The Generalized Dynamic Factor Model: Representation Theory

Author(s): Mario Forni and Marco Lippi

Source: *Econometric Theory*, Dec., 2001, Vol. 17, No. 6 (Dec., 2001), pp. 1113-1141

Published by: Cambridge University Press

Stable URL: https://www.jstor.org/stable/3533246

Econometric Theory, **17**, 2001, 1113–1141. Printed in the United States of America.

THE GENERALIZED DYNAMIC FACTOR MODEL: REPRESENTATION THEORY

MARIO FORNI
Università di Modena
and
CEPR

MARCO LIPPI
Università di Roma
and
CEPR

This paper, along with the companion paper Forni, Hallin, Lippi, and Reichlin (2000, *Review of Economics and Statistics* 82, 540–554), 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, *Econometrica* 51, 1181–1304) and Chamberlain and Rothschild (1983, *Econometrica* 51, 1305–1324) by allowing serial correlation within and across individual processes and of the dynamic factor model of Sargent and Sims (1977, in C.A. Sims (ed.), *New Methods in Business Cycle Research*, pp. 45–109) and Geweke (1977, in D.J. Aigner & A.S. Goldberger (eds.), *Latent Variables in Socio-Economic Models*, pp. 365–383) by allowing for non-orthogonal idiosyncratic terms. Whereas 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 empirical implementation of the model. Moreover, the common factors are obtained as limits of linear combinations of dynamic principal components. Thus the paper reconciles two seemingly unrelated statistical constructions.

1. INTRODUCTION

1.1. Outline of the Problem

Data sets with many data points both over time and across sections are becoming increasingly available. Think for instance of macroeconomic series on out-

This paper is part of research supported by an A.R.C. contract of the Communauté française de Belgique. The first author also acknowledges financial aid by MURST (Ministero per l'Università e la Ricerca Scientifica e Tecnologica). We thank Christine De Mol, Marc Hallin, and Lucrezia Reichlin for constant help and support. The paper, in its present form, owes much to the suggestions of an anonymous referee. Address correspondence to: Marco Lippi, Dipartimento di Scienze Economiche, Via Cesalpino 12, I-00161, Roma, Italy; e-mail: marco.lippi@uniroma1.it.

 0266-4666/01 $9.50

put or employment that 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, because 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 that 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 that 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 that 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, although 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 (2000), a new model, which 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.

1.2. Brief Summary of Results

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 $\{x_{jt},\, t \in \mathbb{Z}\}$ for any j. We do not assume an autoregressive moving average (ARMA) structure for the x's. We only require the existence of a spectral density matrix $\boldsymbol{\Sigma}_n^x$ for the vector $(x_{1t} \quad x_{2t} \quad \cdots \quad x_{nt})'$.

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 $\operatorname{var}(y_i) = \sigma^2$. Taking a sequence of averages $Y_n = \sum_{i=1}^n a_{ni} y_i$, the variance $\operatorname{var}(Y_n) = \sigma^2 \sum_{i=1}^n a_{ni}^2$ 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 $e^{-|i-j|}$, then $\operatorname{var}(Y_n)$ 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,

$$\sum_{i=1}^{n} \sum_{h=-k}^{k} a_{nih} x_{it-h},$$

provided that the sum of the squared weights tends to zero. We prove (in Theorem 1, which follows) that x_{it} is idiosyncratic if and only if the maximum eigenvalue of $\boldsymbol{\Sigma}_n^x$ is dominated by an essentially bounded function defined on $[-\pi, \pi]$ and independent of n.

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, $(u_{1t} \quad u_{2t} \quad \cdots \quad u_{qt})'$ is an orthonormal vector white noise, and ξ_{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, which follows, that a sequence has a generalized dynamic factor structure with q factors if and only if (i) the $(q + 1)$st eigenvalue of $\boldsymbol{\Sigma}_n^x$, in decreasing order, is dominated for any n by an essentially bounded function of the frequency θ; (ii) as n tends to infinity, the qth eigenvalue diverges for θ almost everywhere in $[-\pi, \pi]$.

Thus the unobservable factor structure is completely characterized in terms of properties of the observable matrices $\mathbf{\Sigma}_n^x$. This result, besides its theoretical interest, 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 $\mathbf{\Sigma}_n^x$ 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, first, the variables follow a generalized dynamic factor model and, second, that the number of factors is q. This is the main contribution of the present paper with respect to the companion paper, mentioned previously, in which a generalized dynamic factor model for the x's is assumed to concentrate on identification and estimation of common and idiosyncratic components and on criteria to detect the number of common factors.

Theorems 3 and 4, which follow, establish uniqueness of the idiosyncratic component ξ_{it} and of the common component $\chi_{it} = x_{it} - \xi_{it}$. It must be pointed out that this identifiability result holds for the whole infinite sequence of the variables x_{it}, not for its finite subsets: otherwise stated, identifiability occurs in the limit, when the size of the cross section tends to infinity. Moreover, note that identification of χ_{it} does not imply identification of the u's or of the filters $b_{ij}(L)$, which might be achieved only by imposing further, economically motivated, restrictions. Such an issue will not be discussed in this paper. Finally, in Theorem 5, which follows, 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. Moreover, it is interesting from a theoretical point of view in that, by unveiling the intimate relationship linking common factors to principal components, it provides a reconciliation between two important chapters of statistical analysis.

The case in which the x's are either difference or trend stationary is briefly discussed in Section 5.

1.3. Related Research

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. Also, the link between principal component and factor analysis has been observed by Chamberlain and Rothschild in the static case. Related models can also be found in Quah and Sargent (1993), Forni and Reichlin (1996, 1998), Forni and Lippi (1997), and Stock and Watson (1999).

2. NOTATION, BASIC DEFINITIONS, AND LEMMAS

2.1. Notation and Some Results

Given a complex matrix $\mathbf{D}$, finite or infinite, we denote by $\widetilde{\mathbf{D}}$ the complex conjugate of the transpose of $\mathbf{D}$. Inner product and norm in $\mathbb{C}^s$ are the usual Euclidean entities $(\mathbf{v},\mathbf{w}) = \sum_{i=1}^{s} v_i \bar{w}_i$ and $|\mathbf{v}| = \sqrt{\sum_{i=1}^{s} |v_i|^2}$, respectively. Moreover, Θ denotes the real interval $[-\pi,\pi]$.

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 Ω. We recall that $L_2(\mathcal{P},\mathbb{C})$, with the inner product defined as $\langle x, y\rangle = \mathrm{E}(x\bar{y}) = \mathrm{cov}(x,y)$ and the norm as $\|x\| = \sqrt{\mathrm{E}(|x|^2)} = \sqrt{\mathrm{var}(x)}$, is a complex Hilbert space. If Q is a subset of $L_2(\mathcal{P},\mathbb{C})$ we denote by $\overline{\mathrm{span}}(Q)$ 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 $\mathrm{proj}(x|V)$ the orthogonal projection of x on V.

The paper will deal with a double sequence

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

where $x_{it} \in L_2(\mathcal{P},\mathbb{C})$. We adopt the following notation:

(a) $\mathbf{X} = \overline{\mathrm{span}}(\mathbf{x})$.
(b) $\mathbf{x}_t$ is the infinite column vector $(x_{1t} \quad x_{2t} \quad \cdots \quad x_{it} \quad \cdots \quad)'$.
(c) $\mathbf{x}_{nt}$ is the n-dimensional column vector $(x_{1t} \quad x_{2t} \quad \cdots \quad x_{nt})'$.
(d) $\mathbf{X}_n = \overline{\mathrm{span}}(\{x_{st},\, s = 1,2,\ldots,n,\, t \in \mathbb{Z}\})$. Obviously $\mathbf{X}_n \subseteq \mathbf{X} = \overline{\bigcup_j \mathbf{X}_j}$.

Often, when no confusion can arise, we speak of the process z_t, meaning the process $\{z_t,\, t \in \mathbb{Z}\}$. Moreover, considering an m-dimensional vector process

$$\mathbf{y} = \{(y_{1t} \quad y_{2t} \quad \cdots \quad y_{mt})', \quad t \in \mathbb{Z}\},$$

we say that $\mathbf{y}$ belongs to $\mathbf{W} \subseteq L_2(\mathcal{P},\mathbb{C})$ if y_{jt} belongs to $\mathbf{W}$ for any j and t. In the same way, we use $\overline{\mathrm{span}}(\mathbf{y})$ to indicate $\overline{\mathrm{span}}(\{y_{jt},\, j = 1,2,\ldots,m,\, t \in \mathbb{Z}\})$.

Assumption 1. For any $n \in \mathbb{N}$:

(1) the process $\mathbf{x}_{nt}$ is covariance stationary;
(2) the spectral measure of $\mathbf{x}_{nt}$ is absolutely continuous (with respect to the Lebesgue measure on Θ), i.e., $\mathbf{x}_{nt}$ has a spectral density (see Rozanov, 1967, pp. 19–20).

Assumption 1 will be the basis for all definitions and results that follow and will be tacitly supposed to hold throughout the paper. We denote by $\boldsymbol{\Sigma}_n^x$ the spectral density matrix of $\mathbf{x}_{nt}$ and recall that $\boldsymbol{\Sigma}_n^x$ is Hermitian, nonnegative definite for any $\theta \in \Theta$, integrable, and that $\mathrm{E}(\mathbf{x}_{nt}\tilde{\mathbf{x}}_{nt-k}) = (1/2\pi)\int_{-\pi}^{\pi} e^{ik\theta}\boldsymbol{\Sigma}_n^x(\theta)\,d\theta$. Last, $\boldsymbol{\Sigma}^x$ denotes the infinite matrix whose $n \times n$ top-left submatrix is $\boldsymbol{\Sigma}_n^x$.

Remark 1. Note that our definition of the spectral density is equal to the usual definition (see, e.g., Brockwell and Davis, 1991, p. 120; Rozanov, 1967, pp. 19–20) times the factor 2π. This is a convenience, having the effect that all the orthonormal s-dimensional white-noise vectors appearing in Section 4 will have spectral density $\mathbf{I}_s$, instead of $\mathbf{I}_s/2\pi$.

If $\mathbf{a}$ denotes the infinite row vector $(a_1 \quad a_2 \quad \cdots \quad a_n \quad a_{n+1} \quad \cdots \quad)$, we denote by $\mathbf{a}^{[n]}$ the infinite row vector $(a_1 \quad a_2 \quad \cdots \quad a_n \quad 0 \quad 0 \quad \cdots \quad)$ and by $a^{\{n\}}$ the n-dimensional row vector $(a_1 \quad a_2 \quad \cdots \quad a_n)$.

We denote by $L_2^\infty(\Theta,\mathbb{C},\mathbf{\Sigma}^x)$ the complex linear space of all infinite row vectors $\mathbf{f} = (f_1 \quad f_2 \quad \cdots \quad f_n \quad \cdots \quad)$, such that (i) f_i is a measurable complex function defined on Θ, (ii) $\int_{-\pi}^{\pi} \mathbf{f}(\theta)\mathbf{\Sigma}^x(\theta)\tilde{\mathbf{f}}(\theta)d\theta = \lim_n \int_{-\pi}^{\pi} \mathbf{f}^{\{n\}}(\theta) \times \mathbf{\Sigma}_n^x(\theta)\tilde{\mathbf{f}}^{\{n\}}(\theta)d\theta < \infty$. The terms $\mathbf{f}$ and $\mathbf{g}$ are to be considered equivalent if $\int_{-\pi}^{\pi}(\mathbf{f}(\theta) - \mathbf{g}(\theta))\mathbf{\Sigma}^x(\theta)(\tilde{\mathbf{f}}(\theta) - \tilde{\mathbf{g}}(\theta))d\theta = 0$. Defining the inner product as $\langle \mathbf{f},\mathbf{g}\rangle_{\Sigma^x} = (1/2\pi)\int_{-\pi}^{\pi}\mathbf{f}(\theta)\mathbf{\Sigma}^x(\theta)\tilde{\mathbf{g}}(\theta)d\theta$, and the norm as $\|\mathbf{f}\|_{\Sigma^x} = \sqrt{\langle \mathbf{f},\mathbf{f}\rangle_{\Sigma^x}}$, the space $L_2^\infty(\Theta,\mathbb{C},\mathbf{\Sigma}^x)$ is a Hilbert space.

The space $L_2^\infty(\Theta,\mathbb{C})$ is defined as previously with $\mathbf{\Sigma}^x$ replaced by the infinite identity matrix (i.e., the matrix having $\mathbf{I}_n$ as the $n \times n$ top-left submatrix). Inner product and norm in $L_2^\infty(\Theta,\mathbb{C})$ are indicated by $\langle \mathbf{f},\mathbf{g}\rangle$ and $\|\mathbf{f}\|$, respectively. We will also refer to the Banach space $L_1(\Theta,\mathbb{C})$, whose elements are functions f such that $\int_{-\pi}^{\pi}|f(\theta)|d\theta < \infty$, with norm $\|f\|_1 = (1/2\pi)\int_{-\pi}^{\pi}|f(\theta)|d\theta$. The definition of the spaces $L_2^n(\Theta,\mathbb{C},\mathbf{\Sigma}_n^x)$ and $L_2^n(\Theta,\mathbb{C})$ is obvious, with n-dimensional in place of infinite-dimensional vector functions.

We denote by $\mathcal{L}$ the Lebesgue measure on $\mathbb{R}$. Let us recall that an extended real function $f: \Theta \to \mathbb{R}$ is *essentially bounded* if there exist a real c and a subset D of Θ such that $\mathcal{L}(D) = 0$ and $|f(\theta)| \le c$ for $\theta \in \Theta - D$. Moreover, for any real function f, $\operatorname{ess\,sup}(f) = \inf\{M : \mathcal{L}(\{y : f(y) > M\}) = 0\}$ (Royden, 1988, p. 119). The function f is essentially bounded if and only if $\operatorname{ess\,sup}(f) < \infty$. We denote by $L_\infty^n(\Theta,\mathbb{C})$ the complex linear space of all n-dimensional row vectors $\mathbf{f} = (f_1 \quad f_2 \cdots \quad f_n)$, with f_i measurable, such that $|\mathbf{f}|$ is essentially bounded.

Last, the space $L_2^{m\times s}(\Theta,\mathbb{C},\mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is an $s \times s$ spectral density matrix, is the set of all $m \times s$ matrices $\mathbf{A}$ such that $\mathbf{A}\mathbf{\Sigma}\tilde{\mathbf{A}}$ is integrable. If $\mathbf{A} \in L_2^{m\times s}(\Theta,\mathbb{C},\mathbf{\Sigma})$, then each row of $\mathbf{A}$ belongs to $L_2^s(\Theta,\mathbb{C},\mathbf{\Sigma})$. Analogously for $L_2^{m\times s}(\Theta,\mathbb{C})$. By $L_\infty^{m\times s}(\Theta,\mathbb{C})$ we denote the set of the matrix functions whose entries are essentially bounded.

Strictly speaking, the elements of $L_2^\infty(\Theta,\mathbb{C},\mathbf{\Sigma}^x)$, $L_2^\infty(\Theta,\mathbb{C})$, and so forth, are equivalence classes of functions, not functions (see Royden, 1988, footnote, p. 119). However, because no confusion can arise, we speak of functions, with the understanding that different functions are equivalent if their distance is zero. Thus statements such as $L_\infty^n(\Theta,\mathbb{C}) \subseteq L_2^n(\Theta,\mathbb{C})$, or $L_\infty^{m\times s}(\Theta,\mathbb{C}) \subseteq L_2^{m\times s}(\Theta,\mathbb{C}) \cap L_2^{m\times s}(\Theta,\mathbb{C},\mathbf{\Sigma})$, for any $\mathbf{\Sigma}$, make sense and are obviously true.

The following lemma shows that $L_2^\infty(\Theta,\mathbb{C},\mathbf{\Sigma}^x)$ is the straightforward generalization of the vector-function space occurring in the spectral representation of finite-dimensional vector stochastic processes.

LEMMA 1. *Let* $\hat{\mathbf{X}} = \cup_{s=1}^{\infty} \mathbf{X}_s$ *and* $\hat{L}_2^{\infty} = \cup_{s=1}^{\infty} \hat{L}_2^s$, *where* $\hat{L}_2^s = \{\mathbf{f}^{[s]} : \mathbf{f} \in L_2^{\infty}(\Theta, \mathbb{C}, \boldsymbol{\Sigma}^x)\}$. *Define* $\hat{\boldsymbol{\Omega}} : \hat{\mathbf{X}} \to \hat{L}_2^{\infty}$ *as the linear extension of*

$$\hat{\boldsymbol{\Omega}}(x_{ht}) = e^{it\cdot}(\delta_{h1} \quad \delta_{h2} \quad \cdots \quad \delta_{hk} \quad \cdots \quad), \tag{1}$$

where $\delta_{hk} = 1$ *if* $h = k$, $\delta_{hk} = 0$ *if* $h \neq k$. *The map* $\hat{\boldsymbol{\Omega}}$ *can be extended in a unique way into a map* $\boldsymbol{\Omega} : \mathbf{X} \to L_2^{\infty}(\Theta, \mathbb{C}, \boldsymbol{\Sigma}^x)$. *Moreover,* $\boldsymbol{\Omega}$ *is an isomorphism, i.e., one-to-one, onto and norm-preserving.*

Proof. The restriction of $\hat{\boldsymbol{\Omega}}$ to $\mathbf{X}_s$ is an isomorphism between $\mathbf{X}_s$ and $\hat{L}_2^s$ (see Rozanov, 1967, p. 32). This implies that $\hat{\boldsymbol{\Omega}}$ is an isomorphism between $\hat{\mathbf{X}}$ and $\hat{L}_2^{\infty}$. The conclusion follows from the fact that $\mathbf{X}$ and $L_2^{\infty}(\Theta, \mathbb{C}, \boldsymbol{\Sigma}^x)$ are the closure of $\hat{\mathbf{X}}$ and $\hat{L}_2^{\infty}$, respectively. ∎

The following lemma ensures that all vector stochastic processes belonging to $\mathbf{X}$ and costationary with the x's have a spectral density.

LEMMA 2. *Assume that the s-dimensional vector process* $\mathbf{y} = \{\mathbf{y}_t,\ t \in \mathbb{Z}\}$ *belongs to* $\mathbf{X}$ *and is costationary with the x's. Then*

(1) there exists a sequence of integers k_n *and coefficients* c_{jmkn}*, independent of t, such that* $y_{jt} = \lim_n \sum_{m=1}^{n} \sum_{k=-k_n}^{k_n} c_{jmkn} x_{mt-k}$*;*
(2) $\mathbf{y}$ *has a spectral density; i.e., there exists a Hermitian, nonnegative definite, integrable* $s \times s$ *matrix* $\boldsymbol{\Sigma}^y$ *such that* $E(\mathbf{y}_t \tilde{\mathbf{y}}_{t-k}) = (1/2\pi)\int_{-\pi}^{\pi} e^{ik\theta} \boldsymbol{\Sigma}^y(\theta)\, d\theta$.

Proof. Statement (1) is a trivial consequence of the definition of $\mathbf{X}$ and the costationarity assumption. To prove (2), let $\boldsymbol{\Sigma}^y$ be the matrix whose (i,j) entry is $\boldsymbol{\Omega}(y_{it})\boldsymbol{\Sigma}^x\tilde{\boldsymbol{\Omega}}(y_{jt})$; call it $\mathcal{S}(y_{it}, y_{jt};\theta)$. By the definition of $\boldsymbol{\Omega}$, and by statement (1), $\mathcal{S}(y_{it}, y_{jt};\theta)$ is independent of t. Note that $\boldsymbol{\Sigma}^y$ is Hermitian, nonnegative definite, and integrable. By the definition of $\boldsymbol{\Omega}$ and Lemma 1,

$$E(y_{it}\bar{y}_{jt-k}) = \langle y_{it}, y_{jt-k}\rangle = \langle \boldsymbol{\Omega}(y_{it}), \boldsymbol{\Omega}(y_{jt-k})\rangle_{\Sigma^x} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\theta}\mathcal{S}(y_{it}, y_{jt};\theta)d\theta$$

($\mathcal{S}(y_{it}, y_{jt};\theta)$ is usually referred to as the cross-spectrum between y_{it} and y_{jt}). ∎

DEFINITION 1. *As usual, we denote by L the lag operator, defined on* $\mathbf{X}$ *by linear extension of* $Lx_{it} = x_{it-1}$. *Given* $\mathbf{f} \in L_2^{\infty}(\Theta, \mathbb{C}, \boldsymbol{\Sigma}^x)$, *we define* $\underline{\mathbf{f}}(L)\mathbf{x}_t$ *by*

$$\underline{\mathbf{f}}(L)\mathbf{x}_t = \boldsymbol{\Omega}^{-1}(\mathbf{f}e^{it\cdot}). \tag{2}$$

The spectral density of the process $\{\underline{\mathbf{f}}(L)\mathbf{x}_t,\ t \in \mathbb{Z}\}$ is $\mathbf{f}\boldsymbol{\Sigma}^x\tilde{\mathbf{f}}$. The expression $\underline{\mathbf{f}}(L)$ must be used carefully. Suppose that $\boldsymbol{\Sigma}_n^x = \mathbf{I}_n$ ($\mathbf{x}_{nt}$ is an orthonormal white noise). Then the Fourier expansion $\mathbf{f}(\theta) = \sum_{k=-\infty}^{\infty} \mathbf{F}_k^{\mathbf{f}} e^{-ik\theta}$, where $\mathbf{F}_k^{\mathbf{f}} = (1/2\pi)\int_{-\pi}^{\pi} \mathbf{f}(\theta)e^{ik\theta}d\theta$, converges in $L_2^{\infty}(\Theta, \mathbb{C}, \boldsymbol{\Sigma}^x)$ (which is equal in this case to $L_2^{\infty}(\Theta,\mathbb{C})$). In this case we can define $\underline{\mathbf{f}}(L)$ as the linear filter $\sum_{k=-\infty}^{\infty}\mathbf{F}_k^{\mathbf{f}}L^k$, with

$$\underline{\mathbf{f}}(L)\mathbf{x}_t = \lim_s \sum_{k=-s}^{s} \mathbf{F}_k^{\mathbf{f}}\mathbf{x}_{t-k} \tag{3}$$

being of finite variance and therefore making sense in $\mathbf{X}$. However, in general $\underline{\mathbf{f}}(L)\mathbf{x}_t$, although the limit of finite linear combinations of the variables x_{jt-k}, cannot be represented as the sum of a series as in (3). In other words, in general $\underline{\mathbf{f}}(L)$ does not admit a separate definition as a filter and makes sense only within the expression $\underline{\mathbf{f}}(L)\mathbf{x}_t$, defined in (2). Given $\mathbf{f} \in L_2^n(\Theta,\mathbb{C},\boldsymbol{\Sigma}_n^x)$, $\underline{\mathbf{f}}(L)\mathbf{x}_{nt}$ is defined using the isomorphism $\boldsymbol{\Omega}_n$ between $\mathbf{X}_n$ and $L_2^n(\Theta,\mathbb{C},\boldsymbol{\Sigma}_n^x)$, where $\boldsymbol{\Omega}_n$ is defined, *mutatis mutandis*, as $\boldsymbol{\Omega}$ in equation (1) in Lemma 1.

If $\mathbf{y}_t$ is an s-dimensional vector belonging to $\mathbf{X}$ and costationary with the x's, and $\mathbf{A} \in L_2^{m\times s}(\Theta,\mathbb{C},\boldsymbol{\Sigma}^y)$, the m-dimensional vector $\underline{\mathbf{A}}(L)\mathbf{y}_t$ is defined applying $\boldsymbol{\Omega}_y^{-1}$ to each row of $\mathbf{A}e^{it\cdot}$, where $\boldsymbol{\Omega}_y$ is defined as $\boldsymbol{\Omega}$ in equation (1) in Lemma 1. If $\mathbf{A}$ is $m \times s$ and $\mathbf{B}$ is $n \times m$, and $\mathbf{BA}$ belongs to $L_2^{n\times s}(\Theta,\mathbb{C},\boldsymbol{\Sigma}^y)$, then we write $\underline{\mathbf{B}}(L)\underline{\mathbf{A}}(L)\mathbf{y}_t$ for $\underline{\mathbf{BA}}(L)\mathbf{y}_t$. Last, in expressions such as $\underline{\tilde{C}}(L)$ or $\underline{\tilde{C}}(L)\mathbf{y}_t$ it must be understood that *first* C is transformed by $\tilde{}$ and *second* _ is applied.

Remark 2. Given $y \in \mathbf{X}$, by definition, $y = \lim_n \sum_{m=1}^{n}\sum_{k=-k_n}^{k_n} a_{mkn}L^k x_{m0}$, for some coefficients a_{mkn}. Defining $y_t = \lim_n \sum_{m=1}^{n}\sum_{k=-k_n}^{k_n} a_{mkn}L^k x_{mt}$, the process $\{y_t,\ t \in \mathbb{Z}\}$ belongs to $\mathbf{X}$, is costationary with the x's, and contains y. With the preceding argument in mind, the generic element of $\mathbf{X}$ will often be referred to as y_t, z_t, etc., rather than y, z, etc., where y_t, z_t, etc., are costationary and costationary with the x's. Analogous considerations hold if we consider a vector $\mathbf{y}$ belonging to $\mathbf{X}$.

2.2 Properties of the Spectral Density Matrices $\boldsymbol{\Sigma}_n^x$

Now we give some definitions and results on eigenvalues and eigenvectors of the spectral density matrices $\boldsymbol{\Sigma}_n^x$.

DEFINITION 2. *For $i = 1,2,\ldots,n$, let $\lambda_{ni}^x : \Theta \to \mathbb{R}$ be defined as the function associating with $\theta \in \Theta$ the ith eigenvalue, in descending order, of $\boldsymbol{\Sigma}_n^x(\theta)$. The functions λ_{ni}^x will be called the dynamic eigenvalues of $\boldsymbol{\Sigma}_n^x$.*

Remark 3. We use "dynamic" for eigenvectors and eigenvalues of $\boldsymbol{\Sigma}_n^x$ to insist on the difference between the dynamic analysis developed here and the static approach, based on the eigenvalues of variance-covariance matrices. On eigenvalues and eigenvectors of spectral density matrices, and related filters, see Brillinger (1981, Ch. 9).

The following lemma is an elementary consequence of well-known results.

LEMMA 3. *The functions λ^x_{ni} are Lebesgue-measurable and integrable in Θ for any $n \in \mathbb{N}$ and $i \leq n$.*

Proof. Measurability is a consequence of (a) continuity of the eigenvalues with respect to the entries of $\mathbf{\Sigma}^x_n$ (for continuity of the roots of a polynomial see, e.g., Ahlfors, 1987, pp. 300–306); (b) measurability of the entries of $\mathbf{\Sigma}^x_n$ as functions of θ (recall that $\mathbf{\Sigma}^x_n$ is integrable); (c) measurability of a continuous function of a measurable function (see, for the real case, Royden, 1988, Problem 25, p. 71; extension to the complex case is immediate). For integrability, note that for any n, $i \leq n$ and θ, $0 \leq \lambda^x_{ni}(\theta) \leq \sum_{s=1}^n \lambda^x_{ns}(\theta) = \mathrm{trace}(\mathbf{\Sigma}^x_n(\theta))$ and that $(1/2\pi)\int_{-\pi}^{\pi} \mathrm{trace}(\mathbf{\Sigma}^x_n(\theta))\, d\theta = \mathrm{E}(|\mathbf{x}_{nt}|^2) < \infty$. ∎

Let us recall some properties of the eigenvalues of Hermitian nonnegative definite matrices.

FACT M.

(a) Let $\mathbf{D}$ and $\mathbf{E}$ be $m \times m$ Hermitian nonnegative definite and $\mathbf{F} = \mathbf{D} + \mathbf{E}$. Then

$$\lambda^F_s \leq \lambda^D_s + \lambda^E_1, \quad \lambda^F_s \leq \lambda^D_1 + \lambda^E_s, \quad \lambda^F_s \geq \lambda^D_s, \quad \lambda^F_s \geq \lambda^E_s \tag{4}$$

for any $s = 1,2,\ldots,m$.

(b) Let $\mathbf{D}$ be as in (a) and let $\mathbf{G}$ be the top-left $(m-1) \times (m-1)$ submatrix of $\mathbf{D}$. Then $\lambda^D_s \geq \lambda^G_s$ for $s = 1,2,\ldots,m-1$.

Proof. Because $(\mathbf{D} + \lambda^E_1\mathbf{I}_m) - \mathbf{F} = \lambda^E_1\mathbf{I}_m - \mathbf{E}$ and $\mathbf{F} - \mathbf{D} = \mathbf{E}$ are Hermitian nonnegative definite, the first and third inequalities in (4) follow from Lancaster and Tismenetsky (1985, Theorem 1, p. 301); analogously for the second and fourth; statement (b) follows from Corollary 1, p. 293 in the same book. ∎

Because the spectral density matrices $\mathbf{\Sigma}^x_n$ are nested as in Fact M, statement (b), then we can state the following lemma.

LEMMA 4. *Given i, for $n \geq i$, $\lambda^x_{ni}(\theta)$ is nondecreasing as a function of n for any $\theta \in \Theta$, i.e., $\lambda^x_{ni}(\theta) \leq \lambda^x_{n+1i}(\theta)$.*

A consequence of Lemma 4 is that $\lim_n \lambda^x_{ni}(\theta)$ exists for any i and θ and equals $\sup_n \lambda^x_{ni}(\theta)$.

DEFINITION 3. *For any i we define the function λ^x_i by $\lambda^x_i(\theta) = \sup_n \lambda^x_{ni}(\theta)$.*

It must be pointed out that λ^x_i is an extended real function, i.e., its value may be infinite. Note also that λ^x_i is measurable (see Royden, 1988, Theorem 20, p. 68) and that $\{\theta : \lambda^x_i(\theta) = \infty\}$ may be of null or positive measure and may even coincide with Θ.

Now consider the system of equations

$$\mathbf{p}(\theta)[\mathbf{\Sigma}^x_n(\theta) - \lambda^x_{n1}(\theta)\mathbf{I}_n] = 0, \quad |\mathbf{p}(\theta)| = 1. \tag{5}$$

Because the functions λ^x_{ni} are measurable by Lemma 3, the coefficients of (5) are measurable. Determining a solution to (5) that is continuous with respect to the coefficients, and therefore measurable with respect to θ, is a simple exercise. Call $\mathbf{p}^x_{n1}$ such a solution. Recursively, for $i > 1$, we can determine $\mathbf{p}^x_{ni}$ as a measurable solution to

$$\mathbf{p}(\theta)[\mathbf{\Sigma}^x_n(\theta) - \lambda^x_{ni}(\theta)\mathbf{I}_n] = 0, \quad \mathbf{p}(\theta)\tilde{\mathbf{p}}^x_{nj}(\theta) = 0, \quad 1 \le j < i, \quad |\mathbf{p}(\theta)| = 1.$$

Thus we have the following lemma.

LEMMA 5. *There exist n functions $\mathbf{p}^x_{ni}$, $i = 1,2,\dots,n$, belonging to $L^n_\infty(\Theta,\mathbb{C})$, and therefore to $L^n_2(\Theta,\mathbb{C},\mathbf{\Sigma}^x) \cap L^n_2(\Theta,\mathbb{C})$, such that*

(1) $|\mathbf{p}^x_{ni}(\theta)| = 1$, *for any* $\theta \in \Theta$;
(2) $\mathbf{p}^x_{ni}(\theta)\tilde{\mathbf{p}}^x_{nj}(\theta) = 0$, *for* $i \neq j$ *and any* $\theta \in \Theta$;
(3) $\mathbf{p}^x_{ni}(\theta)\mathbf{\Sigma}^x_n(\theta) = \lambda^x_{ni}(\theta)\mathbf{p}^x_{ni}(\theta)$ *for any* $\theta \in \Theta$.

DEFINITION 4. *An n-tuple of functions $\mathbf{p}^x_{ni}$ fulfilling (1)–(3) of Lemma 5 will be called a set of* dynamic eigenvectors *associated with $\mathbf{x}_{nt}$.*

DEFINITION 5. *If the functions $\mathbf{p}^x_{nj}$, $j = 1,2,\dots,n$, form a set of dynamic eigenvectors, then $\underline{\mathbf{p}}^x_{nj}(L)\mathbf{x}_{nt}$, $j = 1,2,\dots,n$, is a set of* dynamic principal components *associated with $\mathbf{x}_{nt}$.*

Remark 4. Note that dynamic eigenvectors and dynamic principal components associated with $\mathbf{x}_{nt}$ are not unique.

3. DYNAMIC AVERAGING SEQUENCES, AGGREGATION SPACE, IDIOSYNCRATIC VARIABLES

In the Introduction we have considered averages of the x's in which the sum of the squared weights tends to zero. The function spaces introduced in Section 2 now permit a precise definition.

DEFINITION 6. *Let $\mathbf{a}_n \in L^\infty_2(\Theta,\mathbb{C}) \cap L^\infty_2(\Theta,\mathbb{C},\mathbf{\Sigma}^x)$ for $n \in \mathbb{N}$. The sequence $\{\mathbf{a}_n,\ n \in \mathbb{N}\}$ is a* dynamic averaging sequence, DAS *henceforth, if $\lim_n \|\mathbf{a}_n\| = 0$, i.e., if $\mathbf{a}_n$ converges to zero in the norm of $L^\infty_2(\Theta,\mathbb{C})$.*

Example 1

Define $\hat{L}^n_\infty$ as $\{\mathbf{f}^{[n]}: \mathbf{f} \in L^\infty_2(\Theta,\mathbb{C}),\ \mathbf{f}^{\{n\}} \in L^n_\infty(\Theta,\mathbb{C})\}$. Note that $L^\infty_2(\Theta,\mathbb{C}) \cap L^\infty_2(\Theta,\mathbb{C},\mathbf{\Sigma}^x) \supseteq \hat{L}^n_\infty$ for any n and is therefore never trivial. In particular, the sequence

$$\mathbf{d}_n = \frac{1}{n}(\underbrace{1 \quad 1 \quad \dots \quad 1}_{n} \quad 0 \quad 0 \quad \dots) \in \hat{L}^n_\infty,$$

producing arithmetic averages, is obviously a DAS.

DEFINITION 7. *Let* $y_t \in \mathbf{X}$. *We say that* y_t *is an* aggregate *if there exists a DAS* $\{\mathbf{a}_n, n \in \mathbb{N}\}$ *such that* $\lim_n \underline{\mathbf{a}}_n(L)\mathbf{x}_t = y_t$. *The set of all the aggregates will be denoted by* $\mathcal{G}(\mathbf{x})$ *and called the* aggregation subspace *of* $\mathbf{X}$.

LEMMA 6. *The set* $\mathcal{G}(\mathbf{x})$ *is a closed subspace of* $\mathbf{X}$.

Proof. Assume that $z_t = \lim_m y_{mt}$, with $y_{mt} \in \mathcal{G}(\mathbf{x})$. Let $y_{mt} = \lim_n \underline{\mathbf{a}}_{mn}(L)\mathbf{x}_t$, where $\{\mathbf{a}_{mn}, n \in \mathbb{N}\}$ is a DAS for any m. Let m_i be such that $\|z_t - y_{m_i t}\| < 1/i$ and n_i such that $\|\mathbf{a}_{m_i n_i}\| < 1/i$ and $\|y_{m_i t} - \underline{\mathbf{a}}_{m_i n_i}(L)\mathbf{x}_t\| < 1/i$. The sequence

$$\{\mathbf{a}_{m_1 n_1} \quad \mathbf{a}_{m_2 n_2} \quad \cdots \quad \}$$

is a DAS and

$$\|z_t - \underline{\mathbf{a}}_{m_i n_i}(L)\mathbf{x}_t\| \leq \|z_t - y_{m_i t}\| + \|y_{m_i t} - \underline{\mathbf{a}}_{m_i n_i}(L)\mathbf{x}_t\| < 2/i. \qquad \blacksquare$$

DEFINITION 8. *Consider the projection equation*

$$x_{it} = \mathrm{proj}(x_{it}|\mathcal{G}(\mathbf{x})) + \delta_{it}. \tag{6}$$

Decomposition (6) will be called the canonical decomposition *of* $\mathbf{x}$.

DEFINITION 9. *We say that* $\mathbf{x}$ *is* idiosyncratic *if* $\lim_n \underline{\mathbf{a}}_n(L)\mathbf{x}_t = 0$ *for any DAS* $\{\mathbf{a}_n, n \in \mathbb{N}\}$.

If $\mathbf{x}$ is idiosyncratic then obviously $\mathcal{G}(\mathbf{x}) = \{0\}$ and the canonical decomposition is trivial with $\delta_{it} = x_{it}$. However, as the next example shows, the converse does not hold.

Example 2

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

$$\mathbf{c}_n = \frac{1}{\sqrt{n}}(\underbrace{0 \quad 0 \quad \ldots \quad 0 \quad 1}_{n} \quad 0 \quad 0 \quad \ldots).$$

The sequence $\{\mathbf{c}_n, n \in \mathbb{N}\}$ is a DAS. Moreover $\|\mathbf{c}_n\mathbf{x}_t\|^2 = 1$, so that $\mathbf{x}$ is not idiosyncratic. Now let y_t be an aggregate, so that

$$y_t = \lim_n \underline{\mathbf{a}}_n(L)\mathbf{x}_t = \lim_n \sum_{j=1}^{\infty} \underline{a}_{nj}(L)x_{jt} = \lim_n \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} a_{njk}x_{jt-k},$$

where $\{\mathbf{a}_n, n \in \mathbb{N}\}$ is a DAS. Because $y_t \in \mathbf{X}$ and the x_{it}'s are mutually orthogonal white noises, then

$$y_t = \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} b_{jk}x_{jt-k}. \tag{7}$$

Moreover, representation (7) is unique, and $\lim_n a_{njk} = b_{jk}$ for any j and k. On the other hand, because $\{\mathbf{a}_n, n \in \mathbb{N}\}$ is a DAS, $\lim_n \sum_{j=1}^{n} \sum_{k=-\infty}^{\infty} |a_{njk}|^2 = 0$, so that $b_{jk} = 0$ for any j and k, i.e., $y_t = 0$. Thus $\mathcal{G}(\mathbf{x}) = \{0\}$ although $\mathbf{x}$ is not idiosyncratic.

If the vector $\mathbf{x}_{nt}$ is a white noise for any n, i.e., if the matrix $\boldsymbol{\Sigma}_n^x$ and its eigenvalues are constant as functions of θ, then $\mathbf{x}$ is idiosyncratic if and only if λ_{n1}^x is bounded as a function of n (see Chamberlain, 1983; Chamberlain and Rothschild, 1983). Theorem 1 generalizes this result to any $\mathbf{x}$ fulfilling Assumption 1.

THEOREM 1. *The following three statements are equivalent:*

(a) $\mathbf{x}$ *is idiosyncratic.*
(b) λ_1^x *is essentially bounded.*
(c) Consider the space $\boldsymbol{\Psi} = L_2^\infty(\Theta, \mathbb{C}) \cap L_2^\infty(\Theta, \mathbb{C}, \boldsymbol{\Sigma}^x)$, *with the* $\|\cdot\|$ *norm, and define* $\boldsymbol{\Upsilon}: \boldsymbol{\Psi} \to L_2^\infty(\Theta, \mathbb{C}, \boldsymbol{\Sigma}^x)$ *as* $\boldsymbol{\Upsilon}(\mathbf{f}) = \mathbf{f}$. *Here* $\boldsymbol{\Upsilon}$ *is continuous.*

Proof. Let us recall that continuity at 0, continuity everywhere, and boundedness are equivalent for linear maps between normed vector spaces (Royden, 1988, Proposition 2, p. 220). Because $\|\underline{\mathbf{a}}_n(L)\mathbf{x}_t\| = \|\mathbf{a}_n\|_{\Sigma^x} = \|\boldsymbol{\Upsilon}(\mathbf{a}_n)\|_{\Sigma^x}$, Definition (9) is equivalent to continuity of $\boldsymbol{\Upsilon}$ at 0. Thus (a) and (c) are equivalent. Moreover, defining $\|\boldsymbol{\Upsilon}\| = \sup_{\mathbf{f}\in\boldsymbol{\Psi}, \|\mathbf{f}\|=1} \|\boldsymbol{\Upsilon}(\mathbf{f})\|_{\Sigma^x}$, boundedness of $\boldsymbol{\Upsilon}$ is equivalent to $\|\boldsymbol{\Upsilon}\| < \infty$ (Royden, 1988, p. 220). Because we prove that $\|\boldsymbol{\Upsilon}\| = \sqrt{\text{ess sup}(\lambda_1^x)}$, (b) and (c) are equivalent.

To show that $\|\boldsymbol{\Upsilon}\| = \sqrt{\text{ess sup}(\lambda_1^x)}$, let $\psi_n = \sup_{\mathbf{f}\in\boldsymbol{\Psi}, \|\mathbf{f}\|=1} \|\boldsymbol{\Upsilon}(\mathbf{f}^{[n]})\|_{\Sigma^x}$. Because $\boldsymbol{\Sigma}_n^x(\theta)$ is Hermitian for any θ, then (see Lancaster and Tismenetsky, 1985, Theorem 4, p. 285)

$$\psi_n^2 = \sup_{\mathbf{f}\in\boldsymbol{\Psi},\, \|\mathbf{f}\|=1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{f}^{\{n\}}(\theta) \boldsymbol{\Sigma}_n^x(\theta) \tilde{\mathbf{f}}^{\{n\}}(\theta) d\theta$$

$$= \sup_{h\in\Psi_n,\, \|h\|=1} \frac{1}{2\pi} \int_{-\pi}^{\pi} |h(\theta)|^2 \lambda_{n1}^x(\theta) d\theta, \tag{8}$$

where $\Psi_n = L_2(\Theta, \mathbb{C}) \cap L_2(\Theta, \mathbb{C}, \lambda_{n1}^x)$. Applying a standard argument to the last term of (8) (see, e.g., Conway, 1985, Theorem 1.5, p. 28), we have $\psi_n^2 = \text{ess sup}(\lambda_{n1}^x)$. The result follows from $\|\boldsymbol{\Upsilon}\| = \lim_n \psi_n$, and $\lim_n \text{ess sup}(\lambda_{n1}^x) = \text{ess sup}(\lim_n \lambda_{n1}^x) = \text{ess sup}(\lambda_1^x)$. ■

COROLLARY. *If* $\mathbf{x}$ *is idiosyncratic then*

$$\sup_n \int_{-\pi}^{\pi} \lambda_{n1}^x(\theta) d\theta = \lim_n \int_{-\pi}^{\pi} \lambda_{n1}^x(\theta) d\theta < \infty.$$

Proof. Because λ_1^x is essentially bounded, we have $\int_{-\pi}^{\pi} \lambda_1^x(\theta)d\theta < \infty$. Moreover, λ_{n1}^x converges monotonically a.e. in Θ to λ_1^x. Thus, by the monotone convergence theorem (Royden, 1988, p. 87),

$$\lim_n \int_{-\pi}^{\pi} \lambda_{n1}^x(\theta)d\theta = \int_{-\pi}^{\pi} \lambda_1^x(\theta)d\theta < \infty. \qquad \blacksquare$$

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

Example 3

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 any nonnegative function f, independent of i, with $f \in L_1(\Theta,\mathbb{R}) - L_\infty(\Theta,\mathbb{R})$. In this case the matrix $\mathbf{\Sigma}_n^x$ is diagonal, $\lambda_1^x = f$, which is not essentially bounded. Thus $\mathbf{x}$ is not idiosyncratic, even though $\sup_n \int_{-\pi}^{\pi} \lambda_{n1}^x(\theta)d\theta < \infty$.

Adapting the argument used to obtain (8) we have $\sqrt{(2\pi)^{-1}\int_{-\pi}^{\pi} \mathbf{f}(\theta)\mathbf{\Sigma}^x(\theta)\tilde{\mathbf{f}}(\theta)d\theta} \leq \sqrt{\text{ess sup}(\lambda_1^x)}\|\mathbf{f}\|$ for any $\mathbf{f} \in L_2^\infty(\Theta,\mathbb{C})$. Thus if $\mathbf{x}$ is idiosyncratic $L_2^\infty(\Theta,\mathbb{C}) \subseteq L_2^\infty(\Theta,\mathbb{C},\mathbf{\Sigma}^x)$, and therefore $\mathbf{\Psi} = L_2^\infty(\Theta,\mathbb{C})$. The inclusion can be strict, as the following example shows.

Example 4

Let $\mathbf{\Sigma}_n^x(\theta) = |1 - e^{-i\theta}|^2\mathbf{I}_n$. In this case $\mathbf{x}$ is idiosyncratic, so that $L_2^\infty(\Theta,\mathbb{C}) \subseteq L_2^\infty(\Theta,\mathbb{C},\mathbf{\Sigma}^x)$. However, the opposite inclusion relation does not hold. Consider, e.g., $\mathbf{f}(\theta) = (1 - e^{-i\theta})^{-1}(1 \quad 0 \quad 0 \quad \ldots \quad)$. Here $\mathbf{f}$ belongs to $L_2^\infty(\Theta,\mathbb{C},\mathbf{\Sigma}^x)$ but not to $L_2^\infty(\Theta,\mathbb{C})$.

4. A FINITE NUMBER OF DYNAMIC COMMON FACTORS

4.1. The Generalized Dynamic Factor Model

Note that dynamic averaging of $\mathbf{x}$, according to Definition 6, is nothing but averaging simultaneously both in the cross section and the time dimension. It is easy to show that the same aggregation space would result by taking finite averages in one of the two dimensions or in both. In particular, if $y \in \mathcal{G}(\mathbf{x})$, then there exist a sequence of integers s_n and a sequence $\{\mathbf{a}_n,\ n \in \mathbb{N},\ \mathbf{a}_n \in L_2^{s_n}(\Theta,\mathbb{C}) \cap L_2^{s_n}(\Theta,\mathbb{C},\mathbf{\Sigma}_{s_n}^x)\}$ such that $\lim_n\|\mathbf{a}_n\| = 0$ and $\lim_n \underline{\mathbf{a}}_n(L)\mathbf{x}_{s_n t} = y_t$. Thus an equivalent definition of a DAS, which will be used in the present section, is that of a sequence $\mathbf{a}_n \in L_2^{s_n}(\Theta,\mathbb{C}) \cap L_2^{s_n}(\Theta,\mathbb{C},\mathbf{\Sigma}_{s_n}^x)$ such that $\lim_n\|\mathbf{a}_n\| = 0$.

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

DEFINITION 10. *Let q be a nonnegative integer. The double sequence* $\mathbf{x}$ *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*

$$\mathbf{u} = \{(u_{1t} \quad u_{2t} \quad \dots \quad u_{qt})', t \in \mathbb{Z}\} = \{\mathbf{u}_t, t \in \mathbb{Z}\},$$

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

(i) For any $i \in \mathbb{N}$,

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

$$\chi_{it} = \underline{b}_{i1}(L)u_{1t} + \underline{b}_{i2}(L)u_{2t} + \cdots + \underline{b}_{iq}(L)u_{qt} = \underline{\mathbf{b}}_i(L)\mathbf{u}_t, \qquad \mathbf{(9)}$$

where $\mathbf{b}_i \in L_2^q(\Theta,\mathbb{C})$.

(ii) For any $i \in \mathbb{N}$, $j = 1,2,\dots,q$, and $k \in \mathbb{Z}$, we have $\xi_{it} \perp u_{jt-k}$. As a consequence $\xi_{it} \perp \chi_{st-k}$ for any $i \in \mathbb{N}$, $s \in \mathbb{N}$, and $k \in \mathbb{Z}$.

(iii) λ_1^ξ is essentially bounded, i.e., $\boldsymbol{\xi}$ is idiosyncratic.

(iv) Putting $\boldsymbol{\chi} = \{\chi_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$, $\lambda_q^\chi(\theta) = \infty$ a.e. in Θ.

The double sequences $\boldsymbol{\chi}$ and $\boldsymbol{\xi}$ are referred to as the common and the idiosyncratic component of representation (9).

THEOREM 2. *The double sequence* $\mathbf{x}$ *is a q-DFS if and only if*

(i) λ_{q+1}^x is essentially bounded;

(ii) $\lambda_q^x(\theta) = \infty$ a.e. in Θ.

Remark 5. Forni et al. (2000) propose a heuristic criterion to determine in empirical cases the number q such that (i) and (ii) hold. Because 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 factor model and that the number of factors is q.

THEOREM 3. *If* $\mathbf{x}$ *is a q-DFS with representation (9) then*

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

Moreover,

$$\chi_{it} = \text{proj}(x_{it}|\mathcal{G}(\mathbf{x})). \qquad \mathbf{(10)}$$

An immediate but very important consequence of (10) is that if $\mathbf{x}$ is a q-DFS then the components χ_{it} and ξ_{it} are uniquely determined. Precisely, we can state the following theorem.

THEOREM 4. *Suppose that* $\mathbf{x}$ *is a q-DFS with representation (9). Suppose further that there exists an s-dimensional orthonormal white-noise vector process* $\mathbf{v}$, *with* $v_{jt} \in L_2(\mathcal{P}, \mathbb{C})$, *such that*

$$x_{it} = \omega_{it} + \zeta_{it},$$
$$\omega_{it} = \underline{\mathbf{c}}_i(L)\mathbf{v}_t,$$

where $\mathbf{c}_i \in L_2^s(\Theta, \mathbb{C})$, *and that* λ_1^ζ *and* λ_s^ω *fulfill, respectively, conditions* (*iii*) *and* (*iv*) *of Definition* 10. *Then* $s = q$, $\omega_{it} = \chi_{it}$, *and* $\zeta_{it} = \xi_{it}$.

Several observations are in order.

Remark 6. Theorem 3 implies that both χ_{it} and ξ_{it} belong to $\mathbf{X}$. Theorem 4 implies that no representation fulfilling Definition 10 is possible with the common or the idiosyncratic component not belonging to $\mathbf{X}$.

Remark 7. It must be pointed out that the components are unique, not $\mathbf{u}_t$ or the filters $\underline{\mathbf{b}}_i(L)$. Precisely, if (9) holds, all possible representations of χ_{it} are obtained by setting $\chi_{it} = \underline{\mathbf{d}}_i(L)\mathbf{w}_t$, $\mathbf{w}_t = \underline{\mathbf{C}}(L)\mathbf{u}_t$, $\mathbf{d}_i = \mathbf{b}_i\widetilde{\mathbf{C}}$, where $\mathbf{C} \in L_2^{q\times q}(\Theta, \mathbb{C})$ and $\widetilde{\mathbf{C}}\mathbf{C} = \mathbf{I}_q$.

Remark 8. Because $\mathbf{u}$ is an orthonormal white noise the function $\underline{\mathbf{b}}_i(L)\mathbf{u}_t \in \mathbf{L}_2(\mathcal{P}, \mathbb{C})$ if and only if $\mathbf{b}_i \in L_2^q(\Theta, \mathbb{C})$. As a consequence $\underline{\mathbf{b}}_i(L)$ has a representation as a filter (see Section 2.1). Note that Definition 10 does not exclude that the filters $\underline{b}_{ij}(L)$ are two-sided. If representation (9) must have a structural interpretation then it is reasonable to assume that the filters $\underline{b}_{ij}(L)$ are one-sided. However, one-sidedness of the $\underline{b}_{ij}(L)$ has no consequences on the eigenvalues λ_{nj}^χ or λ_{nj}^x, nor does fulfillment of conditions (i) and (ii) have 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 $\mathbf{u}_t$, which is uniquely determined (Theorems 2–4), and with the reconstruction of χ_{it} and ξ_{it} (see Theorem 5, which follows). Existence and identification of one-sided representations of the common component are left to further study.

Remark 9. The result $s = q$ in Theorem 4 can be restated by saying that if $\mathbf{x}$ is a q-DFS, then q is minimal, i.e., no representation fulfilling Definition 10 is possible with a smaller number of factors. It is important to point out that this is no longer true if condition (iv) in Definition 10 does not hold. For example, suppose that

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

with $\boldsymbol{\xi}$ idiosyncratic and $\sum |b_i|^2 < \infty$. In this case $\lambda_1^\chi < \infty$. As a consequence, $b_i u_t + \xi_{it}$ is idiosyncratic, so that a representation with zero factors is possible.

Remark 10. Suppose that $\mathbf{x}_{nt}$ 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 λ_{nj}^x are constant as functions of θ. As a consequence, if $\lambda_s^x < \infty$, the model has q factors, with $q < s$. Unfortunately, in the general dynamic case, there exist cases where λ_s^x is essentially bounded but the sequence does not fulfill Definition 10 for any $q < s$. Consider

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

with $\boldsymbol{\xi}$ idiosyncratic and

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

Here $\lambda_2^x(\theta)$ is essentially bounded, but $\lambda_1^x(\theta)$ is infinite only for $\theta \in [-1,1]$, finite elsewhere. The analysis of such cases is left to further work.

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

$$x_{it} = \text{proj}(x_{it}|\mathcal{G}(\mathbf{x})) + \delta_{it} = \underline{\mathbf{c}}_i(L)\mathbf{z}_t + \delta_{it}.$$

Last, in Section 4.6 we show that $\boldsymbol{\delta}$ is idiosyncratic, thus completing the proof of Theorem 2. In Section 4.7 we prove Theorem 3.

4.2. An Additional Assumption on $\mathbf{x}$

Theorems 2 and 3 will be proved supposing that the following assumption holds.

Assumption 2. For any $n \in \mathbb{N}$, $j \leq n$ and $\theta \in \Theta$, $\lambda_{nj}^x(\theta) \geq 1$.

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

(a) $\boldsymbol{\Sigma}_n^y = \boldsymbol{\Sigma}_n^x + \mathbf{I}_n$, $\lambda_{nj}^y = \lambda_{nj}^x + 1$. Thus if conditions (i) and (ii) in Theorem 2 hold for $\mathbf{x}$, then they hold for $\mathbf{y}$ also. By Theorem 2 $\mathbf{y}$ is a q-DFS with representation $y_{it} = \check{\chi}_{it} + \check{\xi}_{it}$. By Theorem 3, $\check{\chi}_{it} = \text{proj}(y_{it}|\mathcal{G}(\mathbf{y}))$. But the definitions of $\hat{\boldsymbol{\xi}}$ and $\mathbf{y}$ imply that $\check{\chi}_{it} = \text{proj}(x_{it}|\mathcal{G}(\mathbf{x}))$. Therefore

$$x_{it} = \text{proj}(x_{it}|\mathcal{G}(\mathbf{x})) + (\check{\xi}_{it} - \hat{\xi}_{it}). \tag{11}$$

Because $\hat{\xi}_{it}$ is orthogonal to $\mathbf{X}$ and $\check{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}}$ is idiosyncratic, then (11) is a q-DFS representation. Thus if (i) and (ii) hold for $\mathbf{x}$, then $\mathbf{x}$ has a q-DFS representation.

(b) If $\mathbf{x}$ has the q-DFS representation $x_{it} = \chi_{it} + \xi_{it}$, then $\mathbf{y}$ has the q-DFS representation $y_{it} = \chi_{it} + (\xi_{it} + \hat{\xi}_{it})$. Applying Theorem 2 to $\mathbf{y}$, we obtain conditions (i) and (ii) for $\lambda_q^y = \lambda_q^x + 1$ and $\lambda_{q+1}^y = \lambda_{q+1}^x + 1$ and therefore for λ_q^x and λ_{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 $\mathbf{y}$, Theorem 3 can be proved in general.

4.3. The Necessity of Conditions (i) and (ii) for a q-DFS

Let us prove that if $\mathbf{x}$ is a q-DFS then (i) and (ii) in Theorem 2 hold. By Definition 10, $\boldsymbol{\Sigma}_n^x(\theta) = \boldsymbol{\Sigma}_n^\chi(\theta) + \boldsymbol{\Sigma}_n^\xi(\theta)$. By Fact M, the third inequality in (4), $\lambda_{nq}^x(\theta) \geq \lambda_{nq}^\chi(\theta)$, so that (ii) is proved. By the first inequality in (4),

$$\lambda_{nq+1}^x(\theta) \leq \lambda_{nq+1}^\chi(\theta) + \lambda_{n1}^\xi(\theta) = \lambda_{n1}^\xi(\theta), \tag{12}$$

so that (i) is proved. Moreover, (12) implies the following interesting inequality:

$$\lambda_{q+1}^x(\theta) \leq \lambda_1^\xi(\theta) \tag{13}$$

(the opposite inequality is proved in Section 4.7).

4.4. Construction of a q-Dimensional White-Noise Vector

Now we start assuming (i) and (ii) in Theorem 2. First we prove that $\mathcal{G}(\mathbf{x})$ contains a q-dimensional white-noise vector. The proof goes as follows. We start with a q-dimensional orthonormal white noise, call it $\boldsymbol{\psi}_t$, whose entries are linear combinations of the mth order principal components $\underline{\mathbf{p}}_{mj}^x(L)\mathbf{x}_{mt}$, for $j = 1,2,\ldots,q$, $t \in \mathbb{Z}$. Then we project $\boldsymbol{\psi}_t$ on the space spanned by the nth order principal components $\underline{\mathbf{p}}_{nj}^x(L)\mathbf{x}_{nt}$, $j = 1,2,\ldots,q$, $t \in \mathbb{Z}$, for $n > m$, call $\mathbf{y}_t$ the projection. We show that when m and n become large the distance between $\boldsymbol{\psi}_t$ and $\mathbf{y}_t$ becomes small. This leads to the construction of a sequence of q-dimensional white-noise vectors whose components are Cauchy sequences and converge to $\mathcal{G}(\mathbf{x})$.

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

$$x_{it} = (1 - L)u_t + \xi_{it}, \tag{14}$$

with $\boldsymbol{\Sigma}_n^\xi = \mathbf{I}_n$, in which $\boldsymbol{\Sigma}_n^x$ is continuous and $\boldsymbol{\Sigma}_n^x(0) = \mathbf{I}_n$ for any n. Unfortunately, to include cases such as (14) our proofs must be carried over piecewise on Θ.

For $q \le n$, we denote by $\mathbf{P}_n$ the $q \times n$ matrix

$$(\mathbf{p}_{n1}^{x\,\prime} \quad \mathbf{p}_{n2}^{x\,\prime} \quad \cdots \quad \mathbf{p}_{nq}^{x\,\prime})',$$

i.e., the matrix having the dynamic eigenvectors $\mathbf{p}_{nj}^{x}$, $j = 1,2,\ldots,q$, on the rows, and by $\mathbf{Q}_n$ the $(n-q) \times n$ matrix

$$(\mathbf{p}_{nq+1}^{x\,\prime} \quad \mathbf{p}_{nq+2}^{x\,\prime} \quad \cdots \quad \mathbf{p}_{nn}^{x\,\prime})'.$$

Moreover, let us call $\mathbf{\Lambda}_n$ the $q \times q$ diagonal matrix having on the diagonal the eigenvalues λ_{nj}^{x}, $j = 1,2,\ldots,q$, and denote by $\mathbf{\Phi}_n$ the $(n-q) \times (n-q)$ diagonal matrix having on the diagonal the eigenvalues λ_{nj}^{x}, $j = q+1,\ldots,n$. The matrices $\mathbf{\Sigma}_n^x$ and $\mathbf{I}_n$ can be rewritten in their spectral decomposition form (see Lancaster and Tismenetsky, 1985, Exercise 5, p. 175):

$$\mathbf{\Sigma}_n^x = \tilde{\mathbf{P}}_n \mathbf{\Lambda}_n \mathbf{P}_n + \tilde{\mathbf{Q}}_n \mathbf{\Phi}_n \mathbf{Q}_n,$$

$$\mathbf{I}_n = \tilde{\mathbf{P}}_n \mathbf{P}_n + \tilde{\mathbf{Q}}_n \mathbf{Q}_n. \tag{15}$$

Because $\mathbf{\Lambda}_n^{-1}$ is bounded in Θ by Assumption 2, $\mathbf{\Lambda}_n^{-1}\mathbf{P}_n \in L_\infty^{n\times n}(\Theta,\mathbb{C})$, so that the definition

$$\boldsymbol{\psi}_t^n = (\psi_{1t}^n \quad \psi_{2t}^n \quad \cdots \quad \psi_{qt}^n)' = \underline{\mathbf{\Lambda}}_n^{-1/2}(L)\underline{\mathbf{P}}_n(L)\mathbf{x}_{nt}$$

makes sense and $\boldsymbol{\psi}_t^n$ is an orthonormal white noise. Note that the processes ψ_{jt}^n, $j = 1,2,\ldots,q$, are the dynamic principal components associated with the largest q eigenvectors, rescaled so that the spectral density is equal to $\mathbf{I}_n$.

DEFINITION 11. *Let $M \subseteq \Theta$. We denote by K_M the subset of $L_\infty^{q\times q}(\Theta,\mathbb{C})$ whose elements $\mathbf{C}$ are such that*

(i) $\mathbf{C}(\theta) = \mathbf{0}_q$ *for* $\theta \notin M$,
(ii) $\mathbf{C}(\theta)\tilde{\mathbf{C}}(\theta) = \mathbf{I}_q$ *for* $\theta \in M$.

In the sequel, so as not to complicate notation, we write matrix products $\mathbf{AB}$ in which the number of columns of $\mathbf{A}$ is smaller than the number of rows of $\mathbf{B}$. In this case we implicitly assume that $\mathbf{A}$ has been augmented with columns of zeros to match the number of rows of $\mathbf{B}$. For example, we write $\underline{\mathbf{P}}_m(L)\mathbf{x}_{nt}$ for $n > m$, this meaning nothing other than $\underline{\mathbf{P}}_m(L)\mathbf{x}_{mt}$. In the same way, we have equations with a $1 \times m$ matrix on one side and a $1 \times n$ matrix on the other, with $m < n$, this meaning that the $1 \times m$ matrix has been augmented with zeros.

Now let $\mathbf{C} \in K_M$, so that $\underline{\mathbf{C}}(L)\boldsymbol{\psi}_t^m$ makes sense as a vector belonging to $\mathbf{X}$. We want to determine the (element by element) orthogonal projection of the vector $\underline{\mathbf{C}}(L)\boldsymbol{\psi}_t^m$ on the space

$$\overline{\text{span}}(\{\psi_{jt}^n, j = 1,2,\ldots,q, t \in \mathbb{Z}\})$$

for $n > m$. From (15) we get

$$\mathbf{x}_{nt} = \tilde{\underline{\mathbf{P}}}_n(L)\underline{\mathbf{P}}_n(L)\mathbf{x}_{nt} + \tilde{\underline{\mathbf{Q}}}_n(L)\underline{\mathbf{Q}}_n(L)\mathbf{x}_{nt} = \tilde{\underline{\mathbf{P}}}_n(L)\underline{\mathbf{\Lambda}}_n^{1/2}(L)\boldsymbol{\psi}_t^n + \tilde{\underline{\mathbf{Q}}}_n(L)\underline{\mathbf{Q}}_n(L)\mathbf{x}_{nt} \tag{16}$$

(note that integrability of the eigenvalues [see Lemma 3] implies that $\tilde{\mathbf{P}}_n\mathbf{\Lambda}_n^{1/2} \in L_2^{n\times q}(\Theta,\mathbb{C})$). Because $\mathbf{Q}_n(\theta)\mathbf{\Sigma}_n^x(\theta)\tilde{\mathbf{P}}_n(\theta) = \mathbf{\Phi}_n(\theta)\mathbf{Q}_n(\theta)\tilde{\mathbf{P}}_n(\theta) = \mathbf{0}$ for any θ, the two terms on the right-hand side of (16) are orthogonal at any lead and lag element by element, so that the first is the projection of $\mathbf{x}_{nt}$ on $\overline{\text{span}}(\{\psi_{jt}^n, j = 1,2,\dots,q, t \in \mathbb{Z}\})$ and the second is the residual vector. The required projection equation is then obtained by applying on both sides $\underline{\mathbf{C}}(L)\underline{\mathbf{\Lambda}}_m^{-1/2}(L)\underline{\mathbf{P}}_m(L)$ and noting that $\underline{\mathbf{\Lambda}}_m^{-1/2}(L)\underline{\mathbf{P}}_m(L)\mathbf{x}_{nt} = \underline{\mathbf{\Lambda}}_m^{-1/2}(L)\underline{\mathbf{P}}_m(L)\mathbf{x}_{mt} = \boldsymbol{\psi}_t^m$; that is,

$$\underline{\mathbf{C}}(L)\boldsymbol{\psi}_t^m = \underline{\mathbf{D}}(L)\boldsymbol{\psi}_t^n + \underline{\mathbf{R}}(L)\mathbf{x}_{nt},$$

where

$$\mathbf{D} = \mathbf{C}\mathbf{\Lambda}_m^{-1/2}\mathbf{P}_m\tilde{\mathbf{P}}_n\mathbf{\Lambda}_n^{1/2}, \quad \mathbf{R} = \mathbf{C}\mathbf{\Lambda}_m^{-1/2}\mathbf{P}_m\tilde{\mathbf{Q}}_n\mathbf{Q}_n. \tag{17}$$

Note that $\mathbf{R}$ belongs to $L_\infty^{q\times n}(\Theta,\mathbb{C})$ and therefore to $L_2^{q\times n}(\Theta,\mathbb{C},\mathbf{\Sigma}_n^x)$. Moreover, because $\mathbf{\Lambda}_n^{1/2} \in L_2^{q\times q}(\Theta,\mathbb{C})$ and $\mathbf{C}\mathbf{\Lambda}_m^{-1/2}\mathbf{P}_m\tilde{\mathbf{P}}_n \in L_\infty^{q\times q}(\Theta,\mathbb{C})$, then $\mathbf{D} \in L_2^{q\times q}(\Theta,\mathbb{C})$. Note too that $\mathbf{D}$, and also $\mathbf{\Delta}$, $\mathbf{H}$, and $\mathbf{F}$, which are defined subsequently, depend on $\mathbf{C}$, m, and n. However, as no confusion can arise, we do not make explicit this dependence for notational simplicity. The following result holds.

LEMMA 7. *Suppose that (i) and (ii) hold in Theorem 2. Let $n > m$, $M \subseteq \Theta$, and $\mathbf{C} \in K_M$. Consider again the projection equation*

$$\underline{\mathbf{C}}(L)\boldsymbol{\psi}_t^m = \underline{\mathbf{D}}(L)\boldsymbol{\psi}_t^n + \underline{\mathbf{R}}(L)\mathbf{x}_{nt}, \tag{18}$$

where $\mathbf{D}$ and $\mathbf{R}$ are defined as in (17), and call $\mu(\theta)$ the largest eigenvalue of the spectral density matrix of the residual $\underline{\mathbf{R}}(L)\mathbf{x}_{nt}$. Then $\mu(\theta) \le \lambda_{nq+1}^x(\theta)/\lambda_{mq}^x(\theta)$.

Proof. The matrix $\mathbf{I}_n - \tilde{\mathbf{Q}}_n\mathbf{Q}_n$ is nonnegative definite by (15) and $\lambda_{nq+1}^x\tilde{\mathbf{Q}}_n\mathbf{Q}_n - \tilde{\mathbf{Q}}_n\mathbf{\Phi}_n\mathbf{Q}_n$ is nonnegative definite by the definition of $\mathbf{\Phi}_n$, so that $\lambda_{nq+1}^x\mathbf{I}_n - \tilde{\mathbf{Q}}_n\mathbf{\Phi}_n\mathbf{Q}_n$ is also nonnegative definite. Premultiplying by $\mathbf{C}\mathbf{\Lambda}_m^{-1/2}\mathbf{P}_m$ and postmultiplying by $\tilde{\mathbf{P}}_m\mathbf{\Lambda}_m^{-1/2}\tilde{\mathbf{C}}$ it is seen that

$$\lambda_{nq+1}^x\mathbf{C}\mathbf{\Lambda}_m^{-1}\tilde{\mathbf{C}} - \mathbf{R}\mathbf{\Sigma}_n^x\tilde{\mathbf{R}}$$

is also nonnegative definite. The desired inequality follows from Fact M, the third and fourth inequalities in (4). ∎

Now let us begin the construction of our converging sequence. Note that, under assumptions (i) and (ii) in Theorem 2, there exists a set $\Pi \subseteq \Theta$ and a real W such that $\Theta - \Pi$ has null measure and, for $\theta \in \Pi$: (1) $\lambda_{nq+1}^x(\theta) \le W$ 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 Π, then it holds a.e. in Θ and vice versa.

Let M be a positive measure subset of Π such that $\lambda_{nq}^x(\theta) \geq \alpha_n$ for $\theta \in M$, where $\{\alpha_n, n \in \mathbb{N}\}$ is a real positive nondecreasing sequence satisfying $\lim_n \alpha_n = \infty$. Assume $\mathbf{C} \in K_M$ and consider (18). Taking the spectral density of both sides we get, for $\theta \in M$,

$$\mathbf{I}_q = \mathbf{D}(\theta)\widetilde{\mathbf{D}}(\theta) + \mathbf{R}(\theta)\boldsymbol{\Sigma}_n^x(\theta)\widetilde{\mathbf{R}}(\theta). \tag{19}$$

Applying Lemma 7 we obtain $\mu(\theta) \leq \lambda_{nq+1}^x(\theta)/\lambda_{mq}^x(\theta) \leq W/\alpha_m$ for $\theta \in M$. Hence by Fact M, calling $\Delta_j(\theta)$, $j = 1,2,\ldots,q$, the eigenvalues of $\mathbf{D}(\theta)\widetilde{\mathbf{D}}(\theta)$ in descending order, we have

$$1 \geq \Delta_q(\theta) \geq 1 - W/\alpha_m \tag{20}$$

for any θ in M. Thus, if m^* is such that

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

we have

$$\Delta_q(\theta) \geq 1 - W/\alpha_{m^*} > 0 \tag{21}$$

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

Now assume $m \geq m^*$. Denote by $\boldsymbol{\Delta}$ the diagonal matrix having Δ_j in place (j,j) and by $\mathbf{H}(\theta)$ a matrix that is measurable in M and fulfills for any $\theta \in M$ the following conditions: (a) $\mathbf{H}(\theta)\widetilde{\mathbf{H}}(\theta) = \mathbf{I}_q$, (b) $\mathbf{H}(\theta)\boldsymbol{\Delta}(\theta)\widetilde{\mathbf{H}}(\theta) = \mathbf{D}(\theta)\widetilde{\mathbf{D}}(\theta)$. Inequality (21) ensures that $1/\sqrt{\Delta_j(\theta)}$ is bounded in M for $j = 1,2,\ldots,q$, so that the definition

$$\mathbf{F}(\theta) = \begin{cases} \mathbf{H}(\theta)\boldsymbol{\Delta}(\theta)^{-1/2}\widetilde{\mathbf{H}}(\theta)\mathbf{D}(\theta) & \text{if } \theta \in M \\ \mathbf{0}_q & \text{if } \theta \notin M \end{cases} \tag{22}$$

makes sense. Note that $\mathbf{F}$ belongs to K_M.

LEMMA 8. *Suppose that (i) and (ii) in Theorem 2 hold. Let M be a positive measure subset of Π and $\{\alpha_n, n \in \mathbb{N}\}$ a real positive nondecreasing sequence such that $\lim_n \alpha_n = \infty$. Assume that*

(a) $\mathbf{C} \in K_M$;
(b) $\lambda_{nq}^x(\theta) \geq \alpha_n$ for $\theta \in M$;

Then, given τ, such that $2 > \tau > 0$, there exists an integer m_τ such that, first, $W/\alpha_{m_\tau} < 1$ and, second, for $n > m \geq m_\tau$, the largest eigenvalue of the spectral density matrix of

$$\underline{\mathbf{C}}(L)\boldsymbol{\psi}_t^m - \underline{\mathbf{F}}(L)\boldsymbol{\psi}_t^n$$

is less than τ for any $\theta \in \Pi$, where $\mathbf{F}$ is defined as in (22), with $\mathbf{D}$ defined as in (17).

Proof. From (18) we get

$$\underline{\mathbf{C}}(L)\boldsymbol{\psi}_t^m - \underline{\mathbf{F}}(L)\boldsymbol{\psi}_t^n = \underline{\mathbf{R}}(L)\mathbf{x}_{nt} + (\underline{\mathbf{D}}(L) - \underline{\mathbf{F}}(L))\boldsymbol{\psi}_t^n.$$

The terms on the right-hand side 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 $\mathbf{S}$ the spectral density matrix on the left-hand side and using (19), we see that, for $\theta \in M$,

$$\mathbf{S} = 2\mathbf{I}_q - \mathbf{D}\tilde{\mathbf{F}} - \mathbf{F}\tilde{\mathbf{D}} = 2\mathbf{I}_q - 2\mathbf{H}\boldsymbol{\Delta}^{1/2}\tilde{\mathbf{H}} = 2\mathbf{H}(\mathbf{I}_q - \boldsymbol{\Delta}^{1/2})\tilde{\mathbf{H}},$$

whose largest eigenvalue is $2 - 2\sqrt{\Delta_q(\theta)}$, which is less than or equal to $2[1 - \Delta_q(\theta)] \leq 2W/\alpha_m$ by (20). Thus, for $\mathbf{F}$ to make sense and the statement of the lemma to hold we need $2W/\alpha_{m_\tau} < \min(2,\tau)$. Because $\tau < 2$, m_τ must fulfill

$$2W/\alpha_{m_\tau} < \tau. \tag{23}$$

■

The following lemma will be repeatedly employed. Its proof is a consequence of the following statement.

FACT L. *Suppose that $\{f_n,\ n \in \mathbb{N}\}$ is a sequence of functions belonging to $L_k(\Theta,\mathbb{C})$, with k equal to 1 or 2, which is convergent in the norm of $L_k(\Theta,\mathbb{C})$. Then there exists an increasing sequence s_i such that $\lim_i f_{s_i}(\theta) = f(\theta)$ a.e. in Θ (see Apostol, 1974, p. 298).*

LEMMA 9. *Suppose that $A = \{A_{nt},\ t \in \mathbb{Z}\}$ and $B = \{B_{nt},\ t \in \mathbb{Z}\}$ belong to $\mathbf{X}$ and are costationary with the x's, and that $\lim_n A_{nt} = A_t$ and $\lim_n B_{nt} = B_t$. Then, for a sequence of integers s_i,*

$$\lim_i \mathcal{S}(A_{s_i t}, B_{s_i t}; \theta) = \mathcal{S}(A_t, B_t; \theta)$$

($\mathcal{S}$ has been defined in the proof of Lemma 2) a.e. in Θ.

Proof. $\langle A_{nt}, B_{nt}\rangle = (1/2\pi)\int_{-\pi}^{\pi} \mathcal{S}(A_{nt}, B_{nt};\theta)d\theta$ (see the proof of Lemma 2). Continuity of the inner product implies that $(1/2\pi)\int_{-\pi}^{\pi}|\mathcal{S}(A_{nt}, B_{nt};\theta) - \mathcal{S}(A_t, B_t;\theta)|d\theta \to 0$, i.e., that $\mathcal{S}(A_{nt}, B_{nt};\theta)$ converges to $\mathcal{S}(A_t, B_t;\theta)$ in $L_1(\Theta,\mathbb{C})$. The result follows from Fact L. ■

LEMMA 10. *Suppose that (i) and (ii) in Theorem 2 hold and let M and $\{\alpha_n,\ n \in \mathbb{N}\}$ be as in Lemma 8. There exists a q-dimensional vector process $\mathbf{v}$ such that*

(a) v_{jt} is an aggregate for $j = 1,2,\ldots,q$;
(b) the spectral density matrix of $\mathbf{v}$ equals $\mathbf{I}_q$ for θ a.e. in M, $\mathbf{0}_q$ for $\theta \notin M$.

Proof. Let $\mathbf{F}_1$ be any element of K_M. Set $\tau = 1/2^2$ and $s_1 = m_\tau$, where m_τ satisfies (23). Then set $\mathbf{G}_1 = \mathbf{F}_1 \boldsymbol{\Lambda}_{s_1}^{-1/2} \mathbf{P}_{s_1}$ and $\mathbf{v}_t^1 = \underline{\mathbf{G}}_1(L)\mathbf{x}_{nt}$. It is easily seen that the spectral density matrix of $\mathbf{v}_t^1$ equals $\mathbf{I}_q$ for $\theta \in M$, $\mathbf{0}_q$ for $\theta \notin M$.

Now set $\tau = 1/2^4$ and $s_2 = m_\tau$, where m_τ satisfies (23) and $m_\tau \geq s_1$. Then determine $\mathbf{D}$ as in (17), with $\mathbf{F}_1$ in place of $\mathbf{C}$, s_2 in place of n, and s_1 in place of m, and determine $\mathbf{F}_2$ as in (22). Finally set $\mathbf{G}_2 = \mathbf{F}_2 \boldsymbol{\Lambda}_{s_2}^{-1/2} \mathbf{P}_{s_2}$ and $\mathbf{v}_t^2 = \underline{\mathbf{G}}_2(L)\mathbf{x}_{nt}$. The spectral density matrix of $\mathbf{v}_t^2$ equals $\mathbf{I}_q$ for $\theta \in M$, $\mathbf{0}_q$ for $\theta \notin M$. Moreover, by the definition of s_1 and Lemma 8, calling $\mathcal{A}_1$ the largest eigenvalue of the spectral density matrix of $\mathbf{v}_t^1 - \mathbf{v}_t^2$, we have $\mathcal{A}_1(\theta) < 1/2^2$ for any $\theta \in \Pi$, so that $\|v_{jt}^1 - v_{jt}^2\| < \frac{1}{2}$, for $j = 1,2,\ldots,q$.

By recursion, set $\tau = 1/2^{2k}$ and $s_k = m_\tau$, where m_τ satisfies (23) and $m_\tau \geq s_{k-1}$. Then determine $\mathbf{D}$ as in (17), with $\mathbf{F}_{k-1}$ in place of $\mathbf{C}$, s_k in place of n and s_{k-1} in place of m, and determine $\mathbf{F}_k$ as in (22). Finally set $\mathbf{G}_k = \mathbf{F}_k \boldsymbol{\Lambda}_{s_k}^{-1/2} \mathbf{P}_{s_k}$ and $\mathbf{v}_t^k = \underline{\mathbf{G}}_k(L)\mathbf{x}_{nt}$. The spectral density matrix of $\mathbf{v}_t^k$ equals $\mathbf{I}_q$ for $\theta \in M$, $\mathbf{0}_q$ for $\theta \notin M$. Moreover, by the definition of s_{k-1} and Lemma 8, calling $\mathcal{A}_{k-1}$ the largest eigenvalue of the spectral density matrix of $\mathbf{v}_t^{k-1} - \mathbf{v}_t^k$, we have $\mathcal{A}_{k-1}(\theta) < 1/2^{2(k-1)}$ for any $\theta \in \Pi$, so that $\|v_{jt}^{k-1} - v_{jt}^k\| < 1/2^{k-1}$ for $j = 1,2,\ldots,q$.

Because we have

$$\|v_{jt}^k - v_{jt}^{k+h}\| \leq \|v_{jt}^k - v_{jt}^{k+1}\| + \cdots + \|v_{jt}^{k+h-1} - v_{jt}^{k+h}\| < 1/2^{k-1},$$

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

$$\mathbf{G}_n(\theta)\tilde{\mathbf{G}}_n(\theta) = \mathbf{F}_n(\theta)\boldsymbol{\Lambda}_{s_n}^{-1}(\theta)\tilde{\mathbf{F}}_n(\theta),$$

whose diagonal entries $|\mathbf{g}_{nj}(\theta)|^2$ cannot be larger than $1/\lambda_{s_n q}^x(\theta)$ because $\mathbf{F}_n \in K_M$. The latter ratio converges to zero a.e. in Θ and is less than 1 by Assumption 2, so that its integral on Θ converges to zero by the Lebesgue convergence theorem (Royden, 1988, p. 91).

Finally, (b) follows from Lemma 9 and the fact that the spectral density matrix of $\mathbf{v}_t^k$ equals $\mathbf{I}_q$ for $\theta \in M$, $\mathbf{0}_q$ for $\theta \notin M$. ∎

LEMMA 11. *Suppose that (i) and (ii) in Theorem 2 hold. There exists a q-dimensional orthonormal white-noise vector process $\mathbf{z}$ such that z_{jt} is an aggregate for $j = 1,2,\ldots,q$.*

Proof. Define $M_0 = \Pi$. Then, by recursion, define ν_a, $a \in \mathbb{N}$, as the smallest among the integers m such that

$$\mathcal{L}(\{\theta \in M_{a-1},\ \lambda_{mq}^x(\theta) > a\}) > \pi$$

and

$$M_a = \{\theta \in M_{a-1}, \lambda^x_{\nu_a q}(\theta) > a\}.$$

The measure of the set

$$N_1 = M_1 \cap M_2 \cap \cdots \cap M_a \cap \cdots$$

is not less than π. Now define N_2 starting with $\Pi - N_1$ instead of Π, and using $\mathcal{L}(\Pi - N_1)/2$ instead of π, N_b, $b > 2$, starting with $\Pi - N_1 - N_2 - \cdots - N_{b-1}$ and using $\mathcal{L}(\Pi - N_1 - N_2 - \cdots - N_{b-1})/2$, etc. Setting $N = N_1 \cup N_2 \cup \cdots$, we have

$$\mathcal{L}(N) = \mathcal{L}(N_1) + \mathcal{L}(N_2) + \cdots + \mathcal{L}(N_b) + \cdots = 2\pi.$$

Lemma 10 can be applied to the subset N_b, with the sequence α_n defined as $\alpha_n = a$, where a is the only integer such that $\nu_a \leq n < \nu_{a+1}$. We obtain a q-dimensional vector $\mathbf{v}_t^b = (v_{1t}^b \quad v_{2t}^b \quad \dots \quad v_{qt}^b)'$ such that (i) v_{jt}^b is an aggregate for $j = 1,2,\dots,q$; (ii) its spectral density matrix is $\mathbf{I}_q$ a.e. in N_b, $\mathbf{0}_q$ for $\theta \notin N_b$. Now set $\mathbf{z}_t = \sum_{b=1}^{\infty} \mathbf{v}_t^b$. It is easily seen that the spectral density matrix of $\mathbf{z}_t$ is $\mathbf{I}_q$ a.e. in Θ, so that $\mathbf{z}$ is a q-dimensional orthonormal white noise process. ∎

4.5. The Space Spanned by z

We now prove that the space spanned by $\mathbf{z}$ is $\mathcal{G}(\mathbf{x})$. Let y_t be an aggregate and consider the projection

$$y_t = \mathrm{proj}(y_t | \overline{\mathrm{span}}(\mathbf{z})) + r_t.$$

We want to show that r_t is necessarily zero. Consider the $(q+1)$-dimensional vector process $(\mathbf{z}_t \quad r_t)$. Its spectral density, call it $\mathcal{W}$, is diagonal with $\mathbf{I}_q$ in the $q \times q$ upper-left submatrix, so that

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

Because z_{jt} and r_t belong to $\mathcal{G}(\mathbf{x})$, there exist DAS's $\{\mathbf{a}_{nj}, n \in \mathbb{N}\}$, for $j = 1,2,\dots,q+1$ such that

$$\lim_n \underline{\mathbf{a}}_{nj}(L)\mathbf{x}_{s_n t} = z_{jt}, \quad \text{for } j = 1,2,\dots,q,$$

$$\lim_n \underline{\mathbf{a}}_{nq+1}(L)\mathbf{x}_{s_n t} = r_t.$$

Moreover: (1) $\int_{-\pi}^{\pi} |\mathbf{a}_{nj}(\theta)|^2 d\theta$ converges to zero for $j = 1,2,\dots,q+1$, so that a subsequence of $\mathbf{a}_{nj}$ converges to zero a.e. in Θ (Fact L); (2) calling $\mathcal{Z}_n$ the spectral density matrix of the vector process

$$(\underline{\mathbf{a}}_{n1}(L)\mathbf{x}_{s_n t} \quad \underline{\mathbf{a}}_{n2}(L)\mathbf{x}_{s_n t} \quad \dots \quad \underline{\mathbf{a}}_{nq+1}(L)\mathbf{x}_{s_n t}),$$

a subsequence of $\mathcal{Z}_n$ converges to $\mathcal{W}$ a.e. in Θ (Lemma 9). Thus, with no loss of generality we can assume that $\mathbf{a}_{nj}$ converges to zero and $\mathcal{Z}_n$ converges to $\mathcal{W}$ a.e. in Θ.

Now, for $j = 1,2,\dots,q+1$, set $\mathbf{f}_{nj} = \mathbf{a}_{nj}\tilde{\mathbf{P}}_{s_n}$ and $\mathbf{g}_{nj} = \mathbf{a}_{nj} - \mathbf{f}_{nj}\mathbf{P}_{s_n}$, so that

$$\mathbf{a}_{nj} = \mathbf{f}_{nj}\mathbf{P}_{s_n} + \mathbf{g}_{nj}$$

and

$$|\mathbf{a}_{nj}(\theta)|^2 = |\mathbf{f}_{nj}(\theta)|^2 + |\mathbf{g}_{nj}(\theta)|^2.$$

Because $\mathbf{a}_{nj}$ converges to zero a.e. in Θ, then $\mathbf{g}_{nj}$ converges to zero a.e. in Θ. Moreover, the definition of $\mathbf{g}_{nj}$ and $\mathbf{f}_{nj}$ implies that

$$\underline{\mathbf{a}}_{nj}(L)\mathbf{x}_{s_n t} = \underline{\mathbf{f}}_{nj}(L)\underline{\mathbf{P}}_{s_n}(L)\mathbf{x}_{s_n t} + \underline{\mathbf{g}}_{nj}(L)\mathbf{x}_{s_n t}$$

is the orthogonal projection of the left-hand side on the space spanned by $\underline{\mathbf{p}}^x_{s_n k}(L)\mathbf{x}_{s_n t}$, for $k = 1,2,\dots,q$ and $t \in \mathbb{Z}$. As a consequence, the spectral density matrix $\mathcal{Z}_n$ is equal to the spectral density matrix of

$$(\underline{\mathbf{f}}_{n1}(L)\underline{\mathbf{P}}_{s_n}(L)\mathbf{x}_{s_n t} \quad \underline{\mathbf{f}}_{n2}(L)\underline{\mathbf{P}}_{s_n}(L)\mathbf{x}_{s_n t} \quad \cdots \quad \underline{\mathbf{f}}_{nq+1}(L)\underline{\mathbf{P}}_{s_n}(L)\mathbf{x}_{s_n t}),$$

call it $\mathcal{Z}^1_n$, plus the spectral density matrix of

$$(\underline{\mathbf{g}}_{n1}(L)\mathbf{x}_{s_n t} \quad \underline{\mathbf{g}}_{n2}(L)\mathbf{x}_{s_n t} \quad \cdots \quad \underline{\mathbf{g}}_{nq+1}(L)\mathbf{x}_{s_n t}),$$

call it $\mathcal{Z}^2_n$: $\mathcal{Z}_n = \mathcal{Z}^1_n + \mathcal{Z}^2_n$.

Now observe first that $\mathcal{Z}^1_n$ is singular for any θ. Second, because $\mathbf{g}_{nj}(\theta)$ is orthogonal to $\mathbf{p}^x_{s_n k}(\theta)$, for $k = 1,2,\dots,q$, then

$$\mathbf{g}_{nj}(\theta)\boldsymbol{\Sigma}^x_{s_n}(\theta)\tilde{\mathbf{g}}_{nj}(\theta) \le \lambda^x_{s_n q+1}|\mathbf{g}_{nj}(\theta)|^2$$

(Lancaster and Tismenetsky, 1985, Exercise 1, p. 287). Essential boundedness of λ^x_{q+1} along with convergence to zero a.e. of $\mathbf{g}_{nj}$ implies that $\mathcal{Z}^2_n$ converges to zero a.e. in Θ. This implies that $\det \mathcal{Z}_n$ converges to zero a.e. in Θ and therefore that $\det \mathcal{W}(\theta) = \mathcal{S}(r_t, r_t;\theta) = 0$ a.e. in Θ, so that $r_t = 0$.

4.6. Proof of Theorem 2

So far we have proved that if (i) and (ii) in Theorem 2 hold then the canonical decomposition is

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

$$\gamma_{it} = \operatorname{proj}(x_{it}|\mathcal{G}(\mathbf{x})) = \underline{\mathbf{c}}_i(L)\mathbf{z}_t,$$

where $\mathbf{z}$ is a q-dimensional orthonormal white noise and $\mathbf{c}_i \in L^q_2(\Theta, \mathbb{C})$. Suppose that $\boldsymbol{\delta}$ is idiosyncratic. By Fact M(a), $\lambda^\gamma_{nq}(\theta) \ge \lambda^x_{nq}(\theta) - \lambda^\delta_{n1}(\theta)$, so that $\lambda^\gamma_q(\theta) = \infty$ a.e. in Θ. Thus, to complete the proof of Theorem 2 we must only show that $\boldsymbol{\delta}$ is idiosyncratic.

We need some additional preliminary results. Suppose that $\mathbf{v} = \{\mathbf{v}_t, t \in \mathbb{Z}\}$ and $\mathbf{w} = \{\mathbf{w}_t, t \in \mathbb{Z}\}$ are orthonormal q-dimensional white-noise vectors belonging to $\mathbf{X}$. Moreover, suppose that $\mathbf{v}$ and $\mathbf{w}$ are costationary with the x's and therefore with one another. Let $\mathbf{A}$ be the matrix whose (h,k) entry is the cross spectrum $\mathcal{S}(v_{ht}, w_{kt}; \theta)$. Note that all the entries of $\mathbf{A}$ have modulus bounded by 1 for θ a.e. in Θ. The orthogonal projection, element by element, of $\mathbf{v}_t$ on the process $\mathbf{w}$ is $\underline{\mathbf{A}}(L)\mathbf{w}_t$, whereas $\tilde{\underline{\mathbf{A}}}(L)\mathbf{v}_t$ is the orthogonal projection of $\mathbf{w}_t$ on the process $\mathbf{v}$.

DEFINITION 12. *Let $\mathbf{v}_n = \{\mathbf{v}_{nt}, t \in \mathbb{Z}\}$, $n \in \mathbb{N}$, be a q-dimensional orthonormal white-noise process belonging to $\mathbf{X}$ and costationary with the x's, so that $\mathbf{v}_n$ and $\mathbf{v}_m$ are costationary for any n and m. Consider the orthogonal projection*

$$\mathbf{v}_{mt} = \underline{\mathbf{A}}^{mn}(L)\mathbf{v}_{nt} + \boldsymbol{\rho}_t^{mn} \tag{24}$$

and let $\mathcal{D}^{mn}$ be the spectral density of $\boldsymbol{\rho}_t^{mn}$. The sequence $\{\mathbf{v}_n, n \in \mathbb{N}\}$ generates a Cauchy sequence of spaces *if, given $\epsilon > 0$, for θ a.e. in Θ there exists an integer $m_\epsilon(\theta)$ such that for $n, m > m_\epsilon(\theta)$, trace$(\mathcal{D}^{mn}(\theta)) < \epsilon$.*

Remark 11. Note that, if $\mathbf{v}_{nt}$ converges, it generates a Cauchy sequence of spaces, because, denoting by $\mathcal{E}^{mn}(\theta)$ the spectral density matrix of $\mathbf{v}_{mt} - \mathbf{v}_{nt}$, we have $\text{trace}(\mathcal{D}^{mn}(\theta)) \le \text{trace}(\mathcal{E}^{mn}(\theta))$. By contrast, the converse does not necessarily hold. As we show subsequently, the normalized principal components $\boldsymbol{\psi}_t^n$ generate a Cauchy sequence of spaces. However, they do not converge in general: for example, take $q = 1$ and assume that $\boldsymbol{\psi}_t^n$ is a normalized principal component converging to $\boldsymbol{\psi}_t$; then $(-1)^n\boldsymbol{\psi}_t^n$ is also a normalized principal component that does not converge.

LEMMA 12. *Assume that*

(1) the sequence $\{\mathbf{v}_n, n \in \mathbb{N}\}$ fulfills Definition 12;
(2) $y = \{y_t, t \in \mathbb{Z}\}$ belongs to $\mathbf{X}$ and is costationary with the x's.

Let Y_{nt} be the orthogonal projection of y_t on the process $\mathbf{v}_n$, i.e., $Y_{nt} = \text{proj}(y_t|\overline{\text{span}}(\mathbf{v}_n))$. Then Y_{nt} converges in $\mathbf{X}$.

PROOF. We have

$$y_t = Y_{nt} + r_{nt} = \underline{\mathbf{b}}_n(L)\mathbf{v}_{nt} + r_{nt},$$

$$y_t = Y_{mt} + r_{mt} = \underline{\mathbf{b}}_m(L)\mathbf{v}_{mt} + r_{mt},$$

where $\mathbf{b}_s \in L_2^s(\Theta, \mathbb{C})$. Hence

$$\underline{\mathbf{b}}_n(L)\mathbf{v}_{nt} - \underline{\mathbf{b}}_m(L)\mathbf{v}_{mt} = r_{mt} - r_{nt}.$$

The spectral density of the left-hand side is the cross spectrum between the left- and the right-hand sides. The latter, as a result of the definition of r_{mt} and

r_{nt}, is the sum of the cross spectrum between r_{nt} and $\underline{\mathbf{b}}_m(L)\mathbf{v}_{mt}$, call it $\mathcal{S}_1$, and the cross spectrum between r_{mt} and $\underline{\mathbf{b}}_n(L)\mathbf{v}_{nt}$, call it $\mathcal{S}_2$. Using (24), $\mathcal{S}_1$ is the cross spectrum between r_{nt} and $\underline{\mathbf{b}}_m(L)\underline{\mathbf{A}}^{mn}(L)\mathbf{v}_{nt} + \underline{\mathbf{b}}_m(L)\boldsymbol{\rho}_t^{mn}$, which reduces to the cross spectrum between r_{nt} and $\underline{\mathbf{b}}_m(L)\boldsymbol{\rho}_t^{mn}$, call it $\mathcal{C}_{mn}$. Now observe that both the spectral density of r_{nt} and the squared entries of $\mathbf{b}_m$ are bounded in modulus by the spectral density of y_t. Thus, because $\{\mathbf{v}_n,\ n \in \mathbb{N}\}$ generates a Cauchy sequence of spaces, $\mathcal{C}_{mn}$ converges to zero a.e. in Θ as $m, n \to \infty$. The same argument holds for $\mathcal{S}_2$, so that the spectral density of $Y_{nt} - Y_{mt}$ converges to zero a.e. in Θ as $m, n \to \infty$. Because both the spectral density of Y_{nt} and the spectral density of Y_{mt} are dominated by the spectral density of y_t, by the Lebesgue convergence theorem (Royden, 1988, p. 91), the integral of the spectral density of $Y_{nt} - Y_{mt}$ also converges to zero as $m, n \to \infty$, so that Y_{nt} is a Cauchy sequence. ∎

LEMMA 13. *The sequence* $\{\boldsymbol{\psi}^n,\ n \in \mathbb{N}\}$ *generates a Cauchy sequence of spaces.*

Proof. For $n > m$ consider (18) for $\mathbf{C} = \mathbf{I}_q$:

$$\boldsymbol{\psi}_t^m = \underline{\mathbf{D}}(L)\boldsymbol{\psi}_t^n + \boldsymbol{\rho}_t^{mn}. \tag{25}$$

Calling $\mathcal{D}^{mn}$ the spectral density of $\boldsymbol{\rho}_t^{mn}$, convergence to zero of $\text{trace}(\mathcal{D}^{mn}(\theta))$ for θ a.e. in Θ and $n > m$ is a consequence of Lemma 7. On the other hand,

$$\boldsymbol{\psi}_t^n = \underline{\tilde{\mathbf{D}}}(L)\boldsymbol{\psi}_t^m + \boldsymbol{\rho}_t^{nm}. \tag{26}$$

From (25) and (26) we get

$$\mathbf{I}_q = \mathbf{D}(\theta)\tilde{\mathbf{D}}(\theta) + \mathcal{D}^{mn}(\theta) = \tilde{\mathbf{D}}(\theta)\mathbf{D}(\theta) + \mathcal{D}^{nm}(\theta)$$

a.e. in Θ. By taking the trace on both sides and noting that the trace of $\mathbf{D}(\theta)\tilde{\mathbf{D}}(\theta)$ is equal to the trace of $\tilde{\mathbf{D}}(\theta)\mathbf{D}(\theta)$ we get $\text{trace}(\mathcal{D}^{mn}(\theta)) = \text{trace}(\mathcal{D}^{nm}(\theta))$ a.e. in Θ. Finally $\mathcal{D}^{mm}(\theta) = 0$. Thus $\text{trace}(\mathcal{D}^{mn}(\theta))$ converges to zero a.e. in Θ for any diverging n and m. ∎

Now let us go back to equation (16) and concentrate on a single line, i.e., the orthogonal decomposition obtained by projecting x_{it} on the normalized principal components ψ_{jt}^n, $j = 1, 2, \ldots, q$. Calling $\underline{\boldsymbol{\pi}}_{ni}(L)$ the ith (q-dimensional) row of $\underline{\tilde{\mathbf{P}}}_n(L)$ and $\underline{\mathbf{q}}_{ni}(L)$ the ith row of $\underline{\tilde{\mathbf{Q}}}_n(L)$, we get

$$x_{it} = \underline{\boldsymbol{\pi}}_{ni}(L)\underline{\mathbf{\Lambda}}_n^{1/2}(L)\boldsymbol{\psi}_t^n + \underline{\mathbf{q}}_{ni}(L)\underline{\mathbf{Q}}_n(L)\mathbf{x}_{nt}.$$

The following theorem, besides being useful to show that $\boldsymbol{\delta}$ is idiosyncratic, is important per se, because of its implications for the estimation of common and idiosyncratic components (see Forni et al., 1999).

THEOREM 5. *The sequence of projections* $\gamma_{it}^n = \underline{\boldsymbol{\pi}}_{ni}(L)\underline{\boldsymbol{\Lambda}}_n^{1/2}(L)\boldsymbol{\psi}_t^n = \underline{\boldsymbol{\pi}}_{ni}(L)\underline{\mathbf{P}}_n(L)\mathbf{x}_{nt}$, $n \in \mathbb{N}$, *converges in mean square to* $\gamma_{it} = \text{proj}(x_{it}|\mathcal{G}(\mathbf{x}))$, *for any* i.

Proof. By Lemmas 12 and 13 γ_{it}^n converges in mean square to an element γ_{it}^* in $\mathbf{X}$. Therefore the sequence of the residuals $\delta_{it}^n = x_{it} - \gamma_{it}^n$ also converges to an element δ_{it}^* in $\mathbf{X}$. Moreover, γ_{it}^* is an aggregate, because $\boldsymbol{\pi}_{ni}\mathbf{P}_n$ is a DAS. To see this, consider that the spectral density of γ_{it}^n, i.e., $\boldsymbol{\pi}_{ni}\boldsymbol{\Lambda}_n\tilde{\boldsymbol{\pi}}_{ni}$, is not smaller than $\boldsymbol{\pi}_{ni}\tilde{\boldsymbol{\pi}}_{ni}\lambda_{nq}^x$ and is bounded above by the spectral density of x_{it}, call it σ_i, implying $\boldsymbol{\pi}_{ni}(\theta)\tilde{\boldsymbol{\pi}}_{ni}(\theta) \le \sigma_i(\theta)/\lambda_{nq}^x(\theta)$. The latter ratio converges to zero a.e. in Θ and is bounded above by $\sigma_i(\theta)$ by Assumption 2, so that the Lebesgue convergence theorem (Royden, 1988, p. 91) applies.

Last, by construction, δ_{it}^n is orthogonal to $\boldsymbol{\psi}_{t-k}^n$ for any $k \in \mathbb{Z}$. Because $\mathcal{G}(\mathbf{x}) = \overline{\text{span}}(\mathbf{z})$, and because the process $\mathbf{z}$ has been obtained by taking limits of linear combinations of the ψ's (Lemmas 7, 8, 10, 11), continuity of the inner product implies that $\delta_{it}^* \perp \mathcal{G}(\mathbf{x})$. The conclusion follows from uniqueness of the orthogonal decomposition. ■

The following lemma concludes the proof of Theorem 2.

LEMMA 14. $\boldsymbol{\delta}$ *is idiosyncratic.*

Proof. Let us fix m and denote by $\boldsymbol{\Sigma}_m^\delta$ the spectral density matrix of the vector process $\boldsymbol{\delta}_{mt} = (\delta_{1t} \quad \delta_{2t} \quad \dots \quad \delta_{mt})'$. We want to show that the largest eigenvalue of such matrix, i.e., $\lambda_{m1}^\delta(\theta)$, cannot be larger than $\sup_n \lambda_{nq+1}^x(\theta) = \lambda_{q+1}^x(\theta)$ for any $\theta \in \Theta$. Let $\boldsymbol{\Sigma}_m^{\delta^n}$, $n > m$, be the spectral density matrix of $\boldsymbol{\delta}_{mt}^n = (\delta_{1t}^n \quad \delta_{2t}^n \quad \dots \quad \delta_{mt}^n)'$ and $\lambda_{m1}^{\delta^n}$ be its largest eigenvalue. By Theorem 5 δ_{it}^n converges to δ_{it} in mean square for $i = 1,2,\dots,m$, so that, by Lemma 9, a subsequence of $\boldsymbol{\Sigma}_m^{\delta^n}$ converges to $\boldsymbol{\Sigma}_m^\delta$ a.e. in Θ. Assuming that $\lim_n \boldsymbol{\Sigma}_m^{\delta^n} = \boldsymbol{\Sigma}_m^\delta$ a.e. in Θ avoids further complication in notation and does not imply any loss of generality. Continuity of the eigenvalues as functions of the matrix entries (Ahlfors, 1987, pp. 300–306) implies that

$$\lim_n \lambda_{m1}^{\delta^n}(\theta) = \lambda_{m1}^\delta(\theta), \tag{27}$$

a.e. in Θ. Moreover, note that $\boldsymbol{\Sigma}_m^{\delta^n}$ is the $m \times m$ upper-left submatrix of $\boldsymbol{\Sigma}_n^{\delta^n}$, so that, by Fact M(b),

$$\lambda_{m1}^{\delta^n}(\theta) \le \lambda_{n1}^{\delta^n}(\theta) = \lambda_{nq+1}^x(\theta)$$

for any $n \ge m$ and any θ in Θ. Hence by (27) $\lambda_{m1}^\delta(\theta) \le \lambda_{q+1}^x(\theta)$. Because this is true for any m,

$$\lambda_1^\delta(\theta) \le \lambda_{q+1}^x(\theta), \tag{28}$$

so that λ_1^δ is essentially bounded. The statement follows from Theorem 1. ■

4.7. Proof of Theorem 3

Now we prove Theorem 3. Assume that $\mathbf{x}$ fulfills Definition 10, so that

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

$$\chi_{it} = \underline{\mathbf{b}}_i(L)\mathbf{u}_t,$$

where $\mathbf{u}$ is q-dimensional. As we have proved in Section 4.5, $\mathbf{x}$ has also the canonical representation

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

$$\gamma_{it} = \text{proj}(x_{it}|\mathcal{G}(\mathbf{x})) = \underline{\mathbf{c}}_i(L)\mathbf{z}_t,$$

where $\mathbf{z}$ is q-dimensional and $\overline{\text{span}}(\mathbf{z}) = \mathcal{G}(\mathbf{x})$. Because $\boldsymbol{\xi}$ is idiosyncratic then $\mathcal{G}(\mathbf{x}) \subseteq \overline{\text{span}}(\boldsymbol{\chi})$, and obviously $\overline{\text{span}}(\boldsymbol{\chi}) \subseteq \overline{\text{span}}(\mathbf{u})$, so that $\overline{\text{span}}(\mathbf{z}) \subseteq \overline{\text{span}}(\mathbf{u})$. Because both $\mathbf{u}$ and $\mathbf{z}$ are q-dimensional white-noise processes, then $\overline{\text{span}}(\mathbf{z}) = \overline{\text{span}}(\mathbf{u})$, so that

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

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

Remark 12. Because we have proved that $\delta_{it} = \xi_{it}$, (13) and (28) imply that

$$\lambda^x_{q+1}(\theta) = \lambda^\xi_1(\theta)$$

a.e. in Θ.

5. NONSTATIONARY VARIABLES

The case of trend stationary or difference stationary variables can be easily accommodated in our model. Assuming that the nature of nonstationarity 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) in Theorem 2 hold for λ^y_{q+1} and λ^y_q, respectively. Then we have the representation

$$(1 - L)x_{it} = \chi_{it} + \xi_{it},$$

$$\chi_{it} = \underline{\mathbf{b}}_i(L)\mathbf{u}_t,$$

where $\mathbf{u}_t$ is q-dimensional and $\boldsymbol{\xi}$ is idiosyncratic. Now observe that the vectors $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$ are unique, and so are the spectral density matrices $\boldsymbol{\Sigma}^\chi_n$ and $\boldsymbol{\Sigma}^\xi_n$. Therefore all the information necessary to determine whether the χ's, or the ξ's, are $I(1)$ or $I(0)$, and whether cointegration relationships hold among the

χ's or the ξ's, can be recovered starting with the spectral density matrices of the x's.

REFERENCES

Ahlfors, L.V. (1987) *Complex Analysis.* Sidney: McGraw-Hill.

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

Brillinger, D.R. (1981) *Time Series Data Analysis and Theory.* San Francisco: Holden Day.

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag.

Chamberlain, G. (1983) Funds, factors, and diversification in arbitrage pricing models. *Econometrica* 51, 1281–1304.

Chamberlain, G. & M. Rothschild (1983) Arbitrage, factor structure, and mean-variance analysis in large asset markets. *Econometrica* 51, 1305–1324.

Conway, J.B. (1985) *A Course in Functional Analysis.* New York: Springer-Verlag.

Forni, M., M. Hallin, M. Lippi, & L. Reichlin (2000) The Generalized Dynamic Factor Model: Identification and Estimation. *Review of Economics and Statistics* 82, 540–554.

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

Forni, M. & L. Reichlin (1996) Dynamic common factors in large cross-sections. *Empirical Economics* 21, 27–42.

Forni, M. & L. Reichlin (1998) Let's get real: A factor analytic approach to disaggregated business cycle dynamics. *Review of Economic Studies* 65, 453–473.

Geweke, J. (1977) The dynamic factor analysis of economic time series. In D.J. Aigner & A.S. Goldberger (eds.), *Latent Variables in Socio-Economic Models*, pp. 365–383. Amsterdam: North-Holland.

Lancaster, P. & M. Tismenetsky (1985) *The Theory of Matrices*, 2nd ed. Orlando: Academic Press.

Quah, D. & T.J. Sargent (1993) A dynamic index model for large cross sections. In J.H. Stock & M.W. Watson (eds.), *Business Cycles, Indicators, and Forecasting*, pp. 285–309. Chicago: University of Chicago Press.

Royden, H.L. (1988) *Real Analysis*, 3rd ed. New York: Macmillan.

Rozanov, Yu.A. (1967) *Stationary Random Processes.* San Francisco: Holden-Day.

Sargent, T.J. & C.A. Sims (1977) Business cycle modelling without pretending to have too much *a priori* economic theory. In C. A. Sims (ed.), *New Methods in Business Cycle Research*, pp. 45–109. Minneapolis: Federal Reserve Bank of Minneapolis.

Stock, J.H. & M.H. Watson (1999) Diffusion Indexes. Manuscript.