DISCUSSION PAPER SERIES

No. 2338

THE GENERALIZED DYNAMIC FACTOR MODEL: IDENTIFICATION AND ESTIMATION

Mario Forni, Marc Hallin, Marco Lippi and Lucrezia Reichlin

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP2338.asp**

ISSN 0265-8003

THE GENERALIZED DYNAMIC FACTOR MODEL: IDENTIFICATION AND ESTIMATION

Mario Forni, Università di Modena
Marc Hallin, Université Libre de Bruxelles
Marco Lippi, Università di Roma La Sapienza
Lucrezia Reichlin, ECARES, Université Libre de Bruxelles

Discussion Paper No. 2338
December 1999

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

CEPR Discussion Paper No. 2338

December 1999

ABSTRACT

The Generalized Dynamic Factor Model: Identification and Estimation*

This paper analyzes identification conditions, and proposes an estimator, for a dynamic factor model where the idiosyncratic components are allowed to be mutually non-orthogonal. This model, which we call the generalized dynamic factor model, is novel to the literature, and generalizes the static approximate factor model of Chamberlain and Rothschild (1983), as well as the exact factor model à la Sargent and Sims (1977). We propose an estimator of the common components and prove convergence as both time and cross-sectional size go to infinity at appropriate rates. Simulations yield encouraging results in small samples. We use our model to construct an index of the state of the economy for the European currency area. Such an index is defined as the common component of real GDP within a model including several macroeconomic variables for each European country.

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

Mario Forni
Dipartimento Di Economia Politica
University of Modena and Reggio Emilia
Via Berengario 51
41100 Modena
ITALY

Tel: 2056852
Fax: 2056947
Email: forni@unimo.it

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

Marc Hallin
ECARES CP 114
50 Avenue Roosevelt
1050 Bruxelles
BELGIUM

Tel: 650 5886
Fax: 650 5899
Email: mhallin@ulb.ac.be

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

Marco Lippi
Dipartimento di Scienze Economiche
Università di Roma "La Sapienza"
Via Cesalpino 12
00161 Roma
ITALY

Tel: 00 39 06 4428 4202
Fax: 00 39 06 4404 572
Email: marco.lippi@uniroma1.it

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

Lucrezia Reichlin
ECARES
Université Libre de Bruxelles
50, avenue Roosevelt CP 114
B-1050 Brussels
BELGIUM

Tel: 650 4221
Fax: 650 4475
Email: lreichli@ulb.ac.be

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

* This research has been supported by an A.R.C. contract of the Communauté francaise de Belgique. We would like to thank Christine De Mol for generously giving us her time, Serena Ng for pointing out programming problems, Jim Stock for helpful advice and Jorge Rodrigues for very valuable research assistance. 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 30 November 1999

1. Introduction[1]

Economic activity in market economies is characterized by phases of upturns followed by phases of depression which is manifested by the cyclical behavior and comovements of many macroeconomic variables. If comovements are strong, it makes sense to represent the state of the economy by an index —the 'reference cycle'— describing the common behavior of such variables. This idea, first suggested by Burns and Mitchell (1946), is behind the NBER coincident indicator. The formal model which best captures it is the index model, or dynamic factor model, proposed by Sargent and Sims (1977) and Geweke (1977). A vector of n time series is represented as the sum of two unobservable orthogonal components, a common component, driven by few (fewer than n) common factors, and an idiosyncratic component, driven by n idiosyncratic factors. If we have only one common factor, affecting only contemporaneously (i.e. without lags) all of the time series, such factor can be interpreted as the reference cycle (Stock and Watson, 1989).

Factor models can also be used to address different economic issues. For instance, a factor structure is often assumed in both financial and macroeconomic literature to estimate insurable risk. The latter is measured by the variance of the idiosyncratic component of asset prices (finance) or of output (macroeconomic risk sharing). Moreover, factor models can be used to learn about macroeconomic behaviour on the basis of disaggregated data (sectors, regions): Quah and Sargent (1993), Forni and Reichlin (1996, 1998a, 1998b), Forni and Lippi (1997) are useful references. Finally, factor models can be successfully used for prediction (Stock and Watson, 1998).

In the above examples, n, i.e. the number of cross sectional units (different macro variables, returns on different assets, data disaggregated by sector or region), is typically large, possibly larger than the number T of observations over time. VAR or VARMA models are not appropriate in this case, since they imply the estimation of too many parameters. Factor models are an interesting alternative in that they can provide a much more parsimonious parameterization. To address properly all the economic issues cited above, however, a factor model must have two characteristics. First, it must be dynamic, since business cycle questions are typically dynamic questions. Second, it must allow for cross-correlation among idiosyncratic components, since orthogonality is an unrealistic assumption for most applications.

The model we propose in this paper has both characteristics. It encompasses as a special case the 'approximate factor model' of Chamberlain (1983) and Chamberlain and

Rothschild (1983), which allows for correlated idiosyncratic components, but is static. And it generalizes the factor model of Sargent and Sims (1977) and Geweke (1977), which is dynamic, but has orthogonal idiosyncratic components.

An important feature of our model is that the common component is allowed to have an infinite Moving Average (MA) representation, so as to accommodate for both autoregressive (AR) and MA responses to common factors. In this respect, it is more general than a static factor model where lagged factors are introduced as additional static factors, since in such model AR responses are ruled out.

The paper has three parts: population results, estimation, and empirics. In the population part we show that the common and the idiosyncratic components are asymptotically identified. Moreover, we prove that, if we have q dynamic factors, the first q dynamic principal component series of the observable variables converge to the factor space as $n \to \infty$ and the projection of each variable on the leads and lags of these principal components converges to the common component of the variable.

The second part is focused on estimation. Here we propose an estimator of the common components which is the empirical (finite T) counterpart of the projection above. Building on the population results, we show that such estimator converges to the common component as both n and T go to infinity. Simulation results show that our estimator performs well even when T is relatively small, possibly smaller than n.

In the empirical section we use data on several macroeconomic variables for the countries of the European Monetary Union and compute a 'reference Euro-zone business cycle'. This is defined as a weighted average of the common components of the GDPs of the countries of the Union and can be driven by more than one common factor. On the basis of our results, we also evaluate the performance of variables which are usually taken as reference for the European business cycle such as the sentiment indicator and the spread.

This paper is closely related to three recent papers. Forni and Lippi (1999) analyze the generalized dynamic factor model proposed here from a purely theoretical point of view. They do not deal with estimation problems, but, unlike here, where we assume a factor structure from the start, they provide the conditions in population under which such structure exists. Forni and Reichlin (1998) deal with estimation and empirics and show consistency of an estimator for the common component in a dynamic factor model where the idiosyncratic terms are mutually orthogonal. They also analyze identification of the common factors. Stock and Watson (1998) mainly deal with forecasting in a specification which is different from ours in that it allows for time varying factor loadings but not for autoregressive dynamics.

2. The Model

We suppose that all the stochastic variables taken into consideration belong to the Hilbert space $L_2(\Omega, \mathcal{F}, \mathrm{P})$, where $(\Omega, \mathcal{F}, \mathrm{P})$ is a given probability space; thus all first and second moments are finite. We will study a double sequence

$$\{x_{it},\ i \in \mathbb{N},\ t \in \mathbb{Z}\},$$

where

$$x_{it} = b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} + \xi_{it}, \tag{1}$$

L standing for the lag operator, and suppose that the following Assumptions 1–4 hold.

Assumption 1.
(I) The q-dimensional vector process $\{(\, u_{1t} \quad u_{2t} \quad \cdots \quad u_{qt}\,)',\ t \in \mathbb{Z}\}$ is orthonormal white noise, i.e. $\mathrm{E}(u_{jt}) = 0$, $\mathrm{var}(u_{jt}) = 1$ for any j and t, $u_{jt} \perp u_{jt-k}$ for any j, t, and $k \neq 0$, $u_{jt} \perp u_{st-k}$ for any $s \neq j$, t, and k;
(II) $\boldsymbol{\xi} = \{\xi_{it},\ i \in \mathbb{N},\ t \in \mathbb{Z}\}$ is a double sequence such that, firstly,

$$\boldsymbol{\xi}_n = \{(\, \xi_{1t} \quad \xi_{2t} \quad \ldots \quad \xi_{nt}\,)',\ t \in \mathbb{Z}\}$$

is a zero-mean stationary vector process for any n, and, secondly, $\xi_{it} \perp u_{jt-k}$ for any i, j, t, and k;
(III) the filters $b_{ij}(L)$ are one-sided in L and their coefficients are square summable.

Assumption 1 implies that the n-dimensional vector process $\mathbf{x}_n = \{\mathbf{x}_{nt},\ t \in \mathbb{Z}\}$, where

$$\mathbf{x}_{nt} = (\, x_{1t} \quad x_{2t} \quad \ldots \quad x_{nt}\,)',$$

is zero-mean and stationary for any n. Trend stationary processes can be easily treated with the tools developed below, which are applicable to the stationary residuals from deterministic detrending, while in the case of difference stationary processes our analysis can be applied to the result of differencing and mean subtracting.

The variables u_{jt}, $j = 1, \ldots, q$, will be called the *common shocks* of model (1), the variables $\chi_{it} = x_{it} - \xi_{it}$ and ξ_{it} will be called the *common component* and the *idiosyncratic component* of x_{it}, respectively.

Model (1) is a factor analytic model. It is dynamic as the models employed in Geweke (1977), Sargent and Sims (1977). However, here the cross-sectional dimension is infinite. This feature is the same as in the static factor model of Chamberlain (1983) and Chamberlain and Rothschild (1983). An infinite cross-section, together with Assumptions 3 and 4 below, is crucial for the identification of our model: indeed, and this is the third distinctive feature of (1), which differentiates it from the dynamic factor models mentioned above, we are not assuming mutual orthogonality of the idiosyncratic components ξ_{it}. Without orthogonality, for fixed n, reasonable assumptions allowing for identification of the idiosyncratic and the common component would be very hard to find.

We do not assume rational lag distributions in equations (1). Through Section 3.1, we only impose a bounded spectral density for $\{x_{it}\}$, for any i. In Section 3.2 further requirements, allowing for consistent estimation, will be introduced. We denote by $\mathbf{\Sigma}_n(\theta)$ the spectral density matrix of the vector process $\mathbf{x}_{nt}$ and by $\sigma_{ij}(\theta)$ its entries (note that the matrices $\mathbf{\Sigma}_n$ and $\mathbf{\Sigma}_m$, $n < m$, are nested, so that no reference to n is necessary for $\sigma_{ij}(\theta)$).

Assumption 2. For any $i \in \mathbb{N}$, there exists a real $c_i > 0$ such that $\sigma_{ii}(\theta) \leq c_i$ for any $\theta \in [-\pi,\ \pi]$.

Note that we are not assuming that boundedness of $\sigma_{ii}(\theta)$ is uniform in i. Note also that Assumption 2 implies that all the entries of $\mathbf{\Sigma}_n(\theta)$ are bounded in modulus.

Now, denote by λ_{nj} the function associating with any $\theta \in [-\pi,\ \pi]$ the real non-negative j-th eigenvalue of $\mathbf{\Sigma}_n(\theta)$ in descending order of magnitude. The functions λ_{nj} will be called the *dynamic eigenvalues* of $\mathbf{\Sigma}_n$.[2] In the same way, with obvious notation, λ^{χ}_{nj} and λ^{ξ}_{nj} denote the dynamic eigenvalues of $\mathbf{\Sigma}^{\chi}_n$ and $\mathbf{\Sigma}^{\xi}_n$, respectively. The latter will be called common and idiosyncratic eigenvalues respectively.

Assumption 3. The first idiosyncratic dynamic eigenvalue λ^{ξ}_{n1} is uniformly bounded, i.e., there exists a real Λ such that $\lambda^{\xi}_{n1}(\theta) \leq \Lambda$ for any $\theta \in [-\pi,\ \pi]$ and any $n \in \mathbb{N}$.

Assumption 4. The first q common dynamic eigenvalues diverge almost everywhere in $[-\pi,\ \pi]$, i.e., $\lim_{n\to\infty} \lambda^{\chi}_{nj}(\theta) = \infty$ for $j \leq q$, a.e. in $[-\pi,\ \pi]$.

Assumptions 3 and 4 call for some explanation. Assumption 3 is clearly satisfied if the x's are mutually orthogonal at any lead and lag and have uniformly bounded spectral densities, but is more general as it allows, so to speak, for a limited amount of dynamic cross-correlation. Similarly, Assumption 4 guarantees a minimum amount of cross-correlation between the common components. With a slight oversimplification, Assumption 4 implies that each u_{jt} is present in infinitely many cross-sectional units, with non-decreasing importance (on Assumption 4 see also Remark 5 in Section 3.1). On the contrary, Assumption 3 implies that idiosyncratic causes of variation, although possibly shared by many (even all) units, have their effects concentrated on a finite number of units, and tending to zero as i tends to infinity. For example, Assumption 3 is fulfilled if $\text{var}(\xi_{it}) = 1$, $\text{cov}(\xi_{it}, \xi_{i+1t}) = \rho \neq 0$, while $\text{cov}(\xi_{it}, \xi_{i+ht}) = 0$ for $h > 1$.

Note that in Assumption 4 we require divergence "almost everywhere". The reason is twofold. Firstly, we do not need divergence everywhere to prove our results. Secondly, cases in which divergence does not hold everywhere can arise in very elementary situations. Suppose, for example, that $x_{it} = u_t + \xi_{it}$, where ξ_{it} is non-stationary but $(1 - L)\xi_{it}$ is stationary. Then consider the variables $(1 - L)x_{it} = (1 - L)u_t + (1 - L)\xi_{it}$. Assuming that the variables $(1 - L)\xi_{it}$ fulfill Assumption 3, the model for the variables $(1 - L)x_{it}$ fulfills Assumptions 1 through 4 with $\chi_{it} = (1 - L)u_t$ and $\lambda^{\chi}_{n1}(\theta) = n|1 - e^{-i\theta}|^2$, which is divergent in $[-\pi, \ \pi]$ with the exception of $\theta = 0$.

Our first result is the following.

Proposition 1. *Under Assumptions 1 through 4, the first q eigenvalues of $\mathbf{\Sigma}_n$ diverge, as $n \to \infty$, a.e. in $[-\pi, \ \pi]$, whereas the $(q + 1)$-th one is uniformly bounded, i.e. there exists a real M such that $\lambda_{nq+1}(\theta) \leq M$ for any $\theta \in [-\pi, \ \pi]$ and any $n \in \mathbb{N}$.*

Proof. See the Appendix.

The importance of Proposition 1 lies in the fact that it transforms statements on the dynamic eigenvalues associated with the unobservable components $\boldsymbol{\chi}_n$ and $\boldsymbol{\xi}_n$ into statements on the dynamic eigenvalues associated with $\mathbf{x}_n$, which is supposed observable. Moreover, as proved in Forni and Lippi (1999), the converse of Proposition 1 also holds, i.e. if the first q eigenvalues of $\mathbf{\Sigma}_n$ diverge, as $n \to \infty$, a.e. in $[-\pi, \ \pi]$, whereas the $(q+1)$-th one is uniformly bounded, then the x's can be represented as in (1). Thus, if the analysis of the dynamic eigenvalues of the observed process leads to the conclusion that the first q eigenvalues diverge a.e. in $[-\pi, \ \pi]$, whereas the $(q + 1)$-th one is uniformly bounded, then the hypothesis of a model of the form (1) with q factors is plausible.

We call model (1), under Assumptions 1 to 4, the *generalized dynamic factor model.* We will show that, under Assumptions 1 through 4, the common components χ_{it} and the idiosyncratic components ξ_{it} are identified and can be consistently estimated. On the other hand, it must be stressed that in this paper we do not deal with identification and estimation of the shocks u_{jt} or the filters $b_{ij}(L)$. Thus we are not interested here in whether representation (1) has a structural interpretation or not.[3] In this respect, even the assumption that the filters $b_{ij}(L)$ are one-sided could be dropped with no consequence.

3. Recovering the Common Components

3.1 Population results

In this section our task is the construction of an estimator of χ_{it}, for any given i, based on the finite set of variables $\{x_{it},\ i = 1, \ldots, n,\ t = 1, \ldots, T\}$, and prove consistency for such an estimator as n and T tend to infinity. The proof is obtained in two steps: (A) In this subsection, we consider the projection of x_{it} on all leads and lags of the first q dynamic principal components (see the definition below) of $\mathbf{x}_n$, obtained from the *population* spectral density matrix $\mathbf{\Sigma}_n$. We show that this projection, call it $\chi_{it,n}$, converges to χ_{it} in mean square as n tends to infinity. (B) In subsection 3.2, we construct the finite-sample counterpart of $\chi_{it,n}$, which is based on the estimated spectral density $\mathbf{\Sigma}_n^T$, call it $\chi_{it,n}^T$. Then we combine convergence of $\chi_{it,n}$ to χ_{it}, with the fact that $\chi_{it,n}^T$ is a consistent estimator of $\chi_{it,n}$ for any n as T tends to infinity, thus obtaining the desired result.

Let us recall that given the spectral density matrix $\mathbf{\Sigma}_n(\theta)$, there exist n vectors of complex-valued functions

$$\mathbf{p}_{nj}(\theta) = (\, p_{nj,1}(\theta) \quad p_{nj,2}(\theta) \quad \cdots \quad p_{nj,n}(\theta)\,),$$

$j = 1, 2, \ldots, n$, such that

(i) $\mathbf{p}_{nj}(\theta)$ is a row eigenvector of $\mathbf{\Sigma}_n(\theta)$ corresponding to $\lambda_{nj}(\theta)$, i.e.,

$$\mathbf{p}_{nj}(\theta)\mathbf{\Sigma}_n(\theta) = \lambda_{nj}(\theta)\mathbf{p}_{nj}(\theta) \quad \text{for any} \quad \theta \in [-\pi,\ \pi];$$

(ii) $|\mathbf{p}_{nj}(\theta)|^2 = 1$ for any j and $\theta \in [-\pi,\ \pi]$;

(iii) $\mathbf{p}_{nj}(\theta)\tilde{\mathbf{p}}_{ns}(\theta) = 0$ for $j \neq s$ and any $\theta \in [-\pi,\ \pi]$;

(iv) $\mathbf{p}_{nj}(\theta)$ is measurable on $[-\pi,\ \pi]$;

where, as usual, we denote by $\tilde{D}$ the adjoint (transposed, complex conjugate) of a matrix D (for existence and properties of the functions $\mathbf{p}_{nj}(\theta)$, see Brillinger, 1981, Chapter 9, and Forni and Lippi, 1999).

An n-tuple fulfilling properties (i) through (iv) will be called a set of *dynamic eigenvectors* of $\mathbf{\Sigma}_n$. Note that, apart from some inevitable complication, dynamic eigenvectors are nothing else than eigenvectors of the spectral density matrix, as functions of the frequency θ. A consequence of (ii) and (iv) is that dynamic eigenvectors can be expanded in Fourier series:

$$\mathbf{p}_{nj}(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \left[\int_{-\pi}^{\pi} \mathbf{p}_{nj}(\theta) e^{ik\theta} d\theta \right] e^{-ik\theta}$$

(this is the componentwise Fourier expansion of the vector $\mathbf{p}_{nj}(\theta)$), where the series on the right hand side converges in mean square.

Defining

$$\underline{\mathbf{p}}_{nj}(L) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\left[\int_{-\pi}^{\pi}\mathbf{p}_{nj}(\theta)e^{ik\theta}d\theta\right]L^k,$$

the filter $\underline{\mathbf{p}}_{nj}(L)$ is square summable. Moreover, Assumption 2 implies that the scalar $\underline{\mathbf{p}}_{nj}(L)\mathbf{x}_{nt}$ converges in mean square (Brockwell and Davis, 1987, p. 149, Theorem 4.10.1). For $j = 1, \dots, n$, the scalar process $\{\underline{\mathbf{p}}_{nj}(L)\mathbf{x}_{nt},\ t \in \mathbb{Z}\}$, whose spectral density is

$$\mathbf{p}_{nj}(\theta)\mathbf{\Sigma}_n(\theta)\tilde{\mathbf{p}}_{nj}(\theta) = \lambda_{nj}(\theta),$$

will be called the j-th *dynamic principal component* of $\mathbf{x}_n$. A consequence of (iii) is that if $j \neq k$ then the j-th and k-th principal components are orthogonal at any lead and lag.

Now consider the minimal closed subspace of $L_2(\Omega, \mathcal{F}, \mathrm{P})$ containing the first q principal components

$$\mathcal{U}_n = \overline{\mathrm{span}}(\underline{\mathbf{p}}_{nj}(L)\mathbf{x}_{nt}(L),\ j = 1, \dots, q,\ t \in \mathbb{Z}),$$

and the orthogonal projection

$$\chi_{it,n} = \mathrm{proj}(x_{it}|\mathcal{U}_n).$$

We can obtain an explicit formula both for $\chi_{it,n}$ and the residual $\xi_{it,n} = x_{it} - \chi_{it,n}$ by observing that

$$\mathbf{I}_n = \tilde{\mathbf{p}}_{n1}(\theta)\mathbf{p}_{n1}(\theta) + \tilde{\mathbf{p}}_{n2}(\theta)\mathbf{p}_{n2}(\theta) + \cdots + \tilde{\mathbf{p}}_{nn}(\theta)\mathbf{p}_{nn}(\theta)$$

(the vectors $\mathbf{p}_{nj}(\theta)$ are an orthonormal system of eigenvectors for $\mathbf{I}_n$). Therefore

$$\mathbf{x}_{nt} = \underline{\tilde{\mathbf{p}}}_{n1}(L)\underline{\mathbf{p}}_{n1}(L)\mathbf{x}_{nt} + \underline{\tilde{\mathbf{p}}}_{n2}(L)\underline{\mathbf{p}}_{n2}(L)\mathbf{x}_{nt} + \cdots + \underline{\tilde{\mathbf{p}}}_{nn}(L)\underline{\mathbf{p}}_{nn}(L)\mathbf{x}_{nt}.$$

Taking the i-th coordinate:

$$\begin{aligned} x_{it} = &\left[\underline{\tilde{p}}_{n1,i}(L)\underline{\mathbf{p}}_{n1}(L)\mathbf{x}_{nt} + \underline{\tilde{p}}_{n2,i}(L)\underline{\mathbf{p}}_{n2}(L)\mathbf{x}_{nt} + \cdots + \underline{\tilde{p}}_{nq,i}(L)\underline{\mathbf{p}}_{nq}(L)\mathbf{x}_{nt}\right] \\ &+ \left[\underline{\tilde{p}}_{n\,q+1,i}(L)\underline{\mathbf{p}}_{n\,q+1}(L)\mathbf{x}_{nt} + \cdots + \underline{\tilde{p}}_{nn,i}(L)\underline{\mathbf{p}}_{nn}(L)\mathbf{x}_{nt}\right]. \end{aligned}$$

Now, since the dynamic principal components are mutually orthogonal at any lead and lag, then

$$\chi_{it,n} = \underline{\mathbf{K}}_{ni}(L)\mathbf{x}_{nt}, \tag{2}$$

with

$$\mathbf{K}_{ni}(\theta) = \tilde{p}_{n1,i}(\theta)\mathbf{p}_{n1}(\theta) + \tilde{p}_{n2,i}(\theta)\mathbf{p}_{n2}(\theta) + \cdots + \tilde{p}_{nq,i}(\theta)\mathbf{p}_{nq}(\theta).$$

Remark 1. Note that: (a) in (2) the orthogonal projection $\chi_{it,n}$ is expressed as the sum of the orthogonal projections of x_{it} on (leads and lags of) each of the first q dynamic principal components; (b) the coefficients of the j-th orthogonal projection are the coefficients of the filter $\tilde{\underline{\mathbf{p}}}_{nj}(L)$; (c) obviously, analogous formulae and observations hold for $\xi_{it,n}$ and the principal components from $q+1$ to n.

Let us now state and comment our first step toward recovering χ_{it}.

Proposition 2. *Suppose that Assumptions 1 through 4 hold. Then*

$$\lim_{n\to\infty} \chi_{it,n} = \chi_{it}$$

in mean square for any i and t.

Proof. See the Appendix.

Remark 2. Note firstly that $\chi_{it,n}$, that is, the population approximate common component of x_{it}, results from a simple rule involving the dynamic eigenvectors of the matrices $\mathbf{\Sigma}_n$, with no intervention of the unobservable χ's and ξ's. Thus we are ready for the second step, in which we construct an empirical approximate common component based on the observable $\mathbf{x}_{nt}$, for $t = 1, \dots, T$.

Remark 3. An intuitive insight into Proposition 2 can be obtained by considering the following example

$$x_{it} = u_t + \xi_{it}, \tag{3}$$

where all ξ's are white noise, have unit variance, and are mutually orthogonal at any lead and lag. In this one-factor case $\underline{\mathbf{p}}_{n1}(L) = (\,1/\sqrt{n} \quad 1/\sqrt{n} \quad \cdots \quad 1/\sqrt{n}\,)$, so that

$$\chi_{it,n} = \tilde{\underline{p}}_{n1,i}(L)\underline{\mathbf{p}}_{n1}(L)\mathbf{x}_{nt} = (\,1/n \quad 1/n \quad \cdots \quad 1/n\,)\,\mathbf{x}_{nt} = u_t + \frac{1}{n}\sum_{s=1}^{n}\xi_{st}.$$

Convergence of $\chi_{it,n}$ to χ_{it} in mean square thus follows from $\text{var}\,(\sum_{s=1}^{n}\xi_{st}/n) = 1/n$. In this example the filter $\tilde{\underline{p}}_{n1,i}(L)\underline{\mathbf{p}}_{n1}(L)$ is nothing else than the standard arithmetic mean of $\mathbf{x}_{nt}$. In the Appendix we show that in general the filters $\tilde{\underline{p}}_{nj,i}(L)\underline{\mathbf{p}}_{nj}(L)$, for $j = 1, \dots, q$, which average the x's both over the cross-section and over time, share with the standard arithmetic mean the property that the sum of the squared coefficients tends to zero as n tends to infinity. Assumption 3 indeed ensures that $\tilde{\underline{p}}_{nj,i}(L)\underline{\mathbf{p}}_{nj}(L)\boldsymbol{\xi}_{nt}$ vanishes as n tends to infinity (see the Appendix), so that, since

$$\tilde{\underline{p}}_{nj,i}(L)\underline{\mathbf{p}}_{nj}(L)\mathbf{x}_{nt} = \tilde{\underline{p}}_{nj,i}(L)\underline{\mathbf{p}}_{nj}(L)\boldsymbol{\chi}_{nt} + \tilde{\underline{p}}_{nj,i}(L)\underline{\mathbf{p}}_{nj}(L)\boldsymbol{\xi}_{nt},$$

in the limit only the term $\tilde{\underline{p}}_{nj,i}(L)\underline{\mathbf{p}}_{nj}(L)\boldsymbol{\chi}_{nt}$ survives. However, proving that in general $\sum_{j=1}^{q} \tilde{\underline{p}}_{nj,i}(L)\underline{\mathbf{p}}_{nj}(L)\boldsymbol{\chi}_{nt}$ converges to χ_{it} is not as elementary as in model (3).

Remark 4. Assume again, for simplicity, that $q = 1$ but that the model is general: $x_{it} = b_i(L)u_t + \xi_{it}$. Now suppose that we take the standard arithmetic mean $\bar{x}_{nt}$ of $\mathbf{x}_{nt}$, instead of the first dynamic principal component and that we project x_{it} on all leads and lags of $\bar{x}_{nt}$. Call $\bar{\chi}_{it,n}$ this projection. Assumption 3 ensures that the idiosyncratic part of $\bar{x}_{nt}$ tends to zero, so that the projection $\bar{\chi}_{it,n}$ tends to the projection of x_{it} on the space spanned by the common components, i.e. χ_{it}. This estimation method can be extended to $q > 1$ by using q averages with different systems of weights, as in Forni and Reichlin (1998). An advantage of their method is that the coefficients of their averages are independent of the x's and not estimated (as in our case). However, unless *ad hoc* assumptions are introduced, near singularity of the chosen averages for n growing, with the consequence of inaccurate estimation, cannot be excluded. This problem is completely solved with dynamic principal components, which are mutually orthogonal at any lead and lag.

Since $\chi_{it,n}$ depends only on $\mathbf{x}_{nt}$, Proposition 2 has the immediate implication that the components χ_{it} and ξ_{it} are identified. More precisely, we can state the following Corollary.

Corollary 1. *Suppose that x_{it} can be represented as in (1), and that Assumptions 1 through 4 are fulfilled. Suppose that x_{it} admits the alternative representation*

$$x_{it} = \check{b}_{i1}(L)\check{u}_{1t} + \check{b}_{i2}(L)\check{u}_{2t} + \cdots + \check{b}_{i\check{q}}(L)\check{u}_{\check{q}t} + \check{\xi}_{it}, \tag{4}$$

and that Assumptions 1 through 4 are also fulfilled for (4). Then, $\check{\chi}_{it} = \chi_{it}$, so that $\check{\xi}_{it} = \xi_{it}$. Moreover $\check{q} = q$.

Remark 5. An important consequence of Corollary 1 is that representation (1) is *non-redundant*, i.e. no other representation fulfilling Assumptions 1 through 4 is possible with a smaller number of factors. In the following example we have a common-idiosyncratic representation of the form (1) with one factor. However, since Assumption 4 is not fulfilled, another representation with zero factors fulfilling Assumptions 1 through 4 is possible. Specify (1) as

$$x_{it} = b_i u_t + \xi_{it},$$

where $\boldsymbol{\xi}$ is defined as in model (3). Now suppose that the sequence of coefficients b_i, $i \in \mathbb{N}$, is square summable, i.e. that $\sum_{i=1}^{\infty} b_i^2 < \infty$. In this case, as the reader can easily check, the first eigenvalue of $\boldsymbol{\Sigma}_n(\theta)$ is $1 + \sum_{i=1}^{n} b_i^2$, and is therefore bounded as n tends to infinity. Thus the x's, though the correlation between x_{it} and x_{jt} never vanishes, are idiosyncratic.

Naturally, in empirical situations we do not know the number q. However, another implication of Proposition 2 is that assuming a q^* larger than the actual q has no dramatic consequences, since the expected mean-squared difference between the resulting projections $\chi^*_{it,n}$ and $\chi^*_{it,n}$, averaged over the cross-sectional units, is asymptotically zero. Precisely:

Corollary 2. *Under Assumptions 1 through 4, let $\chi^*_{it,n}$ be the projection of x_{it} on the space spanned by all leads and lags of the first q^* dynamic principal components, with $q^* > q$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left[(\chi^*_{it,n} - \chi_{it,n})^2\right] = 0.$$

Proof. See the Appendix.

A dynamic factor model with an infinite cross-sectional dimension is studied in Stock and Watson (1998). Among several differences, let us observe here that their model is more general than ours in that their factor loading coefficients are allowed to be time-varying. On the other hand, in Stock and Watson's paper the common components are modeled (in our notation and assuming for simplicity only one factor) as $c_i(L)c(L)u_t$ with polynomials $c_i(L)$ of finite order, which is dynamically more restrictive than (1). Stock and Watson construct estimated factors that converge to the space spanned by the "true" factors. This corresponds, in this paper, to the statement that the estimated counterparts of $\underline{\mathbf{p}}_{nj}(L)\mathbf{x}_{nt}$ converge to the space spanned by the χ's (or the u's). In this paper we prove this result and go a step further, showing that the estimated $\chi_{it,n}$ converges to χ_{it} for any i (see the comment under Lemma 4, Appendix).

3.2 Estimation Results

Proposition 2 shows that the common component χ_{it} can be recovered asymptotically from the sequence $\underline{\mathbf{K}}_{ni}(L)\mathbf{x}_{nt}$. The filters $\underline{\mathbf{K}}_{nj}(L)$ are obtained as functions of the spectral density matrices $\mathbf{\Sigma}_n(\theta)$. Now, in practice, the population spectral densities $\mathbf{\Sigma}_n(\theta)$ must be replaced by their empirical counterparts based on finite realizations of the form

$$\mathbf{X}_n^T = (\, \mathbf{x}_{n1} \quad \mathbf{x}_{n2} \quad \cdots \quad \mathbf{x}_{nT}\,).$$

On the other hand, consistent estimation of the spectral density requires a strengthening of Assumption 2. Precisely, we replace Assumption 2 by:

Assumption 2′. The vector $\mathbf{x}_{nt}$ has a representation

$$\mathbf{x}_{nt} = \sum_{k=-\infty}^{\infty} C_k Z_{t-k},$$

where Z_t is an n-dimensional white noise with non-singular covariance matrix and fourth order moments, and

$$\sum_{k=-\infty}^{\infty} |C_{ij,k}||k|^{1/2} < \infty,$$

for $i, j = 1, \ldots, n$, where $C_{ij,k}$ is the i, j entry of C_k.

Under Assumption $2'$, if $\boldsymbol{\Sigma}_n^T(\theta)$ denotes any periodogram-smoothing or lag-window estimator of $\boldsymbol{\Sigma}_n(\theta)$, based on $\mathbf{X}_n^T$, we have

$$\lim_{T\to\infty} \mathrm{P}\left[\sup_{\theta\in[-\pi,\ \pi]} |\sigma_{ij}^T(\theta) - \sigma_{ij}(\theta)| > \epsilon\right] = 0, \tag{5}$$

where $\sigma_{ij}^T(\theta)$ denotes the i, j entry of $\boldsymbol{\Sigma}_n^T(\theta)$ (see Brockwell and Davis 1987, p. 433). Under Assumption $2'$, the estimated counterpart of $\mathbf{K}_{ni}(\theta)$ allows for a consistent reconstruction of the factor space. More precisely, we prove that the projection of x_{it} onto the space spanned by the first q empirical principal components converges to the common component χ_{it}.

Denote by $\lambda_{nj}^T(\theta)$ and $\mathbf{p}_{nj}^T(\theta)$, respectively, the eigenvalues and eigenvectors of the matrix $\boldsymbol{\Sigma}_n^T(\theta)$. Since eigenvalues and eigenvectors are continuous functions of the entries of the corresponding matrix, (5) implies that $\lambda_{nj}^T(\theta)$ and $\mathbf{p}_{nj}^T(\theta)$ converge to $\lambda_{nj}(\theta)$ and $\mathbf{p}_{nj}(\theta)$, respectively, in probability, uniformly in $\theta \in [-\pi,\ \pi]$, for $T \to \infty$. Moreover, considering

$$\mathbf{K}_{ni}^T(\theta) = \tilde{p}_{n1,i}^T(\theta)\mathbf{p}_{n1}^T(\theta) + \tilde{p}_{n2,i}^T(\theta)\mathbf{p}_{n2}^T(\theta) + \cdots + \tilde{p}_{nq,i}^T(\theta)\mathbf{p}_{nq}^T(\theta),$$

i.e., the empirical counterpart of $\mathbf{K}_{ni}(\theta)$, $i \leq q$, $\mathbf{K}_{ni}^T(\theta)$ converges to $\mathbf{K}_{ni}(\theta)$ in probability, uniformly in $\theta \in [-\pi,\ \pi]$, for $T \to \infty$. Thus, for all $\epsilon > 0$ and $\eta > 0$, there exists $T_1 = T_1(n, \epsilon, \eta)$ such that, for all $T \geq T_1$,

$$\mathrm{P}\left[\sup_{\theta\in[-\pi,\ \pi]} |\mathbf{K}_{ni}^T(\theta) - \mathbf{K}_{ni}(\theta)| > \epsilon\right] \leq \eta. \tag{6}$$

Now, observe that, in principle, given the estimated spectral density matrix $\boldsymbol{\Sigma}_n^T(\theta)$, $\mathbf{K}_{ni}^T(\theta)$ can be computed for any θ, so that each of the coefficients of the corresponding two-sided filter

$$\underline{\mathbf{K}}_{ni}^T(L) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \left[\int_{-\pi}^{\pi} \mathbf{K}_{ni}^T(\theta) e^{-ik\theta} d\theta\right] L^k$$

can be obtained. However, in practice, the projection $\underline{\mathbf{K}}_{ni}^T(L)\mathbf{x}_{nt}$ of x_{it} onto the space spanned by the first q empirical principal components cannot be computed, since, for $t \leq 0$ and $t > T$, $\mathbf{x}_{nt}$ is not available; $\underline{\mathbf{K}}_{ni}(L)$ actually has to be truncated at lag $t-1$ and lead $T-t$, respectively, yielding the finite-order filter $\underline{\mathbf{K}}_{ni}^{Tt}(L)$. Due to this truncation, the common component χ_{it}, for fixed t, never can be recovered, even as n and T tend to infinity: indeed, part of its variance is lost because of the non-observability of $\mathbf{x}_{nt}$, $t \leq 0$ and $t > T$. We therefore restrict our attention to the "central part" of the observed series, i.e. to values of t of the form $t = t^*(T)$, with

$$0 < a \leq \liminf_{T\to\infty} \frac{t^*(T)}{T} \leq \limsup_{T\to\infty} \frac{t^*(T)}{T} \leq b < 1. \tag{7}$$

The following result then provides the empirical counterpart of Proposition 2.

PROPOSITION 3. *Assume that Assumptions 1, 2′, 3 and 4 are satisfied. Then, for all $\epsilon > 0$ and $\eta > 0$, there exists $N_0(\epsilon, \eta)$ such that*

$$\mathrm{P}\left[|\underline{\mathbf{K}}_{ni}^{Tt}(L)\mathbf{x}_{nt} - \chi_{it}| > \epsilon \right] \leq \eta$$

for all $t = t^*(T)$ satisfying (7), all $n \geq N_0$ and all T larger than some $T_0(n, \epsilon, \eta)$.

PROOF. For any $t \leq T$, we have

$$\begin{aligned}
&\mathrm{P}\left[|\underline{\mathbf{K}}_{ni}^{Tt}(L)\mathbf{x}_{nt} - \chi_{it}| > \epsilon \right] \\
&\quad \leq \mathrm{P}\left[|(\underline{\mathbf{K}}_{ni}^{Tt}(L) - \underline{\mathbf{K}}_{ni}(L))\mathbf{x}_{nt}| > \epsilon/2 \right] + \mathrm{P}\left[|\underline{\mathbf{K}}_{ni}(L)\mathbf{x}_{nt} - \chi_{it}| > \epsilon/2 \right] \\
&\quad = R_{n1}^{Tt} + R_{n2}, \quad \text{say.}
\end{aligned}$$

Proposition 2 ensures the existence of an $N_0(\epsilon, \eta)$ such that, for $n \geq N_0$, $R_{n2} \leq \frac{\eta}{2}$. As for R_{n1}^{Tt}, it follows from Chebyshev's theorem and (6) that, for $T \geq T_1(n, \delta, \frac{\eta}{4})$,

$$\begin{aligned}
R_{n1}^{Tt} \leq\; &\mathrm{P}\left[|(\underline{\mathbf{K}}_{ni}^{Tt}(L) - \underline{\mathbf{K}}_{ni}(L))\mathbf{x}_{nt}| > \epsilon/2 \;\text{ and }\; \sup_{\theta\in[-\pi,\ \pi]} |\mathbf{K}_{ni}^{Tt}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta \right] \\
&+ \mathrm{P}\left[\sup_{\theta\in[-\pi,\ \pi]} |\mathbf{K}_{ni}^{Tt}(\theta) - \mathbf{K}_{ni}(\theta)| > \delta \right] \\
\leq\; &\frac{4}{\epsilon^2}\mathrm{E}\left[|(\underline{\mathbf{K}}_{ni}^{Tt}(L) - \underline{\mathbf{K}}_{ni}(L))\mathbf{x}_{nt}|^2 \;\middle|\; \sup_{\theta\in[-\pi,\ \pi]} |(\mathbf{K}_{ni}^{Tt}(\theta) - \mathbf{K}_{ni}(\theta))| \leq \delta \right] + \frac{\eta}{4}.
\end{aligned}$$

If the filter $\underline{\mathbf{K}}_{ni}^{Tt}(L)$ and the observation $\mathbf{x}_{nt}$ were independent, then, in view of the classical properties of eigenvalues (see the proof of Proposition 1, Appendix), the above expression would reduce to

$$\frac{4}{\epsilon^2}\mathrm{E}\left[|(\underline{\mathbf{K}}_{ni}^{Tt}(L)-\underline{\mathbf{K}}_{ni}(L))\mathbf{x}_{nt}|^2 \,\middle|\, \sup_{\theta\in[-\pi,\ \pi]}|(\mathbf{K}_{ni}^{Tt}(\theta)-\mathbf{K}_{ni}(\theta))\leq\delta\right]+\frac{\eta}{4}$$

$$=\frac{4}{\epsilon^2}\mathrm{E}\Bigg[\int_{-\pi}^{\pi}(\mathbf{K}_{ni}^{Tt}(\theta)-\mathbf{K}_{ni}(\theta))\mathbf{\Sigma}_n(\theta)(\tilde{\mathbf{K}}_{ni}^{Tt}(\theta)-\tilde{\mathbf{K}}_{ni}(\theta))d\theta$$

$$\left|\ \sup_{\theta\in[-\pi,\ \pi]}|(\mathbf{K}_{ni}^{Tt}(\theta)-\mathbf{K}_{ni}(\theta))|\leq\delta\right]+\frac{\eta}{4}$$

$$\leq\frac{4\delta^2}{\epsilon^2}\int_{-\pi}^{\pi}\lambda_{n1}(\theta)d\theta+\frac{\eta}{4}.$$

Thus, for $n\geq N_0(\epsilon,\eta)$, and $T\geq T_1(n,\delta,\eta/4)$, with $\delta^2=\frac{\epsilon^2\eta}{16\int_{-\pi}^{\pi}\lambda_{n1}(\theta)d\theta}$, we would obtain $R_{n1}^{Tt}+R_{n2}\leq\eta$ for all $t=t^*(T)$ satisfying (7). The proposition follows.

This reasoning is essentially correct, and it carries the basic idea of the proof, since the dependence between $\underline{\mathbf{K}}_{ni}^{Tt}(L)$ and $\mathbf{x}_{nt}$ vanishes as $T\to\infty$. A formal treatment, however, requires a slightly more elaborate argument: see the Appendix.

4. The proposed estimator and the choice of q

In the light of the results of the previous section we propose the following estimator. For a fixed integer M, we compute the sample covariance matrix $\mathbf{\Gamma}_{nk}^T$ of $\mathbf{x}_{nt}$ and $\mathbf{x}_{nt-k}$ for $k=0,1,\ldots,M$ and the $(2M+1)$ points discrete Fourier transform of the truncated two-sided sequence $\mathbf{\Gamma}_{n,-M}^T,\ldots,\mathbf{\Gamma}_{n0}^T,\ldots,\mathbf{\Gamma}_{nM}^T$, where $\mathbf{\Gamma}_{n,-k}=\mathbf{\Gamma}'_{nk}$. More precisely, we compute

$$\mathbf{\Sigma}_n^T(\theta_h)=\sum_{k=-M}^{M}\mathbf{\Gamma}_{nk}^T\omega_k e^{-ik\theta_h},\tag{8}$$

where

$$\theta_h=2\pi h/(2M+1),\quad h=0,1,\ldots,2M$$

and $\omega_k=1-\frac{|k|}{(M+1)}$ are the weights corresponding to the Bartlett lag window of size $M=M(T)$. Consistent estimation of $\mathbf{\Sigma}_n(\theta)$ (which is required for the validity of Proposition 3) is ensured, provided that $M(T)\to\infty$ and $M(T)/T\to 0$ as $T\to\infty$.

Then we compute the first q eigenvectors $\mathbf{p}_{nj}^T(\theta_h)$, $j=1,2,\ldots,q$, of $\mathbf{\Sigma}_n^T(\theta_h)$, for $h=0,1,\ldots,2M$.[4] Finally, for $h=0,1,\ldots,2M$, we construct

$$\mathbf{K}_{ni}^T(\theta_h)=\tilde{p}_{n1,i}^T(\theta_h)\mathbf{p}_{n1}^T(\theta_h)+\cdots+\tilde{p}_{nq,i}^T(\theta_h)\mathbf{p}_{nq}^T(\theta_h).$$

The proposed estimator of the filter $\underline{\mathbf{K}}_{nj}(L)$, $j=1,2,\ldots,q$, is obtained by the inverse discrete Fourier transform of the vector

$$(\,\mathbf{K}_{ni}^T(\theta_0),\quad\cdots,\quad\mathbf{K}_{ni}^T(\theta_{2M})\,),$$

i.e. by the computation of

$$\underline{\mathbf{K}}^T_{ni,k} = \frac{1}{2M+1}\sum_{h=0}^{2M}\mathbf{K}^T_{ni}(\theta_h)e^{ik\theta_h}$$

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

$$\underline{\mathbf{K}}^T_{ni}(L) = \sum_{k=-M}^{M}\underline{\mathbf{K}}^T_{ni,k}L^k. \tag{9}$$

In order to render our procedure operational, we need a rule for fixing M. In the simulations of the following section we simply set $M = (\text{round}\sqrt{T}/4)$, a rule which performs remarkably well. As an alternative, we could take a maximum value for M, $M_0(T)$, say, such that $M_0(T) \to \infty$ and $M_0(T)/T \to 0$ as $T \to \infty$, estimate all of the specifications with $0 \leq M \leq M_0(T)$, and choose the one minimizing some dynamic specification criterion. While in principle a data-dependent rule seems preferable, we found that the standard AIC and BIC criteria underestimate the optimal lag-window size, so that this topic is left for further research.

So far, we have assumed that q, the number of non-redundant common factors, is known. In practice of course, q is not predetermined, and also has to be selected from the data. Proposition 1 can be used to this end, since it links the number of factors in (1) to the eigenvalues of the spectral density matrix of $\mathbf{x}_n$: precisely, if the number of factors is q and $\boldsymbol{\xi}$ is idiosyncratic, then the first q dynamic eigenvalues of $\boldsymbol{\Sigma}_n(\theta)$ diverge a.e. in $[-\pi,\ \pi]$ whereas the $(q+1)$-th one is uniformly bounded.

Indeed, no formal testing procedure can be expected for selecting the number q of factors in finite sample situations. Even letting $T \to \infty$ does not help much. The definition of the idiosyncratic component indeed is of an asymptotic nature, where asymptotics are taken as $n \to \infty$, and there is no way a slowly diverging sequence (divergence, under the model, can be arbitrarily slow) can be told from an eventually bounded sequence (for which the bound can be arbitrarily large). Practitioners thus have to rely on a heuristic inspection of the eigenvalues against the number of series n.

More precisely, if T observations are available for a large number n of variables x_{it}, the spectral density matrices $\boldsymbol{\Sigma}^T_r$, $r \leq n$, can be estimated, and the resulting empirical dynamic eigenvalues λ^T_{rj} computed for a grid of frequencies. The following two features of the eigenvalues computed from $\boldsymbol{\Sigma}^T_r$, $r = 1, \ldots, n$, should be considered as reasonable evidence that the data have been generated by (1), with q factors and that $\boldsymbol{\xi}$ is idiosyncratic:

(a) The average over θ of the first q empirical eigenvalues diverges, whereas the average of the $(q+1)$-th one is relatively stable.

(b) Taking $r = n$ there is a substantial gap between the variance explained by the q-th principal component and the variance explained by $(q+1)$-th one. A preassigned minimum, such as 5%, for the explained variance, could be used as a practical criterion for the determination of the number of common factors to retain. This 5% limit is used in the empirical exercise of Section 6.

To illustrate the use of criteria (a) and (b), we have generated data from a two-factor model (model M4 below) with $n = 50$ and $T = 100$. Then, we have estimated the spectral density matrix for a grid of frequencies, using (8) with $M = 10$. Lastly, we have computed the eigenvalues of the upper-left $r \times r$ submatrix, $r = 1, \ldots, n$.

Figure 4.1. Dynamic eigenvalues averaged over frequencies, model M4

Horizontal axis: r; vertical axis: variance/2π.

Figure 4.1 reports the plot of the average over frequencies of the theoretical and estimated eigenvalues. On the horizontal axis we indicate the number of cross-sectional units r, which obviously is maximum when the whole sample $n = 50$ is considered. Features (a) and (b) emerge quite clearly: the first q averaged eigenvalues exhibit an approximately constant positive slope, while the remaining ones are rather flat; moreover, the variance explained by the q-th principal component is substantially larger than the variance explained by the $(q+1)$-th, even for small r.

To conclude this section, let us remark that, when applying criteria (a) and (b), we should keep in mind that, as indicated by Corollary 2, setting a number of factors larger than the true one cannot have dramatic consequences on estimation.

5. Simulation results

In order to evaluate the performance of our estimation procedure for finite values of n and T, we have carried out Monte Carlo experiments on the following four two-factor models.

Static model:

$$x_{it} = a_i u_{1t} + b_i u_{2t} + \sqrt{2}\xi_{it}. \tag{M1}$$

Static with delay:

$$\begin{aligned} x_{it} &= a_i u_{1t} + b_i u_{2t} + \sqrt{2}\xi_{it} && \text{for } i \text{ even} \\ x_{it} &= a_i u_{1t-1} + b_i u_{2t-1} + \sqrt{2}\xi_{it} && \text{for } i \text{ odd.} \end{aligned} \tag{M2}$$

MA(1) common component:

$$x_{it} = a_{0i} u_{1t} + a_{1i} u_{1t-1} + b_{0i} u_{2t} + b_{1i} u_{2t-1} + 2\xi_{it}. \tag{M3}$$

AR(1) common component:

$$x_{it} = \frac{a_i}{1 - c_i L} u_{1t} + \frac{b_i}{1 - d_i L} u_{2t} + \sqrt{2.5}\xi_{it}. \tag{M4}$$

In all these models, u_{1t}, u_{2t}, a_i, a_{0i}, a_{1i}, b_i, b_{0i}, b_{1i} and ξ_{it} are i.i.d. standard normal deviates, while c_i and d_i are uniformly distributed over $[-0.8, 0.8]$, in order to ensure co-stationarity of the x's. Note that the idiosyncratic shocks are multiplied by a constant so that, on the average, the cross sectional units have the same common-idiosyncratic variance ratio 1 in all models.

TABLE 5.1 HERE

We generated data from each model with $n = 10, 20, 50, 100$ and $T = 20, 50, 100, 200$ and applied the estimation procedure described in Section 4 with $M(T) = \text{round}[\frac{\sqrt{T}}{4}]$. Each experiment was replicated 400 times.

We measured the performance of our estimator, $\hat{\chi}_{it}$, by means of the criterion

$$R(\hat{\chi}, \chi) = \frac{\sum_{i,t} (\hat{\chi}_{it} - \chi_{it})^2}{\sum_{i,t} \chi_{it}^2}.$$

Table 5.1 reports the average and the standard deviation (in brackets) of this statistic across the experiments.

For all models, we see that the fit improves as both n and T increase. To better appreciate the results, we add a row reporting $R(\bar{\chi}, \chi)$, where $\bar{\chi}_{it}$ is the infeasible estimate of the common components obtained by performing OLS regressions of the variables on the contemporaneous and lagged values of the unobservable true common factors u_{jt}; $\bar{\chi}_{it}$ is computed only for $n = 100$. The AIC criterion is used for the choice of the number of lags. Note that for the autoregressive model M4, the results obtained with $n \geq 50$ are similar to those obtained with the true factors or even better, indicating that the error involved in approximating the factor space is negligible as compared with the error arising from the MA approximation of the AR dynamic structure implied by the OLS strategy.

6. A coincident indicator for the EURO currency area

In this section, we use our method to compute a coincident indicator for the countries of the European Monetary Union. We estimate the generalized factor model, using a large panel including several macroeconomic variables for each EURO country. The coincident indicator is constructed as the weighted average of the common components of countries' GDPs.

Our approach is similar in spirit to Stock and Watson (1989), who define the 'reference cycle' as an unobserved index, common to many macroeconomic variables. However, one important difference is that we allow for the possibility that more than one single common shock captures the comovements of the macroeconomic variables of interest. This is relevant whenever there is more than one source of aggregate fluctuations.

We proceed as follows.

STEP 1. We construct a panel pooling seven quarterly macroeconomic indicators for all countries of the EURO zone, excluding Luxembourg, from 1985 to 1996 (see Table 6.1, with X and – indicating, respectively, that the series is available or missing). Data are taken in

TABLE 6.1 HERE

logs and differenced (except for the spread, which is not transformed, and the Sentiment Indicator, which is simply taken in logs), and normalized dividing by the standard deviation.

STEP 2. We estimate the spectral density matrix, compute the dynamic eigenvalues and identify $q = 3$, using criterion (b) of Section 4.

STEP 3. We estimate the common component of GDP for each separate country, following the procedure of Section 4. Figure 6.1 reports the resulting estimates.

STEP 4. We construct the weighted average of the common components above using as weights the GDP levels. This is the proposed coincident indicator. We illustrate it in Figure 6.2. Some remarks are needed. Firstly, results from step 2 show that a cycle in the strong sense of Stock and Watson (1989), i.e. a single common factor is not supported by this data set.

Secondly, in our methodology, output plays a prominent role. Since the data do not support a single static factor, the cycle must be defined as the common component of a particular cross-sectional unit. Clearly, GDP is the most natural choice as the reference variable. On the other hand, we are interested in the common component of output, and not in output itself, because we want to disregard that part of GDP variation which is poorly correlated with other variables. Hence the latter also play an indirect role in the construction of the index, through the estimated dynamic principal components.

Finally, with our methodology there is no need to distinguish *a priori* between leading and coincident variables. The weight of each variable in the index depends on the cross-correlations at all leads and lags: a variable which leads with respect to GDP, for example, will have small contemporaneous weight and will be shifted automatically in the appropriate way.

In order to understand better the structure of the multi-country, multivariate correlations, we also compute the common components of variables other than GDP and construct two sets of statistics. First, for each variable, we compute the ratio of the variance of

Figure 6.1. Common component of GDP of nine countries of the EURO zone

Horizontal axis: time.

the common component to total variance, for each country and for the aggregate EURO area (Table 6.2). These ratios measure the 'degree of commonality' of each variable in the system. Second, we compute, for each variable, the average contemporaneous correlation

Figure 6.2. Coincident indicator for the EURO zone

Horizontal axis: time.

coefficient with the common components of the other variables of the same country, for each country and the aggregate EURO area (Table 6.3). These statistics measure the 'degree of synchronization' of each variable with the other variables of the same country. Through these results, we can also evaluate the performance of variables which are typically used to describe the state of the economy, such as the sentiment indicator and the spread, and validate ex-post the choice of GDP as the reference variable for the European cycle.

TABLE 6.2 HERE

There are few interesting findings. First, the common component of GDP has the largest average contemporaneous correlation for almost all countries and for the aggregate. This fact provides an ex-post confirmation of our choice of the GDP as the reference variable for

TABLE 6.3 HERE

the coincident index. Note however that using directly the GDP, rather than the common component of GDP, as the index, would not be a good choice, due to the presence of an idiosyncratic component which accounts for 15% of total variance.

Second, for most countries and for the aggregate, the sentiment indicator has the largest common component. However its synchronization with the other variables is lower than that of GDP, which suggests that the sentiment indicator is not an appropriate coincident index, probably due to its leading behavior.

Note that the correlations between the common components appearing in Table 6.3 are in general unexpectedly small. This is mainly due to the fact that, somewhat surprisingly, the inflation rate has very low or even negative synchronization.

7. Summary and Discussion

The generalized dynamic factor model analyzed in this paper is novel to the literature, in that it allows for both a dynamic representation of the common component and non-orthogonal idiosyncratic components. We have shown that, although for a finite cross-sectional dimension this model is not identified, identification of the common and the idiosyncratic components is obtained asymptotically as the cross-sectional dimension goes to infinity.

Since the idiosyncratic components are correlated, the model cannot be estimated on the basis of traditional methods. We have proposed a new method, yielding consistent estimates of the components as both the cross-section and the time dimensions go to infinity at some rate. More precise information on their rate would be interesting; however, such information typically would require much heavier assumptions on the heterogeneity of cross-sectional units. This is a topic of the ongoing research by the authors. The common components are computed as the projections of the observations onto the leads and lags of the dynamic principal components of the observations and the idiosyncratic components are derived as the orthogonal residuals.

The method is applied to a panel including several macroeconomic indicators for each of the EURO countries, in order to obtain an index describing the state of the economy in the EURO area. The European coincident indicator is defined as the common component of the European GDP.

REFERENCES

Apostol, Tom M., *Mathematical Analysis* (Reading, MA: Addison Wesley, 1974).

Billingsley, Patrick, *Probability and Measure*, 3rd edition (New York: J. Wiley, 1995).

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

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

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

Chamberlain, Gary, "Funds, factors, and diversification in arbitrage pricing models", *Econometrica* 51 (1983), 1281-1304.

Chamberlain, Gary, and Michael Rothschild, "Arbitrage, factor structure and mean-variance analysis in large asset markets", *Econometrica* 51 (1983), 1305-1324.

Forni, Mario, and Marco Lippi, *Aggregation and the Microfoundations of Dynamic Macroeconomics* (Oxford: Oxford University Press, 1997).

Forni, Mario, and Marco Lippi, "The generalized factor model: Representation theory", Working Paper no. 132, Université Libre de Bruxelles (1999).

Forni, Mario, and Lucrezia Reichlin, "Dynamic common factors in large cross-sections", *Empirical Economics* 21 (1996), 27-42.

Forni, Mario, and Lucrezia Reichlin, "National policies and local economies", CEPR Discussion paper series no. 1632, (1997a), forthcoming on *European Economic Review*.

Forni, Mario, and Lucrezia Reichlin, "Let's get real: a factor analytic approach to disaggregated business cycle dynamics", *Review of Economic Studies* 65 (1998), 453-473.

Geweke, John, "The dynamic factor analysis of economic time series", in Dennis J. Aigner and Arthur S. Goldberger (eds.) *Latent Variables in Socio-Economic Models* (Amsterdam: North-Holland, 1977).

Sargent, Thomas J. and Cristopher A. Sims, "Business cycle modelling without pretending to have too much *a priori* economic theory" in Cristopher A. Sims (ed.) *New Methods in Business Research* (Minneapolis: Federal Reserve Bank of Minneapolis, 1977).

Stock, James H., and Mark W. Watson, "New indexes of coincident and leading economic indicators", *NBER Macroeconomic Annual* (1989), 351-94.

Stock, James H., and Mark W. Watson, "Diffusion indexes", manuscript (July, 1998).

APPENDIX

PROOF OF PROPOSITION 1. We need the following result (see Brillinger, 1981, p. 84, Exercise 3.10.16):

Let $\mathbf{\Lambda}$ be an $n\times n$, complex, hermitian non-negative definite matrix, and let λ_k, $k = 1,\ldots,n$, be its (real) eigenvalues in descending order of magnitude. By $\mathbf{D}_k$ we denote an $n \times k-1$ complex matrix for $1 < k \leq n$, the $n \times 1$ null matrix for $k = 1$. The eigenvalue λ_k is the solution of

$$\begin{aligned} &\min_{\mathbf{D}_k} \max_{\mathbf{b}} \ \mathbf{b}\mathbf{\Lambda}\tilde{\mathbf{b}} \\ &s.t.\ |\mathbf{b}| = 1,\ \mathbf{b}\tilde{\mathbf{D}}_k = 0. \end{aligned} \tag{10}$$

Note that for $k = 1$ the only constraint is $|\mathbf{b}| = 1$.

Since $\mathbf{\Sigma}_n = \mathbf{\Sigma}_n^\chi + \mathbf{\Sigma}_n^\xi$, given $\mathbf{D}_j$, for $\mathbf{b}\mathbf{D}_j = 0$ and $|\mathbf{b}| = 1$, then

$$\max_{\mathbf{b}} \mathbf{b}\mathbf{\Sigma}_n(\theta)\tilde{\mathbf{b}} \geq \max_{\mathbf{b}} \mathbf{b}\mathbf{\Sigma}_n^\chi(\theta)\tilde{\mathbf{b}},$$

$$\max_{\mathbf{b}} \mathbf{b}\mathbf{\Sigma}_n(\theta)\tilde{\mathbf{b}} \leq \max_{\mathbf{b}} \mathbf{b}\mathbf{\Sigma}_n^\chi(\theta)\tilde{\mathbf{b}} + \max_{\mathbf{b}} \mathbf{b}\mathbf{\Sigma}_n^\xi(\theta)\tilde{\mathbf{b}} \leq \max_{\mathbf{b}} \mathbf{b}\mathbf{\Sigma}_n^\chi(\theta)\tilde{\mathbf{b}} + \lambda_{n1}^\xi(\theta).$$

¿From (10) and the above inequalities we get

$$\text{(a) } \lambda_{nj}(\theta) \leq \lambda_{nj}^\chi(\theta) + \lambda_{n1}^\xi(\theta), \qquad \text{(b) } \lambda_{nj}(\theta) \geq \lambda_{nj}^\chi(\theta).$$

The statement on the first q eigenvalues of $\mathbf{\Sigma}_n$ follows from (b). The statement on the $(q+1)$-th follows from (a) and the fact that the $(q+1)$-th eigenvalue of $\mathbf{\Sigma}_n^\chi$ vanishes at any frequency. QED

To prove Proposition 2 we need some intermediate results. We suppose that Assumptions 1 through 4 hold.

LEMMA 1. *In Section 3.1 we have defined $p_{nj,i}(\theta)$ as the i-th component of $\mathbf{p}_{nj}(\theta)$. For $j \leq q$, $\lim_{n\to\infty} |p_{nj,i}(\theta)| = 0$ for θ a.e. in $[-\pi,\ \pi]$.*

PROOF. Let $\mathbf{P}_n$ be the $n \times n$ matrix having the eigenvectors $\mathbf{p}_{nj}$ on the rows. From $\tilde{\mathbf{P}}_n \mathrm{diag}\,(\,\lambda_{n1} \quad \lambda_{n2} \quad \cdots \quad \lambda_{nn}\,)\,\mathbf{P}_n = \mathbf{\Sigma}_n$ one obtains

$$\sum_{j=1}^{q} |p_{nj,i}(\theta)|^2 \lambda_{nj}(\theta) + \sum_{j=q+1}^{n} |p_{nj,i}(\theta)|^2 \lambda_{nj}(\theta) = \sigma_i(\theta),$$

where σ_i is the spectral density of x_{it}. By Proposition 1 $\lambda_{nj}(\theta)$ diverges for θ a.e. in $[-\pi,\ \pi]$ for $j \leq q$. But σ_i is finite a.e. in $[-\pi,\ \pi]$. QED

LEMMA 2. *For a given i and $n \in \mathbb{N}$ consider the n-dimensional filters (defined in Section 3.1)*

$$\underline{\mathbf{K}}_{ni}(L) = \sum_{j=1}^{q} \underline{\tilde{p}}_{nj,i}(L)\underline{\mathbf{p}}_{nj}(L).$$

Then

$$\lim_{n\to\infty}\int_{-\pi}^{\pi}|\mathbf{K}_{ni}(\theta)|^2 d\theta = 0,$$

where $|\mathbf{K}_{ni}(\theta)|^2 = \mathbf{K}_{ni}(\theta)\tilde{\mathbf{K}}_{ni}(\theta)$.

PROOF. We have

$$|\mathbf{K}_{ni}(\theta)|^2 = \sum_{j=1}^{q}|p_{nj,i}(\theta)|^2 \leq 1.$$

Moreover, by Lemma 1, $|\mathbf{K}_{ni}(\theta)|^2$ tends to zero a.e. in $[-\pi,\ \pi]$. The result follows applying the Lebesgue dominated convergence theorem (see Apostol, 1974, p. 270). QED

LEMMA 3. *For* $n \in \mathbb{N}$ *let* $\underline{\mathbf{a}}_n(L)$ *be an n-dimensional two-sided square-summable filter. Assume that*

$$\lim_{n\to\infty}\int_{-\pi}^{\pi}|\mathbf{a}_n(\theta)|^2 d\theta = 0.$$

Then

$$\lim_{n\to\infty}\underline{\mathbf{a}}_n(L)\boldsymbol{\xi}_{nt} = 0$$

in mean square.

PROOF. By the result recalled in the proof of Proposition 1,

$$\text{var}(\underline{\mathbf{a}}_n(L)\boldsymbol{\xi}_{nt}) = \int_{-\pi}^{\pi}\mathbf{a}_n(\theta)\boldsymbol{\Sigma}_n^{\xi}(\theta)\tilde{\mathbf{a}}_n(\theta)d\theta \leq \int_{-\pi}^{\pi}\lambda_{n1}^{\xi}(\theta)|\mathbf{a}_n(\theta)|^2 d\theta.$$

The result follows from Assumption 3. QED

Denoting by , $\mathcal{A}_{ni}(\theta)$ the spectral density of $\underline{\mathbf{K}}_{ni}(L)\boldsymbol{\xi}_{nt}$, we have

$$\mathcal{A}_{ni}(\theta) = \mathbf{K}_{ni}(\theta)\boldsymbol{\Sigma}_n^{\xi}(\theta)\tilde{\mathbf{K}}_{ni}(\theta) \leq \lambda_{n1}^{\xi}(\theta)|\mathbf{K}_{ni}(\theta)|^2.$$

Thus Lemma 1 and Assumption 3 imply that $\mathcal{A}_{ni}(\theta)$ converges to zero a.e. in $[-\pi,\ \pi]$. Lemma 2 and Lemma 3 imply that

$$\lim_{n\to\infty}\underline{\mathbf{K}}_{ni}(L)\boldsymbol{\xi}_{nt} = 0$$

in mean square.

With no loss of generality we can assume that

ASSUMPTION A. $\lambda_{nj}^{x}(\theta) \geq 1$ for any j, n, and $\theta \in [-\pi,\ \pi]$.

Indeed, possibly by embedding $L_2(\Omega,\mathcal{F},P)$ into a larger space, we can assume that $L_2(\Omega,\mathcal{F},P)$ contains a double sequence $\{\phi_{it},\ i \in \mathbb{N},\ t \in \mathbb{Z}\}$ such that, firstly, ϕ_{it} is orthogonal to the u's and the ξ's at any lead and lag, and, secondly, $\boldsymbol{\phi}_n = \{\boldsymbol{\phi}_{nt},\ t \in \mathbb{Z}\}$, where

$$\boldsymbol{\phi}_{nt} = (\,\phi_{1t}\quad \phi_{2t}\quad \cdots\quad \phi_{nt}\,)',$$

is an orthonormal white noise. Defining $\hat{\xi}_{it} = \xi_{it} + \phi_{it}$, and

$$y_{it} = \chi_{it} + \hat{\xi}_{it}, \tag{11}$$

for $i \in \mathbb{N}$ and $t \in \mathbb{Z}$, we have:

(1) Model (11) fulfills Assumptions 1 through 4, with

$$\mathbf{\Sigma}_n^{\hat{\xi}}(\theta) = \mathbf{\Sigma}_n^{\xi}(\theta) + \mathbf{I}_n, \qquad \mathbf{\Sigma}_n^{y}(\theta) = \mathbf{\Sigma}_n(\theta) + \mathbf{I}_n,$$

and therefore $\lambda_{nj}^{\hat{\xi}}(\theta) = \lambda_{nj}^{\xi}(\theta) + 1$, $\lambda_{nj}^{y}(\theta) = \lambda_{nj}(\theta) + 1$. Moreover, $\mathbf{p}_{nj}^{y} = \mathbf{p}_{nj}$ for any n and j, so that $\mathbf{K}_{ni}^{y} = \mathbf{K}_{ni}$ for any n and i.

(2) As a consequence, if we prove Proposition 2 for the y's, i.e.

$$\lim_{n\to\infty} \underline{\mathbf{K}}_{ni}(L)\mathbf{y}_{nt} = \chi_{it},$$

then the desired result

$$\lim_{n\to\infty} \underline{\mathbf{K}}_{ni}(L)\mathbf{x}_{nt} = \chi_{it}$$

follows, since $\lim_n \underline{\mathbf{K}}_{ni}^{y}(L)\boldsymbol{\phi}_{nt} = 0$ by Lemma 3.

Under Assumption A, the function $\mu_{nj}(\theta) = [\lambda_{nj}(\theta)]^{-1/2}$ is defined for any $\theta \in [-\pi,\ \pi]$, is bounded and therefore has a mean-square convergent Fourier representation. Let us denote by $\underline{\mu}_{nj}(L)$ the corresponding square-summable filter. Now consider the vector of the first q *normalized* dynamic principal components:

$$\mathbf{W}_{nt} = (\, W_{n1,t} \quad W_{n2,t} \quad \cdots \quad W_{nq,t}\,),$$

where $W_{nj,t} = \underline{\mu}_{nj}(L)\underline{\mathbf{p}}_{nj}(L)\mathbf{x}_{nt}$. The vector process $\{\mathbf{W}_{nt},\ t \in \mathbb{Z}\}$ is an orthonormal q-dimensional white noise.

LEMMA 4. *Consider the orthogonal projection of* $\mathbf{W}_{nt}$ *on* $\mathcal{U} = \overline{\text{span}}(u_{jt},\ j = 1, \ldots, q,\ t \in \mathbb{Z})$*:*

$$(\, W_{n1,t} \quad W_{n2,t} \quad \cdots \quad W_{nq,t}\,)' = \mathbf{A}_n(L)\,(\, u_{1t} \quad u_{2t} \quad \cdots \quad u_{qt}\,)' + \mathbf{R}_{nt}, \tag{12}$$

where $\mathbf{A}_n(L)$ *is an* $n \times n$ *two-sided square-summable filter and* $\mathbf{R}_{nt}$ *is orthogonal to* $\mathcal{U}$*. Then: (A) the spectral density of* $\mathbf{R}_{nt}$ *converges to zero a.e. in* $[-\pi,\ \pi]$*; (B)* $\mathbf{R}_{nt}$ *converges to zero in mean square; (C) considering the projection of* $\mathbf{u}_t$ *on the space spanned by the leads and lags of* $\mathbf{W}_{nt}$*,*

$$(\, u_{1t} \quad u_{2t} \quad \cdots \quad u_{qt}\,)' = \tilde{\mathbf{A}}_n(F)\,(\, W_{n1,t} \quad W_{n2,t} \quad \cdots \quad W_{nq,t}\,)' + \mathbf{S}_{nt}, \tag{13}$$

the spectral density of $\mathbf{S}_{nt}$ converges to zero a.e. in $[-\pi,\ \pi]$ and $\mathbf{S}_{nt}$ converges to zero in mean square.

Proof. Firstly observe that

$$W_{nj,t} = \underline{\mu}_{nj}(L)\underline{\mathbf{p}}_{nj}(L)\boldsymbol{\chi}_{nt} + \underline{\mu}_{nj}(L)\underline{\mathbf{p}}_{nj}(L)\boldsymbol{\xi}_{nt}.$$

Since the χ's belong to $\mathcal{U}$ and the ξ's are orthogonal to $\mathcal{U}$, $\underline{\mu}_{nj}(L)\underline{\mathbf{p}}_{nj}(L)\boldsymbol{\xi}_{nt}$ is the residual of the orthogonal projection of $W_{nj,t}$ on $\mathcal{U}$. By Assumption 3, the spectral density of $\underline{\mu}_{nj}(L)\underline{\mathbf{p}}_{nj}(L)\boldsymbol{\xi}_{nt}$ satisfies

$$f_{nj}(\theta) \leq \mu_{nj}(\theta)^2|\mathbf{p}_{nj}(\theta)|^2\Lambda = \lambda_{nj}(\theta)^{-1}\Lambda.$$

By Assumption 4, $f_{nj}(\theta)$ converges to zero a.e. in $[-\pi,\ \pi]$. Moreover, by Assumption A, $f_{nj}(\theta) \leq \Lambda$, so that the Lebesgue dominated convergence theorem applies and $\int_{-\pi}^{\pi} f_{nj}(\theta)d\theta$ converges to zero. Thus (A) and (B) are proved. To prove (C), from (12) and (13) we obtain

$$\mathbf{I}_q = \mathbf{A}_n(e^{-i\theta})\tilde{\mathbf{A}}_n(e^{i\theta}) + \boldsymbol{\Sigma}_n^R(\theta) = \tilde{\mathbf{A}}_n(e^{i\theta})\mathbf{A}_n(e^{-i\theta}) + \boldsymbol{\Sigma}_n^S(\theta),$$

where $\boldsymbol{\Sigma}_n^R(\theta)$ and $\boldsymbol{\Sigma}_n^S(\theta)$ are the spectral density matrices of $\mathbf{R}_{nt}$ and $\mathbf{S}_{nt}$ respectively. By taking the trace on both sides and noting that the trace of $\mathbf{A}_n(e^{i\theta})\tilde{\mathbf{A}}_n(e^{-i\theta})$ is equal to the trace of $\tilde{\mathbf{A}}_n(e^{-i\theta})\mathbf{A}_n(e^{i\theta})$ we get

$$\text{trace}\left(\boldsymbol{\Sigma}_n^S(\theta)\right) = \text{trace}\left(\boldsymbol{\Sigma}_n^R(\theta)\right).$$

The result follows. QED

Note that Lemma 4 proves that the space spanned by the normalized dynamic principal components, identical to the space $\mathcal{U}_n$ spanned by the dynamic principal components themselves, converges to $\mathcal{U}$, not that $\mathbf{W}_{nt}$ converges to any particular orthonormal white noise in $\mathcal{U}$. Indeed, it is easy to provide examples in which the variables $W_{nj,t}$, though converging to $\mathcal{U}$, do not converge to any vector of $\mathcal{U}$. What is stated in Proposition 2 is that the projection of x_{it} on $\mathcal{U}_n$, i.e. $\chi_{it,n}$, converges and that the limit is χ_{it}.

Proof of Proposition 2. We have

$$x_{it} = \chi_{it} + \xi_{it} = \chi_{it,n} + \xi_{it,n}, \tag{14}$$

and

$$\chi_{it,n} = \underline{\mathbf{K}}_{ni}(L)\mathbf{x}_{nt} = \underline{\mathbf{K}}_{ni}(L)\boldsymbol{\chi}_{nt} + \underline{\mathbf{K}}_{ni}(L)\boldsymbol{\xi}_{nt}. \tag{15}$$

Combining (14) and (15) we obtain

$$[\chi_{it} - \underline{\mathbf{K}}_{ni}(L)\boldsymbol{\chi}_{nt}] + [\xi_{it} - \xi_{it,n}] = \underline{\mathbf{K}}_{ni}(L)\boldsymbol{\xi}_{nt}. \tag{16}$$

Consider the spectral density of the right hand side of (16), which has been denoted by $\mathcal{A}_{ni}(\theta)$ (see the comment under Lemma 3). Since ξ_{it} is orthogonal to the χ's at all leads and lags,

$$\mathcal{A}_{ni}(\theta) = \mathcal{B}_{ni}(\theta) + \mathcal{C}_{ni}(\theta) - 2\Re\mathcal{D}_{ni}(\theta),$$

($\Re$ denoting the real part of a complex number) where $\mathcal{B}_{ni}(\theta)$ is the spectral density of $\chi_{it} - \underline{\mathbf{K}}_{ni}(L)\boldsymbol{\chi}_{nt}$, $\mathcal{C}_{ni}(\theta)$ is the spectral density of $\xi_{it} - \xi_{it,n}$, and $\mathcal{D}_{ni}(\theta)$ is the cross spectrum between $\xi_{it,n}$ and $\chi_{it} - \underline{\mathbf{K}}_{ni}(L)\boldsymbol{\chi}_{nt}$.

In the comment under Lemma 3 we have shown that $\mathcal{A}_{ni}(\theta)$, converges to zero a.e. in $[-\pi,\ \pi]$. If we show that $\mathcal{D}_{ni}(\theta)$ converges to zero a.e. in $[-\pi,\ \pi]$, then both $\mathcal{B}_{ni}(\theta)$ and $\mathcal{C}_{ni}(\theta)$ converge to zero a.e. in $[-\pi,\ \pi]$, and since both are obviously dominated by integrable functions, by the Lebesgue dominated convergence theorem the integrals of $\mathcal{B}_{ni}(\theta)$ and $\mathcal{C}_{ni}(\theta)$ converge to zero and the result is obtained.

Thus we must show that the cross spectrum between $\xi_{it,n}$ and $\chi_{it} - \underline{\mathbf{K}}_{ni}(L)\boldsymbol{\chi}_{nt}$ converges to zero a.e. in $[-\pi,\ \pi]$. Consider firstly the cross spectrum between $\xi_{it,n}$ and χ_{it}. Setting $\mathbf{b}_i(L) = (\, b_{i1}(L) \quad b_{i2}(L) \quad \cdots b_{iq}(L)\,)$, and using (13), we have

$$\chi_{it} = \mathbf{b}_i(L)\,(\, u_{1t} \quad u_{2t} \quad \cdots \quad u_{qt}\,)' = \mathbf{b}_i(L)\tilde{\mathbf{A}}_n(L^{-1})\,(\, W_{n1,t} \quad W_{n2,t} \quad \cdots \quad W_{nq,t}\,)' \\ + \mathbf{b}_i(L)\mathbf{S}_{nt}.$$

Since $\xi_{it,n}$ is orthogonal to the terms $\underline{\mathbf{p}}_{nj}(L)\mathbf{x}_{nt}$, for $j = 1,\ \ldots,\ q$, at any lead and lag, it is also orthogonal at any lead and lag to the terms $W_{nj,t}$. Thus the cross spectrum between $\xi_{it,n}$ and χ_{it} is equal to the cross spectrum between $\xi_{it,n}$ and $\mathbf{b}_i(L)\mathbf{S}_{nt}$, call it $\mathcal{E}_{ni}(\theta)$. The squared modulus of $\mathcal{E}_{ni}(\theta)$ is bounded by the product of the spectral density of $\xi_{it,n}$, which is dominated by the spectral density of x_{it}, and the spectral density of $\mathbf{b}_i(L)\mathbf{S}_{nt}$, i.e., by

$$\mathbf{b}_i(e^{-i\theta})\boldsymbol{\Sigma}_n^S(\theta)\tilde{\mathbf{b}}_i(e^{i\theta}).$$

By Lemma 4, all the entries of $\boldsymbol{\Sigma}_n^S(\theta)$ tend to zero a.e. in $[-\pi,\ \pi]$, so that $\mathcal{E}_n(\theta)$ tends to zero a.e. in $[-\pi,\ \pi]$.

Using the same argument, considering the cross spectrum between $\xi_{it,n}$ and $\underline{\mathbf{K}}_{ni}(L)\boldsymbol{\chi}_{nt}$, we end up with the cross spectrum between $\xi_{it,n}$ and $\underline{\mathbf{K}}_{ni}(L)\mathbf{B}_n(L)\mathbf{S}_{nt}$, where $\mathbf{B}_n(L)$ is the $n \times q$ matrix having the vectors $\mathbf{b}_s(L)$, $s = 1, \ldots, n$, on the rows. As for the spectral density of $\underline{\mathbf{K}}_{ni}(L)\mathbf{B}_n(L)\mathbf{S}_{nt}$, first observe that, since $\mathbf{\Sigma}_n^{\chi}(\theta) = \mathbf{B}_n(e^{-i\theta})\tilde{\mathbf{B}}_n(e^{i\theta})$ and $\mathbf{\Sigma}_n(\theta) = \mathbf{\Sigma}_n^{\chi}(\theta) + \mathbf{\Sigma}_n^{\xi}(\theta)$,

$$\begin{aligned}\mathbf{K}_{ni}(\theta)\mathbf{B}_n(e^{-i\theta})\tilde{\mathbf{B}}_n(e^{i\theta})\tilde{\mathbf{K}}_{ni}(\theta) &= \mathbf{K}_{ni}(\theta)\mathbf{\Sigma}_n^{\chi}(\theta)\tilde{\mathbf{K}}_{ni}(\theta) \leq \mathbf{K}_{ni}(\theta)\mathbf{\Sigma}_n(\theta)\tilde{\mathbf{K}}_{ni}(\theta)\\ &= \sum_{j=1}^{q} |p_{nj,i}(\theta)|^2 \lambda_{nj}(\theta),\end{aligned}$$

which is bounded by the spectral density of x_{it} (see Lemma 1). Next, observe that the maximum eigenvalue of $\mathbf{\Sigma}_n^S(\theta)$, which is a continuous function of the entries, tends to zero a.e. in $[-\pi,\ \pi]$. The result follows from

$$\mathbf{K}_{ni}(\theta)\mathbf{B}_n(e^{-i\theta})\mathbf{\Sigma}_n^S(\theta)\tilde{\mathbf{B}}_n(e^{i\theta})\tilde{\mathbf{K}}_{ni}(\theta) \leq \lambda_{n1}^S(\theta)\mathbf{K}_{ni}(\theta)\mathbf{B}_n(e^{-i\theta})\tilde{\mathbf{B}}_n(e^{i\theta})\tilde{\mathbf{K}}_{ni}(\theta).$$

QED

Proof of Corollaries 1 and 2. Corollary 1 is trivial. For Corollary 2, suppose that there are q factors but we project on the first $q+s$ dynamic principal components. Then

$$\chi_{it,n}^* - \chi_{it,n} = \underline{\tilde{p}}_{n\,q+1,i}(L)\underline{\mathbf{p}}_{n\,q+1}(L)\mathbf{x}_{nt} + \cdots + \underline{\tilde{p}}_{n\,q+s,i}(L)\underline{\mathbf{p}}_{n\,q+s}(L)\mathbf{x}_{nt}.$$

Since different dynamic principal components are orthogonal at any lead and lag,

$$\sum_{i=1}^{n} \mathrm{var}(\chi_{it,n}^* - \chi_{it,n}) \leq \int_{-\pi}^{\pi} \lambda_{n\,q+1}(\theta)d\theta + \cdots + \int_{-\pi}^{\pi} \lambda_{n\,q+s}(\theta)d\theta.$$

The result follows from Assumption 3. QED

Proof of Proposition 3. For fixed T (and n), the T random variables $(\underline{\mathbf{K}}_{ni}^{Tt}(L) - \underline{\mathbf{K}}_{ni}(L))\mathbf{x}_{nt},\ t = 1, \ldots, T$, are not identically distributed, due to two reasons:

(a) The truncation of the filters, which depends on t;

(b) The *boundary effects*, which imply that the joint distributions of $(\underline{\mathbf{K}}_{ni}^{T}(L), \mathbf{x}_{n1})$ and $(\underline{\mathbf{K}}_{ni}^{T}(L), \mathbf{x}_{nT})$ are not the same as, e.g., that of $(\underline{\mathbf{K}}_{ni}^{T}(L), \mathbf{x}_{n\,T/2})$.

However, as the filters $\underline{\mathbf{K}}_{ni}^{T}(L)$ and $\underline{\mathbf{K}}_{ni}(L)$ are both square-summable and $\{\mathbf{x}_{nt},\ t \in \mathbb{Z}\}$ is a stationary process, both effects (a) and (b) are asymptotically nil for *central* values of t. Precisely, given $\eta > 0$, there exists a $T_2 = T_2(n, \eta)$ such that, for $T \geq T_2$ and t such that

$$aT \leq t \leq bT,$$

with $0 < a < b < 1$, so that (7) is automatically fulfilled, we have

$$\begin{aligned}\Bigg| \mathrm{E}\left[|(\underline{\mathbf{K}}_{ni}^{Tt}(L) - \underline{\mathbf{K}}_{ni}(L))\mathbf{x}_{nt}| \;\middle|\; \sup_{\theta\in[-\pi,\pi]} |\mathbf{K}_{ni}^{T}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta \right] \\ - \mathrm{E}\left[|(\underline{\mathbf{K}}_{ni}^{T}(L) - \underline{\mathbf{K}}_{ni}(L))\mathbf{x}_{nt}| \;\middle|\; \sup_{\theta\in[-\pi,\pi]} |\mathbf{K}_{ni}^{T}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta \right] \Bigg| \leq \eta.\end{aligned}$$

It follows that, for $T \geq T_2(n, \frac{\epsilon^2\eta}{32})$,

$$\mathrm{E}\left[|(\underline{\mathbf{K}}_{ni}^{Tt}(L) - \underline{\mathbf{K}}_{ni}(L))\mathbf{x}_{nt}| \,\middle|\, \sup_{\theta\in[-\pi,\pi]} |\mathbf{K}_{ni}^{T}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta\right]$$
$$\leq \mathrm{E}\left[(bT-aT)^{-1}\sum_{t=aT}^{bT} |(\underline{\mathbf{K}}_{ni}^{T}(L) - \underline{\mathbf{K}}_{ni}(L))\mathbf{x}_{nt}| \,\middle|\, \sup_{\theta\in[-\pi,\pi]} |\mathbf{K}_{ni}^{T}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta\right]$$
$$+ \frac{\epsilon^2\eta}{32},$$

where $\sum_{t=aT}^{bT}$ stands for a sum over t running from the smallest integer $\lfloor aT \rfloor$ larger than or equal to aT to the largest integer $\lceil bT \rceil$ smaller than or equal to bT (a window of width $(\lceil bT \rceil - \lfloor aT \rfloor)$. For simplicity, we write $(bT-aT)^{-1}$ for $(\lceil bT \rceil - \lfloor aT \rfloor)^{-1}$. Hence,

$$R_{n1}^{Tt} \leq \frac{4}{\epsilon^2}\mathrm{E}\left[\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}(\underline{\mathbf{K}}_{ni,k}^{T} - \underline{\mathbf{K}}_{ni,k})(bT-aT)^{-1}\right.$$
$$\left.\sum_{t=aT}^{bT}\mathbf{x}_{n\,t-k}\tilde{\mathbf{x}}_{n\,t-l}(\tilde{\underline{\mathbf{K}}}_{ni,l}^{T} - \tilde{\underline{\mathbf{K}}}_{ni,l}) \,\middle|\, \sup_{\theta\in[-\pi,\pi]} |\mathbf{K}_{ni}^{T}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta\right]$$
$$+ \frac{\eta}{8} + \frac{\eta}{4}$$
$$\leq \frac{4\delta^2}{\epsilon^2}\mathrm{E}\left[\int_{-\pi}^{\pi}\overline{\lambda}_{n1}^{T}(\theta)d\theta\right] + \frac{3\eta}{8},$$

where $\overline{\lambda}_{n1}^{T}(\theta)$ denotes the first dynamic eigenvalue associated with the *pseudo-empirical* cross-covariance function $\overline{\mathbf{\Gamma}}_{kl}^{nT} = (bT-aT)^{-1}\sum_{t=aT}^{bT}(\mathbf{x}_{n\,t-k}\tilde{\mathbf{x}}_{n\,t-l})$. We are using "pseudo" because, due to the substitution of the infinite-order bilateral filter $\mathbf{K}_{ni}^{T}(L)$ for the truncated, hence finite-order ones $\mathbf{K}_{ni}^{Tt}(L)$, the sums defining $\overline{\mathbf{\Gamma}}_{kl}^{nT}$ all contain the same number $bT - aT$ of terms, and are not truncated, as in the usual definition of empirical cross-covariances, according to the finite sample length.

Since all covariances are "estimated" using the same number of "observations", then the usual lag-window argument leading from Assumption 2′ to (5) may be invoked (see, e.g., Brockwell and Davis, 1997, Section 10.4). Thus the spectral density $\overline{\mathbf{\Sigma}}_{n}^{T}(\theta)$, corresponding to the covariances $\overline{\mathbf{\Gamma}}_{k,l}^{nT}$, converges in probability to $\mathbf{\Sigma}_n(\theta)$, uniformly in $[-\pi,\ \pi]$, as $T \to \infty$. As a consequence, since eigenvalues are continuous functions of the entries of the corresponding spectral density matrix, $\int_{-\pi}^{\pi}\overline{\lambda}_{n1}^{T}(\theta)d\theta$ converges in probability, as $T \to \infty$, to $\int_{-\pi}^{\pi}\lambda_{n1}(\theta)d\theta$. Moreover, the sequence $\int_{-\pi}^{\pi}\overline{\lambda}_{n1}^{T}(\theta)d\theta$ is bounded by $\int_{-\pi}^{\pi}\mathrm{trace}[\overline{\mathbf{\Sigma}}_{n}^{T}(\theta)]d\theta = \mathrm{trace}[\overline{\mathbf{\Gamma}}_{00,n}^{T}] = (bT-aT)^{-1}\sum_{t=aT}^{bT}\sum_{i=1}^{n}x_{it}^2$. This latter quantity, by Assumption 2′, has uniformly bounded moment of order two. It follows (cf. Billingsley, 1995, p. 338) that $\int_{-\pi}^{\pi}\overline{\lambda}_{n1}^{T}(\theta)d\theta$ is uniformly integrable, as $T \to \infty$, so that

$$\lim_{T\to\infty} \mathrm{E}\left[\int_{-\pi}^{\pi} \overline{\lambda}_{n1}^{T}(\theta)d\theta\right] = \mathrm{E}\left[\int_{-\pi}^{\pi} \lambda_{n1}(\theta)d\theta\right] = \int_{-\pi}^{\pi} \lambda_{n1}(\theta)d\theta.$$

Thus, there exists some $T_3(n)$ such that, for all $T \geq T_3$,

$$\mathrm{E}\left[\int_{-\pi}^{\pi} \overline{\lambda}_{n1}^{T}(\theta)d\theta\right] \leq 2\int_{-\pi}^{\pi} \lambda_{n1}(\theta)d\theta.$$

Summing up, for $n \geq N_0(\epsilon, \eta)$ and $T \geq T_0$, with

$$T_0 = T_0(n, \epsilon, \eta) = \max\bigl(T_1(n, \delta, \frac{\eta}{4}), T_2(n, \frac{\epsilon^2\eta}{32}), T_3(n)\bigr)$$

where $\delta^2 = \frac{\epsilon^2\eta}{128\int_{-\pi}^{\pi}\lambda_{n1}(\theta)d\theta}$, we have proved that $R_{1t}^{nt} \leq \frac{\eta}{2}$. The proposition follows. QED

Footnotes

[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 Christine De Mol for generously giving us her time, Serena Ng for pointing out programming problems, Jim Stock for helpful advise and Jorge Rodrigues for very valuable research assistantship.

[2] We use the term 'dynamic eigenvalues' to insist on the difference between the functions λ and the eigenvalues of the variance-covariance matrix employed in the static principal component analysis. A standard reference for eigenvalues and eigenvectors of the spectral density matrix is Brillinger (1981), Chapter 9.

[3] On the identification and estimation of the common shocks in a related model, see Forni and Reichlin (1998).

[4] Note that, for $M = 0$, $\mathbf{p}_{nj}^{T}(\theta_0)$ is simply the j-th eigenvector of the (estimated) variance-covariance matrix of $\mathbf{x}_{nt}$: the dynamic principal components then reduce to the static principal components.

Table 5.1. Average and standard deviation (in brackets) of $R(\hat{\chi}, \chi)$ across 400 experiments

	$T = 20$	$T = 50$	$T = 100$	$T = 200$
Model M1				
$n = 10$	0.554 (0.281)	0.394 (0.201)	0.343 (0.162)	0.296 (0.131)
$n = 20$	0.372 (0.174)	0.244 (0.091)	0.194 (0.068)	0.162 (0.045)
$n = 50$	0.261 (0.098)	0.150 (0.035)	0.109 (0.024)	0.081 (0.014)
$n = 100$	0.227 (0.069)	0.123 (0.024)	0.084 (0.014)	0.059 (0.008)
$R(\bar{\chi}, \chi)$ with $n = 100$	0.197 (0.105)	0.061 (0.012)	0.030 (0.005)	0.015 (0.002)
Model M2				
$n = 10$	0.671 (0.351)	0.472 (0.206)	0.382 (0.187)	0.317 (0.138)
$n = 20$	0.505 (0.202)	0.295 (0.100)	0.070 (0.081)	0.047 (0.063)
$n = 50$	0.390 (0.117)	0.048 (0.085)	0.026 (0.049)	0.016 (0.032)
$n = 100$	0.353 (0.098)	0.032 (0.071)	0.016 (0.041)	0.009 (0.027)
$R(\bar{\chi}, \chi)$ with $n = 100$	0.366 (0.113)	0.105 (0.019)	0.052 (0.008)	0.025 (0.004)
Model M3				
$n = 10$	0.633 (0.255)	0.436 (0.152)	0.340 (0.108)	0.294 (0.081)
$n = 20$	0.479 (0.160)	0.289 (0.084)	0.211 (0.051)	0.161 (0.029)
$n = 50$	0.384 (0.106)	0.193 (0.039)	0.128 (0.022)	0.092 (0.013)
$n = 100$	0.344 (0.084)	0.163 (0.029)	0.103 (0.014)	0.067 (0.007)
$R(\bar{\chi}, \chi)$ with $n = 100$	0.366 (0.112)	0.103 (0.019)	0.051 (0.007)	0.025 (0.003)
Model M4				
$n = 10$	0.642 (0.360)	0.433 (0.213)	0.352 (0.170)	0.299 (0.132)
$n = 20$	0.459 (0.187)	0.278 (0.093)	0.201 (0.060)	0.166 (0.047)
$n = 50$	0.342 (0.100)	0.193 (0.039)	0.131 (0.025)	0.095 (0.015)
$n = 100$	0.322 (0.083)	0.167 (0.028)	0.108 (0.015)	0.073 (0.008)
$R(\bar{\chi}, \chi)$ with $n = 100$	0.424 (0.118)	0.199 (0.028)	0.114 (0.015)	0.067 (0.007)

Table 6.1. The data

Countries	GDP	Cons.	Inv.	CPI	Spread	Sent.	I.P.
Germany	X	X	X	X	X	X	X
France	X	X	X	X	X	X	X
Italy	X	X	X	X	X	X	X
Netherlands	X	X	X	X	X	X	X
Ireland	–	–	–	X	X	X	X
Spain	X	X	X	X	X	X	X
Finland	X	X	X	X	–	X	X
Austria	X	X	X	X	X	–	X
Belgium	X	–	–	X	X	X	X
Portugal	X	X	X	X	X	X	X

GDP: GDP, s.a., in national currency, at constant (1990) prices. Source: OECD; for Germany and Portugal: IMF.

Cons.: Private final consumption expenditure, s.a., in national currency at constant (1990) prices. Source: OECD; for Germany and Portugal: IMF.

Inv.: Gross fixed capital formation, s.a., in national currency at constant (1990) prices. Source: OECD; for Germany and Portugal: IMF.

CPI: Consumer Price Index, base year 1990.

Spread: Difference between the Government bond yield and the Treasury Bill rate (or the money market rate depending on data availability), in percent per year. Source: IMF

Sent.: Economic Sentiment Indicator. Source: European Commission, DG II.

I.P.: Industrial production, s.a., index number, base year 1990. Source: IMF.

Table 6.2. Percentage of variance explained by the common component

Countries	GDP	Cons.	Inv.	CPI	Spread	Sent.	I.P.
EURO aggregate	85	70	57	74	95	99	80
Germany	68	56	49	69	95	95	68
France	70	43	72	38	96	94	56
Italy	54	65	71	42	69	98	29
Netherlands	35	57	29	57	96	65	27
Ireland	–	–	–	39	51	87	22
Spain	96	73	96	37	34	68	62
Finland	46	–	–	65	–	90	46
Austria	55	–	–	–	99	–	53
Belgium	55	–	–	55	82	97	44
Portugal	54	43	54	69	63	69	–

Table 6.3. Average correlation of the common component with the common components of the other variables of the same country

Countries	GDP	Cons.	Inv.	CPI	Spread	Sent.	I.P.
EURO aggregate	0.58	0.36	0.55	-0.20	0.14	0.41	0.51
Germany	0.55	0.33	0.49	-0.22	0.01	0.34	0.50
France	0.66	0.47	0.68	0.17	0.39	0.56	0.58
Italy	0.49	0.48	0.49	0.09	-0.52	0.46	0.34
Netherlands	0.33	0.09	0.15	0.07	0.26	0.31	0.01
Ireland	–	–	–	-0.11	0.44	0.36	0.41
Spain	0.64	0.62	0.63	0.09	0.17	0.55	0.36
Finland	0.35	–	–	-0.21	–	0.40	0.04
Austria	0.47	–	–	–	0.22	–	0.58
Belgium	0.50	–	–	-0.08	0.30	0.31	0.43
Portugal	0.22	0.50	0.29	0.37	0.32	0.53	–