dynamic principal components average time series cross-sectionally and over time. This double operation allows to weight variables according to their leading-lagging relations. This section develops a stylized example which illustrates the point. To make things simple we shall assume an infinite number of observations over time, so that the error stemming from a finite T will be ignored.

Suppose that $q = 1$ and the filters $\mathbf{b}_j(L)$ appearing in equation (5) are of the form L^{s_j}, with s_j equal to zero, one or two. Thus we have a single common factor u_t; some of the variables load it with lag one, the coincident variables, some with lag zero, the leading variables, some with lag two, the lagging variables. Equation (5′) becomes

$$\mathbf{x}_{nt} = \begin{pmatrix} L^{s_1} \\ L^{s_2} \\ \vdots \\ L^{s_n} \end{pmatrix} u_t + \boldsymbol{\xi}_{nt}.$$

Moreover, assume that the idiosyncratic components ξ_{jt}, $j = 1, \ldots, \infty$, are mutually orthogonal white noises, with the same variance σ^2, so that the spectral density of $\mathbf{x}_{nt}$ is

$$\frac{1}{2\pi} \begin{pmatrix} e^{-is_1\theta} \\ e^{-is_2\theta} \\ \vdots \\ e^{-is_n\theta} \end{pmatrix} \left(e^{is_1\theta} \quad e^{is_2\theta} \quad \cdots \quad e^{is_n\theta} \right) + \frac{\sigma^2}{2\pi} \mathbf{I}_n.$$

In this case, it can be easily verified that the larger eigenvalue is

$$\lambda_{1n}(\theta) = n + \sigma^2 \tag{8}$$

and a valid corresponding row eigenvector is

$$\mathbf{p}_{1n}(e^{-i\theta}) = \frac{1}{\sqrt{n}} \begin{pmatrix} e^{is_1\theta} & e^{is_2\theta} & \cdots & e^{is_n\theta} \end{pmatrix}. \tag{9}$$

The related filter is[10]

$$\mathbf{p}_{1n}(L) = \frac{1}{\sqrt{n}} \begin{pmatrix} F^{s_1} & F^{s_2} & \cdots & F^{s_n} \end{pmatrix},$$

while the first principal component series is

$$\begin{aligned} z_{1t} &= \frac{1}{\sqrt{n}} \begin{pmatrix} F^{s_1} & F^{s_2} & \cdots & F^{s_n} \end{pmatrix} \begin{pmatrix} L^{s_1} \\ L^{s_2} \\ \vdots \\ L^{s_n} \end{pmatrix} u_t \\ &+ \frac{1}{\sqrt{n}} \begin{pmatrix} F^{s_1} & F^{s_2} & \cdots & F^{s_n} \end{pmatrix} \boldsymbol{\xi}_{nt} \\ &= \sqrt{n} u_t + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \xi_{jt+s_j}. \end{aligned}$$

Two observations are in order. First, the idiosyncratic part of the principal component vanishes with respect to the common part as n becomes larger and larger, so that the principal component itself becomes increasingly 'collinear' with the common factor u_t. In other words, if n is sufficiently large, the first principal component captures the information space spanned by the common shock.

Second, the filter $\mathbf{p}_{1n}(L)$ shifts the common components by multiplying each of the L^{s_j} precisely by F^{s_j}, so that time delays and time leads are eliminated and we end up by summing n times the same common shock u_t. This sort of automatic re-alignment is the reason why we do not need to discriminate *a priori* between coincident and non-coincident series and discard the non-coincident ones in the construction of the coincident index. Far from disturbing, leading and lagging series

[10] Note that, in this example, the idiosyncratic components plays no role in the determination of $\mathbf{p}_{1n}(L)$, which would have been identical with zero idiosyncratic terms. This is due to the particular form that we have assumed here for the cross-covariance structure of the idiosyncratic components. However, as shown in Forni and Lippi (1999), the same property holds approximately for large n under Assumption (i) of Section 2, i.e. the boundedness of the first eigenvalue of the spectral density matrix of $\boldsymbol{\xi}_{nt}$.

contribute to a better cleaning of the variables from the idiosyncratic noise. By averaging both across time and across sections, we are able to exploit correctly the information conveyed by the leading and lagging variables, which otherwise would be lost.

This point can be better understood by developing the example further. Applying (4) we see that the filter $\mathbf{K}_n(L)$ is equal to $\mathbf{p}_{1n}(F)'\mathbf{p}_{1n}(L)$, so that the 'estimated' common components, which coincide here with the γ_{jt}'s appearing in equation (1) (because of the simplification $T = \infty$) are

$$\begin{aligned}\gamma_{jt} &= \frac{1}{\sqrt{n}} L^{s_j} \left[\sqrt{n} u_t + \frac{1}{\sqrt{n}} \sum_{h=1}^{n} \xi_{ht+s_h} \right] = L^{s_j} u_t + \frac{1}{n} \sum_{h=1}^{n} \xi_{h(t+s_h-s_j)} \\ &= \chi_{jt} + \frac{1}{n} \sum_{h=1}^{n} \xi_{h(t+s_h-s_j)}.\end{aligned}$$

Note that, when applying the filter $\mathbf{p}_{1n}(F)'$ (i.e. projecting on the leads and lags of z_{1t}), the correct lags of the common components are restored and the leading or lagging nature of each variable emerges again.

To see how the 'phase shift' reveals this information assume $s_1 = 1$, i.e. the first variable is coincident, and take it as the reference point. The spectral density matrix of the common components will be estimated as

$$\mathbf{K}_n(e^{-i\theta})\mathbf{\Sigma}_n(\theta)\mathbf{K}_n(e^{i\theta}) = \mathbf{p}_{1n}(e^{i\theta})'\mathbf{p}_{1n}(e^{-i\theta})\frac{\lambda_{1n}(\theta)}{2\pi}.$$

¿From (8) and (9) it is seen that the estimated cross-spectrum of χ_{jt} and χ_{1t} is $e^{-i\theta(s_j-1)}(1/2\pi + \sigma^2/2n\pi)$. Hence the true cross-spectral density $e^{-i\theta(s_j-1)}/2\pi$ will be obtained for $n \to \infty$ and the angle phase delay is perfectly estimated as

$$\psi_j(\theta) = \theta(s_j - 1).$$

The implied time delay (see note 8, p.9) is $\psi_j(\theta)/\theta = s_j - 1$, i.e. -1 for the leading, 0 for the coincident, and 1 for the lagging common components.

The coincident index ΔC_t is obtained by averaging the coincident common components. Assuming without loss of generality that the coincident variables are the first m ones, we have $s_1 = \cdots = s_m = 1$ so that

$$\Delta C_t = \sum_{h=1}^{m} \gamma_{ht}/m = u_{t-1} + \frac{1}{n} \sum_{j=1}^{n} \xi_{j(t+s_j-1)}.$$

The true coincident index u_{t-1} is approximated with error variance σ^2/n.

Let us now compare our index with the index which we would obtain by simply averaging across the coincident variables, which for simplicity we now assume known in advance. This alternative index, say ΔC_t^{NBER}, is a stylized exampled of the traditional NBER procedure.

We have

$$\Delta C_t^{NBER} = \frac{1}{m}\sum_{j=1}^{m} x_{jt} = u_{t-1} + \frac{1}{m}\sum_{j=1}^{m} \xi_{jt},$$

with error variance σ^2/m. Hence, as anticipated above, as long as leading and lagging variables are available ($n > m$), their inclusion will improve the estimates.

Notice that in this stylized example, the first principal component, suitably shifted to be in phase with GDP, is collinear with the coincident index. However, using the first principal component as the coincident index, as suggested by Bowden and Martin (1993), is not appropriate when we have more than one common shock.

Finally, let us observe that if, as in the empirical application of the next Section, we have more than one country and the GDP of one of them is leading, the coincident index for this country will differ from the general index. More generally, in a model with more than one common shock and heterogenous factor loadings, there is no reason to expect equal coincident indexes for all countries, even in the case in which all GDPs are coincident. Therefore, our method identifies national as well as international indexes. In models with one common shock and contemporaneous loadings this flexibility would be lost.

5. Coincident and leading indexes for the EMU

We will now estimate coincident and leading indexes for all countries of the EMU aggregate and for each different member country following the procedure outlined in Section 3. Data and data sources are described in Appendix B.

The variables selected for the core are all from the real sector. Additional labor market variables, prices, monetary and financial indicators and orders were left out and analyzed one-by-one as described in Step 1 of Section 3. We fixed $\alpha = 0.05$ and found $q^* = 3$, i.e. three common factors. For the core, the degree of commonality, as measured by (3), is 0.50. Table 1 reports results from the selection procedure. It emerges clearly that all financial and monetary variables do not pass the selection criterion and neither do the price indexes or the share indexes. This indicates that the financial sector and prices have a limited commonality with the real sector of the economy. On the other hand, and not surprisingly, male unemployment rate passes the test while the female rate does not. The result for orders is not surprising either, since this variable is traditionally used as a leading indicator by practitioners.

In Table 2 we show variance ratios between each variable common component and its total. These estimates provide an information on the 'degree of commonality' of each variable in the panel. Results vary slightly across countries, but, in general, GDP, labor market variables and orders show the highest degree of commonality.

Table 3 reports the analysis of the time phase lead of each variable with respect to the common European GDP (time leads are expressed in quarters) at frequency $\theta^* = \pi/16$, corresponding to a cycle of period of eight years. On the basis of these results we can then establish which variables are coincident, leading or lagging. Leading variables are defined as those with time phase lead larger than one month (0.33 quarters), corresponding to an angle phase lead $\tau = \pi/48$; lagging variables are defined as those which lag by more than one month and the residual variables are defined as coincident. Starred variables are those found to be in phase with respect to the common component (procyclical). Not surprisingly, unemployment is anti-cyclical and labor market variables are generally lagging. Leading variables are orders for all countries and capacity utilization and energy consumption for some countries. We can also observe that, contrary to what one may have expected, neither German GDP nor German investment are leading. In fact, it is the Finnish GDP and investment which have a leading role. This is explained by the specific characteristics of the Finish recession of the early nineties. The exceptionality of the Finnish cycle will again emerge from country-specific results below.

On the basis of these results we are finally able to aggregate coincident, leading and lagging variables into their respective indexes. Figure 1 shows the resulting EMU coincident and leading indexes in levels (without the drift). We have two main turning points in the first quarter of 1990 (beginning of a recession) and the first quarter of 1994 (beginning of an expansion), followed by a short cycle with small amplitude at the end of the sample, with turning points 1995:1 and 1996:2, which could be interpreted as a minor episode within a basically expansive period.

Finally, let us report results from the country analysis. Figure 2 reports the national coincident indexes against the EMU aggregate coincident index.

6. Summary and conclusions

This paper develops a unified methodology for the construction of coincident and leading indicators and the identification of turning points of the business cycle. The method is based on statistical theory and reconciles dynamic principal component and dynamic factor analysis. We apply the methodology to a data set of many macroeconomic variables for ten EMU countries. Results indicate that the real cycle is not strongly correlated with the monetary-financial cycle and that, contrary to common wisdom, Germany does not have a 'leading' role within the EMU.

References

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

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

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

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

Chamberlain, G. and Rothschild, M. (1983), "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets", *Econometrica* 51, 1305-1324.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (1999) "The Generalized Factor Model: Identification and Estimation", *The Review of Economic and Statistics*, forthcoming.

Forni, M. and Lippi, M., (1999) "The Generalized Factor Model: Representation Theory", Working Paper no. 132, Université Libre de Bruxelles.

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.

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. (1998) "Business Cycles Fluctuations in US Macroeconomic Time Series," NBER Working Paper no. 6528.

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

Appendix A: the estimator

In this Appendix we show how $\boldsymbol{\chi}_{nt}$ and its spectral density matrix can be estimated. We adopt the notation of (5′) and (1′), i.e. we drop the superscript q and explicit the dependence on n. As explained in the main text, χ_{jnt}, the j-th entry of $\boldsymbol{\chi}_{nt}$, can be regarded as an estimator of both γ_{jnt}, the j-th entry of $\boldsymbol{\gamma}_{nt}$, and χ_{jt}. Standard results ensure consistency of the estimator, as an estimator of γ_{jnt}, as T goes to infinity. Consistency with respect to χ_{jt} as both n and T go to infinity is shown in Forni, Hallin, Lippi, Reichlin (1999).

The estimation procedure is in three steps. First, we estimate the spectral density matrix $\boldsymbol{\Sigma}(\theta)$ of $\mathbf{x}_{nt}$ at a number of frequencies, using a Bartlett lag-window estimator of size $M = M(T)$. Precisely, we compute the sample covariance matrix $\boldsymbol{\Gamma}_k^T$ of $\mathbf{x}_{nt}$ and $\mathbf{x}_{nt-k}$ for $k = 0, \dots, M$. Then we compute the $2M+1$ points discrete Fourier transform of the truncated two-sided sequence $\boldsymbol{\Gamma}_{-M}^T, \dots, \boldsymbol{\Gamma}_0^T, \dots, \boldsymbol{\Gamma}_M^T$, where $\boldsymbol{\Gamma}_{-k} = \boldsymbol{\Gamma}_k'$, i.e. we compute

$$\boldsymbol{\Sigma}_n^T(\theta_s) = \sum_{k=-M}^{M} \boldsymbol{\Gamma}_k^T \omega_k e^{-ik\theta_s},$$

where

$$\theta_s = 2\pi s/(2M+1), \quad s = 0, \dots, 2M$$

and $\omega_k = 1 - \frac{|k|}{(M+1)}$ are the weights corresponding to the Bartlett lag window of size M. Consistent estimation of $\boldsymbol{\Sigma}(\theta_s)$ is ensured, provided that $M(T) \to \infty$ and $M(T)/T \to 0$ as $T \to \infty$. The rule $M = \text{round}(\sqrt{T}/4)$ seems to perform well for a number of low order MA and AR models under simulation (see Forni, Hallin, Lippi, Reichlin 1999). The construction of a data-dependent rule, which would be preferable in principle, is still an open question. Here, with $T = 50$, we used both $M = 2$ and $M = 3$ and decided in favor of $M = 3$ because of the existence of consistent phase shifts between the variables.

Second, we compute the first q eigenvectors $\boldsymbol{\pi}_{hn}^T(\theta_s)$, $h = 1, \dots, q$, of $\boldsymbol{\Sigma}_n^T(\theta_s)$, for $s = 0, \dots, 2M$. Note that, for $M = 0$, $\boldsymbol{\pi}_{hn}^T(\theta_0)$ is simply the h-th eigenvector of the (estimated) variance-covariance matrix of $\mathbf{x}_{nt}$: the dynamic principal components then reduce to the static principal components. ¿From the eigenvectors we compute

$$\boldsymbol{\Phi}_n^T(\theta_s) = \tilde{\boldsymbol{\pi}}_{1n}^T(\theta_s)\boldsymbol{\pi}_{1n}^T(\theta_s) + \dots + \tilde{\boldsymbol{\pi}}_{qn}^T(\theta_s)\boldsymbol{\pi}_{qn}^T(\theta_s),$$

where the tilde denotes conjugation and transposition.

Finally, we compute the estimator of the $n \times n$ matrix of filters $\mathbf{K}_n(L)$ by applying the inverse discrete Fourier transform of

$$(\,\boldsymbol{\Phi}_n^T(\theta_0), \quad \cdots, \quad \boldsymbol{\Phi}_n^T(\theta_{2M})\,).$$

Precisely, we compute

$$\mathbf{K}_{kn}^T = \frac{1}{2M+1} \sum_{s=0}^{2M} \mathbf{\Phi}_n^T(\theta_s) e^{ik\theta_s}$$

for $k = -M, \dots, M$. The estimator of the filter is given by

$$\mathbf{K}_n^T(L) = \sum_{k=-M}^{M} \mathbf{K}_{kn}^T L^k.$$

The estimator of the common components is

$$\boldsymbol{\chi}_{nt}^T = \mathbf{K}_n^T(L)\mathbf{x}_t.$$

The spectral density matrix of the common components can be estimated as

$$\mathbf{\Sigma}_n^{\chi T}(\theta_s) = \mathbf{\Phi}_n^T(\theta_s)\mathbf{\Sigma}_n^T(\theta_s)\tilde{\mathbf{\Phi}}_n^T(\theta_s).$$

APPENDIX B: Definition of variables, sources and data treatment

Core variables

Code	Variable	Source	Details	Data Treatment
GDP	Gross domestic product	Eurostat	Mio Ecu 1990 - National accounts (SEC79)	Differences in logs
Inv.	Gross fixed capital formation	Eurostat	Mio ecus,1990 /National accounts (SEC79)	Differences in logs
Cons.	Private national consumption	Eurostat	Mio ecus,1990 /National accounts (SEC79)	Differences in logs
Cap. Util.	BSS capacity utilization rate	Eurostat	Seasonally Adjusted /Business tendency surveys/ Rate of capacity utilisation	Differences - For Italy, regression on 3 seasonal dummies and a trend (as the seas. adjusted series was not available)
Unemp.	Unemployment	Eurostat	Total, seasonally adjusted	Differences
En. Cons.	Energy	Eurostat	Energy - Gross inland consumption - all products - 1000 TOE	Differences - Deseasonalized by a regression on 3 seasonal dummies and a constant. For Germany, regression on 3 seasonal dummies, a trend and a constant with a change in regime in 90:4
IP	Industrial production	Eurostat	Total industry (excluding construction) - 1995=100	Differences in logs

Other variables

Code	Variable	Source	Details	Data Treatment
Emp	Employment	Eurostat	Total industry (excluding construction) -1995=100	Differences
Mrate	Day-to-day money rate	Eurostat	Day-to-day money rate (mean) %	Differences
3Mth rate	3-month money market rate	Eurostat	3-month money market rate (mean) %	Differences
M1	Money supply: M1	Eurostat	Money supply: M1 (end of period) T/T-12	Differences in logs
M2	Money supply: M2	Eurostat	Money supply: M2 (end of period) T/T-12 %	Differences log
M3	Money supply: M3	Eurostat	Money supply: M3 (end of period) T/T-12 %	Differences in logs
Unemp men	Unemployment rate men	Eurostat	Unemployment rates men seasonally adjusted	Differences
Unemp women	Unemployment rate women	Eurostat	Unemployment rates women seasonally adjusted	Differences
Unemp young	Unemployment rate of persons under 25 years	Eurostat	Unemployment rates of persons under 25 years, total seas. adj.	Differences
CPI	CPI	OECD	All items - Index - 1995=100	Differences in logs
PPI	PPI	OECD	Total Index publication base - 1995Y	Differences in logs
Share Price	Share Prices	OECD	SHARE PRICES ALL SHARES - Index	Differences in logs
Real Int.	3Mth rate - Δ(in logsCPI)	Authors' computations	3Mth rate - Δ(logCPI)	Differences
Orders	Orders for manufacturing industries	EC-DGII	Orders for manufacturing industries	Differences

Table 1: Data set

	Austria	Belgium	Finland	France	Germany	Ireland	Italy	Netherlands	Portugal	Spain
GDP	core	core	core	core	core	-	core	core	-	core
Inv.	core	core	core	core	core	-	core	core	-	core
Cons.	core	core	core	core	core	-	core	core	-	core
Cap. Util.	-	core	-	core	core	-	core	core	core	core
Unemp.	-	core	-	core	-	core	core	core	core	core
En. Cons.	-	core	-	core	core	core	core	core	core	core
IP	core	core	core	core	core	core	core	core	core	core
Emp	-	nc	-	nc	nc	-	nc	-	-	nc
Mrate	x	-	-	x	x	-	-	x	-	x
3Mth rate	x	x	x	x	x	x	x	x	x	x
M1	x	x	x	x	-	x	x	-	x	x
M2	-	-	x	x	-	-	-	-	x	x
M3	x	x	x	x	-	-	-	-	x	x
Unemp men	-	nc	-	nc	-	nc	nc	nc	nc	nc
Unemp wom	-	x	-	x	-	x	x	x	x	x
Unemp young	-	nc	-	nc	-	nc	nc	nc	nc	nc
CPI	x	x	x	x	x	x	x	x	x	x
PPI	x	x	x	x	x	x	x	x	-	x
Share Price	x	x	-	x	x	x	x	x	-	x
Real Int.	x	x	x	x	x	x	x	x	x	-
Orders	nc	nc	nc	nc	nc	nc	nc	nc	-	-

core: variables belonging to the core
nc: variables not belonging to the core, used in the final estimation
x: variables not belonging to the core, not used in the final estimation

Table 2: Variance ratios of the common components

	Austria	Belgium	Finland	France	Germany	Ireland	Italy	Netherlands	Portugal	Spain
GDP	0.4422	0.4981	0.4609	0.7704	0.5596	-	0.533	0.3363	-	0.8001
Inv.	0.3793	0.3707	0.4154	0.6663	0.5925	-	0.5546	0.4669	-	0.7773
Cons.	0.1925	0.4314	0.3957	0.3569	0.5012	-	0.597	0.3793	-	0.7001
Cap. Util.	-	0.3281	-	0.5591	0.5482	-	0.3978	0.5026	0.2548	0.4489
Unemp.	-	0.6374	-	0.7032	-	0.7384	0.4189	0.5585	0.5466	0.8213
En. Cons.	-	0.4672	-	0.5999	0.4065	0.4829	0.4414	0.4609	0.1765	0.2151
IP	0.3235	0.223	0.5427	0.5768	0.6552	0.3706	0.2997	0.374	0.3576	0.5464
Emp	-	0.3461	-	0.7915	0.7524	-	0.5315	-	-	0.6316
Unemp men	-	0.6198	-	0.7766	-	0.7199	0.4552	0.583	0.4493	0.7781
Unemp young	-	0.7267	-	0.713	-	0.6246	0.4117	0.389	0.4749	0.8069
Orders	0.6788	0.675	0.532	0.6615	0.7484	0.2905	0.689	0.5931	-	-

Table 3: Phase lead/lag with respect to the common European GDP

	Austria	Belgium	Finland	France	Germany	Ireland	Italy	Netherlands	Portugal	Spain
GDP	-0.0319*	0.0230*	0.4545*	0.0147*	-0.1696*	-	0.0397*	-0.0329*	-	0.0751*
Inv.	0.5427*	-0.1508*	0.0421*	-0.0985*	-0.2316*	-	-0.2717*	0.1786*	-	0.0548*
Cons.	-0.6206*	-0.0434*	0.3407*	-0.1680*	-1.4209*	-	-0.0447*	-0.5661*	-	-0.0947*
Cap. Util.	-	0.6893*	-	-0.1106*	0.3671*	-	0.3657*	0.3702*	0.1902*	0.5323*
Unemp.	-	-0.3591	-	-0.4625	-	-0.0353	-1.0839	-0.416	-0.1499	-0.1988
En. Cons.	-	0.2983*	-	0.5071*	-1.048	1.3540*	0.1548*	1.1242	-0.7777*	0.4682*
IP	-0.1231*	-0.0456*	3.0941*	0.1269*	0.0552*	-0.4841*	0.1749*	-0.2501*	-0.2936*	0.5754*
Emp	-	-0.7197*	-	-0.4074*	-0.4126*	-	-0.1412*	-	-	-0.0443*
Unemp men	-	-0.3804	-	-0.3221	-	-0.0816	-1.0063	-0.2013	-0.1062	-0.0765
Unemp young	-	-0.296	-	-0.3307	-	0.2773	-0.6475	-0.0664	0.0126	-0.1222
Orders	0.3798*	0.4756*	4.1444*	0.4641*	-0.0066*	0.4211*	0.7120*	0.4662*	-	-

Starred variables are procyclical (in phase with respect to the Euro common GDP). Those without a star are countercyclical (in opposition of phase).

Figure 1: European coincident and leading indexes

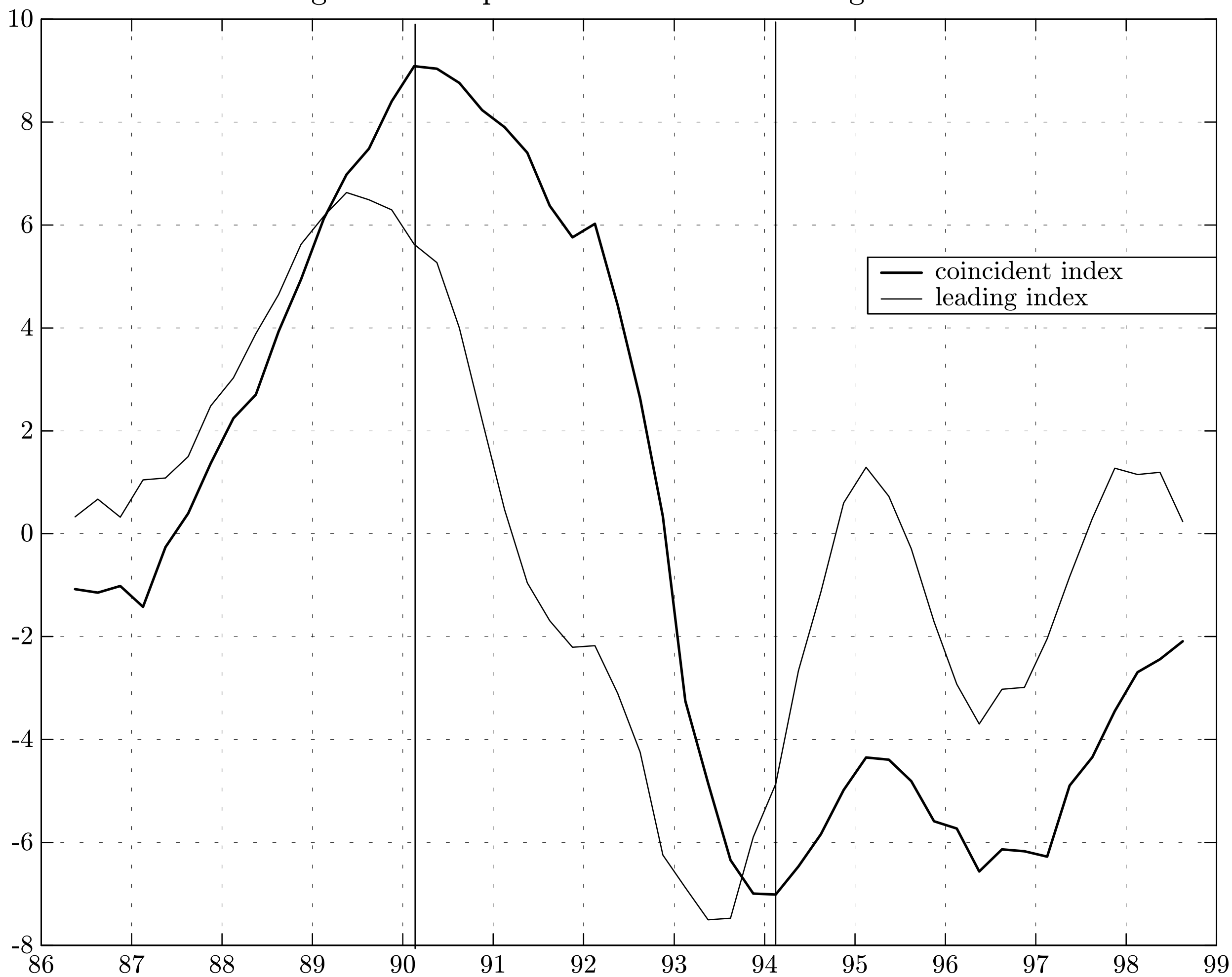

Figure 2: National coincident indexes