

SFB
823

Factor models in high- dimensional time series A time-domain approach

Marc Hallin, Marco Lippi

Nr. 10/2013

Discussion Paper



FACTOR MODELS IN HIGH-DIMENSIONAL TIME SERIES

A TIME-DOMAIN APPROACH

Marc HALLIN^{1*} and Marco LIPPI^{2†}

¹ECARES, Université Libre de Bruxelles and ORFE, Princeton University

²Università di Roma La Sapienza and EIEF, Rome

March 13, 2013

Abstract

High-dimensional time series may well be the most common type of dataset in the so-called “big data” revolution, and have entered current practice in many areas, including meteorology, genomics, chemometrics, connectomics, complex physics simulations, biological and environmental research, finance and econometrics. The analysis of such datasets poses significant challenges, both from a statistical as from a numerical point of view. The most successful procedures so far have been based on dimension reduction techniques and, more particularly, on high-dimensional factor models. Those models have been developed, essentially, within time series econometrics, and deserve being better known in other areas. In this paper, we provide an original time-domain presentation of the methodological foundations of those models (dynamic factor models usually are described via a spectral approach), contrasting such concepts as commonality and idiosyncrasy, factors and common shocks, dynamic and static principal components. That time-domain approach emphasizes the fact that, contrary to the *static factor models* favored by practitioners, the so-called *general dynamic factor model* essentially does not impose any constraints on the data-generating process, but follows from a general representation result.

1 Introduction.

The analysis of high-dimensional data in the past few years has become one of the most active subjects of modern statistical methodology. The reason is that information increasingly often takes the form of T observations with values in n -dimensional real spaces, where n and T both are quite large, often of the same magnitude, possibly with n much larger than T . It is well known that “traditional” T -asymptotics (where T tends to infinity under fixed n) yield poor or misleading results in such context, and that “double” (n, T) -asymptotics (where both n and T tend to infinity) provide more sensible solutions. The literature in the area is huge, and develops at a fast pace: even a brief account of it is impossible in the limits of this short note, and we refer the reader to recent surveys by Bai and Ng (2008) and Stock and Watson (2011) (both focused on *static* factor models—in a sense to be defined below).

*Supported by the Sonderforschungsbereich “Statistical modelling of nonlinear dynamic processes” (SFB 823) of the Deutsche Forschungsgemeinschaft, a Discovery Grant of the Australian Research Council, and an IAP of the Belgian federal Government. E-mail address: <mhallin ulb.ac.be>.

†E-mail address: <ml@lippi.ws>.

All domains of applications, basically, are affected with the developments of such high-dimensional asymptotics: genetics, chemometrics, environmental studies, image analysis, finance, econometrics ... Some are dealing with sampling models, where observations are assumed to be independent and identically distributed, often multivariate Gaussian. In most cases, however, serial dependence is the rule, and the time series aspects of the problem cannot be ignored.

Such situations are quite standard in econometrics, where observations often take the form of an $n \times T$ double-indexed array $\{X_{it}; i = 1, \dots, n, t = 1, \dots, T\}$ of observed random variables, where i is a cross-sectional index and t stands for time. Such an array is called a *panel*; each row in the panel is a univariate time series of length T , while each column can be considered as the observation, at time t , of a n -dimensional time series. The dimension n , in econometric applications, is often of the order of several hundreds, even one thousand. All those series exhibit complex (lagged) cross-correlations, a sensible parametrization of which is infeasible (even the simplest VAR(1) model would involve n^2 autoregression parameters). Traditional multivariate time series methods for such datasets, thus, are totally helpless, and econometricians have developed a variety of methods for coping with high-dimensionality issues. The most successful ones are based on *factor model* techniques, which consist in decomposing the observation X_{it} into $X_{it} = \chi_{it} + \xi_{it}$, the sum of two mutually orthogonal unobserved components—the *common component* χ_{it} and the *idiosyncratic component* ξ_{it} . Various characterizations of *idiosyncrasy* lead to various *factor models*, *exact* or *approximate*, *static* or *dynamic*, etc.

Since their introduction some twelve years ago, factor model methods have been quite successful in the analysis of large panels of econometric data, and have entered daily practice in most national statistical institutes, central banks and business cycle analysis institutions. Whether exact or approximate, static or dynamic (the terminology unfortunately is not entirely fixed; in particular, the word “dynamic” is used quite loosely in the literature), most factor models have the nature of *statistical models*, in the sense that they put restrictions on the underlying data-generating process. As usual, those restrictions may be satisfied, or lead to good approximations, in which case the model is a good model. They also may be dangerously misleading when they do not hold. And, as a rule, they hardly can be checked from the observations.

There is, however, an important exception to that rule: the so-called *general* or *generalized dynamic factor model* introduced in Forni et al. (2000), of which all other factor models (exact or approximate, static or dynamic, ...), under the assumption of second-order stationarity, are particular cases. Beyond very mild and general structural assumptions (such as stationarity), indeed, that “model”, as we shall see, does not place any restriction on the data-generating process; as such, it constitutes a *canonical representation* of the stochastic process under study rather than a *statistical model*. Our purpose here is to emphasize that fact by providing (Section 2) a very general time-domain presentation of the concepts leading to the definition of the general dynamic factor model and the representation result establishing its existence and uniqueness without requiring the existence of a spectrum. Then, in Section 3, under the additional assumption of an absolutely continuous spectral measure, we establish the connection with Brillinger’s spectral concepts of dynamic eigenvalues and eigenvectors, and the more usual frequency domain definition.

Let us conclude this section by pointing out that we only consider here factor models under the assumption of stationarity of the observable process \mathbf{X} , which entails stationarity of the common and idiosyncratic components as well. The non-stationary cases, the I(1) case in particular has not yet been explored systematically (important papers in that direction are Bai and Ng 2004, Bai 2004). A local stationary approach also has been taken in Motta et al. (2011).

2 Factor models

2.1 Panel data.

An $n \times T$ panel is a finite realization

$$\begin{array}{cccc} X_{11}, & X_{12}, & \dots, & X_{1T} \\ \vdots & \vdots & & \vdots \\ X_{n1}, & X_{n2}, & \dots, & X_{nT} \end{array}$$

of a double-indexed stochastic process of the form

$$\mathbf{X} := \{X_{it} | i \in \mathbb{N}, t \in \mathbb{Z}\},$$

hence, a collection of n observed time series of length T , related to n individuals or “cross-sectional items”, or, equivalently, one single time series in dimension n . Throughout, we denote by $\mathbf{X}_t^{(n)}$ the n -dimensional vector $(X_{1t}^{(n)}, \dots, X_{nt}^{(n)})'$, by \mathbf{X}_t the fixed- t collection $\{X_{it} | i \in \mathbb{Z}\}$, by $\mathbf{X}^{(n)}$ the n -dimensional process $\{X_{it} | i \in \{1, \dots, n\}, t \in \mathbb{Z}\}$, and by \mathbf{X} the whole double-indexed process $\{X_{it} | i \in \mathbb{N}, t \in \mathbb{Z}\}$. The following assumption will be made throughout.

ASSUMPTION A1(i). The process \mathbf{X} is second-order time-stationary, that is, for all i, i', i'', t and k , the variances $\text{Var}(X_{it})$ and covariances $\text{Cov}(X_{i't}X_{i'',t-k})$ exist, are finite, and do not depend on t .

For simplicity, we henceforth also assume that all X_{it} 's are centered and, in order to avoid trivialities, nondegenerate:

ASSUMPTION A1(ii). For all $i \in \mathbb{N}$ and $t \in \mathbb{Z}$, $E[X_{it}] = 0$ and $0 < E[X_{it}^2]$.

Let Assumption A1 hold, and denote by $\mathcal{H}^{\mathbf{X}}$ the Hilbert space spanned by \mathbf{X} , equipped with the L_2 covariance scalar product, that is, the set of all L_2 -convergent linear combinations of X_{it} 's and limits of L_2 -convergent sequences thereof. Similarly, we use the notation $\mathcal{H}_t^{\mathbf{X}}$, $\mathcal{H}^{\mathbf{X}^{(n)}}$, and $\mathcal{H}_t^{\mathbf{X}^{(n)}}$ for the subspaces of $\mathcal{H}^{\mathbf{X}}$ spanned by $\{X_{is} | i \in \mathbb{N}, s \leq t\}$, $\{X_{is} | 1 \leq i \leq n, t \in \mathbb{Z}\}$, and $\{X_{is} | 1 \leq i \leq n, s \leq t\}$, respectively.

Let $\eta_0 := \sum_{i=1}^{\infty} \sum_{s=-\infty}^{\infty} a_{is} X_{is} \in \mathcal{H}^{\mathbf{X}}$. Then, $\eta_t := \sum_{i=1}^{\infty} \sum_{s=-\infty}^{\infty} a_{i,s+t} X_{i,s+t} \in \mathcal{H}^{\mathbf{X}}$ for all $t \in \mathbb{Z}$, and we say that the process $\boldsymbol{\eta} := \{\eta_t | t \in \mathbb{Z}\}$ belongs to $\mathcal{H}^{\mathbf{X}}$; note that the processes $\boldsymbol{\eta}$ and \mathbf{X} then are *costationary*.

2.2 Common versus idiosyncratic.

A panel is not a “natural” object, though, but an artificial construction. The n time series constituting the panel indeed have been put together by someone, who did it on purpose—usually, for the reason that those series all carry, or are expected to carry, some information about some unobservable feature or latent process of interest. That unobserved common feature, in general, is the most relevant issue of the analysis. However, its exact relation to the observed X_{it} prior to the analysis is not known, and “commonness” in general is the only way through which that feature is identified. Examples are the *business cycle*, which is *common* to all variables describing an economy, but remains otherwise undefined; the *market liquidity*, which is *common* to a market-wide panel of liquidity measurements, but remains otherwise undefined; etc. As a consequence, although that common feature is present in (almost) all individual series in the

panel, it is not identified on the basis of any single one of them, and a large n is essential. Variables orthogonal to all “common variables” will be called “idiosyncratic”.

Note that the cross-sectional ordering of the panel, which in principle, is arbitrary, should play no role in the characterization of “commonness” and “idiosyncrasy” given below. Sensible concepts and sensible statistical procedures therefore should be invariant under permutation of cross-sectional items.

On the other hand, all individuals or cross-sectional items are exposed, in general, to the influence of the same covariables, all of which cannot be recorded. This induces complex interrelations that are not statistically tractable, or that would involve uncomfortably many parameters. Parametric methods, as a rule, are helpless or quite unrealistic.

2.3 Factor models

Factor model concepts, as we shall see, are tailor-made in this context.

When applied to a process of the form $\{X_{it}|i \in \mathbb{N}, t \in \mathbb{Z}\}$, factor model methods aim at identifying a decomposition of X_{it} into two mutually orthogonal (at all leads and lags) parts

$$X_{it} = \chi_{it} + \xi_{it}, \quad \text{here to be interpreted as “common”}_{it} + \text{“idiosyncratic”}_{it}, \quad i \in \mathbb{N}, t \in \mathbb{Z}. \quad (2.1)$$

Such decompositions are quite standard in statistics, and of course crucially depend on what is imposed on χ_{it} in order to be “common”, on ξ_{it} in order to be “idiosyncratic”. The terminology, moreover, is all but unified, and a large variety of definitions of a factor model can be found in the literature, depending on the characterization of “common” and “idiosyncratic”. If “idiosyncratic” is understood as synonym of “white noise” (as in Lam and Yao 2012), χ_{it} has to account for all auto- and cross-correlations in the panel, and (2.1) yields a decomposition of the “(exogenous) signal plus noise” type, which does not convey the meaning of “commonness” it is expected to focus on. Similarly, decompositions of the type “low dimension plus negligible error”, as in dimension reduction methods, or “reduced rank signal plus sparse residuals” as in high-dimensional signal processing (see Fan et al. 2013 for a recent reference), fail to provide an adequate mathematical translation of the intuitive idea of commonness, which is central to the present context.

If latent variables are common to a subset of series in the panel, these variables should account for the *cross*-correlations within that subset of series—not necessarily for their *auto*-correlations. As the terminology suggests, common components thus should account for “panel-wide cross-correlations”, induced by the panel-wide impact of unobserved latent variables or, equivalently, their innovations, the *common shocks*. On the contrary, idiosyncratic components are expected to be item-specific.

One therefore might be tempted to call idiosyncratic those processes $\{\zeta_{it}\}$ in $\mathcal{H}^{\mathbf{X}}$ that do not exhibit *any* cross-correlation at all: ζ_{it} and $\zeta_{i',t-k}$ mutually orthogonal for all $i' \neq i$ and all $k \in \mathbb{Z}$. Imposing such a condition leads to the so-called *exact* or *strong factor model* considered, for instance, by Sargent and Sims (1997) and Geweke (1997). That requirement, however, is too restrictive for most practical purpose. It is unpleasantly sensitive, in particular, to the possible presence in the panel of two closely related series: if, for instance, X_{2t} is of the form $X_{2t} = a(L)X_{1t}$ for some linear filter $a(L)$, it automatically gets treated as fully “common”, although it could be strictly orthogonal to X_{it} for all t and $i > 2$. The requirement that the idiosyncratic components ξ_{it} be cross-sectionally *strictly* orthogonal to each other at all leads and lags therefore is profitably weakened into a milder requirement of “limited cross-correlation”, yielding an *approximate* or *weak factor model*.

In order to introduce a more precise definition of that idea of “mild cross-correlation”, let us consider two examples of extreme idiosyncrasy/extreme commonness. Let

$$X_{it} = \phi_{it} + \psi_{it}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z} \quad \text{with} \quad \phi_{it} = \phi_t \text{ i.i.d. } \mathcal{N}(0, \sigma_\phi^2), \quad \psi_{it} \text{ i.i.d. } \mathcal{N}(0, \sigma_\psi^2)$$

with ϕ_t and ψ_{t-k} orthogonal for all i, t , and k . Clearly, $\phi_{it} = \lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n X_{jt}$ and $\psi_{it} = X_{it} - \lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n X_{jt}$, where convergence holds in quadratic mean, so that the processes $\{\phi_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ and $\{\psi_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ both are in \mathcal{H} . Since it has no cross-correlations at all, $\{\psi_{it}\}$ is an example of extreme idiosyncrasy, while $\{\phi_{it} = \phi_t, i \in \mathbb{N}, t \in \mathbb{Z}\}$, where the same ϕ_t appears in all cross-sectional items, clearly should qualify as an example of extreme “commonness”.

Now consider arbitrary *normed* (i.e., with coefficients satisfying $\sum_{i=1}^n \sum_{k=-\infty}^{\infty} (a_{ik}^{(n)})^2 = 1$ for all n) sequences of linear combinations $\sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_{ik}^{(n)} \psi_{i,t-k}$ of the ψ 's: their variances are

$$\sum_{i=1}^n \sum_{k=-\infty}^{\infty} (a_{ik}^{(n)})^2 \sigma_\psi^2 = \sigma_\psi^2.$$

It follows that, for fixed n , the maximum, over all normed linear combinations, of those variances is σ_ψ^2 , which remains bounded as $n \rightarrow \infty$. It is easy to check that this behavior is not affected if autocorrelation is introduced among the ψ_{it} 's; nor is it affected by the presence of *mild* cross-correlation—letting, for instance, $\text{Cov}(\psi_{i,t}, \psi_{j,t}) = c_{ij} I[i \text{ odd and } j = i + 1]$ —only affecting finite numbers of cross-sectional items.

The situation is entirely different for the linear combinations

$$w^{(n)} := \sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_{ik}^{(n)} \phi_{i,t-k} = \sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_{ik}^{(n)} \phi_{t-k}$$

involving the ϕ 's. Choosing, for instance, $a_{ik}^{(n)} = a_k^{(n)}$ ($a_k^{(n)}$'s that do not depend on i), $w^{(n)}$ has variance

$$\sigma_{w^{(n)}}^2 = \sum_{k=-\infty}^{\infty} \left(\sum_{i=1}^n a_{ik}^{(n)} \right)^2 \sigma_\phi^2 = \sum_{k=-\infty}^{\infty} (n a_k^{(n)})^2 \sigma_\phi^2 = n \sigma_\phi^2.$$

It immediately follows that the fixed- n maximum, over all normed linear combinations, of these variances, tends to infinity as $n \rightarrow \infty$. The limit in quadratic mean of $w^{(n)}$ thus does not exist. However, the limit in quadratic mean $\lim_{n \rightarrow \infty} w_0^{(n)}$ of the sequence of *standardized* versions

$$w_0^{(n)} := w^{(n)} / \sigma_{w^{(n)}} = n^{-1/2} w^{(n)} / \sigma_\phi$$

of the same $w^{(n)}$'s is a well-behaved standard normal variable of $\mathcal{H}^{\mathbf{X}}$. Moreover, letting

$$w_{\mathbf{X}}^{(n)} := \sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_{ik}^{(n)} X_{i,t-k} \quad \text{and} \quad w_{\mathbf{X}0}^{(n)} := w_{\mathbf{X}}^{(n)} / \text{Var}^{1/2}(w_{\mathbf{X}}^{(n)}),$$

the variance of $w_{\mathbf{X}}^{(n)}$ similarly explodes, while, due to the boundedness as $n \rightarrow \infty$ of the variance of $\sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_k^{(n)} \psi_{i,t-k}$, the difference between $w_{\mathbf{X}0}^{(n)}$ and $w_0^{(n)}$ goes to zero in quadratic mean.

We therefore propose the following definitions.

Definition 2.1 A random variable ζ in $\mathcal{H}^{\mathbf{X}}$, with variance $0 < \sigma_\zeta^2$, is called *common* if ζ/σ_ζ is the limit in quadratic mean of a sequence of standardized elements of $\mathcal{H}^{\mathbf{X}}$, of the form $w_{\mathbf{X}}^{(n)}/(\text{Var}(w_{\mathbf{X}}^{(n)}))^{1/2}$, where $w_{\mathbf{X}}^{(n)} := \sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_{ik}^{(n)} X_{i,t-k}$, with $\sum_{i=1}^n \sum_{k=-\infty}^{\infty} (a_{ik}^{(n)})^2 = 1$, is such that $\lim_{n \rightarrow \infty} \text{Var}(w_{\mathbf{X}}^{(n)}) = \infty$.

Definition 2.2 Define the Hilbert space $\mathcal{H}_{\text{com}}^{\mathbf{X}}$ spanned by the collection of all common variables in $\mathcal{H}^{\mathbf{X}}$ and its orthogonal complement (with respect to $\mathcal{H}^{\mathbf{X}}$) $\mathcal{H}_{\text{idio}}^{\mathbf{X}} := (\mathcal{H}_{\text{com}}^{\mathbf{X}})^\perp$ as \mathbf{X} 's *common* and *idiosyncratic spaces*, respectively. The process \mathbf{X} is called *purely common* if $\mathcal{H}^{\mathbf{X}} = \mathcal{H}_{\text{com}}^{\mathbf{X}}$ (hence, $\mathcal{H}_{\text{idio}}^{\mathbf{X}} = \{0\}$), *purely idiosyncratic* if $\mathcal{H}^{\mathbf{X}} = \mathcal{H}_{\text{idio}}^{\mathbf{X}}$ (hence, $\mathcal{H}_{\text{com}}^{\mathbf{X}} = \{0\}$).

It may be considered desirable that these definitions, and the related existence and uniqueness results below, remain invariant under cross-sectional scale transformations (multiplying each process $\{X_{it}\}$ with a positive constant c_i). Such invariance is achieved if Assumption (A1)(ii) is reinforced into

ASSUMPTION A1(ii') For all $i, t \in \mathbb{Z}$, $E[X_{it}] = 0$, and there exist two constants $0 < C^- < C^+$ such that, for all $i, t \in \mathbb{Z}$, $C^- \leq E[X_{it}^2] \leq C^+$.

This reinforcement has little practical impact (if any), as it does not imply any constraint on the fixed- n dataset at hand; see the comments after Assumption A3. All results below remain valid with A1(ii') substituted for A1(ii).

2.4 The general dynamic factor model

The above definitions then lead to the following extremely general representation result.

Theorem 2.1 Under Assumption A1(i)-(ii), there exist two uniquely defined mutually orthogonal processes $\boldsymbol{\chi} = \{\chi_{it}\}$ and $\boldsymbol{\xi} = \{\xi_{it}\}$ in $\mathcal{H}_{\text{com}}^{\mathbf{X}}$ and $\mathcal{H}_{\text{idio}}^{\mathbf{X}}$, respectively, such that

$$X_{it} = \chi_{it} + \xi_{it} \quad i \in \mathbb{N}, t \in \mathbb{Z}. \quad (2.2)$$

PROOF. The existence and uniqueness of decomposition (2.2) follow from the fact that $\mathcal{H}_{\text{com}}^{\mathbf{X}}$ and $\mathcal{H}_{\text{idio}}^{\mathbf{X}}$ by definition provide a decomposition of $\mathcal{H}^{\mathbf{X}}$ into a subspace and its orthogonal complement; χ_{it} and ξ_{it} then are easily obtained by projecting X_{it} onto those two subspaces. \square

By construction, $\boldsymbol{\chi}$ is purely common, $\boldsymbol{\xi}$ purely idiosyncratic. Our definition of the general dynamic factor model is based on this representation (2.2) of \mathbf{X} as the sum of a purely common and a purely idiosyncratic part.

Definition 2.3 The decomposition (2.2) is called the *general dynamic factor model representation* of \mathbf{X} —in short, a *general dynamic factor model* for \mathbf{X} .

A general dynamic factor model (Forni et al. (2001) and Forni and Lippi (2000) use the terminology *generalized* factor model) thus belongs to the class of “weak” or “approximate” factor models (as first proposed, in a *static* form, by Chamberlain (1983) and Chamberlain and Rothschild (1983)), as opposed to the “strong” or “exact” ones. The price to be paid is that the decomposition (2.2) is only asymptotically identified—since the characterization of an idiosyncratic quantity itself is of an asymptotic nature—a feature which is not uncommon in the literature on high-dimensional data (cf the concept of *sparsity*). As announced in the introduction, the existence and uniqueness of that general dynamic factor model representation of \mathbf{X} does not require (beyond second-order stationarity) any constraint on the data-generating process. Therefore, it does not constitute a statistical model in the usual sense.

2.5 Common shocks

The definition 2.3 of the general dynamic factor model is quite general and allows for extremely weird things. For example, consider the process defined by

$$X_{it} = u_{t+i-1} + \xi_{it} \quad i \in \mathbb{N}, t \in \mathbb{Z}$$

where u_t is scalar white noise, and the variables ξ_{it} are i.i.d. with finite variance, with u_t orthogonal to ξ_{is} for all t, i and s . In this case, each of the variables χ_{it} can be predicted without error using \mathbf{X}_{t-1} . Precisely, $\chi_{it} = \chi_{i+1, t-1}$, so that the infinite-dimensional vector \mathbf{X}_t is *purely deterministic* ($\mathcal{H}_t^{\mathbf{X}} = \mathcal{H}^u$ for all t). An apparently less pathological case is

$$X_{it} = b(L)u_t + \xi_{it} \quad i \in \mathbb{N}, t \in \mathbb{Z}$$

where $b(L)$ be a band-pass filter such that $|b(e^{-i\theta})|$ is strictly positive for $|\theta| < 1$ and vanishes for $|\theta| \geq 1$, with u_t and the variables ξ_{it} as in the previous example, so that $\chi_{it} = b(L)u_t$ for all i . It follows from Kolmogorov's formula (Theorem 5.8.1 in Brockwell and Davis (2009)) that $b(L)u_t$ also is deterministic, and hence has variance innovation zero. We believe that the analysis of such cases is of no interest from the point of view of statistics and its applications.

Further, Definition 2.3 is not a constructive one, and its relation to the frequency-domain approach in Forni et al. (2000) and Forni and Lippi (2001), or the static definitions in Stock and Watson (2002a and b), Bai and Ng (2002), Forni et al. (2009) and many others, is not clear. If that link is to be established, an additional assumption on the complexity of \mathbf{X} 's common space is required. *Reduced rank* common spaces, namely, common spaces driven by a finite number of *common shocks*, will play a major role in the analysis.

Definition 2.4 Let $\boldsymbol{\zeta} := \{\zeta_{it}\}$ denote a double-indexed process in \mathcal{H} . That process is said to have *dynamic dimension* $d \geq 1$ if there exists a purely non deterministic d -dimensional vector process $\{\mathbf{Y}_t := (Y_{1t}, \dots, Y_{dt})' | t \in \mathbb{Z}\}$ with full-rank d -dimensional innovation such that $\mathcal{H}_t^{\boldsymbol{\zeta}} = \mathcal{H}_t^{\mathbf{Y}}$ for all t . We say that $\boldsymbol{\zeta}$ has dynamic dimension $d = 0$ if $\zeta_{it} = 0$ a.s. for all i and t ; if no $d \in \mathbb{N}$ exists such that $\boldsymbol{\zeta}$ has dynamic dimension d , we say that $\boldsymbol{\zeta}$'s dynamic dimension is infinity.

A common component process \mathbf{X} with infinite dynamic dimension is possible, though. Here is an (unavoidably, pathological) example. Consider a partition of \mathbb{N} into an infinite number of infinite subsets. More precisely, let $N_k \subset \mathbb{N}$, $k \in \mathbb{N}$ be a collection of subsets of \mathbb{N} such that (a) $\#N_k$ is infinite for all k , (b) $N_k \cap N_h = \emptyset$ for all $h \neq k$, and (c) $N_0 \cup N_1 \cup \dots = \mathbb{N}$. Such partitions exist: take, for instance, $N_j := \{n(n/1)/2 | n \geq j\}$, yielding $N_0 = \{1, 3, 6, 10, 15, 21, 28, \dots\}$, $N_1 = \{2, 4, 7, 11, 16, 22, 29, \dots\}$, $N_2 = \{5, 8, 12, 17, 23, 30, 38, \dots\}$, etc. Define

$$a_{ik} := \begin{cases} 1 & \text{if } i \in N_k \\ 0 & \text{if } i \notin N_k \end{cases}$$

From part (b) of the definition of the sets N_k ,

$$\sum_{i=1}^n a_{ik} a_{ih} = 0, \quad \text{for all } n \text{ and } h \neq k, \quad (2.3)$$

while (a) implies

$$\sum_{i=1}^n a_{ik}^2 = \sum_{i=1}^n a_{ik} = \#(N_k \cap \{1, 2, \dots, n\}) \rightarrow \infty \quad \text{for all } k, \text{ as } n \rightarrow \infty. \quad (2.4)$$

Next, define

$$X_{it} := \sum_{k=1}^{\infty} a_{ik} u_{kt} + \xi_{it}$$

where $\{u_{1t}\}, \{u_{2t}\}, \dots$ and $\{\xi_{1t}\}, \{\xi_{2t}\}, \dots$ all are mutually independent i.i.d. $\mathcal{N}(0, 1)$ white noises, so that, in view of the definitions of a_{ik} and the sets N_k , $\text{Var}(X_{it}) = 2$ for all i and t . Consider the normalized linear combination

$$\begin{aligned} \left(\sum_{i=1}^n a_{ik} \right)^{-1/2} \sum_{i=1}^n a_{ik} X_{it} &= \left(\sum_{i=1}^n a_{ik} \right)^{-1/2} \left(\sum_{i=1}^n a_{ik} u_{kt} + \sum_{i=1}^n a_{ik} \xi_{it} \right) \\ &= \left(\sum_{i=1}^n a_{ik} \right)^{1/2} u_{kt} + \left(\sum_{i=1}^n a_{ik} \right)^{-1/2} \sum_{i=1}^n a_{ik} \xi_{it} \end{aligned} \quad (2.5)$$

(in view of (2.3)), with variance $\sum_{i=1}^n a_{ik} + 1$ tending to infinity due to (2.4). Therefore, the standardized version of the same linear combination (2.5) takes the form

$$\left(\sum_{i=1}^n a_{ik} / \left(\sum_{i=1}^n a_{ik} + 1 \right) \right)^{1/2} u_{kt} + \sum_{i=1}^n a_{ik} \xi_{it} \left(\sum_{i=1}^n a_{ik} \left(\sum_{i=1}^n a_{ik} + 1 \right) \right)^{-1/2} = w_1^{(n)} + w_2^{(n)}, \quad \text{say.}$$

Still in view of (2.4), we have that $w_1^{(n)}$ tends to u_{kt} and $w_2^{(n)}$ to zero in quadratic mean, as $n \rightarrow \infty$. Hence, the standardized version of (2.5) also converges in mean square to u_{kt} . Since this holds for any $k \in \mathbb{N}$, the dynamic dimension of the common space in this example is not finite.

In order to avoid this, we make the following assumption.

ASSUMPTION A2. The common component process $\boldsymbol{\chi}$ has finite dynamic dimension $q \in \mathbb{N}$ (q otherwise unspecified).

Assumption A2 essentially rules out the “weird” cases described above, while ensuring, as the following result shows, the existence of a finite number of *common shocks*. Ruling out such cases is all but unreasonable. The objective, indeed, when considering n -asymptotics, is not to provide the description of a non-observed “cross-sectional future” associated with growing values of n . The objective, by letting n tend to infinity, is to provide a good and hopefully simple asymptotic approximation to the observed, finite- n situation at hand. A common component with infinite dynamic dimension corresponds means a situation where new common shocks, orthogonal to the previous ones, keep on “entering” the space of the common component as n grows. This way of adding new common shocks, that have no impact on the actual finite- n observation, but hypothetically would pop up “someday”, in the “cross-sectional future”, under a growing- n scenario, is quite unlikely to provide a good approximation for the finite- n dataset at hand. Ruling out the $q = \infty$ “catastrophe scenario” thus does not really restrict the applicability of general dynamic factor methods.

Under Assumption A2, Theorem 2.1 can be reinforced as follows.

Theorem 2.2 *Under Assumptions A1 and A2, there exist a q -tuple $\{\mathbf{U}_t\} = \{(U_{1t}, \dots, U_{qt})'\}$ of mutually orthogonal white noises (namely, $\text{Var}(U_{jt}) = 1$ and $\text{Cov}(U_{jt}, U_{j't'}) = 0$ unless $j' = j$ and $t' = t$), and a collection of one-sided square-summable filters $B_{ij}(L)$, $i \in \mathbb{N}$, $j = 1, \dots, q$ such that*

$$\chi_{it} = \sum_{j=1}^q B_{ij}(L) U_{jt} \quad i \in \mathbb{N}, t \in \mathbb{Z} \quad (2.6)$$

and $\mathcal{H}_t^{\mathbf{U}} = \mathcal{H}_t^{\mathbf{X}}$ for all $t \in \mathbb{Z}$.

PROOF. If Assumption A2 holds, there exists a purely nondeterministic q -dimensional vector process $\{\mathbf{Y}_t\}$ in $\mathcal{H}^{\mathbf{X}}$, with full-rank q -dimensional innovation $\{\mathbf{V}_t\}$, such that $\mathcal{H}_t^{\mathbf{X}} = \mathcal{H}_t^{\mathbf{Y}}$. Denote by $\Sigma^{\mathbf{V}}$ the full-rank covariance of $\{\mathbf{V}_t\}$. Since $\{\chi_{it}\}$ is purely non-deterministic, $\mathcal{H}_t^{\mathbf{Y}} = \mathcal{H}_t^{\mathbf{X}}$. Denoting by $\{\mathbf{U}_t\} := \{(\Sigma^{\mathbf{V}})^{-1/2}\mathbf{V}_{it}\}$ the standardized version of the innovation $\{\mathbf{V}_t\}$, the components $\{U_{jt}\}$, $j = 1, \dots, q$ of $\{\mathbf{U}_t\}$ constitute a q -tuple of mutually orthogonal white noises such that $\mathcal{H}_t^{\mathbf{U}} = \mathcal{H}_t^{\mathbf{Y}}$, hence $\mathcal{H}_t^{\mathbf{U}} = \mathcal{H}_t^{\mathbf{X}}$, and (2.6) is component i of χ_t 's Wold decomposition (not to be mistaken for χ_{it} 's Wold decomposition). \square

Although no *factors* appear in the representation (2.2) of \mathbf{X} , the q white noises $\{U_{jt}\}$, $j = 1, \dots, q$ in (2.6) can be interpreted as *common shocks*, loaded, with lags, by X_{it} via the one-sided filters $B_{ij}(L)$, $\in \mathbb{N}$, $j = 1, \dots, q$. These common shocks remain largely undetermined, but the space they are spanning at time t is uniquely characterized as the *innovation space* of the common component χ .

Theorem 3.2, however, still is a purely theoretical existence result, which does not provide much constructive information on the common and idiosyncratic components $\{\chi_{it}\}$ and $\{\xi_{it}\}$ of a given process $\{X_{it}\}$. No relation is provided with the more familiar frequency-domain definitions of Forni et al. (2000) and Forni and Lippi (2001). Nor does Theorem 3.2 point at any statistical way of consistently reconstructing the common and idiosyncratic components from a finitely observed $\{X_{it}\}$. More insight into this will be provided in the next section, via a link to Brillinger's theory of *dynamic principal components*.

3 Principal components

3.1 Static principal components and static factor models

The most common way of turning a decomposition of the form (2.2) into a statistically tractable model (also satisfying the assumptions of Theorem 3.2) consists in imposing on the common component χ_{it} a simple linear structure, such as

$$\begin{cases} \chi_{it} &= b_{i1}F_{1t} + \dots + b_{ir}F_{rt} \\ \mathbf{F}_t &:= (F_{1t}, \dots, F_{rt})' = \mathbf{A}_1\mathbf{F}_{t-1} + \dots + \mathbf{A}_p\mathbf{F}_{t-p} + \mathbf{R}\mathbf{U}_t \quad i \in \mathbb{N}, t \in \mathbb{Z} \end{cases} \quad (3.1)$$

where $\mathbf{U}_t := (U_{1t}, \dots, U_{qt})'$ is a q -tuple of white noises, the $r \times r$ matrices $\mathbf{A}_1, \dots, \mathbf{A}_p$ define some stationary VAR, and the $r \times q$ matrix \mathbf{R} has rank q . Here, X_{it} depends in a static way (instantaneous dependence: loading constants instead of loading filters) on the unobserved r -dimensional latent process $\{\mathbf{F}_t\}$: the VAR(p) dynamics of the r factors (F_{1t}, \dots, F_{rt}) , driven by the q -dimensional *common shocks* \mathbf{U}_t ($r \geq q$), are to account for the dynamics of the common components throughout the whole panel. Clearly, not everything is identifiable in (3.1), and there is room for a variety of identification constraints; without those, in particular, only the space spanned by the factors is identified.

Denote by

$$\mathbf{A}_r^{(n)} := \begin{pmatrix} \lambda_1^{(n)} & \dots & 0 \\ & \ddots & \\ 0 & \dots & \lambda_r^{(n)} \end{pmatrix} \quad \text{and} \quad \mathbf{P}_r^{(n)} := \begin{pmatrix} \mathbf{P}_1^{(n)'} \\ \vdots \\ \mathbf{P}_r^{(n)'} \end{pmatrix} := \begin{pmatrix} p_{11}^{(n)} & \dots & p_{1n}^{(n)} \\ \vdots & \ddots & \vdots \\ p_{r1}^{(n)} & \dots & p_{rn}^{(n)} \end{pmatrix}$$

the $r \times r$ diagonal matrix containing the r largest eigenvalues of the $n \times n$ covariance matrix $E[\mathbf{X}_t^{(n)} \mathbf{X}_t^{(n)'}]$ and the corresponding $r \times n$ matrix of row eigenvectors. Then $(\mathbf{\Lambda}_r^{(n)})^{-1/2} \mathbf{P}_r^{(n)} \mathbf{X}_t^{(n)}$ is the standardized projection of $\mathbf{X}_t^{(n)}$ onto the r -dimensional space spanned by $\mathbf{X}_t^{(n)}$'s r first principal components. It can be shown, under very general assumptions (see Bai and Ng (2002) or Stock and Watson (2002a and b)), that those projections converge, as both n and T tend to infinity, to the space spanned by \mathbf{F}_t , the only identified feature of the factors in equation (3.1). That consistency property of static principal components is the basis for many methods that have been proposed in the literature (same references as above, and many others, as Forni et al. (2009) and . Other methods, such as Gaussian quasi-maximum likelihood (pretending that the idiosyncratic components are i.i.d. Gaussian), or Kalman filtering also have been considered: see Bai and Li (2012), De Mol et al. (2008), Doz et al. (2011, 2012), Fan et al. (2013).

Model (3.1) is the most refined of *static* factor models for time series in high dimension—the word *static* here does not mean that there is no dynamics in the common component, but that all common dynamics features are loaded in a static way from a finite number of factors. Similarly, the concept of principal components used in this context is the classic, static concept, based on the spectral factorization of the instantaneous covariance matrix $E[\mathbf{X}_t^{(n)} \mathbf{X}_t^{(n)'}]$. This is fine, since loadings are instantaneous, hence reflected in $E[\mathbf{X}_t^{(n)} \mathbf{X}_t^{(n)'}]$. Contrary to the general dynamic factor model (2.2)-(2.6), (3.1) is a statistical model, putting severe restrictions on the data-generating process. Those restrictions at first sight may look quite innocuous, but they are not. Consider, for instance, the very simple case under which $q = 1$ (one single common shock in (2.6)) and the common component χ_{it} satisfies the elementary AR(1) loading scheme

$$\chi_{it} = \rho_i \chi_{i,t-1} + U_{1t} \quad t \in \mathbb{Z}, \rho_i \in (-1, 1), i \in \mathbb{N}.$$

This very simple case does not admit a finite- r static representation of the form (3.1), as each lag of U_{1t} has to be counted as one distinct factor. And the static principal component projections would not consistently reconstruct the (infinite-dimensional) space spanned by the common factors. Static factor models have an indisputable advantage of conceptual simplicity, which explains their success among practitioners. But, just as classical principal components, with which they are associated, are not the adequate principal component concept, static factor models are not the adequate factor model concept in a time series context.

3.2 Brillinger's concept of dynamic principal components

Going back to the definition of a common process, the maximal variance normalized linear combinations involved there have a strong flavor of principal components—with, however, an essential difference: the linear combinations $\sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_{ik} \zeta_{i,t-k}$ taken into consideration are running over all leads and lags, whereas traditional principal components would maximize instantaneous cross-sectional linear combinations of the form $\sum_{i=1}^n a_i \zeta_{i,t}$. That difference, as we now explain, parallels the difference between *dynamic* principal components and the traditional *static* ones. While static principal components are perfectly adapted to the analysis of i.i.d. data, the dynamic ones are much more relevant in a time series context.

The problem with traditional (static) principal components in a time series context is that serial dependencies are overlooked. A static principal component $\sum_{i=1}^n a_i \zeta_{i,t}$ associated with a small static eigenvalue may have a negligible instantaneous impact on ζ_t . But the same linear combination may have a high covariance with ζ_{t+1} , hence a high predictive value: discarding it then results in a significant loss of information. As a result, static principal components in general

do not provide any reasonable solution to the dimension reduction problem in the presence of serial dependence. Besides this conceptual failure, static principal components, when computed from serially dependent observations, also run into technical problems: while cross-sectionally uncorrelated at fixed time t , they typically still exhibit lagged cross-correlations. Therefore, unlike in the i.i.d. case, static principal components cannot be analyzed componentwise, and still need to be treated as vector time series.

The concept of dynamic principal component was introduced by Brillinger (1981), where we refer to for details, in order to overcome these weaknesses of the classical concept.

Dynamic principal components are based on a factorization of spectral density matrices. Therefore, let us make the additional assumption that all $\{X_t, t \in \mathbb{Z}\}$ processes in the panel admit spectral densities. More precisely, we assume the following (see pages 19-20 of Rozanov (1967)).

ASSUMPTION A3. For all $n \in \mathbb{N}$, the spectral measure of $\{\mathbf{X}_t^{(n)}, t \in \mathbb{Z}\}$ is absolutely continuous with respect to the Lebesgue measure on $[-\pi, \pi]$, that is, $\{\mathbf{X}_t^{(n)}\}$ has a spectral density matrix $\boldsymbol{\Sigma}^{(n)}(\theta)$, with entries $(\sigma_{ij}(\theta))$, $\theta \in [-\pi, \pi]$.

Note that the matrices $\boldsymbol{\Sigma}^{(n)}(\theta)$ are nested for all θ (so that $\sigma_{ij}(\theta)$ needs no n superscript), Hermitian and positive semidefinite. Also, spectral densities always are defined up to a set of θ values with Lebesgue measure zero; rather than functions, we are dealing with equivalence classes of a.e. equal functions, thus. By $\boldsymbol{\Sigma}^{(n)}(\theta)$, in the sequel, we tacitly mean a representative of such a class; the same comment applies to $\boldsymbol{\Sigma}^{(n)}(\theta)$'s eigenvalues and eigenvectors.

Since each $\boldsymbol{\Sigma}^{(n)}(\theta)$ is Hermitian positive semidefinite, it has n nonnegative eigenvalues, associated with n eigenvectors

$$\lambda_1^{(n)}(\theta) \geq \lambda_2^{(n)}(\theta) \geq \dots \geq \lambda_n^{(n)}(\theta) \quad \text{and} \quad \mathbf{p}_1^{(n)}(\theta), \mathbf{p}_2^{(n)}(\theta), \dots, \mathbf{p}_n^{(n)}(\theta);$$

call them the *dynamic* eigenvalues and eigenvectors of $\{\mathbf{X}_t^{(n)}\}$, respectively. As eigenvectors, the $\mathbf{p}_j^{(n)}(\theta)$'s are subject to the usual sign indeterminacy: $-\mathbf{p}_j^{(n)}(\theta)$ qualifies as an eigenvector as well as $\mathbf{p}_j^{(n)}(\theta)$. Talking about “the” dynamic eigenvector $\mathbf{p}_j^{(n)}(\theta)$ is thus an abuse of language and notation, since an arbitrary sign can be selected for each value of θ . Adequate choices can be made (see Forni and Lippi, 2001, Section 2) so that the measurability claim in Part (d) of Proposition 3.1 holds.

From their definition as eigenvalues and eigenvectors, dynamic eigenvalues and eigenvectors, for all $\theta \in [-\pi, \pi]$, enjoy the following elementary properties. Write \mathbf{P}^* for the adjoint (transposed convex conjugate) of any complex-valued matrix or vector \mathbf{P} .

Proposition 3.1 *With the above notation,*

$$(a) \quad \boldsymbol{\Sigma}^{(n)}(\theta) \mathbf{p}_j^{(n)}(\theta) = \lambda_j^{(n)}(\theta) \mathbf{p}_j^{(n)}(\theta), \quad j = 1, \dots, n, \quad \theta \in (-\pi, \pi);$$

$$(b) \quad \|\mathbf{p}_j^{(n)}(\theta)\|^2 = \mathbf{p}_j^{(n)*}(\theta) \mathbf{p}_j^{(n)}(\theta) = 1, \quad j = 1, \dots, n, \quad \text{and} \quad \mathbf{p}_j^{(n)*}(\theta) \mathbf{p}_{j'}^{(n)}(\theta) = 0 \quad \text{for} \quad j \neq j' \\ = 1, \dots, n \quad \text{and} \quad \theta \in (-\pi, \pi), \quad \text{so that the } n \times n \text{ matrix } \mathbf{P}^{(n)}(\theta) := (\mathbf{p}_1^{(n)}(\theta), \dots, \mathbf{p}_n^{(n)}(\theta)) \text{ is a} \\ \text{unitary matrix } (\mathbf{P}^{(n)*}(\theta) \mathbf{P}^{(n)}(\theta) = \mathbf{I}_n = \mathbf{P}^{(n)}(\theta) \mathbf{P}^{(n)*}(\theta));$$

$$(c) \quad \lambda_j^{(n)}(\theta) = \max_{\mathbf{p}} \mathbf{p}^* \boldsymbol{\Sigma}^{(n)}(\theta) \mathbf{p}, \quad \text{where the maximum is taken over } \mathbf{p} \in \mathbb{C}^n \text{ such that } \|\mathbf{p}\| = 1 \\ \text{for } j = 1, \text{ such that } \mathbf{p}^* \mathbf{p}_{j'}^{(n)}(\theta) = 0, \quad j' \in \{1, \dots, j-1\} \text{ for } j \geq 2, \quad \theta \in (-\pi, \pi);$$

(d) for all j and θ , $\lambda_j^{(n)}(\theta)$ is monotone nondecreasing in n ;

(e) $\theta \mapsto \lambda_j^{(n)}(\theta)$ are measurable functions on $[-\pi, \pi]$; $\theta \mapsto \mathbf{p}_j^{(n)}(\theta)$ are measurable (see above) and, being bounded, integrable with respect to the Lebesgue measure over $[-\pi, \pi]$.

We refer to Section 2.2 of Forni and Lippi (2001) for details and further properties.

Any matrix or vector $\mathbf{M}(\theta)$ with square-integrable θ -measurable elements defined over $[-\pi, \pi]$ can be expanded (componentwise) into a Fourier series

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

where the right-hand side converges in quadratic mean. That expansion creates a correspondence between the square-integrable matrix-valued function $\mathbf{M}(\theta)$ and the square-summable filter

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

(L , as usual, stands for the *lag operator*); note that $\mathbf{M}(\theta) = \underline{\mathbf{M}}(e^{-i\theta})$. In view of Proposition 2.1(d), the dynamic eigenvectors $\theta \mapsto \mathbf{p}_j^{(n)}(\theta)$, in particular, can be expanded into the Fourier series

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

defining square-summable filters of the form

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

such that $\mathbf{p}_j^{(n)}(\theta) = \underline{\mathbf{p}}_j^{(n)}(e^{-i\theta})$.

The $(m \times n)$ -dimensional function $\theta \mapsto \mathbf{M}(\theta)$ and the filter $\underline{\mathbf{M}}(L)$ moreover are strongly connected by the fact that, if the n -dimensional process $\{\mathbf{X}_t^{(n)}\}$ has spectral density matrix $\boldsymbol{\Sigma}^{(n)}(\theta)$, then the m -variate stochastic process $\{\underline{\mathbf{M}}(L)\mathbf{X}_t^{(n)}\}$ has spectral density matrix

$$\mathbf{M}(\theta)\boldsymbol{\Sigma}^{(n)}(\theta)\mathbf{M}^*(\theta) = \underline{\mathbf{M}}(e^{-i\theta})\boldsymbol{\Sigma}^{(n)}(\theta)\underline{\mathbf{M}}'(e^{i\theta}).$$

It follows that the univariate process $\{W_{jt}^{(n)} \mid t \in \mathbb{Z}\}$, where $W_{jt}^{(n)} := \underline{\mathbf{p}}_j^{(n)'}(L)\mathbf{X}_t^{(n)}$, has spectral density $\mathbf{p}_j^{(n)'}(\theta)\boldsymbol{\Sigma}^{(n)}(\theta)\underline{\mathbf{p}}_j^{(n)}(\theta) = \lambda_j^{(n)}(\theta)$.

Definition 3.1 The univariate process $\{W_{jt}^{(n)} := \underline{\mathbf{p}}_j^{(n)'}(L)\mathbf{X}_t^{(n)} \mid t \in \mathbb{Z}\}$ is called $\mathbf{X}^{(n)}$'s j th *dynamic principal component* ($j = 1, \dots, n$).

As eigenvectors, the $\mathbf{p}_j^{(n)}(\theta)$'s are subject to the usual sign indeterminacy: $-\mathbf{p}_j^{(n)}(\theta)$ qualifies as an eigenvector as well as $\mathbf{p}_j^{(n)}(\theta)$. Talking about “the” dynamic eigenvector $\mathbf{p}_j^{(n)}(\theta)$ is thus an abuse of language and notation, since an arbitrary sign can be selected for each value of θ . The same remark holds for the dynamic principal components. That multiplicity is not a problem, however, in the present context, as the Hilbert space spanned by $\{W_j^{(n)}\}$ (or any collection of $\{W_j^{(n)}\}$'s) is well defined irrespective of the selection of signs.

The properties of $\mathbf{X}^{(n)}$'s dynamic principal components extend to the time-series context the standard properties of traditional principal components associated with the eigenvalues and eigenvectors of $\mathbf{X}^{(n)}$'s covariance matrix $E[\mathbf{X}^{(n)}\mathbf{X}^{(n)'}]$. Here are some of them.

Proposition 3.2 (a) *The n -dimensional process $\{\mathbf{W}_t^{(n)} := (W_{1t}^{(n)}, \dots, W_{nt}^{(n)})'\}$ has diagonal spectral density matrix, with diagonal elements $\lambda_1^{(n)}(\theta), \dots, \lambda_n^{(n)}(\theta)$; hence, the univariate process $\{W_{jt}^{(n)}\}$ has spectral density $\lambda_j^{(n)}(\theta)$, while $\{W_{jt}^{(n)}\}$ and $\{W_{j't}^{(n)}\}$, $j \neq j'$, are mutually orthogonal (at all leads and lags);*

(b) $W_{1t}^{(n)}, \dots, W_{nt}^{(n)}$ therefore constitute an orthogonal basis of $\mathcal{H}^{\mathbf{X}^{(n)}}$;

(c) *the variance of $W_{jt}^{(n)}$ is $\lambda_j^{(n)} := \int_{-\pi}^{\pi} \lambda_j^{(n)}(\theta) d\theta$, and*

$$\lambda_j^{(n)} = \begin{cases} \max\{a_{ik}^{(n)} \mid \sum_{i=1}^n \sum_{k=-\infty}^{\infty} (a_{ik}^{(n)})^2 = 1\} \text{Var}\left(\sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_{ik} X_{i,t-k}\right), & j = 1 \\ \max\{a_{ik}^{(n)} \mid \sum_{i=1}^n \sum_{k=-\infty}^{\infty} (a_{ik}^{(n)})^2 = 1\} \text{Var}\left(\sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_{ik} X_{i,t-k}\right), & j = 2, \dots, n. \\ \text{subject to } \sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_{ik} X_{i,t-k} \text{ orthogonal to } W_{1t}^{(n)}, \dots, W_{j-1,t}^{(n)} \end{cases}$$

It readily follows from Part (c) of Proposition 3.2 that a strong relation exists between idiosyncrasy, the asymptotic behavior of dynamic eigenvalues, and the dynamic principal components of a process.

Theorem 3.1 *A process $\{\zeta_{it}\}$ is purely idiosyncratic if and only if all its dynamic eigenvalues are (equivalently, its first dynamic eigenvalue is) θ -a.e. bounded as $n \rightarrow \infty$.*

This establishes the equivalence, under Assumption A3, of the time-domain definition of a purely idiosyncratic process we are giving in Section 2 and the frequency-domain definition of Forni and Lippi (2001).

3.3 Dynamic principal components and dynamic factor models

Throughout this section, we assume that Assumptions A1 and A3 hold. In view of Theorem 3.1, if $\lambda_1^{(n)}$, hence all the dynamic eigenvalues $\lambda_j^{(n)}$ of $\mathbf{X}^{(n)}$, are θ -a.e. bounded as $n \rightarrow \infty$, then \mathbf{X} is a purely idiosyncratic panel; Assumption A2 holds, but with $q = 0$. The orthogonal complement of the idiosyncratic space $\mathcal{H}_{\text{idio}}^{\mathbf{X}}$ then reduces to $\mathcal{H}_{\text{com}}^{\mathbf{X}} = \{0\}$, and the factor model decomposition (2.2) to a trivial one $X_{it} = 0 + X_{it}$. Such panels are of little interest, as they do not carry any ‘‘common information’’, hence are missing their objective: putting them together does not bring any essential improvement to the analysis and prediction of any of the individual series in the cross-section.

More generally, Assumption A2 holds with $q > 0$. Theorem 2.2 then tells us that the common components are driven by a q -tuple of white noises, and that (2.6) holds. In practice, q is not specified, and consistent identification methods have been proposed in the literature: Hallin and Liška (2007) and, in the static case, Amengual and Watson (2007) and Alessi et al. (2010), improving over the pioneering result by Bai and Ng (2002).

Assumption A2, however, has even stronger consequences on the asymptotic behavior of the dynamic eigenvalues $\lambda_1^{(n)}, \dots, \lambda_q^{(n)}$, when they exist, and their relation to (2.6).

Proposition 3.3 *Let Assumptions A1, A2, and A3 hold. Then,*

- (i) *the q first dynamic eigenvalues of $\{\boldsymbol{\chi}_t^{(n)}\}$ diverge to ∞ , and all other ones are equal to zero, θ -a.e. in $[-\pi, \pi]$, as $n \rightarrow \infty$;*
- (ii) *the q first dynamic eigenvalues of $\{\mathbf{X}_t^{(n)}\}$ diverge to ∞ , and all other ones are bounded, θ -a.e. in $[-\pi, \pi]$, as $n \rightarrow \infty$.*

PROOF. Theorem 2.1 implies that q mutually orthogonal white noises $\{U_{jt}\}$ are spanning the common space. It follows from the definition of the common space that each of those white noises is the limit in quadratic mean of a sequence $U_{\mathbf{X};jt}^{(n)} := \text{Var}^{-1/2}(u_{\mathbf{X};jt}^{(n)})u_{\mathbf{X};jt}^{(n)}$ of standardized versions of normalized linear combinations of the present, past, and future observations X_{it} , of the form

$$u_{\mathbf{X};jt}^{(n)} := \sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_{j;ik}^{(n)} X_{i,t-k} := \mathbf{a}_j^{(n)'}(L)\mathbf{X}_t^{(n)},$$

where

$$\mathbf{a}_j^{(n)'}(L) := \left(\sum_{k=-\infty}^{\infty} a_{j;1k}^{(n)} L^k, \dots, \sum_{k=-\infty}^{\infty} a_{j;nk}^{(n)} L^k \right), \quad \sum_{i=1}^n \sum_{k=-\infty}^{\infty} (a_{j;ik}^{(n)})^2 = 1,$$

and $\lim_{n \rightarrow \infty} \text{Var}(u_{\mathbf{X};jt}^{(n)}) = \infty$. Owing to the fact that the white noises $\{U_{jt}\}$ are mutually orthogonal, those q normalized linear combinations, for n large enough, are linearly independent.

Since $\{\xi_{it}\}$ is purely idiosyncratic, $\sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_{j;ik}^{(n)} \xi_{i,t-k}$ has bounded variance. Hence, U_{jt} is also the limit in quadratic mean of $U_{jt}^{(n)} := \text{Var}^{-1/2}(u_{jt}^{(n)})u_{jt}^{(n)}$, with

$$u_{jt}^{(n)} := \sum_{i=1}^n \sum_{k=-\infty}^{\infty} a_{j;ik}^{(n)} \chi_{i,t-k} = \mathbf{a}_j^{(n)'}(L)\boldsymbol{\chi}_t^{(n)},$$

where $\varsigma_j^{(n)} := \text{Var}^{1/2}(u_{jt}^{(n)})$ also explodes as $n \rightarrow \infty$.

Let $\mathbf{U}_t^{(n)} := (U_{1t}^{(n)}, \dots, U_{qt}^{(n)})'$, $\mathbf{u}_t^{(n)} := (u_{1t}^{(n)}, \dots, u_{qt}^{(n)})'$, $\mathbf{a}^{(n)}(L) := (\mathbf{a}_1^{(n)}(L), \dots, \mathbf{a}_q^{(n)}(L))'$: then, $\mathbf{u}_t^{(n)} = \mathbf{a}^{(n)}(L)\boldsymbol{\chi}_t^{(n)}$. Denote by $\boldsymbol{\Sigma}_{\mathbf{U}}^{(n)}(\theta)$, $\boldsymbol{\Sigma}_{\mathbf{u}}^{(n)}(\theta)$, and $\boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(n)}(\theta)$ the spectral density matrices associated with $\mathbf{U}_t^{(n)}$, $\mathbf{u}_t^{(n)}$, and $\boldsymbol{\chi}_t^{(n)}$, respectively, and let $\boldsymbol{\varsigma}^{(n)} := \text{diag}(\varsigma_1^{(n)}, \dots, \varsigma_q^{(n)})$ be the diagonal matrix with j th diagonal entry $\varsigma_j^{(n)}$. Quadratic mean convergence of a sequence of variables implies L_1 (with respect to Lebesgue on $[-\pi, \pi]$) componentwise convergence of their spectral densities to the spectral density of the limiting variable. That L_1 convergence in turn implies a.e. convergence of a subsequence (see, for instance, pp. 297-298 of [3]). Therefore, θ -a.e. in $[-\pi, \pi]$, the spectral density matrices $\boldsymbol{\Sigma}_{\mathbf{U}}^{(n)}(\theta)$ of the $\mathbf{U}_t^{(n)}$'s are converging to that, $\mathbf{I}_q/2\pi$, of q -dimensional white noise:

$$\boldsymbol{\Sigma}_{\mathbf{U}}^{(n)}(\theta) = \mathbf{I}_q/2\pi + o(1) \quad \text{as } n \rightarrow \infty \quad (3.1)$$

(the $o(1)$ term may depend on θ).

On the other hand, we have

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{U}}^{(n)}(\theta) &= (\boldsymbol{\varsigma}^{(n)})^{-1} \boldsymbol{\Sigma}_{\mathbf{u}}^{(n)}(\theta) (\boldsymbol{\varsigma}^{(n)})^{-1} = (\boldsymbol{\varsigma}^{(n)})^{-1} \mathbf{a}^{(n)}(e^{-i\theta}) \boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(n)}(\theta) \mathbf{a}^{(n)'}(e^{i\theta}) (\boldsymbol{\varsigma}^{(n)})^{-1} \\ &= (\boldsymbol{\varsigma}^{(n)})^{-1} \mathbf{a}^{(n)}(e^{-i\theta}) \mathbf{P}_{\boldsymbol{\chi}}^{(n)'}(\theta) \boldsymbol{\Lambda}_{\boldsymbol{\chi}}^{(n)}(\theta) \bar{\mathbf{P}}_{\boldsymbol{\chi}}^{(n)}(\theta) \mathbf{a}^{(n)'}(e^{i\theta}) (\boldsymbol{\varsigma}^{(n)})^{-1} \\ &=: (\boldsymbol{\varsigma}^{(n)})^{-1} \mathbf{b}^{(n)}(e^{-i\theta}) \boldsymbol{\Lambda}_{\boldsymbol{\chi}}^{(n)}(\theta) \mathbf{b}^{(n)'}(e^{i\theta}) (\boldsymbol{\varsigma}^{(n)})^{-1}, \quad \text{say,} \end{aligned} \quad (3.2)$$

where $\mathbf{P}_{\boldsymbol{\chi}}^{(n)'}(\theta)\boldsymbol{\Lambda}_{\boldsymbol{\chi}}^{(n)}(\theta)\bar{\mathbf{P}}_{\boldsymbol{\chi}}^{(n)}(\theta)$ is $\boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(n)}(\theta)$'s spectral decomposition. The $n \times n$ matrix $\mathbf{P}_{\boldsymbol{\chi}}^{(n)}(\theta)$ is a unitary matrix ($\mathbf{P}_{\boldsymbol{\chi}}^{(n)}(\theta)\mathbf{P}_{\boldsymbol{\chi}}^{(n)*}(\theta) = \mathbf{I}_n = \mathbf{P}_{\boldsymbol{\chi}}^{(n)*}(\theta)\mathbf{P}_{\boldsymbol{\chi}}^{(n)}(\theta)$); since $\boldsymbol{\chi}_t^{(n)}$, being driven by q white noises, has reduced rank q , $\boldsymbol{\Lambda}_{\boldsymbol{\chi}}^{(n)}(\theta)$ is of the form $\text{diag}(\lambda_{\boldsymbol{\chi};1}^{(n)}(\theta), \dots, \lambda_{\boldsymbol{\chi};q}^{(n)}(\theta), 0, \dots, 0)$, and the last $(n - q)$ columns of $\mathbf{P}_{\boldsymbol{\chi}}^{(n)}(\theta)$ are largely indeterminate—an indetermination that carries over to $\mathbf{b}^{(n)}(e^{-i\theta})$ but has no impact on this proof. Because $\mathbf{P}_{\boldsymbol{\chi}}^{(n)}(\theta)$ is unitary, it preserves the norm, and, denoting by $\mathbf{b}_j^{(n)}(e^{-i\theta}) = \mathbf{P}_{\boldsymbol{\chi}}^{(n)}(\theta)\mathbf{a}_j^{(n)}(e^{-i\theta})$ the j th column of $\mathbf{b}^{(n)'}(e^{-i\theta})$, we have, for $j = 1, \dots, q$,

$$\mathbf{b}_j^{(n)'}(e^{-i\theta})\mathbf{b}_j^{(n)}(e^{i\theta}) = \|\mathbf{b}_j^{(n)}(e^{-i\theta})\|^2 = \|\mathbf{a}_j^{(n)}(e^{-i\theta})\|^2.$$

It follows that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \|\mathbf{b}_j^{(n)}(e^{-i\theta})\|^2 d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \|\mathbf{a}_j^{(n)}(e^{-i\theta})\|^2 d\theta = \sum_{i=1}^n \sum_{k=-\infty}^{\infty} (a_{j;ik}^{(n)})^2 = 1, \quad (3.3)$$

so that the set of θ values such that $\limsup_{n \rightarrow \infty} \|\mathbf{b}_j^{(n)}(e^{-i\theta})\| = \infty$ has Lebesgue measure zero.

Putting (3.1) and (3.2) together, we obtain that, θ -a.e. and, possibly, along a subsequence,

$$(\boldsymbol{\zeta}^{(n)})^{-1} \mathbf{b}^{(n)'}(e^{-i\theta}) \boldsymbol{\Lambda}_{\boldsymbol{\chi}}^{(n)}(\theta) \mathbf{b}^{(n)}(e^{i\theta}) (\boldsymbol{\zeta}^{(n)})^{-1} = \mathbf{I}_q / 2\pi + o(1) \quad \text{as } n \rightarrow \infty. \quad (3.4)$$

Now, the columns $\mathbf{b}_j^{(n)}(e^{-i\theta})$ of $\mathbf{b}^{(n)}(e^{-i\theta})$, for n large enough, are linearly independent, and, in view of (3.3), have θ -a.e. bounded modulus. Since $(\boldsymbol{\zeta}^{(n)})^{-1}$ is a $q \times q$ diagonal matrix converging to zero, (3.4), at given θ , is possible only if the q nonzero diagonal elements $\lambda_{\boldsymbol{\chi};1}^{(n)}(\theta), \dots, \lambda_{\boldsymbol{\chi};q}^{(n)}(\theta)$ of $\boldsymbol{\Lambda}_{\boldsymbol{\chi}}^{(n)}(\theta)$ all tend to infinity. Although (3.4), hence also this divergence of $\lambda_{\boldsymbol{\chi};q}^{(n)}(\theta)$, may be restricted to a subsequence, the fact that a dynamic eigenvalue $\lambda_{\boldsymbol{\chi};j}^{(n)}(\theta)$ tends to infinity along a subsequence implies that it tends to infinity in the plain sense, as it is monotonically increasing with n . This takes care of Part (i) of the proposition.

Turning to Part (ii), it is obvious from mutual orthogonality of the common and idiosyncratic components that the spectral density of $\mathbf{X}^{(n)}$ decomposes into $\boldsymbol{\Sigma}^{(n)}(\theta) = \boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(n)}(\theta) + \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(n)}(\theta)$. The claim then readily follows from Weyl's classical inequality, which here takes the form

$$\lambda_{\boldsymbol{\chi};j}^{(n)}(\theta) + \lambda_{\boldsymbol{\xi};n}^{(n)}(\theta) \leq \lambda_j^{(n)}(\theta) \leq \lambda_{\boldsymbol{\chi};j}^{(n)}(\theta) + \lambda_{\boldsymbol{\xi};1}^{(n)}(\theta). \quad \square$$

This Proposition establishes a time-domain alternative to the usual frequency-domain definition of dynamic factors. The main result of Forni and Lippi (2001) indeed can be formulated as follows.

Theorem 3.2 (Forni and Lippi 2001). *Under Assumptions A1 and A3, the following statements are equivalent.*

(a) *The components X_{it} of the panel \mathbf{X} admit a unique decomposition*

$$X_{it} = \chi_{it} + \xi_{it} \quad i \in \mathbb{N}, t \in \mathbb{Z} \quad (3.5)$$

where $\{\chi_{it}\}$ and $\{\xi_{it}\}$ are mutually orthogonal at all leads and lags, $\{\xi_{it}\}$ only has θ -a.e. bounded dynamic eigenvalues, and $\{\chi_{it}\}$ is driven by a q -tuple of white noises, that is, is of the form (2.6) for some q -tuple $\{\mathbf{U}_t = (U_{1t}, \dots, U_{qt})'\}$ of mutually orthogonal white noises, and some collection of bilateral square-summable filters $B_{ij}(L)$, $i \in \mathbb{N}$, $j = 1, \dots, q$. Moreover, denoting by $\lambda_{\boldsymbol{\chi};q}^{(n)}(\theta)$ the q th dynamic eigenvalue of $\{\boldsymbol{\chi}_t^{(n)}\}$, $\lambda_{\boldsymbol{\chi};q}^{(n)}(\theta) \rightarrow \infty$ a.e. in $[-\pi, \pi]$ as $n \rightarrow \infty$.

(b) The q first dynamic eigenvalues of $\{\mathbf{X}_t^{(n)}\}$ diverge to ∞ , and all other ones are bounded, θ -a.e. in $[-\pi, \pi]$, as $n \rightarrow \infty$.

The statement in Part (a) of this claim constitutes what Forni and Lippi (2001), Forni et al. (2000) and their subsequent papers take as a definition of the general(ized) dynamic factor model, while the statement in Part (b) of course coincides with that of Part (ii) of Proposition 3.3.

The fact that Assumptions A1, A2, and A3 jointly imply the more classical assumptions on the divergence of dynamic eigenvalues thus brings to our time-domain approach all the benefits of the statistical results that have been obtained under Assumptions A1, A3 and (a), in the spectral approach. In particular, the results that have been established in Forni et al. (2000) still hold, among which the following constructive characterization of the common component.

Denote by $\chi_{it}^{(n)}$ the projection of X_{it} onto the space spanned by the first q dynamic components $W_{1t}^{(n)}, \dots, W_{qt}^{(n)}$. That projection takes the form $\chi_{it}^{(n)} := \mathbf{K}_i^{(n)}(L)\mathbf{X}_t^{(n)}$, where

$$\mathbf{K}_i^{(n)}(\theta) := p_{1,i}^{(n)*}(\theta)\mathbf{p}_1^{(n)'}(\theta) + \dots + p_{q,i}^{(n)*}(\theta)\mathbf{p}_q^{(n)'}(\theta) \quad \theta \in [-\pi, \pi].$$

Theorem 3.3 (Forni, Hallin, Lippi, and Reichlin 2000). *Under Assumptions A1, A3 and (a), hence also (b), of Theorem 3.2, and, therefore, also under Assumptions A1, A2, and A3 of Proposition 3.3,*

$$\chi_{it} = \chi_{it}^{(n)} + o_{\mathbb{P}}(1) \quad \text{as } n \rightarrow \infty, \quad i \in \mathbb{N}, \quad t \in \mathbb{N}. \quad (3.6)$$

This fundamental convergence result, which serves as the basis of all statistical applications of the general dynamic factor model, now holds for the time-domain definition provided here. Note, however, that Assumptions A1-A3, which imply one-sided representations of the common components in terms of common shocks, are more restrictive than Assumptions A1, A3 and (a), which only guarantee the existence of possibly two-sided representations (on this last issue, see Forni and Lippi (2011)).

4 Conclusions

In this paper, we provide a unified time-domain presentation of the methodological foundations of general dynamic factor models for high-dimensional time series, emphasizing the generality of the approach which, contrary to its static counterparts, relies on a canonical representation result, hence only imposes very general constraints on the data-generating process. All factor models in the literature, whether dynamic or static, under the assumption of second-order stationarity, can be obtained by imposing further modeling assumptions on that canonical representation, justifying the terminology *general dynamic factor model*.

Acknowledgement. The authors are grateful to Rainer Dahlhaus for his constructive editorial comments on a previous version of this manuscript, which led to considerable improvements, and for his patience regarding submission deadlines. Marc Hallin also thanks Manfred Deistler and the participants in a March 2013 course at the Institute for Advanced Studies in Vienna for their questions and remarks, which helped clarifying several points in the exposition.

References

- [1] Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters* 80, 1806-1813.
- [2] Amengual, D. and M. Watson (2007). Consistent estimation of the number of dynamic factors in large n and T panel. *Journal of Business and Economic Statistics* 25, 91-96.
- [3] Apostol, T.M. (1974). *Mathematical Analysis*, Second Edition, Addison-Wesley Publishing Company, Reading, Massachusetts.
- [4] Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics* 122, 137-183.
- [5] Bai, J. and K. Li (2012). Statistical analysis of factor models of high dimension. *Annals of Statistics* 40, 436-465.
- [6] Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70, 191-221.
- [7] Bai, J. and S. Ng (2004). A PANIC attack on unit roots and cointegration. *Econometrica* 72, 1127-1177.
- [8] Bai, J. and S. Ng (2008). *Large dimensional factor analysis. Foundations and Trends in Econometrics* 3, Hanover, Mass. : Now Publishers.
- [9] Brillinger, D.R. (1981). *Time Series Data Analysis and Theory*, Holden Day, San Francisco.
- [10] Brockwell, P.J. and R.A. Davis (2009). *Time Series: Theory and Methods*, Springer, New York.
- [11] Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica* 51, 1281-1304.
- [12] Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* 51, 1305-1324.
- [13] Fan, J., Y. Liao and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B*, forthcoming.
- [14] De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics* 146, 318-328.
- [15] Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics* 164, 188-205.
- [16] Doz, C., D. Giannone, and L. Reichlin (2012). A quasi-maximum likelihood approach for large approximate dynamic factor models. *Review of Economics and Statistics* 94, 1014-1024.

- [17] Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of Royal Statistical Society Series B*, to appear.
- [18] Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009). Opening the black box: structural factor models with large cross-sections. *Econometric Theory* 25, 1319-1347.
- [19] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic factor model: identification and estimation. *The Review of Economics and Statistics* 82, 540- 554.
- [20] Forni, M. and M. Lippi (2001). The generalized dynamic factor model: representation theory. *Econometric Theory* 17, 1113-1341.
- [21] Forni, M. and M. Lippi (2011). The generalized dynamic factor model: one-sided representation results. *Journal of Econometrics* 163, 23-28.
- [22] Geweke, J. (1977). The dynamic factor analysis of economic time series, in Aigner D.J. and Goldberger A.S., Eds., *Latent Variables in Socio-Economic Models*, North Holland, Amsterdam, pp. 365-383.
- [23] Hallin M. and R. Liška (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association* 102, 603- 617.
- [24] Lam, C. and Q. Yao (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *Annals of Statistics* 40, 694-726.
- [25] Lawley, D.N. and A.E. Maxwell (1971). *Factor Analysis as a Statistical Method*, 2nd Edition, Butterworths, London.
- [26] Motta, G., Chr. M. Hafner, and R. von Sachs (2011). Locally stationary factor models: identification and nonparametric estimation. *Econometric Theory* 27, 1279-1319.
- [27] Peña, D. and G.E.P. Box (1987). Identifying a simplifying structure in time series, *Journal of the American Statistical Association* 82, 836-843.
- [28] Rozanov, Yu. A. (1967). *Stationary Random Processes*. Holden-Day, San Francisco.
- [29] Sargent, T.J. and 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*, Federal Reserve Bank of Minneapolis, Minneapolis, pp. 45-109.
- [30] Stock, J.H. and M.W. Watson (2002a). Macroeconomic forecasting using diffusion indexes, *Journal of Business and Economic Statistics* 20, 147-162.
- [31] Stock, J.H. and M.W. Watson (2002b). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* 97, 1167-1179.
- [32] Stock, J.H. and M.W. Watson (2011). Dynamic factor analysis. In M.P. Clements and D.F. Hendry Eds, the *New Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford, pp. 35-60.

