Journal of Econometrics 199 (2017) 74–92

Contents lists available at ScienceDirect

Journal of Econometrics

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

Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis[☆]

Mario Forni[a], Marc Hallin[b,*], Marco Lippi[c], Paolo Zaffaroni[d,e]

[a] Università di Modena e Reggio Emilia, CEPR and RECent, Italy
[b] ECARES, Université libre de Bruxelles CP 114/4, 50 Avenue F.D. Roosevelt, B-1050 Bruxelles, Belgium
[c] Einaudi Institute for Economics and Finance, Roma, Italy
[d] Imperial College London, United Kingdom
[e] Università di Roma La Sapienza, Italy

ARTICLE INFO

Article history:
Received 5 April 2016
Received in revised form 29 November 2016
Accepted 4 April 2017
Available online 24 April 2017

JEL classification:
C0
C01
E0

Keywords:
High-dimensional time series
Generalized dynamic factor models
Vector processes with singular spectral density
One-sided representations of dynamic factor models
Consistency and rates

ABSTRACT

Factor models, all particular cases of the Generalized Dynamic Factor Model (GDFM) introduced in Forni et al., (2000), have become extremely popular in the theory and practice of large panels of time series data. The asymptotic properties (consistency and rates) of the corresponding estimators have been studied in Forni et al. (2004). Those estimators, however, rely on Brillinger's concept of *dynamic* principal components, and thus involve two-sided filters, which leads to rather poor forecasting performances. No such problem arises with estimators based on standard (*static*) principal components, which have been dominant in this literature. On the other hand, the consistency of those *static* estimators requires the assumption that the space spanned by the factors has finite dimension, which severely restricts their generality—prohibiting, for instance, autoregressive factor loadings. This paper derives the asymptotic properties of a semiparametric estimator of the loadings and common shocks based on one-sided filters recently proposed by Forni et al., (2015). Consistency and exact rates of convergence are obtained for this estimator, under a general class of GDFMs that does not require a finite-dimensional factor space. A Monte Carlo experiment and an empirical exercise on US macroeconomic data corroborate those theoretical results and demonstrate the excellent performance of those estimators in out-of-sample forecasting.

1. Introduction

This paper provides consistency results and consistency rates for the estimators recently proposed by Forni et al. (2015) (hereafter, FHLZ) for the *Generalized Dynamic Factor Model* (GDFM). Let

$$\{x_{it},\ 1 \le i \le n_0,\ 1 \le t \le T_0\} \tag{1.1}$$

be an observed $(n_0 \times T_0)$-dimensional panel, namely, a n_0-tuple of time series observed over a time period of length T_0. The GDFM, as introduced in Forni et al. (2000) and Forni and Lippi (2001), consists in modeling that panel as a finite realization of a stochastic process of the form $\{x_{it},\ i \in \mathbb{N},\ t \in \mathbb{Z}\}$, that is, a countable number of univariate processes $\{x_{it},\ t \in \mathbb{Z}\}$, admitting a decomposition of the form

$$x_{it} = \chi_{it} + \xi_{it} = b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} + \xi_{it},\quad i \in \mathbb{N},\ t \in \mathbb{Z}, \tag{1.2}$$

where $\mathbf{u}_t = (u_{1t}\ u_{2t}\ \cdots\ u_{qt})'$ is unobservable q-dimensional orthonormal white noise and the filters $b_{if}(L), i \in \mathbb{N},\ f = 1, \ldots, q$, are square-summable ($L$, as usual, stands for the lag operator). The unobservable processes χ_{it} and ξ_{it} are called the *common* and *idiosyncratic components*, respectively. Detailed assumptions on (1.2) are given below. Let us only recall here that the idiosyncratic components ξ_{it} and the *common shocks* u_{ft}, also called *dynamic factors*, are mutually orthogonal at any lead and lag, and that the idiosyncratic components are "weakly" cross-correlated in a sense to be defined below—cross-sectional orthogonality being an extreme case. The generality of such a representation has been stressed in Hallin and Lippi (2013).

☆ We are grateful to Michael Eichler and Giovanni Motta for suggestions and constructive criticism on early versions of this research, to Wei Biao Wu for helping us with crucial issues in multivariate spectral estimation and to Alessandro Giovannelli for sharing with us his experience in forecasting.

* Correspondence to: ECARES, Université libre de Bruxelles CP114/4, 50 ave. F.D. Roosevelt, B-1050 Bruxelles, Belgium

E-mail address: mhallin@ulb.ac.be (M. Hallin).

http://dx.doi.org/10.1016/j.jeconom.2017.04.002
0304-4076/

Much of the literature on Dynamic Factor Models is based on (1.2) under the assumption that the space spanned by the stochastic variables χ_{it}, for t given and $i \in \mathbb{N}$, is *finite-dimensional*.[1] Under that assumption, model (1.2) can be rewritten in the so-called *static representation*

$$\begin{aligned} x_{it} &= \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \cdots + \lambda_{ir}F_{rt} + \xi_{it} \\ \mathbf{F}_t &= (F_{1t} \ \dots \ F_{rt})' = \mathbf{N}(L)\mathbf{u}_t. \end{aligned} \tag{1.3}$$

The variables $F_{jt}, j = 1, 2, \dots, r$ are usually called the *static factors*, as opposed to the dynamic factors u_{jt}. Criteria to determine r consistently have been given in Bai and Ng (2002) and, more recently, in Alessi et al. (2010), Onatski (2010), and Ahn and Horenstein (2013). The vectors $\mathbf{F}_t$ and the loadings λ_{ij} can be estimated consistently using the first r standard principal components, see Stock and Watson (2002a, b), Bai and Ng (2002). Moreover, the second equation in (1.3) is usually specified as a (possibly singular) VAR, so that (1.3) takes the form

$$\begin{aligned} x_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \cdots + \lambda_{ir}F_{rt} \quad &+ \quad \xi_{it} \\ \mathbf{D}(L)\mathbf{F}_t = (\mathbf{I} - \mathbf{D}_1 L - \mathbf{D}_2 L^2 - \cdots - \mathbf{D}_p L^p)\mathbf{F}_t \quad &= \quad \mathbf{K}\mathbf{u}_t, \end{aligned} \tag{1.4}$$

where the matrices $\mathbf{D}_j$ are $r \times r$ while $\mathbf{K}$ is $r \times q$, $r \geq q$. Under (1.4), Bai and Ng (2007) and Amengual and Watson (2007) provide consistent criteria to determine q. We refer to estimators and predictors based on the existence of the static representation (1.3) and standard principal components as the *static method*, as opposed to the method developed in FHLZ and the present paper, referred to as the *dynamic method*.

The assumption of a finite-dimensional factor space, however, is far from being innocuous. For instance, (1.3) is so restrictive that even the very elementary model

$$x_{it} = a_i(1 - \alpha_i L)^{-1}u_t + \xi_{it}, \tag{1.5}$$

where $q = 1$, u_t is scalar white noise, and the coefficients α_i are drawn from a uniform distribution over the stationary region, is ruled out. In this case, the space spanned, for given t, by the common components χ_{it}, $i \in \mathbb{N}$, is easily seen to be infinite-dimensional unless the α_is take only a finite number of values.

On the other hand, in the absence of the finite-dimensionality assumption, estimation of model (1.2) cannot be based on a finite number r of standard principal components. That situation is the one studied in Forni et al. (2000) and (2004), who are using q principal components in the frequency domain (Brillinger's *dynamic principal components*; see Brillinger (1981)) to estimate the common components χ_{it}.[2] However, their estimators involve the application of two-sided filters acting on the observations x_{it}, and hence perform poorly at the end/beginning of the observation period. As a consequence, they are of little help for prediction.

In FHLZ, which only contains representation results, we show how one-sided filters can be obtained without the finite-dimensionality assumption, under the additional condition that the common components have *rational spectral density*, that is, each filter $b_{if}(L)$ in (1.2) is a ratio of polynomials in L. Elaborating upon recent results by Anderson and Deistler (2008a, b), FHLZ prove that, for generic values of the parameters $c_{if,k}$ and $d_{if,k}$ (i.e. apart from a lower-dimensional subset in the parameter space, see FHLZ for details), the infinite-dimensional idiosyncratic vector $\boldsymbol{\chi}_t = (\chi_{1t} \ \chi_{2t} \ \cdots \ \chi_{nt} \ \cdots)'$ admits *a unique autoregressive representation with block structure* of the form

$$\begin{pmatrix} \mathbf{A}^1(L) & 0 & \cdots & 0 & \cdots \\ 0 & \mathbf{A}^2(L) & \cdots & 0 & \\ & & \ddots & & \\ 0 & 0 & \cdots & \mathbf{A}^k(L) & \\ \vdots & & & & \ddots \end{pmatrix} \boldsymbol{\chi}_t = \begin{pmatrix} \mathbf{R}^1 \\ \mathbf{R}^2 \\ \vdots \\ \mathbf{R}^k \\ \vdots \end{pmatrix} \mathbf{u}_t, \tag{1.6}$$

where $\mathbf{A}^k(L)$ is a $(q+1) \times (q+1)$ polynomial matrix *with finite degree* and $\mathbf{R}^k$ is $(q+1) \times q$.

The contribution of the present paper is the construction of estimators for $\mathbf{A}^k(L)$, $\mathbf{R}^k$ and $\mathbf{u}_t$, and their asymptotic analysis (consistency and rates).

Our consistency results are based on recent advances on spectral estimation: Shao and Wu (2007) and Liu and Wu (2010), as extended to the multivariate case by Wu and Zaffaroni (2017). These papers prove that lag-window estimators of spectra and cross-spectra, under quite general assumptions on the processes and the kernel, are consistent, as $T \to \infty$, *uniformly* with respect to the frequency θ, with rate $\sqrt{T^{-1}B_T \log B_T}$, where B_T is the size of the lag window.

Exploiting those results here requires some enhancement of the FHLZ assumptions on the common shocks and the idiosyncratic components. For example, the vector $\mathbf{u}_t$, which is second-order white noise in FHLZ, is i.i.d. here. This, as well as other changes in the FHLZ assumptions, is discussed in detail in Section 2. Under this enhanced set of assumptions, we prove that the estimators of $\mathbf{A}^k(L)$, $\mathbf{R}^k$ and $\mathbf{u}_t$ are consistent with rate

$$\zeta_{nT} = \max\left(\sqrt{n^{-1}}, \sqrt{T^{-1}B_T \log B_T}\right), \tag{1.7}$$

where B_T diverges as T^δ, $1/3 < \delta < 1$.

As pointed out in FHLZ (end of Section 4.5), despite the fact that the dynamic model studied in this paper is more general than model (1.4), when a dataset with finite $n = n_0$ and $T = T_0$ is given, the static approach might perform well even though the required finite-dimension assumptions are not satisfied. A Monte Carlo study is discussed in Section 4, in which the static and dynamic methods have been applied to simulated data. A very short summary of our results is that (i) when the data are generated by infinite-dimensional models which are simple generalizations of (1.5), the estimation of impulse–response functions and predictions via the dynamic method is by far better than those obtained via the static one; (ii) even when the data are generated by (1.4), the dynamic method still performs slightly better. Though not conclusive, our Monte Carlo results strongly suggest that the FHLZ method may be uniformly competitive. A pseudo out-of-sample forecasting exercise with US quarterly macroeconomic series provides further evidence in favor of the dynamic method, see Section 5.

Lastly, the rate (1.7) should be compared to $\max(\sqrt{n^{-1}}, \sqrt{T^{-1}})$, which is standard in the literature assuming the existence of a static representation, see e.g. Bai and Ng (2002) and Forni et al. (2009), the diverging factor $B_T \log B_T$ being the price we pay to non-parametric estimation of the spectral density of the xs. However, both the Monte Carlo and the empirical results, presented in Sections 4 and 5 respectively, suggest that the slower rate of our estimators has no consequence on their precision for the typical size of macroeconomic datasets.

The paper is organized as follows. In Section 2, we present and comment the main assumptions to be made throughout. Section 3 provides the main asymptotic results. Sections 4 and 5 contain a detailed description and analysis of the Monte Carlo experiments and the empirical exercise respectively. Section 6 concludes. Proofs are concentrated in Appendix A–E. Online Appendices F and G provide the tables of results for the numerical studies of Sections 4 and 5.

[1] The definition of χ_{it} obviously implies that this dimension does not depend on t.

[2] Criteria to determine q without assuming (1.3) or (1.4) are obtained in Hallin and Liška (2007) and Onatski (2009).

2. Main assumptions and some preliminary results

The assumptions in this section reproduce those in FHLZ with some important additional specifications. Assumption 1 decomposes the xs into common and idiosyncratic components. The common components are driven by the q-dimensional i.i.d. vector of common shocks $\mathbf{u}_t$ via rational filters. The idiosyncratic components, unlike in FHLZ, are modeled here as moving averages of an infinite-dimensional i.i.d. vector $\boldsymbol{\eta}_t = (\eta_{1t}\ \eta_{2t}\ \cdots)'$. Assuming that $\mathbf{u}_t$ is i.i.d. instead of second-order white noise (as in FHLZ), as well as modeling the idiosyncratic components, is necessary for the assumptions and results on estimation. Assumption 2 imposes standard conditions on the rational functions in the common components. Assumption 3 is the standard condition on the eigenvalues of the spectral density of the common components as n, the number of series, tends to infinity, see FHLZ, enhanced with their separation, which is necessary in the consistency proof, see in particular Lemma 3, Appendix B. Assumption 4 imposes that serial and cross-sectional dependence of the idiosyncratic components both decline geometrically. As a consequence, the first eigenvalue of the spectral density of the idiosyncratic components is bounded as n tends to infinity, which is the definition of idiosyncratic components in FHLZ.

Assumption 5 is borrowed, together with its motivation, from FHLZ. Its consequence is the existence of a transformation of the dynamic model into a static one. Assumption 6 imposes divergence and separation of the eigenvalues of the common components in the static form, Assumption 7 is used to obtain boundedness of the eigenvalues of the idiosyncratic components.

Assumption 8 imposes a bound on the pth moments of $u_{ft}, f = 1, 2, \ldots, q$, and those of $\eta_{jt}, j \in \mathbb{N}$, uniform with respect to f and j, with $p > 4$. Together with Assumption 9, it implies Proposition 6, which is crucial, stating that the estimated cross-spectral density of x_{it} and x_{jt} converges uniformly with respect to the frequency θ, i and j.

For a comparison of our assumptions with those in the literature assuming the existence of a static representation we consider here Forni et al. (2009), whose setting is closest to the present paper. Assumptions 1 through 4 in the present paper, ensuring pervasiveness and non-pervasiveness of common and idiosyncratic components respectively, closely correspond to Assumptions 1 through 5 in Forni et al. (2009). However, (1) here we model the idiosyncratic components as linear combinations of idiosyncratic shocks, (2) we assume independence, not mere orthogonality, of common and idiosyncratic components, and (3) i.i.d.–ness, not mere whiteness, of the shocks driving common and idiosyncratic components. Assumption 5 of the present paper is borrowed from FHLZ and has no counterpart in Forni et al. (2009). Assumptions 6 and 7 entail pervasiveness and non-pervasiveness for the static model derived from Assumption 5, and slightly enhance Assumptions 1 through 4. Lastly, Forni et al. (2009) in their Assumption 8 *require* that

$$E\left[(\hat{\gamma}^x_{ij} - \gamma^x_{ij})^2\right] < \rho T^{-1}, \tag{2.1}$$

for all $i,\ j,\ T$, where γ^x_{ij} and $\hat{\gamma}^x_{ij}$ denote the covariance and estimated covariance between x_{it} and x_{jt}. The corresponding statement here is inequality (3.4), which is *obtained* as a consequence of "deep" assumptions for the moments of $\mathbf{u}_t$ and $\boldsymbol{\eta}_t$ and a standard assumption on the kernel of the lag-window estimator of the spectrum (Assumptions 8 and 9 of the present paper). Note that obtaining Assumption 8 in Forni et al. (2009) as a consequence of deeper assumptions would require imposing conditions either on the moments of the xs or on those of the common and idiosyncratic components, see the observation in their footnote 2, p. 1341. Assumption 10 is purely technical.

2.1. Common and idiosyncratic components

The Dynamic Factor Model studied in the present paper is a decomposition, of the form

$$x_{it} = \chi_{it} + \xi_{it}, \quad i \in \mathbb{N},\ t \in \mathbb{Z}$$

of an observed random variable x_{it} into a nonobserved *common* component χ_{it} and a nonobserved *idiosyncratic* component ξ_{it}. Throughout, we are assuming that the family of random variables

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

satisfies the assumptions listed below as Assumptions 1 through 10.

Assumption 1. *There exist a natural number $q > 0$ and*

(1) *a q-dimensional stochastic zero-mean process $\mathbf{u}_t = (u_{1t}u_{2t}\cdots u_{q_t})'$, $t \in \mathbb{Z}$, and an infinite-dimensional stochastic process $\boldsymbol{\eta}_t = (\eta_{1t}\eta_{2t}\cdots)'$, $t \in \mathbb{Z}$;*
(2) *square-summable filters $b_{if}(L)$, $i \in \mathbb{N}, f = 1, \ldots, q$;*
(3) *coefficients $\beta_{ij,k}$, $i, j \in \mathbb{N}$, $k = 0, 1, \ldots, \infty$, where $\sum_{j=1}^{\infty}\sum_{k=0}^{\infty}\beta^2_{ij,k} < \infty$ for all $i \in \mathbb{N}$;*

such that

(i) *the vector $\mathbf{S}_t = (\mathbf{u}'_t\ \boldsymbol{\eta}'_t)'$, $t \in \mathbb{Z}$, is i.i.d. and orthonormal; in particular, $\mathrm{var}(u_{ft}) = \mathrm{var}(\eta_{jt}) = 1$, $\mathrm{cov}(u_{ft}, \eta_{j,t-k}) = 0$, $f = 1, \ldots, q, j \in \mathbb{N}, k \in \mathbb{Z}$;*
(ii)

$$\begin{aligned}\chi_{it} &= b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} = \mathbf{b}_i(L)\mathbf{u}_t \\ \xi_{it} &= \sum_{j=1}^{\infty}\sum_{k=0}^{\infty}\beta_{ij,k}\eta_{j,t-k}.\end{aligned} \tag{2.2}$$

Clearly, neither $\mathbf{u}_t$ nor the polynomials $b_{if}(L)$ are identified. Indeed, for any orthogonal matrix $\mathbf{Q}$, the common component χ_{it} has the alternative representation

$$\chi_{it} = \left[\mathbf{b}_i(L)\mathbf{Q}^{-1}\right]\left[\mathbf{Q}\mathbf{u}_t\right] = \mathbf{b}^*_i(L)\mathbf{u}^*_t.$$

Note that (i) and (2.2) imply $\mathrm{cov}(u_{ft}, \xi_{i,t-k}) = 0$ for all $f,\ i,\ k$.

Assumption 2. *Conditions on the filters $b_{if}(L)$.*

(i) *The filters $b_{if}(L)$ are rational. More precisely, for all $i \in \mathbb{N}$ and $f = 1, \ldots, q$, there exist natural numbers $s_1 = s_{i1}$ and $s_2 = s_{i2}$ such that $b_{if}(L) = c_{if}(L)/d_{if}(L)$, where*

$$\begin{aligned}&c_{if}(L) = c_{if,0} + c_{if,1}L + \cdots + c_{if,s_1}L^{s_1} \quad \textit{and} \\ &d_{if}(L) = 1 + d_{if,1}L + \cdots + d_{if,s_2}L^{s_2}.\end{aligned} \tag{2.3}$$

(ii) *There exists $\phi > 1$ such that none of the roots of $d_{if}(L)$ is less than ϕ in modulus, for $i \in \mathbb{N}, f = 1, \ldots, q$.*

(iii) *There exists B^χ, $0 < B^\chi < \infty$, such that $|c_{if,j}| \le B^\chi$, $i \in \mathbb{N}$, $f = 1, \ldots, q, j = 0, \ldots, s_1$.*

Under Assumption 2, the vector $\boldsymbol{\chi}_{nt} = (\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{nt})'$ has a rational spectral density matrix $\Sigma^\chi_n(\theta)$; denote by $\lambda^\chi_{nj}(\theta)$ its jth eigenvalue (in decreasing order).

Assumption 3. *Common component spectral density eigenvalues: divergence and separation.*

There exist continuous functions

$$\theta \mapsto \alpha^\chi_f(\theta),\ \ f = 1, \ldots, q \quad \textit{and}$$
$$\theta \mapsto \beta^\chi_f(\theta), f = 0, \ldots, q-1, \quad \theta \in [-\pi,\ \pi],$$

and a positive integer n^χ such that, for $n > n^\chi$ and all $\theta \in [-\pi,\ \pi]$,

$$\beta^\chi_0(\theta) \ge \frac{\lambda^\chi_{n1}(\theta)}{n} \ge \alpha^\chi_1(\theta) > \beta^\chi_1(\theta) \ge \frac{\lambda^\chi_{n2}(\theta)}{n} \ge \cdots \ge \alpha^\chi_{q-1}(\theta)$$
$$> \beta^\chi_{q-1}(\theta) \ge \frac{\lambda^\chi_{nq}(\theta)}{n} \ge \alpha^\chi_q(\theta) > 0.$$

Assumption 4. *Serial and cross-sectional dependence of idiosyncratic components.*

There exist finite positive numbers B, B_{is}, $i \in \mathbb{N}$, $s \in \mathbb{N}$, and ρ, $0 \le \rho < 1$, such that

$$\sum_{s=1}^{\infty} B_{is} \le B \quad \text{for all } i \in \mathbb{N} \tag{2.4}$$

$$\sum_{i=1}^{\infty} B_{is} \le B \quad \text{for all } s \in \mathbb{N} \tag{2.5}$$

$$|\beta_{is,k}| \le B_{is}\rho^k \quad \text{for all } i, s \in \mathbb{N} \text{ and } k = 0, 1, \ldots \tag{2.6}$$

An immediate consequence of (2.4) and (2.5) is that

$$\sum_{i=1}^{\infty}\sum_{s=1}^{\infty} B_{is}B_{js} \le B^2 \quad \text{for all } j \in \mathbb{N}. \tag{2.7}$$

Conditions (2.4) and (2.5) are obviously satisfied in the "purely idiosyncratic" case $\xi_{it} = \eta_{it}$, and for finite "cross-sectional moving averages" such as $\xi_{it} = \eta_{it} + \eta_{i+1,t}$. It follows from (2.6) that the time dependence of the variables ξ_{it} declines geometrically, at common rate ρ.

Under Assumption 4, setting $\beta_{is}(L) = \sum_{k=0}^{\infty}\beta_{is,k}L^k$ and $\xi_{it} = \sum_{s=1}^{\infty}\beta_{is}(L)\eta_{st}$, and denoting by $\imath$ the imaginary unit,

$$|\boldsymbol{\beta}_{is}(e^{-\imath\theta})| = \left|\sum_{k=0}^{\infty}\beta_{is,k}e^{-\imath k\theta}\right| \le \sum_{k=0}^{\infty}|\beta_{is,k}| \le \sum_{k=0}^{\infty}B_{is}\rho^k \le B_{is}\frac{1}{1-\rho}.$$

Therefore, letting $\sigma_{ij}^{\xi}(\theta)$ denote the cross-spectral density of ξ_{it} and ξ_{jt}, by (2.7),

$$\begin{aligned}\sum_{i=1}^{\infty}|\sigma_{ij}^{\xi}(\theta)| &\le \frac{1}{2\pi}\sum_{i=1}^{\infty}\sum_{s=1}^{\infty}|\beta_{is}(e^{-\imath\theta})\overline{\beta_{js}(e^{-\imath\theta})}| \\ &\le \tfrac{1}{2\pi(1-\rho)^2}\textstyle\sum_{i=1}^{\infty}\sum_{s=1}^{\infty}B_{is}B_{js} \\ &\le B^2\frac{1}{2\pi(1-\rho)^2}.\end{aligned} \tag{2.8}$$

Assumption 4 thus implies that the cross-spectra $\sigma_{ij}^{\xi}(\theta)$ are bounded, in θ, uniformly in i and j. On the other hand, Assumption 2(ii) and (iii) imply that $\sigma_{ij}^{\chi}(\theta)$ is bounded, in θ, uniformly in i and j. Therefore, $\sigma_{ij}^{x}(\theta) = \sigma_{ij}^{\chi}(\theta)+\sigma_{ij}^{\xi}(\theta)$ is bounded, in θ, uniformly in i and j.

The spectral density matrices of the ξs and the xs, and their eigenvalues, ordered in decreasing order, are denoted by $\boldsymbol{\Sigma}_n^{\xi}(\theta)$, $\boldsymbol{\Sigma}_n^{x}(\theta)$, $\lambda_{nj}^{\xi}(\theta)$ and $\lambda_{nj}^{x}(\theta)$, respectively; under the above assumptions, they satisfy the following properties.

Proposition 1. *Under Assumptions 1 through 4,*

(i) *there exists $B^{\xi} > 0$ such that $\lambda_{n1}^{\xi}(\theta) \le B^{\xi}$ for all $n \in \mathbb{N}$ and $\theta \in [-\pi, \pi]$ (thus, the ξ's are idiosyncratic, see FHLZ, Section 2.2);*

(ii) *there exists $n^x \in \mathbb{N}$ such that, for $n > n^x$ and all $\theta \in [-\pi, \pi]$, and with $\alpha_j^{\chi}(\theta)$ as in Assumption 3,*

$$\frac{\lambda_{n1}^{x}(\theta)}{n} > \alpha_1^{\chi}(\theta) > \frac{\lambda_{n2}^{x}(\theta)}{n} > \cdots > \alpha_{q-1}^{\chi}(\theta) > \frac{\lambda_{nq}^{x}(\theta)}{n} > \alpha_q^{\chi}(\theta);$$

(iii) *there exists $B^x > 0$ such that $\lambda_{n,q+1}^{x}(\theta) \le B^x$ for all $n \in \mathbb{N}$ and $\theta \in [-\pi, \pi]$.*

Proof. The column and row norms of $\boldsymbol{\Sigma}_n^{\xi}(\theta)$ are equal, and, by (2.8), satisfy

$$\max_{j=1,2,\ldots,n}\sum_{i=1}^{n}|\sigma_{ij}^{\xi}(\theta)| \le \max_{j=1,2,\ldots,n}\sum_{i=1}^{\infty}|\sigma_{ij}^{\xi}(\theta)| \le B^2\frac{1}{2\pi(1-\rho)^2}.$$

On the other hand, the product of the row and the column norms, the square of the column norm in our case, is greater than or equal to the square of the spectral norm, see Lancaster and Tismenetsky (1985), p. 366, Exercise 11. As a consequence, setting $B^{\xi} = B^2 1/2\pi(1-\rho)^2$, we have $\lambda_{n1}^{\xi}(\theta) \le B^{\xi}$ for all n and θ.

Regarding (ii), $\boldsymbol{\Sigma}_n^{x}(\theta) = \boldsymbol{\Sigma}_n^{\chi}(\theta) + \boldsymbol{\Sigma}_n^{\xi}(\theta)$ implies that

$$\lambda_{nf}^{x}(\theta) \ge \lambda_{nf}^{\chi}(\theta) + \lambda_{nn}^{\xi}(\theta) \quad \text{and} \quad \lambda_{nf}^{x}(\theta) \le \lambda_{nf}^{\chi}(\theta) + \lambda_{n1}^{\xi}(\theta)$$

(these are two of the Weyl inequalities, see Franklin (2000), p. 157, Theorem 1; see also Appendix B). By Assumption 3,

$$\frac{\lambda_{nf}^{x}(\theta)}{n} \ge \frac{\lambda_{nf}^{\chi}(\theta) + \lambda_{nn}^{\xi}(\theta)}{n} > \alpha_f^{\chi}(\theta),$$

for $f = 1, \ldots, q$, and, for $f = 2, \ldots, q$,

$$\begin{aligned}\frac{\lambda_{nf}^{x}(\theta)}{n} &\le \frac{\lambda_{nf}^{\chi}(\theta) + \lambda_{n1}^{\xi}(\theta)}{n} \le \frac{\lambda_{nf}^{\chi}(\theta)}{n} + \frac{B^{\xi}}{n} \\ &\le \beta_{f-1}^{\chi}(\theta) + \frac{B^{\xi}}{n} < \alpha_{f-1}^{\chi}(\theta),\end{aligned}$$

for $n > n^{\chi}$, n^{χ} being such that $B^{\xi}/n^{\chi} < \min_{f=1,2,\ldots,q}[\min_{\theta\in[-\pi,\pi]}(\alpha_f^{\chi}(\theta) - \beta_f^{\chi}(\theta))]$.

As for (iii), $\lambda_{n,q+1}^{x} \le \lambda_{n,q+1}^{\chi}(\theta) + \lambda_{n1}^{\xi}(\theta)$. On the other hand, $\lambda_{n,q+1}^{\chi}(\theta) = 0$ for all θ. The result then follows from (i). □

Proposition 2. *Under Assumptions 1 through 4, the cross-spectral densities $\sigma_{ij}^{x}(\theta)$ possess derivatives of any order and are of bounded variation uniformly in $i, j \in \mathbb{N}$; namely, there exists $A^x > 0$ such that*

$$\sum_{h=1}^{\nu}|\sigma_{ij}^{x}(\theta_h) - \sigma_{ij}^{x}(\theta_{h-1})| \le A^x$$

for all $i, j, \nu \in \mathbb{N}$ and all partitions $-\pi = \theta_0 < \theta_1 < \cdots < \theta_{\nu-1} < \theta_\nu = \pi$ of $[-\pi, \pi]$.

Proof. Denoting by $\gamma_{ij,h}^{\xi}$, $h \ge 0$, the covariance between ξ_{it} and $\xi_{j,t-h}$,

$$\begin{aligned}|\gamma_{ij,h}^{\xi}| &= \left|\sum_{k=0}^{\infty}\sum_{s=1}^{\infty}\beta_{is,k}\beta_{js,k+h}\right| \le \sum_{k=0}^{\infty}\sum_{s=1}^{\infty}B_{is}B_{js}\rho^k\rho^{k+h} \\ &\le \rho^h\sum_{k=0}^{\infty}\rho^{2k}\sum_{s=1}^{\infty}B_{is}B_{js} \le \rho^h\frac{B^2}{1-\rho^2},\end{aligned} \tag{2.9}$$

by (2.7). For $h < 0$, $\gamma_{ij,h}^{\xi} = \gamma_{ji,-h}^{\xi}$, so that $|\gamma_{ij,h}^{\xi}| \le \rho^{|h|}B^2/(1-\rho^2)$. This implies that

$$\sigma_{ij}^{\xi}(\theta) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma_{ij,h}^{\xi}e^{-\imath h\theta}$$

has derivatives of all orders. Moreover,

$$\begin{aligned}\left|\frac{d}{d\theta}\sigma_{ij}^{\xi}(\theta)\right| &= \frac{1}{2\pi}\left|\sum_{h=-\infty}^{\infty}(-\imath h)\gamma_{ij,h}^{\xi}e^{-\imath h\theta}\right| \le \frac{B^2}{\pi(1-\rho^2)}\sum_{h=1}^{\infty}h\rho^h \\ &= \frac{B^2\rho}{\pi(1-\rho^2)(1-\rho)^2},\end{aligned}$$

which entails bounded variation of $\sigma_{ij}^{\xi}(\theta)$ uniformly in i and j. Bounded variation of $\sigma_{ij}^{\chi}(\theta)$, uniformly in i and j, is an obvious consequence of Assumption 2. The conclusion follows from the fact that $\sigma_{ij}^{x}(\theta) = \sigma_{ij}^{\chi}(\theta) + \sigma_{ij}^{\xi}(\theta)$. □

2.2. Transforming the dynamic model into a static one

In FHLZ we prove that, for *generic values* of the parameters $c_{if,k}$ and $d_{if,k}$ in (2.3), the space spanned by $u_{f,t-k}$, $f = 1, 2, \ldots, q$, $k \ge 0$, is equal to the space spanned by any $(q+1)$-dimensional subvector of $\boldsymbol{\chi}_t$ and its lags, and that the $(q+1)$-dimensional subvectors of $\boldsymbol{\chi}_t$ admit a *finite and unique* autoregressive representation (see, in particular, Section 4.1, Lemma 3). Following FHLZ, we use these genericity results as a motivation for the next assumption.

Let $\boldsymbol{\chi}_t^k = \left(\chi_{(k-1)(q+1)+1,t} \quad \cdots \quad \chi_{k(q+1),t}\right)$, $\quad k \in \mathbb{N}$.

Assumption 5. *Each vector $\boldsymbol{\chi}_t^k$, $k \in \mathbb{N}$, admits an autoregressive representation*

$$\mathbf{A}^k(L)\boldsymbol{\chi}_t^k = \mathbf{R}^k\mathbf{u}_t, \tag{2.10}$$

where

(i) $\mathbf{A}^k(L)$ *is* $(q+1)\times(q+1)$ *of degree not greater than* $S = qs_1+q^2s_2$, $\mathbf{A}^k(0) = \mathbf{I}_{q+1}$, *and the roots of* $\det \mathbf{A}^k(L)$ *are greater than one in modulus;*
(ii) $\mathbf{R}^k$ *is* $(q+1)\times q$ *and has rank* q*;*
(iii) *Representation* (2.10) *is unique. Precisely, if* $\tilde{\mathbf{A}}^k(L)\boldsymbol{\chi}_t^k = \tilde{\mathbf{R}}^k\tilde{\mathbf{u}}_t$, *where* (a) *the degree of* $\tilde{\mathbf{A}}^k(L)$ *does not exceed* S, (b) $\tilde{\mathbf{A}}(0) = \mathbf{I}_{q+1}$, (c) $\tilde{\mathbf{u}}_t$ *is* q*-dimensional white noise, and* (d) $\tilde{\mathbf{R}}^k$ *is* $(q+1)\times q$, *then* $\tilde{\mathbf{A}}^k(L) = \mathbf{A}^k(L)$, $\tilde{\mathbf{R}}^k = \mathbf{R}^k\mathbf{Q}'$, $\tilde{\mathbf{u}}_t = \mathbf{Q}\mathbf{u}_t$, *where* $\mathbf{Q}$ *is a* $q\times q$ *orthogonal matrix independent of* k.

Assumption 5 is a weaker version of Assumption A.3 in FHLZ, the difference being that FHLZ assume that all $(q+1)$-dimensional subvectors $(\chi_{i_1t}\ \chi_{i_2t}\ \cdots\ \chi_{i_{q+1}t})$ and their lags span the same space spanned by $\mathbf{u}_t$ and its lags, and have an autoregressive representation as in (2.10).

Writing $\underline{\mathbf{A}}(L)$ for the (infinite) block-diagonal matrix with diagonal blocks $\mathbf{A}^1(L), \mathbf{A}^2(L), \ldots$, and letting $\underline{\mathbf{R}} = (\mathbf{R}^{1'}, \mathbf{R}^{2'}, \ldots)'$, we thus have

$$\underline{\mathbf{A}}(L)\boldsymbol{\chi}_t = \underline{\mathbf{R}}\mathbf{u}_t. \tag{2.11}$$

We assume throughout the paper that $n = m(q+1)$ with $m \in \mathbb{Z}$. This is convenient and does not imply any loss of generality for our asymptotic analysis. Of course it may not hold in real datasets, see Section 4.2 for a discussion.

The upper $n\times n$ submatrix of $\underline{\mathbf{A}}(L)$ and the upper $n\times q$ submatrix of $\underline{\mathbf{R}}$ are denoted by $\mathbf{A}_n(L)$ and $\mathbf{R}_n$, respectively. If $n = m(q+1)$, so that the first m blocks of size $q+1$ are included,

$$\mathbf{A}_n(L)\boldsymbol{\chi}_{nt} = \mathbf{R}_n\mathbf{u}_t. \tag{2.12}$$

Inverting $\underline{\mathbf{A}}(L)$ in (2.11) yields $\boldsymbol{\chi}_t = \underline{\mathbf{A}}(L)^{-1}\underline{\mathbf{R}}\mathbf{u}_t$. Because $\mathbf{u}_t$ is orthonormal white noise, we obtain the following result.

Proposition 3. (i) *The ith row of* $\underline{\mathbf{A}}(L)^{-1}\underline{\mathbf{R}}$ *is* $\mathbf{b}_i(L) = (b_{i1}(L)\ b_{i2}(L)\ \cdots\ b_{iq}(L))$.
(ii) *In particular, the ith row of* $\underline{\mathbf{R}}$ *is* $(c_{i1,0}\ c_{i2,0}\ \cdots\ c_{iq,0})$.

Letting $\mathbf{Z}_t = \underline{\mathbf{A}}(L)\mathbf{x}_t$, we have

$$\mathbf{Z}_t = \boldsymbol{\Psi}_t + \boldsymbol{\Phi}_t, \quad \text{with} \quad \boldsymbol{\Psi}_t = \underline{\mathbf{R}}\mathbf{u}_t \quad \text{and} \quad \boldsymbol{\Phi}_t = \underline{\mathbf{A}}(L)\boldsymbol{\xi}_t. \tag{2.13}$$

This is a static form, linking $\mathbf{Z}_t$ to the common shocks $\mathbf{u}_t$. However, using the standard principal components of $\mathbf{Z}_t$ to estimate $\mathbf{R}$ and $\mathbf{u}_t$ requires further assumptions. Denote by $\boldsymbol{\Gamma}_n^{\phi}$ and $\boldsymbol{\Gamma}_n^{\psi}$ the variance–covariance matrices of $\boldsymbol{\Phi}_{nt}$ and $\boldsymbol{\Psi}_{nt}$, with eigenvalues μ_{nj}^{ϕ} and μ_{nj}^{ψ}, respectively.

Assumption 6. *(Eigenvalues of the covariance matrix of* $\boldsymbol{\Psi}_t$*: divergence and separation) There exist real numbers* $\alpha_f^{\psi}, f = 1,\ldots,q$, β_f^{ψ}, $f = 0,\ldots,q-1$, *and a positive integer* n^{ψ} *such that, for* $n > n^{\psi}$,

$$\beta_0^{\psi} \ge \frac{\mu_{n1}^{\psi}}{n} \ge \alpha_1^{\psi} > \beta_1^{\psi} \ge \frac{\mu_{n2}^{\psi}}{n} \ge \alpha_2^{\psi} > \beta_2^{\psi}$$
$$\ge \cdots \ge \alpha_{q-1}^{\psi} > \beta_{q-1}^{\psi} \ge \frac{\mu_{nq}^{\psi}}{n} \ge \alpha_q^{\psi} > 0.$$

The eigenvalues μ_{nf}^{ψ} depend on the "deep parameters" $c_{if,0}$, see Proposition 3(ii), and are invariant if $\underline{\mathbf{R}}$ and $\mathbf{u}_t$ are replaced by $\underline{\mathbf{R}}\mathbf{Q}'$ and $\mathbf{Q}\mathbf{u}_t$ respectively. Assumption 6 ensures that the static model (2.13) has exactly q common shocks. Note that Assumption 6 is not a consequence of Assumption 3, even enhanced with Assumption 5(ii). For, it is easy to construct an example in which (1) $q = 2$, (2) Assumptions 3 and 5 hold, but the two 3-dimensional columns of R^k asymptotically approach the same vector fast enough to prevent μ_{n2}^{ϕ} from diverging.

In order to introduce the next assumption, we must go over the procedure leading from the spectral density of the χs to the matrices $\mathbf{A}^k(L)$ appearing in (2.10). This procedure, with the population quantities replaced by their estimates, produces our estimator, see Section 3. It proceeds in two steps:

(i) Denoting by $\boldsymbol{\Sigma}_{jk}^{\chi}(\theta)$ the $(q+1)\times(q+1)$ cross-spectral density between $\boldsymbol{\chi}_t^j$ and $\boldsymbol{\chi}_t^k$, and by $\boldsymbol{\Gamma}_{jk,s}^{\chi}$ the covariance between $\boldsymbol{\chi}_t^j$ and $\boldsymbol{\chi}_{t-s}^k$, we have

$$\boldsymbol{\Gamma}_{jk,s}^{\chi} = \mathrm{E}\left(\boldsymbol{\chi}_t^j\boldsymbol{\chi}_{t-s}^{k\ \prime}\right) = \int_{-\pi}^{\pi} e^{\imath s\theta}\boldsymbol{\Sigma}_{jk}^{\chi}(\theta)d\theta. \tag{2.14}$$

(ii) The minimum-lag matrix polynomial $\mathbf{A}^k(L)$ and the variance–covariance function of the unobservable vectors

$$\boldsymbol{\Psi}_t^1 = \mathbf{A}^1(L)\boldsymbol{\chi}_t^1, \quad \boldsymbol{\Psi}_t^2 = \mathbf{A}^2(L)\boldsymbol{\chi}_t^2, \quad \ldots \tag{2.15}$$

follow from (2.14). Indeed, defining

$$\mathbf{A}^k(L) = \mathbf{I}_{q+1} - \mathbf{A}_1^kL - \cdots - \mathbf{A}_S^kL^S,$$

$$\mathbf{A}^{[k]} = \left(\mathbf{A}_1^k\ \mathbf{A}_2^k\ \cdots\ \mathbf{A}_S^k\right), \qquad \mathbf{B}_k^{\chi} = \left(\boldsymbol{\Gamma}_{kk,1}^{\chi}\ \boldsymbol{\Gamma}_{kk,2}^{\chi}\ \cdots\ \boldsymbol{\Gamma}_{kk,S}^{\chi}\right) \tag{2.16}$$

and

$$\mathbf{C}_{jk}^{\chi} = \begin{pmatrix} \boldsymbol{\Gamma}_{jk,0}^{\chi} & \boldsymbol{\Gamma}_{jk,1}^{\chi} & \cdots & \boldsymbol{\Gamma}_{jk,S-1}^{\chi} \\ \boldsymbol{\Gamma}_{jk,-1}^{\chi} & \boldsymbol{\Gamma}_{jk,0}^{\chi} & \cdots & \boldsymbol{\Gamma}_{jk,S-2}^{\chi} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}_{jk,-S+1}^{\chi} & \boldsymbol{\Gamma}_{jk,-S+2}^{\chi} & \cdots & \boldsymbol{\Gamma}_{jk,0}^{\chi} \end{pmatrix}, \tag{2.17}$$

we have

$$\mathbf{A}^{[k]} = \mathbf{B}_k^{\chi}\left(\mathbf{C}_{kk}^{\chi}\right)^{-1} = \mathbf{B}_k^{\chi}\left(\mathbf{C}_{kk}^{\chi}\right)_{\mathrm{ad}}\det\left(\mathbf{C}_{kk}^{\chi}\right)^{-1}, \tag{2.18}$$

where $\left(\mathbf{C}_{kk}^{\chi}\right)_{\mathrm{ad}}$ stands for the adjoint of the square matrix $\mathbf{C}_{kk}^{\chi}$.

Note that the procedure above, the definition of $\mathbf{C}_{kk}^{\chi}$ in particular, requires that $S > 0$. For $S = 0$, that is when the common components are white noise, $\mathbf{A}^k(L) = \mathbf{I}_{q+1}$ and we define $\mathbf{C}_{kk}^{\chi} = \mathbf{I}_{S(q+1)}$.

Non-singularity of $\mathbf{C}_{kk}^{\chi}$ is necessary for the uniqueness of the $\mathbf{A}^{[k]}$s, and is implied by Assumption 5. However, we require a slightly stronger condition to ensure that the $\mathbf{A}^{[k]}$s are (uniformly) bounded, in norm, as n tends to infinity.

Assumption 7. *There exists a real* $d > 0$ *such that* $\left|\det \mathbf{C}_{kk}^{\chi}\right| > d$ *for all* $k \in \mathbb{N}$.

For any fixed n and, in particular, for $n = n_0$ (supposed to be a multiple of $q+1$), the existence of a constant $d_n > 0$ such that $\left|\det \mathbf{C}_{kk}^{\chi}\right| > d_n$ for $1 \le k \le n/(q+1)$ is a consequence of Assumption 5. Assumption 7, however, is more demanding, as it imposes $\left|\det \mathbf{C}_{kk}^{\chi}\right| > d$ for all $k \in \mathbb{N}$ and a value of d that does not depend on n. This is reasonable if we require the (fictitious) "cross-sectional future" of the panel to resemble what has been observed, i.e. the n_0-dimensional panel (1.1)—a form of *cross-sectional stationarity*.

Proposition 4. *Under Assumptions 1 through 7, there exists $B^\Phi > 0$ such that $\mu^\Phi_{n1} \leq B^\Phi$ for all $n \in \mathbb{N}$.*

Proof. Let $\lambda^\Phi_{nj}(\theta)$ be the jth eigenvalue of the spectral density matrix of $\boldsymbol{\Phi}_{nt}$. Let us show that there exists a constant C^Φ such that $\lambda^\Phi_{n1}(\theta) \leq C^\Phi$ for all n and θ. Because $\lambda^\Phi_{n1}(\theta)$, for all θ, is non-decreasing with n (see (Forni and Lippi, 2001), our assumption that $n = m(q+1)$ does not cause any loss of generality. The spectral density of $\boldsymbol{\Phi}_{nt}$ is

$$\mathbf{A}_n(e^{-\iota\theta})\Sigma^\xi_n(\theta)\mathbf{A}'_n(e^{\iota\theta}),$$

where $\mathbf{A}_n(L)$ (see Eq. (2.12)) is block-diagonal with diagonal blocks $\mathbf{A}^k(L)$. If $\mathbf{a}(\theta)$ is an n-dimensional complex column vector such that $\mathbf{a}(\theta)'\overline{\mathbf{a}(\theta)} = 1$ for all θ, we have

$$\mathbf{a}(\theta)'\mathbf{A}_n(e^{-\iota\theta})\Sigma^\xi_n(\theta)\mathbf{A}'_n(e^{\iota\theta})\overline{\mathbf{a}(\theta)} \leq \lambda^\xi_{n1}(\theta)\left(\mathbf{a}'(\theta)\mathbf{A}_n(e^{-\iota\theta})\mathbf{A}'_n(e^{\iota\theta})\overline{\mathbf{a}(\theta)}\right)$$
$$\leq \lambda^\xi_{n1}(\theta)\lambda^{A_n}_1(\theta),$$

where $\lambda^{A_n}_1(\theta)$ is the first eigenvalue of $\mathbf{A}_n(e^{-\iota\theta})\mathbf{A}'_n(e^{\iota\theta})$, which is Hermitian, non-negative definite. By Proposition 1, $\sup_n\lambda^\xi_{n1}(\theta) \leq B^\xi$. Moreover, given the diagonal structure of $\mathbf{A}_n(L)$, $\lambda^{A_n}_1(\theta) = \max_{k=1,2,\dots,m}\lambda^{A^k}_1(\theta) \leq \sup_{k\in\mathbb{N}}\lambda^{A^k}_1(\theta)$, where $\lambda^{A^k}_1(\theta)$ is the first eigenvalue of $\mathbf{A}^k(e^{-\iota\theta})\mathbf{A}^{k\prime}(e^{\iota\theta})$. Assumptions 2 and 7 imply that $\sup_{k\in\mathbb{N}}\lambda^{A^k}_1(\theta) \leq D^\Phi$ for some $D^\Phi > 0$ and all θ. On the other hand,

$$\lambda^\Phi_{n1}(\theta) = \sup \mathbf{a}(\theta)'\mathbf{A}_n(e^{-\iota\theta})\Sigma^\xi_n(\theta)\mathbf{A}'_n(e^{\iota\theta})\overline{\mathbf{a}(\theta)} \leq B^\xi D^\Phi,$$

the sup being over all the vectors $\mathbf{a}(\theta)$ such that $\mathbf{a}(\theta)'\overline{\mathbf{a}(\theta)} = 1$. Lastly,

$$\mu^\Phi_{n1} = \sup \mathbf{b}'\Gamma^\Phi_n\mathbf{b} = \int_{-\pi}^{\pi}\left(\mathbf{b}'\Sigma^\Phi_n(\theta)\mathbf{b}\right)d\theta$$
$$\leq \int_{-\pi}^{\pi}\lambda^\Phi_{n1}(\theta)d\theta \leq 2\pi B^\xi D^\Phi,$$

the sup being over all the n-dimensional column vectors $\mathbf{b}$ such that $\mathbf{b}'\mathbf{b} = 1$. $\square$

Proposition 4 ensures that $\boldsymbol{\Psi}_t$ is a genuine idiosyncratic component. Because $\boldsymbol{\Phi}_t$ and $\boldsymbol{\Psi}_s$ are independent for all t and s in $\mathbb{Z}$, a consequence of Assumption 1(i), the model (2.13) is a factor model with a static representation—a special case of (1.4), with $r = q$ and $\mathbf{N}(L) = \mathbf{I}_q$.

3. Estimation: asymptotics

Our estimation procedure follows the same steps as the population construction in Section 2.2, with the population spectral density of the xs replaced with an estimator $\hat{\Sigma}^x_n(\theta)$ fulfilling Assumption 9. Based on Forni et al. (2000), we obtain the estimator $\hat{\Sigma}^\chi_n(\theta)$ by means of the first q frequency-domain principal components of the xs (using the first q eigenvectors of $\hat{\Sigma}^x_n(\theta)$). Then the matrices $\hat{\Gamma}^\chi_{jk}$, $\hat{\mathbf{B}}^\chi_{jk}$, $\hat{\mathbf{C}}^\chi_{jk}$ and $\hat{\mathbf{A}}_n(L)$ are computed as natural counterparts of their population versions in Section 2.2. Finally, estimators for $\mathbf{R}_n$ and $\mathbf{u}_t$ are obtained via a standard principal component analysis of $\hat{\mathbf{Z}}_{nt} = \hat{\mathbf{A}}(L)\mathbf{x}_{nt}$. Consistency results with exact rates of convergence ζ_{nT}, as defined in Eq. (1.7), are provided in Propositions 7 through 11 for all those estimators.

Explicit dependence on the index n has been necessary in Section 2. From now on, it will be convenient to introduce a minor change in notation, dropping n whenever possible. In particular,

(i) $\Sigma^x(\theta) = \left(\sigma^x_{ij}(\theta)\right)_{i,j=1,\dots,n}$ and $\lambda^x_f(\theta)$ replace $\Sigma^x_n(\theta)$ and $\lambda^x_{nf}(\theta)$, respectively;
(ii) $\Lambda^x(\theta)$ denotes the $q \times q$ diagonal matrix with diagonal elements $\lambda^x_f(\theta)$;
(iii) $\mathbf{P}^x(\theta)$ denotes the $n \times q$ matrix the q columns of which are the unit-modulus eigenvectors corresponding to $\Sigma^x(\theta)$s first q eigenvalues; the columns and entries of $\mathbf{P}^x(\theta)$ are denoted by $\mathbf{P}^x_f(\theta)$ and $p^x_{if}(\theta), f = 1,\dots,q, i = 1,\dots,n$, respectively;
(iv) $\Sigma^\chi(\theta) = \left(\sigma^\chi_{ij}(\theta)\right)_{i,j=1,\dots,n}$, $\lambda^\chi_f(\theta)$, $\Lambda^\chi(\theta)$, $\mathbf{P}^\chi(\theta)$, $\Sigma^\xi(\theta)$, etc. are defined as in (i);
(v) all the above matrices and scalars depend on n; the corresponding estimators,

$$\hat{\Sigma}^x(\theta),\ \hat{\lambda}^x_f(\theta),\ \hat{\Lambda}^x(\theta),\ \hat{\mathbf{P}}^x(\theta)\ \text{ and }\ \hat{\Sigma}^\chi(\theta),\ \hat{\lambda}^\chi_f(\theta),$$
$$\hat{\Lambda}^\chi(\theta),\ \hat{\mathbf{P}}^\chi(\theta)$$

(precise definitions are provided below) depend both on n and the observed values $x_{it}, i = 1,\dots,n, t = 1,\dots,T$. For simplicity, we say that they depend on n and T;
(vi) the same notational change applies to Γ^ψ_n and related eigenvalues and eigenvectors;
(vii) $\mathbf{A}(L)$ and $\mathbf{R}$, denoting the upper left $n\times n$ and $n\times q$ submatrices of $\underline{\mathbf{A}}(L)$ and $\underline{\mathbf{R}}$, respectively, are used instead of $\mathbf{A}_n(L)$ and $\mathbf{R}_n$; $\hat{\mathbf{A}}(L)$ and $\hat{\mathbf{R}}$ stand for their estimated counterparts;
(viii) to avoid confusion, however, we keep explicit reference to n in $\mathbf{x}_{nt}$, $\boldsymbol{\chi}_{nt}$, $\mathbf{Z}_{nt}$, etc., with estimated counterparts of the form $\hat{\boldsymbol{\chi}}_{nt}$, $\hat{\mathbf{Z}}_{nt}$, etc.; thus, we write, for instance,

$$\mathbf{Z}_{nt} = \mathbf{A}(L)\mathbf{x}_{nt} = \mathbf{R}\mathbf{u}_t + \boldsymbol{\Phi}_{nt};$$

(ix) lastly, if $\mathbf{F}$ is a matrix, we denote by $\tilde{\mathbf{F}}$ its conjugate transpose, and by $\|\mathbf{F}\|$ its spectral norm (see Appendix B).

3.1. Estimation of $\Sigma^x(\theta)$

The following definition, coined by Wu (2005), generalizes the usual measures of time dependence for stochastic processes.

Definition 1. Physical dependence. Let $\boldsymbol{\epsilon}_t$ be an i.i.d. stochastic vector process, possibly infinite-dimensional, and let $z_t = F(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1},\dots)$, where $F : [\mathbb{R}\times\mathbb{R}\times\cdots] \to \mathbb{R}$ is a measurable function; assume that z_t has finite pth moment for $p > 0$. Let $\boldsymbol{\epsilon}^*$ be a stochastic vector with the same dimension and distribution as the $\boldsymbol{\epsilon}_t$s, such that $\boldsymbol{\epsilon}^*$ and $\boldsymbol{\epsilon}_t$ are independent for all t. For $k \geq 0$ the physical dependence $\delta^{[z_t]}_{kp}$ is defined as

$$\delta^{[z_t]}_{kp} = \left(E\left(|F(\boldsymbol{\epsilon}_k,\dots,\boldsymbol{\epsilon}_0,\boldsymbol{\epsilon}_{-1},\dots) - F(\boldsymbol{\epsilon}_k,\dots,\boldsymbol{\epsilon}^*,\boldsymbol{\epsilon}_{-1},\dots)|^p\right)\right)^{1/p}.$$

Assumption 8. *There exist p and A, with $p > 4$ and $0 < A < \infty$, such that*

$$E\left(|u_{ft}|^p\right) \leq A,\quad E\left(|\eta_{it}|^p\right) \leq A, \tag{3.1}$$

for all $i \in \mathbb{N}$ and $f = 1,\dots,q$.

The main result of the section, that the estimate of the cross-spectral density between x_{it} and x_{jt} converges uniformly with respect to the frequency and to i and j, see Proposition 6, requires the following results on the pth moments and the physical dependence of the xs.

Proposition 5. *Under Assumptions 1 through 8, there exist $\rho_1 \in (0,1)$ and $A_1 \in (0,\infty)$ such that, for all $i \in \mathbb{N}$,*

$$E\left(|x_{it}|^p\right) \leq A_1 \quad \text{and} \quad \delta^{[x_{it}]}_{kp} \leq A_1\rho_1^k. \tag{3.2}$$

Proof. By the Minkowski inequality,

$$\left(E\left(|x_{it}|^p\right)\right)^{\frac{1}{p}} = \left(E\left(|\chi_{it}+\xi_{it}|^p\right)\right)^{\frac{1}{p}} \leq \left(E\left(|\chi_{it}|^p\right)\right)^{\frac{1}{p}} + \left(E\left(|\xi_{it}|^p\right)\right)^{\frac{1}{p}}.$$

Using the Minkowski inequality again, condition (2.4) and Assumption 8, we obtain

$$\begin{aligned}\left(E\left(|\xi_{it}|^p\right)\right)^{\frac{1}{p}} &= \left(E\left(\left|\sum_{s=1}^{\infty}\sum_{k=0}^{\infty}\beta_{is,k}\eta_{s,t-k}\right|^p\right)\right)^{\frac{1}{p}}\\ &\le \sum_{s=1}^{\infty}\sum_{k=0}^{\infty}\left(E\left(|\beta_{is,k}\eta_{s,t-k}|\right)^p\right)^{\frac{1}{p}}\\ &\le \sum_{s=1}^{\infty}\sum_{k=0}^{\infty}|\beta_{is,k}|E\left(|\eta_{s,t-k}|^p\right)^{\frac{1}{p}}\\ &\le A^{\frac{1}{p}}\sum_{s=1}^{\infty}\sum_{k=0}^{\infty}B_{is}\rho^k \le A^{\frac{1}{p}}B\frac{1}{1-\rho}.\end{aligned}$$

An analogous inequality can be obtained for the common components, using Assumption 2 and the first inequality in (3.1). The first inequality in (3.2) follows.

Turning to the second inequality, for $k \ge 0$,

$$\xi_{ik} - \xi^*_{ik} = \sum_{s=1}^{\infty}\beta_{is,k}(\eta_{sk} - \eta^*_s),$$

where ξ^*_{ik} has the same definition as ξ_{ik}, with η_{s0} replaced by η^*_s. The Minkowski inequality, condition (2.4) and Assumption 8 imply

$$\begin{aligned}\delta^{[\xi_{it}]}_{k,p} &= \left(E\left(\left|\sum_{s=1}^{\infty}\beta_{is,k}(\eta_{sk} - \eta^*_s)\right|^p\right)\right)^{\frac{1}{p}}\\ &\le \sum_{s=1}^{\infty}\left(E\left(|\beta_{is,k}(\eta_{sk} - \eta^*_s)|^p\right)\right)^{\frac{1}{p}}\\ &\le \rho^k\sum_{s=1}^{\infty}B_{is}\left(E(|\eta_{sk} - \eta^*_s|)^p\right)^{\frac{1}{p}} \le \rho^k 2BA^{\frac{1}{p}}.\end{aligned}$$

A similar inequality can be obtained for the common components, using Assumption 2 and the first of inequalities (3.1), with ρ replaced by ϕ^{-1}, ϕ being defined in Assumption 2. Then,

$$\begin{aligned}\delta^{[x_{it}]}_{kp} &= (E|x_{it} - x^*_{it}|^p)^{\frac{1}{p}} = \left(E\left(|(\chi_{it} - \chi^*_{it}) + (\xi_{it} - \xi^*_{it})|^p\right)^{\frac{1}{p}}\right)\\ &\le \left(E\left(|\chi_{it} - \chi^*_{it}|\right)^p\right)^{\frac{1}{p}} + \left(E\left(|\xi_{it} - \xi^*_{it}|\right)^p\right)^{\frac{1}{p}} = \delta^{[\chi_{it}]}_{kp} + \delta^{[\xi_{it}]}_{kp}.\end{aligned}$$

The conclusion follows. □

Consider now the lag-window estimator

$$\hat{\Sigma}^x(\theta) = \frac{1}{2\pi}\sum_{k=-T+1}^{T-1}K\left(\frac{k}{B_T}\right)e^{-\imath k\theta}\hat{\Gamma}^x_k, \tag{3.3}$$

of the spectral density $\Sigma^x(\theta)$, where $\hat{\Gamma}^x_k = \frac{1}{T}\sum_{t=|k|+1}^{T}\mathbf{x}_t\mathbf{x}_{t-|k|}$.

Assumption 9. *Lag-window estimation of $\Sigma^x(\theta)$.*

(i) *The kernel function K is even, bounded, with support $[-1, 1]$; moreover,*

  (1) *for some $\kappa > 0$, $|K(u) - 1| = O(|u|^\kappa)$ as $u \to 0$,*
  (2) *$\int_{-\infty}^{\infty}K^2(u)du < \infty$,*
  (3) *$\sum_{j\in\mathbb{Z}}\sup_{|s-j|\le 1}|K(jw) - K(sw)| = O(1)$ as $w \to 0$;*

(ii) *for some $c_1, c_2 > 0$, δ and $\underline{\delta}$ such that $0 < \underline{\delta} < \delta < 1 < \underline{\delta}(2\kappa + 1)$,*

  $c_1T^{\underline{\delta}} \le B_T \le c_2T^{\delta}$

Proposition 6. *Under Assumptions 1 through 9, there exists $C > 0$ such that*

$$E\left(\max_{|h|\le B_T}\left|\hat{\sigma}^x_{ij}(\theta^*_h) - \sigma^x_{ij}(\theta^*_h)\right|^2\right) \le C\left(T^{-1}B_T\log B_T\right), \tag{3.4}$$

where $\theta^*_h = \pi h/B_T$, for all T, i and j in $\mathbb{N}$.

See Appendix A for the proof.

3.2. Estimation of $\sigma^\chi_{ij}(\theta)$ and $\gamma^\chi_{ij,k}$

Our estimator of the spectral density matrix of the common components $\boldsymbol{\chi}_{nt}$ is the Forni et al. (2000) estimator

$$\hat{\Sigma}^\chi(\theta) = \hat{\mathbf{P}}^x(\theta)\hat{\Lambda}^x(\theta)\tilde{\hat{\mathbf{P}}}^x(\theta_h). \tag{3.5}$$

Proposition 7. *Under Assumptions 1 through 7,*

$$\max_{|h|\le B_T}|\hat{\sigma}^\chi_{ij}(\theta^*_h) - \sigma^\chi_{ij}(\theta^*_h)| = O_P(\zeta_{nT}),$$

where $\theta^*_h = \pi h/B_T$, as $T \to \infty$ and $n \to \infty$, uniformly in i and j. Precisely, for any $\epsilon > 0$, there exists $\eta(\epsilon)$, independent of i and j, such that, for all n and T,

$$P\left(\frac{\max_{h\le B_T}|\hat{\sigma}^\chi_{ii}(\theta^*_h) - \sigma^\chi_{ii}(\theta^*_h)|}{\zeta_{nT}} \ge \eta(\epsilon)\right) < \epsilon.$$

See Appendix B for the proof.

Our estimator of the covariance $\gamma^\chi_{ij,\ell}$ of χ_{it} and $\chi_{j,t-\ell}$ is, as in Forni et al. (2005),

$$\hat{\gamma}^\chi_{ij,\ell} = \frac{\pi}{B_T}\sum_{|h|\le B_T}e^{\imath\ell\theta^*_h}\hat{\sigma}^\chi_{ij}(\theta^*_h), \tag{3.6}$$

where $\theta^*_h = \pi h/B_T$. Recalling that $\gamma^\chi_{ij,\ell} = \int_{-\pi}^{\pi}e^{\imath\ell\theta}\sigma^\chi_{ij}(\theta)d\theta$, we have

$$\begin{aligned}|\hat{\gamma}^\chi_{ij,\ell} - \gamma^\chi_{ij,\ell}| \le{}& \frac{\pi}{B_T}\sum_{|h|\le B_T}|e^{\imath\ell\theta^*_h}\hat{\sigma}^\chi_{ij}(\theta^*_h) - e^{\imath\ell\theta^*_h}\sigma^\chi_{ij}(\theta^*_h)|\\ &+\left|\frac{\pi}{B_T}\sum_{|h|\le B_T}e^{\imath\ell\theta^*_h}\sigma^\chi_{ij}(\theta^*_h) - \int_{-\pi}^{\pi}e^{\imath\ell\theta}\sigma^\chi_{ij}(\theta)d\theta\right|\\ \le{}& \frac{\pi}{B_T}\sum_{|h|\le B_T}|\hat{\sigma}^\chi_{ij}(\theta^*_h) - \sigma^\chi_{ij}(\theta^*_h)|\\ &+\frac{\pi}{B_T}\sum_{|h|\le B_T}\max_{\theta^*_{h-1}\le\theta\le\theta^*_h}|e^{\imath\ell\theta^*_h}\sigma^\chi_{ij}(\theta^*_h) - e^{\imath\ell\theta}\sigma^\chi_{ij}(\theta)|\\ \le{}& \pi\max_{|h|\le B_T}|\hat{\sigma}^\chi_{ij}(\theta^*_h) - \sigma^\chi_{ij}(\theta^*_h)|\\ &+\frac{\pi B}{B_T}\sum_{|h|\le B_T}\max_{\theta^*_{h-1}\le\theta\le\theta^*_h}|e^{\imath\ell\theta^*_h} - e^{\imath\ell\theta}|\\ &+\frac{\pi}{B_T}\sum_{|h|\le B_T}\max_{\theta^*_{h-1}\le\theta\le\theta^*_h}|\sigma^\chi_{ij}(\theta^*_h) - \sigma^\chi_{ij}(\theta)|\\ \le{}& \pi\max_{|h|\le B_T}|\hat{\sigma}^\chi_{ij}(\theta^*_h) - \sigma^\chi_{ij}(\theta^*_h)|\\ &+\frac{\pi B}{B_T}\sum_{|h|\le B_T}\left(|e^{\imath\ell\theta^*_{h-1}} - e^{\imath\ell\tilde{\theta}^*_{h-1}}| + |e^{\imath\ell\tilde{\theta}^*_{h-1}} - e^{\imath\ell\theta^*_{h-1}}|\right)\\ &+\frac{\pi}{B_T}\sum_{|h|\le B_T}\left(|\sigma^\chi_{ij}(\theta^*_{h-1}) - \sigma^\chi_{ij}(\check{\theta}^*_{h-1})|\right.\\ &\left.+|\sigma^\chi_{ij}(\check{\theta}^*_{h-1}) - \sigma^\chi_{ij}(\theta^*_h)|\right),\end{aligned} \tag{3.7}$$

where B is the bound in Proposition 1(i), and $\tilde{\theta}^*_{h-1}$ and $\check{\theta}^*_{h-1}$ are points in the interval $[\theta_{h-1}, \theta_h]$ where the functions of θ, $|e^{i\ell\theta^*_s} - e^{i\ell\theta}|$ and $|\sigma_{ij}(\theta^*_s) - \sigma_{ij}(\theta)|$, respectively, attain a maximum. Of course, the function $e^{\imath\ell\theta}$ is of bounded variation, while the functions $\sigma^\chi_{ij}(\theta)$ are of bounded variation by Assumption 2, so that the second and third terms are $O(1/B_T)$.

Using Proposition 7, we obtain that $|\hat{\gamma}^\chi_{ij,\ell} - \gamma^\chi_{ij,\ell}|$ is $O_P(\zeta_{nT}) + O(1/B_T)$. Since $\zeta_{nT} = \max(1/\sqrt{n},\ 1/\sqrt{T/B_T\log T})$, the latter term is absorbed in the former under Assumption 10. Proposition 8 follows.

Assumption 10. *The lower bound $\underline{\delta}$ in Assumption 9 satisfies $\underline{\delta} > 1/3$.*

Proposition 8. *Under Assumption 1 through 10, for each $\ell \geq 0$,*

$$|\hat{\gamma}^{\chi}_{ij,\ell} - \gamma^{\chi}_{ij,\ell}| = O_P(\zeta_{nT}) \quad \text{as } T \to \infty \text{ and } n \to \infty. \tag{3.8}$$

3.3. Estimation of $\mathbf{A}^k(L)$

Under our assumptions, the common component admits the block-diagonal finite-order vector autoregressive representation (2.11). If the $\boldsymbol{\chi}_t$s were observed, estimation by OLS would be appropriate. However, although we do not observe the $\boldsymbol{\chi}_t$s, we do have (consistent) estimates of their autocovariance function. This naturally leads to Yule–Walker estimators of the autoregressive coefficients and the innovation covariance matrix. The definition of $\hat{\mathbf{A}}^{[k]}$ then is straightforward from (2.16), (2.17) and (2.18).

Proposition 9. *Under Assumption 1 through 10, $\|\hat{\mathbf{A}}^{[k]} - \mathbf{A}^{[k]}\| = O_P(\zeta_{nT})$ as $T \to \infty$ and $n \to \infty$.*

See Appendix C for the proof.

3.4. Estimation of $\mathbf{R}$ and $\mathbf{u}_t$

We start with $\mathbf{Z}_{nt} = \boldsymbol{\Psi}_{nt} + \boldsymbol{\Phi}_{nt} = \mathbf{R}\mathbf{u}_t + \boldsymbol{\Phi}_{nt}$. The covariance matrix of $\boldsymbol{\Psi}_{nt}$ is

$$\mathbf{R}\mathbf{R}' = \mathbf{P}^{\psi}\boldsymbol{\Lambda}^{\psi}\mathbf{P}^{\psi\prime} = \mathbf{P}^{\psi}(\boldsymbol{\Lambda}^{\psi})^{1/2}(\boldsymbol{\Lambda}^{\psi})^{1/2}\mathbf{P}^{\psi\prime},$$

where $\boldsymbol{\Lambda}^{\psi}$ is $q \times q$ with the non-zero eigenvalues of $\mathbf{R}\mathbf{R}'$ on the main diagonal, while $\mathbf{P}^{\psi}$ is $n \times q$ with the corresponding eigenvectors on the columns. Thus, we have the representation

$$\mathbf{Z}_{nt} = \mathbf{P}^{\psi}(\boldsymbol{\Lambda}^{\psi})^{1/2}\mathbf{v}_t + \boldsymbol{\Phi}_{nt} = \mathcal{R}\mathbf{v}_t + \boldsymbol{\Phi}_{nt},$$

say, where $\mathbf{v}_t = \mathbf{H}\mathbf{u}_t$, with $\mathbf{H}$ orthogonal. Note that, for given i and f, the (i, f) entry of $\mathcal{R}$ depends on n, so that the matrices $\mathcal{R}$ are not nested; nor is $\mathbf{v}_t$ independent of n. However, the product of each row of $\mathcal{R}$ by $\mathbf{v}_t$ yields the corresponding coordinate of $\boldsymbol{\Psi}_{nt}$, which does not depend on n.

Our estimator of $\mathcal{R} = \mathbf{P}^{\psi}(\boldsymbol{\Lambda}^{\psi})^{1/2}$ is $\hat{\mathcal{R}} = \hat{\mathbf{P}}^{z}(\hat{\boldsymbol{\Lambda}}^{z})^{1/2}$, where $\hat{\mathbf{P}}^{z}$ and $\hat{\boldsymbol{\Lambda}}^{z}$ are the eigenvectors and eigenvalues, respectively, of the empirical variance–covariance matrix of $\hat{\mathbf{Z}}_{nt} = \hat{\mathbf{A}}(L)\mathbf{x}_{nt}$, that is, $\mathbf{x}_{nt}$ filtered with the *estimated* matrices $\hat{\mathbf{A}}(L)$. This is the reason for the complications we have to deal with in Appendix D.

Proposition 10. *Under Assumptions 1 through 10, $\|\hat{\mathcal{R}}_i - \mathcal{R}_i\hat{\mathbf{W}}_q\| = O_P(\zeta_{nT})$, as $T \to \infty$ and $n \to \infty$, where $\mathcal{R}_i$ is the ith row of $\mathcal{R}$, and $\hat{\mathbf{W}}_q$ is a $q \times q$ diagonal matrix, depending on n and T, whose diagonal entries are either 1 or -1.*

See Appendix D for the proof.

Let us point out again that the ith row of $\mathcal{R}$ depends on n. Therefore, Proposition 10 only states that the differences between the entries of $\hat{\mathcal{R}}$ and those of $\mathcal{R}$ converge to zero (upon sign correction), not that the estimated entries converge. Now, suppose that the common shocks can be identified by means of economically meaningful statements. For example, suppose that we have good reasons to claim that the upper $q \times q$ matrix of the "structural" representation is lower triangular with positive diagonal entries (an iterative scheme for the first q common components). As is well known, such conditions determine a unique representation, denote it by $\mathbf{Z}_t = \underline{\mathbf{R}}^{*}\mathbf{u}^{*}_t + \boldsymbol{\Phi}_t$ or $\mathbf{Z}_{nt} = \mathbf{R}^{*}\mathbf{u}^{*}_t + \boldsymbol{\Phi}_t$, where the $n \times q$ matrices $\mathbf{R}^{*}$ are nested. In particular, starting with $\mathbf{Z}_{nt} = \mathcal{R}\mathbf{v}_t + \boldsymbol{\Phi}_{nt}$, there exists exactly one orthogonal matrix $\mathbf{G}(\mathcal{R})$ (actually $\mathbf{G}(\mathcal{R})$ only depends on the $q \times q$ upper submatrix of $\mathcal{R}$) such that $\mathbf{R}^{*} = \mathcal{R}\mathbf{G}(\mathcal{R})$. Thus, while the entries of $\mathcal{R}$ depend on n, those of $\mathcal{R}\mathbf{G}(\mathcal{R})$ do not.

Applying the same rule to $\hat{\mathcal{R}}$, we obtain the matrices $\hat{\mathbf{R}}^{*} = \hat{\mathcal{R}}\mathbf{G}(\hat{\mathcal{R}})$. It is easily seen that each entry of $\hat{\mathbf{R}}^{*}$ (depending on n and T) converges to the corresponding entry of $\mathbf{R}^{*}$ (independent of n and T) at rate ζ_{nT}.

Lastly, define the population *impulse–response functions* as the entries of the $n \times q$ matrix $\mathbf{B}^{*}(L) = \mathbf{A}(L)^{-1}\mathbf{R}^{*}$, and their estimators as those of $\hat{\mathbf{B}}^{*}(L) = \hat{\mathbf{A}}(L)^{-1}\hat{\mathbf{R}}^{*}$. Denoting by $b^{*}_{if}(L) = b^{*}_{if,0} + b^{*}_{if,1}L + \cdots$ and $\hat{b}^{*}_{if}(L) = \hat{b}^{*}_{if,0} + \hat{b}^{*}_{if,1}L + \cdots$, respectively, such entries, Propositions 9 and 10 imply that $|\hat{b}^{*}_{if,k} - b^{*}_{if,k}| = O_P(\zeta_{nT})$ for all i, f and k.

An iterative identification scheme will be used in Section 4 to compare different estimates of the impulse–response functions.[3]

Our estimator of $\mathbf{v}_t$ is simply the projection of $\hat{\mathbf{z}}_t$ on $\hat{\mathbf{P}}^{z}(\hat{\boldsymbol{\Lambda}}^{z})^{-1/2}$, namely,

$$\hat{\mathbf{v}}_t = ((\hat{\boldsymbol{\Lambda}}^{z})^{1/2}\hat{\mathbf{P}}^{z\prime}\hat{\mathbf{P}}^{z}(\hat{\boldsymbol{\Lambda}}^{z})^{1/2})^{-1}(\hat{\boldsymbol{\Lambda}}^{z})^{1/2}\hat{\mathbf{P}}^{z\prime}\hat{\mathbf{z}}_t = (\hat{\boldsymbol{\Lambda}}^{z})^{-1/2}\hat{\mathbf{P}}^{z\prime}\hat{\mathbf{z}}_t.$$

For that estimator $\hat{\mathbf{v}}_t$, we have the following consistency result.

Proposition 11. *Under Assumption 1 through 10, $\|\hat{\mathbf{v}}_t - \hat{\mathbf{W}}_q\mathbf{v}_t\| = O_P(\zeta_{nT})$, as $T \to \infty$ and $n \to \infty$, where $\hat{\mathbf{W}}_q$ is a $q \times q$ diagonal matrix, depending on n and T, whose diagonal entries equal either 1 or -1.*

See Appendix E for the proof.

3.5. Estimation and cross-sectional ordering

Let us now focus on the observed $(n_0 \times T_0)$-dimensional panel (1.1), where as usual it is convenient to assume $n_0 = m_0(q+1)$. Because the cross-sectional ordering of the n_0 variables is arbitrary, sensible concepts and sensible inference methods, as a rule, should be invariant (equivariant) under cross-sectional permutations.

Let $\mathbf{p} = (i_1\ i_2\ \cdots\ i_{n_0})$ denote a permutation of $\{1, \ldots, n_0\}$. The cross-spectral densities $\sigma^{x}_{ij}(\theta)$ and their estimators $\hat{\sigma}^{x}_{ij}(\theta)$, of course, are equivariant under $\mathbf{p}$, that is, for any $k,\ \ell \in \{1, \ldots, n_0\}$, $\sigma^{x}_{i_k i_\ell}(\theta) = \sigma^{x}_{k\ell}(\theta)$, etc.: except for their indexation, they do not depend on the cross-sectional ordering of the panel. Similarly, the eigenvalues of $\boldsymbol{\Sigma}^{x}(\theta)$ and $\hat{\boldsymbol{\Sigma}}^{x}(\theta)$ are permutation-invariant, while the corresponding eigenvectors are equivariant, so that $\hat{\sigma}^{\chi}_{ij}(\theta)$ and $\hat{\gamma}^{\chi}_{ij,\ell}$ are permutation-equivariant as well.

On the other hand, the matrices $\mathbf{A}^k(L)$ and their estimators strongly depend on the ordering of the cross-section. However, we are not interested in such matrices *per se* but only insofar as they enter the impulse–response functions $\mathbf{A}(L)^{-1}\mathbf{R}^{*}$ and their estimators. We argue below that although the population impulse–response functions are permutation-equivariant, their estimators are not.

Given $\mathbf{p} = (i_1\ i_2\ \cdots\ i_{n_0})$, let $\boldsymbol{\chi}^{[\mathbf{p}]}_{n_0 t} = (\chi_{i_1 t}\ \chi_{i_2 t}\ \cdots\ \chi_{i_{n_0} t})' = \boldsymbol{\Pi}_{\mathbf{p}}\boldsymbol{\chi}_{n_0 t}$, where $\boldsymbol{\Pi}_{\mathbf{p}}$ stands for the permutation matrix associated with $\mathbf{p}$. The permuted vector $\boldsymbol{\chi}^{[\mathbf{p}]}_{n_0 t}$ has a block-diagonal VAR representation (of the form (2.12))

$$\mathbf{A}^{[\mathbf{p}]}(L)\boldsymbol{\chi}^{[\mathbf{p}]}_{n_0 t} = \mathbf{R}^{*[\mathbf{p}]}\mathbf{u}_t$$

(the index n_0 has been dropped in $\mathbf{A}^{[\mathbf{p}]}$, see the beginning of Section 3). That representation rewrites as

$$\mathbf{A}^{(\mathbf{p})}(L)\boldsymbol{\chi}_{n_0 t} = \mathbf{R}^{*(\mathbf{p})}\mathbf{u}_t$$

with $\mathbf{A}^{(\mathbf{p})}(L) = \boldsymbol{\Pi}_{\mathbf{p}}^{-1}\mathbf{A}^{[\mathbf{p}]}(L)\boldsymbol{\Pi}_{\mathbf{p}}$ and $\mathbf{R}^{*(\mathbf{p})} = \boldsymbol{\Pi}_{\mathbf{p}}^{-1}\mathbf{R}^{*[\mathbf{p}]}$. The matrix $\mathbf{A}^{(\mathbf{p})}(L)$ has $q+1$ non-zero entries in each row and column but, unlike $\mathbf{A}_n(L)$ in (2.12), is not block-diagonal. The following statement is a direct consequence of the assumption that $\mathbf{u}_t$ is orthonormal white noise.

[3] All just-identifying rules considered in the SVAR literature can be dealt with along the same lines, see Forni et al. (2009).

Proposition 12. *Suppose that Assumption 5 holds for all $(q+1)$-dimensional subvectors $(\chi_{j_1t}\ \chi_{j_2t}\ \cdots\ \chi_{j_{q+1}t})$, with $j_s \le n_0$. Then, for all permutations $\mathbf{p}$,*

$$\text{(i) } \mathbf{A}^{(\mathbf{p})}(L)^{-1}\mathbf{R}^{*(\mathbf{p})} = \mathbf{A}(L)^{-1}\mathbf{R}^*, \quad \textit{and} \quad \text{(ii) } \mathbf{R}^{*(\mathbf{p})} = \mathbf{R}^*.$$

It follows that the population impulse–response functions $\mathbf{A}^{(\mathbf{p})}(L)^{-1}\mathbf{R}^{*(\mathbf{p})}$ do not depend on the arbitrary ordering of the panel.

Unfortunately such independence does not hold anymore for $\hat{\mathbf{A}}^{(\mathbf{p})}(L)^{-1}\hat{\mathbf{R}}^{*(\mathbf{p})}$. In particular,

(I) owing to finiteness of n and T, the estimation error of $\hat{\mathbf{b}}_i^{*(\mathbf{p})}(L)$ partly depends on ξ_{it} and the other idiosyncratic components involved in the block of x_{it} under $\mathbf{p}$; the correlation between such permutation-specific sources of error, those arising with, say, $\mathbf{p}_1$ and $\mathbf{p}_2$, depends on the overlapping between the blocks containing x_{it} under $\mathbf{p}_1$ and $\mathbf{p}_2$, and is therefore fairly weak on average;

(II) the expected size of the permutation-specific source of error of $\hat{\mathbf{b}}_i^{*(\mathbf{p})}(L)$ obviously depends on the variance of the idiosyncratic components involved in the block of x_{it}, and also on the "amount of collinearity" of the block, as measured by $\det\left(\hat{\mathbf{C}}^{\chi}_{kk}\right)^{-1}$.

That dependence on the arbitrary ordering of the cross-section is, of course, highly undesirable. As a remedy, we propose to average the estimated impulse–response functions (or forecasts) over the $n_0^* = n_0!/m_0![(q+1)!]^{m_0}$ possible partitions into $(q+1)$-tuples of the n_0 cross-sectional items. Averaging over those n_0^* orderings is equivalent to averaging over the $n_0!$ possible cross-sectional permutations, as two permutations leading to the same partition yield the same estimators, and thus restores permutational invariance (more precisely, permutational equivariance).

Averaging, be it over the n_0^* partitions or over the $n_0!$ permutations, also improves statistical performance. For, let us assume that the ordering of the panel is the result of a random draw under which all $n_0!$ possible permutations are equally likely, each with probability $1/n_0!$ (hence also all n_0^* partitions, each with probability $1/n_0^*$)—equivalently, start by randomly labeling the n_0 cross-sectional items in the panel. The value of the likelihood for that random permutation, whatever its form, is invariant under further permutation. It follows that the un-ordered panel – equivalently, any arbitrarily chosen ordering of the same – is a *sufficient statistic*. To fix the ideas, the panel $\{X_{(i)t},\ i = 1, \dots, n,\ t = 1, \dots, T\}$, where (i) is such that $X_{(1)1} \le X_{(2)1} \le \dots \le X_{(n)1}$, is sufficient, very much in the same way as the *order statistic*[4] is sufficient in an i.i.d. sample—therefore, let us call it the *panel order statistic*.

Conditional on this sufficient statistic, the randomly ordered panel our estimators were computed from is uniformly distributed over the $n_0!$ permutations of the panel order statistic, and averaging the estimators computed from those $n_0!$ permutations (equivalently, averaging the estimators computed from the n_0^* possible partitions into $(q+1)$-tuples) yields the conditional expectation of any of them conditional on the (sufficient) panel order statistic.

The Rao–Blackwell Theorem (see e.g. Lehmann and Casella (1998), p. 47) then tells us that, provided that our estimators have finite expectation and quadratic risk, the quadratic risk of their averaged version is uniformly less than or equal to the original one.

This is a sound theoretical justification for the averaging method. However, computing the estimators for $n_0!$ distinct permutations (for n_0^* distinct partitions) is, even for moderately large values of n_0, numerically infeasible. Fortunately, simulations provide convincing evidence that, selecting a small number of permutations at random and averaging the corresponding estimators of the impulse–response functions $\mathbf{b}_i^*(L)$ leads to

(i) rapidly stabilizing results: the infeasible computation of all n_0^* possible estimators thus is not required for achieving the desired Rao–Blackwellization;

(ii) a substantial reduction of the expected Mean Square Estimation Error (MSE), which is consistent with observation (I) above, that different permutations produce estimators of $b_i^*(L)$ that are affected by weakly correlated sets of idiosyncratic components; see Section 4.3 for details on the simulation and results.

Lastly, if we consider n_0^* infinite sequences of the form

$$x_{i_1,t},\ x_{i_2,t}, \dots, x_{i_{n_0},t},\ x_{n_0+1,t},\ x_{n_0+2,t}, \dots$$

that is, the original infinite sequence with reordering of its first n_0 items, enhancing Assumption 5 within the panel (1.1), see Proposition 12, all the consistency results hold for each of the corresponding n_0^* estimators, and therefore for the average of any subset of them.

4. Simulation experiments

In this section, we use simulated data to compare the estimator proposed in the present paper with estimators based on the existence of a static representation. We focus on (i) estimation of impulse–response functions, (ii) estimation of structural shocks and (iii) one-step-ahead forecasts. Regarding (i) and (ii), we compare FHLZ with the method proposed in Forni et al. (2009), referred to as FGLR. As regards (iii), the results of FHLZ are compared to the method in Stock and Watson (2002a), referred to as SW. Let us recall that both FGLR and SW assume the existence of the static factor representation (1.4), and are based on ordinary principal components. We generate artificial data according to two simple models: (I) a dynamic factor model with no static factor model representation (so that neither FGLR nor SW are consistent) and (II) a model admitting a static factor model representation (under which all methods are consistent).

In our exercises we generate panels with increasing numbers of variables and observations. As the panels are independent (and therefore non-nested), they must be considered as unrelated examples of the observed panel (1.1). However, we use here the notation (n, T) instead of the heavy (n_0, T_0) of Section 3.5.

4.1. Data-generating processes

We consider the following data-generating processes.

Model I (no static factor model representation)

$$x_{it} = a_{i1}(1-\alpha_{i1}L)^{-1}u_{1t} + a_{i2}(1-\alpha_{i2}L)^{-1}u_{2t} + \xi_{it}. \tag{4.9}$$

We generate u_{jt}, $j = 1, 2$ and ξ_{it}, $i = 1, \dots, n$, $t = 1, \dots, T$ as i.i.d. standard Gaussian variables; a_{ij} as independent variables, uniformly distributed on the interval $[-1, 1]$; α_{ij} as independent variables, uniformly distributed on the interval $[-0.8, 0.8]$.

Estimation of the shocks and the impulse–response functions requires an identification rule. Our exercise is based on a Cholesky identification scheme on the first q variables. Precisely, denote by $\mathbf{B}_q(0)$ the matrix with $b_{if}(0)$, $i = 1, 2, \dots, q$, $f = 1, 2, \dots, q$, in the (i, f) entry, and let $\mathbf{H}$ be the lower triangular matrix with positive diagonal entries such that $\mathbf{H}\mathbf{H}' = \mathbf{B}_q(0)\mathbf{B}_q(0)'$. Then, the "structural" shocks, denoted by $\mathbf{u}_t^*$, and the impulse–response functions, denoted by $\mathbf{b}_i^*(L)$, are $\mathbf{b}_i^*(L) = \mathbf{b}_i(L)\mathbf{B}_q(0)^{-1}\mathbf{H}$ and $\mathbf{u}_t^* = \mathbf{H}'\mathbf{B}_q(0)'\mathbf{u}_t$, respectively.

Model II (with static factor representation)

$$\begin{aligned} x_{it} &= \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \dots + \lambda_{ir}F_{rt} + \xi_{it} \\ \mathbf{F}_t &= \mathbf{D}\mathbf{F}_{t-1} + \mathbf{K}\mathbf{u}_t. \end{aligned} \tag{4.10}$$

Here $\mathbf{F}_t = (F_{1t}\ \dots\ F_{rt})'$ and $\mathbf{u}_t = (u_{1t}\ \dots\ u_{qt})'$, $\mathbf{D}$ is $r \times r$ and $\mathbf{K}$ is $r \times q$. Again, u_{jt}, $j = 1, \dots, q$ and ξ_{it}, $i = 1, \dots, n$, $t = 1, \dots, T$ are i.i.d. standard Gaussian and mutually independent white noises. Moreover, λ_{hi}, $h = 1, \dots, r$, $i = 1, \dots, n$ and

[4] Actually, $X_{(1)1}, X_{(2)1}, \dots \le, X_{(n)1}$ is the order statistic of $\mathbf{X}_1 = (X_{11}, X_{21}, \dots, X_{n1})$.

the entries of $\mathbf{K}$ are independently, uniformly distributed on the interval $[-1, 1]$. Finally, the entries of $\mathbf{D}$ are generated as follows: first we generated entries independently, uniformly distributed on the interval $[-1, 1]$; second, we divided the resulting matrix by its spectral norm to obtain unit norm; third, we multiplied the resulting matrix by a random variable uniformly distributed on the interval $[0.4, 0.9]$, to ensure stationarity while preserving sizable dynamic responses. Precisely, $\mathbf{b}_i(L) = \boldsymbol{\lambda}_i(\mathbf{I} - \mathbf{D}L)^{-1}\mathbf{K}$, $\boldsymbol{\lambda}_n$ being the $1 \times r$ matrix having λ_{ih} as its (i, h) entry. To identify the "structural" shocks $\mathbf{u}_t^*$ and the corresponding impulse–response functions $\mathbf{b}_i^*(L)$, we impose a Cholesky identification scheme on the first q variables as in Model I.

4.2. Estimation details and accuracy evaluation

Let $b_{if}^*(L) = \sum_{k=0}^{\infty} b_{if,k}^* L^k$ be the fth entry of $\mathbf{b}_i^*(L)$. Our target is the estimation of $b_{if,k}^*$, $i = 1, \ldots, n$, $f = 1, \ldots, q$, $k = 0, \ldots, K$ and u_{ft}^*, $f = 1, \ldots, q$, $t = 1, \ldots, T$, as well as the forecast of $x_{i,T+1}$, $i = 1, \ldots, n$.

The structural impulse–response functions and the structural shocks are estimated by FHLZ and FGLR. Both methods require the calibration of some parameters. As regards FHLZ, we must specify three things.

(i) The lag-window size in the estimation of the spectral density $\boldsymbol{\Sigma}^x(\theta)$. We use a Bartlett lag window of size $B_T = \sqrt{T}$. Then the spectral density $\boldsymbol{\Sigma}^\chi(\theta)$ and the covariances $\gamma_{ij,\ell}^\chi$ are estimated as described in Section 3.2.

(ii) The number q of structural shocks. This number is assumed to be known when estimating the structural shocks and impulse–response functions. Identification is obtained by imposing the same Cholesky scheme as above.

(iii) The order of each $(q + 1)$-dimensional VAR matrix $\mathbf{A}^k(L)$. These orders are determined via the BIC criterion.

As regards FGLR, we estimate a VAR for the principal components of the data. The number of principal components is either assumed known or determined by Bai and Ng's IC_{p2} criterion, the number of lags is determined by the BIC criterion.

FHLZ forecasts are computed by filtering the estimated shocks with the estimated impulse–response functions

$$\hat{x}_{i,T+1} = \sum_{f=1}^{q} \left(\hat{b}_{if,1}^* \hat{u}_{fT}^* + \hat{b}_{if,2}^* \hat{u}_{f,T-1}^* + \cdots \right).$$

The number of structural shocks is no longer assumed to be known. Rather, it is estimated by the Hallin and Liška (2007) method.[5] SW forecasts are obtained by regressing $x_{i,T+1}$ onto either the ordinary principal components at T and x_{iT}, or the principal components at T alone. The former method corresponds to the original Stock and Watson (2002a) method; the latter is motivated by the fact that, in both models (4.9) and (4.10) above, the idiosyncratic components are serially uncorrelated. The number of principal components is determined with Bai and Ng's IC_{p2} criterion.

The estimation error for the impulse–response functions is defined as the normalized sum of the squared deviations of the estimated from the "structural" impulse–response coefficients. Precisely, let $\hat{b}_{if,k}^*$ be the estimated impulse–response coefficient of variable i for shock f at lag k: the estimation error on the impulse–response functions is measured by

$$\frac{\sum_{i=1}^n \sum_{f=1}^q \sum_{k=0}^K \left(\hat{b}_{if,h}^* - b_{if,h}^* \right)^2}{\sum_{i=1}^n \sum_{f=1}^q \sum_{k=0}^K (b_{if,k}^*)^2}. \tag{4.11}$$

The truncation lag K is set to 60. Similarly, denoting by $\hat{u}_{ft}^*$ the estimate of u_{ft}^*, the estimation error on the "structural" shocks is measured by

$$\frac{\sum_{f=1}^q \sum_{t=1}^T \left(\hat{u}_{ft}^* - u_{ft}^* \right)^2}{\sum_{f=1}^q \sum_{t=1}^T (u_{ft}^*)^2}.$$

Finally, the accuracy of the forecast is measured by the sum of the squared deviations of the forecasts from the unfeasible target obtained by filtering the true structural shocks with the true structural impulse–response functions, i.e. $x_{i,T+1}^P = \sum_{f=1}^q \sum_{k=1}^T b_{if,k}^* u_{f,T+1-k}^*$. Again, we normalize by dividing by the sum of squared target values:

$$\frac{\sum_{i=1}^n \left(\hat{x}_{i,T+1} - x_{i,T+1}^P \right)^2}{\sum_{i=1}^n (x_{i,T+1}^P)^2}.$$

Model I is evaluated for different sample size combinations, with $n = 30, 60, 120, 240$ and $T = 60, 120, 240, 480$. Model II is evaluated for a fixed sample size of $n = 120$ and $T = 240$, but different configurations of q and r, i.e. $r = 4, 6, 8, 12$ and $q = 2, 4, 6$, with $r > q$.[6] For each couple (n, T) in Model I, and each (r, q) in Model II, the MSE's corresponding to different estimators are averaged over 500 replications.

Regarding FHLZ, note that in Model I we chose the values of n such that $n = m(q + 1)$. However, the same does not hold for Model II when $n = 6$, nor of course in empirical applications, the one in Section 5 in particular. In general, when $n = m(q + 1) + \tilde{n}$, with $\tilde{n} \neq 0$, each permutation, and subsequent estimation, is obtained by selecting randomly $n - \tilde{n}$ series from the whole n-sized dataset. Then the estimates of impulse–response functions or forecasts relative to variable i are averaged according to the number of times the variable i has been selected.

4.3. Cross-sectional permutations

As argued in Section 3.5, the estimators obtained by the FHLZ method should be averaged over different permutations of cross-sectional items. In order to study the influence of such permutations, for various values of n and T, we simulated 500 datasets from Model I, denoted by d_k, $k = 1, \ldots, 500$. For each of the resulting panels, we computed (with the Cholesky identification rule described in Section 4.1) the estimated impulse–response functions averaged over $\mu = 1, \ldots, M$ randomly chosen permutations:

$$\text{MSE}_{n,T,\mu,d_k} = \frac{1}{\mu} \sum_{j=1}^{\mu} \text{MSE}_{n,T,d_k}^{\mathbf{p}_j},$$

where $\text{MSE}_{n,T,d_k}^{\mathbf{p}_j}$ is the ratio (4.11) relative to the panel d_k and permutation $\mathbf{p}_j$. The expectation of MSE_{n,T,μ,d_k}, given n, T and μ, is estimated by the average of MSE_{n,T,μ,d_k} over the 500 panels d_k and denoted by $\text{MSE}_{n,T,\mu}$. The graphs of $\text{MSE}_{n,T,\mu}$ as a function of μ, for some values of n and T, are plotted in Fig. 1. We see that

(i) as expected, estimates corresponding to different random permutations do differ;
(ii) averaging those estimates yields a clear improvement of the expected MSE;
(iii) the rate of that improvement declines steadily as the number μ of permutations increases, and rapidly stabilizes until additional permutations produce a negligible effect;
(iv) as n and T increase, the improvement decreases, both in absolute and relative terms, and the number of permutations required for "stabilization" decreases: 10 for ($n = 60$ and $T = 120$), only 5 for ($n = 240$ and $T = 480$).

[5] We used the log criterion $IC_{2;n}^T$ with penalty function p_1 and lag window equal to $\sqrt{T}$. The "second stability interval" was evaluated over the grid $n_j = \lfloor(3n/4 + jn/40)\rfloor$, $T_j = T$, $j = 1, \ldots, 10$.

[6] We impose $r > q$ since for the case $r = q$, in the FHLZ method, the regressors of the $(q + 1)$-dimensional VARs are asymptotically collinear.

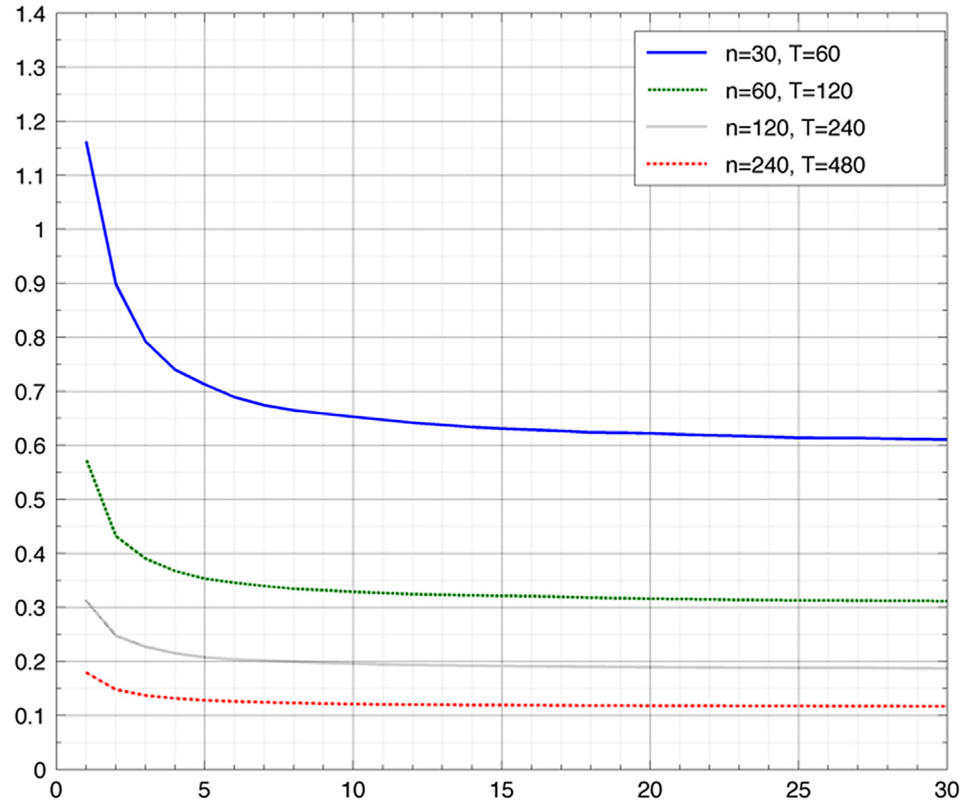

Fig. 1. Model I, MSE for impulse–response functions, averaged firstly over μ permutations of the n variables, and secondly over 500 simulated datasets, as a function of $\mu = 1, 2, \ldots, 30$, four combinations of n and T.

Summing up, in our model the estimator obtained by averaging over random permutations stabilizes quite rapidly and provides a substantial reduction of the expected MSE. Using the US monthly macroeconomic dataset known as the Stock and Watson dataset, Forni et al. (2016) also obtain stabilization of the average estimators with a small number of permutations.

4.4. Results

We now turn to a performance comparison between the FHLZ method and its competitors. Table 1 in online Appendix F, reports the results for the estimation of impulse–response functions and structural shocks, Model I. The upper panel reports results for the FHLZ method without averaging; the central panel for the FHLZ with averaging over 30 reorderings; the lower panel for the FGLR method. The estimates obtained with FGLR, although theoretically inconsistent, do approach the target as n and T get larger. This is because the number of estimated static factors increases with n and T, so that the static model achieves a fairly good approximation of the underlying "infinite-factor" model.[7] Despite of this, FHLZ clearly outperforms FGLR. Regarding impulse–response functions, FHLZ, with and without averaging, dominates the static method for all (n, T) configurations. The error is up to 50%–60% smaller than the corresponding one for FGLR. As for the shocks, the performance of FHLZ with averaging is similar to that of FGLR for large T, but dominates FGLR for small T. Forecast results are reported in Table 2. Not surprisingly, the SW method (central and lower panels) performs better when lagged xs are not included among the regressors, owing to the fact that the idiosyncratic components are serially uncorrelated. Indeed, we are comparing forecasts of the common components of the xs, i.e. the χs, rather than the xs themselves. FHLZ forecasts (with averaging) outperform SW for all (n, T) configurations, with an improvement ranging from 20 to 40%.[8] Observe that here we no longer impose the correct q, but estimate it via the Hallin and Liška (2007) criterion, so that both forecasts in the upper and central panels are feasible.

Table 3 reports results for Model II, estimation of impulse–response functions and structural shocks. Here both FHLZ and FGLR are consistent. Somewhat surprisingly, FHLZ (with averaging, upper panel) outperforms FGLR for all (r, q) configurations. With this model, Bai and Ng's criterion tends to underestimate the number of factors.[9] Hence, we computed the (unfeasible) FGLR estimation obtained by imposing the correct r (lower panel), to see whether the above result can be blamed on the underestimation of r. In general, FGLR performs better when imposing the correct number of factors; nonetheless, FHLZ still exhibits the best performance in most cases.

Forecast errors, reported in Table 4, confirm the result that FHLZ performs better than SW for most (r, q) configurations.

5. Empirical application

In this section, we present a pseudo real-time forecast evaluation exercise with US quarterly data. We take as target variables real GDP, real private fixed investment, real consumption expenditures, the number of unemployed and the consumer price index. We compare results obtained with FHLZ, SW and a simple univariate autoregressive model. The forecasts are computed within a rolling window scheme. An extensive pseudo real-time forecasting analysis based on US monthly data is found in Forni et al. (2016).

7 The average $\hat{r}$ is 2.01 for $n = 30, T = 60$ and 4.00 for $n = 240, T = 480$.

8 FHLZ without averaging, not reported here, performs better than SW but not as well as FHLZ with averaging, in line with the results in Table 1.

9 On average, $\hat{r}$ is smaller than r for all n and T configurations.

5.1. Data and methods

We use the same dataset as in Forni and Gambetti (2014), complemented with the inclusion of twelve additional series, taken from the Survey of Professional Forecasters, thus obtaining $n = 73$. The time span is 1968:Q4–2010:Q4.[10] The data set includes NIPA series, industrial production, employment and unemployment data, prices, interest rates, money, credit and financial data, as well as leading indicators and survey series. To get stationarity, we take first differences of logs for real variables and second differences of logs for price indexes and money aggregates. The complete list of the series, along with data treatment details, is reported in online Appendix G.

After transformation, our series range from 1969:Q1 to 2010:Q4, thus $T = 168$. We choose $t = 1985{:}Q4$ as the starting date for forecasting, so that 68 observations are used for the first estimation. We then proceed with a rolling window of length 68 quarters (17 years). At each t, $t = 1985{:}Q4, \ldots, 2009{:}Q4$, we compute h-quarter ahead forecasts for horizons $h = 1, 2, 3, 4$, thus $101 - h$ forecasts for each h.

If x_{it} denotes the transformed variable, our target is $x_{i,t+h} - x_{it}$; hence, we predict the growth rate of GDP, investment and consumption, the percentage change of the number of unemployed and the inflation rate variation, between t and $t + h$.

We compare the forecasts obtained by the dynamic method (FHLZ), the static method (SW), and a univariate AR model. Following Stock and Watson (2012), we use an AR(4) model for all variables.

As regards the calibration of the parameters in FHLZ (see Section 4.2),

(i) We use, as in the simulation exercise, the rule of thumb $B_T = \sqrt{T}$, which gives $B_T = 8$. However, we also report some results obtained with $B_T = 12$.[11]

(ii) The number of factors is kept fixed across all t. The Hallin and Liška log criterion applied in the first sub-sample 1969:Q1–1985:Q4 gives $q = 4$.[12] The results with 2, 3, 4 and 5 factors are similar, those with $q = 2$ being slightly worse. In Tables 5 and 6 we report results for three factor-window combinations: $(q = 3, B_T = 8)$, $(q = 4, B_T = 8)$ and $(q = 3, B_T = 12)$.

(iii) As in the previous section, we use the BIC criterion to set the number of lags in the $(q + 1)$-dimensional VARs.

FHLZ forecasts are computed by averaging across 30 random permutations of the original ordering, as in the previous Section.[13]

Static factor forecasts are computed by regressing the target variable onto the constant, the first r ordinary principal components of the standardized data, and p lags of the dependent variable.

(I) We keep the number of static factors r fixed for all t. The Bai and Ng IC_{p2} criterion finds 6 factors for the first sub-sample 1969:Q1–1985:Q4. However, when using the refinement proposed by Alessi et al. (2010), we are left with just 2 factors. The same result of 2 factors is found with the Ahn and Horenstein Eigenvalue Ratio and Growth Ratio criteria. We tried 2, 4 and 6 factors.

(II) As for p, we tried a fixed $p = 0, 1, 4$ and a floating p determined by the BIC criterion. We found that generally the inclusion of the dependent variable does not improve results.[14]

10 The starting date of the sample is that of the series reported in the Survey.

11 We found very similar results for $B_T = 12$ and $B_T = 16$, whereas results are considerably worse for $B_T = 4$.

12 We used the log criterion $IC^T_{2;n}$ with penalty function p_1 and $B_T = 8$. The "second stability interval" was evaluated over the grid $n_j = n - 15 + j$, $T_j = 68$, $j = 1, \ldots, 15$. We kept fixed T_j since the number of time observations is relatively small.

13 We first standardize the variables and estimate the model. Then we filter the estimated shocks with the estimated impulse–response functions (truncation lag = 60) and compute the simple average of the forecasts obtained with different permutations. Finally we restore the original mean and standard deviation.

14 A similar result is found in D'Agostino and Giannone (2012).

In Tables 5 and 6, we report results for four (r, p) specifications: $(r = 2, p = 0)$, $(r = 4, p = 0)$, $(r = 6, p = 0)$, and $(r = 4, p = 1)$.

The forecasting performance is evaluated by the mean square forecast error, normalized by the mean square error of the autoregressive forecasts. We refer to this measure as the Relative Mean Square Forecast Error (RMSFE). For example, a RMSFE equal to 0.8 means that the mean square error is 20% smaller than that of AR(4) forecasts.

5.2. Results

The results are reported in online Appendix F. Inspection of Table 5 reveals the following main facts.

(I) Factor models outperform the autoregressive model, according to the RMSFE, with the noticeable exception of consumption.

(II) FHLZ generally outperforms the static method. For GDP and inflation, FHLZ has the best RMSFE at all horizons, whereas for investment and unemployment results are mixed.

(III) The improvement of FHLZ with respect to the benchmark AR(4) model is substantial, particularly at longer horizons, reaching about 30% for investment and inflation and 15%–20% for GDP and unemployment. In several cases, the improvement is significant according to the Diebold–Mariano test, particularly for GDP and unemployment.

In Table 6, we report results for government spending, imports, exports, labor productivity, total factor productivity, the federal funds rate and stock prices, along with the average RMSFE for the NIPA series included in the data set. Results are broadly in line with the above findings. The performance of the dynamic method is very good for all variables, as compared to the AR(4) benchmark, with the exception of government spending and the industrial production index, $h = 1$. Moreover FHLZ outperforms the static method for most variables, particularly for $h > 1$.

We conclude that FHLZ, in this real-life dataset, outperforms standard techniques in the prediction of most macroeconomic variables, including GDP and inflation.

6. Conclusions

An estimate of the common-component spectral density matrix $\hat{\Sigma}^\chi$ is obtained using the frequency-domain principal components of the observations x_{it}. The central idea of the present paper is that, because $\hat{\Sigma}^\chi$ has large dimension but small rank q, a factorization of $\hat{\Sigma}^\chi$ can be obtained piecewise. Precisely, the factorization of $\hat{\Sigma}^\chi$ only requires the factorization of $(q + 1)$-dimensional subvectors of $\boldsymbol{\chi}_t$. Under our assumption of a rational spectral density for the common components, this implies that the number of parameters to be estimated grows as n, not n^2.

The rational spectral density assumption also has the important consequences that $\boldsymbol{\chi}_t$ has a finite autoregressive representation and that the dynamic factor model can be transformed into the static one $\mathbf{z}_t = \mathbf{R}\mathbf{v}_t + \boldsymbol{\phi}_t$, where $\mathbf{z}_t = \mathbf{A}(L)\mathbf{x}_t$. We construct estimators for $\mathbf{A}(L)$, $\mathbf{R}$ and $\mathbf{v}_t$ starting with a standard non-parametric estimator of the spectral density of the xs. This implies a slower rate of convergence as compared to the usual $T^{-1/2}$. However, in Section 3, we prove that our estimators for $\mathbf{A}(L)$, $\mathbf{R}$ and $\mathbf{v}_t$ do not undergo any further reduction in their speed of convergence.

The main difference of the present paper with respect to the previous literature on GDFM's is that although we make use of a parametric structure for the common components, we do not make the standard, but quite restrictive assumption that our dynamic factor model has a static representation of the form (1.4). Sections 4 and 5 provide important empirical support to the richer dynamic structure of unrestricted GDFM's.

Acknowledgments

The first and third authors' research was supported by the PRIN-MIUR Grant 2010J3LZEN. Second author's research was supported by an Interuniversity Attraction Pole P7/06 (2012–2017) of the Belgian Science Policy Office. Fourth author's research was supported by the ESRC Grant RES-000-22-3219.

Appendix A. Proof of Proposition 6

Adding and subtracting $E(\hat{\sigma}^x_{ij}(\theta^*_h))$ within the absolute value in $E\left(\max_{|h|\le B_T}\left|\hat{\sigma}^x_{ij}(\theta^*_h)-\sigma^x_{ij}(\theta^*_h)\right|^2\right)$ and re-arranging yields

$$E\Big(\max_{|h|\le B_T}\left|\hat{\sigma}^x_{ij}(\theta^*_h)-\sigma^x_{ij}(\theta^*_h)\right|^2\Big)\le 2\Big[E\Big(\max_{|h|\le B_T}\left|\hat{\sigma}^x_{ij}(\theta^*_h)-E\hat{\sigma}^x_{ij}(\theta^*_h)\right|^2\Big)$$
$$+E\Big(\max_{|h|\le B_T}\left|E\hat{\sigma}^x_{ij}(\theta^*_h)-\sigma^x_{ij}(\theta^*_h)\right|^2\Big)\Big]$$

(C_r-inequality for $r=2$). The first term (variance) on the right-hand side of the above inequality satisfies

$$E\Big(\max_{|h|\le B_T}\left|\hat{\sigma}^x_{ij}(\theta^*_h)-E\hat{\sigma}^x_{ij}(\theta^*_h)\right|^2\Big)\le C^*(B_T\log B_T/T),$$

where C^* depends only on p (see Assumption 8), ρ_1 (see Proposition 5), δ (see Assumption 9). This is proved by letting $\nu=2$ in Lemma 10 of Wu and Zaffaroni (2017).

As for the second term (the squared bias), simple calculations give

$$\begin{aligned}\mathcal{S}_{ij}(\theta)&=2\pi\left(E\hat{\sigma}^x_{ij}(\theta)-\sigma^x_{ij}(\theta)\right)\\&=\textstyle\sum_{l=-T+1}^{T-1}\left(1-\frac{|l|}{T}\right)K\left(\frac{l}{B_T}\right)\gamma^x_{ij,l}\,e^{-\imath l\theta}-\sum_{l=-\infty}^{\infty}\gamma^x_{ij,l}\,e^{-\imath l\theta}\\&\le\left|\textstyle\sum_{l=-T+1}^{T-1}\left(K\left(\frac{l}{B_T}\right)-1\right)\gamma^x_{ij,l}\,e^{-\imath l\theta}\right|\\&\quad+\left|\textstyle\sum_{l=-T+1}^{T-1}K\left(\frac{l}{B_T}\right)\frac{|l|}{T}\gamma^x_{ij,l}\,e^{-\imath l\theta}\right|+\left|\textstyle\sum_{|l|\ge T}\gamma^x_{ij,l}\,e^{-\imath l\theta}\right|\\&=\mathcal{A}^T_{ij}(\theta)+\mathcal{B}^T_{ij}(\theta)+\mathcal{C}^T_{ij}(\theta).\end{aligned}$$

Assumptions 2 and 4 imply that, for some $\phi\in(0,1)$ and some D, $|\gamma^x_{ij,l}|\le|\gamma^\chi_{ij,l}|+|\gamma^\xi_{ij,l}|\le D\phi^{|l|}$, for all i and j (see Eq. (2.9))). This inequality and Assumption 9(i) ensure that, for some F and all i, j and θ, $\mathcal{A}^T_{ij}(\theta)\le FD\sum_{l=-\infty}^{\infty}\phi^{|l|}(|l|/B_T)^\kappa\le[2DF\phi/(1-\phi)^2]T^{-\delta\kappa}=HT^{-\delta\kappa}$. Moreover, $\mathcal{B}^T_{ij}(\theta)\le DT^{-1}\sum_{l=-\infty}^{\infty}\phi^{|l|}|l|=[2D\phi/(1-\phi)^2]T^{-1}=KT^{-1}$, for all i, j, and θ. Finally, $\mathcal{C}^T_{ij}(\theta)\le D\sum_{|l|\ge T}\phi^{|l|}|l|^\kappa/T^\kappa$, since $|l|^\kappa/T^\kappa\ge1$ for $|l|\ge T$. Hence, it follows that $\mathcal{C}^T_{ij}(\theta)\le MT^{-\kappa}\le MT^{-\delta\kappa}$ for all i,j and θ. Thus,

$$\mathcal{S}_{ij}(\theta)/2\pi\le KT^{-1}+(H+M)T^{-\delta\kappa}\le PT^{-\mu},$$

where $\mu=\min(\delta\kappa,1)$, for all i,j and θ. Now, $2\delta\kappa>1-\underline{\delta}>1-\delta$, by Assumption 9(ii). Hence, $\max_{|h|\le B_T}\left|E\hat{\sigma}^x_{ij}(\theta^*_h)-\sigma^x_{ij}(\theta^*_h)\right|^2\le P^2T^{-2\mu}\le C^{**}(B_T\log B_T/T)$ for all i and j. □

Appendix B. Proof of Proposition 7

The proof below closely follows Forni et al. (2009). Denote by $\mu_j(\mathbf{A}), j=1,2,\dots,s$, the (real) eigenvalues, in decreasing order, of an $s\times s$ Hermitian matrix $\mathbf{A}$, and by $\|\mathbf{B}\|=\sqrt{\mu_1(\tilde{\mathbf{B}}\mathbf{B})}$ the spectral norm of an $s_1\times s_2$ matrix $\mathbf{B}$. The norm $\|\mathbf{B}\|$ coincides with the Euclidean norm of $\mathbf{B}$ when $\mathbf{B}$ is a column matrix and is equal to $|\mu_1(\mathbf{B})|$ when $\mathbf{B}$ is square and Hermitian. Recall that, if $\mathbf{B}_1$ is $s_1\times s_2$ and $\mathbf{B}_2$ is $s_2\times s_3$, then

$$\|\mathbf{B}_1\mathbf{B}_2\|\le\|\mathbf{B}_1\|\|\mathbf{B}_2\|.\tag{B.1}$$

We will use the fact that, for any two $s\times s$ Hermitian matrices $\mathbf{A}_1$ and $\mathbf{A}_2$,

$$|\mu_j(\mathbf{A}_1+\mathbf{A}_2)-\mu_j(\mathbf{A}_1)|\le\|\mathbf{A}_2\|,\quad j=1,\dots,s.\tag{B.2}$$

This fact is an obvious consequence of Weyl's inequality $\mu_j(\mathbf{A}_1+\mathbf{A}_2)\le\mu_j(\mathbf{A}_1)+\mu_1(\mathbf{A}_2)$ (Franklin, 2000, p. 157, Theorem 1). The proof of Proposition 7 is divided into several intermediate propositions. Denote by $\mathcal{S}_i$ the $n\times1$ matrix with 1 in entries $(i,1)$ and 0 elsewhere, so that $\mathcal{S}'_1\mathbf{A}$ is the ith row of $\mathbf{A}$, and define $\rho_T=T/B_T\log B_T$.

As most of the arguments below depend on equalities and inequalities that hold for all $\theta\in[-\pi,\pi]$, the notation has been simplified by dropping θ. Also, properties holding for $\max_{|h|\le B_T}F(\theta_h)$, where F is some function of θ, are often phrased as holding for *F uniformly in* θ. The meaning of uniformity in i, or i and j, has been clarified in the statement of Proposition 7.

All lemmas in this Appendix hold and are proved under Assumption 1 through 10.

Lemma 1. *As $T\to\infty$ and $n\to\infty$,*
(i) $\max_{|h|\le B_T}n^{-1}\|\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^x\|=O_P(\rho_T^{-1/2})$;
(ii) $\max_{|h|\le B_T}n^{-1/2}\|\mathcal{S}'_i(\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^x)\|=O_P(\rho_T^{-1/2})$ *uniformly in i;*
(iii) $\max_{|h|\le B_T}n^{-1}\|\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^\chi\|=O_P(\max(n^{-1},\rho_T^{-1/2}))$;
(iv) $\max_{|h|\le B_T}n^{-1/2}\|\mathcal{S}'_i(\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^\chi)\|=O_P(\max(n^{-1/2},\rho_T^{-1/2}))=O_P(\zeta_{nT})$ *uniformly in i.*

Proof. We have

$$\mu_1((\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^x)(\tilde{\hat{\mathbf{\Sigma}}}^x-\tilde{\mathbf{\Sigma}}^x))\le\operatorname{trace}((\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^x)(\tilde{\hat{\mathbf{\Sigma}}}^x-\tilde{\mathbf{\Sigma}}^x))$$
$$=\sum_{i=1}^n\sum_{j=1}^n|\hat{\sigma}^x_{ij}-\sigma^x_{ij}|^2.$$

Using (3.4) and the Markov inequality,

$$n^{-2}\max_{|h|\le B_T}\sum_{i=1}^n\sum_{j=1}^n|\hat{\sigma}^x_{ij}-\sigma^x_{ij}|^2\le n^{-2}\sum_{i=1}^n\sum_{j=1}^n\max_{|h|\le B_T}|\hat{\sigma}^x_{ij}-\sigma^x_{ij}|^2$$
$$\le C\rho_T^{-1}.$$

Statement (i) follows. In the same way,

$$n^{-1}\mathcal{S}'_i(\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^x)(\tilde{\hat{\mathbf{\Sigma}}}^x-\tilde{\mathbf{\Sigma}}^x)\mathcal{S}_i=n^{-1}\sum_{j=1}^n|\hat{\sigma}^x_{ij}-\sigma^x_{ij}|^2\le C\rho_T^{-1},$$

where C is independent of i. Statement (ii) follows. As regards (iii), $\mathbf{\Sigma}^x=\mathbf{\Sigma}^\chi+\mathbf{\Sigma}^\xi$ implies that $\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^\chi=\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^x+\mathbf{\Sigma}^\xi$, so that, by the triangle inequality for matrix norm,

$$\|\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^\chi\|\le\|\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^x\|+\|\mathbf{\Sigma}^\xi\|.$$

The statement follows from (i) and the fact that $\|\mathbf{\Sigma}^\xi\|=\lambda^\xi_1$ is bounded. Statement (iv) is obtained in a similar way, using (ii) instead of (i). □

Lemma 2. *As $T\to\infty$ and $n\to\infty$,*
(i) $\max_{|h|\le B_T}n^{-1}\left|\hat{\lambda}^x_f-\lambda^\chi_f\right|=O_P(\max(n^{-1},\rho_T^{-1/2}))$ *for* $f=1,2,\dots,q$;
(ii) *letting*

$$\mathbf{G}^\chi=\begin{cases}\mathbf{I}_q & \text{if } \lambda^\chi_q=0,\\ n(\mathbf{\Lambda}^\chi)^{-1} & \text{otherwise,}\end{cases}\quad\text{and}$$
$$\hat{\mathbf{G}}^x=\begin{cases}\mathbf{I}_q & \text{if } \hat{\lambda}^x_q=0,\\ n(\hat{\mathbf{\Lambda}}^x)^{-1} & \text{otherwise,}\end{cases}$$

$\max_{|h|\le B_T}n^{-1}\|\mathbf{\Lambda}^\chi\|$ *and* $\max_{|h|\le B_T}\|\mathbf{G}^\chi\|$ *are* $O(1)$, $\max_{|h|\le B_T}n^{-1}\|\hat{\mathbf{\Lambda}}^x\|$ *and* $\max_{|h|\le B_T}\|\hat{\mathbf{G}}^x\|$ *are* $O_P(1)$.

Proof. Setting $\mathbf{A}_1=\mathbf{\Sigma}^\chi$ and $\mathbf{A}_2=\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^\chi$, (B.2) yields $|\hat{\lambda}^x_f-\lambda^\chi_f|\le\|\hat{\mathbf{\Sigma}}^x-\mathbf{\Sigma}^\chi\|$; hence, statement (i) follows from Lemma 1(iii). Boundedness of $n^{-1}\|\mathbf{\Lambda}^\chi\|$ and $\|\mathbf{G}^\chi\|$, uniformly in θ, is a consequence of Assumption 3. Boundedness in probability of $n^{-1}\|\hat{\mathbf{\Lambda}}^x\|$ and $\|\hat{\mathbf{G}}^x\|$, uniformly in θ, follows from statement (i). □

Lemma 3. *As $T \to \infty$ and $n \to \infty$,*

(i) $\max_{|h|\le B_T} n^{-1}\|\tilde{\tilde{\mathbf{P}}}^x\hat{\mathbf{P}}^x\hat{\Lambda}^x - \Lambda^x\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x\| = O_P(\max(n^{-1}, \rho_T^{-1/2}))$;

(ii) $\max_{|h|\le B_T}\|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x - \mathbf{I}_q\| = O_P(\max(n^{-1}, \rho_T^{-1/2}))$;

(iii) *there exist diagonal complex orthogonal matrices* $\hat{\mathbf{W}}_q = \mathrm{diag}\,(\hat{w}_1\ \hat{w}_2\ \cdots\ \hat{w}_q)$, $|\hat{w}_j|^2 = 1, j = 1, \dots, q$ *depending on n and T, such that* $\max_{|h|\le B_T}\|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x - \hat{\mathbf{W}}_q\| = O_P(\max(n^{-1}, \rho_T^{-1/2}))$.

Proof. Using inequality (B.1) and the fact that $\|\tilde{\mathbf{P}}^x\| = \|\hat{\mathbf{P}}^x\| = 1$, we have

$$\|\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x\hat{\Lambda}^x - \Lambda^x\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x\| = \|\tilde{\mathbf{P}}^x(\hat{\Sigma}^x - \Sigma^x)\hat{\mathbf{P}}^x\| \le \|\hat{\Sigma}^x - \Sigma^x\|.$$

Statement (i) thus follows from Lemma 1(iii). Turning to (ii), set

$$\mathbf{a} = \tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x,$$
$$\mathbf{b} = \left[\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x\right]n^{-1}\hat{\Lambda}^x\hat{\mathbf{G}}^x = \tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x\left[\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x n^{-1}\hat{\Lambda}^x\right]\hat{\mathbf{G}}^x,$$
$$\mathbf{c} = \tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x\left[n^{-1}\Lambda^x\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x\right]\hat{\mathbf{G}}^x = \left[n^{-1}\tilde{\hat{\mathbf{P}}}^x\Sigma^x\hat{\mathbf{P}}^x\right]\hat{\mathbf{G}}^x,$$
$$\mathbf{d} = \left[n^{-1}\tilde{\hat{\mathbf{P}}}^x\hat{\Sigma}^x\hat{\mathbf{P}}^x\right]\hat{\mathbf{G}}^x = n^{-1}\hat{\Lambda}^x\hat{\mathbf{G}}^x,$$

and $\mathbf{f} = \mathbf{I}_q$: we have

$$\left\|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x - \mathbf{I}_q\right\| \le \|\mathbf{a}-\mathbf{b}\| + \|\mathbf{b}-\mathbf{c}\| + \|\mathbf{c}-\mathbf{d}\| + \|\mathbf{d}-\mathbf{f}\|. \tag{B.3}$$

Using Lemma 2, statement (i), and the boundedness in probability, uniformly in θ, of $\|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x\|$, $\|\hat{\mathbf{G}}^x\|$ and $\|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x\|$, all terms on the right-hand side of inequality (B.3) can be shown to be $O_P(\max(n^{-1}, \rho_T^{-1/2}))$, uniformly in θ.

Turning to (iii), note that, from statement (i), $n^{-1}\tilde{\hat{\mathbf{P}}}_h^x\mathbf{P}_k^x(\lambda_k^x - \hat{\lambda}_h^x) = O_P(\max(n^{-1}, \rho_T^{-1/2}))$. Assumption 3 (asymptotic separation of the eigenvalues $\lambda_f^x(\theta)$) implies that, for $h \ne k$, $\tilde{\hat{\mathbf{P}}}_h^x\mathbf{P}_k^x = O_P(\max(n^{-1}, \rho_T^{-1/2}))$. Moreover, $\sum_{f=1}^q|\tilde{\hat{\mathbf{P}}}_h^x\mathbf{P}_f^x|^2 - 1 = O_P(\max(n^{-1}, \rho_T^{-1/2}))$ from statement (ii). Therefore,

$$|\tilde{\hat{\mathbf{P}}}_h^x\mathbf{P}_h^x|^2 - 1 = (|\tilde{\hat{\mathbf{P}}}_h^x\mathbf{P}_h^x| - 1)(|\tilde{\mathbf{P}}_h^x\hat{\mathbf{P}}_h^x| + 1) = O_P(\max(n^{-1}, \rho_T^{-1/2})).$$

The conclusion follows. □

Note that Lemma 3 clearly also holds for $n^{-1}\|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x\Lambda^x - \hat{\Lambda}^x\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x\|$, $\|\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x - \mathbf{I}_q\|$ and $\|\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x - \tilde{\hat{\mathbf{W}}}_q\|$.

Lemma 4. *As $T \to \infty$ and $n \to \infty$,*

$$\max_{|h|\le B_T}\|\mathcal{S}_i'\big(\mathbf{P}^x(\Lambda^x)^{1/2}\hat{\mathbf{W}}_q - \hat{\mathbf{P}}^x(\hat{\Lambda}^x)^{1/2}\big)\| = O_P(\zeta_{nT}), \tag{B.4}$$

uniformly in i.

Proof. We have

$$\begin{aligned}&\|\mathcal{S}_i'(\mathbf{P}^x(\Lambda^x)^{1/2}\hat{\mathbf{W}}_q - \hat{\mathbf{P}}^x(\hat{\Lambda}^x)^{1/2})\|\\ &\quad\le \|\mathcal{S}_i'(n^{1/2}\mathbf{P}^x\hat{\mathbf{W}}_q - n^{1/2}\hat{\mathbf{P}}^x)(n^{-1}\Lambda^x)^{1/2}\|\\ &\qquad+ \|\mathcal{S}_i'\hat{\mathbf{P}}^x(n^{-1/2}(\Lambda^x)^{1/2} - n^{-1/2}(\hat{\Lambda}^x)^{1/2})\|.\end{aligned}$$

By Lemma 2(i), thus, we only need to prove that

$$\|n^{1/2}\mathcal{S}_i'\mathbf{P}^x\hat{\mathbf{W}}_q - n^{1/2}\mathcal{S}_i'\hat{\mathbf{P}}^x\| = O_P(\max(n^{-1/2}, \rho_T^{-1/2})).$$

Firstly, we show that

$$\|n^{1/2}\mathcal{S}_i'\mathbf{P}^x\| \le \mathcal{A}, \tag{B.5}$$

for some $\mathcal{A}$ and all θ and i. Assumption 2 implies that $\sigma_{ii}^x = \sum_{f=1}^q\lambda_f^x|p_{if}^x|^2 \le \mathcal{B}$, for some $\mathcal{B}$ and all θ and i. As all the terms in the sum are positive, $\lambda_f^x|p_{if}^x|^2 = (\lambda_f^x/n)n|p_{if}^x|^2 \le \mathcal{B}$, for all θ and i. By Assumption 3, $\lambda_f^x/n \ge \mathcal{C} > 0$ for all θ and f, so that $n|p_{if}^x|^2 \le \mathcal{D}$ for all θ and i. Hence, $n\,\mathcal{S}_i'\mathbf{P}^x\tilde{\mathbf{P}}^x\mathcal{S}_i$ is bounded uniformly in θ and i; (B.5) follows. Next, define

$$\mathbf{g} = n^{1/2}\mathcal{S}_i'\mathbf{P}^x\left[\hat{\mathbf{W}}_q\right],$$
$$\mathbf{h} = n^{1/2}\mathcal{S}_i'\mathbf{P}^x\left[\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x\right] = n^{1/2}\mathcal{S}_i'\mathbf{P}^x[\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x\hat{\mathbf{\Lambda}}^x/n](\hat{\mathbf{\Lambda}}^x/n)^{-1},$$
$$\mathbf{i} = n^{1/2}\mathcal{S}_i'\mathbf{P}^x[(\mathbf{\Lambda}^x/n)\tilde{\mathbf{P}}^x\hat{\mathbf{P}}^x](\hat{\mathbf{\Lambda}}^x/n)^{-1} = [n^{-1/2}\mathcal{S}_i'\Sigma^x]\hat{\mathbf{P}}^x(\hat{\mathbf{\Lambda}}^x/n)^{-1},$$

and

$$\mathbf{j} = [n^{-1/2}\mathcal{S}_i'\hat{\Sigma}^x]\hat{\mathbf{P}}^x(\hat{\mathbf{\Lambda}}^x/n)^{-1} = n^{1/2}\mathcal{S}_i'\hat{\mathbf{P}}^x.$$

Lemma 3(iii) and inequality (B.5) imply that $\|\mathbf{g}-\mathbf{h}\|$ is $O_P(\max(n^{-1}, \rho_T^{-1/2}))$ uniformly in θ and i. Inequality (B.5), Lemma 3(i) and Lemma 2(ii) imply that $\|\mathbf{h}-\mathbf{i}\|$ is $O_P(\max(n^{-1}, \rho_T^{-1/2}))$ uniformly in θ and i. Moreover, $\|\hat{\mathbf{P}}^x(\hat{\mathbf{\Lambda}}^x/n)^{-1}\| = O_P(1)$, uniformly in θ, by Lemma 2(ii) and the fact that $\|\hat{\mathbf{P}}^x\| = 1$. Thus, using Lemma 1(iv), it is seen that, uniformly in θ and i, $\|\mathbf{i}-\mathbf{j}\|$ is $O_P(\max(n^{-1/2}, \rho_T^{-1/2}))$. The result follows. □

Proposition 7 now follows from

$$\hat{\Sigma}^x = \left[\hat{\mathbf{P}}^x\left(\hat{\Lambda}^x\right)^{1/2}\right]\left[\left(\hat{\Lambda}^x\right)^{1/2}\tilde{\hat{\mathbf{P}}}^x\right] = \hat{\mathbf{P}}^x\hat{\Lambda}^x\tilde{\hat{\mathbf{P}}}^x$$

and

$$\Sigma^x = \left[\mathbf{P}^x\left(\Lambda^x\right)^{1/2}\hat{\mathbf{W}}_q\right]\left[\tilde{\hat{\mathbf{W}}}_q\left(\Lambda^x\right)^{1/2}\tilde{\mathbf{P}}^x\right] = \mathbf{P}^x\Lambda^x\tilde{\mathbf{P}}^x. \quad \square$$

Note that the eigenvectors $\mathbf{P}^x$ are defined up to post-multiplication by a complex diagonal matrix with unit modulus diagonal entries. In particular, using the eigenvectors $\Pi^x = \mathbf{P}^x\hat{\mathbf{W}}_q$, (B.4) would hold for $\Pi^x(\Lambda^x)^{1/2} - \hat{\mathbf{P}}^x(\hat{\Lambda}^x)^{1/2}$. For the sake of simplicity, we avoid introducing a new symbol and henceforth refer to the result of Lemma 4 as

$$\max_{|h|\le B_T}\|\mathcal{S}_1'(\mathbf{P}^x(\Lambda^x)^{1/2} - \hat{\mathbf{P}}^x(\hat{\Lambda}^x)^{1/2})\| = O_P(\max(n^{-1/2}, \rho_T^{-1/2})) \tag{B.6}$$

and the result of Lemma 3(iii) as

$$\|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^x - \mathbf{I}_q\| = O_P(\max(n^{-1}, \rho_T^{-1/2})).$$

In the same way, we drop $\hat{\mathbf{W}}_q$ in Lemmas 6–8, though not in the conclusion of Appendix D, nor in Appendix E.

Appendix C. Proof of Proposition 9

To start with, note that, as the extreme right-hand side in (3.7) contains the term

$$\frac{\pi B}{B_T}\sum_{|h|\le B_T}\left(|e^{\iota\ell\theta_{h-1}^*} - e^{\iota\ell\tilde{\theta}_{h-1}^*}| + |e^{\iota\ell\tilde{\theta}_{h-1}^*} - e^{\iota\ell\theta_{h-1}^*}|\right),$$

convergence in (3.8) is not uniform with respect to ℓ. However, estimation of the matrices $\mathbf{B}_k^x$ and $\mathbf{C}_{jk}^x$ only requires the covariances $\hat{\gamma}_{ij,\ell}^x$ with $\ell \le S$, where S is finite. Therefore, Proposition 8 implies that $\|\hat{\mathbf{B}}_k^x - \mathbf{B}_k^x\|$ and $\|\hat{\mathbf{C}}_{jk}^x - \mathbf{C}_{jk}^x\|$ are $O_P(\max(n^{-1/2}, \rho_T^{-1/2}))$. From (2.18), applying (B.1),

$$\|\hat{\mathbf{A}}^{[k]} - \mathbf{A}^{[k]}\| \le \|\hat{\mathbf{B}}_k^x\|\|(\hat{\mathbf{C}}_{kk}^x)^{-1} - \left(\mathbf{C}_{kk}^x\right)^{-1}\| + \|\hat{\mathbf{B}}_k^x - \mathbf{B}_k^x\|\|\left(\mathbf{C}_{kk}^x\right)^{-1}\|.$$

By Assumption 2, $\|\mathbf{B}_k^x\| \le W$ for some constant $W > 0$, so that $\|\hat{\mathbf{B}}_k^x\|$ is bounded in probability. By Assumptions 2 and 7, $\|\left(\mathbf{C}_{kk}^x\right)^{-1}\| \le W_1$ for some $W_1 > 0$. Observing that the entries of $\left(\mathbf{C}_{kk}^x\right)^{-1}$ are rational functions of the entries of $\mathbf{C}_{kk}^x$, and that $\det\left(\mathbf{C}_{kk}^x\right) > 0$ by Assumption 7, Proposition 8 implies that $\|(\hat{\mathbf{C}}_{kk}^x)^{-1} - \left(\mathbf{C}_{kk}^x\right)^{-1}\|$ is $O_P(\max(n^{-1/2}, \rho_T^{-1/2}))$.

The conclusion follows. □

Appendix D. Proof of Proposition 10

Consider the static model $\mathbf{Z}_{nt} = \mathcal{R}\mathbf{v}_t + \boldsymbol{\Phi}_{nt}$. If $\mathbf{Z}_{nt} = \mathbf{A}(L)\mathbf{x}_{nt}$ were observed, i.e. if the matrices $\mathbf{A}(L)$ were known, then Proposition 10, with an estimator of $\mathcal{R}$ based on the empirical covariance $\boldsymbol{\Gamma}^z$ of the $\mathbf{Z}_{nt}$, would be straightforward. However, we only have access to $\hat{\mathbf{Z}}_{nt} = \hat{\mathbf{A}}(L)\mathbf{x}_t$ and its empirical covariance matrix $\hat{\boldsymbol{\Gamma}}^z$, which makes the estimation of $\mathcal{R}$ significantly more difficult. The consistency properties of our estimator follow from the convergence result (D.4) in Lemma 11, which establishes the asymptotic behavior of the difference $\boldsymbol{\Gamma}^z - \hat{\boldsymbol{\Gamma}}^z$; Lemmas 5–10 are but a preparation for that key result. All lemmas in this Appendix hold, and are proved under Assumptions 1–10.

Lemma 5. *For $f = 1, \ldots, q$, as $T \to \infty$ and $n \to \infty$,*

(i) $|p^{\chi}_{if}| = O(n^{-1/2})$ *and* $|\hat{p}^{x}_{if}| = O_P(n^{-1/2})$*, uniformly in θ and i;*
(ii) *for any positive integer d, $n^{-1}\sum_{i=1}^n |p^{\chi}_{if}|^d$ and $n^{-1}\sum_{i=1}^n |\hat{p}^{x}_{if}|^d$ are $O(n^{-d/2})$ and $O_P(n^{-d/2})$, respectively, uniformly in θ.*

Proof. The first part of (i) follows from (B.5). As regards the second part, let us first prove that $\hat{\sigma}^x_{ii}$ is $O_P(1)$ uniformly in θ and i. We have

$$\max_h \hat{\sigma}^x_{ii}(\theta_h) \le \max_h \sigma^x_{ii}(\theta_h) + \max_h |\hat{\sigma}^x_{ii}(\theta_h) - \sigma^x_{ii}(\theta_h)|.$$

By Assumptions 2 and 4, the first term on the right-hand side is bounded uniformly in i. By the Markov inequality and (3.4),

$$P(\max_h |\hat{\sigma}^x_{ii}(\theta_h) - \sigma^x_{ii}(\theta_h)| \ge \eta) \le \eta^{-2} E\left(\max_{|h|\le B_T} \left|\hat{\sigma}^x_{ii}(\theta^*_h) - \sigma^x_{ii}(\theta^*_h)\right|^2\right) \le \eta^{-2} C(T^{-1}B_T \log B_T).$$

Thus, for any $\epsilon > 0$, we can set

$$\eta(\epsilon) \ge \left[\frac{\max_T C(T^{-1}B_T \log B_T)}{\epsilon}\right]^{1/2},$$

irrespective of θ_h and i. Because $\hat{\sigma}^{\chi}_{ii} \le \hat{\sigma}^x_{ii}$, we have that $\hat{\sigma}^{\chi}_{ii} = \sum_{f=1}^q \hat{\lambda}^x_f |\hat{p}^x_{if}|^2 = O_P(1)$ uniformly in θ and i. As all the terms in the sum are positive, $\hat{\lambda}^x_f |\hat{p}^x_{if}|^2 = (\hat{\lambda}^x_f/n) n |\hat{p}^x_{if}|^2$ is $O_P(1)$ as well, uniformly in θ and i. Lemma 2(i) and Assumption 3 imply that $\hat{\lambda}^x_f/n$ is $O_P(1)$ and bounded away from zero in probability uniformly in θ. The conclusion follows.

Statement (ii) is proved by induction. Consider $\mathbf{P}^{\chi}_f$. It follows from statement (i) that $n^{-1}\sum_{i=1}^n |p^{\chi}_{if}|$ is $O(n^{-1/2})$, uniformly in θ. Assume now that the result holds for $d-1$, with $d \ge 2$. Using the first part of (i), uniformity in i in particular, we have

$$\begin{aligned} n^{-1}\sum_{i=1}^n |p^{\chi}_{if}|^d &= n^{-1}\sum_{i=1}^n |p^{\chi}_{if}|^{d-1}|p^{\chi}_{if}| \\ &\le \left(\max_{i\le n}|p^{\chi}_{if}|\right) n^{-1}\sum_{i=1}^n |p^{\chi}_{if}|^{d-1} = O(n^{-1/2}\, n^{-(d-1)/2}) \\ &= O\left(n^{-d/2}\right). \end{aligned}$$

The same argument applies to $\hat{\mathbf{P}}^x_f$. $\square$

Lemma 6. *As $T \to \infty$ and $n \to \infty$,*

$$\max_{|h|\le B_T} \left\| \mathbf{P}^{\chi}\left(\boldsymbol{\Lambda}^{\chi}\right)^{1/2} - \hat{\mathbf{P}}^x\left(\hat{\boldsymbol{\Lambda}}^x\right)^{1/2} \right\| = O_P(n^{1/2}\max(n^{-1}, \rho_T^{-1/2})). \quad \text{(D.1)}$$

Proof. The left-hand side of (D.1) equals the left-hand side of (B.4) when $\mathcal{S}_i$ is replaced by $\mathbf{I}_n$. The proof goes along the same lines as that of Lemma 4. Firstly, $\|n^{1/2}\mathbf{P}^{\chi}\|$ is $O(n^{1/2})$. Both $\|\mathbf{g} - \mathbf{h}\|$ and $\|\mathbf{h} - \mathbf{i}\|$ are $O_P(n^{1/2}\max(n^{-1}, \rho_T^{-1/2}))$. As for $\|\mathbf{i} - \mathbf{j}\|$, the conclusion follows from Lemma 1(iii). $\square$

Lemma 7. *For $f = 1, \ldots, q$, as $T \to \infty$ and $n \to \infty$, $|p^{\chi}_{if} - \hat{p}^x_{if}| = O_P(n^{-1/2}\max(n^{-1/2}, \rho_T^{-1/2}))$, uniformly in θ and i.*

Proof. By (B.6), $p^{\chi}_{if}(\lambda^{\chi}_f)^{1/2} - \hat{p}^x_{if}(\hat{\lambda}^x_f)^{1/2} = O_P(\max(n^{-1/2}, \rho_T^{-1/2}))$, uniformly in θ and i. Now,

$$\begin{aligned} p^{\chi}_{if}(\lambda^{\chi}_f)^{1/2} - \hat{p}^x(\hat{\lambda}^x_f)^{1/2} &= p^{\chi}_{if}\left((\lambda^{\chi}_f)^{1/2} - (\hat{\lambda}^x_f)^{1/2}\right) \\ &\quad + (\hat{\lambda}^x_f)^{1/2}\left(p^{\chi}_{if} - \hat{p}^x_{if}\right). \end{aligned} \quad \text{(D.2)}$$

The former term on the right-hand side can be written as

$$n^{1/2} p^{\chi}_{if} \frac{(\lambda^{\chi}_f - \hat{\lambda}^x_f)/n}{\left((\lambda^{\chi}_f)^{1/2} + (\hat{\lambda}^x_f)^{1/2}\right)/n^{1/2}},$$

which is $O_P(\max(n^{-1}, \rho_T^{-1/2}))$, uniformly in θ and i, since the numerator is $O_P(\max(n^{-1}, \rho_T^{-1/2}))$, uniformly in θ, by Lemma 2(i); the denominator is bounded away from zero, uniformly in θ, by Assumption 3 and $n^{1/2}p^{\chi}_{if}$ is $O(1)$, uniformly in θ and i, by Lemma 5(i). It follows that the latter term in (D.2), $(\hat{\lambda}^x_f)^{1/2}\left(p^{\chi}_{if} - \hat{p}^x_{if}\right)$, is $O_P(\max(n^{-1/2}, \rho_T^{1/2}))$, uniformly in θ and i. By Lemma 2(ii), $n^{-1/2}(\hat{\lambda}^x_f)^{1/2}$ is bounded away from zero in probability, uniformly in θ. The result follows. $\square$

Lemma 8. *For any integer $d \in \mathbb{N}$, for $f = 1, \ldots, q$, as $T \to \infty$ and $n \to \infty$,*

$$n^{-1}\sum_{i=1}^n |p^{\chi}_{if} - \hat{p}^x_{if}|^d = O_P((n^{-1}\max(n^{-1}, \rho_T^{-1}))^{d/2}), \quad \text{(D.3)}$$

uniformly in θ.

Proof. Lemma 7 implies that $\left(\max_{i\le n}|p^{\chi}_{if} - \hat{p}^x_{if}|\right)$, and therefore $n^{-1}\sum_{i=1}^n |\hat{p}^x_{if} - p^{\chi}_{if}|$, are $O_P((n^{-1}\max(n^{-1}, \rho_T^{-1}))^{1/2})$, uniformly in θ. By induction, assume now that the result holds for $d-1$, $d \ge 2$. We have

$$\begin{aligned} n^{-1}\sum_{i=1}^n |\, p^{\chi}_{if} - \hat{p}^x_{if}|^d &= n^{-1}\sum_{i=1}^n |\, p^{\chi}_{if} - \hat{p}^x_{if}|^{d-1} \,|\, p^{\chi}_{if} - \hat{p}^x_{if}\,| \\ &\le \left(\max_{i\le n} |\, p^{\chi}_{if} - \hat{p}^x_{if}\,|\right) n^{-1}\sum_{i=1}^n |\, p^{\chi}_{if} - \hat{p}^x_{if}|^{d-1} \\ &= O_P((n^{-1}\max(n^{-1}, \rho_T^{-1}))^{1/2}) \\ &\quad \times O_P\left((n^{-1}\max(n^{-1}, \rho_T^{-1}))^{(d-1)/2}\right), \end{aligned}$$

uniformly in θ, as was to be shown. $\square$

Lemma 9. *For $n \to \infty$ and $T \to \infty$, uniformly in θ,*

(i) $n^{-2}\sum_{i=1}^n\sum_{j=1}^n |\hat{\sigma}^{\chi}_{ij}(\theta) - \sigma^{\chi}_{ij}(\theta)|^d = O_P((\max(n^{-1}, \rho_T^{-1}))^{d/2})$*;*
(ii) $n^{-1}\sum_{i=1}^n |\, \hat{\sigma}^{\chi}_{ij}(\theta) - \sigma^{\chi}_{ij}(\theta)|^d = O_P((\max(n^{-1}, \rho_T^{-1}))^{d/2})$ *for any $1 \le j \le n$;*
(iii) $n^{-1}\sum_{i=1}^n |\, \hat{\sigma}^{\chi}_{ii}(\theta) - \sigma^{\chi}_{ii}(\theta)|^d = O_P((\max(n^{-1}, \rho_T^{-1}))^{d/2})$.

Proof. We have

$$\hat{\sigma}_{ij}^{\chi} - \sigma_{ij}^{\chi} = (\hat{\lambda}_1^{\chi} - \lambda_1^{\chi})\hat{p}_{i1}^{\chi}\bar{\hat{p}}_{j1}^{\chi} + \cdots + (\hat{\lambda}_q^{\chi} - \lambda_q^{\chi})\hat{p}_{iq}^{\chi}\bar{\hat{p}}_{jq}^{\chi} + \lambda_1^{\chi}\hat{p}_{i1}^{\chi}(\bar{\hat{p}}_{j1}^{\chi} - \bar{p}_{j1}^{\chi}) + \lambda_1^{\chi}\bar{p}_{j1}^{\chi}(\hat{p}_{i1}^{\chi} - p_{i1}^{\chi}) + \cdots + \lambda_q^{\chi}\hat{p}_{iq}^{\chi}(\bar{\hat{p}}_{jq}^{\chi} - \bar{p}_{jq}^{\chi}) + \lambda_q^{\chi}\bar{p}_{jq}^{\chi}(\hat{p}_{iq}^{\chi} - p_{iq}^{\chi}).$$

Using the triangular and C_r inequalities, by Lemmas 2, 5 and 8,

$$\begin{aligned} &n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} |\,\hat{\sigma}_{ij}^{\chi} - \sigma_{ij}^{\chi}\,|^d \\ &\le (3q)^{d-1}\Bigg(|\,\lambda_1^{\chi} - \hat{\lambda}_1^{\chi}|^d\Big(n^{-1}\sum_{i=1}^{n}|\,\hat{p}_{i1}^{\chi}|^d\Big)^2 + \cdots + |\,\lambda_q^{\chi} - \hat{\lambda}_q^{\chi}|^d\Big(n^{-1}\sum_{i=1}^{n}|\,\hat{p}_{iq}^{\chi}|^d\Big)^2\Bigg) \\ &+ (3q)^{d-1}(\lambda_1^{\chi})^d\Bigg(n^{-2}\sum_{i=1}^{n}|\,\hat{p}_{i1}^{\chi}|^d\sum_{j=1}^{n}|\,p_{j1}^{\chi} - \hat{p}_{j1}^{\chi}|^d + n^{-2}\sum_{j=1}^{n}|\,p_{j1}^{\chi}|^d\sum_{i=1}^{n}|\,p_{i1}^{\chi} - \hat{p}_{i1}^{\chi}|^d\Bigg) \\ &+ \cdots \\ &+ (3q)^{d-1}(\lambda_q^{\chi})^d\Bigg(n^{-2}\sum_{i=1}^{n}|\,\hat{p}_{iq}^{\chi}|^d\sum_{j=1}^{n}|\,p_{jq}^{\chi} - \hat{p}_{jq}^{\chi}|^d + n^{-2}\sum_{j=1}^{n}|\,p_{jq}^{\chi}|^d\sum_{i=1}^{n}|\,p_{iq}^{\chi} - \hat{p}_{iq}^{\chi}|^d\Bigg) \\ &= O_P((\max(n^{-1}, \rho_T^{-1/2}))^d) + O_P((\max(n^{-1}, \rho_T^{-1}))^{d/2}) \\ &= O_P((\max(n^{-1}, \rho_T^{-1}))^{d/2}). \end{aligned}$$

Statement (i) follows. For statement (ii),

$$\begin{aligned} &n^{-1}\sum_{i=1}^{n} |\,\hat{\sigma}_{ij}^{\chi} - \sigma_{ij}^{\chi}\,|^d \\ &\le (3q)^{d-1}\Bigg(|\,\lambda_1^{\chi} - \hat{\lambda}_1^{\chi}|^d|\hat{p}_{j1}^{\chi}|^d\, n^{-1}\sum_{i=1}^{n}|\hat{p}_{i1}^{\chi}|^d + \cdots + |\,\lambda_q^{\chi} - \hat{\lambda}_q^{\chi}|^d|\hat{p}_{jq}^{\chi}|^d\, n^{-1}\sum_{i=1}^{n}|\,\hat{p}_{iq}^{\chi}|^d\Bigg) \\ &+ (3q)^{d-1}(\lambda_1^{\chi})^d\Bigg(\left|p_{j1}^{\chi} - \hat{p}_{j1}^{\chi}\right|^d n^{-1}\sum_{i=1}^{n}\left|\hat{p}_{i1}^{\chi}\right|^d + \left|p_{j1}^{\chi}\right|^d n^{-1}\sum_{i=1}^{n}|\,p_{i1}^{\chi} - \hat{p}_{i1}^{\chi}|^d\Bigg) \\ &+ \cdots \\ &+ (3q)^{d-1}(\lambda_q^{\chi})^d\Bigg(\left|p_{jq}^{\chi} - \hat{p}_{jq}^{\chi}\right|^d n^{-1}\sum_{i=1}^{n}\left|\hat{p}_{iq}^{\chi}\right|^d + \left|p_{jq}^{\chi}\right|^d n^{-1}\sum_{i=1}^{n}|\,p_{iq}^{\chi} - \hat{p}_{iq}^{\chi}|^d\Bigg) \\ &= O_P((\max(n^{-1}, \rho_T^{-1/2}))^d) + O_P((\max(n^{-1}, \rho_T^{-1}))^{d/2}) \\ &= O_P((\max(n^{-1}, \rho_T^{-1}))^{d/2}). \end{aligned}$$

Statement (iii) follows along the same lines, by setting $j = i$. □

Lemma 10. *For $n \to \infty$ and $T \to \infty$, $n^{-2}\sum_{\ell=0}^{S}\sum_{i=1}^{n}\sum_{j=1}^{n} |\,\hat{\gamma}_{ij,\ell}^{\chi} - \gamma_{ij,\ell}^{\chi}|^d$ and, for any given j in $\{1, \ldots, n\}$, $n^{-1}\sum_{\ell=0}^{S}\sum_{i=1}^{n} |\,\hat{\gamma}_{ij,\ell}^{\chi} - \gamma_{ij,\ell}^{\chi}|^d$, are $O_P\left((\max(n^{-1}, \rho_T^{-1}))^{d/2}\right)$.*

Proof. We have $|\hat{\gamma}_{ij,\ell}^{\chi} - \gamma_{ij,\ell}^{\chi}| \le \mathcal{U}_{ij} + \mathcal{V}_{\ell} + \mathcal{W}_{ij}$, where $\mathcal{U}_{ij}$, $\mathcal{V}_{\ell}$ and $\mathcal{W}_{ij}$ are the terms in the extreme right-hand side of (3.7). Using the C_r inequality, we get

$$n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}|\hat{\gamma}_{ij,0}^{\chi} - \gamma_{ij,0}^{\chi}|^d \le n^{-2}3^{d-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathcal{U}_{ij}^d + n^{-2}3^{d-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathcal{V}_{\ell}^d + n^{-2}3^{d-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathcal{W}_{ij}^d.$$

The first term on the right-hand side is bounded in view of Lemma 9. Since ℓ takes only a finite number of values, the second term is $O(B_T^{-d})$ (see the proof of Proposition 9). Because the functions σ_{ij}^{χ} are of bounded variation uniformly in i and j, see Proposition 2, the third term is $O(B_T^{-d})$. The same argument used to obtain Proposition 8 applies. The second statement is proved in the same way. □

We are now able to state and prove the main lemma of this section. We keep assuming that $n = m(q + 1)$, so that the dataset increases by blocks of size $q + 1$.

Lemma 11. *Denoting by $\hat{\mathbf{Z}}$ the $T \times n$ matrix with $\hat{Z}_{it}$ in entry $(t,\ i)$, let $\hat{\Gamma}^z = \hat{\mathbf{Z}}'\hat{\mathbf{Z}}/T$. Then, as $n \to \infty$ and $T \to \infty$,*

$$n^{-1}\|\hat{\Gamma}^z - \Gamma^z\| = O_P(\zeta_{nT}) \quad \text{and} \quad n^{-1/2}\|\mathcal{S}_i'(\hat{\Gamma}^z - \Gamma^z)\| = O_P(\zeta_{nT}), \tag{D.4}$$

where $\mathbf{\Gamma}^z$ is the population covariance matrix of $\mathbf{Z}_{nt}$.

Proof. Denote by $\check{\mathbf{\Gamma}}^z = \mathbf{Z}'\mathbf{Z}/T$ the empirical covariance matrix we would compute from the $\mathbf{Z}_{nt}$s if the matrices $\mathbf{A}(L)$ were known. We have

$$\|\hat{\Gamma}^z - \Gamma^z\| \le \|\hat{\Gamma}^z - \check{\Gamma}^z\| + \|\check{\Gamma}^z - \Gamma^z\|, \tag{D.5}$$

so that the lemma can be proved by showing that (D.4) holds with $\|\hat{\Gamma}^z - \Gamma^z\|$ replaced by any of the two terms on the right-hand side of (D.5).

First consider $\|\check{\Gamma}^z - \Gamma^z\|$. Since $\mathbf{A}(L) = \mathbf{I}_n - \mathbf{A}_1 L - \cdots - \mathbf{A}_S L^S$, where

$$\mathbf{A}_s = \begin{pmatrix} \mathbf{A}_s^1 & 0 & \cdots & 0 \\ 0 & \mathbf{A}_s^2 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \mathbf{A}_s^m \end{pmatrix}, \qquad s = 1, \ldots, S$$

and $\mathbf{A}_0 = \mathbf{I}_n$, we obtain

$$\begin{aligned} \|\check{\Gamma}^z - \Gamma^z\|^2 &\le \sum_{s=0}^{S}\sum_{r=0}^{S}\|\mathbf{A}_s\hat{\Gamma}_{s-r}^x\mathbf{A}_r' - \mathbf{A}_s\Gamma_{s-r}^x\mathbf{A}_r'\|^2 \\ &= \sum_{s=0}^{S}\sum_{r=0}^{S}\|\mathbf{A}_s\left(\hat{\Gamma}_{s-r}^x - \Gamma_{s-r}^x\right)\mathbf{A}_r'\|^2, \end{aligned} \tag{D.6}$$

which is a sum of $(S + 1)^2$ terms, where we set $\hat{\Gamma}_{s-r}^x = T^{-1}\sum_{t=1}^{T}\mathbf{x}_{t-r}\mathbf{x}_{t-s}'$. Inspection of the right-hand side of (D.6) shows that (D.4) holds, with $\|\hat{\Gamma}^z - \Gamma^z\|$ replaced with $\|\check{\Gamma}^z - \Gamma^z\|$, under Assumptions 2 and 7, and in view of Propositions 2 and 6.

Turning to $\|\hat{\Gamma}^z - \check{\Gamma}^z\|$, since

$$\|\hat{\Gamma}^z - \check{\Gamma}^z\|^2 \le \sum_{s=0}^{S}\sum_{r=0}^{S}\|\hat{\mathbf{A}}_s\hat{\Gamma}_{s-r}^x\hat{\mathbf{A}}_r' - \mathbf{A}_s\hat{\Gamma}_{s-r}^x\mathbf{A}_r'\|^2,$$

it is sufficient to prove that (D.4) still holds with $\|\hat{\Gamma}^z - \Gamma^z\|$ replaced with any of the $\|\hat{\mathbf{A}}_s\hat{\Gamma}_{s-r}^x\hat{\mathbf{A}}_r' - \mathbf{A}_s\hat{\Gamma}_{s-r}^x\mathbf{A}_r'\|$s. Denoting by $\mathbf{a}_{s\alpha}^j$, $1 \le \alpha \le$

$q+1$, the αth column of $\mathbf{A}_s^{j\prime}$, we have

$$\begin{aligned}
&\|\hat{\mathbf{A}}^s\hat{\varGamma}^x_{s-r}\hat{\mathbf{A}}^{r\prime} - \mathbf{A}^s\hat{\varGamma}^x_{s-r}\mathbf{A}^{r\prime}\|^2 \\
&\quad\le \sum_{j=1}^m\sum_{k=1}^m\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}\left(\hat{\mathbf{a}}^{j\prime}_{s\alpha}\hat{\varGamma}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta} - \mathbf{a}^{j\prime}_{s\alpha}\hat{\varGamma}^x_{jk,s-r}\mathbf{a}^k_{r\beta}\right)^2 \\
&\quad\le 2\sum_{j=1}^m\sum_{k=1}^m\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}\left((\hat{\mathbf{a}}^{j}_{s\alpha} - \mathbf{a}^{j\prime}_{s\alpha})\hat{\varGamma}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta}\right)^2 \\
&\qquad + 2\sum_{j=1}^m\sum_{k=1}^m\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}\left(\mathbf{a}^{j\prime}_{s\alpha}\hat{\mathbf{\Gamma}}^x_{jk,s-r}(\hat{\mathbf{a}}^k_{r\beta} - \mathbf{a}^k_{r\beta})\right)^2,
\end{aligned} \tag{D.7}$$

where $\hat{\varGamma}^x_{jk,s-r}$ is the (j,k)-block of $\hat{\varGamma}^x_{s-r}$, and the second inequality follows from applying the C_r inequality to each term of the form

$$\begin{aligned}
&(\hat{\mathbf{a}}^{j\prime}_{s\alpha}\hat{\varGamma}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta} - \mathbf{a}^{j\prime}_{s\alpha}\hat{\varGamma}^x_{jk,s-r}\mathbf{a}^k_{r\beta})^2 \\
&\quad = ((\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'\hat{\varGamma}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta} - \mathbf{a}^{j\prime}_{s\alpha}\hat{\varGamma}^x_{jk,s-r}(\hat{\mathbf{a}}^k_{r\beta} - \mathbf{a}^k_{r\beta}))^2.
\end{aligned}$$

The two terms on the right-hand side of (D.7) can be dealt with in the same way. Let us focus on the first of them. Using twice the Cauchy–Schwarz inequality, then subsequently the C_r and Jensen inequalities, we obtain

$$\begin{aligned}
&\sum_{j=1}^m\sum_{k=1}^m\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}((\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'\hat{\varGamma}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta})^2 \\
&\quad\le \sum_{j=1}^m\sum_{k=1}^m\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}((\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\varGamma}^{x\prime}_{jk,s-r}\hat{\varGamma}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta} \\
&\quad= \sum_{k=1}^m\sum_{\beta=1}^{q+1}\sum_{j=1}^m\sum_{\alpha=1}^{q+1}(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\varGamma}^{x\prime}_{jk,s-r}\hat{\varGamma}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta} \\
&\quad\le \sum_{k=1}^m\sum_{\beta=1}^{q+1}\Big[\sum_{j=1}^m\Big[\sum_{\alpha=1}^{q+1}(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})\Big]^2\Big]^{1/2} \\
&\qquad\times\Big[\sum_{j=1}^m(\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\varGamma}^{x\prime}_{jk,s-r}\hat{\varGamma}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta})^2\Big]^{1/2} \\
&\quad= m\Big[\sum_{j=1}^m\Big[\sum_{\alpha=1}^{q+1}(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^{j\prime}_{s\alpha})(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})\Big]^2\Big]^{1/2} \\
&\qquad\times\frac{1}{m}\sum_{k=1}^m\sum_{\beta=1}^{q+1}\Big[\sum_{j=1}^m(\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\varGamma}^{x\prime}_{jk,s-r}\hat{\varGamma}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta})^2\Big]^{1/2} \\
&\quad\le \mathcal{A}\mathcal{B},\quad\text{say,}
\end{aligned}$$

where

$$\mathcal{A} = m(q+1)^{1/2}\Big[\sum_{j=1}^m\sum_{\alpha=1}^{q+1}\left((\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})\right)^2\Big]^{1/2}$$

and

$$\begin{aligned}
\mathcal{B} &= \frac{1}{m}\sum_{k=1}^m\sum_{\beta=1}^{q+1}\Big[\sum_{j=1}^m\left(\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\varGamma}^{x\prime}_{jk,s-r}\hat{\varGamma}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta}\right)^2\Big]^{1/2} \\
&\le \Big[(q+1)/m\sum_{k=1}^m\sum_{\beta=1}^{q+1}\sum_{j=1}^m\left(\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\varGamma}^{x\prime}_{jk,s-r}\hat{\varGamma}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta}\right)^2\Big]^{1/2} = \mathcal{C},\ \text{say.}
\end{aligned}$$

First consider $\mathcal{A}$. Letting $\mathbf{a}^{j\prime}_{s\alpha} = (a^j_{s\alpha,1}\ a^j_{s\alpha,2}\ \cdots\ a^j_{s\alpha,q+1})$, note that $a^j_{s\alpha,\delta} = \mathbf{e}'_\alpha\mathbf{A}^{[j]}\mathbf{g}_{s\delta}$, where $\mathbf{e}_\alpha$ and $\mathbf{g}_{s\delta}$ stand for the αth and $(s-1)(q+1)+\delta$th unit vectors in the $(q+1)$- and $(q+1)S$-dimensional canonical bases, respectively. Writing, for the sake of simplicity, $\mathbf{B}_j$ and $\mathbf{C}_j$ instead of $\mathbf{B}^\chi_j$ and $\mathbf{C}^\chi_{jj}$, as defined in (2.16) and (2.17), we obtain, from (B.1), and applying subsequently the C_r, the triangular, the C_r again and then twice the Cauchy–Schwarz inequalities,

$$\begin{aligned}
&\Big[\sum_{j=1}^m\sum_{\alpha=1}^{q+1}\left((\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})\right)^2\Big]^{1/2} \\
&\quad\le (q+1)^{1/2}\Big(\sum_{j=1}^m\sum_{\alpha=1}^{q+1}\sum_{\delta=1}^{q+1}(\hat{a}^j_{s\alpha,\delta} - a^j_{s\alpha,\delta})^4\Big)^{1/2} \\
&\quad= (q+1)^{1/2}\Big(\sum_{j=1}^m\sum_{\alpha=1}^{q+1}\sum_{\delta=1}^{q+1}\Big[\mathbf{e}_\alpha\left((\hat{\mathbf{B}}_j - \mathbf{B}_j)\hat{\mathbf{C}}_j^{-1} + \mathbf{B}_j\hat{\mathbf{C}}_j^{-1}(\hat{\mathbf{C}}_j - \mathbf{C}_j)\mathbf{C}_j^{-1}\right)\mathbf{g}_{s\delta}\Big]^4\Big)^{1/2} \\
&\quad\le 2^{3/2}(q+1)^{3/2}\Big(\sum_{j=1}^m\|(\hat{\mathbf{B}}_j - \mathbf{B}_j)\hat{\mathbf{C}}_j^{-1}\|^4 + \|\mathbf{B}_j\hat{\mathbf{C}}_j^{-1}(\hat{\mathbf{C}}_j - \mathbf{C}_j)\mathbf{C}_j^{-1}\|^4\Big)^{1/2} \\
&\quad\le 2^{3/2}(q+1)^{3/2}\Big(\Big[\sum_{j=1}^m\|\hat{\mathbf{B}}_j - \mathbf{B}_j\|^8\Big]^{1/2}\Big[\sum_{j=1}^m\|\hat{\mathbf{C}}_j^{-1}\|^8\Big]^{1/2} \\
&\qquad+\Big[\sum_{j=1}^m\|\hat{\mathbf{C}}_j - \mathbf{C}_j\|^8\Big]^{1/2}\Big[\sum_{j=1}^m\|\hat{\mathbf{B}}_j\hat{\mathbf{C}}_j^{-1}\|^8\|\mathbf{C}_j^{-1}\|^8\Big]^{1/2}\Big)^{1/2} \\
&\quad\le 2^{3/2}(q+1)^{3/2}\Big(\Big[\sum_{j=1}^m\|\hat{\mathbf{B}}_j - \mathbf{B}_j\|^8\Big]^{1/2}\Big[\sum_{j=1}^m\|\hat{\mathbf{C}}_j^{-1}\|^8\Big]^{1/2} \\
&\qquad+\Big[\sum_{j=1}^m\|\hat{\mathbf{C}}_j - \mathbf{C}_j\|^8\Big]^{1/2}\Big[\sum_{j=1}^m\|\hat{\mathbf{B}}_j\|^{16}\Big]^{\frac14}\Big[\sum_{j=1}^m\|\hat{\mathbf{C}}_j^{-1}\|^{16}\|\mathbf{C}_j^{-1}\|^{16}\Big]^{\frac14}\Big)^{1/2}.
\end{aligned}$$

Denoting by $b^j_{i\delta}$ the entries of $\mathbf{B}_j$, $i=1,\dots,q+1$, $\delta=1,\dots,S(q+1)$, the C_r inequality and Lemma 10 entail

$$\begin{aligned}
\sum_{j=1}^m\|\hat{\mathbf{B}}_j - \mathbf{B}_j\|^8 &\le \sum_{j=1}^m\Big(\sum_{i=1}^{q+1}\sum_{\delta=1}^{S(q+1)}(\hat{b}^j_{i\delta} - b^j_{i\delta})^2\Big)^4 \\
&\le (q+1)^6S^3\sum_{j=1}^m\sum_{i=1}^{q+1}\sum_{\delta=1}^{S(q+1)}(\hat{b}^j_{i\delta} - b^j_{i\delta})^8 \\
&= O_P(m(\max(n^{-1},\rho_T^{-1}))^4).
\end{aligned}$$

In a similar way, one can prove that $\sum_{j=1}^m\|\hat{\mathbf{C}}_j - \mathbf{C}_j\|^8$ is $O_P(m(\max(n^{-1},\rho_T^{-1}))^4)$. Moreover, Assumptions 2 and 7 together with Lemma 10 imply that $\sum_{j=1}^m\|\hat{\mathbf{B}}_j\|^{16}$ and $\sum_{j=1}^m\|\mathbf{C}_j^{-1}\|^{16}$, as well as $\sum_{j=1}^m\|\hat{\mathbf{C}}_j^{-1}\|^8$ and $\sum_{j=1}^m\|\hat{\mathbf{C}}_j^{-1}\|^{16}$, are $O_P(m)$.

Collecting terms yields

$$\begin{aligned}
\mathcal{A} &= m(q+1)^{1/2}\Big[\sum_{j=1}^m\sum_{\alpha=1}^{q+1}((\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha}))^2\Big]^{1/2} \\
&\le 2^{3/2}(q+1)^2m\Big(\sum_{i=1}^m\|\hat{\mathbf{A}}^i_s - \mathbf{A}^i_s\|^4\Big)^{1/2} \\
&= O_P\left(m^{3/2}\max(n^{-1},\rho_T^{-1})\right).
\end{aligned} \tag{D.8}$$

Turning to $\mathcal{C}$, we obtain, by means of similar methods,

$$\begin{aligned}
\mathcal{C} &\le ((q+1)/m)^{1/2}\Bigg\{\Big[\sum_{k=1}^m\Big(\sum_{\beta=1}^{q+1}(\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\mathbf{a}}^k_{r\beta})^2\Big)^2\Big]^{1/2} \\
&\qquad\times\Big[\sum_{j=1}^m\Big(\sum_{k=1}^m\Big(\mathrm{trace}[\hat{\varGamma}^{x\prime}_{jk,s-r}\hat{\varGamma}^x_{jk,s-r}]\Big)^4\Big)^{1/2}\Big]\Bigg\}^{1/2} \\
&\le ((q+1)/m)^{1/2}\Bigg\{\Big[(q+1)\sum_{k=1}^m\sum_{\beta=1}^{q+1}(\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\mathbf{a}}^k_{r\beta})^4\Big]^{1/2} \\
&\qquad\times\Big[\sum_{j=1}^m\Big(\sum_{k=1}^m\Big(\mathrm{trace}[\hat{\varGamma}^{x\prime}_{jk,s-r}\hat{\varGamma}^x_{jk,s-r}]\Big)^4\Big)^{1/2}\Big]\Bigg\}^{1/2}
\end{aligned}$$

$$\leq (q+1)^{1/2}\left[(q+1)^4\sum_{k=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}(\hat{a}^k_{r,\alpha\beta})^8\right]^{1/4}$$
$$\times\left[m^{-1}\sum_{j=1}^{m}\sum_{k=1}^{m}\left(\mathrm{trace}[\hat{\Gamma}^{x\prime}_{jk,s-r}\hat{\Gamma}^{x}_{jk,s-r}]\right)^4\right]^{1/4}$$
$$\leq (q+1)^{3/2}\left[\sum_{k=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}(\hat{a}^k_{r,\alpha\beta})^8\right]^{1/4}$$
$$\times\left[((q+1)^6/m)\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}(\hat{\gamma}^x_{jk,\alpha\beta}(s-r))^8\right]^{\frac{1}{4}}$$
$$= O_P(m^{1/2}),$$

where $\hat{\gamma}^x_{jk,\alpha\beta}(s-r)$ stands for the (α,β) entry of $\hat{\Gamma}^x_{jk,s-r}$. Collecting terms again, we get

$$m^{-1}\|\hat{\mathbf{A}}_s\hat{\mathbf{\Gamma}}^x_{s-r}\hat{\mathbf{A}}'_r - \mathbf{A}_s\hat{\mathbf{\Gamma}}^x_{s-r}\mathbf{A}'_r\| \leq \left(\frac{1}{m^2}\mathcal{AC}\right)^{1/2}$$
$$= O_P(\zeta_{nT}), \quad r,s = 0,\ldots,S.$$

Now consider the second statement in (D.4). Again, it is sufficient to prove that it holds with $\|\hat{\Gamma}^z - \Gamma^z\|$ replaced with any of the $\|\hat{\mathbf{A}}_s\hat{\Gamma}^x_{s-r}\hat{\mathbf{A}}'_r - \mathbf{A}_s\hat{\Gamma}^x_{s-r}\mathbf{A}'_r\|$s. The two terms on the right-hand side of (D.7) must be dealt with separately. In the first of those two terms, dropping one of the summations for $k = 1,\ldots,m$ and setting $k = i$,

$$\sum_{j=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}\left((\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'\hat{\Gamma}^x_{ji,s-r}\hat{\mathbf{a}}^i_{r\beta}\right)^2 = O_P\left(m(\max(n^{-1},\rho_T^{-1}))\right).$$

Indeed, the left-hand side is bounded by a product $\mathcal{DE}$, say, where

$$\mathcal{D} = m^{1/2}(q+1)^{1/2}\left[\sum_{j=1}^{m}\sum_{\alpha=1}^{q+1}\left((\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})\right)^2\right]^{1/2}$$

and

$$\mathcal{E} = \sum_{\beta=1}^{q+1}\left(\frac{1}{m}\sum_{j=1}^{m}\left(\hat{\mathbf{a}}^{i\prime}_{r\beta}\hat{\Gamma}^{x\prime}_{ji,s-r}\hat{\Gamma}^x_{jk,s-r}\hat{\mathbf{a}}^i_{r\beta}\right)^2\right)^{1/2}$$

can be bounded along the same lines as $\mathcal{A}$ and $\mathcal{B}$ in the proof of the first statement.

As for the second term of (D.7), using arguments similar to those used in the first part of the proof, we obtain

$$\sum_{j=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}\left((\hat{\mathbf{a}}^i_{s\alpha} - \mathbf{a}^i_{s\alpha})'\hat{\Gamma}^{x\prime}_{jk,s-r}\mathbf{a}^j_{r\beta}\right)^2$$
$$\leq m\left[\left[\sum_{\alpha=1}^{q+1}(\hat{\mathbf{a}}^i_{s\alpha} - \mathbf{a}^i_{s\alpha})'(\hat{\mathbf{a}}^i_{s\alpha} - \mathbf{a}^i_{s\alpha})\right]^2\right]^{1/2}$$
$$\times\left[\frac{1}{m}\sum_{j=1}^{m}\sum_{\beta=1}^{q+1}(\mathbf{a}^{j\prime}_{r\beta}\hat{\Gamma}^x_{ji,s-r}\hat{\Gamma}^{x\prime}_{ji,s-r}\mathbf{a}^j_{r\beta})\right]$$
$$= \mathcal{FG}, \text{ say.}$$

It easily follows from Proposition 9 that $\mathcal{F} = O_P(m\,\zeta^2_{nT})$, while $\mathcal{G} = O_P(1)$ can be obtained from the arguments used to bound $\mathcal{C}$ in the proof of the first statement. Collecting terms, we obtain, as desired,

$$m^{-1/2}\|\mathcal{S}'_i(\hat{\mathbf{A}}_s\hat{\mathbf{\Gamma}}^x_{s-r}\hat{\mathbf{A}}'_r - \mathbf{A}_s\hat{\mathbf{\Gamma}}^x_{s-r}\mathbf{A}'_r)\| = O_p(\zeta_{nT}), \quad r,s = 0,\ldots,S. \quad \square$$

Starting with Lemma 11, which plays here the same role as Proposition 6 does for the proof of Proposition 7, we can easily prove statements that replicate in this context Lemmas 1–4, using the same arguments as in Appendix B, with x, χ and ξ replaced by Z, Ψ and Φ, respectively. More precisely,

(I) In the results corresponding to Lemma 1 we obtain the rate ζ_{nT} for (i), (ii), (iii) and (iv). Note that no reduction from $1/n$ to $1/\sqrt{n}$ occurs between (iii) and (iv), as in Lemma 1. For, (iii) has $O_P(\zeta_{nT}) + O(1/n) = O_P(\zeta_{nT})$, while (iv) has $O_P(\zeta_{nT}) + O(1/\sqrt{n})$, which is $O_P(\zeta_{nT})$.

(II) The same rate ζ_{nT} is obtained for the results of Lemma 2.

(III) The same holds for Lemma 3. The orthogonal matrix in point (iii), call it again $\hat{\mathbf{W}}_q$, has either 1 or -1 on the diagonal; thus $\tilde{\hat{\mathbf{W}}}_q = \hat{\mathbf{W}}_q$.

(IV) Lastly, Lemma 4 becomes

$$\left\|\mathcal{S}'_i\left(\hat{\mathbf{P}}^z\left(\hat{\Lambda}^z\right)^{1/2} - \mathbf{P}^\psi\left(\Lambda^\psi\right)^{1/2}\hat{\mathbf{W}}_q\right)\right\| = \left\|\hat{\mathcal{R}}_i - \mathcal{R}_i\hat{\mathbf{W}}_q\right\|$$
$$= O_P(\zeta_{nT}). \qquad (D.9)$$

Going over the proof of Lemma 4, we see that $\|\mathbf{i}-\mathbf{j}\|$ has the worst rate, whereas here $\|\mathbf{g}-\mathbf{h}\|$, $\|\mathbf{h}-\mathbf{i}\|$ and $\|\mathbf{i}-\mathbf{j}\|$ all have rate $O_P(\zeta_{nT})$. This completes the proof of Proposition 10. $\square$

Finally, in the same way as the proof of Lemma 4 can be replicated to obtain (D.9), the proof of Lemma 6 can be replicated to obtain

$$\|\hat{\mathbf{P}}^z(\hat{\Lambda}^z)^{1/2} - \mathbf{P}^\psi(\mathbf{\Lambda}^\psi)^{1/2}\hat{\mathbf{W}}_q\| = O_P\left(n^{1/2}\zeta_{nT}\right). \qquad (D.10)$$

Appendix E. Proof of Proposition 11

We have

$$\hat{\mathbf{v}}_t = \left((\hat{\mathbf{\Lambda}}^z)^{1/2}\hat{\mathbf{P}}^{z\prime}\hat{\mathbf{P}}^z(\hat{\mathbf{\Lambda}}^z)^{1/2}\right)^{-1}(\hat{\mathbf{\Lambda}}^z)^{1/2}\hat{\mathbf{P}}^{z\prime}\hat{\mathbf{Z}}_t = (\hat{\mathbf{\Lambda}}^z)^{-1/2}\hat{\mathbf{P}}^{z\prime}\hat{\mathbf{Z}}_t$$
$$= (\hat{\mathbf{\Lambda}}^z)^{-1/2}\hat{\mathbf{P}}^{z\prime}\left(\hat{\mathbf{A}}(L) - \mathbf{A}(L)\right)\mathbf{x}_t$$
$$+ \left((\hat{\mathbf{\Lambda}}^z)^{-1/2}\hat{\mathbf{P}}^{z\prime} - \hat{\mathbf{W}}_q(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\right)\mathbf{A}(L)\mathbf{x}_t$$
$$+ \hat{\mathbf{W}}_q(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\mathbf{A}(L)\xi_t + \hat{\mathbf{W}}_q(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\mathbf{P}^\psi(\mathbf{\Lambda}^\psi)^{1/2}\mathbf{v}_t. \qquad (E.11)$$

Considering the first term on the right-hand side of (E.11),

$$\|(\hat{\mathbf{\Lambda}}^z)^{-1/2}\hat{\mathbf{P}}^{z\prime}\left(\hat{\mathbf{A}}(L) - \mathbf{A}(L)\right)\mathbf{x}_t\|$$
$$= \|(\hat{\mathbf{\Lambda}}^z/n)^{-1/2}\hat{\mathbf{P}}^{z\prime}n^{-1/2}\left(\hat{\mathbf{A}}(L) - \mathbf{A}(L)\right)\mathbf{x}_t\|$$
$$\leq \|(\hat{\mathbf{\Lambda}}^z/n)^{-1/2}\|\|\hat{\mathbf{P}}^{z\prime}\|\|n^{-1/2}\left(\hat{\mathbf{A}}(L) - \mathbf{A}(L)\right)\mathbf{x}_t\|.$$

Since $\|(\hat{\mathbf{\Lambda}}^z/n)^{-1/2}\| = O_P(1)$ and $\|\hat{\mathbf{P}}^z\| = 1$, by (D.8), we get

$$\|n^{-1/2}\left(\hat{\mathbf{A}}(L) - \mathbf{A}(L)\right)\mathbf{x}_t\|$$
$$\leq n^{-1/2}\sum_{r=0}^{p}\left(\sum_{i=1}^{m}\mathbf{x}^{i\prime}_{t-r}(\hat{\mathbf{A}}^i_r - \mathbf{A}^i_r)'(\hat{\mathbf{A}}^i_r - \mathbf{A}^i_r)\mathbf{x}^i_{t-r}\right)^{1/2}$$
$$\leq \sum_{r=0}^{p}\left(n^{-1}\sum_{i=1}^{m}(\mathbf{x}^{i\prime}_{t-r}\mathbf{x}^i_{t-r})^2\right)^{1/4}$$
$$\times\left(n^{-1}\sum_{i=1}^{m}\left(\sum_{j=1}^{q+1}\sum_{h=1}^{q+1}(\hat{a}^i_{r,jh} - a^i_{r,jh})^2\right)^2\right)^{1/4}$$
$$\leq \sum_{r=0}^{p}\left(n^{-1}\sum_{i=1}^{m}(\mathbf{x}^{i\prime}_{t-r}\mathbf{x}^i_{t-r})^2\right)^{1/4}\left((q+1)^3n^{-1}\sum_{i=1}^{m}\|\hat{\mathbf{A}}^i_r - \mathbf{A}^i_r\|^4\right)^{1/4}$$
$$= O_P(\zeta_{nT})$$

where $\mathbf{x}_t = (\mathbf{x}^{1\prime}_t \ldots \mathbf{x}^{i\prime}_t \ldots \mathbf{x}^{m\prime}_t)'$ stands for sub-vectors $\mathbf{x}^i_t$ of size $(q+1)\times 1$.

Next, considering the second term on the right-hand side of (E.11),

$$\left\|\left((\hat{\mathbf{\Lambda}}^z)^{-1/2}\hat{\mathbf{P}}^{z\prime} - \hat{\mathbf{W}}_q(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\right)\mathbf{A}(L)\mathbf{x}_t\right\|$$
$$= \left\|(\hat{\mathbf{\Lambda}}^z/n)^{-1}\left((\hat{\mathbf{\Lambda}}^z)^{1/2}\hat{\mathbf{P}}^{z\prime} - \hat{\mathbf{W}}_q\hat{\mathbf{\Lambda}}^z(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\right)\mathbf{A}(L)\mathbf{x}_t/n\right\|$$

$$= \left\|(\hat{\boldsymbol{\Lambda}}^z/n)^{-1}\left((\hat{\boldsymbol{\Lambda}}^z)^{1/2}\hat{\mathbf{P}}^{z\prime} - \hat{\mathbf{W}}_q[\hat{\boldsymbol{\Lambda}}^z - \boldsymbol{\Lambda}^\psi + \boldsymbol{\Lambda}^\psi](\boldsymbol{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\right)\mathbf{A}(L)\mathbf{x}_t/n\right\|$$
$$\leq \|(\hat{\boldsymbol{\Lambda}}^z/n)^{-1}\|\left\|\left((\hat{\boldsymbol{\Lambda}}^z)^{1/2}\hat{\mathbf{P}}^{z\prime} - \hat{\mathbf{W}}_q(\boldsymbol{\Lambda}^\psi)^{1/2}\mathbf{P}^{\psi\prime}\right)\right\|\|\mathbf{A}(L)\mathbf{x}_t/n\|$$
$$+\|(\hat{\boldsymbol{\Lambda}}^z/n)^{-1}\|\|\hat{\mathbf{W}}_q(\hat{\boldsymbol{\Lambda}}^z - \boldsymbol{\Lambda}^\psi)(\boldsymbol{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\|\|\mathbf{A}(L)\mathbf{x}_t/n\| = O_P(\zeta_{nT}),$$

since, by (D.10), $\|(\hat{\mathbf{P}}^z(\hat{\boldsymbol{\Lambda}}^z)^{1/2} - \mathbf{P}^\psi(\boldsymbol{\Lambda}^\psi)^{1/2}\hat{\mathbf{W}}_q)\| = O_P\left(n^{1/2}\zeta_{nT}\right)$, and

$$\|\hat{\mathbf{A}}(L)\mathbf{x}_t/n\| = n^{-1/2}\left(\mathbf{x}_t'\hat{\mathbf{A}}'(L)\hat{\mathbf{A}}(L)\mathbf{x}_t/n\right)^{1/2}$$
$$\leq n^{-1/2}\sum_{r=0}^{p}\left(\mathbf{x}_{t-r}'\hat{\mathbf{A}}_r'\hat{\mathbf{A}}_r\mathbf{x}_{t-r}/n\right)^{1/2}$$
$$\leq n^{-1/2}\sum_{r=0}^{p}(\mathbf{x}_{t-r}'\mathbf{x}_{t-r}/n)^{1/2}(\lambda_1(\hat{\mathbf{A}}_r'\hat{\mathbf{A}}_r))^{1/2} = O_P(n^{-1/2}),$$

boundedness of $\lambda_1(\hat{\mathbf{A}}_r'\hat{\mathbf{A}}_r)$ being a consequence of Assumptions 2 and 7. As for the third term on the right-hand side of (E.11), $(\boldsymbol{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\mathbf{A}(L)\boldsymbol{\xi}_t$ is $O_P(n^{-1/2})$. To conclude, note that the last term $\hat{\mathbf{W}}_q(\boldsymbol{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\mathbf{P}^\psi(\boldsymbol{\Lambda}^\psi)^{1/2}\mathbf{v}_t$ is equal to $\hat{\mathbf{W}}_q\mathbf{v}_t$. The conclusion follows. □

Appendix F and G. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jeconom.2017.04.002.

References

Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203–1227.

Alessi, L., Barigozzi, M., Capasso, M., 2010. Improved penalization for determining the number of factors in approximate factor models. Statist. Probab. Lett. 80, 1806–1813.

Amengual, D., Watson, M., 2007. Consistent estimation of the number of dynamic factors in a large *N* and *T* panel. J. Bus. Econom. Statist. 25, 91–96.

Anderson, B., Deistler, M., (2008a). Generalized linear dynamic factor models—A structure theory, 2008 IEEE Conference on Decision and Control.

Anderson, B., Deistler, M., 2008b. Properties of zero-free transfer function matrices. SICE J. Control Meas. Syst. Integr. 1, 1–9.

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. J. Bus. Econom. Statist. 25, 52–60.

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

D'Agostino, A., Giannone, D., 2012. Comparing alternative predictors based on large panel factor models. Oxford Bull. Econ. Stat. 74, 306–326.

Forni, M., Gambetti, L., 2014. Sufficient information in structural VARs. J. Monetary Econ. 66, 124–136.

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.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic factor model: identification and estimation. Rev. Econ. Stat. 82, 540–554.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2004. The generalized dynamic factor model: consistency and rates. J. Econometrics 119, 231–255.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2005. The generalized factor model: one-sided estimation and forecasting. J. Amer. Statist. Assoc. 100, 830–840.

Forni, M., Hallin, M., Lippi, M., Zaffaroni, P., 2015. Dynamic factor models with infinite-dimensional factor space: one-sided representations. J. Econometrics 185, 359–371.

Forni, M., Giovannelli, A., Lippi, M., Soccorsi, S., 2016. Dynamic factor model with infinite-dimensional factor space: forecasting, CEPR DP 11161.

Forni, M., Lippi, M., 2001. The generalized dynamic factor model: representation theory. Econometric Theory 17, 1113–1341.

Franklin, J.N., 2000. Matrix Theory. Dover Publications, New York.

Hallin, M., Lippi, M., 2013. Factor models in high-dimensional time series: a time-domain approach. Stochastic Process. Appl. 123, 2678–2695.

Hallin, M., Liška, R., 2007. Determining the number of factors in the general dynamic factor model. J. Amer. Statist. Assoc. 102, 603–617.

Lancaster, P., Tismenetsky, M., 1985. The Theory of Matrices, second edition. Academic Press, New York.

Lehmann, E.L., Casella, G., 1998. Theory of Point Estimation, second edition. Springer, New York.

Liu, W.D., Wu, W.B., 2010. Asymptotics of spectral density estimates. Econometric Theory 26, 1218–1245.

Onatski, A., 2009. Testing hypotheses about the number of factors in large factor models. Econometrica 77, 1447–1479.

Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. Rev. Econ. Stat. 92, 1004–1016.

Shao, W., Wu, W.B., 2007. Asymptotic spectral theory for nonlinear time series. Ann. Statist. 35, 1773–1801.

Stock, J.H, Watson, M.W., 2002a. Forecasting using principal components from a large number of predictors. J. Amer. Statist. Assoc. 97, 1167–1179.

Stock, J.H., Watson, M.W., 2002b. Macroeconomic forecasting using diffusion indexes. J. Bus. Econ. Statist. 20, 147–162.

Stock, J.H, Watson, M.W., 2012. Generalized shrinkage methods for forecasting using many predictors. J. Bus. Econ. Statist. 30, 481–493.

Wu, W.B., 2005. Nonlinear system theory: Another look at dependence. Proc. Natl. Acad. Sci. USA 102, 14150–14154.

Wu, W.B., Zaffaroni, P., 2017. Asymptotic theory for spectral density estimates of general multivariate time series. Econometric Theory 1–22.