Journal of Econometrics 163 (2011) 29–41

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Dynamic factors in the presence of blocks

Marc Hallin [a,b,c,d,e,*], Roman Liška [a,1]

[a] ECARES (European Centre for Advanced Research in Economics and Statistics), Université libre de Bruxelles, CP 114, B-1050 Bruxelles, Belgium
[b] ORFE, Princeton University, United States
[c] CentER, Tilburg University, Netherlands
[d] ECORE, Bruxelles and Louvain-la-Neuve, Belgium
[e] Académie Royale de Belgique

ARTICLE INFO

Article history:
Available online 11 November 2010

JEL classification:
C13
C33
C43

Keywords:
Panel data
High dimensional time series data
Dynamic factor model
Dynamic principal components

ABSTRACT

Macroeconometric data often come under the form of large panels of time series, themselves decomposing into smaller but still quite large subpanels or blocks. We show how the dynamic factor analysis method proposed in Forni et al. (2000), combined with the identification method of Hallin and Liška (2007), allows for identifying and estimating joint and block-specific common factors. This leads to a more sophisticated analysis of the structures of dynamic interrelations within and between the blocks in such datasets, along with an informative decomposition of explained variances. The method is illustrated with an analysis of a dataset of Industrial Production Indices for France, Germany, and Italy.

1. Introduction

1.1. Panel data and dynamic factor models

In many fields – macroeconometrics, finance, environmental sciences, chemometrics, … – information comes under the form of a large number of observed time series or *panel data*. Panel data consist of series of observations (length T) made on n individuals or "cross-sectional items" that have been put together on purpose, because, mainly, they carry some information about some common feature or unobservable process of interest, or are expected to do so. This "commonness" is a distinctive feature of panel data: mutually independent cross-sectional items, in that respect, do not constitute a panel (or then, a degenerate one). Cross-sectional heterogeneity is another distinctive feature of panels: n (possibly non-independent) replications of the same time series would be another form of degeneracy. Moreover, the impact of item-specific or *idiosyncratic* effects, which have the role of a nuisance, very often dominate, quantitatively, that of the common features one is interested in.

Finally, all individuals in a panel are exposed to the influence of unobservable or unrecorded covariates, which create complex interdependencies, both in the cross-sectional as in the time dimension, which cannot be modelled, as this would require questionable modelling assumptions and a prohibitive number of nuisance parameters. These interdependencies may affect all (or almost all) items in the panel, in which case they are "common"; they also may be specific to a small number of items, hence "idiosyncratic".

The idea of separating "common" and "idiosyncratic" effects is thus at the core of panel data analysis. The same idea is the cornerstone of factor analysis. There is little surprise, thus, to see a time series version of factor analysis emerging as a powerful tool in the context of panel data. This dynamic version of factor models, however, requires adequate definitions of "common" and "idiosyncratic". This definition should not simply allow for identifying the decomposition of the observation into a "common" component and an "idiosyncratic" one, but also should provide an adequate translation of the intuitive meanings of "common" and "idiosyncratic".

Denote by X_{it} the observation of item i ($i = 1, \ldots, n$) at time t ($t = 1, \ldots, T$); the factor model decomposition of this observation takes the form

$$X_{it} = \chi_{it} + \xi_{it}, \quad i = 1, \ldots, n, \; t = 1, \ldots, T,$$

* Corresponding author at: ECARES (European Centre for Advanced Research in Economics and Statistics), Université libre de Bruxelles, CP 114, B-1050 Bruxelles, Belgium.
E-mail address: marc.hallin@ulb.ac.be (M. Hallin).
[1] Present address: Institute for Health and Consumer Protection (IHCP), European Commission Joint Research Centre, I-21027 Ispra (VA), Italy.

0304-4076/$ – see front matter © 2011 Published by Elsevier B.V.
doi:10.1016/j.jeconom.2010.11.004

where the *common component* χ_{it} and an the idiosyncratic one ξ_{it} are mutually orthogonal (at all leads and lags) but unobservable. Some authors identify this decomposition by requiring the idiosyncratic components to be "small" or "negligible", as in dimension reduction techniques. Some others require that the n idiosyncratic processes be mutually orthogonal white noises. Such characterizations do not reflect the fundamental nature of factor models: idiosyncratic components indeed can be "large" and strongly autocorrelated, while white noise can be common. For instance, in a model of the form $X_{it} = \chi_t + \xi_{it}$, where χ_t is white noise and orthogonal to $\xi_{it} = \varepsilon_{it} + a_i\varepsilon_{i,t-1}$, with i.i.d. ε_{it}'s, the white noise component χ_t, which is present in all cross-sectional items, very much qualifies as being "common", while the cross-sectionally independent autocorrelated ξ_{it}'s, being item-specific, exhibit all the attributes one would like to see in an "idiosyncratic" component.

A possible characterization of commonness/idiosyncrasy is obtained by requiring the common component to account for all cross-sectional correlations, leading to possibly autocorrelated but cross-sectionally orthogonal idiosyncratic components. This yields the so-called "exact factor models" considered, for instance, by Sargent and Sims (1977) and Geweke (1977). These exact models, however, are too restrictive in most real life applications, where it often happens that two (or a small number of) cross-sectional items, being neighbours in some broad sense, exhibit cross-sectional correlation also in variables that are orthogonal, at all leads and lags, to all other observations throughout the panel. A "weak" or "approximate factor model", allowing for mildly cross-sectionally correlated idiosyncratic components, therefore also has been proposed (Chamberlain, 1983; Chamberlain and Rothschild, 1983), in which, however, the common and idiosyncratic components are only asymptotically (as $n \to \infty$) identified. Under its most general form, the characterization of idiosyncrasy, in this weak factor model, can be based on the behavior, as $n \to \infty$, of the eigenvalues of the spectral density matrices of the unobservable idiosyncratic components, but also (Forni and Lippi, 2001) on the asymptotic behavior of the eigenvalues of the spectral density matrices of the observations themselves: see Section 2 for details. This general characterization is the one we are adopting here.

Finally, once the common and idiosyncratic components are identified, two types of factor models can be found in the literature, depending on the way *factors* are driving the common components. In *static* factor models, it is assumed that common components are of the form

$$\chi_{it} = \sum_{l=1}^{q} b_{il} f_{lt}, \quad i = 1, \ldots, n, \ t = 1, \ldots, T, \tag{1}$$

that is, the χ_{it}'s are driven by q factors $f_{1t}, \ldots, f_{qt}$ which are loaded instantaneously. This static approach is the one adopted by Chamberlain (1983), Chamberlain and Rothschild (1983), Stock and Watson (1989, 2002a,b, 2005), Bai and Ng (2002, 2007), and a large number of applied studies. The so-called *general dynamic model* decomposes common components into

$$\chi_{it} = \sum_{l=1}^{q} b_{il}(L) u_{lt}, \quad i = 1, \ldots, n, \ t = 1, \ldots, T, \tag{2}$$

where $u_{1t}, \ldots, u_{qt}$, the unobservable *common shocks,* are loaded via one-sided linear filters $b_{il}(L)$. That "fully dynamic" approach (the terminology is not unified and the adjective "dynamic" is often used in an ambiguous way) goes back, under exact factor form, to Chamberlain (1983) and Chamberlain and Rothschild (1983), but was developed, mainly, by Forni et al. (2000, 2003, 2004, 2005, 2009); Forni and Lippi (2001) and Hallin and Liška (2007).

The static model (1) clearly is a particular case of the general dynamic one (2). Its main advantage is simplicity. On the other hand, both models share the same assumption on the asympotic behavior of spectral eigenvalues—the plausibility of which is confirmed by empirical evidence. But the static model (1) places an additional and rather severe restriction on the data-generating process, while the dynamic one (2), as shown by Forni and Lippi (2001), does not—we refer to Section 2 for details. Moreover, the synchronization of clocks and calendars across the panel is often quite approximative, so that the concept of "instantaneous loading" itself may be questionable.

Both the static and the general dynamic models are receiving increasing attention in finance and macroeconometric applications where information usually is scattered through a (very) large number n of interrelated time series (n values of the order of several hundreds, or even one thousand, are not uncommon). Classical multivariate time series techniques are totally helpless in the presence of such values of n, and factor model methods, to the best of our knowledge, are the only ones that can handle such datasets. In macroeconomics, factor models are used in business cycle analysis (Forni and Reichlin, 1998; Giannone et al., 2006), in the identification of economy-wide and global shocks, in the construction of indices and forecasts exploiting the information scattered in a huge number of interrelated series (Altissimo et al., 2001), in the monitoring of economic policy (Giannone et al., 2005), and in monetary policy applications (Bernanke and Boivin, 2003; Favero et al., 2005). In finance, factor models are at the heart of the extensions proposed by Chamberlain and Rothschild (1983) and Ingersol (1984) of the classical arbitrage pricing theory; they also have been considered in performance evaluation and risk measurement (Chapters 5 and 6 of Campbell et al., 1997), and in the statistical analysis of the structure of stock returns (Yao, 2008).

Factor models in the recent years also generated a huge amount of applied work: see d'Agostino and Giannone (2005), Artis et al. (2005), Bruneau et al. (2007), Den Reijer (2005), Dreger and Schumacher (2004), Schumacher (2007), Nieuwenhuyzen (2004), Schneider and Spitzner (2004), Giannone and Matheson (2007), and Stock and Watson (2002b) for applications to data from UK, France, the Netherlands, Germany, Belgium, Austria, New Zealand, and the US, respectively; Altissimo et al. (2001), Angelini et al. (2001), Forni et al. (2003), and Marcellino et al. (2003) for the Euro area and Aiolfi et al. (2006) for South American data—to quote only a few. Dynamic factor models also have entered the practice of a number of economic and financial institutions, including several central banks and national statistical offices, who are using them in their current analysis and prediction of economic activity. A real time coincident indicator of the EURO area business cycle (EuroCOIN), based on Forni et al. (2000), is published monthly by the London-based Center for Economic Policy Research and the Banca d'Italia: see (http://www.cepr.org/data/EuroCOIN/). A similar index, based on the same methods, is established for the US economy by the Federal Reserve Bank of Chicago.

1.2. Dynamic factor models in the presence of blocks: outline of the paper

Although heterogeneous, panel data very often are obtained by pooling together several "blocks" which themselves constitute "large" subpanels. In macroeconometrics, for instance, data typically are organized either by country or sectoral origin: the database which is used in the construction of EuroCOIN, the monthly indicator of the euro area business cycle published by CEPR, includes almost 1000 time series that cover six European countries and are organized into eleven subpanels including industrial production, producer prices, monetary aggregates, etc. Depending on the objectives of the analysis, such a panel could be divided into six blocks (one for each country), or into eleven blocks (one for each subpanel). When these blocks are large enough, several dynamic factor models can be considered and analyzed, allowing for a refined analysis of interblock relations. In the simple two-block case, "marginal common factors" can be defined for each block, and need not coincide with the "joint common factors" resulting from pooling the two blocks.

The objective of this paper is to provide a theoretical basis for that type of analysis. For simplicity, we start with the simple case of two blocks. We show (Section 2) how a factorization of the Hilbert space spanned by the n observed series leads to a decomposition of each of them into four mutually orthogonal (at all leads and lags) components: a *strongly idiosyncratic* one, a *strongly common* one, a *weakly common*, and a *weakly idiosyncratic* one. In Sections 3 and 4, we show how projections onto appropriate subspaces provide consistent data-driven reconstructions of those various components. Section 5 is devoted to the general case of $K \geq 2$ blocks, allowing for various decompositions of each observation into mutually orthogonal (at all leads and lags) components. The tools we are using throughout are Brillinger's theory of dynamic principal components, a key result (Proposition 2) by Forni et al. (2000), and the identification method developed by Hallin and Liška (2007). Proofs are concentrated in Appendix.

The potential of the method is briefly illustrated, in Section 6, with a panel of Industrial Production Index data for France and Germany ($K = 2$ blocks, $q = 3$ factors), then France, Germany and Italy ($K = 3$ blocks, $q = 4$ factors). Simple as it is, the analysis of that dataset reveals some striking facts. For instance, both Germany and Italy exhibit a "national common factor" which is idiosyncratic to the other two countries, while France's common factors are included in the space spanned by Germany's. The (estimated) percentages of explained variation associated with the various components also are quite illuminating: Germany, with 25% of common variation, is the "most common" out of the three countries. But it also is, with only 6.4% of its total variation, the "least strongly common" one. France has the highest proportion (82.4%) of marginal idiosyncratic variation but also the highest proportions of strongly and weakly idiosyncratic variations (72.6% and 9.8%, respectively). We do not attempt here to provide an economic interpretation for such facts. Nor do we apply the method to a more sophisticated dataset.[2] But we feel that the simple application we are proposing provides sufficient evidence of the potential power of the method, both from a structural as from a quantitative point of view.[3]

2. The dynamic factor model in the presence of blocks

We throughout assume that all stochastic variables in this paper belong to the Hilbert space $L_2(\Omega, \mathcal{F}, \mathrm{P})$, where $(\Omega, \mathcal{F}, \mathrm{P})$ is some given probability space. We will study two double-indexed sequences $\mathbf{Y} := \{Y_{it};\, i \in \mathbb{N}, t \in \mathbb{Z}\}$ and $\mathbf{Z} := \{Z_{jt};\, j \in \mathbb{N}, t \in \mathbb{Z}\}$, where t stands for time and i, j are cross-sectional indices, of observed random variables. Let $\mathbf{Y}_{n_y} := \{\mathbf{Y}_{n_y,t};\, t \in \mathbb{Z}\}$ and $\mathbf{Z}_{n_z} := \{\mathbf{Z}_{n_z,t};\, t \in \mathbb{Z}\}$ be the n_y- and n_z-dimensional subprocesses of $\mathbf{Y}$ and $\mathbf{Z}$, respectively, where $\mathbf{Y}_{n_y,t} := \left(Y_{1t}, \ldots, Y_{n_yt}\right)'$ and $\mathbf{Z}_{n_z,t} := \left(Z_{1t}, \ldots, Z_{n_zt}\right)'$, and write

$$\mathbf{X}_{\mathbf{n},t} := (Y_{1t}, \ldots, Y_{n_yt}, Z_{1t}, \ldots, Z_{n_zt})' := (\mathbf{Y}'_{n_y,t}\mathbf{Z}'_{n_z,t})'$$

with $\mathbf{n} := (n_y, n_z)$ and $n := n_y + n_z$. The Hilbert subspaces spanned by the processes $\mathbf{Y}, \mathbf{Z}$ and $\mathbf{X}$ are denoted by $\mathcal{H}_y$, $\mathcal{H}_z$ and $\mathcal{H}$, respectively.

The following assumption is made throughout the paper.

Assumption A1. For all $\mathbf{n}$, the vector process $\{\mathbf{X}_{\mathbf{n},t};\, t \in \mathbb{Z}\}$ is a zero-mean second-order stationary process.

Denoting by $\mathbf{\Sigma}_{y;n_y}(\theta)$ and $\mathbf{\Sigma}_{z;n_z}(\theta)$ the $(n_y \times n_y)$ and $(n_z \times n_z)$ spectral density matrices of $\mathbf{Y}_{n_y,t}$ and $\mathbf{Z}_{n_z,t}$, and by $\mathbf{\Sigma}_{yz;\mathbf{n}}(\theta) = \mathbf{\Sigma}'_{zy;\mathbf{n}}(\theta)$ their $(n_y \times n_z)$ cross-spectrum matrix, write

$$\mathbf{\Sigma}_{\mathbf{n}}(\theta) =: \begin{pmatrix} \mathbf{\Sigma}_{y;n_y}(\theta) & \mathbf{\Sigma}_{yz;\mathbf{n}}(\theta) \\ \mathbf{\Sigma}_{zy;\mathbf{n}}(\theta) & \mathbf{\Sigma}_{z;n_z}(\theta) \end{pmatrix}, \quad \theta \in [-\pi, \pi]$$

for the $(n \times n)$ spectral density matrix of $\mathbf{X}_{\mathbf{n},t}$, with elements $\sigma_{i_1i_2}(\theta), \sigma_{j_1j_2}(\theta)$ or $\sigma_{kk}(\theta)$, $k = 1, \ldots, n, i_1, i_2 = 1, \ldots, n_y, j_1, j_2 = 1, \ldots, n_z$. On these matrices, we make the following assumption.

Assumption A2. For any $k \in \mathbb{N}$, there exists a real $c_k > 0$ such that $\sigma_{kk}(\theta) \leq c_k$ for any $\theta \in [-\pi, \pi]$.

For any $\theta \in [-\pi, \pi]$, let $\lambda_{y;n_y,i}(\theta)$ be $\mathbf{\Sigma}_{y;n_y}(\theta)$'s i-th eigenvalue (in decreasing order of magnitude). The function $\theta \mapsto \lambda_{y;n_y,i}(\theta)$ is called $\mathbf{\Sigma}_{y;n_y}(\theta)$'s i-th *dynamic eigenvalue*. The notation $\theta \mapsto \lambda_{z;n_z,j}(\theta)$ and $\theta \mapsto \lambda_{\mathbf{n},k}(\theta)$ is used in an obvious way for the dynamic eigenvalues of $\mathbf{\Sigma}_{z;n_z}(\theta)$ and $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$, respectively. The corresponding *dynamic eigenvectors*, of dimensions $(n_y \times 1)$, $(n_z \times 1)$, and $(n \times 1)$, are denoted by $\mathbf{p}_{y;n_y,i}(\theta)$, $\mathbf{p}_{z;n_z,j}(\theta)$, and $\mathbf{p}_{\mathbf{n},k}(\theta)$, respectively.

Throughout, we repeatedly use the classical correspondence

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

between a matrix-values function $\mathbf{M}(\theta)$ and the filter $\underline{\mathbf{M}}(L)$: for instance,

$$\underline{\lambda}^{-1}_{1;k}(L) := \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\left[\int_{-\pi}^{\pi}\lambda^{-1}_{1;k}(\theta)\mathrm{e}^{\mathrm{i}s\theta}\mathrm{d}\theta\right]L^s,$$

$$\underline{\mathbf{\Sigma}}^{-1/2}(L) := \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\left[\int_{-\pi}^{\pi}\mathbf{\Sigma}^{-1/2}(\theta)\mathrm{e}^{\mathrm{i}s\theta}\,\mathrm{d}\theta\right]L^s, \quad \text{etc.}$$

Dynamic eigenvectors, in particular, can be expanded in Fourier series, e.g.

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

where the series on the right-hand side converge in quadratic mean, which in turn defines square summable filters of the form

$$\underline{\mathbf{p}}_{\mathbf{n},k}(L) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\left[\int_{-\pi}^{\pi}\mathbf{p}_{\mathbf{n},k}(\omega)\mathrm{e}^{\mathrm{i}s\omega}\mathrm{d}\omega\right]L^s.$$

Similarly define $\underline{\mathbf{p}}_{y;n_y,i}(L)$ and $\underline{\mathbf{p}}_{z;n_z,j}(L)$ from $\mathbf{p}_{y;n_y,i}(\theta)$ and $\mathbf{p}_{z;n_z,j}(\theta)$, respectively.

On these dynamic eigenvalues, we make the following assumptions.

Assumption A3. For some $q_y, q_z \in \mathbb{N}$,

(i) the q_y-th dynamic eigenvalue of $\mathbf{\Sigma}_{y;n_y}(\theta)$, $\lambda_{y;n_y,q_y}(\theta)$, diverges as $n_y \to \infty$, a.e. in $[-\pi, \pi]$, while the (q_y+1)-th one, $\lambda_{y;n_y,q_y+1}(\theta)$, is θ-a.e. bounded;

(ii) the q_z-th dynamic eigenvalue of $\mathbf{\Sigma}_{z;n_z}(\theta)$, $\lambda_{z;n_z,q_z}(\theta)$, diverges as $n_z \to \infty$, a.e. in $[-\pi, \pi]$, while the (q_z+1)-th one, $\lambda_{z;n_z,q_z+1}(\theta)$, is θ-a.e. bounded.

2 Another application, of an entirely different nature and scope, is considered in Hallin et al. (2011).

3 A similar problem is considered in two recent working papers (Ng et al., 2008; Ng and Moench, in press), where the authors adopt a hierarchical model with "national shocks" that are pervasive throughout all blocks, while "regional shocks" are region-specific; their approach, however, resorts to the *static* factor model literature.

The following lemma shows that this behavior of the dynamic eigenvalues of the subpanel spectral matrices $\mathbf{\Sigma}_{y;n_y}(\theta)$ and $\mathbf{\Sigma}_{z;n_z}(\theta)$ entails a similar behavior for the dynamic eigenvalues $\lambda_{\mathbf{n},k}(\theta)$ of $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$.

Lemma 1. *Let Assumptions A1–A3 hold. Then, there exists $q \in \mathbb{N}$, with $\max(q_y, q_z) \leq q \leq q_y + q_z$, such that $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$'s q-th dynamic eigenvalue $\lambda_{\mathbf{n},q}(\theta)$ diverges as $\min(n_y, n_z) \to \infty$, a.e. in $[-\pi, \pi]$, while the $(q+1)$-th one, $\lambda_{\mathbf{n},q+1}(\theta)$, is θ-a.e. bounded.*

Proof. See the Appendix. □

Theorem 2 in Forni and Lippi (2001) establishes that the behavior of dynamic eigenvalues described in Assumption A3 and Lemma 1 characterizes the existence of a *dynamic factor representation*. We say that a process $\mathbf{X} := \{X_{kt}, k \in \mathbb{N}, t \in \mathbb{Z}\}$ admits a *dynamic factor representation* with q factors if X_{kt} decomposes into a sum

$$X_{kt} = \chi_{kt} + \xi_{kt}, \quad \text{with } \chi_{kt} := \sum_{l=1}^{q} b_{kl}(L)u_{lt} \quad \text{and}$$

$$b_{kl}(L) := \sum_{m=1}^{\infty} b_{klm}L^m, \quad k \in \mathbb{N}, \ t \in \mathbb{Z},$$

such that

(i) the q-dimensional vector process $\{\mathbf{u}_t := (u_{1t}u_{2t} \ldots u_{qt})'; t \in \mathbb{Z}\}$ is orthonormal white noise;
(ii) the (unobservable) n-dimensional processes $\{\boldsymbol{\xi}_n := (\xi_{1t}\xi_{2t} \cdots \xi_{nt})'; t \in \mathbb{Z}\}$ are zero-mean stationary for any n, with *(idiosyncrasy)* θ-a.e. bounded (as $n \to \infty$) dynamic eigenvalues;
(iii) ξ_{k,t_1} and u_{l,t_2} are mutually orthogonal for any k, l, t_1 and t_2;
(iv) the filters $b_{kl}(L)$ are square summable: $\sum_{m=1}^{\infty} b_{klm}^2 < \infty$ for all $k \in \mathbb{N}$ and $l = 1, \ldots, q$, and
(v) q is minimal with respect to (i)–(iv).

The processes $\{u_{lt}; t \in \mathbb{Z}\}$, $l = 1, \ldots, q$, are called the *common shocks* or *factors*, the random variables ξ_{kt} and χ_{kt} the *idiosyncratic* and *common components* of X_{kt}, respectively. Actually, Forni and Lippi define *idiosyncrasy* via the behavior of dynamic aggregates, then show (their Theorem 1) that this definition is equivalent to the condition on dynamic eigenvalues given here.

That result of Forni and Lippi (2001), along with Lemma 1, leads to the following proposition.

Proposition 1. *Let Assumptions A1 and A2 hold. Then,*

(a) *Assumption A3(i) is satisfied iff the process **Y** has dynamic factor representation (q_y factors; call them the (common) y-factors, spanning the y-common space $\mathcal{H}_y^\chi$, with orthogonal complement $\mathcal{H}_y^\xi$)*

$$Y_{it} = \chi_{y;it} + \xi_{y;it} = \sum_{l=1}^{q_y} b_{y;il}(L)u_{y;lt} + \xi_{y;it}, \quad i \in \mathbb{N}, \ t \in \mathbb{Z}; \ (3)$$

(b) *Assumption A3(ii) is satisfied iff the process **Z** has dynamic factor representation (q_z factors; call them the (common) z-factors, spanning the z-common space $\mathcal{H}_z^\chi$, with orthogonal complement $\mathcal{H}_z^\xi$)*

$$Z_{jt} = \chi_{z;jt} + \xi_{z;jt} = \sum_{l=1}^{q_z} b_{z;jl}(L)u_{z;lt} + \xi_{z;jt}, \quad j \in \mathbb{N}, \ t \in \mathbb{Z}; \ (4)$$

(c) *Assumption A3 is satisfied iff the process **X** has dynamic factor representation (q factors, with q as in Lemma 1; call them the joint common factors, spanning the joint common space $\mathcal{H}_x^\chi$, with orthogonal complement $\mathcal{H}_x^\xi$).*

$$X_{kt} = \begin{cases} Y_{it} = \chi_{xy;it} + \xi_{xy;it} = \sum_{l=1}^{q} b_{xy;il}(L)u_{lt} + \xi_{xy;it}, \\ \quad k \in \mathbb{N}, \ t \in \mathbb{Z} \text{ in case } X_{kt} = Y_{it} \\ Z_{jt} = \chi_{xz;jt} + \xi_{xz;jt} = \sum_{l=1}^{q} b_{xz;jl}(L)u_{lt} + \xi_{xz;jt}, \\ \quad k \in \mathbb{N}, \ t \in \mathbb{Z} \text{ in case } X_{kt} = Z_{jt}. \end{cases} \quad (5)$$

All filters involved have square-summable coefficients.

Proof. The proof follows directly from the characterization theorem of Forni and Lippi (2001). □

It follows that, under Assumption A3, the processes **Y** and **Z** admit two distinct decompositions each: the *marginal factor models* (a) and (b), with *marginal common shocks* $u_{y;lt}$ ($l = 1, \ldots, q_y$) and $u_{z;lt}$ ($l = 1, \ldots, q_z$), respectively, and the *joint factor model* (c), with *joint common shocks* u_{lt} ($l = 1, \ldots, q$). This double representation allows for refining the factor decomposition. Call x-, y-, or z-*idiosyncratic* a process which is orthogonal (at all leads and lags) to the x-, y-, or z-factors (to $\mathcal{H}_x^\chi$, $\mathcal{H}_y^\chi$, or $\mathcal{H}_z^\chi$), respectively. Similarly, call x-, y-, or z-*common* any process belonging to $\mathcal{H}_x^\chi$, $\mathcal{H}_y^\chi$, or $\mathcal{H}_z^\chi$. The x-common components $\chi_{xy;it}$ and $\chi_{xz;jt}$ further decompose into

$$\chi_{xy;it} = \chi_{y;it} + \nu_{y;it} \quad \text{and} \quad \chi_{xz;jt} = \chi_{z;jt} + \nu_{z;jt},$$

where $\nu_{y;it} := \xi_{y;it} - \xi_{xy;it}$ is y-idiosyncratic, hence orthogonal to $\chi_{y;it}$, and $\nu_{z;jt} := \xi_{z;jt} - \xi_{xz;jt}$ is z-idiosyncratic, hence orthogonal to $\chi_{z;jt}$; since we also have $\nu_{y;it} := \chi_{xy;it} - \chi_{y;it}$ where both $\chi_{xy;it}$ and $\chi_{y;it}$ are orthogonal to $\xi_{xy;it}$, it follows that $\nu_{y;it}$ and $\xi_{xy;it}$ also are mutually orthogonal.

Define $\mathcal{H}_{y\cap z}^\chi := \mathcal{H}_y^\chi \bigcap \mathcal{H}_z^\chi$ and $q_{y\cap z} := q_y + q_z - q$: $\mathcal{H}_{y\cap z}^\chi$ is spanned by a $(q_{y\cap z})$-tuple of white noises, which are both y- and z-common (in case $q_y + q_z = q$, $\mathcal{H}_{y\cap z}^\chi = \{0\}$). Denoting by $\phi_{y;it}$ and $\phi_{z;jt}$ the projections of $\chi_{y;it}$ and $\chi_{z;jt}$ onto $\mathcal{H}_{y\cap z}^\chi$, and by $\psi_{y;it}$ and $\psi_{z;jt}$ the corresponding residuals, we obtain the decompositions

$$Y_{it} = \underbrace{\overbrace{\phi_{y;it} + \psi_{y;it}}^{} }_{\chi_{y;it}} \!\!\!\!\!\!\!\!\!\!\!\!\!\! \overbrace{\phi_{y;it} + \psi_{y;it} + \nu_{y;it}}^{\chi_{xy;it}} \! $$

$$Y_{it} = \overbrace{\underbrace{\phi_{y;it} + \psi_{y;it}}_{\chi_{y;it}} + \nu_{y;it}}^{\chi_{xy;it}} \!\!\!\!\!\!\!\!\! \underbrace{\phantom{\nu_{y;it}} + \xi_{xy;it}}_{\xi_{y;it}} \quad \text{and}$$

$$Z_{jt} = \overbrace{\underbrace{\phi_{z;jt} + \psi_{z;jt}}_{\chi_{z;jt}} + \nu_{z;jt}}^{\chi_{xz;jt}} \!\!\!\!\!\!\!\!\! \underbrace{\phantom{\nu_{z;jt}} + \xi_{xz;jt}}_{\xi_{z;jt}}, \quad i, j \in \mathbb{N}, \ t \in \mathbb{Z} \quad (6)$$

of the original observations into four mutually orthogonal components. The $\phi_{y;it}$ and $\phi_{z;jt}$ components will be called *strongly common*, $\xi_{xy;it}$ and $\xi_{xz;jt}$ *strongly idiosyncratic*, $\psi_{y;it}$ and $\psi_{z;jt}$ *weakly common*, $\nu_{y;it}$ and $\nu_{z;jt}$ *weakly idiosyncratic*. These decompositions induce additive decompositions of the variances of the observations into a sum of four terms indicating the relative contributions of each component.

In the following sections, we propose a procedure that provides consistent estimates of $\phi_{y;it}$, $\psi_{y;it}$, $\nu_{y;it}$, $\xi_{xy;it}$ and $\phi_{z;jt}$, $\psi_{z;jt}$, $\nu_{z;jt}$, $\xi_{xz;jt}$, hence $\xi_{y;it}$, $\xi_{xy;it}$, $\xi_{z;jt}$, and $\xi_{xz;jt}$.

3. Identifying the factor structure; population results

Based on the n-dimensional vector process $\mathbf{X}_{\mathbf{n},t} = \left(\mathbf{Y}'_{n_y,t}, \mathbf{Z}'_{n_z,t}\right)'$, we first asymptotically identify $\phi_{y;it}$, $\psi_{y;it}$, $\nu_{y;it}$, $\phi_{z;jt}$, $\psi_{z;jt}$ and $\nu_{z;jt}$ as $\min(n_y, n_z) \to \infty$. More precisely, we show that, under specified spectral structure, all those quantities can be consistently recovered from the finite-$\mathbf{n}$ subpanels $\{\mathbf{X}_{\mathbf{n},t}\}$ as $\min(n_y, n_z) \to \infty$.

3.1. Recovering the joint common and strongly idiosyncratic components

Under the joint factor model, Proposition 2 in Forni et al. (2000) provides $\mathbf{X}_{\mathbf{n},t}$-measurable reconstructions – denoted by $\chi^{\mathbf{n}}_{xy;it}$ and $\chi^{\mathbf{n}}_{xz;jt}$, respectively – of the joint common components $\chi_{xy;it}$ and $\chi_{xz;jt}$, which converge in quadratic mean for any i, j and t, as $\min(n_y, n_z) \to \infty$; we are using the terminology "reconstruction" rather than "estimation" to emphasize that spectral densities here, unlike in Section 4, are assumed to be known.

Write $\mathbf{M}^*$ for the adjoint (transposed, complex conjugate) of a matrix $\mathbf{M}$. The scalar process $\{V_{\mathbf{n},kt} := \underline{\mathbf{p}}^*_{\mathbf{n},k}(L)\mathbf{X}_{\mathbf{n},t};\, t \in \mathbb{Z}\}$, the spectral density of which is $\lambda_{\mathbf{n},k}(\theta)$, will be called $\mathbf{X}_{\mathbf{n},t}$'s (equivalently, $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$'s) *k-th dynamic principal component*, $k = 1, \ldots, n$. The basic properties of dynamic principal components imply that $\{V_{\mathbf{n},k_1 t}\}$ and $\{V_{\mathbf{n},k_2 t}\}$, for $k_1 \neq k_2$, are mutually orthogonal at all leads and lags. Forni et al. (2000) show that the projections of Y_{it} and Z_{jt} onto the closed space spanned by the present, past and future values of $V_{\mathbf{n},kt}$, $k = 1, \ldots, q$ yield the desired reconstructions of $\chi_{xy;it}$ and $\chi_{xz;jt}$. They also provide (up to minor changes due to the fact that they consider row- rather than column-eigenvectors, as we do here) the explicit forms

$$\chi^{\mathbf{n}}_{xy;it} = \underline{\mathbf{K}}^*_{y;\mathbf{n},i}(L)\mathbf{X}_{\mathbf{n},t} \quad \text{and}$$
$$\chi^{\mathbf{n}}_{xz;jt} = \underline{\mathbf{K}}^*_{z;\mathbf{n},j}(L)\mathbf{X}_{\mathbf{n},t} \quad i = 1, \ldots, n_y,\ j = 1, \ldots, n_z, \tag{7}$$

with

$$\underline{\mathbf{K}}_{y;\mathbf{n},i}(L) := \sum_{k=1}^{q} \underline{p}^*_{\mathbf{n},k,i}(L)\underline{\mathbf{p}}_{\mathbf{n},k}(L) \quad \text{and}$$
$$\underline{\mathbf{K}}_{z,\mathbf{n},j}(L) = \sum_{k=1}^{q} \underline{p}^*_{\mathbf{n},k,j}(L)\underline{\mathbf{p}}_{\mathbf{n},k}(L),$$

where $\underline{p}_{\mathbf{n},k,i}(L)$ denotes the i-th component of $\underline{\mathbf{p}}_{\mathbf{n},k}(L)$, i such that X_{it} belongs to the y-subpanel and $\underline{p}_{\mathbf{n},k,j}(L)$ the j-th component of $\underline{\mathbf{p}}_{\mathbf{n},k}(L)$, j such that X_{jt} belongs to the z-subpanel. We then can state a first consistency result.

Proposition 2. *Let Assumptions A1–A3 hold. Then,*

$$\lim_{\min(n_y,n_z)\to\infty} \chi^{\mathbf{n}}_{xy;it} = \chi_{xy;it} \quad \text{and} \quad \lim_{\min(n_y,n_z)\to\infty} \chi^{\mathbf{n}}_{xz;jt} = \chi_{xz;jt}$$

in quadratic mean, for any i, j, and t.

Proof. The proof consists in applying Proposition 2 in Forni et al. (2000) to the joint panel. □

It follows from (7) that $\chi^{\mathbf{n}}_{xy;it}$ has variance

$$\mathrm{Var}(\chi^{\mathbf{n}}_{xy;it}) = \sum_{k=1}^{q} \int_{-\pi}^{\pi} |p_{\mathbf{n},k,i}(\theta)|^2 \lambda_{\mathbf{n},k}(\theta)\mathrm{d}\theta.$$

Averaging this variance over the subpanel produces a measure

$$\frac{1}{n_y}\sum_{i=1}^{n_y} \mathrm{Var}(\chi^{\mathbf{n}}_{xy;it}) = \frac{1}{n_y}\sum_{i=1}^{n_y}\sum_{k=1}^{q} \int_{-\pi}^{\pi} |p_{\mathbf{n},k,i}(\theta)|^2 \lambda_{\mathbf{n},k}(\theta)\mathrm{d}\theta$$

of the contribution of joint common factors in the variability of the y-subpanel. Dividing it by the averaged variance

$$\frac{1}{n_y}\sum_{i=1}^{n_y} \mathrm{Var}(Y_{it}) = \frac{1}{n_y}\sum_{i=1}^{n_y} \int_{-\pi}^{\pi} \lambda_{y;n_y,i}(\theta)\mathrm{d}\theta$$

of the y-subpanel yields an evaluation

$$\sum_{i=1}^{n_y}\sum_{k=1}^{q} \int_{-\pi}^{\pi} |p_{\mathbf{n},k,i}(\theta)|^2 \lambda_{\mathbf{n},k}(\theta)\mathrm{d}\theta \Big/ \sum_{i=1}^{n_y} \int_{-\pi}^{\pi} \lambda_{y;n_y,i}(\theta)\mathrm{d}\theta \tag{8}$$

of its "degree of commonness" within the joint panel. For the z-subpanel, this measure takes the form

$$\sum_{j=1}^{n_z}\sum_{k=1}^{q} \int_{-\pi}^{\pi} |p_{\mathbf{n},k,j}(\theta)|^2 \lambda_{\mathbf{n},k}(\theta)\mathrm{d}\theta \Big/ \sum_{j=1}^{n_z} \int_{-\pi}^{\pi} \lambda_{z;n_z,j}(\theta)\mathrm{d}\theta.$$

As for the strongly idiosyncratic components $\xi_{xy;it}$, and $\xi_{xz;jt}$, they are consistently recovered, as $\min(n_y, n_z) \to \infty$, by $\xi^{\mathbf{n}}_{xy;it} := Y_{it} - \chi^{\mathbf{n}}_{xy;it}$ and $\xi^{\mathbf{n}}_{xz;jt} := Z_{jt} - \chi^{\mathbf{n}}_{xz;jt}$, respectively. In view of the mutual orthogonality of common and idiosyncatic components, the variance of $\xi^{\mathbf{n}}_{xy;it}$ writes

$$\mathrm{Var}(\xi^{\mathbf{n}}_{xy;it}) = \mathrm{Var}(Y_{it}) - \sum_{k=1}^{q} \int_{-\pi}^{\pi} |p_{\mathbf{n},k,i}(\theta)|^2 \lambda_{\mathbf{n},k}(\theta)\mathrm{d}\theta;$$

the complement to one of (8) therefore constitutes a measure of the "degree of idiosyncrasy" of the y-subpanel within the joint panel. Similar formulas hold for the strongly idiosyncratic component $\xi^{\mathbf{n}}_{xz;jt}$.

3.2. Recovering the marginal common, marginal idiosyncratic, and weakly idiosyncratic components

If $q_y = q$, the marginal common and idiosyncratic components $\chi_{y;it}$ and $\xi_{y;it}$ coincide with their joint counterparts $\chi_{xy;it}$ and $\xi_{xy;it}$, which were taken care of in the previous section. Assume therefore that $q > q_y$; the marginal and joint y-common spaces then do not coincide anymore.

Applying to the y-subpanels the same type of technique as in Section 3.1, consistent reconstructions of the $\chi_{y;it}$'s could be obtained from the spectral submatrices $\mathbf{\Sigma}_{y;n_y}(\theta)$. Now, $\chi_{y;it}$ is also the common component of $\chi_{xy;it}$, so that the same result can be obtained from factorizing the joint common spectral density matrices. As a reconstruction of $\chi_{y;it}$ we therefore consider the projection $\chi^{\mathbf{n}}_{y;it}$ of $\chi^{\mathbf{n}}_{xy;it}$ onto the space spanned by the first q_y dynamic principal components

$$\mathbf{V}^{\mathbf{n}}_{y;t} := \left(V^{\mathbf{n}}_{y;1t}, \ldots, V^{\mathbf{n}}_{y;q_y t}\right)', \quad \text{with } V^{\mathbf{n}}_{y;kt} := \underline{\mathbf{p}}^*_{\chi_{xy};\mathbf{n},k}(L)\boldsymbol{\chi}^{\mathbf{n}}_{xy;t}, \tag{9}$$

of the spectral density matrix $\mathbf{\Sigma}_{\chi_{xy};\mathbf{n}}(\theta)$ of $\boldsymbol{\chi}^{\mathbf{n}}_{xy;t} = (\chi^{\mathbf{n}}_{xy;1t}, \ldots, \chi^{\mathbf{n}}_{xy;n_y t})'$; $\mathbf{p}_{\chi_{xy};\mathbf{n},k}(\theta)$ here denotes the dynamic eigenvector associated with $\mathbf{\Sigma}_{\chi_{xy};\mathbf{n}}(\theta)$'s k-th dynamic eigenvalue $\lambda_{\chi_{xy};\mathbf{n},k}(\theta)$. This projection takes the form

$$\chi^{\mathbf{n}}_{y;it} = \underline{\mathbf{K}}^*_{\chi_{xy};\mathbf{n},i}(L)\boldsymbol{\chi}^{\mathbf{n}}_{xy;t} = \underline{\mathbf{K}}^*_{\chi_{xy};\mathbf{n},i}(L)\underline{\mathbf{K}}^*_{y;\mathbf{n},i}(L)\mathbf{X}_{\mathbf{n},t}, \tag{10}$$

with $\underline{\mathbf{K}}_{\chi_{xy};\mathbf{n},i}(L) := \sum_{k=1}^{q_y} \underline{p}_{\chi_{xy};\mathbf{n},k,i}(L)\underline{\mathbf{p}}_{\chi_{xy};\mathbf{n},k}(L)$, where $\underline{p}_{\chi_{xy};\mathbf{n},k,i}(L)$ stands for $\underline{\mathbf{p}}_{\chi_{xy};\mathbf{n},k}(L)$'s i-th component.

Similarly, the reconstruction $\chi^{\mathbf{n}}_{z;jt}$ of $\chi_{z;jt}$ is, with obvious notation,

$$\chi^{\mathbf{n}}_{z;jt} = \underline{\mathbf{K}}^*_{\chi_{xz};\mathbf{n},i}(L)\boldsymbol{\chi}^{\mathbf{n}}_{xz;t} = \underline{\mathbf{K}}^*_{\chi_{xz};\mathbf{n},j}(L)\underline{\mathbf{K}}^*_{y;\mathbf{n},j}(L)\mathbf{X}_{\mathbf{n},t}.$$

We then have a second consistency result.

Proposition 3. *Let Assumptions A1–A3 hold. Then*

$$\lim_{\min(n_y,n_z)\to\infty} \chi^{\mathbf{n}}_{y;it} = \chi_{y;it} \quad \text{and} \quad \lim_{\min(n_y,n_z)\to\infty} \chi^{\mathbf{n}}_{z;jt} = \chi_{z;jt}$$

in quadratic mean, for any i, j, and t.

Proof. The proof again is a direct application of Proposition 2 in Forni et al. (2000) to the y- and z-subpanels, respectively. □

The variance of the reconstructed marginal y-common component $\chi^{\mathbf{n}}_{y;it}$ is

$$\mathrm{Var}(\chi^{\mathbf{n}}_{y;it}) = \sum_{k=1}^{q_y} \int_{-\pi}^{\pi} |p_{\chi_{xy};\mathbf{n},k,i}(\theta)|^2 \lambda_{\chi_{xy};\mathbf{n},k}(\theta)\mathrm{d}\theta.$$

The averaged variance explained by the y-common factors in the y-subpanel is thus

$$\frac{1}{n_y}\sum_{i=1}^{n_y}\mathrm{Var}(\chi^{\mathbf{n}}_{y;it}) = \frac{1}{n_y}\sum_{i=1}^{n_y}\sum_{k=1}^{q_y}\int_{-\pi}^{\pi}|p_{\chi_{xy};\mathbf{n},k,i}(\theta)|^2\lambda_{\chi_{xy};\mathbf{n},k}(\theta)\mathrm{d}\theta = \frac{1}{n_y}\sum_{k=1}^{q_y}\int_{-\pi}^{\pi}\lambda_{\chi_{xy};\mathbf{n},k}(\theta)\mathrm{d}\theta. \quad (11)$$

Similarly, the averaged variance explained by the z-common factors in the z-subpanel is

$$\frac{1}{n_z}\int_{-\pi}^{\pi}\sum_{k=1}^{q_z}\lambda_{\chi_{xz};\mathbf{n},k}(\theta)\mathrm{d}\theta. \quad (12)$$

Consistent reconstructions of the marginal idiosyncratic components $\xi_{y;it}$ and $\xi_{z;jt}$ are straightforwardly obtained as $\xi^{\mathbf{n}}_{y;it} := Y_{it} - \chi^{\mathbf{n}}_{y;it}$ and $\xi^{\mathbf{n}}_{z;jt} := Z_{jt} - \chi^{\mathbf{n}}_{z;jt}$, whereas the weakly idiosyncratic components $\nu_{y;it}$ and $\nu_{z;it}$ can be recovered as $\nu^{\mathbf{n}}_{y;it} := \chi^{\mathbf{n}}_{xy;it} - \chi^{\mathbf{n}}_{y;it} = \xi^{\mathbf{n}}_{y;it} - \xi^{\mathbf{n}}_{xy;it}$ and $\nu^{\mathbf{n}}_{z;jt} := \chi^{\mathbf{n}}_{xz;jt} - \chi^{\mathbf{n}}_{z;jt} = \xi^{\mathbf{n}}_{z;jt} - \xi^{\mathbf{n}}_{xz;jt}$, respectively.

The averaged variance of weakly idiosyncratic components (or its ratio to $\sum_{i=1}^{n_y}\mathrm{Var}(Y_{it})$) can be interpreted as measuring the extent to which the z-common factors contribute to y-idiosyncratic variation. Clearly, since

$$\mathrm{Var}(Y_{it}) = \mathrm{Var}(\chi_{xy;it}) + \mathrm{Var}(\xi_{xy;it}) \quad \text{and}$$
$$\mathrm{Var}(\chi_{xy;it}) = \mathrm{Var}(\chi_{y;it}) + \mathrm{Var}(\nu_{y;it}),$$

we have

$$\mathrm{Var}(\nu_{y;it}) = \mathrm{Var}(\chi_{xy;it}) - \mathrm{Var}(\chi_{y;it}) = \mathrm{Var}(\xi_{y;it}) - \mathrm{Var}(\xi_{xy;it}).$$

In the finite-$\mathbf{n}$ panel, $\mathrm{Var}(\nu_{y;it})$ is estimated by

$$\mathrm{Var}(\nu^{\mathbf{n}}_{y;it}) = \mathrm{Var}(\xi^{\mathbf{n}}_{y;it}) - \mathrm{Var}(\xi^{\mathbf{n}}_{xy;it}), \quad (13)$$

so that the evaluation of the z-common factors contribution to y-idiosyncratic variation is

$$\frac{1}{n_y}\sum_{i=1}^{n_y}\mathrm{Var}(\nu^{\mathbf{n}}_{y;it}) = \frac{1}{n_y}\Big[\sum_{i=1}^{n_y}\sum_{k=1}^{q}\int_{-\pi}^{\pi}|p_{\mathbf{n},k,i}(\theta)|^2\lambda_{\mathbf{n},k}(\theta)\mathrm{d}\theta - \sum_{k=1}^{q_y}\int_{-\pi}^{\pi}\lambda_{\chi_{xy};\mathbf{n},k}(\theta)\mathrm{d}\theta\Big].$$

Similar formulas hold for the z-subpanel.

3.3. *Disentangling the strongly and weakly common components*

By definition, $\phi_{y;it}$ is obtained as the projection of $\chi_{y;it}$ onto $\mathcal{H}^{\chi}_{y\cap z}$, and $\psi_{y;it}$ follows as the residual $\chi_{y;it} - \phi_{y;it}$. Unlike $\mathcal{H}^{\chi}_{y}$ and $\mathcal{H}^{\chi}_{z}$, however, $\mathcal{H}^{\chi}_{y\cap z}$ is not characterized via an explicit sequence of orthonormal bases. The methods developed in the previous sections, thus, do not apply unless such a sequence can be computed first. This however requires some preparation: Proposition 4 is adapted from Theorem 8.3.1 in Brillinger (1981); Proposition 5, to the best of our knowledge, is new.

Proposition 4. *Assume that the $(r + s)$-dimensional second-order mean-zero stationary process $\{(\boldsymbol{\zeta}'_t, \boldsymbol{\eta}'_t)'; t \in \mathbb{Z}\}$ is such that the spectral density matrix $\boldsymbol{\Sigma}_{\eta\eta}(\theta)$ of $\boldsymbol{\eta}_t$ is nonsingular. Then, the projection of $\boldsymbol{\zeta}_t$ onto the closed space $\mathcal{H}_\eta$ spanned by $\{\boldsymbol{\eta}_t; t \in \mathbb{Z}\}$—that is, the r-tuple $\mathbf{A}^*(L)\boldsymbol{\eta}_t$ of square-summable linear combinations of the present, past and future of $\boldsymbol{\eta}_t$ minimizing $\mathrm{E}[(\boldsymbol{\zeta}_t - \mathbf{A}^*(L)\boldsymbol{\eta}_t)(\boldsymbol{\zeta}_t - \mathbf{A}^*(L)\boldsymbol{\eta}_t)']$ is $\underline{\boldsymbol{\Sigma}}_{\zeta\eta}(L)\underline{\boldsymbol{\Sigma}}^{-1}_{\eta\eta}(L)\boldsymbol{\eta}_t$, where $\boldsymbol{\Sigma}_{\zeta\eta}(\theta)$ denotes the cross-spectrum of $\boldsymbol{\zeta}_t$ and $\boldsymbol{\eta}_t$.*

Actually, Brillinger also requires $(\boldsymbol{\zeta}'_t, \boldsymbol{\eta}'_t)'$ to have absolutely summable autocovariances, so that the filter $\underline{\boldsymbol{\Sigma}}_{\zeta\eta}(L)\underline{\boldsymbol{\Sigma}}^{-1}_{\eta\eta}(L)$ also is absolutely summable. We, however, do not need this here.

Next, let $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_{12}$ be the Hilbert spaces spanned by $\{\mathbf{V}_{1;t}; t \in \mathbb{Z}\}$, $\{\mathbf{V}_{2;t}; t \in \mathbb{Z}\}$, and $\{(\mathbf{V}'_{1;t}, \mathbf{V}'_{1;t})'; t \in \mathbb{Z}\}$, respectively, where $\mathbf{V}_{1;t} := (V_{1;1,t}, \ldots, V_{1;q_1,t})'$ is a q_1-tuple (resp. $\mathbf{V}_{2;t} := (V_{2;1,t}, \ldots, V_{2;q_2,t})'$ a q_2-tuple) of mutually orthogonal (at all leads and lags) nondegenerate stochastic processes: the dynamic dimensions of $\mathcal{H}_1$ and $\mathcal{H}_2$ are thus q_1 and q_2, respectively. Denoting by

$$\boldsymbol{\Sigma}(\theta) =: \begin{pmatrix}\boldsymbol{\Sigma}_{11}(\theta) & \boldsymbol{\Sigma}_{12}(\theta)\\ \boldsymbol{\Sigma}_{21}(\theta) & \boldsymbol{\Sigma}_{22}(\theta)\end{pmatrix}, \quad \theta \in [-\pi, \pi],$$

the spectral density matrix of $(\mathbf{V}'_{1;t}, \mathbf{V}'_{2;t})'$, with $\boldsymbol{\Sigma}_{11}(\theta) = \mathrm{diag}(\lambda_{1;1}(\theta), \ldots, \lambda_{1;q_1}(\theta))$ and $\boldsymbol{\Sigma}_{22}(\theta) = \mathrm{diag}(\lambda_{2;1}(\theta), \ldots, \lambda_{2;q_2}(\theta))$, assume that $\boldsymbol{\Sigma}(\theta)$ has rank $q_{12}\theta$-a.e., so that $\mathcal{H}_{12}$ has dynamic dimension q_{12}, and the intersection $\mathcal{H}_{1\cap 2} := \mathcal{H}_1 \cap \mathcal{H}_2$ dynamic dimension $q_{1\cap 2} = q_1 + q_2 - q_{12}$. We then have the following result.

Proposition 5. (i) *The spectral density*

$$\boldsymbol{\Sigma}^{-1/2}_{22}(\theta)\boldsymbol{\Sigma}_{21}(\theta)\mathrm{diag}(\lambda^{-1}_{1;1}(\theta), \ldots, \lambda^{-1}_{1;q_1}(\theta)) \times \boldsymbol{\Sigma}_{12}(\theta)\boldsymbol{\Sigma}^{-1/2}_{22}(\theta), \quad \theta \in [-\pi, \pi] \quad (14)$$

of the q_2-dimensional process $\underline{\boldsymbol{\Sigma}}^{-1/2}_{22}(L)\underline{\boldsymbol{\Sigma}}_{21}(L)\underline{\boldsymbol{\Sigma}}^{-1}_{11}(L)\mathbf{V}_{1;t}$ has a maximal eigenvalue equal to one, with multiplicity $q_{1\cap 2}$.

(ii) *Denoting by $\mathbf{p}_{1\cap 2;1}(\theta), \ldots, \mathbf{p}_{1\cap 2;q_{1\cap 2}}(\theta)$ an arbitrary orthonormal basis of the corresponding $q_{1\cap 2}$-dimensional eigenspace, the process $\{\boldsymbol{\Upsilon}_t := (\Upsilon_{1,t}, \ldots, \Upsilon_{q_{1\cap 2},t})'; t \in \mathbb{Z}\}$, with*

$$\Upsilon_{k,t} := \underline{\mathbf{p}}^*_{1\cap 2;k}(L)\underline{\boldsymbol{\Sigma}}^{-1/2}_{22}(L)\mathbf{V}_{2;t}, \quad k = 1, \ldots, q_{1\cap 2}, \quad (15)$$

provides an orthonormal basis for $\mathcal{H}_{1\cap 2}$.

Proof. A random variable $\Upsilon \in \mathcal{H}_2$, that is, a variable of the form $\Upsilon = \underline{\mathbf{a}}^*_{\Upsilon}(L)\mathbf{V}_{2;t}$ with $\underline{\mathbf{a}}^*_{\Upsilon}(L) := (\underline{a}_{\Upsilon,1}(L), \ldots, \underline{a}_{\Upsilon,q_2}(L))$ belongs to $\mathcal{H}_{1\cap 2}$ iff it coincides with its projection onto the space $\mathcal{H}_1$ spanned by the $\mathbf{V}_{1;t}$'s. In view of Proposition 4, that projection is

$$\underline{\mathbf{a}}^*_{\Upsilon}(L)\underline{\boldsymbol{\Sigma}}_{21}(L)\mathrm{diag}(\underline{\lambda}^{-1}_{1;1}(L), \ldots, \underline{\lambda}^{-1}_{1;q_1}(L))\mathbf{V}_{1;t}. \quad (16)$$

The variance of a projection being less than or equal to the variance of the projected variable, the variance of (16) is less than or equal to the variance of Υ itself:

$$\int_{-\pi}^{\pi}\mathbf{a}^*_{\Upsilon}(\theta)\boldsymbol{\Sigma}_{21}(\theta)\mathrm{diag}(\lambda^{-1}_{1;1}(\theta), \ldots, \lambda^{-1}_{1;q_1}(\theta))\boldsymbol{\Sigma}_{12}(\theta)\mathbf{a}_{\Upsilon}(\theta)\mathrm{d}\theta$$
$$= \int_{-\pi}^{\pi}\mathbf{a}^*_{\Upsilon}(\theta)\boldsymbol{\Sigma}^{1/2}_{22}(\theta)\boldsymbol{\Sigma}^{-1/2}_{22}(\theta)\boldsymbol{\Sigma}_{21}(\theta)\mathrm{diag}(\lambda^{-1}_{1;1}(\theta), \ldots, \lambda^{-1}_{1;q_1}(\theta)) \times \boldsymbol{\Sigma}_{12}(\theta)\boldsymbol{\Sigma}^{-1/2}_{22}(\theta)\boldsymbol{\Sigma}^{1/2}_{22}(\theta)\mathbf{a}_{\Upsilon}(\theta)\mathrm{d}\theta$$
$$\leq \int_{-\pi}^{\pi}\mathbf{a}^*_{\Upsilon}(\theta)\boldsymbol{\Sigma}_{22}(\theta)\mathbf{a}_{\Upsilon}(\theta)\mathrm{d}\theta,$$

irrespective of $\boldsymbol{\Sigma}^{1/2}_{22}(\theta)\mathbf{a}_{\Upsilon}(\theta)$. It follows that the spectral matrix (14) has eigenvalues less than or equal to one (θ-a.e.), and that Υ is in $\mathcal{H}_{1\cap 2}$ iff $\boldsymbol{\Sigma}^{1/2}_{22}(\theta)\mathbf{a}_{\Upsilon}(\theta)$ belongs to the eigenspace with eigenvalue one. The $q_{1\cap 2}$ random variables $\Upsilon_{k,t}$ defined in (15) clearly satisfy that condition, and it is easy to check that the spectral density of $\{\boldsymbol{\Upsilon}_t; t \in \mathbb{Z}\}$ moreover is the $q_{1\cap 2} \times q_{1\cap 2}$ identity matrix. The result follows. □

The strongly common component $\phi_{y;it}$ is defined as the projection of $\chi_{y;it}$ onto $\mathcal{H}^{\chi}_{y\cap z}$: recovering it as the projection $\phi^{\mathbf{n}}_{y;it}$ of $\chi^{\mathbf{n}}_{y;it}$ onto the intersection of $\mathcal{H}^{\chi}_{y;\mathbf{n}}$ (spanned by $\mathbf{V}^{\mathbf{n}}_{y;t}$) and $\mathcal{H}^{\chi}_{z;\mathbf{n}}$ (spanned by $\mathbf{V}^{\mathbf{n}}_{z;t}$) seems a natural idea. Proposition 5, with $\mathcal{H}_1 = \mathcal{H}^{\chi}_{y;\mathbf{n}}$

and $\mathcal{H}_2 = \mathcal{H}^{\chi}_{z;\mathbf{n}}$, hence $q_1 = q_y, q_2 = q_z$ and $q_{1\cap 2} = q_{y\cap z}$, provides an orthonormal basis for that intersection. More precisely, denote by

$$\boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}}(\theta) := \begin{pmatrix} \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_y}(\theta) & \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_z}(\theta) \\ \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_z\mathbf{V}_y}(\theta) & \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_z\mathbf{V}_z}(\theta) \end{pmatrix}$$

the spectrum of $(\mathbf{V}^{\mathbf{n}\prime}_{y;t}, \mathbf{V}^{\mathbf{n}\prime}_{z;t})'$. The matrix (14) here takes the form

$$[\boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_z\mathbf{V}_z}(\theta)]^{-1/2}\boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_z\mathbf{V}_y}(\theta)[\boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_y}(\theta)]^{-1}\boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_z}(\theta)[\boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_z\mathbf{V}_z}(\theta)]^{-1/2};$$

denote by $\mathbf{p}_{y\cap z;\mathbf{n},1}(\theta), \ldots, \mathbf{p}_{y\cap z;\mathbf{n},q_{y\cap z}}(\theta)$ its $q_{y\cap z}$ first eigenvectors. An orthonormal basis of $\mathcal{H}^{\chi}_{y;\mathbf{n}} \cap \mathcal{H}^{\chi}_{z;\mathbf{n}}$ is $\mathbf{V}^{\mathbf{n}}_{y\cap z;t} := (V^{\mathbf{n}}_{y\cap z;1,t}, \ldots, V^{\mathbf{n}}_{y\cap z;q_{y\cap z},t})'$ where, in view of (15),

$$V^{\mathbf{n}}_{y\cap z;k,t} := \underline{\mathbf{p}}^*_{y\cap z;\mathbf{n},k}(L)\left(\underline{\boldsymbol{\Sigma}}^{\mathbf{n}}_{\mathbf{V}_z\mathbf{V}_z}\right)^{-1/2}(L)\mathbf{V}^{\mathbf{n}}_{z;t}, \quad k = 1, \ldots, q_{y\cap z}.$$

Since

$$\chi^{\mathbf{n}}_{y;it} = \sum_{k=1}^{q_y} \underline{p}_{\chi_{xy};\mathbf{n},k,i}(L)V^{\mathbf{n}}_{y:k,t} = (\underline{p}_{\chi_{xy};\mathbf{n},1,i}(L), \ldots, \underline{p}_{\chi_{xy};\mathbf{n},q_y,i}(L))\mathbf{V}^{\mathbf{n}}_{y:t},$$

we first compute the projection onto $\mathcal{H}^{\chi}_{y;\mathbf{n}} \cap \mathcal{H}^{\chi}_{z;\mathbf{n}}$ of $\mathbf{V}^{\mathbf{n}}_{y:t}$. That projection is obtained by applying Proposition 4 to the $(q_y + q_{y\cap z})$-dimensional random vector $(\mathbf{V}^{\mathbf{n}\prime}_{y:t}, \mathbf{V}^{\mathbf{n}\prime}_{y\cap z;t})'$, with spectral density

$$\begin{pmatrix} \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_y}(\theta) & \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_{y\cap z}}(\theta) \\ \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_{y\cap z}\mathbf{V}_y}(\theta) & \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_{y\cap z}\mathbf{V}_{y\cap z}}(\theta) \end{pmatrix} = \begin{pmatrix} \mathrm{diag}(\lambda_{\chi_{xy};\mathbf{n},1}(\theta), \ldots, \lambda_{\chi_{xy};\mathbf{n},q_y}(\theta)) & \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_{y\cap z}}(\theta) \\ \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_{y\cap z}\mathbf{V}_y}(\theta) & \mathbf{I}^{\mathbf{n}}_{y\cap z\times y\cap z}(\theta) \end{pmatrix}.$$

Letting

$$\begin{aligned}
\mathbf{P}_{y\cap z;\mathbf{n}}(\theta) &:= \mathrm{diag}(\lambda^{-1/2}_{\chi_{xz};\mathbf{n},1}(\theta), \ldots, \lambda^{-1/2}_{\chi_{xz};\mathbf{n},q_z}(\theta)) \\
&\quad\times \left(\mathbf{p}_{y\cap z;\mathbf{n},1}(\theta), \ldots, \mathbf{p}_{y\cap z;\mathbf{n},q_{y\cap z}}(\theta)\right), \\
\mathbf{P}_{\chi_{xy};\mathbf{n}}(\theta) &:= \left(\mathbf{p}_{\chi_{xy};\mathbf{n},1}(\theta), \ldots, \mathbf{p}_{\chi_{xy};\mathbf{n},q_y}(\theta)\right), \\
\mathbf{P}_{\chi_{xz};\mathbf{n}}(\theta) &:= \left(\mathbf{p}_{\chi_{xz};\mathbf{n},1}(\theta), \ldots, \mathbf{p}_{\chi_{xy};\mathbf{n},q_z}(\theta)\right), \\
\mathbf{P}_{\mathbf{n}(y)}(\theta) &:= \left(\mathbf{p}_{\mathbf{n},1(y)}(\theta), \ldots, \mathbf{p}_{\mathbf{n},q(y)}(\theta)\right), \quad \text{and} \\
\mathbf{P}_{\mathbf{n}(z)}(\theta) &:= \left(\mathbf{p}_{\mathbf{n},1(z)}(\theta), \ldots, \mathbf{p}_{\mathbf{n},q(z)}(\theta)\right),
\end{aligned}$$

with $\mathbf{p}_{\mathbf{n},k(y)}(\theta)$ collecting the components $p_{\mathbf{n},k,i}(\theta)$ of $\mathbf{p}_{\mathbf{n},k}(\theta)$ such that X_{it} belongs to the y-subpanel (resp. $\mathbf{p}_{\mathbf{n},k,(z)}(\theta)$ collecting the components $p_{\mathbf{n},k,j}(\theta)$ such that X_{jt} belongs to the z-subpanel), we have

$$\begin{aligned}
\boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_{y\cap z}}(\theta) &= \mathbf{P}^*_{\chi_{xy};\mathbf{n}}(\theta)\mathbf{P}_{\mathbf{n}(y)}(\theta)\mathrm{diag}(\lambda_{\mathbf{n},1}(\theta), \ldots, \lambda_{\mathbf{n},q}(\theta)) \\
&\quad\times \mathbf{P}^*_{\mathbf{n}(z)}(\theta)\mathbf{P}_{\chi_{xz};\mathbf{n}}(\theta)\mathbf{P}_{y\cap z;\mathbf{n}}(\theta).
\end{aligned}$$

The desired projection of $\mathbf{V}^{\mathbf{n}}_{y:t}$, in view of Proposition 4, is $\underline{\boldsymbol{\Sigma}}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_{y\cap z}}(L)\mathbf{V}^{\mathbf{n}}_{y\cap z;t}$; hence, the reconstructions we are proposing are, for the strongly common component $\phi_{y;it}$,

$$\begin{aligned}
\phi^{\mathbf{n}}_{y;it} &:= (\underline{p}_{\chi_{xy};\mathbf{n},1,i}(L), \ldots, \underline{p}_{\chi_{xy};\mathbf{n},q_y,i}(L))\underline{\boldsymbol{\Sigma}}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_{y\cap z}}(L)\mathbf{V}^{\mathbf{n}}_{y\cap z;t} \\
&= (\underline{p}_{\chi_{xy};\mathbf{n},1,i}(L), \ldots, \underline{p}_{\chi_{xy};\mathbf{n},q_y,i}(L))\underline{\mathbf{P}}^*_{\chi_{xy};\mathbf{n}}(L)\underline{\mathbf{P}}_{\mathbf{n}(y)}(L) \\
&\quad\times \mathrm{diag}(\underline{\lambda}_{\mathbf{n},1}(L), \ldots, \underline{\lambda}_{\mathbf{n},q}(L)) \\
&\quad\times \underline{\mathbf{P}}^*_{\mathbf{n}(z)}(L)\underline{\mathbf{P}}_{\chi_{xz};\mathbf{n}}(L)\underline{\mathbf{P}}_{y\cap z;\mathbf{n}}(L)\underline{\mathbf{P}}^*_{y\cap z;\mathbf{n}}(L)\mathbf{V}^{\mathbf{n}}_{z;t} \\
&=: \underline{\mathbf{H}}^*_{y;\mathbf{n},i}(L)\mathbf{V}^{\mathbf{n}}_{z;t},
\end{aligned} \tag{17}$$

and, for the weakly common one $\psi_{y;it}$, $\psi^{\mathbf{n}}_{y;it} := \chi^{\mathbf{n}}_{y;it} - \phi^{\mathbf{n}}_{y;it}$. With obvious changes, we similarly define $\phi^{\mathbf{n}}_{z;jt}$ and $\psi^{\mathbf{n}}_{z;jt}$.

Parallel with Propositions 2 and 3, we then have the following consistency result for $\phi^{\mathbf{n}}_{y;it}$ and $\phi^{\mathbf{n}}_{z;jt}$ (hence $\psi^{\mathbf{n}}_{y;it}$ and $\psi^{\mathbf{n}}_{z;jt}$).

Proposition 6. *Let Assumptions A1–A3 hold. Then*

$$\lim_{\min(n_y,n_z)\to\infty} \phi^{\mathbf{n}}_{y;it} = \phi_{y;it} \quad \text{and} \quad \lim_{\min(n_y,n_z)\to\infty} \phi^{\mathbf{n}}_{z;jt} = \phi_{z;jt}$$

in quadratic mean, for any i, j, and t.

Proof. The proof still follows from Proposition 2 of Forni et al. (2000), and the fact that all spectral densities involved, for given $\mathbf{n}$, are locally continuous functions of $\boldsymbol{\Sigma}_{\mathbf{n}}(\theta)$. □

It follows from (17) that the reconstructed strongly common component $\phi^{\mathbf{n}}_{y,it}$ has variance

$$\begin{aligned}
\mathrm{Var}(\phi^{\mathbf{n}}_{y,it}) &= \int_{-\pi}^{\pi} (p_{\chi_{xy};\mathbf{n},1,i}(\theta), \ldots, p_{\chi_{xy};\mathbf{n},q_y,i}(\theta))\boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_{y\cap z}}(\theta) \\
&\quad\times \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_{y\cap z}\mathbf{V}_y}(\theta)(p_{\chi_{xy};\mathbf{n},1,i}(\theta), \ldots, p_{\chi_{xy};\mathbf{n},q_y,i}(\theta))'\mathrm{d}\theta.
\end{aligned}$$

The average $\frac{1}{n_y}\sum_{i=1}^{n_y}\mathrm{Var}(\phi^{\mathbf{n}}_{y;it})$ measures the contribution of the strongly common factors in the total variation of the y-subpanel.

Similar quantities are easily computed for the z-subpanel.

4. Recovering the factor structure; estimation results

The previous section shows how all components of Y_{it} and Z_{jt} can be recovered asymptotically as $\min(n_y, n_z) \to \infty$, provided that the spectral density $\boldsymbol{\Sigma}_{\mathbf{n}}$ and the numbers q, q_y, and q_z of factors are known. The estimates $\phi^{\mathbf{n}}_{y;it}$, $\psi^{\mathbf{n}}_{y;it}$ and $\nu^{\mathbf{n}}_{y;it}$ all take the form of a filtered series of the observed process $\mathbf{X}_{\mathbf{n},t}$. We have indeed

$$\begin{aligned}
\phi^{\mathbf{n}}_{y;it} &= \underline{\mathbf{H}}^*_{y;\mathbf{n},i}(L)\mathbf{V}^{\mathbf{n}}_{z;t} =: \underline{\mathbf{K}}^*_{\phi_y;\mathbf{n},i}(L)\mathbf{X}_{\mathbf{n},t} \\
\psi^{\mathbf{n}}_{y;it} &= \chi^{\mathbf{n}}_{y;it} - \phi^{\mathbf{n}}_{y;it} = [\underline{\mathbf{K}}^*_{\chi_{xy};\mathbf{n},i}(L)\underline{\mathbf{K}}^*_{y;\mathbf{n},i}(L) - \underline{\mathbf{K}}^*_{\phi_y;\mathbf{n},i}(L)]\mathbf{X}_{\mathbf{n},t} \\
&=: \underline{\mathbf{K}}^*_{\psi_y;\mathbf{n},i}(L)\mathbf{X}_{\mathbf{n},t}, \quad \text{and} \\
\nu^{\mathbf{n}}_{y;it} &= \chi^{\mathbf{n}}_{xy;it} - \chi^{\mathbf{n}}_{y;it} = [\underline{\mathbf{K}}^*_{y;\mathbf{n},i}(L) - \underline{\mathbf{K}}^*_{\chi_{xy};\mathbf{n},i}(L)\underline{\mathbf{K}}^*_{y;\mathbf{n},i}(L)]\mathbf{X}_{\mathbf{n},t} \\
&=: \underline{\mathbf{K}}^*_{\nu_y;\mathbf{n},i}(L)\mathbf{X}_{\mathbf{n},t},
\end{aligned}$$

with

$$\underline{\mathbf{K}}^*_{\phi_y;\mathbf{n},i}(L) := \underline{\mathbf{H}}^*_{y;\mathbf{n},i}(L)\underline{\mathbf{P}}^*_{\chi_{xz};\mathbf{n}}(L)\underline{\mathbf{P}}_{\mathbf{n}(z)}(L)\underline{\mathbf{P}}^*_{\mathbf{n}}(L)\mathbf{X}_{\mathbf{n},t},$$

and $\underline{\mathbf{P}}_{\mathbf{n}}(L) := (\underline{\mathbf{p}}_{\mathbf{n},1}(L), \ldots, \underline{\mathbf{p}}_{\mathbf{n},q}(L))$.

These three filters all are functions of the spectral density matrix $\boldsymbol{\Sigma}_{\mathbf{n}}(\theta)$ which of course in practice is unknown, as we only observe a finite realization $\mathbf{X}^T_{\mathbf{n}} := (\mathbf{X}_{\mathbf{n}1}, \mathbf{X}_{\mathbf{n}2}, \ldots, \mathbf{X}_{\mathbf{n}T})$ of $\mathbf{X}_{\mathbf{n}}$. We therefore need an estimator $\boldsymbol{\Sigma}^T_{\mathbf{n}}(\theta)$ of $\boldsymbol{\Sigma}_{\mathbf{n}}(\theta)$, the consistency of which requires strengthening slightly Assumption A1 into the following Assumption A1′:

Assumption A1′. For all $\mathbf{n}$, the vector process $\{\mathbf{X}_{\mathbf{n},t};\, t \in \mathbb{Z}\}$ admits a linear representation $\mathbf{X}_{\mathbf{n},t} = \sum_{k=-\infty}^{\infty}\mathbf{C}_k\boldsymbol{\zeta}_{t-k}$, where $\boldsymbol{\zeta}_t$ is full-rank n-dimensional white noise with finite fourth order moments, and the $n \times n$ matrices $\mathbf{C}_k = (C_{ij,k})$ are such that $\sum_{k=-\infty}^{\infty}|k||C_{ij,k}|^{1/2} < \infty$ for all i, j.

Under Assumption A1′, if $\boldsymbol{\Sigma}^T_{\mathbf{n}}(\theta)$, with elements $\sigma^T_{\mathbf{n},ij}(\theta)$, denotes any periodogram-smoothing or lag-window estimator of $\boldsymbol{\Sigma}_{\mathbf{n}}(\theta)$,

$$\lim_{T\to\infty} \mathrm{P}\left[\sup_{\theta\in[-\pi,\pi]}\left|\sigma^T_{\mathbf{n},ij}(\theta) - \sigma_{ij}(\theta)\right| > \varepsilon\right] = 0$$

for all $\mathbf{n}$, i, j, and $\varepsilon > 0$, (see Forni et al., 2000).[4] In Section 6, we consider the lag-window estimators

$$\boldsymbol{\Sigma}^T_{\mathbf{n}}(\theta) := \sum_{k=-M_T}^{M_T}\boldsymbol{\Gamma}^T_{\mathbf{n}k}\omega_k \mathrm{e}^{-\mathrm{i}k\theta} \tag{18}$$

[4] Actually, Forni et al. (2000) wrongly borrow the result from Brockwell and Davis (1987); a more appropriate reference is Robinson (1991), Theorem 2.1.

where $\mathbf{\Gamma}^T_{\mathbf{n}k}$ is the sample covariance matrix of $\mathbf{X}_{\mathbf{n},t}$ and $\mathbf{X}_{\mathbf{n},t-k}$ and $\omega_k := 1-|k|/(M_T+1)$ are the weights corresponding to the Bartlett lag window of size M_T. Consistency then is achieved provided that the following assumption holds:

Assumption B. $M_T \to \infty$, and $M_T T^{-1} \to 0$, as $T \to \infty$.

For simplicity, we consider a single M_T for the y- and z-subpanels. One also could base the estimation of $\mathbf{\Sigma}_{y;n_y}$, $\mathbf{\Sigma}_{z;n_z}$ and $\mathbf{\Sigma}_{yz;\mathbf{n}}$ on three distinct bandwidth parameter values M^y_T, M^z_T, and M^{zy}_T.

A consistent estimator $\mathbf{\Sigma}^T_{\mathbf{n}}(\theta)$ of $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$ however is not sufficient here. Deriving, from this estimator $\mathbf{\Sigma}^T_{\mathbf{n}}(\theta)$, estimated versions $\underline{\mathbf{K}}^T_{\phi_y;\mathbf{n},i}(L)$, $\underline{\mathbf{K}}^T_{\psi_y;\mathbf{n},i}(L)$ and $\underline{\mathbf{K}}^T_{\nu_y;\mathbf{n},i}(L)$, of the filters $\underline{\mathbf{K}}_{\phi_y;\mathbf{n},i}(L)$, $\underline{\mathbf{K}}_{\psi_y;\mathbf{n},i}(L)$ and $\underline{\mathbf{K}}_{\nu_y;\mathbf{n},i}(L)$ indeed also requires an estimation of the numbers of factors q, q_y and q_z involved. The only method allowing for such estimation is the identification method developed in Hallin and Liška (2007), which we now briefly describe, with a few adjustments taking into account the notation of this paper. For a detailed description of the procedure, we refer to the section entitled "A practical guide to the selection of q" in Hallin and Liška (2007).

The lag window method described in (18) provides estimations $\mathbf{\Sigma}^T_{\mathbf{n}}(\theta_l)$ of the spectral density at frequencies $\theta_l := \pi l/(M_T + 1/2)$ for $l = -M_T, \ldots, M_T$. Based on these estimations, consider the information criterion

$$IC^T_{\mathbf{n};c}(k) := \log\left[\frac{1}{n}\sum_{i=k+1}^{n}\frac{1}{2M_T+1}\sum_{l=-M_T}^{M_T}\lambda^T_{\mathbf{n}i}(\theta_l)\right] + kcp(n,T), \quad 0 \le k \le q_{\max}, c \in \mathbb{R}^+_0, \tag{19}$$

where the penalty function $p(n, T)$ is $o(1)$ while $p^{-1}(n, T)$ is $o\left(\min(n, M_T^2, M_T^{-1/2}T^{1/2})\right)$ as both n and T tend to infinity, and $q_{\max}$ is some predetermined upper bound; the eigenvalues $\lambda^T_{\mathbf{n}i}(\theta_l)$ are those of $\mathbf{\Sigma}^T_{\mathbf{n}}(\theta_l)$. Depending on $c > 0$, the estimated number of factors, for given $\mathbf{n}$ and T, is

$$q^T_{\mathbf{n};c} := \underset{0\le k\le q_{\max}}{\operatorname{argmin}}\; IC^T_{\mathbf{n};c}(k).$$

Hallin and Liška (2007) prove that this $q^T_{\mathbf{n};c}$ is consistent for any $c > 0$. "Optimal" tuning c^* of c is then performed as follows. Consider a J-tuple of the form $q^{T_j}_{c,\mathbf{n}_j}$, $j = 1, \ldots, J$, where $\mathbf{n}_j = (n_{y;j}, n_{z;j})$ with $0 < n_{y;1} < \cdots < n_{y;J} = n_y$, $0 < n_{z;1} < \cdots < n_{z;J} = n_z$, and $0 < T_1 \le \cdots \le T_J = T$. This J-tuple can be interpreted as a "history" of the identification procedure, and characterizes, for each $c > 0$, a sequence $q^{T_j}_{c,\mathbf{n}_j}$, $j = 1, \ldots, J$ of estimated factor numbers. In order to keep a balanced representation of the two blocks, we only consider J-tuples along which $n_{y;j}/n_{z;j}$ is as close as possible to n_y/n_z.

The selection of c^* is based on the inspection of two mappings: $c \to q^T_{\mathbf{n};c}$, and $c \to S_c$, where

$$S^2_c := J^{-1}\sum_{j=1}^{J}\left(q^{T_j}_{\mathbf{n}_j;c} - J^{-1}\sum_{j=1}^{J} q^{T_j}_{\mathbf{n}_j;c}\right)^2$$

measures the variability of $q^{T_j}_{\mathbf{n}_j;c}$ over the "history". For n and T large enough, S_c exhibits "stability intervals", that is, intervals of c values over which $S_c = 0$. The definition of S_c implies that $c \mapsto q^T_{\mathbf{n};c}$ is constant over such intervals. Starting in the neighbourhood of $c = 0$, a first stability interval $(0, c^+_1)$ corresponds to $q^T_{\mathbf{n};c} = q_{\max}$; choose c^* as any point in the next one, (c^-_2, c^+_2). The selected number of factors is then $q^T_{\mathbf{n}} = q^T_{\mathbf{n};c^*}$. The same method, applied to the y- and z-subpanels, yields estimators $q^T_{n_y}$ and $q^T_{n_z}$ of q_y and q_z; $q^T_{\mathbf{n};yz} := q^T_{n_y} + q^T_{n_z} - q^T_{\mathbf{n}}$ provides a consistent estimator of q_{yz}.

The success of this identification method however also requires strengthening somewhat the assumptions; from now on, we reinforce Assumption A1′ into Assumption A1″ and Assumptions A2 and A3 into Assumptions A2′ and A3′:

Assumption A1″. Same as Assumption A1′, but (i) the convergence condition on the $C_{ij,k}$'s is uniform, $\sup_{i,j\in\mathbb{N}}\sum_{k=-\infty}^{\infty}|C_{ij,k}||k|^{1/2} < \infty$, and (ii) writing $c_{i_1,\ldots,i_\ell}(k_1, \ldots, k_{\ell-1})$ for the cumulant of order ℓ of $X_{i_1}(t + k_1), \ldots, X_{i_{\ell-1}}(t + k_{\ell-1}), X_{i_\ell}(t)$, for all $1 \le \ell \le 4$ and $1 \le j < \ell$, $\sup_{i_1,\ldots,i_\ell}[\sum_{k_1=-\infty}^{\infty}\cdots\sum_{k_{\ell-1}=-\infty}^{\infty}|c_{i_1,\ldots,i_\ell}(k_1, \ldots, k_{\ell-1})|] < \infty$.

Assumption A2′. The entries $\sigma_{ij}(\theta)$ of $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$ (i) are bounded, uniformly in $\mathbf{n}$ and θ – that is, there exists a real $c > 0$ such that $|\sigma_{ij}(\theta)| \le c$ for any $i, j \in \mathbb{N}$ and $\theta \in [-\pi, \pi]$ – and (ii) they have bounded, uniformly in $\mathbf{n}$ and θ, derivatives up to the order two—namely, there exists $Q < \infty$ such that $\sup_{i,j\in\mathbb{N}}\sup_\theta\left|\frac{\mathrm{d}^k}{\mathrm{d}\theta^k}\sigma_{ij}(\theta)\right| \le Q$, $k = 0, 1, 2$.

Assumption A3′. Same as Assumption A3, but moreover (i) $\lambda_{y;n_y,q_y}(\theta)$ and $\lambda_{z;n_z,q_z}(\theta)$ diverge at least linearly in n_y and n_z, respectively, that is, $\liminf_{n_y\to\infty}\inf_\theta n_y^{-1}\lambda_{y;n_y,q_y}(\theta) > 0$, and $\liminf_{n_z\to\infty}\inf_\theta n_z^{-1}\lambda_{z;n_z,q_z}(\theta) > 0$, and (ii) both n_y/n_z and n_z/n_y are $O(1)$ as $\min(n_y, n_z) \to \infty$.

This "at least linear" divergence assumption is also made in Hallin and Liška (2007), and can be considered as a form of cross-sectional stability of the two panels under study.

Once estimated values of the numbers q, q_y and q_z of factors are available, the estimated counterparts of of $\underline{\mathbf{K}}_{\phi_y;\mathbf{n},i}(L)$, $\underline{\mathbf{K}}_{\psi_y;\mathbf{n},i}(L)$ and $\underline{\mathbf{K}}_{\nu_y;\mathbf{n},i}(L)$ are obtained by substituting $\mathbf{\Sigma}^T_{\mathbf{n}}(\theta)$, $q^T_{\mathbf{n}}$, $q^T_{n_y}$ and $q^T_{n_z}$ for $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$, q, q_y and q_z in all definitions of Section 3, then truncating infinite sums as explained in Section B of Forni et al. (2000) (a truncation which depends on t, which explains the notation), yielding $\underline{\mathbf{K}}^{Tt}_{\phi_y;\mathbf{n},i}(L)$, $\underline{\mathbf{K}}^{Tt}_{\psi_y;\mathbf{n},i}(L)$ and $\underline{\mathbf{K}}^{Tt}_{\nu_y;\mathbf{n},i}(L)$. Parallel with Proposition 3 in Forni et al. (2000), we then have the following result.

Proposition 7. *Let Assumptions A1″, A2′, A3′ and B hold. Then, for all* $\epsilon_k > 0$ *and* $\eta_k > 0$, $k = 1, 2, 3$, *there exists* $N_0(\epsilon_1, \epsilon_2, \epsilon_3, \eta_1, \eta_2, \eta_3)$ *such that* $\mathrm{P}\left[\left|\underline{\mathbf{K}}^{Tt\,*}_{\phi_y;\mathbf{n},i}(L)\mathbf{X}_{\mathbf{n},t} - \phi_{y;it}\right| > \epsilon_1\right] \le \eta_1$, $\mathrm{P}\left[\left|\underline{\mathbf{K}}^{Tt*}_{\psi_y;\mathbf{n},i}(L)\mathbf{X}_{\mathbf{n},t} - \psi_{y;it}\right| > \epsilon_2\right] \le \eta_2$, *and* $\mathrm{P}\left[\left|\underline{\mathbf{K}}^{Tt*}_{\nu_y;\mathbf{n},i}(L)\mathbf{X}_{\mathbf{n},t} - \nu_{y;it}\right| > \epsilon_3\right] \le \eta_3$, *for all* $t = \check{t}(T)$ *satisfying* $a \le \liminf_{T\to\infty}(\check{t}(T)/T) \le \limsup_{T\to\infty}(\check{t}(T)/T) \le b$, *for some* a, b *such that* $0 < a < b < 1$, *all* $n \ge N_0$ *and all T larger than some* $T_0(\mathbf{n}, \epsilon_1, \epsilon_2, \epsilon_3, \eta_1, \eta_2, \eta_3)$.

Proof. The proof consists in reproducing, for each projection involved in the reconstruction of $\phi_{y;it}$, $\psi_{y;it}$ and $\nu_{y;it}$, the proof of Proposition 3 in Forni et al. (2000). Lengthy but obvious details are omitted. □

Consistent estimations of the various contributions to the total variance of each subpanel can be obtained either by substituting estimated spectral eigenvalues and eigenvectors for the exact ones in the formulas of Section 3, and replacing integrals with the corresponding finite sums over Fourier frequencies, or by computing the empirical variances of the estimated strongly and weakly common, strongly and weakly idiosyncratic components.

5. Dynamic factors in the presence of K blocks ($K > 2$)

The ideas developed in the previous sections extend to the more general case of $K > 2$ blocks, with, however, rapidly increasing complexity. Each subset $\{k_1, \ldots, k_\ell\}$, $\ell = 0, 1, \ldots, K$ of $\{1, \ldots, K\}$ indeed characterizes a decomposition of $\mathcal{H}$ into mutually orthogonal common and idiosyncratic subspaces, $\mathcal{H}^{\chi}_{\{k_1,\ldots,k_\ell\}}$ and $\mathcal{H}^{\xi}_{\{k_1,\ldots,k_\ell\}}$, say, leading to 2^K distinct implementations of the Hallin–Liška and Forni et al. procedures.

Instead of Y_{it} and Z_{jt}, denote all observations as $X_{k;it}$ ($i = 1, \ldots, n$; $k = 1, \ldots, K$), with the additional label k specifying that $X_{k;it}$ belongs to block k. Projecting onto $\mathcal{H}^{\chi}_{\{k_1,\ldots,k_\ell\}}$ and $\mathcal{H}^{\xi}_{\{k_1,\ldots,k_\ell\}}$ yields the orthogonal decomposition (for simplicity, we are dropping the subscripts n and T) $X_{k;it} = \chi_{k;\{k_1,\ldots,k_\ell\};it} + \xi_{k;\{k_1,\ldots,k_\ell\};it}$ of $X_{k;it}$ into a $\{k_1, \ldots, k_\ell\}$-common and a $\{k_1, \ldots, k_\ell\}$-idiosyncratic component; in particular, $\xi_{k;\{1,\ldots,K\};it}$, as in Section 2, can be called *strongly idiosyncratic*. For $\{l_1, \ldots, l_m\} \subset \{k_1, \ldots, k_\ell\}$ and $k \in \{k_1, \ldots, k_\ell\} \setminus \{l_1, \ldots, l_m\}$, call *weakly idiosyncratic* the difference

$$\begin{aligned}
&\nu_{k;\{l_1,\ldots,l_m\},\{k_1,\ldots,k_\ell\}\setminus\{l_1,\ldots,l_m\};it} \\
&\quad := \xi_{k;\{k_1,\ldots,k_\ell\}\setminus\{l_1,\ldots,l_m\};it} - \xi_{k;\{k_1,\ldots,k_\ell\};it} \\
&\quad = \chi_{k;\{k_1,\ldots,k_\ell\};it} - \chi_{k;\{k_1,\ldots,k_\ell\}\setminus\{l_1,\ldots,l_m\};it}.
\end{aligned}$$

Since, for $A \subset B$, $\mathcal{H}^{\chi}_A$ is a subspace of $\mathcal{H}^{\chi}_B$ and $\mathcal{H}^{\xi}_B$ a subspace of $\mathcal{H}^{\xi}_A$, this weakly idiosyncratic component is orthogonal to both $\chi_{k;\{l_1,\ldots,l_m\};it}$ and $\xi_{k;\{k_1,\ldots,k_\ell\};it}$.

Further projections can be performed, yielding components with various degrees of commonness/idiosyncrasy. We restrict to projections onto common spaces of the form

$$\mathcal{H}^{\chi}_{l_1\cap\cdots\cap l_m} := \mathcal{H}^{\chi}_{l_1} \bigcap \ldots \bigcap \mathcal{H}^{\chi}_{l_m}$$

decomposing $\chi_{k;\{k_1,\ldots,k_\ell\};it}$, $k \in \{l_1, \ldots, l_m\} \supseteq \{k_1, \ldots, k_\ell\}$, into

$$\phi_{k;\{l_1\cap\cdots\cap l_m\};it} \quad \text{and}$$

$$\psi_{k;\{k_1,\ldots,k_\ell\},\{l_1\cap\cdots\cap l_m\};it} := \chi_{k;\{k_1,\ldots,k_\ell\};it} - \phi_{k;\{l_1\cap\cdots\cap l_m\};it}$$

which, in analogy with the two-block case, we respectively call *strongly* and *weakly common* components. Projections onto (nondegenerate) subspaces of the form $\mathcal{H}^{\chi}_{l_1\cap\cdots\cap l_m}$ can be obtained via the method described in Section 3.3. Sequences of projections onto decreasing sequences of $\mathcal{H}^{\chi}$'s yield decompositions of the original observations into sums of mutually orthogonal components, along with decompositions of their variances; these decompositions, however, depend on the sequence of projections adopted.

In view of the rapidly increasing notational burden, we will not pursue any further with formal developments; an application for $K = 3$ is considered in Section 6.2.

6. Real data applications

We applied our method to a dataset of monthly Industrial Production Indexes for France, Germany, and Italy, observed from January 1995 through December 2006. All data were preadjusted by taking a log-difference transformation ($T = 143$ throughout—one observation is lost due to differencing), then centered and normalized using their sample means and standard errors.

In practice, some care has to be taken, however, due to the fact that, for finite n and T, the joint and marginal common spaces reconstructed from estimated spectral densities need not being nested. When defined from population spectral densities, the y-common space $\mathcal{H}^{\chi}_y$ associated with the Y_{it}'s and the common space associated with the $\chi_{xy;it}$'s coincide, and are a subspace of the joint common space $\mathcal{H}^{\chi}_{xy}$. When based on finite n and T estimations, those two spaces in general are distinct, and only the second one is a subspace of $\mathcal{H}^{\chi}_{xy}$. In order to avoid ambiguities and inconsistencies in sums of squares, it is important, under finite n and T, to proceed with projections and spectral estimation in a sequence. For instance, one first should estimate the global spectrum, and project the Y_{it}'s onto the reconstruction of $\mathcal{H}^{\chi}_{xy}$ based on that global spectral estimation. From those projections, $\chi^{\mathbf{n};T}_{xy;it}$, say, one should then estimate the spectrum of the $\chi_{xy;it}$'s, the dynamic principal components of which in turn yield the reconstruction of $\mathcal{H}^{\chi}_y$, etc. This sequence of projections and estimations is carefully described in the applications we are treating below.

6.1. A two-block analysis

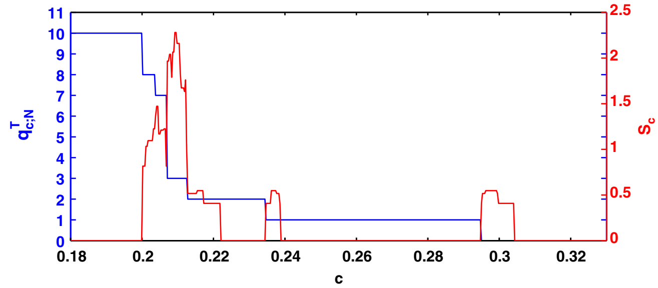
First consider the data for France and Germany. Using $X_{F;it}$ for the French observations Y_{it} and $X_{G;jt}$ for the German ones Z_{jt}, we have $n_y = n_F = 96$, $n_z = n_G = 114$, hence $n = n_{FG} = 210$. Spectral densities were estimated from the pooled panel using a lag-window estimators of the form (18), with truncation parameter $M_T = 0.5\sqrt{T} = 5$. Based on this estimation, we ran the Hallin and Liška (2007) identification method on the French and German subpanels, with sequences $n_{F,j} = 96 - 2j$, $j = 1, \ldots, 5$ and $n_{G,j} = 96 - 2j$, $j = 1, \ldots, 5$, respectively, then on the pooled panel, with sequence $n_{FG,j} = 210 - 2j$, $j = 1, \ldots, 8$ and an "almost constant" proportion 96/210, 114/210 of French and German observations (namely, $\lceil 96n_{FG,j}/210 \rceil$ French observations, and $\lfloor 114n_{FG,j}/210 \rfloor$ German ones. In all cases, we put $T_j = T = 143$, $j = 1, \ldots, 5$. The range for c values, after some preliminary exploration, was taken as $[0, 0.0002, 0.0004, \ldots, 0.5]$, and $q_{\max}$ was set to 10. In all cases, the panels were randomly ordered prior to the analysis. The penalty function was $p(n, T) = \left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)^{-1/2}$.

The results are shown in Fig. 1, and very clearly conclude for $q^T_{(n_F,n_G)} = 3$ (for $c \in [0.1798, 0.1894]$), $q^T_{n_F,F} = 2$ (for $c \in [0.2222, 0.2344]$), and $q^T_{n_G,G} = 3$ (for $c \in [0.2032, 0.2138]$). Since $q^T_{(n_F,n_G)} = \max(q^T_{n_F,F}, q^T_{n_G,G})$, this identification yields a block structure with 3 joint common factors, 3 German-common and 2 French-common factors, the French-common space being included in the German-common one. The French-common components thus are strongly common (no weakly common French components), whereas one German-common factor is French-idiosyncratic (no weakly idiosyncratic German components).

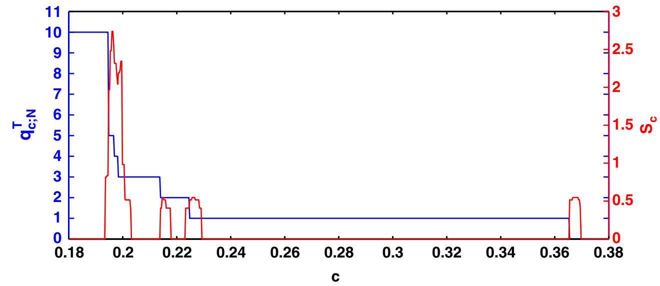
Taking these facts into account, the three factor analyses described in Sections 2 and 3 yield

(a) (global analysis) an analysis based on spectral estimation in the global panel (three factors) decomposes the French observation $X_{F;it}$ (resp. the German observation $X_{G;jt}$) into a strongly idiosyncratic $\xi_{xF;it}$ (resp. $\xi_{xG;jt} = \xi_{G;jt}$) and a joint common $\chi_{xF;it}$ (resp. $\chi_{xG;jt} = \chi_{G;jt}$) component;
(b) (French block analysis) an analysis based on spectral estimation in the subpanel consisting of the $\chi_{xF;it}$'s obtained under (a) (two factors) decomposes the French joint common $\chi_{xF;it}$ into a French-common component $\chi_{F;it}$ (coinciding, since $\mathcal{H}^{\chi}_F \subset \mathcal{H}^{\chi}_G$ with the strongly common one $\phi_{F;it}$) and a French-weakly-idiosyncratic one $\nu_{F;it}$; the same projection decomposes the German joint common component $\chi_{xG;jt} = \chi_{G;jt}$ into a strongly common $\phi_{G;jt}$ and a weakly common $\psi_{G;jt}$ (we already know that $\nu_{G;jt} = 0$).

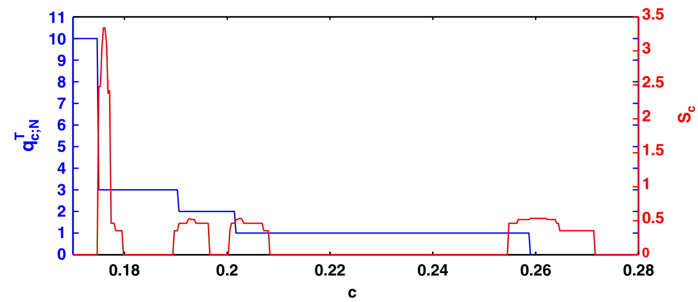
The results, along with the corresponding percentages of explained variance, are provided in Fig. 2. For each of the four mutually orthogonal subspaces appearing in the decomposition, we provide the percentage of total variation explained in each country. The two strongly common (Franco-German) factors jointly account for 3.5% only of the German total variability, but 14.6% of the French total variability. Germany's "all-German" (French-idiosyncratic) common factor explains 22.5% of Germany's total variance. Although French-idiosyncratic, that German factor nevertheless still accounts for 8.5% of the French total variability. These estimated percentages of explained variation were obtained via estimated eigenvectors and eigenvalues.



(a) France.



(b) Germany.



(c) France & Germany.

Fig. 1. Identification of the numbers of factors for the France–Germany Industrial Production dataset. The three figures show the simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^T_{c,n}$ needed for this identification, ((a) and (b)) in the marginal French and German subpanels, and (c) in the complete panel, respectively.

6.2. A three-block analysis

Next, consider the three-block case resulting from adding to the French and German data the corresponding Italian Industrial Production indices, with $n_I = 91$, yielding a panel with $K = 3$ blocks. The series length is still $T = 143$. Adapting the notation of Section 5, let $X_{F;it}$, $X_{G;it}$ and $X_{I;it}$ correspond to the French, the German, and the Italian subpanels, respectively.

From the global panel ($n = n_{FGI} = 301$ series), we can extract six subpanels—the three panels we already analyzed in Section 6.1 (the two-block French–German panel, the French and the German one-block subpanels), one new one-block subpanel (the marginal Italian one, with $n_I = 91$), two new two-block subpanels (the French–Italian one, with $n_{FI} = 187$ and the German–Italian one, with $n_{GI} = 205$, respectively). Analyzing these new subpanels along the same lines as in the previous section (with, using obvious notation, $n_{I,j} = 91 - 2j$, $j = 1, \ldots, 5$, $n_{GI,j} = 191 - 2j$, $j = 1, \ldots, 8$, $n_{FI,j} = 187 - 2j$, $j = 1, \ldots, 8$, and $n_{FGI,j} = 301 - 2j$, $j = 1, \ldots, 15$), still with $M_T = 0.5\sqrt{T} = 5$, the same penalty function and the same $q_{max} = 10$ as before, we obtain the identification results shown in the four graphs of Fig. 3.

These graphs again very clearly identify a total number of $q^T_{\mathbf{n},FGI} = 4$ joint common factors (for $c \in [0.1710, 0.1718]$), $q^T_{(nF,n_I),FI} = 3$ (for $c \in [0.1838, 0.1886]$) French–Italian, and $q^T_{(n_G,n_I),GI} = 4$ (for $c \in [0.1786, 0.1800]$) German–Italian marginal "binational" factors, and $q^T_{n_I,I} = 2$ (for $c \in [0.2118, 0.22218]$) marginal Italian factors. Along with the figures obtained in Section 6.1 for France and Germany, this implies that $\mathcal{H}^\chi_F \subset \mathcal{H}^\chi_G$, hence $\mathcal{H}^\chi_F \cap \mathcal{H}^\xi_G = \{0\} = \mathcal{H}^\chi_{GI} \cap \mathcal{H}^\xi_F$. The relations between those various (dynamic) dimensions are easily obtained; for instance, $q_{(n_F,n_G),FG} = q_{n_F,F} + q_{n_G,G} - q_{(n_F,n_G)}$, a relation we already used in Section 6.1, or

$$q_{(n_F,n_G),FG} = q_{n_F,F} + q_{n_G,G} + q_{n_I,I} - q_{(n_F,n_G),FG} - q_{(n_F,n_I),FI} - q_{(n_G,n_I),GI} + q_{(n_F,n_G,n_I),FGI}.$$

These relations imply that the three countries share one strongly common factor. As already noted, France (two factors) has no

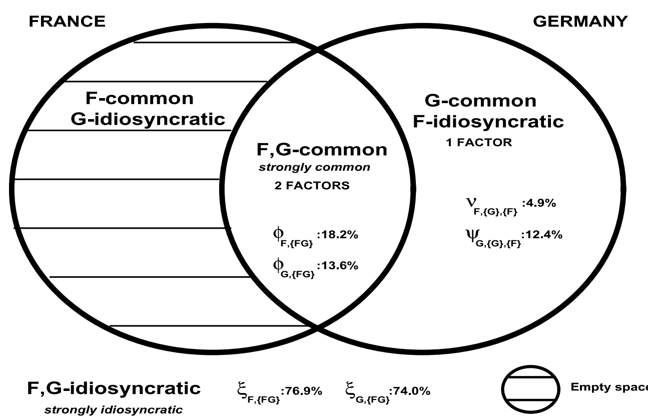


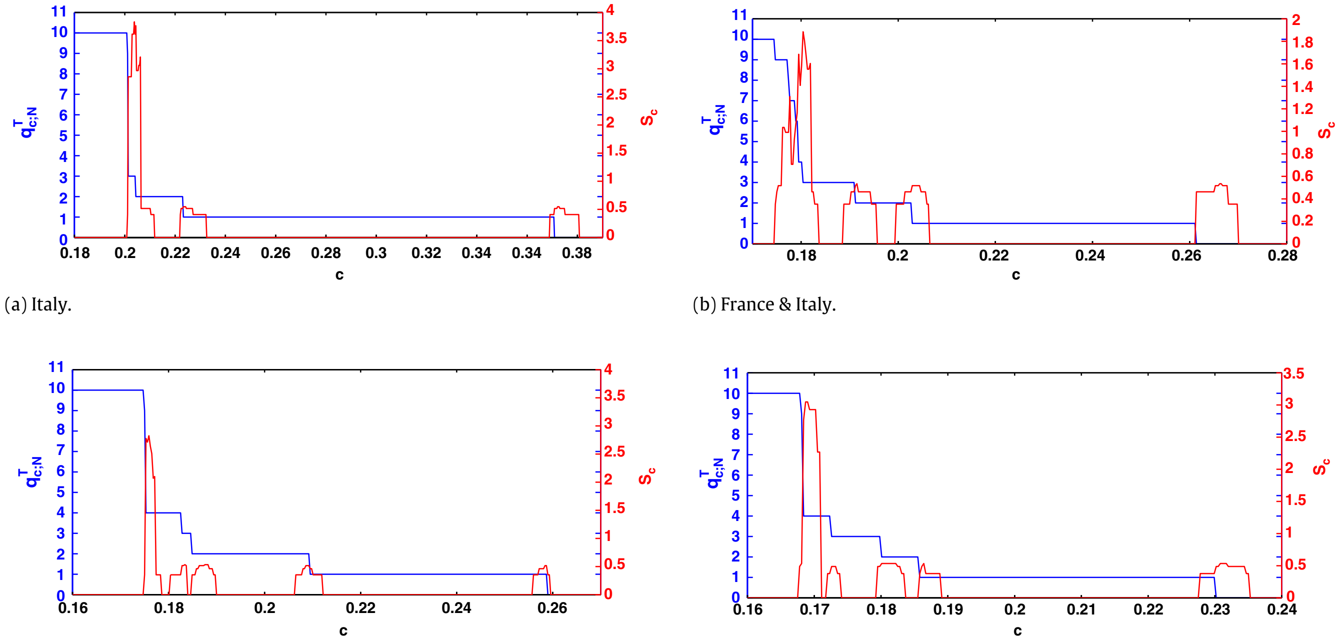
Fig. 2. Decomposition of the France–Germany panel data into four mutually orthogonal components, with the corresponding percentages of explained variation.

specific common factor, but one (the strongly common one) shared with Germany and Italy, and one shared with Germany alone. Both Italy (two factors) and Germany (three factors) have a "national" factor. Italy's "non-national" factor is the strongly common one; Germany's "non-national" factors are those shared with France, and include the strongly common one. The Italian and German "national" factors need not be mutually orthogonal.

Proceeding with the various projections described in Section 5, we successively obtain

(a) (global analysis) a four-factor analysis based on spectral estimation in the global panel: projecting onto the resulting reconstruction of $\mathcal{H}^\xi_{FGI}$ and $\mathcal{H}^\chi_{FGI}$ decomposes $X_{F;it}$, $X_{G;it}$, and $X_{I;it}$ into their strongly idiosyncratic components $\xi_{F;\{FGI\};it}$, $\xi_{G;\{FGI\};it}$, and $\xi_{I;\{FGI\};it}$, with respective orthogonal complements $\chi_{F;\{FGI\};it}$, $\chi_{G;\{FGI\};it}$, and $\chi_{I;\{FGI\};it}$;

(b1) (French–German block analysis) a three-factor analysis based on spectral estimation in the subpanel consisting of the $\chi_{F;\{FGI\};it}$'s and $\chi_{G;\{FGI\};it}$'s obtained under (a): projection onto the resulting reconstructions of $\mathcal{H}^\chi_{FG} = \mathcal{H}^\chi_G$ decomposes



(a) Italy.

(b) France & Italy.

(c) Germany & Italy.

(d) France & Germany & Italy.

Fig. 3. Identification of the numbers of factors for the French–German–Italian Industrial Production dataset. The four figures show the simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^T_{c,n}$ needed for this identification: (a) for the marginal Italian subpanel, ((b) and (c)) for the France–Italy and Germany–Italy subpanels, and (d) for the complete three-country panel, respectively.

$\chi_{F;\{FGI\};it}$ into $\chi_{F;\{FG\};it}$ and $\nu_{F;\{I\},\{FG\};it}$, and $\chi_{G;\{FGI\};it}$ into $\chi_{G;\{FG\};it}$ and $\nu_{G;\{I\},\{FG\};it}$, respectively;

(b2) (French–Italian block analysis) a three-factor analysis, similar to that in (b1), based on spectral estimation in the subpanel consisting of the $\chi_{F;\{FGI\};it}$'s and $\chi_{I;\{FGI\};it}$'s obtained under (a): projection onto the resulting reconstruction of $\mathcal{H}^{\chi}_{FI}$ decomposes $\chi_{I;\{FGI\};it}$ into $\chi_{I;\{FI\};it}$ and $\nu_{I;\{G\},\{FI\};it}$;

(c1) (French block analysis) a two-factor analysis, similar to step (b) in Section 6.1: projecting $\chi_{F;\{FG\};it}$ obtained in (b1) onto $\mathcal{H}^{\chi}_{F}$ yields $\chi_{F;\{F\};it} = \chi_{F;\{F\cap G\};it}$ and $\nu_{F;\{G\},\{FG\};it}$; for $\chi_{G;\{FG\};it}$, the same projection actually coincides with a projection onto $\mathcal{H}^{\chi}_{F\cap G}$, yielding $\phi_{G;\{F\cap G\};it}$ and $\psi_{G;\{FG\},\{F\cap G\};it} = \psi_{G;\{FG\},\{F\};it}$;

(c2) (Italian block analysis) a two-factor analysis based on spectral estimation in the subpanel consisting of the $\chi_{I;\{FI\};it}$'s obtained in step (b2): projecting them onto the resulting reconstruction of $\mathcal{H}^{\chi}_{I}$ yields $\chi_{I;\{I\};it}$ and $\nu_{I;\{FG\},\{I\};it}$;

(d) (French-and-Italian block analysis) a final projection of $\chi_{F;\{F\};it}$ and $\phi_{G;\{F\cap G\};it}$ obtained in (c1), and $\chi_{I;\{I\};it}$ obtained in (c2) onto $\mathcal{H}^{\chi}_{F\cap G\cap I} = \mathcal{H}^{\chi}_{F\cap I}$ yields the strongly common components $\phi_{F;\{F\cap G\cap I\};it}$, $\phi_{G;\{F\cap G\cap I\};it}$ and $\phi_{I;\{F\cap G\cap I\};it}$, along with the weakly common ones $\psi_{F;\{F\},\{F\cap G\cap I\};it}$, $\psi_{G;\{F\},\{F\cap G\cap I\};it}$, and $\psi_{I;\{I\},\{F\cap G\cap I\};it}$.

Conclusions are summarized in the diagram of Fig. 4, along with the various percentages of explained variances. Inspection of that diagram reveals that the three countries all exhibit a high percentage of about 71% of strongly idiosyncratic variation. France and Italy, with about 9% of strongly common variation, are the "most strongly common" in the group of three.

Acknowledgements

Marc Hallin gratefully acknowledges the support of the Sonderforschungsbereich "Statistical modelling of nonlinear dynamic processes" (SFB 823) of the Deutsche Forschungsgemeinschaft and a Discovery Grant of the Australian Research Council. Part of this work was completed while visiting the Economics Department at the European University Institute in Florence under a Fernand Braudel Fellowship. The authors are grateful to Christine De Mol, Richard Spady, Marco Lippi and Mario Forni for many stimulating discussions and helpful suggestions. They also thank Alexei Onatski for pointing out a flaw in an earlier stage of the manuscript, as well as Franz Palm, Jean-Pierre Urbain, and two anonymous referees for their comments on the original version of the paper.

Appendix. Proof of Lemma 1

Denote by $\bar{\Theta}_y$ the set (with Lebesgue measure zero) of θ values for which divergence in Assumption A2(i) does not hold. Similarly define $\bar{\Theta}_z$, and let $\bar{\Theta} := \bar{\Theta}_y \cup \bar{\Theta}_z$: $\bar{\Theta}$ also has Lebesgue measure zero. Since $\mathbf{\Sigma}_{y;n_y}(\theta)$ is a principal submatrix of $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$, a classical result (see Corollary 1, page 293, in Lancaster and Tismenetsky, 1985) implies that, for any $\mathbf{n} = (n_y, n_z)$ and $\theta \in [-\pi, \pi]$, $\lambda_{y;n_y,i}(\theta) \le \lambda_{\mathbf{n},i}(\theta)$, $i = 1, \dots, n_y$. Since $\lambda_{y;n_y,q_y}(\theta)$ diverges for all $\theta \in \Theta$ as $n_y \to \infty$, so does $\lambda_{\mathbf{n},q_y}(\theta)$. A similar result also holds for the $\lambda_{z;n_z,j}$'s, so that, for all $\theta \in \Theta$, $\lambda_{\mathbf{n},\max(q_y,q_z)}(\theta)$ diverges as $\min(n_y, n_z) \to \infty$. Note that the same result by Lancaster and Tismenetsky (1985) also implies that, for all θ and k, $\lambda_{\mathbf{n},k}(\theta)$ is a monotone nondecreasing function of both n_y and n_z and, therefore, either is bounded or goes to infinity as either n_y or $n_z \to \infty$.

Next, let us show that $\lambda_{\mathbf{n},q_y+q_z+1}(\theta)$ is bounded as $\min(n_y, n_z) \to \infty$, for all $\theta \in \Theta$. For all $\theta \in \Theta$, consider the sequences of n-dimensional vectors $\boldsymbol{\zeta}_{\mathbf{n}}(\theta) := (\boldsymbol{\zeta}'_{y;n_y}(\theta), \boldsymbol{\xi}'_{z;n_z}(\theta))'$ which are orthogonal to the $q_y + q_z$ vectors $(\mathbf{p}'_{y;n_y,1}(\theta), 0, \dots, 0)', \dots, (\mathbf{p}'_{y;n_y,q_y}(\theta), 0, \dots, 0)'$ and $(0, \dots, 0, \mathbf{p}'_{z;n_z,1}(\theta))', \dots, (0, \dots, 0, \mathbf{p}'_{z;n_z,q_z}(\theta))'$. The collection of all such $\boldsymbol{\xi}_{\mathbf{n}}$'s is a linear subspace $\boldsymbol{\Xi}_{\mathbf{n}}(\theta)$ of dimension at least $n - q_y - q_z$. For any such $\boldsymbol{\xi}_{\mathbf{n}}(\theta)$, in view of the orthogonality of $\boldsymbol{\xi}_{y;n_y}(\theta)$ and $\mathbf{p}_{y;n_y,1}(\theta), \dots, \mathbf{p}_{y;n_y,q_y}(\theta)$ (resp., of $\boldsymbol{\xi}_{z;n_z}(\theta)$ and $\mathbf{p}_{z;n_z,1}(\theta), \dots, \mathbf{p}_{z;n_z,q_z}(\theta)$),

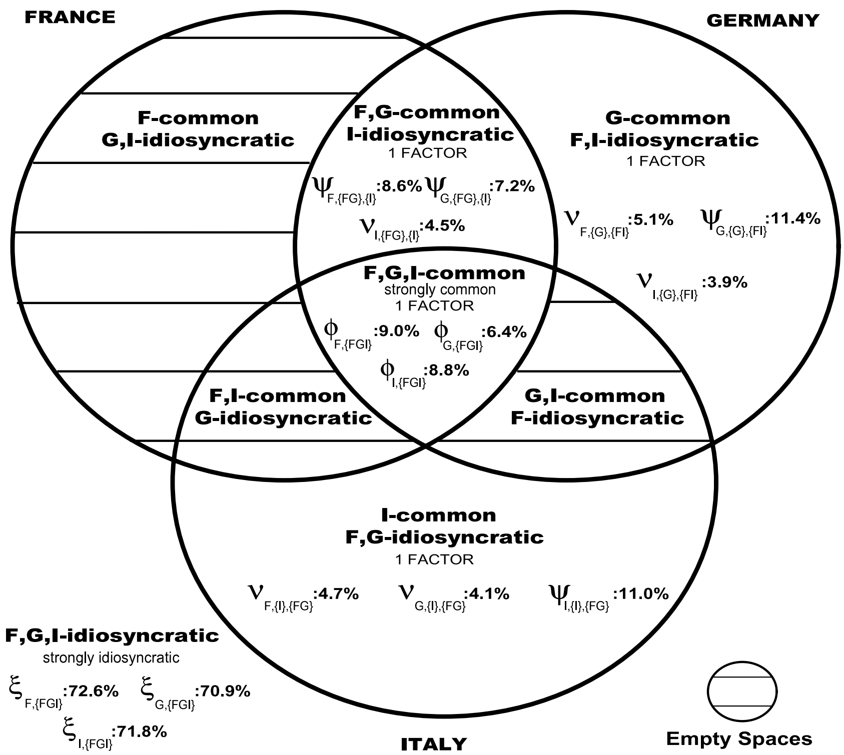


Fig. 4. Decomposition of the France–Germany–Italy panel data into eight components, with the corresponding percentages of explained variation.

$$
\begin{aligned}
&\|\boldsymbol{\xi}_{\mathbf{n}}(\theta)\|^{-2}\boldsymbol{\xi}^{*}_{\mathbf{n}}(\theta)\boldsymbol{\Sigma}_{\mathbf{n}}(\theta)\boldsymbol{\xi}_{\mathbf{n}}(\theta)\\
&\quad = \|\boldsymbol{\xi}_{\mathbf{n}}\|^{-2}\boldsymbol{\xi}^{*}_{y;n_y}(\theta)\boldsymbol{\Sigma}_{y;n_y}(\theta)\boldsymbol{\xi}_{y;n_y}(\theta)\\
&\qquad + \|\boldsymbol{\xi}_{\mathbf{n}}(\theta)\|^{-2}\boldsymbol{\xi}^{*}_{z;n_z}(\theta)\boldsymbol{\Sigma}_{z;n_z}(\theta)\boldsymbol{\xi}_{z;n_z}(\theta)\\
&\qquad + \|\boldsymbol{\xi}_{\mathbf{n}}(\theta)\|^{-2}\boldsymbol{\xi}^{*}_{y;n_y}(\theta)\boldsymbol{\Sigma}_{yz;\mathbf{n}}(\theta)\boldsymbol{\xi}_{z;n_z}(\theta)\\
&\qquad + \|\boldsymbol{\xi}_{\mathbf{n}}(\theta)\|^{-2}\boldsymbol{\xi}^{*}_{z;n_z}(\theta)\boldsymbol{\Sigma}_{zy;\mathbf{n}}(\theta)\boldsymbol{\xi}_{y;n_y}(\theta)\\
&\quad \le 2(\|\boldsymbol{\xi}_{y;n_y}(\theta)\|^{-2}\boldsymbol{\xi}^{*}_{y;n_y}(\theta)\boldsymbol{\Sigma}_{y;n_y}\boldsymbol{\xi}_{y;n_y}(\theta)\\
&\qquad + \|\boldsymbol{\xi}_{z;n_z}(\theta)\|^{-2}\boldsymbol{\xi}^{*}_{z;n_z}(\theta)\boldsymbol{\Sigma}_{z;n_z}\boldsymbol{\xi}_{z;n_z}(\theta))\\
&\quad \le 2(\lambda^2_{y;n_y,q_y+1}(\theta)+\lambda^2_{z;n_z,q_z+1}(\theta))
\end{aligned}
$$

for all $\theta \in \Theta$ and $\mathbf{n} = (n_y, n_z)$. Since $\lambda^2_{y;n_y,q_y+1}(\theta)$ and $\lambda^2_{z;n_z,q_z+1}(\theta)$ are bounded, for any $\theta \in \Theta$, as $\min(n_y, n_z) \to \infty$, so is $\boldsymbol{\xi}^{*}_{\mathbf{n}}(\theta)\boldsymbol{\Sigma}_{\mathbf{n}}(\theta)\boldsymbol{\xi}_{\mathbf{n}}(\theta)$. Hence, for all $\theta \in \Theta$ and $\mathbf{n} = (n_y, n_z)$, $\boldsymbol{\Xi}_{\mathbf{n}}$ (with dimension at least $n - q_y - q_z$) is orthogonal to any eigenvector associated with a diverging sequence of eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{n}}(\theta)$. It follows that the number of such eigenvalues cannot exceed $q_y + q_z$. Summing up, for all $\theta \in \Theta$, the number of diverging eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{n}}(\theta)$ is finite – denote it by q – and comprised between $\max(q_y, q_z)$ and $q_y + q_z$, as was to be shown. □

References

d'Agostino, A., Giannone, D., 2005. Comparing alternative predictors based on large-panel dynamic factor models. ECB Working Paper 680.
Aiolfi, M., Catão, L., Timmerman, A., 2006. Common factors in Latin America's business cycle. Working Paper 06/49. International Monetary Fund.
Altissimo, F., Bassanetti, A., Cristadoro, R., Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2001. A real time coincident indicator of the euro area business cycle. CEPR Discussion Paper 3108.
Angelini, E., Henry, J., Mestre, R., 2001. Diffusion index based inflation forecasts for the euro area. ECB Working Paper 61.
Artis, M., Banerjee, A., Marcellino, M., 2005. Factor forecasts for the UK. Journal of Forecasting 24, 279–298.
Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.
Bai, J., Ng, S., 2007. Determining the number of primitive shocks in factor models. Journal of Business and Economic Statistics 25, 52–60.
Bernanke, B.S., Boivin, J., 2003. Monetary policy in a data rich environment. Journal of Monetary Economics 50, 525–546.
Brillinger, D.R., 1981. Time Series: Data Analysis and Theory. Holden-Day, San Francisco.
Bruneau, C., De Bandt, A., Flageollet, A., Michaux, E., 2007. Forecasting inflation using economic indicators: the case of France. Journal of Forecasting 26, 1–22.
Campbell, J.Y., Lo, A.W., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton.
Chamberlain, G., 1983. Funds, factors, and diversification in arbitrage pricing models. Econometrica 51, 1281–1304.
Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure and mean-variance analysis in large asset markets. Econometrica 51, 1305–1324.
Den Reijer, A.H.J., 2005. Forecasting Dutch GDP using large scale factor models. DNB Working Paper 028. De Nederlandsche Bank, Research Department.
Dreger, C., Schumacher, C., 2004. Estimating large scale factor models for economic activity in Germany. In: Wagner, A. (Ed.), Jahrbücher für Nationalökonomie und Statistik. Lucius & Lucius Verlagsgesellschaft, Stuttgart, pp. 731–750.
Favero, C., Marcellino, M., Neglia, F., 2005. Principal components at work: empirical analysis of monetary policy with large datasets. Journal of Applied Econometrics 20, 603–620.
Forni, M., Lippi, M., 2001. The generalized factor model: representation theory. Econometric Theory 17, 1113–1141.
Forni, M., Reichlin, L., 1998. Let's get real: a factor analytical approach to disaggregated business cycle dynamics. Review of Economic Studies 65, 453–473.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic factor model: identification and estimation. Review of Economics and Statistics 82, 540–554.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2003. Do financial variables help forecasting inflation and real activity in the euro area? Journal of Monetary Economics 50, 1243–1255.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2004. The generalized dynamic factor model: consistency and rates. Journal of Econometrics 119, 231–255.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2005. The generalized dynamic factor model: one-sided estimation and forecasting. Journal of the American Statistical Association 100, 830–840.
Forni, M., Giannone, D., Lippi, M., Reichlin, L., 2009. Opening the black box: structural factor models with large cross sections. Econometric Theory 25, 1319–1347.
Geweke, J., 1977. The dynamic factor analysis of economic time series. In: Aigner, D.J., Goldberger, A.S. (Eds.), Latent Variables in Socio-Economic Models. North-Holland, Amsterdam.
Giannone, D., Matheson, T., 2007. A new core inflation index for New Zealand. International Journal of Central Banking 3, 145–180.
Giannone, D., Reichlin, L., Sala, L., 2005. Monetary policy in real time. In: Gertler, M., Rogoff, K. (Eds.), NBER Macroeconomic Annual 2004. MIT Press, Cambridge, Mass, pp. 161–200.

Giannone, D., Reichlin, L., Sala, L., 2006. VARs, factor models and the empirical validation of equilibrium business cycle models. Journal of Econometrics 132, 257–279.

Hallin, M., Liška, R., 2007. Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association 102, 603–617.

Hallin, M., Mathias, Ch., Pirotte, H., Veredas, D., 2011. Market liquidity as dynamic factors. Journal of Econometrics 163, 42–50.

Ingersol, J., 1984. Some results in the theory of arbitrage pricing. The Journal of Finance 39, 1021–1039.

Lancaster, P., Tismenetsky, M., 1985. The Theory of Matrices, 2nd ed. Academic Press, Orlando.

Marcellino, M., Stock, J., Watson, M.W., 2003. Macroeconomic forecasting in the euro area: country specific versus area wide information. European Economic Review 47, 1–18.

Ng, S., Moench, E., 2009. A hierarchical factor analysis of US housing market. Econometrics Journal (in press).

Ng, S., Moench, E., Potter, S., 2008. Dynamic hierarchical factor models. Mimeo. Columbia University.

Nieuwenhuyzen, C., 2004. A generalized dynamic factor model for the Belgian economy. Journal of Business Cycle Measurement and Analysis 2, 213–248.

Robinson, P., 1991. Automatic frequency domain inference on semiparametric and nonparametric problems. Econometrica 59, 1329–1363.

Sargent, T.J., Sims, C.A., 1977. Business cycle modelling without pretending to have too much a priori economic theory. In Sims, C.A. (Ed.), New Methods in Business Research. Federal Reserve Bank of Minneapolis, Minneapolis.

Schneider, M., Spitzner, M., 2004. Forecasting Austrian GDP using the generalized dynamic factor model. Working Paper 89. Österreichische Nationalbank. Vienna.

Schumacher, C., 2007. Forecasting German GDP using alternative factor models based on large data sets. Journal of Forecasting 26, 271–302.

Stock, J.H., Watson, M.W., 1989. New indices of coincident and leading indicators. In: Blanchard, O.J., Fischer, S. (Eds.), NBER Macroeconomics Annual 1989. MIT Press, Cambridge, Mass.

Stock, J.H., Watson, M.W., 2002a. Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics 20, 147–162.

Stock, J.H., Watson, M.W., 2002b. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Stock, J.H., Watson, M.W., 2005. Implications of dynamic factor models for VAR analysis. NBER Working Paper 11467.

Yao, T., 2008. Dynamic factors and the source of momentum profits. Journal of Business & Economic Statistics 26, 211–226.