DISCUSSION PAPER SERIES

No. 3285

**FACTOR MODELS IN
LARGE CROSS-SECTIONS
OF TIME SERIES**

Lucrezia Reichlin

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP3285.asp**

ISSN 0265-8003

FACTOR MODELS IN LARGE CROSS-SECTIONS OF TIME SERIES

Lucrezia Reichlin, ECARES, Université Libre de Bruxelles and CEPR

Discussion Paper No. 3285
March 2002

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. 3285

March 2002

ABSTRACT

Factor Models in Large Cross-Sections of Time Series*

This Paper reviews recent econometric work on factor models in large cross-sections of time series. In this literature, traditional factor analysis is adapted to develop parsimonious estimation methods for high dimension time series models. The review covers problems of consistency and rates – as the dimension of the cross-section and the time dimension become large – identification and forecasting. We also review empirical applications on measuring and interpreting business cycles.

JEL Classification: C22, C23, E32 and E37
Keywords: business cycles, factor analysis and panel data

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

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

* This Paper is produced under the auspices of the 'Structure of the Euro Area Business Cycle' network, a joint project between CEPR and ECARES and supported by Banque Nationale de Belgique, Deutsche Bundesbank, Bank of England, Banco de España, Bank of Finland, Banque de France, Bank of Greece, Central Bank of Ireland, Banca d'Italia, Banque Centrale du Luxembourg, De Nederlandsche Bank, Oesterreichische Nationalbank, Banco de Portugal, the European Commission and the European Central Bank. We also thank the TMR on fluctuations for support.

Submitted 08 March 2002

1. Motivation*

Business cycles are characterized by two features: comovements and regular phases of expansion and depression. Comovements are observed between aggregate variables – output and inflation, for example – and between disaggregates – individual consumption and regional output, for example. The time series literature has typically analyzed these two characteristics in a separate way. Starting with the seminal contribution of Burns and Mitchell, 1946, a huge amount has been written on the "regularity" of cycles, asymmetries and non-linearities, on the basis of estimation of aggregate output or few relevant macroeconomic time series. A separate literature has addressed the issue of comovements, typically between few key aggregate time series and typically concentrating on long-run comovements (cointegration). Behind this literature, there is the implicit idea that the essential characteristics of the business cycle are captured by few relevant aggregate variables and that the information contained in disaggregate time series or in all the potentially available aggregate time series are not particularly useful to understand macroeconomic behavior. This is also the implicit idea behind VAR modeling where the propagation of "identified" aggregate shocks is analyzed in models typically containing a small number of variables.

On the other hand, there is a large number of econometric studies, which analyze the behavior of many consumers or many firms in order to understand the microeconomic mechanisms behind fluctuations. In these studies the cross-section is typically large and the time series dimension either absent or small. Economic theory is sufficiently heterogenous so as not to give us clear guidance on what is the level of aggregation relevant for macroeconomic questions and on what is the appropriate stochastic dimension for macroeconomic models. In the last fifteen years, the taste of the profession has been oriented towards the dynamic analysis of "small' models, but central banks and statistical institutes are still using macroeconometric models containing a large number of time series. Modern macroeconomic theory is based on the representative agent assumption, but macroeconomic empirics is mostly based on aggregate data. What is the cost of simplicity, i.e. are we losing valuable information by working with econometrics models containing few aggregate variables ? How detailed do our models have to be to have a chance to provide the essential information on

* This survey heavily draws from joint work with Mario Forni, Marc Hallin and Marco Lippi whom I thank. I also wish to thank my exceptional research assistants, Jeorges Rodrigues and Luca Sala.

the macroeconomy ?

To try to answer these questions there is a need to develop econometric models which (a) are able to handle the analysis of many time series by reducing the number of the essential parameters to estimate; (b) can provide an answer on what is the relevant stochastic dimension of a large economy, i.e. on how many aggregate shocks are needed to study the macroeconomy which emerges from the behavior of many agents ; (c) can help us identifying these (possibly few) shocks and studying the propagation mechanism through agents or through geographical space. This is what will help to bridge the gap between purely time series studies and the cross-sectional approach.

A natural starting point is the dynamic index (factor) model of Sargent and Sims (1977), Geweke (1977) and Geweke and Singleton (1981). In this framework, the dynamic of individual variables are represented as the sum of a component which is common to all variables in the economy and an orthogonal idiosyncratic residual. This approach will be briefly reviewed in Section 3. Where our survey really starts, however, is Section 4, where the recent literature which has developed the index approach is analyzed.

The new developments are in different directions. First of all, the model is adapted to the analysis of large cross sections (n large). This is partly a purely econometric development since it deals with the issue of finding consistent estimators to the common components of each variable in the panel as n and T become larger and investigates the required relative rates of n and T at which consistency is achieved. Partly, it is a development in identification and representation theory which is deeply related with fundamental macroeconomic questions. If there are comovements in the economy, few macroeconomic shocks should explain most of the variance of relevant variables and, if comovements characterize cyclical fluctuations, they should be observed at business cycle frequencies (corresponding to cycles of period between three and ten years, say). Comovements imply fewer shocks than variables and therefore – loosely speaking – a factor model type of structure. For a given n this implies conditions on the spectral density of the observations. Now, imagine that more and more information become available (larger n, i.e. more variables and/or more disaggregate information). If the few shocks are "pervasive", i.e. they remain common to all variables in a progressively larger panel, this implies conditions on the spectral density of the observations as n becomes large which can be exploited to define precisely the notion of common and idiosyncratic sources of dynamics. Finally, if the data support a structure in which many n variables are led by few q macro shocks, the information contained in the n variables should help identifying the q shocks providing insights for the use of generalized factor models for policy analysis.

2. Plan of the paper

This paper is not a survey on factor analysis. Rather, it is a survey of the relevant papers in the small but growing recent literature which has analyzed dynamic factor models in large cross sections. I will also discuss the relation between the latter and the static factor approach which has been developed for the study of financial data. The problems that will be covered are the identification and representation of the model in population, estimation and consistency rates, forecasting, identification of aggregate shocks and individual propagation mechanisms. Section 3 establishes the background by briefly reviewing classical dynamic factor analysis. Section 4 reviews the main results for factor models in large cross-sections. Section 5 outlines open research questions and Section 6 reports results from some empirical applications.

3. Index models for a given cross-section

The dynamic factor (index) model has been introduced in macroeconomics by Sargent and Sims (1977) and Geweke (1977). Other relevant papers in the early literature are Geweke and Singleton (1981) and Watson and Engle (1983).

The $n \times 1$ vector of stationary variables is represented as the sum of two orthogonal components:

$$\begin{aligned}\mathbf{X}_t &= \sum_{k=-\infty}^{\infty} \mathbf{A}_k \mathbf{f}_{t-k} + \boldsymbol{\xi}_t \\ &= \mathbf{A}(L)\mathbf{f}_t + \boldsymbol{\xi}_t\end{aligned} \tag{1}$$

where the $\mathbf{A}_k$'s are $n \times q$ matrices, $\mathbf{f}_t$ is a stationary stochastic process of dimension $q \times 1$ and diagonal variance covariance matrix, ξ_t is stationary of dimension $n \times 1$ with diagonal spectral density matrix (its elements are orthogonal at all leads and lags). Both processes are allowed to be temporally correlated.

We can rewrite the model so as to express the common component $\boldsymbol{\chi}_t = \mathbf{A}(L)\mathbf{f}_t$ in terms of uncorrelated common shocks. Consider a moving average representation of $\mathbf{f}_t$, $\mathbf{f}_t = \mathbf{D}(L)\mathbf{u}_t$; we can rewrite (1) as

$$\mathbf{X}_t = \mathbf{B}(L)\mathbf{u}_t + \boldsymbol{\xi}_t \tag{2}$$

where the common component, $\boldsymbol{\chi}_t = \mathbf{A}(L)\mathbf{D}(L)\mathbf{u}_t = \mathbf{B}(L)\mathbf{u}_t$, is expressed in terms of a q-dimension white noise (which we normalize so as to have unit variance).

The model implies that covariation among the observable variables X's is due entirely to the common effect of the q latent factors while variation of an individual variable X_{it} is due to the variation of the specific variable ξ_{it} as well as the variation and covariation of the common factor.

The restrictions on the covariance properties of the data are best understood by writing the spectral density matrix of the X's (see appendix A for an introduction to the spectral representation of stationary time series and the spectral density). Under the assumptions above, the spectral density of $\mathbf{X}_t$ can be written as:

$$\mathbf{\Sigma}_x(\theta) = \mathbf{B}(e^{-i\theta})\tilde{\mathbf{B}}(e^{-i\theta}) + \mathbf{\Sigma}_\xi(\theta), \quad \mid \theta \mid \leq \pi \tag{3}$$

where $\mathbf{B}(e^{-i\theta})$ denotes the Fourier transform of the function $\mathbf{B}(L)$ and $\tilde{\mathbf{B}}(e^{-i\theta})$ its conjugate transpose (from now on $\tilde{}$ will indicate conjugation and transposition). The model implies that the spectral density of the common shocks, $\mathbf{\Sigma}_u(\theta)$ is equal to the $q \times q$ identity matrix $\mathbf{I}_q$ and that the spectral density of $\boldsymbol{\xi}_t$, $\mathbf{\Sigma}_\xi(\theta)$, is diagonal.

The literature has analyzed three problems : identification of the component, identification of the coefficients, estimation.

(i) *Identification of the component.* Since the components are not observable, prior to estimation, identification conditions have to be spelled out. In the general case in which $n >> q$, model (2) is over-identified, the components $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ can be estimated by maximum likelihood techniques and over-identified restrictions tested by standard likelihood ratio tests.

(ii) *Estimation of the component.* Estimation is usually performed under a normality hypothesis and is based on maximum likelihood. There are two methods which have been used in classical factor analysis. The first, based on time domain analysis, is the EM algorithm (e.g. Watson and Engle, 1983 and Quah and Sargent, 1993). The second, is based on frequency domain analysis (e.g. Sargent and Sims, 1977 and Geweke and Singleton, 1981).

(iii) *Identification of the factors and the coefficients.* The common factors, $\mathbf{u}_t$ are only identified up to an orthonormal rotation of dimension q. Expression (3) is observationally equivalent to

$$\mathbf{\Sigma}_x(\theta) = \mathbf{B}(e^{-i\theta})\mathbf{Q}(e^{-i\theta})\tilde{\mathbf{Q}}(e^{-i\theta})\tilde{\mathbf{B}}(e^{-i\theta}) + \mathbf{\Sigma}_\xi(\theta)$$

where $\mathbf{Q}(e^{-i\theta})\tilde{\mathbf{Q}}(e^{-i\theta}) = \mathbf{I}_q$. The identification of the factors and the coefficients has been analyzed by Geweke and Singleton (1981) who provide sufficient conditions for identification. As noticed by these authors, unitary transformation of $\mathbf{\Sigma}_u(\theta)$ are possible because the common factors are not dated. That is, if, in (2), we replace $\mathbf{u}_{j,t-s}$ by $\mathbf{u}_{j,t-s+\tau_j}$, for all s and t and $j = 1, \ldots, q$, the modified model will be observationally equivalent to (2), since a pure delay in the time domain is equivalent to a phase shift (rotation around the unit circle) in the frequency domain.

4. Large economies : an infinite dimensional cross-section

When the number of time series is large (large n), traditional factor analysis based on maximum likelihood estimation involves computational problems since to the number of parameters increases with n. The research question explored by the papers reviewed here is whether one can consistently extract the indexes as $n \to \infty$ and how n must be related to T as $(n, T) \to \infty$. The analysis for n going to infinity is what we call the analysis of large economies. The case of large n is of great practical interest since many relevant business cycle questions involve the analysis of many variables for possibly many sectors, individuals or regions.

The analysis for $n \to \infty$ helps not only for estimation, but also for the identification of the model. Interesting specifications implies a factor structure where the idiosyncratic components are not mutually orthogonal, but allow for some mild cross-correlation. This is typically the case in asset pricing models (approximate factor structure), but also in macroeconomics when the object is to study the local effect of regional or sectoral specific shocks. For those specifications, if n is fixed, identification restrictions are not easily found. As we will see, they will come quite naturally when we analyze the model in population as $n \to \infty$.

Another point is that the analysis for $n \to \infty$ suggests a characterization of macroeconomic behavior. If, as n increases, $q < n$ sources of variations remain common to all variables, we can say that the economy is driven by q macroeconomic shocks and that this is the relevant stochastic dimension of macroeconomic models.

¿From now on we will think of the X's as an infinite dimensional sequence indexed by n and study the properties of the model as n and T go to infinity.

Denote by $\{X_{it};\ t \in \mathbb{Z}\}$ a double array of random variables, assume that the n one dimensional series $\{X_{it};\ t \in \mathbb{Z}\}$, $\quad i = 1, \ldots, n$ have been observed over the period $t = 1, \ldots, T$, and write $\mathbf{X}_t^{(n)} = (X_{1t}, \ldots, X_{nt})'$ for the observation made at time t. The object of the study is the model :

$$X_{it} = \chi_{it} + \xi_{it}, \qquad t \in \mathbb{Z} \tag{4}$$

where χ_{it}, the *common component*, can be represented as a dynamic linear combination

$$\chi_{it} = \sum_{j=1}^{q} b_{ij}(L)u_{jt}, \qquad t \in \mathbb{Z}, \qquad i \in= 1, \cdots, n$$

of (a small number q of) unobservable *common shocks* $u_{1t}, \ldots, u_{qt}$, and ξ_{it}, the *idiosyncratic component*. The latter is orthogonal, at all leads and lags, to those common shocks, hence to the dynamic space spanned by the common components.

We assume:

ASSUMPTION (B)

(B1) $\{\mathbf{u}_t = (u_{1t}, \ldots, u_{qt})\,; t \in \mathbb{Z}\}$ *is a zero-mean, normal white noise* ;

(B2) *the coefficients of the* nq *filters* $b_{ij}(L) = \sum_{k=-\infty}^{\infty} b_{ijk}L^k$ *are square-summable* : $\sum_{k=\infty}^{\infty} b_{ijk}^2 < \infty$*;*

(B3) $\left\{\boldsymbol{\xi}_t^{(n)} = (\xi_{1t}, \ldots, \xi_{nt})'\,; t \in \mathbb{Z}\right\}$ *is a stationary process such that*

$$\mathrm{E}\left[\boldsymbol{\xi}_t^{(n)}\right] = \mathbf{0} \quad \text{and} \quad \mathrm{E}\left[\xi_{it}u'_{jt}\right] = 0, \quad i = 1\ldots, n, \quad j = 1\ldots, q, \quad t, t' \in \mathbb{Z}.$$

The corresponding statistical model for the vector of the observables is :

$$\begin{aligned} \mathbf{X}_t^{(n)} &= \mathbf{B}^{(n)}(L)\mathbf{u}_t + \boldsymbol{\xi}_t^{(n)} \\ &= \boldsymbol{\chi}_t^{(n)} + \boldsymbol{\xi}_t^{(n)} \end{aligned} \tag{5}$$

where $\mathbf{B}^{(n)}(L) = \left(\mathbf{b}_1^{(n)}(L), \ldots, \mathbf{b}_q^{(n)}(L)\right)'$ is the $n \times q$ matrix with i column $\mathbf{b}_i^{(n)}(L) = (b_{1i}^{(n)}(L), \ldots, b_{ni}^{(n)}(L))'$ and $\boldsymbol{\chi}_t^{(n)} = (\chi_{1t}, \ldots, \chi_{nt})'$.

Assumption (B) implies stationarity of $\mathbf{X}_t^{(n)}$ for any n. This can accommodate processes which are stationary after some transformation such as differencing or deterministic detrending.

Denote by $\boldsymbol{\Sigma}^{(n)}(\theta)$ the spectral density of the observations and assume that its entries, $\sigma_{ij}(\theta)$, are bounded in modulus. Model (5) implies, as (2), that $\boldsymbol{\Sigma}^{(n)}(\theta)$ can be written as the sum of the spectral density of the χ's, $\boldsymbol{\Sigma}_\chi^{(n)}(\theta)$, which has reduced rank $q < n$, and the spectral density of the ξ's, $\boldsymbol{\Sigma}_\xi^{(n)}(\theta)$, which has rank n. However, since we have not imposed conditions on cross-sectional orthogonality of the idiosyncratic components, without further assumptions the model is not well specified.

The questions that have been discussed in the "large economies" literature are the same as those analyzed in the traditional factor models literature, but the analysis in population is for $n \to \infty$ while the statistical analysis establishes convergence for $(n, T) \to \infty$. This survey will review results on five issues: (i) *identification* : under what conditions (for $n \to \infty$) on the variance-covariance structure of the data is the common component identifiable ? (ii) *representation* : under what conditions (for $n \to \infty$) does a factor representation (5) exists ? (iii) *estimation* : find a (n, T) consistent estimator of the common component and derive relative rates of convergence for n and T; (iv) *forecasting*: find a (n, T) consistent estimator of the minimum mean square linear forecast; (v) *identification* of $\mathbf{u}_t$ and the $\mathbf{b}_i^{(n)}$ coefficients (for $n \to \infty$); .

4.1 Identification of the components

Without cross-sectional orthogonality of the ξ_i's, χ_{it} is not only non-observed, but also not identified. Forni, Hallin, Lippi and Reichlin (2000) have defined conditions on the spectral density matrix of the X's under which the components are identified as n goes to infinity. The asymptotic identification conditions are preconditions to develop an estimator for the component which is consistent for n and T going to infinity.

The concept of asymptotic identification has been first developed by Chamberlain (1983) and Chamberlain and Rothschild (1983) for a static factor model. Here I will develop the analysis for the more general dynamic model and then state the static result as a special case.

The essential assumption for identification is:

ASSUMPTION (A). *The non-zero q eigenvalues of $\mathbf{\Sigma}_\chi^{(n)}(\theta)$ diverge whereas all eigenvalues of $\mathbf{\Sigma}_\xi^{(n)}(\theta)$ remain bounded, almost everywhere in $[-\pi,\ \pi]$, as $n \to \infty$.*

(A) is clearly satisfied if the ξ_{it}'s are mutually orthogonal at any lead and lag (and have uniformly bounded spectral densities) as in model (2), but is more general as it allows, so to speak, for a limited amount of dynamic cross-correlation. Bounded eigenvalues of the spectral density matrix of the ξ's imply 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, the assumption 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 $h > 1$.

Similarly, the assumption of unbounded eigenvalues of the spectral density of the χ's guarantees a minimum amount of cross-correlation between the common components. With a slight oversimplification, this implies that each u_{jt} is present in infinitely many cross-sectional units with non-decreasing importance.

Model (5) under (A) is the *generalized dynamic factor model* of Forni, Hallin, Lippi and Reichlin (2000) and Forni and Lippi (2001). This model is very general in the dynamic specification and flexible in the assumption on the idiosyncratic component (from now on when refering to model (5) we will mean (5) under assumptions (A) and (B)).

Identification of the common component is shown in two steps. First, observe that by a law of large number argument, q "appropriately chosen" linear combinations of the X's become increasingly collinear with the common factor as $n \to \infty$. This implies that q averages of the observations converge to the dynamic space spanned by the q common factors. As a result, and this is the second step, by projecting the X's onto the present, past and future of these averages we converge to the common component and we identify it uniquely (non-redundancy).

In Forni, Hallin, Lippi and Reichlin (2000), it is shown that the first q dynamic ($q < n$) principal components of the X's are "appropriate aggregates". The latter are a generalization of the well known static concept (see appendix B for details). They are q processes $z_{jt},\ j = 1, \ldots, q$, linear combinations of the leads and lags of the variables in $\mathbf{X}_t$, i.e.

$$z_{jt} = \underline{\mathbf{p}}_j(L)\mathbf{X}_t, \quad j = 1, \ldots, q,$$

where L is the lag operator and $\underline{\mathbf{p}}_j^{(n)}(L)$ is a $1 \times n$ row vector of two-sided filters which has been normalized in such a way that $\underline{\mathbf{p}}_j^{(n)}(L)\underline{\mathbf{p}}_k^{(n)}(F)' = 0$ for $h \neq k$ and $\underline{\mathbf{p}}_j^{(n)}(L)\underline{\mathbf{p}}_j^{(n)}(F)' = 1$, where prime denotes transposition and $F = L^{-1}$.

As shown by Brillinger (1981), the projection of $\mathbf{X}_t$ on the present, past and future of the first q (population) dynamic principal components, provides, for given n, the best approximation (in mean square) to $\mathbf{X}_t$ by means of q linear combinations of the leads and lags of the observations.[1] The projection, can be written as :

$$\begin{aligned} \chi_{it} &= [\underline{\tilde{p}}_{1,i}(L)\underline{\mathbf{p}}_1(L) + \underline{\tilde{p}}_{2,i}(L)\underline{\mathbf{p}}_2(L) + \cdots + \underline{\tilde{p}}_{q,i}(L)\underline{\mathbf{p}}_q(L)]\mathbf{X}_t \\ &= \underline{\mathbf{K}}_i(L)\mathbf{X}_t \end{aligned} \tag{6}$$

where the k coefficient of $\underline{\mathbf{p}}_j(L)$ is

$$p_{jk} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathbf{p}_j(\theta)e^{ik\theta}d\theta \quad j = 1, \ldots, q$$

with $\mathbf{p}_j(\theta)$ being the j q-row eigenvector of $\mathbf{\Sigma}(\theta)$ corresponding to the eigenvalue $\lambda_j(\theta)$. The coefficients of the projections, $\underline{\tilde{p}}_{j,i}(L)$, are the i elements of the vector $\underline{\tilde{\mathbf{p}}}_j(L)$ which is the conjugate transpose of $\underline{\mathbf{p}}_j^{(n)}(L)$, which we can also write as $\underline{\mathbf{p}}_j^{(n)}(F)'$.

Let us now index the projection (6) by n. Forni, Hallin, Lippi and Reichlin (2000) show that, as $n \to \infty$, the $\chi_{it}^{(n)}$ converges in mean square to the common component χ_{it} of the generalized dynamic factor model (4).

[1] It is worth noting that the projection solving the maximization problem is unique, whereas the principal components themselves are not.

RESULT FHLR1 *(Forni, Hallin, Lippi and Reichlin, 2000). Under Assumptions (*B*) and (*A*), we have*

$$\lim_{n\to\infty} \chi_{it}^{(n)} = \chi_{it}$$

in mean square for any i and t.

Since $\chi_{it}^{(n)}$ depends only on $\mathbf{X}_t^{(n)}$, Result FHLR1 has the immediate implication that the components χ_{it} and ξ_{it} are identified. Moreover, it is shown that representation (5) is *non-redundant*, i.e. no other representation fulfilling Assumptions (A) and (B) is possible with a smaller number of factors. Notice that so far we have been assuming $T = \infty$, so $\chi_{it}^{(n)}$ is not an estimator, but a population quantity. Result FHLR1, however, is a basic building block for proving consistency to the common component of the corresponding empirical projection (see Section 4.3).

4.2 Representation

Unbounded eigenvalues of the spectral density of the common component and bounded eigenvalues of the spectral density of the idiosyncratic component (assumption (A)) imply that the first q eigenvalues of the observable spectral density of the observations are unbounded while the last $n - q$ are bounded (see Forni, Hallin, Lippi and Reichlin, 2000). Let us then make the assumption:

ASSUMPTION (A*). *The first q eigenvalues of $\mathbf{\Sigma}^{(n)}(\theta)$ diverge as $n \to \infty$ almost everywhere in $[-\pi,\ \pi]$, that is $\lim_{n\to\infty} \lambda_q^{(n)}(\theta) = \infty,\ \theta -$ a.e., whereas all other eigenvalues are bounded as $n \to \infty$: there exists a $\Lambda > 0$ such that $\lambda_{q+1}^{(n)}(\theta) \leq \Lambda$ almost everywhere in $[-\pi,\ \pi]$.*

We have seen that Result FHLR1 in Forni, Hallin, Lippi and Reichlin (2000) establishes that if assumption (A) and therefore (A*) is satisfied, the generalized dynamic factor model (i.e. representation (5) under assumptions (A) and (B)) exists. Forni and Lippi (2001) complete the representation theory for the generalized dynamic model by establishing an 'if and only if' result.

RESULT FL *(Forni and Lippi, 2001). The double sequence $\{X_{it};\ t \in \mathbb{Z}\}$ is a q-generalized dynamic factor model if and only if assumption (*A**) is satisfied.*

A consequence of this result is that evidence that, for some q, assumption (A*) holds becomes evidence both that the series follow a generalized dynamic factor model, and that the number of factors is q.

Result FL generalizes to the dynamic case, the result shown in Chamberlain (1983) and Chamberlain and Rothschild (1983) for the static approximate factor model:

$$X_i = c_{i1}v_1 + c_{i2}v_2 + \ldots + c_{iq}v_q + \rho_i. \tag{7}$$

Model (7) has no time dimension and is "isomorphic" to model (5) under the assumption that $b_{ij}(L)$ is trivial (does not contain leads nor lags) and $\boldsymbol{\xi}_n$ is a white noise process. If this is the case the spectral density of $\mathbf{X}^{(n)}$, its eigenvalues and eigenvectors do not depend on θ, and all coincide with the variance-covariance matrix of $\mathbf{X}^{(n)}$, its eigenvalues and eigenvectors respectively, which are the tools employed in Chamberlain and Rothschild's analysis.

They show the following:

RESULT CR *(Chamberlain, 1983 and Chamberlain and Rothschild, 1983). If the vector $\mathbf{X}_t^{(n)}$ is a white noise for any n, i.e. if the matrix $\mathbf{\Sigma}^{(n)}(\theta)$ and its eigenvalues are constant as functions of θ, then a q-factor representation exists if and only if the first q eigenvalues of the spectral density of the X's are unbounded as function of n, while the last $n-q$ are bounded.*

Results FL and CR establish a firm link between principal components and factor analysis.

As we will see, static analysis is sometimes used to treat the dynamic case. Assuming finite moving average representation for the common component, lagged factors can be treated as additional factors in a static setup such as (7). To appreciate the difference between this approach and the fully dynamic approach, consider for instance

$$X_{it} = u_t + \alpha_i u_{t-1} + \xi_{it}, \tag{8}$$

with $1 \geq \alpha_i \geq 1/2$ and $\boldsymbol{\xi}^{(n)}$ orthonormal white noise. Defining $v_{1t} = u_t$ and $v_{2t} = u_{t-1}$, model (8) is isomorphic to (7) with $q = 2$. As a consequence, the first two eigenvalues of the variance-covariance matrix of $\mathbf{X}^{(n)}$ diverge, while the third is bounded. However, analysis of the eigenvalues of the variance-covariance matrix of (8) does not allow distinction between (8) and

$$X_{it} = v_{1t} + \alpha_i v_{2t} + \rho_{it}, \tag{9}$$

where v_{1t} and v_{2t} are orthogonal at any leads and lags. By contrast, analysis of dynamic eigenvalues of the spectral density matrix gives:

(A) Model (9) has constant spectral density. Therefore, dynamic and static analysis coincide. The first two eigenvalues diverge, whereas the third one is bounded.

(B) The first eigenvalue of the spectral density matrix of (8) is not smaller than

$$\| \sum_{i=1}^{n} (1 + \alpha_i e^{-i\theta}) \|^2 = |n(1 + \bar{\alpha}_n e^{-i\theta})|^2,$$

where $\bar{\alpha}_n = \sum_{i=1}^{n} \alpha_i / n$, and therefore diverges for any $\theta \in [-\pi, \ \pi]$. The second dynamic eigenvalue is uniformly bounded.

Moreover, as soon as a model as simple as

$$X_{it} = \frac{1}{1 - \alpha_i L} u_t + \xi_{it}$$

is considered, a dynamic analysis reveals that the first eigenvalue diverges everywhere in $[-\pi, \ \pi]$, whereas the second one is uniformly bounded. By contrast, static analysis leads to the conclusion that all eigenvalues of the variance-covariance matrix diverge. This is consistent with an infinite number of static factors.

4.3 Estimation and convergence rates

Result FHLR1 shows that the common component χ_{it} can be recovered asymptotically from the sequence $\underline{\mathbf{K}}_i^{(n)}(L)\mathbf{X}_t^{(n)}$. The filters $\underline{\mathbf{K}}_j^{(n)}(L)$ 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 $\mathbf{X}_t^{(T,n)}$ (see Appendix C on details on estimation of the spectral density).

Let us assume

Assumption (C). $\mathbf{X}_t^{(n)}$ *admits a linear representation of the form*

$$\mathbf{X}_t^{(n)} = \sum_{k=\infty}^{\infty} \mathbf{c}_k^{(n)} \mathbf{Z}_{t-k}^{(n)}$$

where $\left\{ \mathbf{Z}_t^{(n)}; \quad t \in \mathbb{Z} \right\}$ *is white noise with nonsingular covariance matrix and finite fourth-order moments, and* $\sum_{k=-\infty}^{\infty} (\mathbf{c}_k^{(n)})_{ij} |k|^{1/2} < \infty \quad \text{for } i, j = 1, \cdots, n.$

Under assumption (C) any periodogram-smoothing or lag-window estimator $\mathbf{\Sigma}^{(n,T)}(\theta)$ is a consistent (for $T \to \infty$) estimator of $\mathbf{\Sigma}^{(n)}(\theta)$ (see Brockwell and Davis, 1987. Appendix C provides details on the particular estimator used in the work reviewed here). Denote by $\lambda_j^{(n,T)}(\theta)$, $\underline{\mathbf{K}}_i^{(n,T)}(L)$ the estimated counterparts of $\lambda_j^{(n)}(\theta)$, $\underline{\mathbf{K}}_i^{(n)}(L)$, respectively, and put

$$\chi_{it}^{(n,T)} = \underline{\mathbf{K}}_i^{(n,T)}(L) \mathbf{X}_t^{(n)}(L).$$

Consider a truncation of this quantity (which we denote by the same symbols so as not to burden the notation). The following result has been proved:

RESULT FHLR2 *(Forni, Hallin, Lippi and Reichlin, 2000)*

$$\lim \underline{\mathbf{K}}_i^{(n,T)}(L)\mathbf{X}_t^{(n)} = \chi_{it}$$

in probability for n and T going to infinity at some rate.

To prove Result FHLR2, write $|\mathbf{K}_i^{(n,T}(L)\mathbf{X}_t^{(n)} - \chi_{it}|$ as the sum of $R_1^{(n,T)} = |\underline{\mathbf{K}}_i^{(n,T)}(L)\mathbf{X}_t^{(n)} - \underline{\mathbf{K}}_i^{(n)}(L)\mathbf{X}_t^{(n)}|$ and $R_2^{(n)} = |\underline{\mathbf{K}}_i^{(n)}(L)\mathbf{X}_t^{(n)} - \chi_{it}|$. $R_2^{(n)}$ depends on n only, so that convergence is guaranteed by result FHLR1; $R^{(n,T)}$ depends on both T and n and convergence can be proved using Chebyshev's theorem.

Result FHLR2 is a simple consistency result, and provides no consistency rates : it merely asserts the existence of paths in $\mathbb{N} \times \mathbb{N}$, of the form $\{(n, T(n)); n \in \mathbb{N}\}$, where $n \mapsto T(n)$ is monotone increasing and $T(n) \uparrow \infty$ as $n \to \infty$, such that the difference between $\chi_{it}^{(nT)}$ and χ_{it} goes to zero in probability as n and T tend to infinity along such paths. No information is provided about the form of these paths (about $n \mapsto T(n)$).

A *consistency-with-rates* reinforcement of Result FHLR2 relates n, T, and the magnitude of the difference $\mid \chi_{it}^{(n,T)} - \chi_{it} \mid$ as n and T tend to infinity along certain paths $(n, T(n))$.

The difficulties in obtaining rates are that the estimator depends on the estimated spectral density matrix of the first n series, $\mathbf{\Sigma}^{(n,T)}(\theta)$. The latter is governed by a classical root-T rate, but the constants associated with such rates might depend on n, so that a given distance between estimated and population spectral density might require a T increasing faster than n.

To obtain results on consistency rates, additional assumptions on the model are needed. First, it suffices to impose an assumption of uniformity, as $n \to \infty$, on the variance of each of the common terms $b_{ij}(L)u_{jt}$. This can be interpreted as an assumption on the unobservable filters $b_{ij}(L)$: for instance, in the very simple model $X_{it} = u_t + \xi_{it}$, with i.i.d. idiosyncratic components ξ_{it}, the common shock u_t has a uniform impact on the n observed series. As a result, the (unique) diverging eigenvalue is $\lambda_1^{(n)}(\theta) = n + \sigma_\xi^2$, where σ_ξ^2 is the idiosyncratic variance, with eigenvector $(n^{-1/2}, \ldots, n^{-1/2})$. Hence, $\Delta^{(n)} = \sqrt{n}$. No further restriction is needed on the idiosyncratic components. A second assumption is that the convergence of the estimated spectral density $\mathbf{\Sigma}^{(n,T)}(\theta)$ to the population spectral density $\mathbf{\Sigma}^{(n)}(\theta)$ is uniform with respect to n (both the rates of convergence and the constants associated with these rates are independent of n). $\mathbf{\Sigma}^{(n,T)}(\theta)$ is obtained from any smoothed periodogram method with bandwidth B_T that only depends on T. The latter should tend to zero as $T \to \infty$ - neither too fast nor too slow, if consistency of $\mathbf{\Sigma}^{(n,T)}(\theta)$ is to be achieved at appropriate rates - while $B_T \uparrow \infty$.

Now let us $\delta^{(n)} > 0$ be an increasing sequence such that $\lim_{n\to\infty} \delta^{(n)} = \infty$.

DEFINITION. *We say that* $\chi_{it}^{(n,T)}$ *converges*

(A) *to the common factor space* $\mathcal{U}^{(n)2}$ *at rate* $\delta^{(n)}$ *along the path* $(n, T(n))$ *if, for all* $\epsilon > 0$, *there exists a* $B_\epsilon > 0$ *and a* $N_\epsilon \in \mathbb{N}$ *such that*

$$\mathrm{P}[\delta^{(n)} \mid \chi_{it}^{(n,T(n))} - \mathrm{proj}_{\mathcal{U}^{(n)}} \chi_{it}^{(n,T(n))} \mid > B_\epsilon] < \epsilon$$

for all $n \geq N_\epsilon$.

(B) *to the common component at rate* $\delta^{(n)}$ *along the path* $(n, T(n))$ *if, for all* $\epsilon > 0$, *there exists a* $B_\epsilon > 0$ *and a* $N_\epsilon \in \mathbb{N}$ *such that*

$$\mathrm{P}\left[\delta^{(n)} \mid \chi_{it}^{(n,T(n))} - \chi_{it} \mid > B_\epsilon\right] < \epsilon$$

for all $n \geq N_\epsilon$.

(We say that $\chi_{it}^{(nT)}$ achieves consistency rate $\delta^{(n)}$ along the path $(n, T(n))$).

RESULT FHLR3 *(Forni, Hallin, Lippi and Reichlin, 2002a). Both rates of convergence are proved as long as* $T(n)$ *diverge, i.e. at any relative rates for* n *and* T. *In the "classical" case of linearly diverging eigenvalues* ($\Delta^{(n)} = n^{1/2}$ *case), the rate of consistency, for convergence to the space of common components, only depends on* n *and is equal to* $n^{1/2}$. *The precision of the rate of convergence to the component, on the other hand, depends also on* T *and the rate is of the form* $\min\{\sqrt{n}, \sqrt{B_T T}\}$.

The results can be illustrated by using the elementary example:

$$X_{it} = a_i u_t + \xi_{it}, \tag{10}$$

with ξ_{it} i.i.d. and the assumption that $0 < a \leq a_i \leq b$ for any i. The estimator of, say, $a_1 u_t$ is an average of the X_{it-k}'s, $i = 1, 2, \ldots, n$, thus an average of the u_{t-k} plus an average of the ξ_{it-k}. The second average tends to vanish in variance because the ξs are orthogonal and this is quite independently of the speed at which T diverges. Thus the result that the rate

[2] $\mathcal{U}^{(n)}$ is the minimal closed subspace of $L_2(\Omega, \mathcal{F}, \mathrm{P})$ containing the first q dynamic principal components.

at which the space spanned by the common shock is approached depends crucially on the rate at which n diverges, with the rate at which T diverges playing no role.

By contrast, in most interesting cases, a fast diverging n does not help when the task is approaching the true common component $a_1 u_t$ since the estimated coefficients of the projections of the observations onto the common factors must also converge and the precision of their convergence depends on T. The cross-sectional dimension n efficiently contributes to the accuracy of the estimation, which, however, depends on the total number $nT(n)$ of observations.[3]

(n, T)-consistent estimators (and rates) to the common component of factor models slightly different than (5) have been proved by Forni and Reichlin (1998) and Stock and Watson (1999).

Consider model (5) where assumption (A) is replaced by the assumption of mutually orthogonal idiosyncratic elements. Consider then the minimal closed subspace of $L_2(\Omega, \mathcal{F}, \mathrm{P})$ containing q cross-sectional averages $\bar{\mathcal{U}}_n$ and the orthogonal projection

$$\bar{\chi}_{it}^{(n)} = \mathrm{proj}(x_{it} | \bar{\mathcal{U}}^{(n)}).$$

Forni and Reichlin (1998) have shown :

RESULT FR1 *(Forni and Reichlin, 1998)*

$$\lim \bar{\chi}_{it}^{(n)} = \chi_{it}$$

in probability for any i and t as $\min(n, T) \to \infty$.

The Forni and Reichlin's estimator is the empirical projection on the present, past and future of q cross-sectional averages. The consistency proof is constructed in a similar way than for Result FHLR2. First, it is shown, than under some conditions needed to avoid near singularities of the averages as n increases, the latter converge to the dynamic space spanned by the q common factors. Second, convergence to the common component is shown for the projections onto these aggregates. Consistency depends on the convergence properties of the coefficients of these projections, which depend on T. Forni and Reichlin's result tells us that consistency holds at any rate (independently of the relative speed at which n and T go to infinity), but do not derive explicit rates. Note than an advantage of using averages rather than dynamic principal components is that the latter are independent of the X's and not estimated (as in the case of static or dynamic principal components). However, unless

[3] Note that, though its growth can be arbitrarily slow, the series length $T(n)$ has to go to ∞.

ad hoc assumptions are introduced, near singularity of the chosen averages for n growing, with the consequence of inaccurate estimation, cannot be excluded.

Let us now consider a different restriction of model (5), namely assume that $b_{ij}(L)$ is of finite order m. Then model (5) can be written in its static form as a r-factor model where $r = (m+1)q$:

$$\mathbf{X}_t^{(n)} = \mathbf{B}^{(n)}\mathbf{V}_t + \boldsymbol{\xi}_t^{(n)}. \tag{11}$$

where $\mathbf{V}_t = (\mathbf{u}_t, \ldots, \mathbf{u}_{t-q})$ is $r \times 1$ and the i-th row of $\mathbf{B}$ is $(b_{i10}, \ldots, b_{iqm})$.

In this framework, lags are treated as additional factors and (11) can be analyzed as a static factor model. For this model, and for fixed T, Connor and Korajczyk (1986, 1992) have shown that the first r static principal components of the X's are (n)-consistent to the factor space. They show that, when T is fixed the problem of estimating the $n \times n$ variance-covariance matrix of the X's can be turned into a $T \times T$ problem and compute static principal components of the $T \times T$ variance-covariance matrix of the X's. Stock and Watson (1999) analyze the same estimator and show consistency for n and T going to infinity as well as providing results on relative rates of convergence.

Under suitable conditions on the covariance matrix of the ξ_{it}, which limit the cross-covariance of the idiosyncratic components, the results shown by Stock and Watson (1999) are

RESULT SW1 *(Stock and Watson, 1999). The static projection on the first r static principal components of the X's converge in probability to the common component in (11) at any relative rate of n and T.*

RESULT SW2 *(Stock and Watson, 1999). The static projection on the first k static principal components of the X's (where $k \geq r$) converges in probability to the space spanned by the r factors $\mathbf{V}_t$ for $n >> T$.*

Result SW1 is a stronger result than SW2 since it shows convergence to the component rather than to the space. As in FR1, convergence is shown to occur independently of the relative rate of n and T, but, unlike in Result FHLR3, no explicit rates are derived. Result SW2, on the other hand, is only for convergence to the space, but here explicit rates are derived. This is done for an estimator which is computed for a possibly misspecified number of factors and for a model which is more general than (11) in the sense that time varying factor loadings are allowed for. Convergence rates, however, require the unpleasant condition $n >> T$.

4.4 An illustrative example

To understand the intuition of Result FHLR1 and FHLR2, let us develop an example.

Let assume that we have a panel of normalized time series. Suppose that $q = 1$ and the filters $\mathbf{b}_i(L)$ 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

$$\mathbf{X}_t^{(n)} = \begin{pmatrix} L^{s_1} \\ L^{s_2} \\ \vdots \\ L^{s_n} \end{pmatrix} u_t + \boldsymbol{\xi}_t^{(n)}.$$

Moreover, assume that the idiosyncratic components ξ_{jt}, $j = 1, \dots, \infty$, are mutually orthogonal white noises, with the same variance σ^2, so that the spectral density of $\mathbf{X}_t^{(n)}$ is

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

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

$$\lambda_1^{(n)}(\theta) = n + \sigma^2 \tag{12}$$

and a valid corresponding row eigenvector is

$$\mathbf{p}_1^{(n)}(e^{-i\theta}) = \frac{1}{\sqrt{n}} \begin{pmatrix} e^{is_1\theta} & e^{is_2\theta} & \cdots & e^{is_n\theta} \end{pmatrix}. \tag{13}$$

The related filter is[4]

$$\underline{\mathbf{p}}_1^{(n)}(L) = \frac{1}{\sqrt{n}} \begin{pmatrix} F^{s_1} & F^{s_2} & \cdots & F^{s_n} \end{pmatrix},$$

while the first principal component series is

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

[4] Note that, in this example, the idiosyncratic components plays no role in the determination of $\underline{\mathbf{p}}_1^{(n)}(L)$, 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, the same property holds approximately for large n under Assumption (A), i.e. the boundedness of the idiosyncratic eigenvalues of the spectral density matrix of the X's.

Three 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 $\mathbf{u}_t$. In other words, if n is sufficiently large, the first principal component captures the information space spanned by the common shock (convergence to the space).

Second, the filter $\mathbf{p}_1^{(n)}(L)$ 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.

Third, the 'estimated' common components, which coincide here with the χ_{jt}'s appearing in equation (5) (because of the simplification $T = \infty$) are

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

Therefore, when applying the filter $\mathbf{p}_1^{(n)}(F)'$ (i.e. projecting on the leads and lags of z_{1t} as in (6)), the correct lags of the common components are restored and the leading or lagging nature of each variable emerges again. This is the convergence to the component result (Result FHLR1).

Let us comment on these three observations. The first result, "killing the idiosyncratic", could have been achieved by considering other linear combinations of the X's than dynamic principal components. Any weighted average whose square coefficients tends to zero as $n \to \infty$ does the job (on this point, see Forni and Lippi, 2001). Forni and Reichlin (1998, 2001) have used simple cross-sectional averages and generalized static principal components, Chamberlain (1983) and Chamberlain and Rothschild (1983) have used static principal components to make the same point. Notice also that for the more general case of non-orthogonal idiosyncratic, assumption (A) allows that $\underline{p}_{j,i}^{(n)}(F)'\underline{\mathbf{p}}_j^{(n)}(L)\boldsymbol{\xi}_t^{(n)}$ vanishes as n tends to infinity so that in the limit only the term $\underline{p}_{j,i}^{(n)}(F)'\underline{\mathbf{p}}_j^{(n)}(L)\boldsymbol{\chi}_t^{(n)}$ survives.

The second result , "realignment", it is due to the fact that dynamic principal components weight leading and lagging variables appropriately. Notice that had we considered, as Forni and Reichlin (1998), cross-sectional averages as the aggregates onto which to project, the implied filter would have been $(1/n, \ldots, 1/n)'$ and all variables would have been weighted equally. Had we instead used the first static principal component, as in Stock and Watson (1999), the filter would have been $(\frac{1}{\sqrt{n*}}, \ldots, \frac{1}{\sqrt{n*}})'$ where $n*$ is the number of the most numerous amongst the groups of leading, lagging and coincident variables. This implies that we would have ended up selecting only the variables belonging to the most numerous group instead of picking variables amongst all groups of coincident, leading and lagging variables, possibly losing valuable information. In order to take into account the information on the less numerous groups we would have had to consider more than one static principal components, possibly many. This poses a potential problem of efficiency and affects the approximation error which is likely to increase as the number of aggregate increases since it is inversely related to the size of the aggregates (on this point see Forni, Hallin, Lippi and Reichlin, 2002b).

The third result, "reestablishment of correct leads and lags", is what shows that for convergence to the component – which is a stronger result than convergence to the space – we need not only to find appropriate aggregates, but also to project on the present, past and future. Result FHLR1 shows that this ensures us to capture the entire dynamic space spanned by the q factors. Projecting on static principal components does not ensure the result, as shown by the example $\chi_{it} = u_{t-i}$.

4.5 One-sided estimation and forecasting

In a factor model, multivariate information can be exploited for forecasting the common component, while the idiosyncratic, being mildly cross-correlated, can be reasonably well predicted by means of traditional univariate methods (or based on low dimension models such as VARs). Therefore, the forecast of an element of the panel can be derived as the sum of the forecast of the common component, where we exploit multivariate information, and the forecast of the idiosyncratic component, where multivariate information can be disregarded. The common component being of reduced stochastic dimension, its forecast can be expressed as a projection on the span of few appropriately constructed aggregates. The method proposed by Stock and Watson, 1999 and Forni, Hallin, Lippi and Reichlin, 2002b are both based on this general idea. Notice that in the forecasting context, multivariate information is valuable provided that the panel contains many leading variables with respect to the variable to be forecasted since leading variables, loosely speaking, can help predict coincident and lagging ones.

To introduce the general idea let us consider the following example :

$$(1-L)\mathbf{X}_t^{(4)} = \begin{pmatrix} 1 \\ L \\ -1 \\ -L \end{pmatrix} u_t + \boldsymbol{\xi}_t^{(4)}$$

where we have a panel of four variables driven by one common factor which loads with different coefficients on different elements of the cross section: the first variable is pro-cyclical and leading, the second is pro-cyclical and lagging, the third is counter-cyclical and leading while the fourth is counter-cyclical and lagging.

This model can also be written in its static version as

$$(1-L)\mathbf{X}_t^{(4)} = \begin{pmatrix} u_{1t} \\ u_{2t} \\ -u_{1t} \\ -u_{2t} \end{pmatrix} + \boldsymbol{\xi}_t^{(4)}$$

where $u_{1t} = u_t$ and $u_{2t} = u_{t-1}$.

In the example, the number of dynamic factors is $q = 1$ while the number of static factors is $r = 2$. Using the argument developed in Section 4.3, this implies that one aggregate should converge to the space spanned by u_t and two aggregates should converge to the space spanned by u_{1t} and u_{2t}. Consistent estimation of the common component should be obtained with a projection onto the span of one dynamic aggregate or through a projection onto the span of two static aggregates. As discussed in Section 4.3, the former method is what suggested by Forni, Hallin, Lippi and Reichlin, 2000 and the latter is what suggested by Stock and Watson, 1999. These ideas can be applied to the forecasting problem.

Defining $\mathcal{G}(\mathbf{u}, T)$ as the span of the common factors u_{ht}, $h = 1, \cdots, q$, the optimal linear forecast of the common component (i.e. the minimum square error forecast) is:

$$\phi_{i,T+h} = \mathrm{Proj}[\chi_{i,T+h} \mid \mathcal{G}(\mathbf{u}, T)]. \tag{14}$$

The optimal h-step ahead forecast of $X_{i,T+h}$ is the sum of the above projection and the optimal linear forecast of the idiosyncratic. Given the small cross-sectional correlation of the elements of the idiosyncratic components, the latter can be expressed as a univariate autoregressive model, so that the optimal linear forecast of the variable i is:

$$X_{i,T+h} = \phi_{i,T+h} + \sum_{j=1}^{p} \alpha_i \xi_{i,t-j}.$$

The problem is to obtain a (n, T) consistent estimate of the projection, i.e. a consistent estimate of $\mathcal{G}(\mathbf{u}, T)$ and of the covariances $\boldsymbol{\Gamma}_{\chi,k}$ and $\boldsymbol{\Gamma}_{\xi,k}$.

Let us consider model (5) with the restriction of finite lag structure for the factors (as in model (11)). The method of aggregation based on static principal component advocated by SW can be applied to the forecast problem. Stock and Watson, 1999's result follows directly from result SW1 discussed in Section 4.3. Since the static rank of $\mathcal{G}(\mathbf{u}, T)$ is $r = q(m+1)$

(equal to two in the example), they can estimate $\mathcal{G}(\mathbf{u}, T)$ by the first r static principal components of the X's. Calling the vector of the first r static principal components, $\mathbf{Z}_t$, we have:

$$\phi_{i,T+h}^{(n),T} = (\mathbf{\Gamma}_{\chi,h}^{(n),T} \mathbf{Z}^{(n),T} (\mathbf{Z}^{(n),T\prime} \mathbf{\Gamma}_{\chi,0}^{(n),T} \mathbf{Z}^{(n),T})^{-1} (\mathbf{X}^{(n),T})_i$$

and

RESULT SW1 BIS *(Stock and Watson, 1999).*

$$\phi_{i,T+h}^{(n),T} \to \phi_{i,T+h}$$

in probability for $(n, T) \to \infty$ *at any relative rate of* n *and* T*.*

Notice that the estimated projection is obtained by static aggregation of the observations, i.e. as a linear combination of current values of the observations. Under finite lag structure, the result can be easily generalized to the stacked case where the weights are obtained from the covariance matrix of stacked observations.

As we have seen, Stock and Watson, 1999's method is potentially less efficient than the method based on dynamic principal components advocated by Forni, Hallin, Lippi and Reichlin, 2000 (two regressors instead than one for the example above). However, estimation results from Forni , Hallin, Lippi and Reichlin, 2000 cannot easily extended to obtain consistent forecasts since the dynamic principal component method produces an estimator to the common component which is a two-sided filter of the observations. Although their method has the advantage of exploiting the dynamic structure of the data and needs relatively few dynamic regressors (aggregates) to approximate the space spanned by the common factors, two-sideness is obviously an unpleasant characteristics for forecasting.

Forni, Hallin, Lippi and Reichlin, 2002b propose a refinement of the original procedure which retains the advantages of the dynamic approach while obtaining a consistent estimate of the optimal forecast as one-sided filter of the observations. The method consists in two steps. In the first step, they follow Forni, Hallin, Lippi and Reichlin (2000) and obtain the cross-covariances for common and idiosyncratic components at all leads and lags from the inverse Fourier transforms of the estimated spectral density matrices. Let us define them as, respectively, $\mathbf{\Gamma}_{\chi,h}^{(n),T}$ and $\mathbf{\Gamma}_{\xi,h}^{(n),T}$. In the second step, they use these estimates to obtain the r contemporaneous linear combinations of the observations with the smallest idiosyncratic-common variance ratio. The resulting aggregates can be obtained as the solution of a *generalized principal component* problem.

More precisely, they compute the generalized eigenvalues μ_j, i.e. the n complex numbers solving $\det(\mathbf{\Gamma}_{\chi,0}^{(n),T} - z\mathbf{\Gamma}_{\xi,0}^{(n),T}) = 0$ and the corresponding generalized eigenvectors $\mathbf{V}_j^{(n),T}$, $j = 1, \ldots, n$, i.e. the vectors satisfying

$$\mathbf{V}_j^{(n),T}\mathbf{\Gamma}_{\chi,0}^{(n),T} = \mu_j \mathbf{V}_j^{(n),T}\mathbf{\Gamma}_{\xi,0}^{(n),T},$$

and the normalizing condition

$$\mathbf{V}_j^{(n),T}\mathbf{\Gamma}_{\xi,0}^{(n),T}\mathbf{V}_i^{(n),T\prime} = \begin{matrix} 0 & \text{for} & j \neq i, \\ 1 & \text{for} & j = i \end{matrix}.$$

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

$$v_{jt}^{(n),T} = \mathbf{V}_j^{(n),T\prime}\mathbf{X}_t^{(n),T}, \quad j = 1, \ldots, r.$$

The generalized principal components have a useful efficiency property: they are the linear combinations of the X_{jt}'s having the smallest idiosyncratic-common variance ratio (for a proof see Forni, Hallin, Lippi and Reichlin, 2002b).

By using the generalized principal components and the covariances estimated in the first step Forni, Hallin, Lippi and Reichlin obtain an alternative estimate of $\mathcal{G}(\mathbf{u}, T)$. The estimated projection is:

$$\phi_{i,T+h}^{(n),T} = (\mathbf{\Gamma}_h^{(n),\chi}\mathbf{V}^{(n),T}(\mathbf{V}^{(n),T\prime}\mathbf{\Gamma}_0^{(n),\chi}\mathbf{V}^{(n),T})^{-1}(\mathbf{X}^{(n),T})_i$$

and :

FHLR4 *(Forni, Hallin, Lippi and Reichlin, 2002b)*

$$\phi_{i,T+h}^{(n),T} \to \phi_{i,T+h}$$

in probability for $(n, T) \to \infty$ *at a suitable rate of* n *and* T*.*

Notice that the same method can be used to reestimate the within sample common component, thus improving the accuracy of the estimator based on the first step. These projections do not involve future observations and hence do not suffer the end-of-sample problems of the Forni, Hallin, Lippi and Reichlin (2000) method.

Both Stock and Watson's estimators and Forni, Hallin, Lippi and Reichlin's are linear combinations of present and past observations, but the weighting schemes used are different. Theoretically, they both provide a consistent forecast. Empirical relative performance, however, are difficult to establish *a priori*. Indeed, though the Forni , Hallin, Lippi and Reichlin's weighting scheme, being tailored with the purpose of minimizing the impact of the idiosyncratic in the aggregate, should perform better in approximating the common factor space, relative performance depends in a very complicated way on the underlying model, on T, and on n. Forni, Hallin, Lippi and Reichlin, 2002b report simulation results comparing small sample performance of the two methods.

4.6 Identification of the factors and the coefficients: factor models and Structural VARs (SVARs)

Model (5) has no structural meaning and the factors $\mathbf{u}_t$ are identified up to (dynamic) orthogonal rotations (see Section 3). Forni and Reichlin (1998) have made the point that, as far as identification is concerned, there is a close similarity between factor models and SVAR models since, in both cases, we face the same kind of rotational indeterminacy. They have also noticed that, as long as n is finite, in the factor model the problem is to identify the joint moments of the factors, the factors loadings, not the factors themselves. The common shocks are inherently unobservable and cannot be identified. By contrast, when n is infinite, the factors can be identified and the similarity with SVARs is even closer. Since the idiosyncratic disturbances die out in the aggregate, for the aggregate variables the factor model collapses in a VAR model so that we can use the same identification strategy used for SVARs to identify the shocks and therefore the parameters of the factor model.

Forni and Reichlin's argument is as follows. Assume that the $\mathbf{b}_j(L)$'s filters in (5) are one-sided and let $\mathbf{z}_t = \mathbf{D}(L)\mathbf{u}_t$ denote a fundamental infinite moving average representation of a $q \times 1$ vector of aggregates. Fundamentalness implies

$$\overline{\text{span}}\big(\mathbf{z}_{t-k}, k \geq 0\big) \supseteq \overline{\text{span}}\big(\mathbf{u}_{t-k}, k \geq 0\big)$$

i.e. that $\mathbf{u}_t$ can be recovered from a projection onto the present and past of $\mathbf{z}_t$.

The vector $\mathbf{z}_t$ could be any weighted averages of the X's whose square coefficients tend to zero as $n \to \infty$ (simple cross-sectional averages or principal components). This condition ensures $\mathbf{z}_t$ to be the n-mean square limit of $\mathbf{z}_t^{(n)}$ (see Forni and Lippi, 2001). The Wold representation of the vector of the aggregates can be recovered as

$$\mathbf{z}_t = \mathbf{D}(L)\mathbf{D}(0)^{-1}\mathbf{v}_t$$

where $\mathbf{v}_t = \mathbf{D}(0)\mathbf{u}_t$.

By inverting the Wold representation and reordering terms, one can obtain the VAR

$$\mathbf{z}_t = \mathcal{D}(L)\mathbf{z}_{t-1} + \mathbf{v}_t.$$

Having obtained an estimate for $\mathbf{v}_t$ and $\mathcal{D}(L)$, the authors suggest to identify $\mathbf{D}(L)$ as in the SVAR literature. This implies the following steps. Orthonormailty of the shocks is imposed by switching to representation

$$\mathbf{z}_t = \big(\mathbf{D}(L)\mathbf{D}(0)^{-1}\mathbf{U}^{-1}\big)\mathbf{U}\mathbf{v}_t = \hat{\mathbf{D}}(L)\hat{\mathbf{u}}_t,$$

where $\mathbf{U}$ is the upper triangular matrix such that $\mathbf{U}\boldsymbol{\Sigma}_v\mathbf{U}^{(-1)} = \mathbf{I}_q$. Any alternative fundamental MA representation corresponding to a given structural model can be obtained from an orthonomal rotation of $\hat{\mathbf{D}}(L)\hat{\mathbf{u}}_t$, i.e. it will be of the form:

$$\mathbf{z}_t = \mathbf{C}(L)\mathbf{e}_t,$$

with $\boldsymbol{\Sigma}_e = \mathbf{I}_q$, $\mathbf{e}_t = \mathbf{R}'\hat{\mathbf{u}}_t$ and $\mathbf{C}(L) = \hat{\mathbf{D}}(L)\mathbf{R}$, where $\mathbf{R}$ is a constant orthonormal matrix such that $\mathbf{R}\mathbf{R}' = \mathbf{I}_q$.

Hence, in order to identify the model, it suffices to select a $q \times q$ static rotation $\mathbf{R}$.

Forni and Reichlin (1996) establish the following results.

RESULT FR2 *(Forni and Reichlin, 1996). If $\mathbf{u}_t$ is fundamental for $\mathbf{z}_t$, then none of the X_{it} Granger causes $\mathbf{z}_t$.*

The intuition of this result is that, if none of the individual variables are leading with respect to vector of aggregates, which spans the same space as the common component, then the common shocks $\mathbf{u}_t$ must belong to the space spanned by the present and past of the aggregates. Therefore, fundamentalness, unlike for SVARs, is testable.

Forni, Lippi and Reichlin (2002) explore this point further and show (a) that conditions under which the common factors can be recovered from the present and past of the observations ($\mathbf{u}_t$ fundamental for X_{it}, for all i's) are easily met in factor models; (b) how to derive the fundamental representation and identify the loadings $\mathbf{b}_i(L)$ and the factors $\mathbf{u}_t$.

Result (a) can be understood by the fact that, in factor models, the problem is to identify few common shocks from information on many variables. This implies that fundamentalness does not require conditions on each n rows of $\mathbf{B}(L)$ in (5)[5] , but just the existence of a $n \times q$ one-sided filter $\mathbf{C}(L)$ such that:

$$\mathbf{C}(L)'\mathbf{B}(L) = \mathbf{I}_q.$$

[5] This is under the assumption that (each element of) $\mathbf{b}_i(L)$, $i = 1, \ldots, \infty$ in (4), is unilateral toward the past.

This is indeed a very mild condition. If a $q \times q$ invertible submatrix of $\mathbf{B}(L)$ exists, there is no problem. If not, the left inverse could still exists. Consider the following example. Assume that $q = 1$ and that each line of $\mathbf{B}(L)$ is non-fundamental moving average of order one, e.g. $b_i(L) = 1 - a_i L$ with $a_i > 1$ so that none of them is invertible. Nonentheless, if $a_i \neq a_j$:

$$\frac{(1 - a_i L)a_j - (1 - a_j L)a_i}{a_j - a_i} = 1.$$

Therefore we can set $c_h(L) = 0$ for $h \neq i, j$; $c_i(L) = \frac{a_j}{a_j - a_i}$; $c_j(L) = \frac{a_i}{a_j - a_i}$. Hence the only case in which we have non-fundamentalness is

$$\mathbf{B}(L) = \mathbf{B}(0)(1 - \frac{1}{a}L)$$

and the fundamental noise is

$$w_t = \frac{1 - aL}{1 - (1/a)L} u_t.$$

Once fundamentalness is established, the problem of identification is reduced to the selection of a static rotation of dimension q. Notice that, in this framework, the number of required restrictions depends on q and not on n. What matters is the number of 'common shocks', not the number of variables. This contrasts with the SVAR framework where the rotation problem has dimension n and the number of required restrictions for just identification grows with the number of included variables.

This point strongly advocates for the use of factor models for structural macroeconomic analysis. Typically, economic theory does not provide clear guidance on what variables to consider and we know from the SVAR literature that results on impulse response functions are not robust with respect to the conditioning variables. Not only the rotation is arbitrary, but the choice of variables also is. In the factor framework we can in principle include all potentially useful information and then extract the common shocks q which are likely to be a small number as compared with n. Once the relevant dimension is known, identification is a $q \times q$ problem. Moreover, conditions for fundamentalness are easier to find than in SVARs.[6]

5. Econometric problems

Identification of the number of dynamic factors q and of the number of static factors r

Criteria for the choice of q in the large economies literature are heuristic. In the discussion of Section 4 we have assumed that q, the number of non-redundant common factors, is

[6] For an analysis of this problem, see Lippi and Reichlin (1993).

known. In practice of course, q is not predetermined, and also has to be selected from the data. Result FL can be used to this end, since it links the number of factors in (5) 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 $\boldsymbol{\Sigma}_r^T(\theta)$, $r \leq n$, can be estimated, and the resulting empirical dynamic eigenvalues $\lambda_{rj}^T(\theta)$ computed for a grid of frequencies. The following two features of the eigenvalues computed from $\boldsymbol{\Sigma}_r^T(\theta)$, $r = 1, \ldots, n$, should be considered as reasonable evidence that the data have been generated by (4), 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.

Connor and Korajczyk (1993), on the other hand, developed a formal test for the number of factors in a static factor model under sequential limit asymptotics, i.e., n converges to infinity with a fixed T and then let T converge to infinity. Their test is valid under assumptions on cross-sectional stationarity (mixing conditions). The problem with this strategy is that a cross-sectional mixing condition is hard to justify given that the cross-section, unlike time, has no natural ordering.

Stock and Watson (1999) show that, assuming $n, T \to \infty$ with $\sqrt{n}/T \to \infty$, a modification to the BIC criterion can be used to select the number of factors optimal for forecasting a single series. As observed by Bai and Ng (2002), their criterion is restrictive not only because it requires $n >> T$, but also because there can be factors which are pervasive for a set of data and yet have no predictive ability for an individual data series. Thus, their rule may not be appropriate outside the forecasting framework.

Bai and Ng (2002) consider the static factor model and set up the determination of the number of factors as a model selection problem. The proposed criteria depends on the usual trade-off between good fit and parsimony.

Liska (2001) extends the Bai and Ng's criterion to the dynamic factor model case.

Asymptotic distribution

Results described so far establish consistency of the estimators. Bai, 2001 has started developing inferential theory for static factor models of large dimension and derived asymptotic distribution (and rates of convergence) of the estimated factors, factor loadings and the common components.

Testing

The GDFM provides a parsimonious way to take into account the large information set assumption when estimating shocks and propagation mechanisms. As we have seen in Section 4.6, one can identify few shocks from many variables by imposing a minimal set of economically motivated restrictions. Since the shocks to be identified are less numerous than the variables one is conditioning on, with a set of minimal restrictions overidentification restrictions can easily be obtained. Giannone, 2002 has developed testing procedures for long run restrictions in this context.

The paper deals with hypothesis testing problems involving the loading coefficients of the common shocks and proposes a Wald type test statistics that convergences to a chi-square distribution as both the number of series and the number of observations go to infinity.

Bai and Ng, 2001b explore unit root testing in large factor models and propose a new approach. The idea is to exploit the information contained in a large panel of time series to perform a common-idiosyncratic decomposition and then to test stationarity of the common and idiosyncratic components separately. While unit root tests are imprecise when the common components and the idiosyncratic have different order of integration, direct testing on the two components are found to be more accurate in simulations.

6. Applications

To illustrate some of the potential uses of the methodology discussed for revealing features of business cycle behavior, we will describe three applications on different data sets.

6.1 The US manufacturing sector (Forni and Reichlin, 1998).

This paper investigates the behavior of output and productivity for 450 manufacturing sectors in the US from 1958 to 1986.

• *What is the degree of cross-sectional commonality in the time series behavior of sectoral output and productivity ?*

A first answer to this question can be given by assessing the relative importance of the common and idiosyncratic components. An overall measure of fit is the ratio of the sum of the variances of the common components to the sum of the total variances of the variables. This, which is the weighted mean of the sectoral R^2 with weights proportional to the total variances, gives us a percentage of 41% for output and of 29% for productivity.

• *Is the degree of commonality stronger at business cycle frequencies ?*

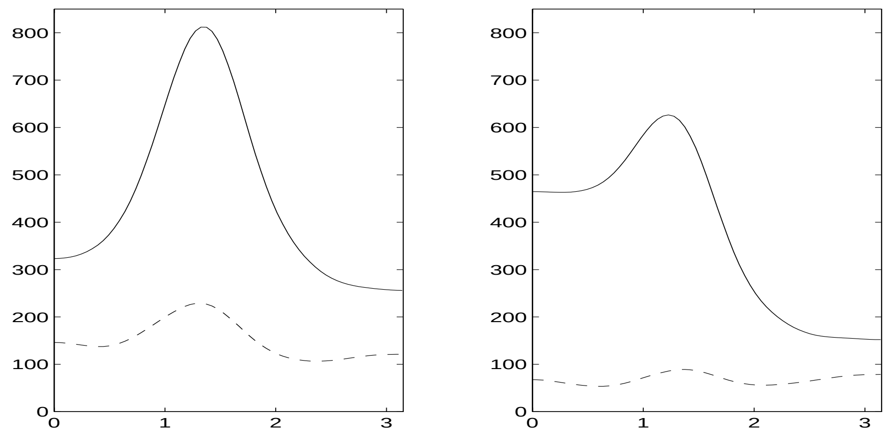
Overall variance ratios are not sufficiently informative about the role of idiosyncratic shocks for business cycle fluctuations. For this we must look at the distribution across frequencies of the variances of the common and sectoral components. This is captured by the sum of the spectra for the common and the idiosyncratic component (Figure 1).

Notice that, for both variables, while the common component has a typical business cycle shape with a peak corresponding to a period of just over four years, the bulk of the variance for the idiosyncratic component is spread equally over frequencies, suggesting that the latter can be characterized as a white noise, probably capturing measurement error. We should conclude that the business cycle features in manufacturing are mostly explained by economy-wide shocks and that, although the sectoral dynamics is more sizeable than the economy-wide one, it cannot account for the cyclical behavior of output and productivity.

• *Number of common shocks and their propagation over the 450 sectors of output and productivity.*

The heuristic method proposed in the paper identifies two common shocks one of which is identified as the technology shock (see the paper for details and Section 4.5). To analyze the weight of the substitution or reallocation effects in the total variability of output and productivity, the authors look at the correlation structure of the impulse response functions associated to the two aggregate shocks. Let us call *substitution effects* the negative sectoral

Figure 1: Sum of the spectra of the common and idiosyncratic components of output (a) and productivity (b)



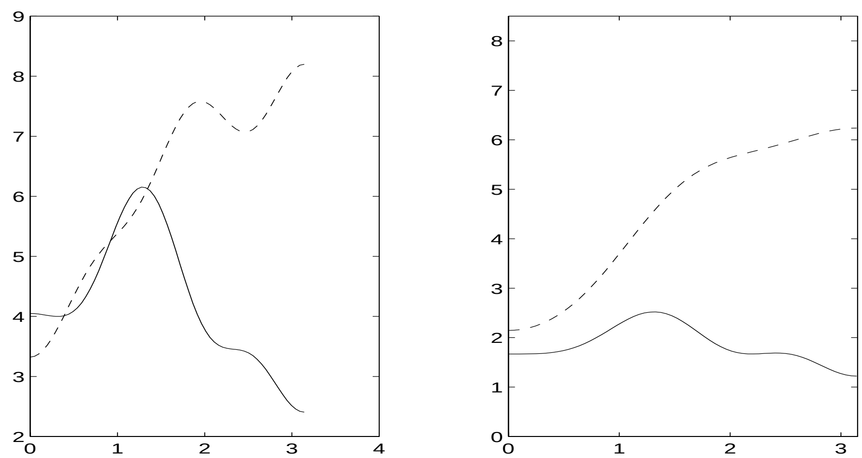
common component (solid line), idiosyncratic component (dashed line). Horizontal axis: frequencies $\theta \in [0, \pi]$ where $\theta = 2\pi/P$ and P is the cycle's period

comovements generated by the aggregate shocks and *complementary effects* the positive sectoral comovements. The former can be measured by the absolute value of the sum of the negative cospectra $|\sum \sigma_{ij}(\theta)_-|$ while the latter by the absolute value of the sum of the positive cospectra $|\sum \sigma_{ij}(\theta)_+|$.

Figures 2 reports $|\sum \sigma_{ij}(\theta)_-|$ and $|\sum \sigma_{ij}(\theta)_+|$ for the technology shock and the non-technology shock (the technology shock here is identified by selecting a particular orthonormal rotation of the moving average representation of the vector of the cross-sectional averages of output and productivity).

The Figures illustrate nicely the business cycle features of our data: all the series of the sums of the positive cospectra have peaks at business cycle frequencies, while the series of

Figure 2: Absolute sum of positive (solid lines) and negative (dashed lines) cospectra



technological component (a), non technological component (b). Horizontal axis: frequencies $\theta \in [0, \pi]$ where $\theta = 2\pi/P$ and P is the cycle's period

the negative cospectra are rather flat. Moreover the business cycle is partly real since the technology shock generates positive cospectra at a period of about four years.

6.2 Europe versus the US: regions (counties), nations (states) and aggregate behavior (Forni and Reichlin, 2001).

This study analyzes a data set on output growth for 138 regions of nine European countries and 3075 counties for 48 US states. The period considered, for most the data set, is 1970-1993.

Let us denote with y_t^{ij} the growth rate of output for the i-th region of nation j, expressed in deviation from the time-series mean. The authors assume

$$y_t^{ij} = E_t^{ij} + N_t^{ij} + \mathcal{L}_t^{ij} = a^{ij}(L)e_t + b^{ij}(L)n_t^j + c^{ij}(L)l_t^{ij}, \tag{1}$$

for $j = 1, \ldots, J$ and $i = 1, \ldots, I^j$. E_t^{ij}, N_t^{ij} and $\mathcal{L}_t^{ij}$ are the European component, the national component and the local component respectively. $a^{ij}(L)$, $b^{ij}(L)$ and $c^{ij}(L)$ are functions in the lag operator L. The European shock e_t, the national shocks n_t^j and the local shocks l_t^{ij} are unobserved unit-variance white noises, mutually uncorrelated at all leads and lags.

The difference of this model with respect to the traditional dynamic factor model with orthogonal idiosyncratic components is that the factor n_t^j is neither common nor idiosyncratic. It is an intermediate-level factor, common for regions belonging to the same country but orthogonal across countries. The difference with the generalized factor model is that additional restrictions on of the spectral density matrix of the observations are imposed.

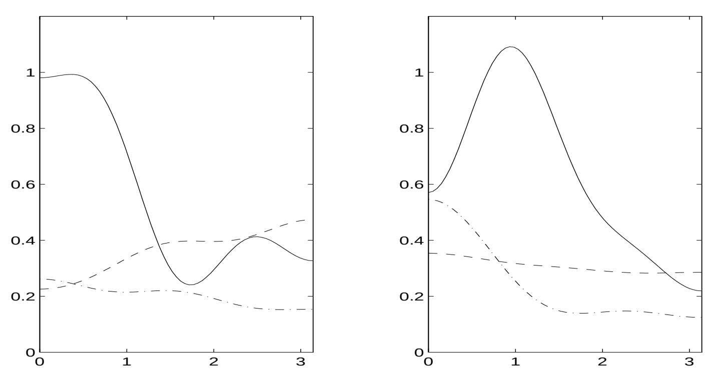
- *Regional, national and over-national commonality*

For most European countries, the common European component explains the bulk of output variance. Exceptions are Greece, Portugal and the UK which have a large nation-specific volatility, well above 50%. Idiosyncratic regional variance is also sizeable (about 30 % on average) while the nation-specific volatility is the smallest. Results for the US are strikingly similar. The average size of the US-wide component is similar to that of the European component when Greece, Portugal and the UK are excluded and the US state component seems to be of the same order of magnitude than the national component in Europe. In both cases, global and local dynamics seem to prevail over national dynamics. This result is especially surprising for the European case. It indicates that pre-monetary union Europe had already a high degree of business cycle commonality and that nation-specific cycles do not account for much of total output volatility.

- *Cross-sectional dynamics*

Figure 3 shows the average spectra of the three components for Europe (UK excluded) and the US (medium and large counties). Although, as we have seen, the variances of the output components are similar, the dynamic profiles are very different. There are two main differences between Europe and the US. First, the European common component is very persistent, whereas the US-wide component exhibits a typical business cycle shape, which peaks at a period of around six years. By looking at low frequencies we see that the total long-run variance is similar in Europe and the US, so that the uncertainty about the future income level at, say, ten or twenty years is nearly the same. However—and this is the second difference—while in the US the main bulk of long-run fluctuations is state specific or local, in Europe the long-run variance is mainly common. This implies that European regions have a larger long-run covariance, i.e. larger cospectra near the vertical axis. In other words, European regions have a "common destiny" in the long-run, more than US counties. Notice that this does not imply that European countries have similar income levels, or that they are converging to the same income level. The result only says that persistent shocks are mainly common. Still, both drift and initial levels, which are not analyzed here, could be rather different, leading to permanent gaps, convergence or even divergence.[7]

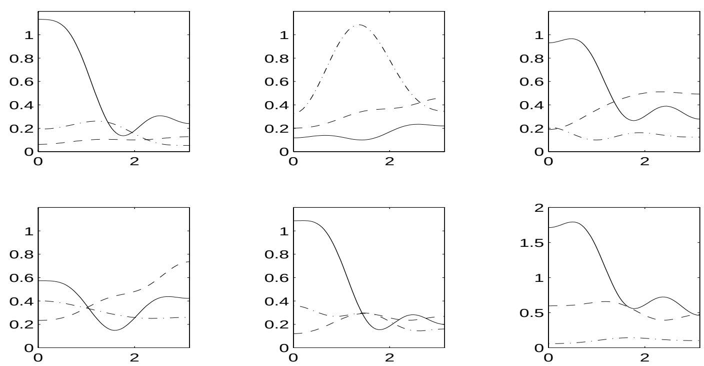
Figure 3. The spectral shape of the three components for Europe and the US



Horizontal axis: frequency; vertical axis: spectral density. Common component: solid line; National (state) component: dotted and dashed line; local component: dotted line. Europe (first data set, UK excluded) is on the left; US (large and medium counties) is on the right.

[7] Note also that we are analyzing total income, as opposed to per-capita income. Different dynamics of total income in the US and Europe could be compensated by migrations, leading to similar dynamics of per-capita income.

Figure 4. The spectral shape of the three components for six European countries



Horizontal axis: frequency; vertical axis: spectral density. Common component: solid line; National (state) component: dotted and dashed line; local component: dotted line. From the left to the right: Germany, UK, France; Italy, Belgium, Netherland (Groningen excluded). Netherland have a different scale on the vertical axis.

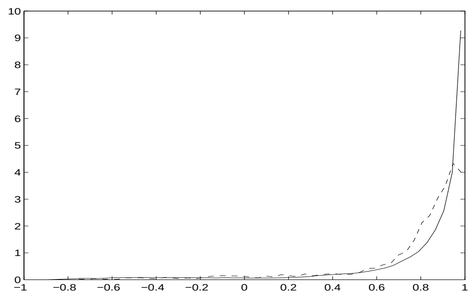
The analysis disaggregated by country (Figure 4) shows that the UK is the only European nation which, like the US, has a typical business cycle. The other countries confirm the aggregate result of a large European wide component with most of the variance concentrated at low frequency (although Italy has a lot of high frequency variation in the idiosyncratic).

• *Synchronization and symmetry*

Before we can interpret these results on variance decomposition in terms of degree of integration, we must study the degree of synchronization and symmetry of the propagation mechanisms of the common shock. If the response functions of the common shock had different signs, the interpretation of a large common component at low frequency would imply that regions are diverging, exactly the opposite of what concluded so far.

A rough measure of symmetry and synchronization is given by the regional distribution of the correlation coefficients. In Figure 5 we report the frequency density distribution of a grid of correlation intervals for European regions and US medium and large counties. In both cases, we can see that most correlation coefficients are positive and large.[8] This confirms the interpretation of variance decomposition results given above, even if we can notice that European regional dynamics is slightly more asymmetric than in the US case.

Figure 5. Density distribution of correlations among the common components



Europe: dashed line; US: solid line. Horizontal axis: correlation coefficients.

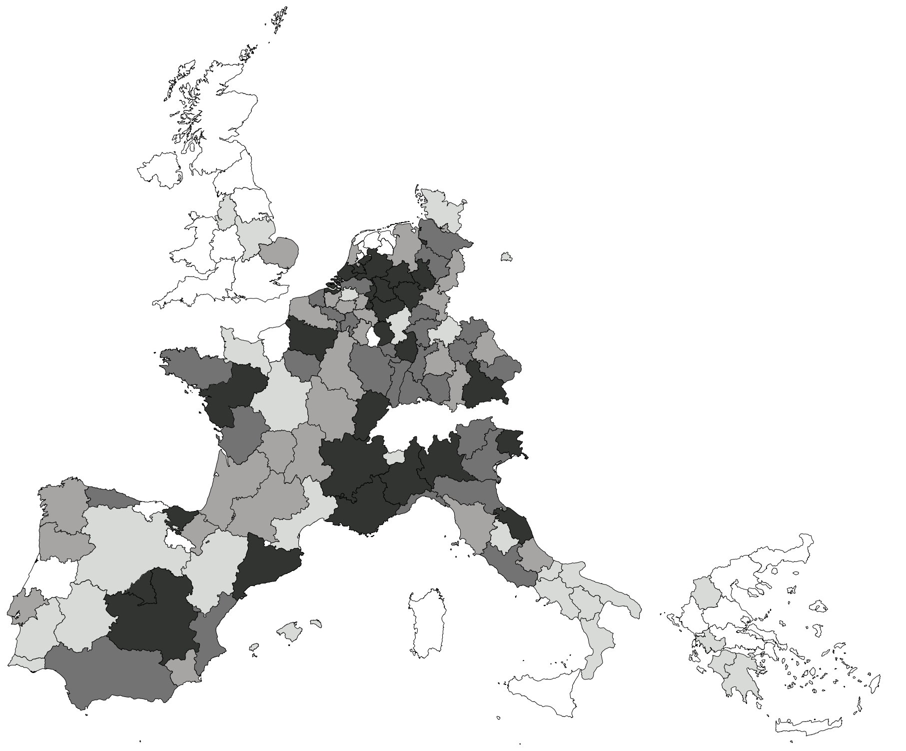
- *Geography*

Finally, to complete the picture on European integration, we want to ask the question of whether regions which are "more European" (larger relative variance of the European component) belong to a particular geographical area. Figure 6 reports the geographical distribution of variance ratios between European-wide components and total variance. Light gray indicates a small European component while dark gray indicates a large European component.

The Figure shows that a core made by the key countries France, Germany and the Benelux does not exist. Dark and light spots are sparse, indicating that almost all countries are partly in and partly out. The only exceptions are Greece and the UK, which are clearly less integrated with the rest of Europe. In general, heterogeneity within nations seems at least as large as heterogeneity across nations.

[8] The negative values are accounted for by Sicily, Sardinia, some UK regions and Groningen, the Dutch outlier.

Figure 6. Percentage of output variance explained by the European component



Dark regions have a large European component. Limits for color changes are 0.23, 0.42, 0.58, 0.70.

In summary, European regions are already highly integrated and are expected to "move together" in the long-run more than US counties. The common shock exhibits high persistence in Europe and a typical business-cycle shape in the US. If we exclude Greece and the UK, Europe appears a continent of regions rather than nations, as far as output fluctuations are concerned.

6.3 EuroCOIN: A real time coincident indicator for the euro area business cycle, Altissimo, Bassanetti, Cristadoro, Forni, Hallin, Lippi, Reichlin, Veronese, 2001

The paper develops a methodology for the construction of a monthly indicator of the euro area business cycle which is released every month without waiting for the current month publication of industrial production or current quarter publication of GDP. It refines idea in Forni, Hallin, Lippi and Reichlin, 2001a. The index is based on approximately 1000 monthly time series for the six major countries of the euro area: Belgium, France, Germany, Italy, The Netherlands and Spain since 1987. It is now published by the Center of European Policy

Research (CEPR) the 28th of every month (consult www.cepr.org).

The basic idea driving the construction of the coincident indicator is that the GDP is a good summary measure of economic activity. However, GDP is affected by errors and noise disturbing the true underlying signal. Such disturbances are measurement errors, local and sectoral shocks and high-frequency, low-persistence movements. The procedure, based on the generalized dynamic factor model, is designed to clean the GDP from such disturbances.

Let the first variable in our panel be the European real GDP. Then EuroCoIN is defined as χ_1^C, i.e. the cyclical, common component of the European GDP.

Why smoothing the GDP by eliminating the short-run part of the common component? Monthly data are typically affected by large seasonal and higher frequency sources of variation. Both economic agents and policy makers are not particularly interested in such high-fequency changes because of their transitory nature. Washing out temporary oscillations is necessary to unveil the true underlying long-lasting tendency of the economic activity.

Why cleaning the GDP from the idiosyncratic component? There are two reasons for this: eliminating measurement errors and produce a better signal for policy makers. Firstly, national GDPs are not obtained by means of direct observation, but are at least in part the result of estimation procedures and therefore are affected by estimation errors. Moreover, data on GDP are provided quarterly by the statistical institutes; monthly figures can only be obtaind by interpolating original data, which entails additional errors. Finally, the European GDP stems from aggregation of data provided by heterogeneous sources, not all equally reliable and perfectly comparable. Summing up, the European GDP is affected by large measurement and estimation errors. Such errors are mainly idiosyncratic, since they are poorly correlated across different variables and independent from the common shocks. Secondly, the idiosyncratic component should capture both variable-specific shocks, such as shocks affecting, say, the output of a particular industrial sector, and local-specific shocks, such as for instance a natural disaster, having possibly large but geographically concentrated effects. Distinguishing between such shocks and common shocks, affecting all sectors and areas, can be useful for policy makers, who have to decide whether to carry out local and sectoral measures or common, Europe-wide interventions.

The full estimation procedure is in four steps:

- **Estimating the covariances of the unobservable components**

Covariances are obtained by applying the inverse Fourier transform to the estimated spectral density (see Appendix D). The covariances for the cyclical and the non-cyclical components χ^C_{jt} and χ^S_{jt} are obtained by applying such transformation to the selected frequency band of the estimated spectra and cross-spectra (see appendix D).

• **Estimating the static factors**

In the second step, the authors compute an estimate of the static factors, following Forni, Hallin, Lippi and Reichlin, 2002b. With the term "static factors" we mean the $r = q(m+1)$ variables appearing contemporaneously in representation (11), including the lagged u_t's, so that, say, u_{1t} and u_{1t-1} are different static factors.

• **Estimating the cyclical common components**

In the third step they use contemporaneous, past and future values of the static factors to obtain an estimate of χ^C_{1t}, the cyclical component of the GDP. Precisely, they project χ^C_{1t} on $\mathbf{v}_{t-m}, \ldots \mathbf{v}_{t+m}$ where $\mathbf{v}_t = (v_{1t}, \cdots, v_{rt})'$. The lag-window size m should increase with the sample size T, but at a slower rate. Consistency of such estimator is ensured, for appropriate relative rates of m, T and n, by the fact that (a) the projection of χ^C_{1t} on the first m leads and lags of χ_{1t} is consistent beacause of consistency of $\chi^{(n),T}_{1t}$ and the estimated covariances involved; (b) χ_{1t} is a linear combination of the factors in $\mathbf{v}_t$, so that projecting on the factors cannot be worse than projecting on the common component itself.

Notice that this method is something like a multivariate version of the procedure by Christiano and Fitzgerald (2001) to approximate the band-pass filter. Exploiting the superior information embedded in the cross-sectional dimension makes it possible to obtain a very good smoothing by using a very small window ($m = 1$). This has the important consequence that a timely and reliable end-of-sample estimation is obtained without having to revise the estimates for a long time (say 12 months or more) after the first release, as with the univariate procedure. To get an intuition of the reason why they get good results with a narrow window, consider the extreme case $m = 0$. Clearly with univariate prediction one cannot get any smoothing at all. By contrast, the static factors will include in general both contemporaneous and past values of the common shocks and can therefore produce smooth linear combinations.

• **End-of-the sample unbalance**

Finally, data become available with different delays. Had we to wait until the last updating arrives, we would be able to compute the indicator only with a delay of four or five months. The authors propose a procedure to handle this problem, which allows to get provisional estimates by exploiting, for each series, the most updated information. Once the missing data become available, the final estimate is computed. This is why the indicator is subject to revision for a few months.

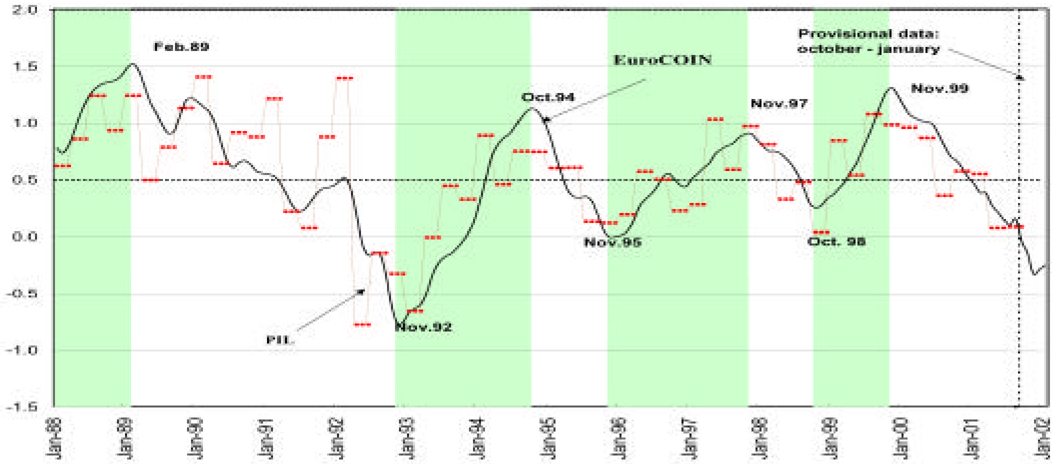
The procedure to handle this problem is the following. Let T be the last date for which the whole dataset is available. Until T they estimate the static factors as explained above, i.e. by taking the generalized principal components of the vector $\mathbf{X}_T^{(n),T}$. From T onward, they use the generalized principal components of a modified n-dimensional vector $\mathbf{Y}_T^{(n),T}$ which includes, for each process in the data set, only the last observed variable, in such a way as to exploit for each process the most recent information. Clearly computation will involve the estimated covariance matrices of the common and the idiosincratic component of $\mathbf{Y}_T^{(n),T}$ in place of those of $\mathbf{X}_T^{(n),T}$.

The idea is simply to shift the variables in such a way as to retain, for each one of them, only the most updated observation, and compute the generalized principal components for the re-aligned vector. In such a way one is able to get information on the factors u_{hT+j}, $h = 1, \dots, q$, $j = 1, \dots, w$, and to exploit it in prediction. The forecasts are then used to replace missing data and to get the forecasts of $\boldsymbol{\chi}_{T+h}$, $h > w$.

Figure 7 below reports the EuroCOIN since 1987:1 plotted against monthly growth of GDP (both at quarterly rate). The chart also marks the dating of the euro business cycle. Shaded areas are expansions, i.e. periods of prolonged increasing growth. Notice that, as opposed to the classifical definition used by the NBER, which is in terms of business cycle levels, the definition of recessions and expansions are in terms of rates of growth (growth cycle concept).

A by-productof the methodology is that one can track the contribution of different variables or block of variables (sectors, nations) to the aggregate result and evaluate leadership and commonality of different blocks. These "facts" are reported in the paper and a selection of them regularly published on the web page of the cepr. Obviously, at the end of the sample, the estimates are based on a reduced number of variables where the weight of financial variables and survey data, which are released with a minimum delay, is particularly large.

Figure 7. EuroCOIN and Euro area GDP (quarterly rate of change)



6.4 Other applications

Other applications of the model to macroeconomic and financial variables are the following:

• *Business cycles*: the construction of a monthly index of core inflation for the euro area (Cristadoro, Forni, Reichlin and Veronese, 2001); the identification of the world business cycle (Malek-Mansour, 2001).

• *Monetary policy*: The propagation of monetary policy shocks across European countries (Sala, 2000) and the analysis of systematic and unsystematic monetary policy in the US (Giannone, Reichlin and Sala, 2002), Bernanke and Boivin (2001).

• *Links between financial and real variables*: the role of financial variables in predicting output and inflation in the euro area (Forni, Hallin, Lippi and Reichlin, 2001c and D'Agostino, 2001), the use of factor models for measuring the degree of real and financial markets integration (Emiris, 2001)

• *Exchange rate dynamics*: The cross-sectional predictability of exchange rates (Rodrigues, 2002)

REFERENCES

Altissimo, Filippo, Bassanetti, Antonio, Cristadoro, Riccardo, Forni, Mario, Lippi, Marco, Reichlin, Lucrezia and Giovanni Veronese (2001), "A real time coincident indicator of the Euro area business cycle", CEPR working paper.

Bai, Jusham (2001), "Inference on factor models of large dimension", manuscript.

Bai, Jusham and Serena Ng, (2001a), "A new look at panel testing of stationarity and the PPP hypothesis", October, manuscript.

Bai, Jusham and Serena Ng, (2001b), "A panick attack on unit roots and cointegration", December, manuscript.

Bai, Jusham and Serena Ng (2002), "Determining the number of factors in approximate factor models", vol. 70, issue 1, pp. 191-221, *Econometrica*.

Bernanke, Ben and Jean Boivin, "Monetary policy in a datarich enviroment" *Journal of Monetary Economics*.

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

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

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

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

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

Connor, G. and Korajczyk, R.A. (1988) "Risk and return in an equilibrium APT. Application of a new test methodology", *Journal of Financial economics* 21, pp. 255–89.

Connor, G. and Korajczyk, R.A. (1993) "A test for the number of factors in an approximate factor model",*Journal of Finance* XLVIII:4, pp. 1263-1291.

Christiano, Lawrence J. and Terry Fitzgerald, J. (2001), "The band pass filter", manuscript, July.

Cristadoro, Riccardo, Forni, Mario, Reichlin, Lucrezia and Giovanni Veronese (2001), "A core inflation index for for the Euro area", CEPR working paper.

D'Agostino, Antonello (2002), "Understanding the leading-lagging structure of sectoral returns", January, manuscript.

Emiris, Marina (2001), "Measuring capital market integration", May, manuscript.

Favero, Carlo and Massimiliano Marcellino (2001), "Large data sets, small models and monetary policy in Europe", CEPR working paper, December.

Forni, Mario, Hallin, Marc, Lippi, Marco and Lucrezia Reichlin, (2000) "The generalized dynamic factor model : identification and estimation", *The Review of Economics and Statistics*, November.

Forni, Mario, Hallin, Marc, Lippi, Marco and Lucrezia Reichlin (2001a), "Coincident and leading indicators for the Euro area", *The Economic Journal* **111**, 62-85.

Forni, Mario, Hallin, Marc, Lippi, Marco and Lucrezia Reichlin, (2001b) "Do Financial Variables Help Forecasting Inflation and real Activity in the Euro Area?", CEPR working paper.

Forni, Mario, Hallin, Marc, Lippi, Marco and Lucrezia Reichlin, (2002a), "The generalized dynamic factor model : consistency and convergence rates", *Journal of Econmetrics*, forthcoming.

Forni, Mario, Lippi, Marco and Lucrezia Reichlin, (2002),"Opening the black box: identifying shocks and propagation mechanisms in VARs and factor models", manuscript, January.

Forni, Mario, Hallin, Marc, Lippi, Marco and Lucrezia Reichlin, (2002b) "The generalized dynamic factor model : one-sided estimation and forecasting", manuscript, January.

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

Forni, Mario, and Marco Lippi, (2001) "The generalized factor model: Representation theory", *Econometric Theory*,**17**, 1113-1141.

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

Forni, Mario, and Lucrezia Reichlin, (2001) "National policies and local economies", *European Economic Review*, January.

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

Geweke, John, (1977) "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).

Geweke, John and Kevin J. Singleton (1981) "Maximum likelihood 'confirmatory' factor analysis of economic time series", *International Economic Review* 22, pp. 37-54.

Giannone, Domenico (2002), "Testing for linear restrictions using with large panels of time series", manuscript, January.

Giannone, Domenico, Reichlin, Lucrezia and Luca Sala (2002), "Tracking Greenspan: systematic and unsystematic monetary policy revisited", manuscript, January.

Lippi, Marco and Lucrezia Reichlin (1993) "The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment", *American Economic Review* 83, pp. 644–52.

Liska, Roman (2001) "Dynamic factor analysis: the number of factors and related issues", manuscript, December.

Malek-Mansour, Jeoffrey, (2001) "Do national business cycles have an international origin?", *Empirical Economics*.

Marcellino, Massimiliano, James H. Stock, Mark W. Watson, (2002) "Macroeconomic forecasting in the euro area: country specific versus area-wide information", *European economic Review*, forthcoming.

Quah, Danny, and Thomas J. Sargent (1993). A dynamic index model for large cross sections. In James H. Stock and Mark W. Watson, Eds, *Business Cycles, Indicators, and Forecasting*, N.B.E.R. and University of Chicago Press, Chicago.

Rodrigues, Jorge (2002) "Common and idiosyncratic shocks to the dynamics of exchange rate volatility", February, manuscript.

Sala, Luca (2001), "Monetary transmission in the euro area: a factor model approach", October, manuscript.

Sargent, Thomas J. and Cristopher A. Sims, (1977) "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).

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

Stock, James H., and Mark W. Watson, (1999) "Diffusion indexes", manuscript.

Watson, Mark W. and Robert F. Engle (1983) “Alternative algorithms for the estimation of dynamic factors, MIMIC, and varying coefficient regression models”, *Journal of Econometrics* 23, pp.385-400.

Appendix A

The spectral representation

Any stationary variable can be represented as the integral of waves of different frequency, each having a random amplitude; this is the so called "spectral representation".

$$x_t = \int_{-\pi}^{\pi} e^{i\theta t} dZ(\theta),$$

where $dZ(\theta)$ is an "orthogonal increment process" such that $\text{cov}(dZ(\theta), dZ(\lambda)) = 0$ for $\lambda \neq \theta$.

The spectral density function of a stationary process is defined as:

$$\sigma(\theta) = (1/2\pi) \sum_{h=-\infty}^{\infty} e^{ih\theta} \gamma_h, \quad -\infty < \theta < \infty$$

where γ_h is the covariance function. Conversely, the covariance function can be obtained from the spectral density function using Fourier techniques.

In the multivariate case, there are analogous definitions. The spectral density matrix of a stationary vector process is defined as:

$$\mathbf{\Sigma}(\theta) = (1/2\pi) \sum_{h=-\infty}^{\infty} e^{ih\theta} \mathbf{\Gamma}_h, \quad -\infty < \theta < \infty$$

where $\mathbf{\Gamma}_k$ is the covariance function. The latter can be obtained by Fourier inversion as:

$$\mathbf{\Gamma}_k = \int_{-\pi}^{\pi} e^{i\theta h} \mathbf{\Sigma}(\theta) d\theta.$$

(see e.g. Brockwell and Davis, Chap. 4).

Appendix B

Estimation of the spectral density

The estimation of the spectral density in Forni, Hallin, Lippi and Reichlin's various papers is constructed using a Bartlett lag-window estimator of size $M = M(T)$.

The sample covariance matrix $\mathbf{\Gamma}_k^{(T,n)}$ of $\mathbf{X}_t^{(n)}$ and $\mathbf{X}_{t-k}^{(n)}$ is computed for $k = 0, 1, \dots, M$. Then, the authors compute the $(2M+1)$ points discrete Fourier transform of the truncated two-sided sequence $\mathbf{\Gamma}_{-M}^{(T,n)}, \dots, \mathbf{\Gamma}_0^{(T,n)}, \dots, \mathbf{\Gamma}_M^{(T,n)}$, where $\mathbf{\Gamma}_{-k}^{(n,T} = \mathbf{\Gamma}_k^{(n,T)'}$. More precisely, they compute

$$\mathbf{\Sigma}^{(T,n)}(\theta_h) = \sum_{k=-M}^{M} \mathbf{\Gamma}_k^{(T,n)} \omega_k e^{-ik\theta_h}, \tag{15}$$

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 $M = M(T)$. The choice of M represents the tradeoff between small bias (large M) and small variance (small M). To ensure consistency of $\mathbf{\Sigma}_{(T,n)}(\theta)$ (which is required for the validity of Result FHLR2) the condition $M(T) \to \infty$ and $M(T)/T \to 0$ as $T \to \infty$ must be fulfilled.

In Forni, Hallin, Lippi and Reichlin (2000) a fixed rule $M = \text{round}(\sqrt{T}/4)$ is used and shown to perform well in simulations.

Appendix C

Dynamic principal components and the Forni, Hallin, Lippi and Reichlin, 2000's filter

Then dynamic principal component decomposition are obtained as in Brillinger, 1981. From now on, let us drop the super-script $(n), T$ for notational simplicity. For each frequency of the grid, the eigenvalues and eigenvectors of $\mathbf{\Sigma}(\theta)$ are computed. By ordering the eigenvalues in descending order for each frequency and collecting values corresponding to different frequencies, the eigenvalue and eigenvector functions $\lambda_j(\theta)$ and $\mathbf{p}_j(\theta)$, $j = 1, \ldots, n$, are obtained. The function $\lambda_j(\theta)$ can be interpreted as the (sample) spectral density of the j-th principal component series and, in analogy with the standard static principal component analysis, the ratio

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

represents the contribution of the j-th principal component series to the total variance in the system. Letting $\mathbf{\Lambda}_q(\theta)$ be the diagonal matrix having on the diagonal $\lambda_1(\theta), \ldots, \lambda_q(\theta)$ and $\mathbf{P}(\theta)$ be the $(n \times q)$ matrix $(\mathbf{p}_1(\theta) \ \cdots \ \mathbf{p}_q(\theta))$, the estimate of the spectral density matrix of the vector of the common components $\boldsymbol{\chi}_t = (\chi_{1t} \ \cdots \ \chi_{nt})'$ is given by

$$\mathbf{\Sigma}_{\chi}(\theta) = \mathbf{P}(\theta)\mathbf{\Lambda}(\theta)\tilde{\mathbf{P}}(\theta) \tag{16}$$

where tilde denotes conjugation and transposition.

The filter is computed from the first q eigenvectors $\mathbf{p}_j(\theta_h)$, $j = 1, 2, \ldots, q$, of $\mathbf{\Sigma}(\theta_h)$, for $h = 0, 1, \ldots, 2M$. For $h = 0, 1, \ldots, 2M$, Forni, Hallin, Lippi and Reichlin, 2000 construct

$$\mathbf{K}_i(\theta_h) = \tilde{p}_{1,i}(\theta_h)\mathbf{p}_1(\theta_h) + \cdots + \tilde{p}_{q,i}(\theta_h)\mathbf{p}_q(\theta_h).$$

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

$$\left(\mathbf{K}_i^{(T,n)}(\theta_0), \quad \cdots, \quad \mathbf{K}_i^{(T,n)}(\theta_{2J}) \right),$$

i.e. by the computation of

$$\underline{\mathbf{K}}_{i,k}^{(T,n)} = \frac{1}{2J+1} \sum_{h=0}^{2J} \mathbf{K}_i^{(T,n)}(\theta_h) e^{ik\theta_h}$$

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

$$\underline{\mathbf{K}}_i^{(T,n)}(L) = \sum_{k=-J}^{J} \underline{\mathbf{K}}_{i,k}^{(T,n)} L^k. \tag{17}$$

Appendix D

Estimates of medium and long-run covariances of the components

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

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

(where the super-script $(n), T$ has been dropped for notational simplicity and will be dropped from now on).

Estimates of the covariance matrices of the medium- and long-run component $\boldsymbol{\chi}_t^C = (\chi_{1t}^C, \ldots, \chi_{nt}^C)'$ can be computed by applying the inverse transform to the frequency band of interest, i.e. $[-\theta^*,\ \theta^*]$, where $\theta^* = 2\pi/24$, so that all periodicities shorter than one year are cut off (this is the cut-off point in Altissimo et al., 2001). Precisely, letting $\mathbf{\Gamma}_{\chi^C}(k) = \mathrm{E}\left(\boldsymbol{\chi}_t^C \boldsymbol{\chi}_{t-k}^{C}{}'\right)$, the corresponding estimate will be

$$\mathbf{\Gamma}_{\chi^L}(k) = (2\pi/2H+1) \sum_{h=-H}^{H} \mathbf{\Sigma}_{\chi}(\theta_h) e^{i\theta_h k},$$

where H is defined by the conditions $\theta_H \leq 2\pi/24$ and $\theta_{H+1} > 2\pi/24$.