The Generalized Dynamic-Factor Model: Identification and Estimation

Author(s): Mario Forni, Marc Hallin, Marco Lippi and Lucrezia Reichlin

Source: *The Review of Economics and Statistics*, Nov., 2000, Vol. 82, No. 4 (Nov., 2000), pp. 540-554

Published by: The MIT Press

Stable URL: https://www.jstor.org/stable/2646650

THE GENERALIZED DYNAMIC-FACTOR MODEL: IDENTIFICATION AND ESTIMATION

Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin*

Abstract—This paper proposes a factor model with infinite dynamics and nonorthogonal idiosyncratic components. The model, which we call the *generalized dynamic-factor model,* is novel to the literature and generalizes the static approximate factor model of Chamberlain and Rothschild (1983), as well as the exact factor model à la Sargent and Sims (1977). We provide identification conditions, propose an estimator of the common components, prove convergence as both time and cross-sectional size go to infinity at appropriate rates, and present simulation results. We use our model to construct a coincident index for the European Union. Such index is defined as the common component of real GDP within a model including several macroeconomic variables for each European country.

I. Introduction

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

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

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

The model we propose in this paper has both characteristics. It encompasses as a special case the approximate-factor model of Chamberlain (1983) and Chamberlain and Rothschild (1983), which allows for correlated idiosyncratic components but is static. And it generalizes the factor model of Sargent and Sims (1977) and Geweke (1977), which is dynamic but has orthogonal idiosyncratic components.

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

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

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

In the empirical section, we use data on several macroeconomic variables for the countries of the European Monetary Union and compute a reference Euro-zone business cycle, which is defined as a weighted average of the common

Received for publication January 14, 1999. Revision accepted for publication December 14, 1999.

* Università di Modena; Université Libre de Bruxelles, ECARES, and ISRO; Università di Roma; and Université Libre de Bruxelles, ECARES, and ISRO, respectively.

This research has been supported by an A.R.C. contract of the Communauté française de Belgique and by the European Commission under the Training and Mobility of Researchers Programme (Contract No. ERBFMRXCT98-0213). We would like to thank Christine De Mol for generously giving us her time, Serena Ng for pointing out programming problems, Jim Stock for helpful advise, and Jorge Rodrigues for very valuable research assistance. The authors are grateful to an anonymous referee and to the Editor, James Stock, for pointing out an error in an earlier version of the paper.

The Review of Economics and Statistics, November 2000, 82(4): 540–554

components of the GDPs of the countries of the Union and can be driven by more than one common factor. On the basis of our results, we also evaluate the performance of variables (such as the sentiment indicator and the spread) that are usually taken as reference for the European business cycle.

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

II. The Model

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

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

where

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

L standing for the lag operator, and suppose that assumptions (1) through (4) hold.

Assumption (1):

(I) The q-dimensional vector process $\{(u_{1t} \quad u_{2t} \quad \cdots \quad u_{qt})', t \in \mathbb{Z}\}$ is orthonormal white noise. That is, $E(u_{jt}) = 0$; var $(u_{jt}) = 1$ for any j and t; $u_{jt} \perp u_{jt-k}$ for any j, t, and $k \neq 0$; $u_{jt} \perp u_{s,t-k}$ for any $s \neq j$, t, and k;

(II) $\boldsymbol{\xi} = \{\xi_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ is a double sequence such that, firstly,

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

is a zero-mean stationary vector process for any n, and, secondly, $\xi_{it} \perp u_{j,t-k}$ for any i, j, t, and k;

(III) the filters $b_{ij}(L)$ are one-sided in L and their coefficients are square summable.

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

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

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

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

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

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

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

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

Now, denote by λ_{nj} the function associating with any $\theta \in [-\pi, \pi]$ the real nonnegative jth eigenvalue of $\boldsymbol{\Sigma}_n(\theta)$ in descending order of magnitude. The functions λ_{nj} will be called the *dynamic eigenvalues* of $\boldsymbol{\Sigma}_n$.[1] In the same way, with obvious notation, λ^{χ}_{nj} and λ^{ξ}_{nj} denote the dynamic eigenvalues of $\boldsymbol{\Sigma}^{\chi}_n$ and $\boldsymbol{\Sigma}^{\xi}_n$, respectively. The latter will be called *common* and *idiosyncratic eigenvalues,* respectively.

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

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

Assumption (4): The first q common dynamic eigenvalues diverge almost everywhere in $[-\pi, \pi]$. That is,

$$\lim_{n\to\infty} \lambda^{\chi}_{nj}(\theta) = \infty \qquad \text{for } j \leq q, \qquad \text{a.e. in } [-\pi, \pi].$$

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

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

Our first result is the following.

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

Proof: See the appendix.

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

We call model (1), under assumptions (1) to (4), the generalized dynamic-factor model. We will show that, under assumptions (1) through (4), the common components χ_{it} and the idiosyncratic components ξ_{it} are identified and can be consistently estimated. On the other hand, it must be stressed that in this paper we do not deal with identification and estimation of the shocks u_{jt} or the filters $b_{ij}(L)$. Thus, we are not interested here in whether representation (1) has a structural interpretation or not.[2] In this respect, even the assumption that the filters $b_{ij}(L)$ are one-sided could be dropped with no consequence.

III. Recovering the Common Components

A. Population Results

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

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

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

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

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

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

(ii) $|\mathbf{p}_{nj}(\theta)|^2 = 1$ for any j and $\theta \in [-\pi, \pi]$;
(iii) $\mathbf{p}_{nj}(\theta)\tilde{\mathbf{p}}_{ns}(\theta) = 0$ for any $j \neq s$ and any $\theta \in [-\pi, \pi]$;
(iv) $\mathbf{p}_{nj}(\theta)$ is θ − measurable on $[-\pi, \pi]$;

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

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

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

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

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

Defining

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

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

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

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

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

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

and the orthogonal projections

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

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

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

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

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

Taking the i_{th} coordinate,

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

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

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

with

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

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

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

Proposition (2): *Suppose that assumptions (1) through (4) hold. Then,*

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

in mean square for any i and t.

Proof: See the appendix.

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

Remark (3): An intuitive insight into proposition (2) can be obtained by considering the following example:

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

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

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

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

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

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

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

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

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

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

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

Remark (5): An important consequence of corollary (1) is that representation (1) is nonredundant; that is, no other representation fulfilling assumptions (1) through (4) is possible with a smaller number of factors. In the following example, we have a common-idiosyncratic representation of the form (1) with one factor. However, because assumption (4) is not fulfilled, another representation with zero factors fulfilling assumptions (1) through (4) is possible. Specify equation (1) as

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

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

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

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

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

Proof: See the appendix.

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

B. Estimation Results

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

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

On the other hand, consistent estimation of the spectral density requires a strengthening of assumption (2). Precisely, we replace assumption (2) by

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

$$\mathbf{x}_{nt} = \sum_{k=-\infty}^{\infty} \mathbf{C}_k \mathbf{Z}_{t-k},$$

where $\mathbf{Z}_t$ is an n-dimensional white noise with non-singular covariance matrix and fourth-order moments, and

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

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

Under assumption (2′), if $\mathbf{\Sigma}_n^T(\theta)$ denotes any periodogram-smoothing or lag-window estimator of $\mathbf{\Sigma}_n(\theta)$, based on $\mathbf{X}_n^T$, we have

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

where $\sigma_{ij}^T(\theta)$ denotes the i, j entry of $\mathbf{\Sigma}_n^T(\theta)$. (See Brockwell and Davis (1987, p. 433).) Under assumption (2′), the estimated counterpart of $\mathbf{K}_{ni}(\theta)$ allows for a consistent reconstruction of the factor space. More precisely, we prove that the projection of x_{it} onto the space spanned by the first q empirical principal components converges to the common component χ_{it}.

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

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

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

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

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

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

can be obtained. However, in practice, the projection $\underline{\mathbf{K}}_{ni}^T(L)\mathbf{x}_{nt}$ of x_{it} onto the space spanned by the first q empirical principal components cannot be computed, because, for $t \le 0$ and $t > T$, $\mathbf{x}_{nt}$ is not available. Therefore, a truncated version of the estimated filters $\underline{\mathbf{K}}_{ni}^T(L)$, of the form $\sum_{k=-M(T)}^{M(T)} \underline{\mathbf{K}}_{ni,k}^T L^k$, where $M(T) \to \infty$ is such that $\limsup_{T\to\infty} M^3(T)/T < \infty$ as $T \to \infty$, is considered. This $T^{-1/3}$ rate for the window width $M(T)$ is related to the method of proof (see the proof of Proposition (3) in the Appendix), and is probably not essential to the consistency result itself. Even this truncated version of $\underline{\mathbf{K}}_{ni}^T(L)$ however has to be truncated further when acting on $\boldsymbol{x}_{nt}$, $t \le M(T)$ or $t > T - M(T)$, yielding

$$\underline{\mathbf{K}}_{ni}^{Tt}(L) = \sum_{k=\max(T-t,-M(T))}^{\min(t-1,M(T))} \underline{\mathbf{K}}_{ni,k}^T L^k.$$

Due to this unavoidable truncation, the common component χ_{it}, for fixed t, never can be recovered, even as n and T tend to infinity. Indeed, part of its variance is lost because of the non-observability of $\mathbf{x}_{nt}$, $t \le 0$ and $t > T$. We therefore restrict our attention to the "central part" of the observed series, the values of t of the form $t = t^*(T)$, with

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

The following result then provides the empirical counterpart of proposition (2).

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

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

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

Proof: Throughout, we write aT instead of $\lfloor aT \rfloor$ (the smallest integer larger than or equal to aT), and bT for $\lceil bT \rceil$ (the largest integer smaller than or equal to bT). We also tacitly assume that T is large enough for $M(T)$ being strictly less than min $(aT, T - bT + 1, (bT - aT)/2)$.

For any $t \in [aT, bT]$, $\underline{\mathbf{K}}_{ni}^{Tt}(L)$ then reduces to $\sum_{k=-M(T)}^{M(T)} \underline{\mathbf{K}}_{ni,k}^T L^k$. We have

$$P[|\underline{\mathbf{K}}_{ni}^{Tt}(L)\mathbf{x}_{nt} - \chi_{it}| > \epsilon] \le P[|(\underline{\mathbf{K}}_{ni}^{Tt}(L) - \underline{\mathbf{K}}_{ni}(L))\mathbf{x}_{nt}| > \epsilon/2] + P[|\underline{\mathbf{K}}_{ni}(L)\mathbf{x}_{nt} - \chi_{it}| > \epsilon/2] = R_{n1}^{Tt} + R_{n2}, \qquad \text{say.}$$

Proposition (2) ensures the existence of an $N_0(\epsilon, \eta)$ such that, for $n \geq N_0$, $R_{n2} \leq \frac{\eta}{2}$. As for R_{n1}^{Tt}, we have, from the definition of $\underline{\mathbf{K}}_{ni}^{Tt}$,

$$\begin{aligned} R_{n1}^{Tt} &\leq P\Bigg[\Bigg|\Bigg(\sum_{h=-M(T)}^{M(T)} (\mathbf{K}_{ni,h}^{T} - \mathbf{K}_{ni,h})L^h \\ &\quad - \sum_{h=-\infty}^{-M(T)-1} \mathbf{K}_{ni,h}L^h - \sum_{h=M(T)+1}^{\infty} \mathbf{K}_{ni,h}L^h\Bigg)\mathbf{x}_{nt}\Bigg| > \frac{\epsilon}{2}\Bigg] \\ &\leq P\Bigg[\Bigg|\sum_{h=-M(T)}^{M(T)} (\mathbf{K}_{ni,h}^{T} - \mathbf{K}_{ni,h})L^h\mathbf{x}_{nt}\Bigg| > \frac{\epsilon}{4}\Bigg] \\ &\quad + P\Bigg[\Bigg|\Bigg(\sum_{h=-\infty}^{-M(T)-1} \mathbf{K}_{ni,h}L^h - \sum_{h=M(T)+1}^{\infty} \mathbf{K}_{ni,h}L^h\Bigg)\mathbf{x}_{nt}\Bigg| > \frac{\epsilon}{4}\Bigg] \\ &= R_{n11}^{Tt} + R_{n12}^{T}. \end{aligned}$$

Since $\underline{\mathbf{K}}_{ni}(L)$ is a nonrandom square-summable filter, there exists $T_2(n, \epsilon, \eta)$ such that $R_{n12}^{T} \leq \frac{\eta}{4}$ for all $T \geq T_2$ and $aT < t < bT$. Turning to R_{n11}^{Tt}, it follows from Chebyshev's theorem and equation (6) that, for $T \geq T_1(n, \delta, \frac{\eta}{8})$,

$$\begin{aligned} R_{n11}^{Tt} &\leq P\Bigg[\Bigg|\Bigg(\sum_{h=-M(T)}^{M(T)} (\mathbf{K}_{ni,h}^{T} - \mathbf{K}_{ni,h})L^h\mathbf{x}_{nt}\Bigg| > \frac{\epsilon}{4} \\ &\qquad \text{and} \quad \sup_{\theta\in[-\pi,\pi]} |\mathbf{K}_{ni}^{T}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta\Bigg] \\ &\quad + P\Bigg[\sup_{\theta\in[-\pi,\pi]} |\mathbf{K}_{ni}^{T}(\theta) - \mathbf{K}_{ni}(\theta)| > \delta\Bigg] \\ &\leq \frac{16}{\epsilon^2} E\Bigg[\Bigg|\sum_{h=-M(T)}^{M(T)} (\mathbf{K}_{ni,h}^{T} - \mathbf{K}_{ni,h})L^h\mathbf{x}_{nt}\Bigg|^2 \\ &\qquad \times \Big| \sup_{\theta\in[-\pi,\pi]} |\mathbf{K}_{ni}^{T}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta\Bigg] + \frac{\eta}{8}. \end{aligned}$$

If the filter $\underline{\mathbf{K}}_{ni}^{T}$ and the observation $\mathbf{x}_{nt}$ were independent, then, in view of the classical properties of dynamic principal components (see the proof of Proposition 1 in the appendix), we would have (denoting by $\mathbf{\Gamma}_{n,h} = E[\mathbf{x}_{n,t}\mathbf{x}'_{n,t-h}]$, $h = 0, \pm 1, \ldots$ the autocovariance function of $\{\mathbf{x}_{nt}\}$)

$$\begin{aligned} &E\Bigg[\Bigg|\sum_{h=-M(T)}^{M(T)} (\mathbf{K}_{ni,h}^{T} - \mathbf{K}_{ni,h})L^h\mathbf{x}_{nt}\Bigg|^2 \\ &\qquad \times \Big| \sup_{\theta\in[-\pi,\pi]} |\mathbf{K}_{ni}^{T}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta\Bigg] \\ &= E\Bigg[\sum_{k=-M(T)}^{M(T)} \sum_{l=-M(T)}^{M(T)} (\mathbf{K}_{ni,k}^{T} - \mathbf{K}_{ni,k})\mathbf{\Gamma}_{n,l-k} \\ &\qquad \times (\tilde{\mathbf{K}}_{ni,l}^{T} - \tilde{\mathbf{K}}_{ni,l}) \Big| \sup_{\theta\in[-\pi,\pi]} |\mathbf{K}_{ni}^{T}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta\Bigg] \\ &\leq E\left[\delta^2 \int_{-\pi}^{\pi} \lambda_{n1}(\theta)\, d\theta\right] = \delta^2 \int_{-\pi}^{\pi} \lambda_{n1}(\theta)\, d\theta. \end{aligned}$$

Thus, letting $\delta^2 = \epsilon^2\eta/128 \int_{-\pi}^{\pi} \lambda_{n1}(\theta)\, d\theta$, for $n \geq N_0(\epsilon, \eta)$ and $T \geq \max(T_1(n, \delta, \frac{\eta}{8}), T_2(n, \epsilon, \eta))$, we would obtain $R_{n11}^{Tt} \leq \frac{\eta}{4}$, hence $R_{n1}^{Tt} + R_{n2}^{T} \leq \eta$ for all $t = t^*(T)$ satisfying condition (7). The proof of Proposition 3 then would be complete—without any rate assumption on $M(T)$.

Unfortunately, $\underline{\mathbf{K}}_{ni}^{T}$ and $\mathbf{x}_{nt}$ are not independent. For T large enough, they are "almost independent," though, so that the above reasoning is essentially correct; moreover, it provides the right insight into the intuitive ideas underlying the proof. A more formal treatment, taking into account the non-independence between $\underline{\mathbf{K}}_{ni}^{T}$ and $\mathbf{x}_{nt}$, is given in the Appendix.

IV. The Proposed Estimator and the Choice of q

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

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

where

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

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

Then we compute the first q eigenvectors $\mathbf{p}_{nj}^{T}(\theta_h)$, $j = 1, 2, \ldots, q$, of $\mathbf{\Sigma}_{n}^{T}(\theta_h)$, for $h = 0, 1, \ldots, 2M$.[3] Finally, for $h = 0, 1, \ldots, 2M$, we construct

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

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

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

that is, by the computation of

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

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

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

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

Note that the same integer M has been used as the size of the Bartlett window in the estimation of $\boldsymbol{\Sigma}_n^T(\theta)$, and as the truncation length of $\underline{\mathbf{K}}_{ni}^{Tt}(L)$, so that imposing $M(T) = O(T^{1/3})$ ensures both consistency of the estimated spectrum and consistency of the estimated common component (see Proposition 3). In particular, it appears that $M(T) =$ round $(\frac{2}{3}T^{1/3})$ performs remarkably well in the simulations reported in Section V. As an alternative, we could take any sequence $M_0(T)$ such that $M_0(T) \to \infty$ as $T \to \infty$, and $M_0(T) = O(T^{1/3})$, estimate all of the specifications with $0 \leq M \leq M_0(T)$, and choose the one minimizing some dynamic specification criterion. Although a data-dependent rule seems preferable in principle, we found that the standard AIC and BIC criteria underestimate the optimal lag-window size, so this topic is left for further research.

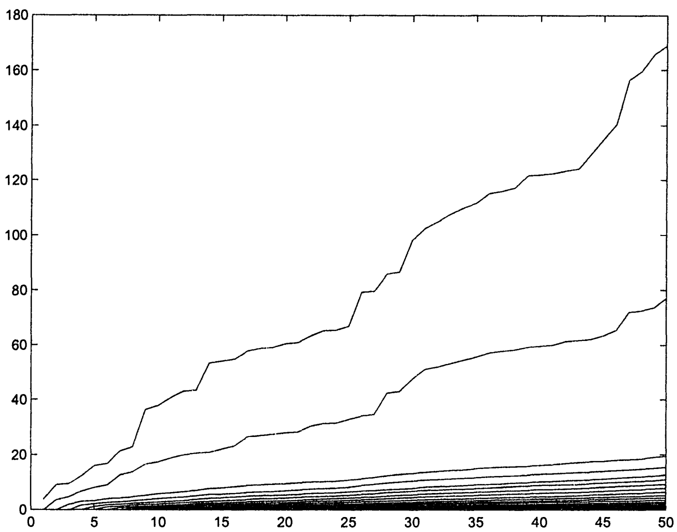
So far, we have assumed that q, the number of non-redundant common factors, is known. In practice of course, q is not predetermined and also has to be selected from the data. Proposition (1) can be used to this end, because it links the number of factors in equation (1) to the eigenvalues of the spectral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

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

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

1. The average over θ of the first q empirical eigenvalues diverges, whereas the average of the $(q + 1)_{\text{th}}$ one is relatively stable.
2. Taking $r = n$, there is a substantial gap between the variance explained by the q_{th} principal component and the variance explained by $(q + 1)_{\text{th}}$ one. A preassigned minimum, such as 5%, for the explained variance, could be used as a practical criterion for the determination of the number of common factors to be retained. This 5% limit is used in the empirical exercise of section VI.

FIGURE 1.—DYNAMIC EIGENVALUES AVERAGED OVER FREQUENCIES, MODEL M4



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

To illustrate the use of criteria (1) and (2), we have generated data from a two-factor model (model M4 below) with $n = 50$ and $T = 100$. Then, we have estimated the spectral density matrix for a grid of frequencies, using equation (8) with $M = 10$. Lastly, we have computed the eigenvalues of the upper-left $r \times r$ submatrices, $r = 1, \ldots, n$.

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

To conclude this section, let us remark that, when applying criteria (1) and (2), we should keep in mind that, as indicated by corollary (2), setting a number q^* of factors larger than the true one q cannot have dramatic consequences on estimation.

V. Simulation Results

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

Static model:

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

Static with delay:

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

MA(1) common component:

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

AR(1) common component:

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

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

We generated data from each model with $n = 10, 20, 50, 100$, and $T = 20, 50, 100, 200$, and applied the estimation procedure described in section IV with $M(T) = \text{round}\,[\frac{2}{3}T^{1/3}]$. Each experiment was replicated 400 times.

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

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

Table 1 reports the average and the standard deviation (in brackets) of this statistic across the experiments.

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

TABLE 1.—AVERAGE AND STANDARD DEVIATION (IN BRACKETS) OF $R(\hat{\chi},\chi)$ ACROSS 400 EXPERIMENTS

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

VI. A Coincident Indicator for the EURO Currency Area

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

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

We proceed as follows.

Step (1): We construct a panel pooling seven quarterly macroeconomic indicators for all countries of the EURO zone, excluding Luxembourg, from 1985 to 1996. (See table 2, with X and − indicating, respectively, whether the series is available or missing.) Data are taken in logs and differenced (except for the spread, which is not transformed, and the sentiment indicator, which is simply taken in logs), and normalized dividing by the standard deviation. Thus, $n = 63$ and $T = 51$.

Step (2): We estimate the spectral density matrix, compute the dynamic eigenvalues, and identify $q = 3$, using criterion (2) of section IV.

TABLE 2.—THE DATA

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

GDP: GDP, s.a., in national currency, at constant (1990) prices. Source: OECD; for Germany and Portugal: IMF.
Cons.: Private final consumption expenditure, s.a., in national currency at constant (1990) prices. Source: OECD; for Germany and Portugal: IMF.
Inv.: Gross fixed-capital formation, s.a., in national currency at constant (1990) prices. Source: OECD; for Germany and Portugal: IMF.
CPI: Consumer Price Index, base year 1990.
Spread: Difference between the government bond yield and the Treasury Bill rate (or the money market rate depending on data availability), in percentage per year. Source: IMF.
Sent.: economic sentiment indicator. Source: European Commission, DG II.
I.P.: Industrial production, s.a., index number, base year 1990. Source: IMF.

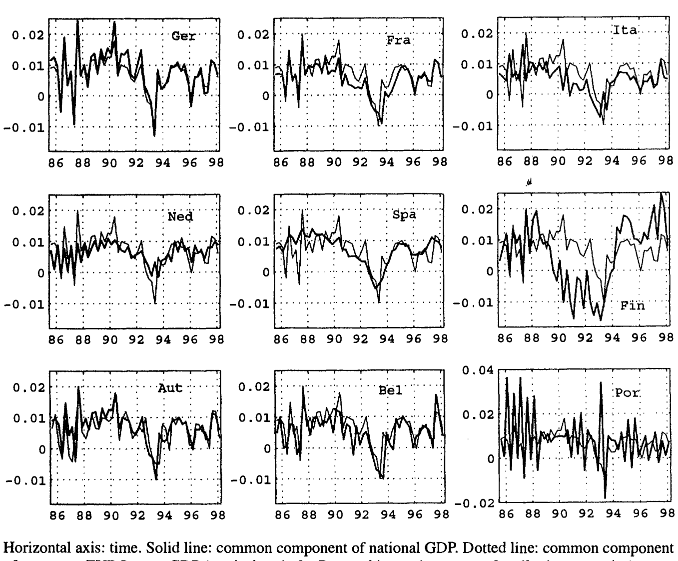
Step (3): We estimate the common component of GDP for each separate country, following the procedure of section IV. Figure 2 reports the resulting estimates.

Step (4): We construct the weighted average of the common components above using as weights the GDP levels. This is the proposed coincident indicator. We illustrate it in figure 3.

Some remarks are needed. Firstly, results from step (2) show that the presence of a cycle in the strong sense of Stock and Watson (1989)—that is, a single common factor—is not supported by this data set.

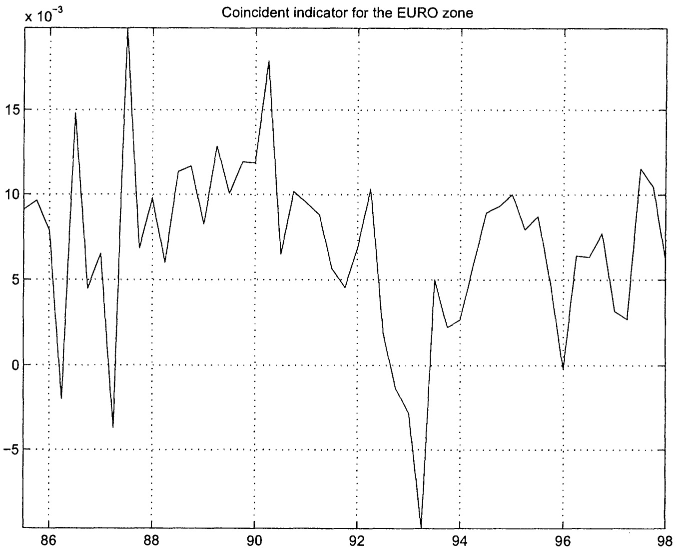
Secondly, in our methodology, output plays a prominent role. Because the data do not support a single static factor, the cycle must be defined as the common component of a particular cross-sectional unit. Clearly, GDP is the most natural choice as the reference variable. On the other hand,

FIGURE 2.—COMMON COMPONENT OF GDP OF NINE COUNTRIES OF THE EURO ZONE



Horizontal axis: time. Solid line: common component of national GDP. Dotted line: common component of aggregate EURO-zone GDP (vertical scale for Portugal is not the same as for all other countries).

FIGURE 3.—COINCIDENT INDICATOR FOR THE EURO ZONE



Horizontal axis: time.

we are interested in the common component of output and not in output itself, because we want to disregard that part of GDP variation that is poorly correlated with other variables. Hence, the latter also play an indirect role in the construction of the index, through the estimated dynamic principal components.

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

In order to understand better the structure of the multicountry, multivariate correlations, we also compute the common components of variables other than GDP and construct two sets of statistics. First, for each variable, we compute the ratio of the variance of the common component to total variance, for each country and for the aggregate EURO area (table 3). These ratios measure the degree of commonality of each variable in the system. Second, we compute, for each variable, the average contemporaneous correlation coeffi-

TABLE 3.—PERCENTAGE OF VARIANCE EXPLAINED BY THE COMMON COMPONENT

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

TABLE 4.—AVERAGE CORRELATION OF THE COMMON COMPONENT WITH THE COMMON COMPONENTS OF THE OTHER VARIABLES OF THE SAME COUNTRY

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

cient with the common components of the other variables of the same country, for each country and the aggregate EURO area (table 4). These statistics measure the degree of synchronization of each variable with the other variables of the same country. Through these results, we can also evaluate the performance of variables that are typically used to describe the state of the economy, such as the sentiment indicator and the spread, and validate ex post the choice of GDP as the reference variable for the European cycle.

There are a few interesting findings. First, the common component of GDP has the largest average contemporaneous correlation for almost all countries and for the aggregate. This fact provides an ex post confirmation of our choice of the GDP as the reference variable for the coincident index. Note, however, that using directly the GDP, rather than the common component of GDP, as the index would not be a good choice, due to the presence of an idiosyncratic component that accounts for 15% of total variance.

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

Note that the correlations between the common components appearing in table 4 are, in general, unexpectedly small. This is mainly due to the fact that, somewhat surprisingly, the inflation rate has very low or even negative synchronization.

VII. Summary and Discussion

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

Because the idiosyncratic components are correlated, the model cannot be estimated on the basis of traditional methods. We have proposed a new method, yielding consistent estimates of the components as both the cross section and the time dimensions go to infinity at some rate. More precise information on these rates would be interesting; however, such information typically would require much heavier assumptions on the heterogeneity of cross-sectional units. This is a topic of the authors' ongoing research; see Forni, Hallin, Lippi, & Reichlin (2000). The common components are computed as the projections of the observations onto the leads and lags of the dynamic principal components of the observations and the idiosyncratic components are derived as the orthogonal residuals.

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

REFERENCES

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

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

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

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

Chamberlain, Gary, "Funds, Factors, and Diversification in Arbitrage Pricing Models," *Econometrica* 51 (1983), 1281–1304.

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

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

——— "The Generalized Factor Model: Representation Theory," I.S.R.O. working paper no. 132, Université Libre de Bruxelles (1999) and CEPR discussion paper no. 2509. Forthcoming in *Econometric Theory*.

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin, "The Generalized Dynamic-Factor Model: Consistency and Rates," I.S.R.O. working paper, Université Libre de Bruxelles (2000).

Forni, Mario, and Lucrezia Reichlin, "Dynamic Common Factors in Large Cross-Sections," *Empirical Economics* 21 (1996), 27–42.

——— "Let's Get Real: A Factor Analytic Approach to Disaggregated Business Cycle Dynamics," this REVIEW 65 (1998), 453–473.

——— "Federal Policies and Local Economies: Europe and the U.S.," CEPR discussion paper series no. 1632, (1997), forthcoming in *European Economic Review* (2000).

Geweke, John, "The Dynamic Factor Analysis of Economic Time Series," in Dennis J. Aigner and Arthur S. Goldberger (Eds.), *Latent Variables in Socio-Economic Models* (Amsterdam: North-Holland, 1977).

Quah, Danny, and Thomas J. Sargent, "A Dynamic Index Model for Large Cross Sections," ch. 7 in J. H. Stock and M. W. Watson (Eds.), *Business Cycles, Indicators and Forecasting, NBER* (1993) 285–309.

Sargent, Thomas J., and Christopher A. Sims, "Business Cycle Modeling without Pretending to Have Too Much *a Priori* Economic Theory," in Christopher A. Sims (Ed.), *New Methods in Business Research* (Minneapolis: Federal Reserve Bank of Minneapolis, 1977).

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

——— "New Indexes of Coincident and Leading Economic Indicators," *NBER Macroeconomic Annual* (1989), 351–394.

APPENDIX A

Proof of Proposition (1). We need the following result (see Brillinger, 1981, p. 84, Exercise 3.10.16):
Let $\mathbf{\Lambda}$ be an $n \times n$, complex, hermitian nonnegative definite matrix, and let λ_k, $k = 1, \dots, n$, be its (real) eigenvalues in descending order of magnitude. Denote by $\mathbf{D}_k$, an $n \times (k-1)$ complex matrix for $1 < k \leq n$, the $n \times 1$ null matrix for $k = 1$. The eigenvalue λ_k is the solution of

$$\min_{\mathbf{D}_k} \max_{\mathbf{b}} \mathbf{b}\mathbf{\Lambda}\tilde{\mathbf{b}} \qquad \text{s.t. } |\mathbf{b}| = 1, \qquad \mathbf{b}\tilde{\mathbf{D}}_k = \mathbf{0}. \tag{10}$$

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

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

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

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

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

From equation (10) and the above inequalities, we get

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

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

To prove proposition (2) we need some intermediate results. We suppose that assumptions (1) through (4) hold.

Lemma (1): *Denote by $p_{nj,i}(\theta)$ the i_{th} component of* $\mathbf{p}_{nj}(\theta)$, as defined in Section III.A. For $j \leq q$, $\lim_{n\to\infty} |p_{nj,i}(\theta)| = 0$ *a.e. in* $[-\pi, \pi]$.

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

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

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

Lemma (2): *For given i and $n \in \mathbb{N}$, consider the n-dimensional filters (defined in section III.A)*

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

Then

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

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

Proof: We have

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

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

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

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

Then

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

in mean square.

Proof: From the same argument as in the proof of proposition (1), we have

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

The result follows from assumption (3). QED

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

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

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

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

in mean square.

With no loss of generality we can assume that

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

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

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

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

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

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

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

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

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

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

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

then the desired result

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

follows, since $\lim_{n\to\infty} \underline{\mathbf{K}}_{ni}(L)\boldsymbol{\phi}_{nt} = 0$ by lemma 3.

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

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

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

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

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

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

$$\begin{aligned}(u_{1t} \quad u_{2t} \quad \ldots \quad u_{qt})'\\ = \tilde{\mathbf{A}}_n(L^{-1})(W_{n1,t} \quad W_{n2,t} \quad \ldots \quad W_{nq,t})' + \mathbf{S}_{nt},\end{aligned} \tag{13}$$

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

Proof: Firstly, observe that

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

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

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

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

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

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

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

The result follows. QED

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

Proof of Proposition (2): We have

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

with

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

Combining equations (14) and (15), we obtain

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}&\mathbf{K}_{ni}(\theta)\mathbf{B}_n(e^{-i\theta})\boldsymbol{\Sigma}_n^S(\theta)\tilde{\mathbf{B}}_n(e^{i\theta})\tilde{\mathbf{K}}_{ni}(\theta)\\ &\le \lambda_{n1}^S(\theta)\mathbf{K}_{ni}(\theta)\mathbf{B}_n(e^{-i\theta})\tilde{\mathbf{B}}_n(e^{i\theta})\tilde{\mathbf{K}}_{ni}(\theta).\end{aligned}$$

QED

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

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

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

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

The result follows from assumption (3). QED

Proof of Proposition (3): As mentioned at the end of Section III, $\underline{\mathbf{K}}^T_{ni}(L)$ and $\mathbf{x}_{nt}$, for fixed T (and n), are mutually dependent. Their dependence structure however is pretty intricate, and hardly can be explicitated. The random variables $\Sigma_{h=-M(T)}^{M(T)}(\mathbf{K}^T_{ni,h} - \mathbf{K}_{ni,h})\mathbf{x}_{n,t-h}$ are not identically distributed, since boundary effects imply that the joint distributions of $(\mathbf{K}^T_{ni,h}, \mathbf{x}_{nt})$ are not the same for $t = 1$ or $t = T$ as for $t = (T/2)$, for instance. However, such boundary effects are asymptotically negligible for *central* values of t, satisfying (7) for some a and b. More precisely, there exists a $T_3 = T_3(n, \delta, \eta)$ such that, for all $s, t \in [at, bT]$,

$$\left| E\left[\left| \sum_{h=-M(T)}^{M(T)} (\mathbf{K}^T_{ni,h} - \mathbf{K}_{ni,h})L^h \mathbf{x}_{nt} \right|^2 \Big| \sup_{\theta\in[-\pi,\pi]} |\mathbf{K}^T_{ni}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta \right] - E\left[\left| \sum_{h=-M(T)}^{M(T)} (\mathbf{K}^T_{ni,h} - \mathbf{K}_{ni,h})L^h \mathbf{x}_{ns} \right|^2 \Big| \sup_{\theta\in[-\pi,\pi]} |\mathbf{K}^T_{ni}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta \right] \right| \leq \eta.$$

Averaging over $s \in [aT, bT]$, it follows that, for $T \geq T_3(n, \delta, \epsilon^2\eta/256)$,

$$\begin{aligned}
&E\left[\left| \sum_{h=-M(T)}^{M(T)} (\mathbf{K}^T_{ni,h} - \mathbf{K}_{ni,h})\mathbf{x}_{n,t-h} \right|^2 \Big| \sup_{\theta\in[-\pi,\pi]} |\mathbf{K}^T_{ni}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta \right] \\
&\leq (bT - aT)^{-1} \sum_{s=aT}^{bT} E\left[\left| \sum_{h=-M(T)}^{M(T)} (\mathbf{K}^T_{ni,h} - \mathbf{K}_{ni,h})L^h \mathbf{x}_{ns} \right|^2 \Big| \cdots \leq \delta \right] \\
&\quad + \frac{\epsilon^2\eta}{256} = (bT - aT)^{-1} \sum_{s=aT}^{bT} E\left[\sum_{k=-M(T)}^{M(T)} \sum_{l=-M(T)}^{M(T)} \right. \qquad (17)\\
&\quad \times \left. (\mathbf{K}^T_{ni,k} - \mathbf{K}_{ni,k})\mathbf{x}_{n,s-k}\mathbf{x}'_{n,s-l}\, (\tilde{\mathbf{K}}^T_{ni,l} - \tilde{\mathbf{K}}_{ni,l}) \Big| \cdots \leq \delta \right] \\
&\quad + \frac{\epsilon^2\eta}{256} = E\left[\sum_{k=-M(T)}^{M(T)} \sum_{l=-M(T)}^{M(T)} (\mathbf{K}^T_{ni,k} - \mathbf{K}_{ni,k})\bar{\mathbf{\Gamma}}^T_{n,kl} \right. \\
&\quad \times \left. (\tilde{\mathbf{K}}^T_{ni,l} - \tilde{\mathbf{K}}_{ni,l}) \Big| \cdots \leq \delta \right] + \frac{\epsilon^2\eta}{256},
\end{aligned}$$

where

$$\bar{\mathbf{\Gamma}}^T_{n,kl} = \begin{cases} \dfrac{1}{bT - aT} \displaystyle\sum_{s=aT}^{bT} \mathbf{x}_{n,s-k}\mathbf{x}'_{n,s-l} & |k| \text{ and } |l| \leq M(T) \\ \mathbf{0} & |k| \text{ or } |l| > M(T) \end{cases}.$$

Letting $s' = s - k$, the matrices $\bar{\mathbf{\Gamma}}^T_{n,kl}$, for $|k| \leq M(T)$, $|l| \leq M(T)$, and $l - k \geq 0$ take the form

$$\begin{aligned}
\bar{\mathbf{\Gamma}}^T_{n,kl} &= \frac{1}{bT - aT} \sum_{s'=aT+M(T)}^{bT-M(T)} \mathbf{x}_{n,s'}\mathbf{x}'_{n,s'-(l-k)} \\
&\quad + \frac{1}{bT - aT} \sum_{s=aT}^{aT+M(T)+k-1} \mathbf{x}_{n,s-k}\mathbf{x}'_{n,s-l} + \frac{1}{bT - aT} \\
&\quad \times \sum_{s=bT-M(T)+k+1}^{bT} \mathbf{x}_{n,s-k}\mathbf{x}'_{n,s-l} = \bar{\mathbf{\Gamma}}^{T*}_{n,(l-k)} + \bar{\boldsymbol{\gamma}}^{T1}_{n,kl} + \bar{\boldsymbol{\gamma}}^{T2}_{n,kl}, \text{ say;}
\end{aligned}$$

whenever $|k| \leq M(T)$, $|l| \leq M(T)$ but $l - k < 0$, letting $s'' = s - l$, the same matrices similarly decompose into

$$\begin{aligned}
\bar{\mathbf{\Gamma}}^T_{n,kl} &= \frac{1}{bT - aT} \sum_{s''=aT+M(T)}^{bT-M(T)} \mathbf{x}_{n,s''+(l-k)}\mathbf{x}'_{n,s''} + \frac{1}{bT - aT} \\
&\quad \times \sum_{s=aT}^{aT+M(T)+l-1} \mathbf{x}_{n,s-k}\mathbf{x}'_{n,s-l} + \frac{1}{bT - aT} \sum_{s=bT-M(T)+l+1}^{bT} \mathbf{x}_{n,s-k}\mathbf{x}'_{n,s-l} \\
&= \bar{\mathbf{\Gamma}}^{T*}_{n,(l-k)} + \bar{\boldsymbol{\gamma}}^{T1}_{n,kl} + \bar{\boldsymbol{\gamma}}^{T2}_{n,kl}, \text{ say.}
\end{aligned}$$

The conditional expectation in the right-hand side of (17) similarly decomposes into

$$\begin{aligned}
&E\left[\sum_{k=-M(T)}^{M(T)} \sum_{l=-M(T)}^{M(T)} (\mathbf{K}^T_{ni,k} - \mathbf{K}_{ni,k})\bar{\mathbf{\Gamma}}^{T*}_{n,(l-k)}(\tilde{\mathbf{K}}^T_{ni,l} - \tilde{\mathbf{K}}_{ni,l}) \Big| \cdots \leq \delta \right] \\
&+ E\left[\sum_{k=-M(T)}^{M(T)} \sum_{l=-M(T)}^{M(T)} (\mathbf{K}^T_{ni,k} - \mathbf{K}_{ni,k})\bar{\boldsymbol{\gamma}}^{T1}_{n,kl}(\tilde{\mathbf{K}}^T_{ni,l} - \tilde{\mathbf{K}}_{ni,l}) \Big| \cdots \leq \delta \right] \\
&+ E\left[\sum_{k=-M(T)}^{M(T)} \sum_{l=-M(T)}^{M(T)} (\mathbf{K}^T_{ni,k} - \mathbf{K}_{ni,k})\bar{\boldsymbol{\gamma}}^{T2}_{n,kl}(\tilde{\mathbf{K}}^T_{ni,l} - \tilde{\mathbf{K}}_{ni,l}) \Big| \cdots \leq \delta \right] \\
&= E^{T*}_{n,\delta} + E^{T1}_{n,\delta} + E^{T2}_{n,\delta}, \text{ say.}
\end{aligned}$$

Due to the fact that the sums (running over s) defining $\bar{\boldsymbol{\gamma}}^{T\nu}_{n,kl}$, $\nu = 1, 2$, which are divided by $(bT - aT)$, either involve $M(T) + k \leq 2M(T)$ or $M(T) + l \leq 2M(T)$ terms, the corresponding expectations $E^{T\nu}_{n,\delta}$, $\nu = 1, 2$ are "small" as $T \to \infty$. More precisely, assuming that $T \geq T_1(n, \delta, \frac{1}{2})$, we have (denoting by $\Sigma^*_{s=aT}$ a sum running over s, either from aT to $aT + M(T) + k - 1$, or from aT to $aT + M(T) + l - 1$, depending on the sign of $(l - k)$),

$$\begin{aligned}
E^{T1}_{n,\delta} &= \frac{1}{bT - aT} E\left[\sum_{k=-M(T)}^{M(T)} \sum_{l=-M(T)}^{M(T)} \sum_{u=1}^{n} \sum_{v=1}^{n} (\mathbf{K}^T_{ni,k} - \mathbf{K}_{ni,k})_u \right. \\
&\quad \times \left. \sum_{s=aT}^{*} (\mathbf{x}_{n,s-k})_u(\mathbf{x}_{n,s-l})_v(\tilde{\mathbf{K}}^T_{ni,l} - \tilde{\mathbf{K}}_{ni,l})_v \Big| \cdots \leq \delta \right] \\
&\leq \frac{2\delta^2}{bT - aT} \sum_{u=1}^{n} \sum_{v=1}^{n} \sum_{s=aT}^{aT+M(T)+k-1} \sum_{k=-M(T)}^{M(T)} \sum_{l=-M(T)}^{M(T)} E[|(\mathbf{x}_{n,s-k})_u(\mathbf{x}_{n,s-l})_v|] \\
&\leq \frac{2\delta^2}{bT - aT} n^2 2M(T)(2M(T) + 1)^2 \max_{1\leq u\leq n} \mathrm{var}\,((\mathbf{x}_{n,t})_u). \qquad (18)
\end{aligned}$$

The assumption that T is larger than $T_1(n, \delta, \frac{1}{2})$ has been used in substituting twice the unconditional expectations $E[\ldots]$ for the conditional ones $E[\ldots| \ldots \leq \delta]$; in view of (6), $T \geq T_1(n, \delta, \frac{1}{2})$ indeed implies that $P[\sup_{\theta\in[-\pi,\pi]} |\mathbf{K}^T_{ni}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta] \leq \frac{1}{2}$, so that

$$\begin{aligned}
&E[|(\mathbf{x}_{n,s-k})_u(\mathbf{x}_{n,s-l})_v|\, | \cdots \leq \delta] \\
&= \frac{E[|(\mathbf{x}_{n,s-k})_u(\mathbf{x}_{n,s-l})_v|] - E[|(\mathbf{x}_{n,s-k})_u(\mathbf{x}_{n,s-l})_v|\, | \cdots > \delta]P[\cdots > \delta]}{P[\sup_{\theta\in[-\pi,\pi]} |\mathbf{K}^T_{ni}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta]} \\
&\leq \frac{E[|(\mathbf{x}_{n,s-k})_u(\mathbf{x}_{n,s-l})_v|]}{P[\sup_{\theta\in[-\pi,\pi]} |\mathbf{K}^T_{ni}(\theta) - \mathbf{K}_{ni}(\theta)| \leq \delta]} \\
&\leq 2E[|(\mathbf{x}_{n,s-k})_u(\mathbf{x}_{n,s-l})_v|].
\end{aligned}$$

It follows from (18) that $E^{T1}_{n,\delta}$ is $O(\delta^2 M^3(T)/T)$ as $T \to \infty$. A similar conclusion holds for $E^{T2}_{n,\delta}$. Hence, in view of the rate assumption on $M(T)$, we may conclude that there exists $T_4(n)$ and a constant $M^*(n)$ such that $T > T_4^*(n, \delta) = \max(T_4(n), T_1(n, \delta, \frac{1}{2}))$ entails $E_{n,\delta}^{T1} + E_{n,\delta}^{T2} \le \delta^2 M^*(n)$.

Turning to the main term $E_{n,\delta}^{T*}$, note that $\bar{\Gamma}_{n,h}^{T*}$, $h \in \mathbb{Z}$ is an empirical autocovariance function, in which each covariance matrix is computed from the same number $(bT - aT - 2M(T))$ of observations. In contrast with $\bar{\gamma}_{n,kl}^{T1}$ and $\bar{\gamma}_{n,kl}^{T2}$, $\bar{\Gamma}_{n,h}^{T*}$ is based on sums which involve a number of terms of the order of T, and are not "small" compared with $bT - aT$. Associated with this empirical autocovariance function is the empirical spectral density $\bar{\Sigma}_n^{T*}$, with dynamic eigenvalues $\bar{\lambda}_{nj}^{T*}(\theta)$, $j = 1, \ldots, n$. The properties of dynamic eigenvalues imply that

$$E\left[\sum_{k=-M(T)}^{M(T)} \sum_{l=-M(T)}^{M(T)} (\mathbf{K}_{ni,k}^{T} - \mathbf{K}_{ni,k})\bar{\Gamma}_{n,l-k}^{T*}(\tilde{\mathbf{K}}_{ni,l}^{T} - \tilde{\mathbf{K}}_{ni,l})\bigg| \cdots \le \delta\right]$$
$$\le E\left[\delta^2 \int_{-\pi}^{\pi} \bar{\lambda}_{n1}^{T*}(\theta)\, d\theta \bigg| \cdots \le \delta\right] \le 2\delta^2 E\left[\int_{-\pi}^{\pi} \bar{\lambda}_{n1}^{T*}(\theta)\, d\theta\right]$$

provided that $T \ge T_1(n, \delta, \frac{1}{2})$ (the factor 2, as in (18), is due to the substitution of unconditional expectations for the conditional ones). Now, the random sequence $\int_{-\pi}^{\pi} \bar{\lambda}_{n1}^{T*}(\theta)\, d\theta$ is a.s. bounded by

$$\int_{-\pi}^{\pi} \mathrm{tr}\,[\bar{\Sigma}_n^{T*}(\theta)]\, d\theta = \mathrm{tr}\left[\int_{-\pi}^{\pi} [\bar{\Sigma}_n^{T*}(\theta)]\, d\theta\right]$$
$$= \mathrm{tr}\,[\bar{\Gamma}_{n,0}^{T*}] = (bT - aT)^{-1} \sum_{i=1}^{n} \sum_{s=aT+M(T)}^{bT-M(T)} (\mathbf{x}_{ns})_i^2$$

($\mathrm{tr}\,(\Gamma)$ stands for the trace of Γ). Hence (noting that $E[(\mathbf{x}_{ns})_i^2]$ does not depend on s),

$$E\left[\int_{-\pi}^{\pi} \bar{\lambda}_{n1}^{T*}(\theta)\, d\theta\right] \le E\left[(bT - aT)^{-1} \sum_{i=1}^{n} \sum_{s=aT+M(T)}^{bT-M(T)} (\mathbf{x}_{ns})_i^2\right]$$
$$= \frac{bT - aT - 2M(T) + 1}{bT - aT}\,\mathrm{tr}\,(\Gamma_{n,0}) < \mathrm{tr}\,(\Gamma_{n,0})$$

for all T. It follows that, for $T \ge T_1(n, \delta, \frac{1}{2})$,

$$E_{n,\delta}^{T*} \le 2\delta^2 E\left[\int_{-\pi}^{\pi} \bar{\lambda}_{n1}^{T*}(\theta)\, d\theta\right] \le 2\delta^2\, \mathrm{tr}\,(\Gamma_{n,0}).$$

Summing up, for any $\epsilon, \eta, \delta > 0$, we have shown that

$$P[|\underline{\mathbf{K}}_{ni}^{T_i}(L)\mathbf{x}_{nt} - \chi_{it}| > \epsilon] \le \frac{\eta}{2} + \frac{\eta}{4} + \frac{\eta}{8}$$
$$+ \frac{16}{\epsilon^2}\left(\frac{\epsilon^2\eta}{256} + \delta^2(M^*(n) + 2\,\mathrm{tr}\,(\Gamma_{n,0}))\right)$$

provided that $n \ge N_0(\epsilon, \eta)$ and $T \ge T_0^* = T_0^*(n, \epsilon, \eta, \delta)$, where

$$T_0^* = \max\left(T_1\left(n, \delta, \frac{\eta}{8}\right), T_1\left(n, \delta, \frac{1}{2}\right), T_2(n, \epsilon, \eta),\right.$$
$$\left. T_3\left(n, \delta, \frac{\epsilon^2\eta}{256}\right), T_4^*(n, \delta), T_5^*(n, \delta)\right).$$

Proposition (3) follows, with $T_0(n, \epsilon, \eta) = T_0^*(n, \epsilon, \eta, \delta)$, letting

$$\delta^2 = \delta^2(n, \epsilon, \eta) = \frac{\epsilon^2\eta}{256(M^*(n) + 2\,\mathrm{tr}\,(\Gamma_{n,0}))}.$$

QED