The generalized dynamic factor model: representation theory

by

Mario Forni *
Marco Lippi **

Febbraio 1999

* Università degli Studi di Modena
Dipartimento di Economia Politica
Via Berengario, 51
41100 Modena (Italia)
e-mail: forni@unimo.it

** Università di Roma
Dipartimento di Scienze Economiche
Via Cesalpino, 12
00161 Roma (Italia)
e-mail: lippi@giannutri.caspur.it
ABSTRACT

This paper, along with the companion paper Forni, Hallin, Lippi and Reichlin (1998), introduces a new model—the **generalized dynamic factor model**—for the empirical analysis of financial and macroeconomic data sets characterized by a large number of observations both cross-section and over time. This model provides a generalization of the static approximate factor model of Chamberlain (1983) and Chamberlain and Rothschild (1983) by allowing serial correlation within and across individual processes, and of the dynamic factor model (or index model) of Sargent and Sims (1977) and Geweke (1977) by allowing for non-orthogonal idiosyncratic terms. While the companion paper concentrates on identification and estimation, here we give a full characterization of the generalized dynamic factor model in terms of observable spectral density matrices, thus laying a firm basis for the empirical implementation of the model.

JEL classification nos.: C13, C33, C43.

Keywords: dynamic factor structure, idiosyncratic components, large cross-sections, Hilbert space.
1. Introduction

1.1 Data sets with many data points both over time and across sections are becoming increasingly available. Think for instance of macroeconomic series on output or employment which are observed for a large number of countries, regions or sectors, or of financial time series such as the returns on many different assets. Such data sets typically present a good deal of regularity along the time dimension, so that each time series, taken in isolation, can be successfully handled by using standard stationary models or their extensions. By contrast, along the cross sectional dimension, data do not have a natural ordering and correlations do not present any regular structure. Yet, the series are strongly dependent on each other, implying that univariate modeling would waste information.

We do not have a satisfactory theoretical framework for extracting and analyzing the enormous amount of information embedded in such large cross sections of time series. VAR models would be suitable for a small subset of time series, but are inadequate for the whole data set, because of the huge number of parameters to estimate. The dynamic factor analytic or index model (Sargent and Sims, 1977, Geweke, 1977) is much better suited, since it is both flexible and parsimonious: each variable is represented as the sum of a common component—i.e. a term depending, possibly with heterogeneous dynamic responses, on a small number of unobserved factors which are common to all variables—and an idiosyncratic component, which is orthogonal at any lead and lag both to the common factors and to the idiosyncratic components of all the other variables.

This feature, mutual orthogonality of the idiosyncratic components at any lead and lag, represents a serious weakness of the index model. The assumption is necessary for identification, but is severely restrictive. As a first example, consider the output of different industries linked to each other by input-output relations. The output of sector $A$ may well be related to the output of sector $B$ in a way which is intimately 'cross-regressive', so that an idiosyncratic shock originated in $B$ propagates, possibly with a lag, to sector $A$. Similar local interactions can also arise when there are 'intermediate' shocks, i.e. shocks which are neither common nor strictly idiosyncratic, such as local events affecting directly more than one area or technological shocks affecting a few sectors. Finally, consider a data set including both employment and income for many regions, and assume that each variable is driven by a national and a regional shock, the second being orthogonal to the first. The regional components of employment and income, while being orthogonal for different regions, are likely to be correlated for the same region. In such a case, although employment, or income, taken in isolation would satisfy the orthogonality assumption, the index model could not be used to handle the whole data set.

In this paper, and in the companion paper Forni, Hallin, Lippi and Reichlin (1998), a new model, that we will call the generalized dynamic factor model, is introduced and analyzed. The model has three important features: (1) it is a finite dynamic factor model, i.e. the variables depend on a finite number of factors with a quite general lag structure; (2) it is based on an infinite sequence of variables and is therefore specifically designed for the analysis of large cross sections of time series; (3) it allows for both contemporaneous and lagged correlation between the idiosyncratic terms, and is therefore more general than the traditional index model.

This paper is part of a research supported by an A.R.C. contract of the Communauté française de Belgique. The first Author also wishes to acknowledge financial aid by MURST (Ministero per l'Università e la Ricerca Scientifica e Tecnologica). We would like to thank Christine De Mol, Marc Hallin and Lucrezia Reichlin for constant help and support.
1.2 Let us briefly summarize the results of the paper. In Section 2 we give our basic definitions and assumptions. We start with a double sequence of stochastic variables \( \{x_{it}, \ i \in \mathbb{N}, \ t \in \mathbb{Z} \} \). We assume that \( \{x_{it}, \ t \in \mathbb{Z} \} \) is stationary for any \( i \) and costationary with \( \{x_{it}, \ t \in \mathbb{Z} \} \) for any \( j \). We do not assume an ARMA structure for the \( x \)'s. We only require the existence of a spectral density matrix \( \Sigma_n^x \) for the vector \( (x_{1t}, x_{2t}, \ldots, x_{nt})' \).

In Section 3 we introduce idiosyncratic sequences. To give a simple illustration of the definition of idiosyncratic sequences adopted here, let us consider a sequence \( \{y_i, \ i \in \mathbb{N} \} \) of mutually orthogonal variables, such that \( \text{var}(y_i) = \sigma^2 \). Taking a sequence of averages \( Y_n = \sum_{i=1}^{n} a_{n} y_i \), the variance \( \text{var}(Y_n) = \sigma^2 \sum_{i=1}^{n} a_{ni}^2 \) tends to zero if and only if \( \sum_{i=1}^{n} a_{ni}^2 \) tends to zero; this occurs typically with the arithmetic mean, \( a_{ni} = 1/n \). Now, the property of a vanishing variance for sequences of averages whose squared weights tend to zero does not require that the \( y \)'s be mutually orthogonal: for example, if \( y_i \) and \( y_j \) are correlated with the correlation declining as \( |i - j| \), then the \( \text{var}(Y_n) \) vanishes asymptotically. This vanishing variance of averages, not orthogonality, is precisely what we need in our construction. Thus, in our definition, the sequence of the \( x \)'s is idiosyncratic if convergence to zero occurs for any weighted average, both cross-section and over time,

\[
\sum_{i=1}^{n} \sum_{h=-k}^{k} a_{ni} x_{it-h},
\]

provided that the sum of the squared weights tends to zero. We prove, Theorem 1, that \( x_{it} \) is idiosyncratic if and only if the maximum eigenvalue of \( \Sigma_n^x \) is dominated by an essentially bounded function of the frequency \( \theta \).

In Section 4 we introduce our generalized dynamic factor model, i.e. a sequence \( \{x_{it}, \ i \in \mathbb{N}, \ t \in \mathbb{Z} \} \) such that

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

where \( b_{ij}(L) \) is a square-summable filter, \( (u_{1t}, u_{2t}, \ldots, u_{qt})' \) is an orthonormal vector white noise, \( \xi_{it} \) is idiosyncratic and orthogonal to the \( u \)'s at any lead and lag, with the filters \( b_{ij}(L) \) fulfilling a condition ensuring that no representation with a smaller number of "common factors" is possible. We prove in Theorem 2 that a sequence has a generalized dynamic factor structure with \( q \) factors if and only if: (I) the \((q+1)\)-th eigenvalue of \( \Sigma_n^x \), in decreasing order, is dominated for any \( n \) by an essentially bounded function of the frequency \( \theta \); (II) as \( n \) tends to infinity, the \( q \)-th eigenvalue diverges for \( \theta \) almost everywhere in \([-\pi, \pi]\).

Thus the unobservable factor structure is completely characterized in terms of properties of the observable matrices \( \Sigma_n^x \). This result has a very important consequence for empirical analysis, as it provides the theoretical basis for heuristic criteria or formal tests in which the sequence of nested matrices \( \Sigma_n^x \) is employed to determine whether the model has a finite dynamic factor structure and what is the number of factors. More precisely, evidence in favor of conditions (I) and (II), with the eigenvalues computed from estimated spectral density matrices, can be interpreted, given the "if" part of Theorem 2, as evidence that, firstly, the variables follow a generalized dynamic factor model, and, secondly, that the number of factors is \( q \).

Theorems 3 and 4 establish uniqueness of the idiosyncratic component \( \xi_{it} \) and of the common component \( x_{it} - \xi_{it} \). It must be pointed out that this identification result holds for the whole infinite sequence of the variables \( x_{it} \), not for its finite subsets: otherwise stated,
identification occurs in the limit, when the size of the cross-section tends to infinity. Moreover, note that identification of $\chi_{it}$ does not imply identification of the $u's$ or of the filters $b_i(L)$, that might be achieved only by imposing further, economically motivated, restrictions. Such an issue will not be discussed in this paper. Finally, in Theorem 5 we show that the common component of $x_{it}$ can be recovered as the limit of the projection of $x_{it}$ on the dynamic principal components. This result provides a firm basis for estimation theory.

The case in which the $x's$ are either difference or trend stationary is shortly discussed in Section 5. Finally, some results that appear as common knowledge but for which no easy reference was available are gathered in the Appendix.

1.3 In spite of some inevitable overlapping this paper is complementary with respect to the companion paper mentioned above. The latter concentrates on identification and estimation of the common and idiosyncratic components, while the main aim of this paper is the full characterization given in Theorem 2. Moreover, the assumption of rational spectral density, made in the companion paper, is dropped completely here.

Correlated idiosyncratic factors, along with infinite cross sectional size, have been introduced in a static model for asset markets by Chamberlain (1983) and Chamberlain and Rothschild (1983). Our Theorem 2 is a generalization to stochastic processes of results proved in the static case by Chamberlain and Rothschild. Related models can also be found in Quah and Sargent (1993), Forni and Reichlin (1996, 1998), Forni and Lippi (1997), and Stock and Watson (1998).

2. Notation, Basic Definitions and Lemmas

Given a complex matrix $D$, we denote by $\bar{D}$ the complex conjugate of the transpose of $D$. Inner product and norm in $\mathbb{C}^s$ are the usual Euclidean entities $(v, w) = \sum_{i=1}^{s} v_i \bar{w_i}$ and $|v| = \sqrt{\sum_{i=1}^{s} |v_i|^2}$ respectively.

Let $\mathcal{P} = (\Omega, \mathcal{F}, P)$ be a probability space and let $L_2(\mathcal{P}, \mathbb{C})$ be the linear space of all complex-valued, zero-mean, square-integrable random variables defined on $\Omega$. We recall that $L_2(\mathcal{P}, \mathbb{C})$, with the inner product defined as $\langle x, y \rangle = E(xy) = \text{cov}(x, y)$, and the norm as $|x| = \sqrt{E(|x|^2)} = \sqrt{\text{var}(x)}$, is a Hilbert space on the complex field $\mathbb{C}$. If $Q$ is a subset of $L_2(\mathcal{P}, \mathbb{C})$ we shall denote by $\text{span}(Q)$ the minimum closed linear subspace of $L_2(\mathcal{P}, \mathbb{C})$ containing $Q$. If $V$ is a closed linear subspace of $L_2(\mathcal{P}, \mathbb{C})$ and $x \in L_2(\mathcal{P}, \mathbb{C})$, we denote by $\text{proj}(x|V)$ the orthogonal projection of $x$ on $V$.

Now consider a double sequence $x = \{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$, where $x_{it} \in L_2(\mathcal{P}, \mathbb{C})$. We adopt the following notation: $X = \text{span}(x)$; $x_{nt}$ is the $n$-dimensional column vector $(x_{1t}, x_{2t}, \ldots, x_{nt})'$. Often, when no confusion can arise, we speak of the process $z_t$, meaning the process $\{z_t, t \in \mathbb{Z}\}$. We also speak of the spectral density of $z_t$. Moreover, considering the $m$-dimensional vector process $Y = \{(y_{1t}, y_{2t}, \ldots, y_{mt})', t \in \mathbb{Z}\}$, we say that $y$ belongs to $Y \subseteq L_2(\mathcal{P}, \mathbb{C})$ if $y_{jt}$ belongs to $Y$ for any $j$ and $t$. In the same way, we use $\text{span}(y)$ to indicate $\text{span}\{y_{jt}, j = 1, \ldots, m, t \in \mathbb{Z}\}$.

**Assumption 1.** For any $n \in \mathbb{N}$: (1) the process $x_{nt}$ is covariance stationary; (2) the spectral measure of $x_{nt}$ is absolutely continuous (with respect to the Lebesgue measure on $[-\pi, \pi]$), i.e. $x_{nt}$ has a spectral density (see Rozanov, 1967, pp. 19-20).
We denote by $\Sigma_n^x$ the spectral density matrix of $x_{nt}$ and recall that $\Sigma_n^x$ is Hermitian, non-negative definite for any $\theta \in [-\pi, \pi]$, Lebesgue-measurable, and that $\int_{-\pi}^{\pi} \Sigma_n^x(\theta)d\theta$ is equal to the variance-covariance matrix of $X_{nt}$.

**Remark 1.** Given $y$ and $z$ in $X$, by definition $||y - z||^2 = \text{var}(y - z)$. On the other hand, we can define in a natural way two stochastic processes $y_t$ and $z_t$ such that $||y_t - z_t||^2 = \int_{-\pi}^{\pi} f(\theta)d\theta$, where $f$ is the spectral density of $y_t - z_t$. For, the definition of $X$ implies the existence of square-summable filters $a_{nj}(L)$ and $b_{nj}(L)$ such that $\sum_{n=1}^{\infty} \sum_{j=1}^{\infty} a_{nj}(L)x_{jn}$ and $\sum_{n=1}^{\infty} \sum_{j=1}^{\infty} b_{nj}(L)x_{jn}$. The processes $y_t$ and $z_t$ are stationary and costationary with $X_{nt}$ for any $n \in \mathbb{N}$. Moreover, $y = y_0$, $z = z_0$, and $||y - z||^2 = \int_{-\pi}^{\pi} f(\theta)d\theta$, where $f$ is the spectral density of $y_t - z_t$ (for the existence of the spectral density of $y_t$, $z_t$ and $y_t - z_t$, see the Appendix, Fact S). With the above argument in mind, the generic element of $X$ will be often referred to as $y_t$, $z_t$, etc., rather than $y$, $z$, etc., where $y_t$, $z_t$, etc. are stationary and costationary with $x_{nt}$, for any $n$.

**Definition 1.** For $i = 1, \ldots, n$ let

$$\lambda_{ni}^x : [-\pi, \pi] \mapsto \mathbb{R}$$

be defined as the function associating with $\theta \in [-\pi, \pi]$ the $i$-th eigenvalue, in descending order, of $\Sigma_n^x(\theta)$. The functions $\lambda_{ni}^x$ will be called the **dynamic eigenvalues** of $\Sigma_n^x$.

In the Appendix we prove the following lemmas.

**Lemma 1.** The functions $\lambda_{ni}^x$ are Lebesgue-measurable in $[-\pi, \pi]$ for any $n \in \mathbb{N}$ and $i \leq n$ (in Remark 3 below we show that $\lambda_{ni}^x$ has finite integral).

**Lemma 2.** Given $i$, for $n \geq i$, $\lambda_{ni}^x(\theta)$ is non-decreasing as a function of $n$ for any $\theta \in [-\pi, \pi]$, i.e. $\lambda_{ni}^x(\theta) \leq \lambda_{n+i}^x(\theta)$.

A consequence of Lemma 2 is that $\lim_n \lambda_{ni}^x(\theta)$ exists for any $i$ and $\theta$, and equals $\sup_n \lambda_{ni}^x(\theta)$.

**Definition 2.** For any $i$ we define the function $\lambda_i^x$ by

$$\lambda_i^x(\theta) = \sup_n \lambda_{ni}^x(\theta).$$

It must be pointed out that $\lambda_i^x$ is an extended real function, i.e. its value may be infinite. Note also that $\{ \theta : \lambda_i^x(\theta) = \infty \}$ may be of null or positive measure, and even coincide with $[-\pi, \pi]$.

Now consider the space $L_2^x([-\pi, \pi], \mathbb{C})$ of all functions $f : [-\pi, \pi] \mapsto \mathbb{C}^n$, with

$$f(\theta) = (f_1(\theta) \quad f_2(\theta) \quad \cdots \quad f_n(\theta))$$

4
(a row vector) and \( f_i : [-\pi, \pi] \rightarrow \mathbb{C} \), such that
\[
\int_{-\pi}^{\pi} |f(\theta)|^2 d\theta = \sum_{i=1}^{n} |f_i(\theta)|^2 d\theta < \infty.
\]

With the inner product of \( f \) and \( g \) and the norm of \( f \) defined respectively as
\[
(f, g) = (2\pi)^{-1} \int_{-\pi}^{\pi} f(\theta)g(\theta) d\theta \quad \text{and} \quad ||f|| = \sqrt{(2\pi)^{-1} \int_{-\pi}^{\pi} |f(\theta)|^2 d\theta},
\]
\( L^2([-\pi, \pi], \mathbb{C}) \) is a Hilbert space. In the Appendix we prove the following results.

**Lemma 3.** There exist \( n \) functions \( p_{ni}^x, i = 1, \ldots, n, \) belonging to \( L^2([-\pi, \pi], \mathbb{C}) \), such that

1. \( |p_{ni}^x(\theta)| = 1, \) for any \( \theta \in [-\pi, \pi] \);
2. \( p_{ni}^x(\theta)p_{nj}^x(\theta) = 0, \) for \( i \neq j \) and any \( \theta \in [-\pi, \pi] \);
3. \( p_{ni}^x(\theta)\Sigma_{ni}^x(\theta) = \lambda_{ni}^x(\theta)p_{ni}^x(\theta) \) for any \( \theta \in [-\pi, \pi] \).

**Definition 3.** An \( n \)-tuple of functions \( p_{ni}^x \) fulfilling (1), (2) and (3) of Lemma 3 will be called a set of dynamic eigenvectors.

Given \( f \in L^2([-\pi, \pi], \mathbb{C}) \), consider the Fourier expansion
\[
f(\theta) = \sum_{k=\infty} \mathbf{F}_k e^{-ik\theta},
\]
where \( \mathbf{F}_k = \int_{-\pi}^{\pi} f(\theta)e^{ik\theta} d\theta \in \mathbb{C}^n \). We shall indicate by \( f(L) \) the square summable \( n \)-dimensional filter
\[
\sum_{k=\infty} \mathbf{F}_k^T L^k.
\]
We have
\[
\sum_{k=\infty} |\mathbf{F}_k|^2 = (2\pi)^{-1} \int_{-\pi}^{\pi} |f(\theta)|^2 d\theta = ||f||^2.
\]
In particular, \( p_{nj}^x(L) \) is the filter associated with the dynamic eigenvector \( p_{nj}^x : [-\pi, \pi] \rightarrow \mathbb{C}^n \).

Conversely, if
\[
\mathbf{A}(L) = \sum_{k=\infty} \mathbf{A}_k L^k
\]
is a square summable \( n \)-dimensional filter, consider the function \( \mathbf{A}^\circ \) defined as
\[
\mathbf{A}^\circ(\theta) = \mathbf{A}(e^{-i\theta}) = \sum_{k=\infty} \mathbf{A}_k e^{-ik\theta}.
\]

---

2 We use "dynamic" for eigenvectors and eigenvalues of \( \Sigma^x \) and related matrices to insist on the difference between the dynamic analysis developed here and the static approach, based on the eigenvalues of variance-covariance matrices. For a general treatment of eigenvalues and eigenvectors of spectral density matrices, and related filters, see Brillinger (1981).
Obviously $A^0$ belongs to $L_2^2([-\pi, \pi], \mathbb{C})$. Moreover $A^0(L) = A(L)$.

**Definition 4.** If the functions $p^i_j$, $j = 1, \ldots, n$ form a set of dynamic eigenvectors, then $p^i_j(L)x_{nt}$, $j = 1, \ldots, n$ is a set of **dynamic principal components** of $x_{nt}$.

The main results presented below will employ the filters $p^i_j(L)$, derived from the dynamic eigenvectors. In general, being obtained from the Fourier expansion of expressions involving polynomial roots, such filters are bilateral and infinite, even when the vector process $x_{nt}$ is a finite moving average.

### 3. Dynamic averaging sequences, aggregation space, idiosyncratic variables

**Definition 5.** Consider a sequence of positive integers $\{s_n$, $n \in \mathbb{N}\}$ and a sequence

$$\{a_n(L), n \in \mathbb{N}\},$$

where

$$a_n^0 = (a_{n1}^0 \ a_{n2}^0 \ \cdots \ a_{ns_n}^0) \in L_2^s([-\pi, \pi], \mathbb{C}).$$

We say that $\{a_n(L), n \in \mathbb{N}\}$ is a **dynamic averaging sequence**, DAS henceforth, if

$$\lim_{n \rightarrow \infty} ||a_n^0||^2 = \lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} |a_{n}^0(\theta)|^2 d\theta = 0.$$

**Definition 6.** Let $y_t \in X$. We say that $y_t$ is an **aggregate** if there exists a DAS $\{a_n(L), n \in \mathbb{N}\}$ such that

$$\lim_{n \rightarrow \infty} a_n(L)x_{s_n t} = y_t.$$

The set of all the aggregates will be denoted by $\mathcal{G}(X)$ and called the **aggregation subspace** of $X$.

**Remark 2.** In Assumption 1 and in Definition 5 we have not supposed that the entries of $\Sigma_n^x$ or of $a_n^0$ are bounded in $[-\pi, \pi]$. As a consequence, the elements of the sequence $a_n(L)x_{s_n t}$ does not necessarily have finite variance, i.e. the integrals $\int_{-\pi}^{\pi} a_n^0(\theta)\Sigma_n^x(\theta)\tilde{a}_n^0(\theta)d\theta$ are not necessarily finite, which means that $a_n(L)x_{s_n t}$ is not necessarily an element of $L_2(\mathcal{P}, \mathbb{C})$. However, convergence of $\{a_n(L)x_{s_n t}, n \in \mathbb{N}\}$ has an obvious definition: $\{a_n(L)x_{s_n t}, n \in \mathbb{N}\}$ converges if $a_n(L)x_{s_n t}$ has finite variance for $n$ greater than some $n_1$ and if the sequence $\{a_n(L)x_{s_n t}, n \in \mathbb{N}$, $n > n_1\}$ converges.

**Remark 3.** Dynamic averaging of $X$, according to Definition 5, is nothing other than averaging simultaneously both in the cross-section and the time dimension. Precisely

$$a_n(L)x_{s_n t} = \sum_{i=1}^{s_n} a_{ni}(L)x_{it} = \sum_{i=1}^{s_n} \sum_{k=-\infty}^{\infty} a_{nik}x_{it-k},$$

where the coefficients $a_{nik}$ are complex numbers fulfilling the condition

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{s_n} \sum_{k=-\infty}^{\infty} |a_{nik}|^2 = 0.$$
Infinite averaging in the time dimension is convenient since, as anticipated above, the averaging sequences that we are going to employ are typically infinite, but not strictly necessary. It is easily seen that the same aggregation space would be obtained by taking finite averages of any length in the time dimension

$$\sum_{i=1}^{s_n} \sum_{k=-\rho_n}^{\tau_n} a_{nki} x_{it-k},$$

with the condition

$$\lim_n \sum_{i=1}^{s_n} \sum_{k=-\rho_n}^{\tau_n} |a_{nki}|^2 = 0.$$

**Lemma 4.** The set $G(x)$ is a closed subspace of $X$.

**Proof.** Assume that $z_t = \lim_m y_{mt}$, with $y_{mt} \in G(x)$. Let $y_{mt} = \lim_n a_{mn}(L)x_{smn}$, where

$$\{a_{mn}(L), n \in \mathbb{N}\}$$

is a DAS. Let $m_i$ be such that $||z_t - y_{m_i,t}|| < 1/i$ and $n_i$ such that $||a^2_{m_i,n_i}|| < 1/i$ and $||y_{m_i,t} - a_{m_i,n_i}(L)x_{s_{m_i,n_i}}|| < 1/i$. The sequence

$$(a_{m_1,n_1}(L) \ a_{m_2,n_2}(L) \ \ldots)$$

is a DAS and

$$||z_t - a_{m_i,n_i}(L)x_{s_{m_i,n_i}}|| \leq ||z_t - y_{m_i,t}|| + ||y_{m_i,t} - a_{m_i,n_i}(L)x_{s_{m_i,n_i}}|| < 2/i.$$

**QED**

**Definition 7.** Suppose that $x$ fulfills Assumption 1. We say that $x$ is idiosyncratic if

$$\lim_n a_n(L)x_{s_n} = 0$$

for any DAS $\{a_n(L), n \in \mathbb{N}\}$.

If $x$ is idiosyncratic then obviously $G(x) = \{0\}$. However, as the next example shows, the converse does not hold.

**Example 1.** Assume that $x_{it} \perp x_{jt-k}$ for any $i \neq j$ and any $k \in \mathbb{Z}$, that $x_{it}$ is a white noise for any $i$, and that $||x_{it}||^2 = i$. Define

$$c_n(L) = \frac{1}{\sqrt{i}} (0 \ 0 \ \ldots \ 0 \ 1).$$

The sequence $\{c_n(L), n \in \mathbb{N}\}$ is a DAS. Moreover $||c_n(L)x_{nt}||^2 = 1$, so that $x$ is not idiosyncratic. Now let $y_t$ be an aggregate, so that

$$y_t = \lim_n a_n(L)x_{s_n,t} = \lim_n \sum_{j=1}^{s_n} a_{nj}(L)x_{jt} = \lim_n \sum_{j=1}^{s_n} \sum_{k=-\infty}^{\infty} a_{njk} x_{jt-k}, \quad (1)$$

where $\{a_n(L), n \in \mathbb{N}\}$ is a DAS. Since $y_t \in X$ and the $x_{it}$'s are mutually orthogonal white noises, then

$$y_t = \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} b_{jk} x_{it-k}, \quad (2)$$

7
Moreover, representations (1) and (2) are unique and $a_{n,jk} = b_{jk}$ for any $j$ and $k$. On the other hand,
\[ \lim_{n} \int_{-\pi}^{\pi} |a_n^\alpha(\theta)|^2 d\theta = \lim_{n} \left[ \sum_{j=1}^{s_n} \sum_{k=-\infty}^{\infty} |a_{n,jk}|^2 \right] = 0, \]
so that $b_{jk} = 0$ for any $j$ and $k$, i.e. $y_t = 0$. Thus $\mathcal{G}(x) = \{0\}$ although $x$ is not idiosyncratic.

If the vector $x_{nt}$ is a white noise for any $n$, i.e. if the matrix $\Sigma^x_n$ and its eigenvalues are constant as functions of $\theta$, then $x$ is idiosyncratic if and only if $\lambda_n^{x^2}$ is bounded as a function of $n$ (see Chamberlain, 1983, Chamberlain and Rothschild, 1983). The theorem below generalizes this result to any $x$ fulfilling Assumption 1: $x$ is idiosyncratic if and only if the functions $\lambda_n^{x^2}$ are uniformly bounded in $[-\pi, \pi] - D$, where $D$ has null Lebesgue measure, i.e. if there exists a real $M$ such that $\lambda_n^{x^2}(\theta) \leq M$ for any $n \in \mathbb{N}$ and $\theta \in [-\pi, \pi] - D$.

In the sequel $\mathcal{L}$ will denote the Lebesgue measure on $\mathbb{R}$. Let us recall that an extended real function $f$ is essentially bounded if there exists a real $c$ and a subset $D$ of $[-\pi, \pi]$ such that $\mathcal{L}(D) = 0$ and $|f(\theta)| \leq c$ for $\theta \in [-\pi, \pi] - D$ (Halmos, 1958, p. 86). Obviously, if $f$ is essentially bounded the set where $f$ is not finite has null measure.

**Remark 4.** If $c \in L^2([-\pi, \pi], \mathcal{C})$ and $|c(\theta)|^2$ is essentially bounded, then, since the entries of $\Sigma^x_n$ have finite integral, the variance of $\mathcal{G}(L)x_{nt}$ is finite. In particular,
\[ \text{var}(\mathbf{p}_{(n)}^{x^2}(L)x_{nt}) = \int_{-\pi}^{\pi} \mathbf{p}_{(n)}^{x^2}(\theta)\Sigma^x_n(\theta)\mathbf{p}_{(n)}^{x^2}(\theta) d\theta = \int_{-\pi}^{\pi} \lambda_n^{x^2}(\theta) d\theta < \infty. \]

We will invoke the following two results, the first known as the Lebesgue dominated convergence theorem.

**Fact L1.** Assume that $\{f_n, n \in \mathbb{N}\}$ is a sequence of integrable functions (i.e. having finite integral) defined on $[-\pi, \pi]$, such that (a) $\lim_n f_n(\theta) = f(\theta)$ a.e. in $[-\pi, \pi]$, and (b) $|f_n(\theta)| \leq g(\theta)$ a.e. in $[-\pi, \pi]$, where $g$ is non-negative and integrable. Then $f$ is integrable and
\[ \lim_{n} \int_{-\pi}^{\pi} f_n(\theta) d\theta = \int_{-\pi}^{\pi} f(\theta) d\theta \]
(see Apostol, 1974, p. 270).

**Fact L2.** Suppose that $\{f_n, n \in \mathbb{N}\}$ is a sequence of integrable functions defined on $[-\pi, \pi]$, that $f$ is an integrable function defined on $[-\pi, \pi]$, and that
\[ \lim_{n} \int_{-\pi}^{\pi} |f_n(\theta) - f(\theta)|^2 d\theta = 0 \quad \text{or} \quad \lim_{n} \int_{-\pi}^{\pi} |f_n(\theta) - f(\theta)| d\theta = 0. \]

Then there exists an increasing sequence $s_n$ such that $\lim_n f_{s_n}(\theta) = f(\theta)$ a.e. in $[-\pi, \pi]$ (see Apostol, 1974, p. 298; Halmos, 1958, p. 103, Theorem A, p. 93, Theorem D, p. 89, Theorem B).

**Theorem 1.** $x$ is idiosyncratic if and only if $\lambda_n^{x^2}$ is essentially bounded.

**Proof.** Let $\{a_n(L), n \in \mathbb{N}\}$ be a DAS. By Fact M1 (Appendix),
\[ ||a(L)n_{x_{n,t}}||^2 = \int_{-\pi}^{\pi} a_n^\alpha(\theta)\Sigma^x_n(\theta)a_n^\alpha(\theta) d\theta \leq \int_{-\pi}^{\pi} \lambda_n^{x^2}(\theta)|a_n^\alpha(\theta)|^2 d\theta \leq \int_{-\pi}^{\pi} \lambda_n^{x^2}(\theta)|a_n^\alpha(\theta)|^2 d\theta. \]
If $\lambda^x_1$ is essentially bounded, the RHS of (3) tends to zero, so that $x$ is idiosyncratic.

Conversely, assume that $\lambda^x_1$ is not essentially bounded. This means that for any positive integer $n$, there exists an integer $s_n$ such that

$$\mu_n = L(\{ \theta : \lambda^x_{s_n}(\theta) \geq n \}) > 0.$$ 

Define $h_n$ by

$$h_n(\theta) = \begin{cases} 1/\sqrt{\mu_n} & \text{if } \lambda^x_{s_n}(\theta) \geq n \\ 0 & \text{otherwise}, \end{cases}$$

and $H_n = \int_{-\pi}^\pi \lambda^x_{s_n}(\theta) h_n(\theta)^2 d\theta$, so that $F_n \geq n$. Then define $b_n(L) = h_n(L) \mathbf{p}^x_{s_n}(L) / \sqrt{H_n}$. Clearly $b_n(L)$ is a DAS, while $||b_n(L)x_{s_n,t}||^2 = 1$, so that $x$ is not idiosyncratic. QED

**Corollary.** If $x$ is idiosyncratic then

$$\sup_n \int_{-\pi}^\pi \lambda^x_{s_n}(\theta)d\theta = \lim_n \int_{-\pi}^\pi \lambda^x_{s_n}(\theta)d\theta < \infty.$$ 

**Proof.** Since $\lambda^x_1$ is essentially bounded, we have $\int_{-\pi}^\pi \lambda^x_1(\theta)d\theta < \infty$. $\lambda^x_{s_n}$ converges a.e. in $[-\pi, \pi]$ to $\lambda^x_1$ and is bounded a.e. in $[-\pi, \pi]$ by $\lambda^x_1$. Applying Fact L1,

$$\lim_n \int_{-\pi}^\pi \lambda^x_{s_n}(\theta)d\theta = \int_{-\pi}^\pi \lambda^x_1(\theta)d\theta < \infty.$$ 

QED

The following example shows that the converse of the Corollary is false.

**Example 2.** Assume that $x_{it}$ is orthogonal to $x_{jt-k}$ for any $k$ and any $i \neq j$, and suppose that the spectral density of the stationary process $x_{it}$ is $|\theta|^\alpha$, for $|\theta| > 1/i$, zero otherwise, with $-1 < \alpha < 0$. In this case the matrix $\Sigma^x_n$ is diagonal, $\lambda^x_1(\theta) = |\theta|^\alpha$ for $\theta \neq 0$, zero for $\theta = 0$, which is not essentially bounded. Thus $x$ is not idiosyncratic, even though $\sup_n \int_{-\pi}^\pi \lambda^x_{s_n}(\theta)d\theta = \int_{-\pi}^\pi |\theta|^\alpha d\theta < \infty$.

**Definition 8.** Assume that $x$ fulfills Assumption 1. Consider the orthogonal projection

$$x_{it} = \text{proj}(x_{it} | G(x)) + \delta_{it}. \quad (4)$$

Decomposition (4) will be called the canonical decomposition of $x$.

4. A Finite Number of Dynamic Common Factors

4.1 Let us now give a formal definition of the generalized dynamic factor model and now state our main results.

**Definition 9.** Let $q$ be a non-negative integer. The double sequence $x$ is a $q$-dynamic factor sequence, $q$-DFS henceforth, if $L_2(\mathcal{P}, \mathbb{C})$ contains an orthonormal $q$-dimensional white-noise vector process

$$u = \{(u_{1t}, u_{2t}, \ldots, u_{qt})', \ t \in \mathbb{Z}\} = \{u_t, \ t \in \mathbb{Z}\},$$

and a double sequence $\xi = \{\xi_{it}, \ i \in \mathbb{N}, \ t \in \mathbb{Z}\}$ fulfilling Assumption 1, such that:
(i) for any $i \in \mathbb{N}$,

$$x_{it} = X_{it} + \xi_{it}$$

$$X_{it} = b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} = b_i(L)u_t,$$

where $b_i \in L_2([\pi, \pi], \mathbb{C}^q)$;

(ii) $\lambda_i^f$ is essentially bounded, i.e. $\xi$ is idiosyncratic;

(iii) putting $\chi = \{X_{it}, \; i \in \mathbb{N}, \; t \in \mathbb{Z}\}$, $\lambda_q^x(\theta) = \infty$ a.e. in $[-\pi, \pi]$.

The double sequences $\chi$ and $\xi$ are referred to as the common and the idiosyncratic component of representation (5).

Theorem 2. The double sequence $\chi$ is a q-DFS if and only if:

(I) $\lambda_{i+1}^f$ is essentially bounded;

(II) $\lambda_q^x = \infty$ a.e. in $[-\pi, \pi]$.

Remark 5. Forni, Hallin, Lippi and Reichlin (1998) propose a heuristic criterion to determine in empirical cases the number $q$ such that (I) and (II) hold. Since they only rely on the 'only if' part of Theorem 2, their criterion provides evidence on the number of common factors, under the assumption of a generalized dynamic factor model. Once the 'if' part is proved, evidence that for some $q$ (I) and (II) hold becomes evidence both that the series follow a generalized dynamic factors, and that the number of factors is $q$. Theorem 3. If $\chi$ is a q-DFS with representation (5) then

$$\text{span}(\chi) = \text{span}(u) = \mathcal{G}(x).$$

Moreover

$$X_{it} = \text{proj}(x_{it}|\mathcal{G}(x)).$$

An immediate but very important consequence of (6) is that if $\chi$ is a q-DFS then the components $X_{it}$ and $\xi_{it}$ are uniquely determined. Precisely:

Theorem 4. Suppose that $\chi$ is a q-DFS with representation (5). Suppose further that there exists an $s$-dimensional orthonormal white-noise vector process $v$, with $v_{jt} \in L_2(\mathcal{P}, \mathbb{C})$, such that

$$x_{it} = \omega_{it} + \zeta_{it}$$

$$\omega_{it} = c_i(L)v_i,$$

where $c_i^t \in L_2([\pi, \pi], \mathbb{C}^s)$, and that $\lambda_i^f$ and $\lambda_q^x$ fulfill, respectively, conditions (ii) and (iii) of Definition 9. Then $s = q$, $\omega_{it} = X_{it}$ and $\zeta_{it} = \xi_{it}$.

Remark 6. It must be pointed out that the components are identified, not $u_t$ or the filters $b_i(L)$. If (5) holds, and $C(L)$ is $q \times q$ and such that $C^o(\theta)C^o(\theta) = I_q$, then $\chi_{it} = b_i^*(L)u_i^*$, with $b_i^*(L) = b_i(L)C(L)$ and $u_i^* = C(L^{-1})u_t$.

Remark 7. Note that in Definition 9 the filters $b_{ij}(L)$ are in general bilateral. If representation (5) must have a structural interpretation then it is reasonable to assume that the filters $b_{ij}(L)$ are one-sided. However, one-sidedness of the $b_{ij}(L)$ has no consequences on the eigenvalues $\lambda_{nj}$, nor fulfillment of conditions (I) and (II) has implications on the existence of one-sided representations of the common component. In this paper we deal only with
the number of common shocks, i.e. the dimension of $u_t$, which is identified (Theorems 2, 3, 4), and with the reconstruction of $x_{it}$ and $\xi_t$ (Theorem 5). Existence and identification of one-sided representations of the common component are left to further study.

**Remark 8.** The result in Theorem 4 can be restated by saying that if $x$ is $q$-DFS, then $q$ is minimal, i.e. no representation fulfilling Definition 9 is possible with a smaller number of factors. It is important to point out that condition (iii) in Definition 9 is crucial. For example, suppose that $x_{it} = b_i u_t + \xi_{it}$, with $\xi$ idiosyncratic and $\sum |b_i|^2 < \infty$. In this case $\lambda_{ij}^x < \infty$. As a consequence, $b_i u_t + \xi_{it}$ is idiosyncratic, so that a representation with zero factor is possible.

**Remark 9.** Suppose that $x_{nt}$ is a vector white noise for any $n$, so that the model is “isomorphic” to the static model in Chamberlain and Rothschild (1983). Then the eigenvalues $\lambda_{n,j}^x$ are constant as functions of $\theta$. As a consequence, if $\lambda_{n,j}^x < \infty$, the model has $q$ factors, with $q < s$. Unfortunately, in the general dynamic case, there exist cases where $\lambda_{n,j}^x$ is essentially bounded, but the sequence does not fulfill Definition 9 for any $q < s$. Consider

$$x_{it} = b(L)u_t + \xi_{it},$$

with $\xi$ idiosyncratic and

$$b^\circ(\theta) = \begin{cases} 1 & \text{if } \theta \in [-1,1] \\ 0 & \text{otherwise.} \end{cases}$$

In this case $\lambda_{n,j}^x(\theta)$ is essentially bounded, but $\lambda_{n,j}^x(\theta)$ is infinite for $\theta \in [-1,1]$, finite elsewhere. However, such cases do not seem to deserve further consideration.

The proof of Theorems 2 and 3 will require several steps. In Section 4.2 we introduce an additional assumption on $x$ and show that it does not imply any loss of generality. In Section 4.3 we prove that conditions (I) and (II) are necessary for a $q$-DFS, which is relatively easy. The converse is much more complicated. In 4.4 we prove that $G(x)$ contains a $q$-dimensional orthonormal white-noise vector process $z$, so that $G(x) \supseteq \text{span}(z)$. In 4.5 we prove that actually $G(x) = \text{span}(z)$, so that the canonical decomposition has the form

$$x_{it} = \text{proj}(x_{it} | G(x)) + \delta_{it} = c_i(L)z_i + \delta_{it}.$$ 

Lastly, in 4.6 we show that $\delta$ is idiosyncratic, thus completing the proof of Theorem 2. In 4.7 we prove Theorem 3.

**4.2 Theorems 2 and 3 will be proved supposing that**

**Assumption 2.** For any $n, j \leq n$ and $\theta \in [-\pi, \pi]$, $\lambda_{n,j}^x(\theta) \geq 1$.

To show that Assumption 2 does not imply any loss of generality, observe that, possibly by embedding $P$ into a larger probability space, we can assume that $L_2(P, C)$ contains a stationary sequence $\{\xi_{it}, i \in N, t \in Z\}$ such that $\xi_{it} \perp X$ for any $i$ and $t$, $\text{var}(\xi_{it}) = 1$ for any $i$ and $t$, and $\xi_{it} \perp \xi_{jt-k}$ for any $t$ and $i \neq j$. Now define $y_i = \{x_{it} + \xi_{it}, i \in N, t \in Z\}$, and suppose that Theorems 2 and 3 have been proved under Assumption 2. We have:

(a) $\Sigma_n^y = \Sigma_n^x + I_n$, $\lambda_{n,j}^y = \lambda_{n,j}^x + 1$. Thus if conditions (I) and (II) hold for $x$, then they hold for $y$ as well. By Theorem 2 $y$ is a $q$-DFS with representation $y_{it} = \chi_{it} + \xi_{it}$. By Theorem...
3. \( \tilde{x}_{it} = \text{proj}(y_{it}|G(y)) \). But the definitions of \( \tilde{\xi} \) and \( y \) imply that \( \tilde{x}_{it} = \text{proj}(x_{it}|G(x)) \).

Therefore

\[
x_{it} = \text{proj}(x_{it}|G(x)) + (\tilde{\xi}_{it} - \tilde{\xi}_{it}).
\]

Since \( \tilde{\xi}_{it} \) is orthogonal to \( X \) and \( \tilde{\xi} - \tilde{\xi} \) is idiosyncratic, then (7) is a q-DFS representation. Therefore (I) and (II) hold for \( x \), then \( x \) has a q-DFS representation.

(b) If \( x \) has the q-DFS representation \( x_{it} = \chi_{it} + \xi_{it} \), then \( y \) has the q-DFS representation \( y_{it} = \chi_{it} + (\xi_{it} + \xi_{it}) \). Applying Theorem 2 to \( y \), we obtain conditions (I) and (II) for \( \lambda_q^x = \lambda_q^x + 1 \) and \( \lambda_{q+1}^x = \lambda_{q+1}^x + 1 \) and therefore for \( \lambda_q^x \) and \( \lambda_{q+1}^x \). In conclusion, if Theorems 2 and 3 hold under Assumption 2, then Theorem 2 holds in general.

(c) In the same way, applying Theorems 2 and 3, supposedly proved under Assumption 2, to \( y \), Theorem 3 can be proved in general.

4.3 Let us prove that if \( x \) is a q-DFS then (I) and (II) hold. By Definition 9, \( \Sigma_n^x(\theta) = \Sigma_m^x(\theta) + \Sigma_n^\xi(\theta) \). By Fact M3 (Appendix), \( \lambda_{nq}^x(\theta) \geq \lambda_{nq}^\xi(\theta) \), so that (II) is proved. Moreover, by Fact M3,

\[
\lambda_{nq+1}^x(\theta) \leq \lambda_{nq+1}^x(\theta) + \lambda_{n1}^\xi(\theta) = \lambda_{n1}^\xi(\theta),
\]

so that (I) is proved. Moreover, (8) implies the following interesting inequality:

\[
\lambda_{q+1}^x(\theta) \leq \lambda_1^\xi(\theta)
\]

(9)

(the opposite inequality is proved in 4.7).

4.4 Now we start assuming (I) and (II). Firstly we prove that \( G(x) \) contains a q-dimensional white-noise vector. The proof goes as follows. We start with a q-dimensional orthonormal white noise, call it \( \psi_t \), whose entries are linear combinations of the principal components of \( x_{nt} \), i.e. \( \mathbf{p}_{n,j}^x(L)x_{nt} \), for \( j = 1, \ldots, q, t \in \mathbb{Z} \). Then we project \( \psi_t \) on the space spanned by the principal components of \( x_{nt} \), i.e. \( \mathbf{p}_{n,j}^x(L)x_{nt} \), for \( j = 1, \ldots, q, t \in \mathbb{Z} \), for \( n > m \), call \( y_t \) the projection. We show that when \( m \) and \( n \) become large the distance between \( \psi_t \) and \( y_t \) becomes small. This leads to the construction of a sequence of q-dimensional white noise vectors whose components are Cauchy sequences and converge to \( G(x) \).

The proofs would be considerably easier if we could assume that \( \lambda_{nq}^x(\theta) \geq \alpha_n \) a.e. in \([-\pi, \pi]\), where \( \lim_n \alpha_n = \infty \). However, this condition is false in this 1-factor model:

\[
x_{it} = (1 - L)u_t + \xi_{it},
\]

with \( \Sigma_n^x = I_n \), in which \( \Sigma_n^x \) is continuous and \( \Sigma_n^x(0) = I_n \) for any \( n \). Unfortunately, to include cases like (10) our proofs must be carried over piecewise on \([-\pi, \pi]\).

We need some further notation and definitions. For \( n \geq q \), we denote by \( \mathbf{P}_n \) the \( q \times n \) matrix

\[
(\mathbf{p}_{n,1}^x, \mathbf{p}_{n,2}^x, \ldots, \mathbf{p}_{n,q}^x)',
\]

i.e. the matrix having the dynamic eigenvectors \( \mathbf{p}_{n,j}^x, j = 1, \ldots, q \), on the rows, and by \( \mathbf{Q}_n \) the \( (n-q) \times n \) matrix

\[
(\mathbf{p}_{n,q+1}^x, \mathbf{p}_{n,q+2}^x, \ldots, \mathbf{p}_{nn}^x)',
\]

Moreover, let us call \( \Lambda_n \) the \( q \times q \) diagonal matrix having on the diagonal the eigenvalues \( \lambda_{n,j}^x, j = 1, \ldots, q \) and by \( \Phi_n \) the \( (n-q) \times (n-q) \) diagonal matrix having on the diagonal the
eigenvalues $\lambda_{j}^{e}, j = q + 1, \ldots, n$. The matrices $\Sigma_{n}^{e}$ and $I_{n}$ can be rewritten in their spectral decomposition form:

$$
\Sigma_{n}^{e} = \tilde{P}_{n} A_{n} P_{n} + \tilde{Q}_{n} F_{n} Q_{n},
$$

$$
I_{n} = \tilde{P}_{n} P_{n} + \tilde{Q}_{n} Q_{n}.
$$

Since $A_{n}^{-1}(\theta)$ is bounded in $[-\pi, \pi]$ by Assumption 2, the definition

$$
\psi_{n}^{t} = (\psi_{1}^{n} \cdots \psi_{q}^{n})' = A_{n}^{-1/2}(L)P_{n}(L)X_{nt}
$$

makes sense and $\psi_{n}^{t}$ is an orthonormal white noise. Note that the processes $\psi_{jt}^{n}, j = 1, \ldots, q$, are the first $q$ dynamic principal components of $X_{nt}$, rescaled to get unit spectral density at any frequency.

In the sequel it will be convenient to write matrix products $AB$ in which the number of columns of $A$ is smaller than the number of rows of $B$. In this case we implicitly assume that $A$ has been augmented with columns of zeros to match the number of rows of $B$. For example, we write $P_{m}(L)X_{nt}$ for $n > m$, this meaning nothing other than $P_{m}(L)X_{rnt}$.

Now let $C$ be a $q \times q$ matrix such that the entries $c_{ij}$ belong to $L_{2}([-\pi, \pi], C)$ and are essentially bounded in modulus, so that the linear combination $C(L)\psi_{n}^{t}$ has finite variance-covariance matrix (see Remark 4). We want to determine the (element by element) orthogonal projection of the vector $C(L)\psi_{n}^{t}$ on the space

$$
\text{span} \{ \{ \psi_{jt}, j = 1, \ldots, q, t \in \mathbb{Z} \} \}
$$

for $n > m$. From (11) we get

$$
x_{nt} = \tilde{P}_{n}(L)P_{n}(L)X_{nt} + \tilde{Q}_{n}(L)Q_{n}(L)X_{nt} = \tilde{P}_{n}(L)A_{n}^{-1/2}(L)\psi_{n}^{t} + \tilde{Q}_{n}(L)Q_{n}(L)X_{nt}.
$$

Since $Q_{n}(\theta)\Sigma_{n}^{e}(\theta)\tilde{P}_{n}(\theta) = \Phi_{n}(\theta)Q_{n}(\theta)\tilde{P}_{n}(\theta) = 0$ for any $\theta$, the two terms on the RHS of (12) are orthogonal at any lead and lag element by element, so that the first is the projection of $X_{nt}$ on $\text{span} \{ \{ \psi_{jt}, j = 1, \ldots, q, t \in \mathbb{Z} \} \}$ and the second is the residual. The required projection equation is then obtained by applying on both sides the operator $C(L)A_{n}^{-1/2}(L)P_{m}(L)$ and noting that $A_{m}^{-1/2}(L)P_{m}(L)X_{nt} = A_{m}^{-1/2}(L)P_{m}(L)X_{nt} = \psi_{m}^{t}$, i.e.

$$
C(L)\psi_{n}^{t} = D(L)\psi_{n}^{t} + R(L)X_{nt},
$$

where

$$
D = C A_{m}^{-1/2} P_{m} \tilde{P}_{n} A_{n}^{1/2}, \quad R = C A_{m}^{-1/2} P_{m} \tilde{Q}_{n} Q_{n}.
$$

Note that $D$, as well as $\Delta$, $\mathbf{H}$ and $\mathbf{F}$, which are defined below, depend on $C$, $m$ and $n$. However, as no confusion can arise, we do not explicit this dependence for notational simplicity. The following result holds.

**Lemma 5.** Suppose that (I) and (II) hold. Assume $n > m$ and let $C(\theta)$ be a unitary matrix, i.e. $C(\theta)C(\theta) = I_{q}$ for any $\theta \in [-\pi, \pi]$. Consider the projection equation

$$
C(L)\psi_{n}^{t} = D(L)\psi_{n}^{t} + R(L)X_{nt},
$$

(15)
where $D$ and $R$ are defined as in (14), and call $\mu(\theta)$ the first eigenvalue of the spectral density matrix of the residual $R(L) x_n$. Then $\mu(\theta) \leq \lambda_{nq+1}^x(\theta)/\lambda_{mq}^x(\theta)$.

**Proof.** The matrix $I_n - \bar{Q}_n \Phi_n$ is non-negative definite by (11) and $\lambda_{nq+1}^x \bar{Q}_n \Phi_n \bar{Q}_n$ is non-negative definite by the definition of $\Phi_n$, so that $\lambda_{nq+1}^x I_n - \bar{Q}_n \Phi_n \bar{Q}_n$ is also non-negative definite. Premultiplying by $C \Lambda_m^{-1/2} P_m$ and postmultiplying by $P_m \Lambda_m^{-1/2} \tilde{C}$ it is seen that

$$\lambda_{nq+1}^x C \Lambda_m^{-1} \tilde{C} - R \Sigma_n^x \bar{R}$$

is also non-negative definite. The desired inequality follows from Fact M3 (third and fourth inequality).

QED

Now let us begin the construction of our converging sequence. Note that, under assumptions (I) and (II), there exists a set $\Pi \subseteq [-\pi, \pi]$ and a real $W$ such that $[-\pi, \pi] - \Pi$ has null measure and, for $\theta \in \Pi$: (1) $\lambda_{nq+1}^x(\theta) \leq W$ for any $n \in \mathbb{N}$ and any $\theta \in \Pi$; (2) $\lambda_q^x(\theta) = \infty$ for $\theta \in \Pi$. Obviously, if a statement holds a.e. in $\Pi$, then it holds a.e. in $[-\pi, \pi]$, and vice versa.

Let $M$ be a positive measure subset of $\Pi$ such that $\lambda_{nq}^x(\theta) \geq \alpha_n$ for $\theta \in M$, where $\{\alpha_n, \ n \in \mathbb{N}\}$ is a real positive non-decreasing sequence satisfying $\lim_n \alpha_n = \infty$. Lastly, denote by $K_M$ the set including all of the $q \times q$ matrices $C$ with elements in $L_2([-\pi, \pi], \mathbb{C})$ such that (i) $C(\theta) = 0$ for $\theta \notin M$, (ii) $C(\theta)$ is unitary, i.e. $C(\theta) \bar{C}(\theta) = I_q$, for $\theta \in M$.

Consider (13) and assume $C \in K_M$. Taking the spectral density of both sides we get, for $\theta \in M$,

$$I_q = D \tilde{D} + R \Sigma_n^x \bar{R}.$$  

Applying Lemma 5 we obtain $\mu(\theta) \leq \lambda_{nq+1}^x(\theta)/\lambda_{mq}^x(\theta) < W/\alpha_m$ for $\theta \in M$. Hence by Fact M3, calling $\Delta_j(\theta), j = 1, \ldots, q, \theta$ the eigenvalues of $D(\theta) \tilde{D}(\theta)$ in descending order, we have

$$1 \geq \Delta_j(\theta) > 1 - W/\alpha_m$$

for any $\theta$ in $M$. Thus, if $m^*$ is such that

$$W/\alpha_{m^*} < 1,$$

we have

$$\Delta_j(\theta) > 1 - W/\alpha_{m^*} > 0$$

everywhere in $M$ for any $m \geq m^*$.

Now assume $m > m^*$. Denote by $\Delta$ the diagonal matrix having $\Delta_j$ in place $(j, j)$ and by $H(\theta)$ a matrix which is measurable in $M$ and fulfills for any $\theta \in M$: (a) $H(\theta) \bar{H}(\theta) = I_q$, (b) $H(\theta) \Delta(\theta) \bar{H}(\theta) = D(\theta) \tilde{D}(\theta)$. Inequality (18) ensures that $1/\sqrt{\Delta_j(\theta)}$ is bounded in $M$ for $j = 1, \ldots, q$, so that the definition

$$F(\theta) = \begin{cases} H(\theta) \Delta(\theta)^{-1/2} \bar{H}(\theta)D(\theta) & \text{if } \theta \in M \\ 0 & \text{if } \theta \notin M \end{cases}$$

makes sense. Note that $F$ belongs to $K_M$.

**Lemma 6.** Suppose that (I) and (II) hold. Let $M$ be a positive measure subset of $\Pi$ and $\{\alpha_n, \ n \in \mathbb{N}\}$ a real positive non-decreasing sequence such that $\lim_n \alpha_n = \infty$. Assume that
(a) $C \in K_M$;
(b) $\lambda_{q,n}(\theta) \geq \alpha_n$ for $\theta \in M$;

Then, given $\tau$, such that $2 > \tau > 0$, there exists an integer $m_\tau$ such that, firstly, $W/m_\tau < 1$, and, secondly, for $n > m \geq m_\tau$, the first eigenvalue of the spectral density matrix of

$$C(L)\psi_t^m - F(L)\psi_t^n$$

is less than $\tau$ for any $\theta \in \Pi$, where $F$ is defined as in (19), with $D$ defined as in (14).

**Proof.** From (13) we get

$$C(L)\psi_t^m - F(L)\psi_t^n = R(L)xnt + (D(L) - F(L))\psi_t^n.$$

The terms on the RHS are orthogonal at any lead and lag, so that the spectral density matrix of the sum is equal to the sum of the spectral density matrices. Hence, calling $S$ the spectral density matrix on the LHS and using (16), we see that, for $\theta \in M$,

$$S = 2I_q - DF - FD = 2I_q - 2H^{1/2}\tilde{H} = 2H(I_q - \Delta^{1/2})\tilde{H},$$

whose larger eigenvalue is $2 - 2\sqrt{\Delta_q(\theta)}$, which is less than $2W/\alpha_m$ by (17). Thus, in order for $F$ to make sense and the statement of the lemma to hold we need $2W/\alpha_{m_\tau} < \min(2, \tau)$.

Since $\tau < 2$, $m_\tau$ must fulfill

$$2W/\alpha_{m_\tau} < \tau$$

(20)

QED

We need the following result, whose proof is given in the Appendix. Given the costationary processes $A = \{A_t, t \in \mathbb{Z}\}$, and $B = \{B_t, t \in \mathbb{Z}\}$ we denote by $S(A, B; \theta)$ the value at the frequency $\theta$ of the cross spectrum between $A$ and $B$.

**Lemma 7.** If $A = \{A_{nt}, t \in \mathbb{Z}\}$ and $B = \{B_{nt}, t \in \mathbb{Z}\}$ are costationary for $n \in \mathbb{N}$, $\lim_{n} A_{nt} = A_t$ and $\lim_{n} B_{nt} = B_t$ (in variance), then, for a sequence of integers $s_i$,

$$\lim_i S(A_{s_i}, B_{s_i}; \theta) = S(A, B; \theta),$$

a.e. in $[-\pi, \pi]$.

**Lemma 8.** Suppose that (I) and (II) hold and let $M$ and $\{\alpha_n, n \in \mathbb{N}\}$ be as in Lemma 6. There exists a $q$-dimensional vector process $v$ such that

(a) $v_{jt}$ is an aggregate for $j = 1, \ldots, q$;
(b) the spectral density matrix of $v$ equals $I_q$ for $\theta$ a.e. in $M$, zero for $\theta \notin M$.

**Proof.** Let $F_1$ be any element of $K_M$. Set $\tau = 1/2^3\pi$ and $s_1 = m_\tau$, where $m_\tau$ satisfies (20). Then set $G_1(L) = F_1(L)\Delta_{s_1}^{-1/2}P_{s_1}(L)$ and $v_1^j = G_1(L)x_{nt}$. It is easily seen that the spectral density matrix of $v_1^j$ equals $I_q$ for $\theta \in M$, zero for $\theta \notin M$.

Now set $\tau = 1/2^3\pi$ and $s_2 = m_\tau$, where $m_\tau$ satisfies (20). Then determine $D$ as in (14), with $F_1$ in place of $C$, $s_2$ in place of $n$ and $F_2$ as in (19). Finally set $G_2(L) = F_2(L)\Delta_{s_2}^{-1/2}P_{s_2}(L)$ and $v_2^j = G_2(L)x_{nt}$. The spectral density matrix of $v_2^j$ equals $I_q$ for $\theta \in M$, zero for $\theta \notin M$. Moreover, by the definition of $s_1$ and Lemma 6, calling $A_1$ the first eigenvalue of the spectral density matrix of $v_1^j - v_2^j$, we have $A_1(\theta) < 1/2^3\pi$ for any $\theta \in \Pi$, so that $\|v_1^j - v_2^j\| < 1/2$, for $j = 1, \ldots, q$. 

15
By recursion, set \( \tau = 1/2^{k+1} \pi \) and \( s_k = m_\tau \), where \( m_\tau \) satisfies (20). Then determine \( D \) as in (14), with \( F_{k-1} \) in place of \( C \) and \( s_k \) in place of \( n \), and \( F_k \) as in (19). Finally set \( G_k(L) = F_k(L)A_{n_k}^{-1/2}P_{s_k} \) \( (L) \) and \( v^k_t = G_k(L)X_{kt} \). The spectral density matrix of \( v^k_t \) equals \( I_q \) for \( \theta \in M \), zero for \( \theta \notin M \). Moreover, by the definition of \( s_{k-1} \) and Lemma 6, calling \( A_{k-1} \) the first eigenvalue of the spectral density matrix of \( v^{k-1}_t - v^k_t \), we have \( A_{k-1}(\theta) < 1/2^{2k-1} \pi \) for any \( \theta \in \Pi \), so that \( ||v^{k-1}_t - v^k_t|| < 1/2^{k-1} \) for \( j = 1, \ldots, q \).

Since we have

\[
||v^k_t - v^{k+h}_t|| \leq ||v^k_t - v^{k+1}_t|| + \cdots + ||v^{k+h-1}_t - v^{k+h}_t|| < 1/2^{k-1},
\]

then each component of \( \{v^k_t, k \in \mathbb{N}\} \) is a Cauchy sequence. Call \( v_t \) the vector of the limits. To prove (a), we have to show that each row of \( \{G_n(L), n \in \mathbb{N}\} \) is a DAS. We have

\[
|G^*_n(\theta)|^2 = F_n(\theta)A^{-1}(\theta)F_n(\theta),
\]

whose diagonal entries cannot be larger than \( 1/\chi_{m,q}(\theta) \) by Fact M1, since \( F_n(\theta) \in K_M \). The latter ratio converges to zero a.e. in \([-\pi, \pi]\) and is less than 1 by Assumption 2, so that its integral on \([-\pi, \pi]\) converges to zero by Fact 1.

Finally, (b) follows from Lemma 7 and the fact that the spectral density matrix of \( v^k_t \) equals \( I_q \) for \( \theta \in M \), zero for \( \theta \notin M \).

QED

Lemma 9. Suppose that (I) and (II) hold. There exists a \( q \)-dimensional orthonormal white-noise vector process \( z \) such that \( z_{jt} \) is an aggregate for \( j = 1, \ldots, q \).

Proof. Define \( v_1 \) as the smallest among the integers \( m \) such that the measure of

\[
\mathcal{L} \{ \theta \in \Pi, \lambda_{m,q}(\theta) > 1 \} \geq \pi
\]

and

\[
M_1 = \{ \theta \in \Pi, \lambda_{m,q}(\theta) > 1 \}.
\]

By recursion define \( v_a, a \in \mathbb{N}, \) as the smallest among the integers \( m \) such that

\[
\mathcal{L} \{ \theta \in M_{a-1}, \lambda_{m,q}(\theta) > a \} \geq \pi
\]

and

\[
M_a = \{ \theta \in M_{a-1}, \lambda_{m,q}(\theta) > a \}.
\]

The measure of the set

\[
N_1 = M_1 \cap M_2 \cap \cdots \cap M_a \cap \cdots
\]

is not less than \( \pi \). Now define \( N_2 \) starting with \( \Pi - N_1 \) and using \( \mathcal{L}(\Pi - N_1)/2 \) instead of \( \pi \), \( N_a \) starting with \( \Pi - N_1 - N_2 - \cdots - N_{a-1} \) and using \( \mathcal{L}(\Pi - N_1 - N_2 - \cdots - N_{a-1})/2 \). We have

\[
\Pi = N_1 \cup N_2 \cup \cdots \cup N_a \cup \cdots
\]

and

\[
2\pi = \mathcal{L}(N_1) + \mathcal{L}(N_2) + \cdots + \mathcal{L}(N_a) + \cdots
\]

Lemma 8 can be applied to the subset \( N_a \), with the sequence \( \alpha_n \) suitably defined. We obtain a \( q \)-dimensional vector \( v^a_t = (v^a_{1t}, v^a_{2t}, \ldots, v^a_{qt}) \) such that (i) \( v^a_{jt} \) is an aggregate for \( j = 1, \ldots, q \); (ii) its spectral density matrix is \( I_q \) a.e. in \( N_a \), zero for \( \theta \notin N_a \). Now set
\[ z_t = \sum_{a=1}^{\infty} v_t^a \]. It is easily seen that the spectral density matrix of \( z_t \) is \( I_q \) a.e. in \([-\pi, \pi]\), so that \( z \) is a \( q \)-dimensional orthonormal white noise process.

**QED**

4.5 We now prove that the space spanned by \( z \) is \( G(x) \).

Let \( y_t \) be an aggregate. Consider the projection

\[ y_t = \text{proj}(y_t|\text{span}(z)) + r_t. \]

We want to show that \( r_t \) is necessarily zero. Consider the \((q+1)\)-dimensional vector process \((z_t, r_t)\). Its spectral density, call it \( W \), is diagonal with \( I_q \) in the \( q \times q \) upper-left submatrix, so that

\[ \det W(\theta) = \mathcal{S}(r_t, r_t; \theta). \]

Since \( z_jt \) and \( r_t \) belong to \( G(x) \), let \( \{a_{nj}(L), n \in \mathbb{N}\} \), for \( j = 1, \ldots, q+1 \) be DAS's such that

\[ \lim_n a_{nj}(L)x_{s_n,t} = z_jt, \text{ for } j = 1, \ldots, q, \]
\[ \lim_n a_{nq+1}(L)x_{s_n,t} = r_t. \]

Note that, possibly by augmenting the filters \( a_{nj}(L) \) with zeros, we can assume the same sequence \( \{s_n, n \in \mathbb{N}\} \) for all \( j = 1, \ldots, q+1 \). Moreover: (1) \( \int_{-\pi}^{\pi} |a_{nj}(\theta)|^2 d\theta \) converges to zero for \( j = 1, \ldots, q+1 \), so that a subsequence of \( a_{nj}^o \) converges to zero a.e. in \([-\pi, \pi]\) (Fact L2); (2) calling \( Z_n \) the spectral density matrix of the vector process

\[ (a_{n1}(L)x_{s_n,t} \ a_{n2}(L)x_{s_n,t} \ \cdots \ a_{nq+1}(L)x_{s_n,t}), \]

a subsequence of \( Z_n \) converges to \( W \) a.e. in \([-\pi, \pi]\) (Lemma 7). Thus, with no loss of generality we can assume that \( a_{nj}^o \) converges to zero and \( Z_n \) converges to \( W \) a.e. in \([-\pi, \pi]\).

Now, for \( j = 1, \ldots, q+1 \), set \( f_{nj} = a_{nj}^o \hat{P}_{s_n} \) and \( g_{nj} = a_{nj}^o - f_{nj} P_{s_n} \), so that

\[ a_{nj}^o = f_{nj} P_{s_n} + g_{nj} \]

and

\[ |a_{nj}^o(\theta)|^2 = |f_{nj}(\theta)|^2 + |g_{nj}(\theta)|^2. \]

Since \( a_{nj}^o \) converges to zero a.e. in \([-\pi, \pi]\), then \( g_{nj} \) converges to zero a.e. in \([-\pi, \pi]\). Moreover, the definition of \( g_{nj} \) and \( f_{nj} \) implies that

\[ a_{nj}(L)x_{s_n,t} = f_{nj}(L)\hat{P}_{s_n}(L)x_{s_n,t} + g_{nj}(L)x_{s_n,t} \]

is the orthogonal projection of the LHS on the space spanned by \( \hat{P}_{s_nk}(L)x_{s_n,t} \), for \( k = 1, \ldots, q \) and \( t \in \mathbb{Z} \). As a consequence, the spectral density matrix \( Z_n \) is equal to the spectral density matrix of

\[ (f_{n1}(L)\hat{P}_{s_n}(L)x_{s_n,t} \ f_{n2}(L)\hat{P}_{s_n}(L)x_{s_n,t} \ \cdots \ f_{nq+1}(L)\hat{P}_{s_n}(L)x_{s_n,t}), \]

call it \( Z_n^1 \), plus the spectral density matrix of

\[ (g_{n1}(L)x_{s_n,t} \ g_{n2}(L)x_{s_n,t} \ \cdots \ g_{nq+1}(L)x_{s_n,t}), \]

call it \( Z_n^2 \); \( Z_n = Z_n^1 + Z_n^2 \).
Now observe firstly that $Z_n^1$ is singular for any $\theta$. Secondly, since $g_{nj}(\theta)$ is orthogonal to $p_{\theta,k}^x$, for $k = 1, \ldots, q$, by Fact M$_1$ (Appendix),

$$g_{nj}(\theta)\Sigma_{sn}(\theta)g_{nj}(\theta) \leq \lambda_{s,q+1}^x |g_{nj}(\theta)|^2.$$ 

Essential boundedness of $\lambda_{q+1}^x$ along with convergence to zero a.e. of $g_{nj}$ imply that $Z_n^2$ converges to zero a.e. in $[-\pi, \pi]$. This implies that det $W(\theta) = S(r_t, r_\theta; \theta) = 0$ a.e. in $[-\pi, \pi]$, so that $r_t = 0$.

4.6 So far we have proved that if (I) and (II) hold then

$$x_{it} = \gamma_{it} + \delta_{it},$$

$$\gamma_{it} = \text{proj}(x_{it}; G(x)) = c_t(L)z_t,$$

where $z$ is a $q$-dimensional orthonormal white noise, and $c_t \in L^2([-\pi, \pi], C)$. Suppose that $\delta$ is idiosyncratic. By Fact M$_3$, $\lambda_{n,q}(\theta) \geq \lambda_{n,1}(\theta) - \lambda_{n,1}^\delta(\theta)$, so that $\lambda_{n,q}(\theta) = \infty$ a.e. in $[-\pi, \pi]$. Thus, to complete the proof of Theorem 2 we must only show that $\delta$ is idiosyncratic.

We need some additional preliminary results. Suppose that $v = \{v_t, t \in Z\}$ and $w = \{w_t, t \in Z\}$ are orthonormal $q$-dimensional white-noise vectors whose components belong to $L^2(P, C)$. Moreover, suppose that $v$ and $w$ are costationary, i.e. $E(v_t w_{t-k})$ does not depend on $t$, so that we can use the notation $v_t(w_{t-k})$. Setting $$(X) A(L) = \sum_{k=-\infty}^{\infty} \Gamma_{k}\ L^k,$$

it is easily seen that $A(L)w$ is the orthogonal projection of $v$ on the process $w$. Moreover, $A(F)v_t$, where $F = L^{-1}$, is the orthogonal projection of $w$ on the process $v$. Lastly, consider the matrix $A(e^{it}\theta)$. Its $(i, j)$ entry is the cross spectrum between $v_{it}$ and $w_{jt}$ and has therefore modulus bounded by $1/2\pi$ for any $\theta \in [-\pi, \pi]$.

**Definition 10.** For $n = 1, 2, \ldots, \infty$, let $v_n = \{v_{nt}, t \in Z\}$ be a sequence of $q$-dimensional orthonormal white-noise vectors. Assume that $v_n$ and $v_m$ are costationary for all $n$ and $m$. Consider the orthogonal projection

$$v_{nt} = A^{mn}(L)v_{nt} + \rho_{t}^{mn},$$

and let $D^{mn}$ be the spectral density of $\rho_{t}^{mn}$. The sequence $\{v_n, n \in N\}$ generates a **Cauchy sequence of spaces** if, given $\epsilon > 0$, for $\theta$ a.e. in $[-\pi, \pi]$ there exists an integer $m_\epsilon(\theta)$ such that for $m > m_\epsilon(\theta)$, $\text{trace}(D^{mn}(\theta)) < \epsilon$.

**Remark 10.** Note that, if $v_{nt}$ converges, it generates a Cauchy sequence of spaces, whereas the converse does not necessarily hold. As we shall show, the normalized principal components $\psi_{n}^\theta$ generate a Cauchy sequence of spaces. However, they do not converge in general: for example, take $q = 1$ and assume that $\psi_{n}^\theta$ is a normalized principal component converging to $\psi_t$; then $(-1)^n \psi_{t}^\theta$ is also a normalized principal component which does not converge.

**Lemma 10.** Assume that $\{v_n, n \in N\}$ generates a Cauchy sequence of spaces and let $y = \{y_t, t \in Z\}$, with $y_t \in L^2(P, C)$, be costationary with $v_n \in L^2(P, C)$ for any $n$. Consider the orthogonal projections of $y_t$ on the process $v_n$, i.e. $Y_{nt} = \text{proj}(y_t; \text{span}(v_n))$. Then $Y_{nt}$ converges in mean square to an element $Y_t$ in $L^2(P, C)$. 18
Proof. We have
\[ Y_t = Y_{nt} + r_{nt} = b_n(L)v_{nt} + r_{nt}, \]
where \( b_n(L) \) and \( b_m(L) \) are square summable \( q \)-dimensional filters, so that
\[ b_n(L)v_{nt} - b_m(L)v_{nt} = r_{mt} - r_{nt}. \]

The spectral density of the LHS is the cross spectrum between the LHS and the RHS. The latter, due to the definition of \( r_{nt} \) and \( r_{nt} \), is the sum of the cross spectrum between \( r_{nt} \) and \( b_n(L)v_{nt} \), call it \( S_1 \), and the cross spectrum between \( r_{mt} \) and \( b_n(L)v_{nt} \), call it \( S_2 \). Using (21), \( S_1 \) is the cross spectrum between \( r_{nt} \) and \( b_m(L)A^{mn}(L)v_{nt} + b_m(L)\rho_{mn}^{nm} \), which reduces to the cross spectrum between \( r_{nt} \) and \( b_m(L)\rho_{mn}^{nm} \), call it \( C_{mn} \). Now observe that both the spectral density of \( r_{nt} \) and the entries of \( b_m(e^{-i\theta}) \) are bounded in modulus by the spectral density of \( y_t \). Thus, since \( \{v_n, n \in \mathbb{N}\} \) generates a Cauchy sequence of spaces, \( C_{mn} \) converges to zero a.e. in \([-\pi, \pi]\) as \( n, m \to \infty \). The same argument holds for \( S_2 \), so that the spectral density of \( Y_{nt} - Y_{nt} \) converges to zero a.e. in \([-\pi, \pi]\) as \( m, n \to \infty \). Since both the spectral densities of \( Y_{nt} \) and of \( Y_{nt} \) are dominated by the spectral density of \( y_t \), by the Lebesgue dominated convergence theorem (Fact L1), the integral of the spectral density of \( Y_{nt} - Y_{nt} \) also converges to zero as \( m, n \to \infty \), so that \( Y_{nt} \) is a Cauchy sequence. QED

Now let us go back to equation (12) and concentrate on a single line, i.e. the orthogonal decomposition obtained by projecting \( x_{it} \) on the normalized principal components \( \pi_j(t) \), \( j = 1, \ldots, q \). Calling \( \pi_{ni}(L) \) the \( i \)-th (\( q \)-dimensional) row of \( \Pi_n(L) \) and \( q_{ni}(L) \) the \( i \)-th row of \( Q_n(L) \), we get
\[ x_{it} = \pi_{ni}(L)\Delta_n^{1/2}(L)\psi^i_t + q_{ni}(L)Q_n(L)x_{nt}. \]

Lemma 11. The sequence \( \{\psi^n, n \in \mathbb{N}\} \) generates a Cauchy sequence of spaces.
Proof. For \( n > m \) consider (15) for \( Q(L) = I_q \):
\[ \psi^n_t = D(L)\psi^n + \rho^n_t. \]

Calling \( D^{mn} \) the spectral density of \( \rho_t^{mn} \), convergence to zero of trace(\( D^{mn}(\theta) \)) for \( \theta \) a.e. in \([-\pi, \pi]\) and \( n > m \) is a consequence of Lemma 5. On the other hand,
\[ \psi^n_t = -\frac{1}{2}D(F)\psi^n + \rho^n_t. \]

From (22) and (23) we get
\[ I_q = D(e^{-i\theta})D(e^{i\theta}) + D^{mn}(\theta) = D(e^{i\theta})D(e^{-i\theta}) + D^{nm}(\theta). \]

By taking the trace on both sides and noting that the trace of \( D(e^{i\theta})D(e^{-i\theta}) \) is equal to the trace of \( D(e^{-i\theta})D(e^{i\theta}) \) we get
\[ \text{trace}(D^{mn}(\theta)) = \text{trace}(D^{nm}(\theta)). \]

Thus \( \text{trace}(D^{mn}(\theta)) \) converges to zero for any diverging \( n \) and \( m \). QED

19
The following theorem, besides being useful to show that $\delta$ is idiosyncratic, is important per se, because of its implications for the estimation of common and idiosyncratic components (see Forni, Hallin, Lippi and Reichlin, 1998).

**Theorem 5.** The sequence of projections $\gamma^n_{it} = \pi_{ni}(L)\Delta_{n}^{1/2}(L)\psi^n_t = \pi_{ni}(L)P_{n}(L)x_{nt}, n \in \mathbb{N}$ converges in mean square to $\gamma^*_{it} = \text{proj}(x_{it}|\mathcal{G}(x))$, for any $i$.

**Proof.** By Lemmas 10 and 11 $\gamma^n_{it}$ converges in mean square to an element $\gamma^*_{it}$ in $\mathbf{X}$. Therefore the sequence of the residuals $\delta^n_{it} = x_{it} - \gamma^n_{it}$ also converges to an element $\delta^*_{it}$ in $\mathbf{X}$. By Lemma 7, $\delta^*_{it}$ must be orthogonal to $\gamma^*_{it}$ at all leads and lags. Moreover, $\gamma^*_{it}$ is an aggregate, since $\pi_{ni}(L)P_{n}(L)$ is a DAS. To see this, consider that the spectral density of $\gamma^n_{it}$, i.e. $\pi_{ni}\Delta_{ni}\pi_{ni}$, is not larger than $\pi_{ni}\pi_{ni}\lambda_{ni}^{x}$, and is bounded above by the spectral density of $x_{it}$, call it $\sigma_{i}$, implying $\pi_{ni}(\theta)\pi_{ni}(\theta) \leq \sigma_{i}(\theta)/\lambda_{ni}^{x}(\theta)$. The latter ratio converges to zero a.e. in $[-\pi, \pi]$ and is bounded above by $\sigma_{i}(\theta)$ by Assumption 2, so that Fact L1 applies. Summing up, $\gamma^*_{it}$ belongs to $\mathcal{G}(x)$ and is orthogonal to $\delta^*_{it}$, so that $\gamma^*_{it} = \gamma^*_{it}$. QED

The following Lemma concludes the proof of Theorem 2.

**Lemma 12.** $\delta$ is idiosyncratic.

**Proof.** Let us fix $m$ and denote by $\Sigma^x_{m}$ the spectral density matrix of the vector process $\delta_{m}(\theta) = (\delta_{1t} \delta_{2t} \ldots \delta_{mt})'$. We want to show that the first eigenvalue of such matrix, i.e. $\lambda_{m1}(\theta)$, cannot be larger than $\Sigma^x_{m}$ for any $\theta \in [-\pi, \pi]$. Let $\Sigma^x_{m}, n > m$, be the spectral density matrix of $\delta_{m}(\theta) = (\delta_{1t} \delta_{2t} \ldots \delta_{mt})'$ and $\lambda_{m1}^{x}$ be its first eigenvalue. By Theorem 5 $\delta_{m}^{x}$ converges to $\delta_{m}^{\ast}$ in mean square for $i = 1, \ldots, m$, so that, by Lemma 7, a subsequence of $\Sigma^x_{m}$ converges to $\Sigma^x_{m}$ a.e. in $[-\pi, \pi]$. Assuming that $\lim_n \Sigma^x_{m} = \Sigma^x_{m}$ a.e. in $[-\pi, \pi]$ avoids further complication in notation and does not imply any loss of generality. Continuity of the eigenvalues as functions of the matrix entries (Ahlfors, pp. 300-6; see also the proof of Lemma 1) implies that

$$\lim_n \lambda_{m1}^{x}(\theta) = \lambda_{m1}^{\ast}(\theta),$$

(24) a.e. in $[-\pi, \pi]$. Moreover, note that $\Sigma^x_{m}$ is the $m \times m$ upper-left submatrix of $\Sigma^x_{n}$, so that, by Fact M1,

$$\lambda_{m1}^{x}(\theta) \leq \lambda_{m1}^{n}(\theta) = \lambda_{ni}^{x}(\theta)$$

for any $n \geq m$ and any $\theta$ in $[-\pi, \pi]$. Hence by (24) $\lambda_{m1}^{\ast}(\theta) \leq \lambda_{ni}^{x}(\theta)$. Since this is true for any $m$,

$$\lambda_{i1}^{\ast}(\theta) \leq \lambda_{n1}^{x}(\theta),$$

(25)

so that $\lambda_{i1}^{\ast}$ is essentially bounded. The statement follows from Theorem 1. QED

**4.7** Now we prove Theorem 3. Assume that $x$ fulfills Definition 9, so that

$$x_{it} = \chi_{it} + \xi_{it}$$

$$\chi_{it} = b_{i}(L)u_{t},$$

where $u$ is $q$-dimensional. By Theorem 2, $x$ has also the representation

$$x_{it} = \gamma_{it} + \delta_{it}$$

$$\gamma_{it} = \text{proj}(x_{it}|\mathcal{G}(x)) = c_{i}(L)\xi_{t},$$

20
where $z$ is $q$-dimensional and 
\[ \text{span}(z) = \mathcal{G}(x). \]

Since $\xi$ is idiosyncratic
\[ \mathcal{G}(x) \subset \text{span}(\chi), \]
and obviously
\[ \text{span}(\chi) \subset \text{span}(u). \]

On the other hand, since both $u$ and $z$ are $q$-dimensional, then
\[ \text{span}(z) = \mathcal{G}(x) = \text{span}(\chi) = \text{span}(u). \quad (26) \]

Now, (26) implies that $x_{it} \in \mathcal{G}(x)$ and $\xi_{it} \perp \mathcal{G}(x)$, so that $x_{it} = \text{proj}(x_{it}|\mathcal{G}(x))$ and $\xi_{it} = \delta_{it}$.

Remark 11. Since we have proved that $\delta_{it} = \xi_{it}$, (9) and (25) imply that
\[ \lambda_{q+1}^x(\theta) = \lambda_{q}^{\xi}(\theta) \]
a.e. in $[-\pi, \pi]$.

5. Non-stationary variables

The case of trend stationary or difference stationary variables can be easily accommodated in our model. Assuming that the nature of non-stationarity is correctly detected, then, in the first case, i.e. $x_{it} = T_t + z_{it}$, where $T_t$ is a deterministic trend, our results should be applied to the stationary components $z_{it}$. In the second case, assume, for the sake of simplicity, that the variables $x_{it}$ are I(1). Consider the differences $y_{it} = (1-L)x_{it}$ and suppose that (I) and (II) hold for $\lambda_{q+1}^{y}$ and $\lambda_{q}^{\xi}$ respectively. Then we have the representation
\[
(1-L)x_{it} = \chi_{it} + \xi_{it}
\chi_{it} = b_i(L)u_t,
\]
where $u_t$ is $q$-dimensional and $\xi$ is idiosyncratic. Now observe that the vectors $x_{nt}$ and $\xi_{nt}$ are identified, and so are the spectral density matrices $\Sigma_{x}^n$ and $\Sigma_{\xi}^n$. Therefore all the information necessary to determine whether the $\chi$'s, or the $\xi$'s, are I(1) or I(0), and whether cointegration relationships hold among the $\chi$'s or the $\xi$'s, can be recovered starting with the spectral density matrices of the $x$'s.
REFERENCES

Appendix

Proof of Lemma 1. Since the leading coefficient of the characteristic polynomial of $\Sigma_n^x$ never vanishes, the eigenvalues, as functions of the entries of $\Sigma_n^x$, are continuous. Precisely, let $\mathcal{M}$ be the range of the function $\phi : [-\pi, \pi] \mapsto \mathbb{C}^{n \times n}$ that associates $\Sigma_n^x(\theta)$ with $\theta$. There exist $n$ functions $\delta_j : \mathcal{M} \mapsto \mathbb{R}$, that are continuous in $\mathcal{M}$ and such that $\delta_j(\phi(\theta))$ is a root of the characteristic equation of $\Sigma_n^x(\theta)$ (see Ahlfors, 1987, pp. 300-6). Now, consider the function $\nu_1 : \mathcal{M} \mapsto \mathbb{R}$ defined as

$$\nu_1(\mu) = \max_{j=1,n} \delta_j(\mu).$$

Obviously $\lambda_{n1}^x(\theta) = \nu_1(\phi(\theta))$. By a standard argument $\nu_1$ is continuous on $\mathcal{M}$. Measurability of $\phi$ (see Assumption 1) ensures measurability of $\lambda_{n1}^x$. Now let $k(1, \mu)$ be the integer such that $\delta_{k(1, \mu)}(\mu) = \nu_1(\mu)$. Define

$$\nu_2(\mu) = \max_{j \neq 1,n \mu} \delta_j(\mu).$$

Obviously $\lambda_{n2}^x(\theta) = \nu_2(\phi(\theta))$ and the argument used for the first eigenvalue applies. Iteratively we obtain that all the eigenvalues $\lambda_{nj}^x$ are measurable. QED

Proof of Lemma 2. Let $y = (y_1, y_2, \ldots, y_n)'$ be an $n$-dimensional stochastic vector with variance-covariance matrix $\Sigma$. Let $\lambda_k$ be the $k$-th eigenvalue of $\Sigma$, in descending order, and $p_k$ an eigenvector of $\Sigma$ associated with $\lambda_k$. We recall that:

Fact M1. For $k = 1, \ldots, n$, $\lambda_k$ is

$$\max_b \|by\|^2 = \max_b b\Sigma b$$

s.t. $b \in \mathbb{C}^n$, $|b| = 1$, $b \perp p_j$ for $j < k$.

Moreover, if $b = p_k$, then the orthogonality condition is fulfilled and $\|p_k y\|^2 = \lambda_k$ (see ???).

Fact M2. For $k = 1, \ldots, n$, let $c$ be any $(k-1)$-tuple $\{c_j, j = 1, \ldots, k-1\}$, where $c_j \in \mathbb{C}^n$. The eigenvalue $\lambda_k$ is

$$\min_c \max_b \|b\Sigma c\|$$

s.t. $|b| = 1$, $b \perp c_j$, $j = 1, \ldots, k-1$

(see Brillinger, 1981, p. 84, Exercise 3.10.16).

Going back to the proof, using Fact M1 the statement of Lemma 2 is trivial for $k = 1$. For $k > 1$ consider:

$$\min_c \max_b b\Sigma_{n+1}^x(\theta) b$$

s.t. (I) $|b| = 1$, $b \perp c_j$, $j = 1, \ldots, k-1$; (II) the last component of $b$ is zero. (27)

Trivially, the value of (27) cannot exceed $\lambda_{n+1,k}^x(\theta)$, as obtained using Fact M2. On the other hand, the constraint on the last component of $b$ implies that if $b \perp c_j$ then $b$ is also orthogonal to the $k-1$ vectors of $\mathbb{C}^n$ whose components are the first $n$ components of $c_j$, for $j = 1, \ldots, k-1$. Thus the value of (27) is $\lambda_{nk}^x(\theta)$.

QED

The following is a useful consequence of Fact M2.
Fact M3. Let $D$ and $E$ be $m \times m$ Hermitian non-negative definite, and $F = D + E$. Then

$$\lambda_s^f \leq \lambda_s^d + \lambda_s^e$$
$$\lambda_s^f \leq \lambda_s^d + \lambda_s^e$$
$$\lambda_s^f \geq \lambda_s^d$$
$$\lambda_s^f \geq \lambda_s^e$$

for any $s = 1, \ldots, m$.

Proof. For the first inequality observe that, calling $V_s$ the set of all vectors $b \in \mathbb{C}^m$ such that $|b| = 1$ and $b \perp c_j$, where $c_j, j = 1, \ldots, s - 1$, is a given $(s - 1)$-tuple of vectors in $\mathbb{C}^m$,

$$\max_{b \in V_s} b^* \Sigma b \leq \max_{b \in V_s} b^* \Sigma^d b + \max_{b \in V_s} b^* \Sigma^e b \leq \max_{b \in V_1} b^* \Sigma b + \lambda_1^e.$$

Then take the minimum over all $(s - 1)$-tuples. The second inequality follows in the same way. The second and third are trivial. QED

Proof of Lemma 3. Let $S^1(\theta) = I_n \lambda_{n1}^x(\theta) - \Sigma_n^x(\theta)$ and consider the system of equations

$$b(\theta)S^1(\theta) = 0,$$  \hspace{1cm} (28)

with $b(\theta) = (b_1(\theta) \ b_2(\theta) \ \cdots \ b_n(\theta))$. The subset of $[-\pi, \pi]$ where $\text{rank}(S^1(\theta)) = 0$, call it $M_0$, is measurable, possibly empty. In $M_0$ put $p_{n1}^x(\theta) = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} / \sqrt{n}$. Let $M_1$ be the measurable subset of $[-\pi, \pi]$ where $\text{rank}(S^1(\theta)) = 1$. Then let $M_1^1$ be the measurable subset of $M_1$ where $S^1_{11}(\theta) \neq 0$. Put $b_2(\theta) = 1, b_j(\theta) = 0$ for $j > 2$, then obtain the unique solution for $b_1(\theta)$, i.e. $b_1(\theta) = S^1_{11}(\theta) / S^1_{11}(\theta)$. Putting $p_{n1}^x(\theta) = b(\theta) / |b(\theta)|$ in $M_1^1$ we have a measurable function in $M_1^1$ (notice that the choice $-b(\theta) / |b(\theta)|$ would be also valid). Now consider the subset of $M_1 - M_1^1$ where $S^1_{22}(\theta) = 0$, and repeat the construction by taking $b_1(\theta) = 1, b_j(\theta) = 0$ for $j > 2$. It is clear how to proceed to cover $M_1$. Then consider $M_2$, the set where $\text{rank}(S^1(\theta)) = 2$, and so on until $M_{n-1}$, the subset of $[-\pi, \pi]$ where $\text{rank}(S^1(\theta)) = n - 1$. Define the subset $M_1^{n-1}$ as that in which the top-left submatrix of order $n - 1$ of $S^1(\theta)$ is non singular. Put $b_n(\theta) = 1$ in $M_1^{n-1}$ and find the unique solution to (28). Then again put $p_{n1}^x(\theta) = b(\theta) / |b(\theta)|$. Continuing in this way until $M_{n-1}^{n-1}$ has been covered we obtain a measurable function for $p_{n1}^x$ defined on $[-\pi, \pi]$. Now consider $S^2(\theta) = I_n \lambda_{n2}^x(\theta) - \Sigma_n^x$ and the system

$$b(\theta)S^2(\theta) = 0$$
$$b(\theta)p_{n1}^x(\theta) = 0.$$  \hspace{1cm} (29)

The procedure above can be applied to (29) so that we obtain a measurable $p_{n2}^x$ fulfilling (1) and (2). Iterating we reach the last step, when the system is

$$b(\theta)S^{n-1}(\theta) = 0$$
$$b(\theta)p_{nj}^x(\theta) = 0 \text{ for } j \leq n - 1.$$

QED
Proof of Lemma 7. We have

\[
|S(A_n, B_n; \theta) - S(A, B; \theta)| \\
\leq |S(A_n, B_n; \theta) - S(A_n, B; \theta)| + |S(A_n, B; \theta) - S(A, B; \theta)| \\
= |S(A_n, B_n - B; \theta)| + |S(A_n - A, B; \theta)| \\
\leq \sqrt{S(A_n, A_n; \theta)} \sqrt{S(B_n - B, B_n - B; \theta)} + \sqrt{S(B, B; \theta)}\sqrt{(A_n - A, A_n - A; \theta)} \\
\leq \left[ \sqrt{S(A, A; \theta)} + \sqrt{S(A_n - A, A_n - A; \theta)} \right] \sqrt{S(B_n - B, B_n - B; \theta)} \\
+ \sqrt{S(B, B; \theta)}\sqrt{S(A_n - A, A_n - A; \theta)}.
\]

Since \(S(A_n - A, A_n - A; \theta)\) and \(S(B_n - B, B_n - B; \theta)\) converge to zero in the mean, by Fact L2 there exists a sequence \(s_i\) such that \(S(A_{s_i} - A, A_{s_i} - A; \theta)\) and \(S(B_{s_i} - B, B_{s_i} - B; \theta)\) converge to zero a.e. in \([-\pi, \pi]\).

QED

Fact S. If \(y_t \in X\), then \(y_t\) has a spectral density.

Sketch of the proof. Let

\[
y_t = \lim_{n} \sum_{j=1}^{n} a_{nj}(L)x_{jt},
\]

let \(f_n\) be the spectral density of \(\sum_{j=1}^{n} a_{nj}(L)x_{jt}\) and \(g_{nm}\) the spectral density of

\[
\sum_{j=1}^{n} a_{nj}(L)x_{jt} - \sum_{j=1}^{m} a_{mj}(L)x_{jt}.
\]

Equation (30) implies that \(g_{nm}\) converges to zero in mean as \(n\) and \(m\) tend to infinity. The same argument employed in Lemma 7 leads to the conclusion that \(f_n\) converges in the mean, call \(f\) the limit. Proving that \(f\) actually is the spectral density of \(y_t\) is not difficult but rather tiresome. The interested reader may request the proof from the authors.

Maria Cristina Marcuzzo (a cura di) [1988] "Richard F. Kahn. A discipline of Keynes", pp. 118

Paolo Bosi [1988] "MICROMOD. Un modello dell'economia italiana per la didattica della politica fiscale", pp. 34


Carlo Grillenzoni [1988] "Identification, estimations of multivariate transfer functions", pp. 33


Andrea Ginzburg [1988] "Il comitato italiano", pp. 30


Leonardo Paggi [1988] "Rappresentismo e riformismo. La socialdemocrazia europea nell'economia mondiale aperta", pp. 120


Giovanna Procacci [1989] "State coercion and worker solidarity in Italy (1915-1918): the moral and political content of social unrest", pp. 41

Carlo Alberto Magni [1989] "Reputazione e credibilità di una minaccia in un gioco bargaining", pp. 56

Giovanni Mottra [1989] "Agricultura familiare e sistema agroalimentare in Italia", pp. 84


Paolo Bosi, Roberto Golinelli, Anna Stagni [1989] "Le origini del debito pubblico e il costo della stabilizzazione", pp. 26

Roberto Golinelli [1989] "Nota sulla struttura e sull'impiego dei modelli macroeconometrici", pp. 21

Marco Lippi [1989] "A Short Note on Contegretation and Aggregation", pp. 11


Gabriele Pastrello [1989] "Francois quesnay dal Tableau Zig-zag al Tableau Formule: una ricostruzione", pp. 48

Paolo Silvestri [1989] "Il bilancio dello stato", pp. 54


Michele Lalla [1990] "La spesa pubblica e l'efficienza delle risorse: un esempio fra riguardo e output", pp. 25

Paolo Silvestri [1990] "Sull'autonomia finanziaria dell'università", pp. 11

Tindara Addabbo [1998] "Probability di occupazione e aspettative individuali" pp 36
222. Luigi Brighi e Marcello D’Amato [1998] "Optimal Procurement in Multiproduct Monopoly" pp 23
223. Paolo Bortolussi, Mauro Cecilia Guerra e Paolo Silvestri [1998] "La spesa sociale nel comune Modena" Rapporto intermedio pp 37
226. Tommaso Minerva [1998] "Internet Domande e Risposte" pp 183
229. Tommaso Minerva e Irene Poli [1998] "Building an ARMA Model by using a Genetic Algorithm" pp. 60
231. Enrico Cavazzini e Nicoletta Pacchiarotti [1998] "How to play an hotelling game in a square town" pp 12
232. Alberto Roverato e Irene Poli [1998] "Un algoritmo genetico per la selezione di modelli grafici" pp 11