Dynamic Factor Model with Infinite-Dimensional Factor Space: Estimation

Mario Forni

Università di Modena e Reggio Emilia, CEPR and RECent

Marc Hallin*†

ECARES, Université Libre de Bruxelles and ORFE, Princeton University

Marco Lippi

Università di Roma La Sapienza and EIEF

Paolo Zaffaroni‡

Imperial College London and Università di Roma La Sapienza

November 8, 2014

*Research supported by the Sonderforschungsbereich "Statistical modelling of nonlinear dynamic processes" (SFB 823) of the Deutsche Forschungsgemeinschaft, and by a Discovery Grant of the Australian Research Council.

†Académie Royale de Belgique and CentER, Tilburg University.

‡Research supported by the ESRC Grant RES-000-22-3219.

Abstract. Factor models, all particular cases of the Generalized Dynamic Factor Model (GDFM) introduced in Forni, Hallin, Lippi and Reichlin (2000), have become extremely popular in the theory and practice of large panels of time series data. The corresponding estimators rely on Brillinger's dynamic principal components thus involving two-sided filters leading to an extremely modest forecasting performance. No such problem arises with estimators based on standard principal components, which have been dominant in this literature. On the other hand, those other estimators require the assumption that the space spanned by the factors has finite dimension, severely limiting the generality afforded by the GDFM. This paper derives the asymptotic properties, namely consistency with exact rates of convergence, for a semiparametric estimator of the parameters and common shocks for a general class of GDFMs without relying on two-sided filters nor on the finite dimension assumption. A Monte Carlo experiment corroborates our theoretical results and shows the potential of these one-sided infinite dimensional GDFM, recently studied in Forni, Hallin, Lippi and Zaffaroni (2014), for the purpose of out-of-sample forecasting.

JEL subject classification : C0, C01, E0.
Key words and phrases : Generalized dynamic factor models. Vector processes with singular spectral density. One-sided representations for dynamic factor models. consistency and rates for estimators of dynamic factor models.

1 Introduction

In the present paper we provide consistent estimation results for the *Generalized Dynamic Factor Model* (GDFM) recently studied in Forni, Hallin, Lippi and Zaffaroni (2014) (FHLZ).

A GDFM, as introduced in Forni *et al.* (2000) and Forni and Lippi (2001), is a countably infinite set of observable stochastic processes x_{it} with the following decomposition:

$$x_{it} = \chi_{it} + \xi_{it} = b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} + \xi_{it}, \quad i \in \mathbb{N},\ t \in \mathbb{Z}, \tag{1.1}$$

where $\mathbf{u}_t = (u_{1t}\ u_{2t}\ \cdots\ u_{qt})'$ is a q-dimensional orthonormal unobservable white noise vector and $b_{if}(L), i \in \mathbb{N},\ f = 1, \ldots, q$, are square-summable filters (L, as usual, stands for the lag operator). Detailed assumptions on the *common components* χ_{it} and the *idiosyncratic components* ξ_{it} are given below. Let us only recall here that the idiosyncratic components and the *common shocks* u_{ft} are orthogonal at any lead and lag, and that the idiosyncratic components are "weakly" cross-correlated (orthogonality is an extreme case).

The recent literature on Dynamic Factor Models is based on (1.1) under the assumption that the space spanned by the stochastic variables χ_{it}, for t given and $i \in \mathbb{N}$, is *finite dimensional* (the definition of the common components obviously implies that the dimension of the space spanned by the variables χ_{it}, for t given and $i \in \mathbb{N}$, is independent of t). Seminal papers are Stock and Watson (2002a,b), Bai and Ng (2002). Under that assumption, model (1.1) can be rewritten in the so-called *static representation*

$$\begin{aligned} x_{it} &= \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \cdots + \lambda_{ir}F_{rt} + \xi_{it} \\ \mathbf{F}_t &= (F_{1t}\ \ldots\ F_{rt})' = \mathbf{N}(L)\mathbf{u}_t. \end{aligned} \tag{1.2}$$

Criteria to determine r consistently are given in Bai and Ng (2002) (see also Alessi et al. 2010 and the literature therein). 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 (1.2) is usually specified as a singular VAR, so that (1.2) becomes

$$\begin{aligned} x_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \cdots + \lambda_{ir}F_{rt} &\quad+\quad \xi_{it} \\ \mathbf{D}(L)\mathbf{F}_t = (\mathbf{I} - \mathbf{D}_1 L - \mathbf{D}_2 L^2 - \ldots - \mathbf{D}_p L^p)\mathbf{F}_t &\quad=\quad \mathbf{K}\mathbf{u}_t, \end{aligned} \tag{1.3}$$

where the matrices $\mathbf{D}_j$ are $r \times r$ while $\mathbf{K}$ is $r \times q$. Under (1.3), Bai and Ng (2007) and Amengual and Watson (2007) provide consistent criteria to determine q.

In GHLZ we argue that the finite-dimensional assumption is far from being innocuous. For instance, (1.2) is so restrictive that even the very elementary model

$$x_{it} = a_i(1 - \alpha_i L)^{-1}u_t + \xi_{it}, \tag{1.4}$$

where $q = 1$, u_t is scalar white noise, and the coefficients α_i are drawn from a uniform distribution, is ruled out. In this case the space spanned, for a given t, by the common components χ_{it}, $i \in \mathbb{N}$, is easily seen to be infinite-dimensional unless the α_i's take only a finite number of values.

The problem with models like (1.1), when the space spanned by the common components is infinite-dimensional, is that estimation cannot be based on a finite number r of standard principal components. The GDFM without finite-dimensional assumptions is studied in Forni et al. (2000). They use q principal components in the frequency domain (Brillinger 1981) to estimate the common components χ_{it} (criteria to determine q without assuming (1.2) or (1.3) are obtained in Hallin and Liška, 2007 and Onatski, 2009). However, their estimator involves the application of two-sided filters to the observable variables x_{it} and hence is useless at the end of the sample or for prediction.

In FHLZ we show how to obtain one-sided estimators without the finite-dimension assumption. We impose the weaker condition that the common components have a *rational spectral density*, that is, each filter $b_{if}(L)$ in (1.1) is a ratio of polynomials in L:

$$\chi_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt}, \quad i \in \mathbb{N},\ f = 1, 2, \ldots, q, \tag{1.5}$$

where

$$c_{if}(L) = c_{if,0} + c_{if,1}L + \ldots + c_{if,s_1}L^{s_1} \quad \text{and} \quad d_{if}(L) = d_{if,0} + d_{if,1}L + \ldots + d_{if,s_2}L^{s_2}$$

(the degrees s_1 and s_2 of the polynomials are assumed to be independent of i for the sake of simplicity).

Denote by $\mathbf{x}_t$, $\boldsymbol{\chi}_t$, $\boldsymbol{\xi}_t$ the infinite-dimensional column vectors whose coordinates are the variables x_{it}, χ_{it}, ξ_{it}, respectively. Elaborating upon recent results by Anderson and Deistler (2008a, b), in FHLZ we 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 $\boldsymbol{\chi}_t = (\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{nt}\ \cdots)'$ has a *unique autoregres-*

sive 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 $(q+1)\times q$. Denoting by $\underline{\mathbf{A}}(L)$ and $\underline{\mathbf{R}}$ the (infinite) matrices on the left- and right-hand sides of (1.6), using $\boldsymbol{\chi}_t = \mathbf{x}_t - \boldsymbol{\xi}_t$, and setting $\mathbf{Z}_t = \underline{\mathbf{A}}(L)\mathbf{x}_t$:

$$\mathbf{Z}_t = \underline{\mathbf{R}}\mathbf{u}_t + \underline{\mathbf{A}}(L)\boldsymbol{\xi}_t. \tag{1.7}$$

Under the assumptions of the present paper the term $\underline{\mathbf{A}}(L)\boldsymbol{\xi}_t$ is an idiosyncratic component, so that (1.7) is a a static representation of the form (4.8) with $\mathbf{D}(L) = \mathbf{I}$.

Thus, under the specification of a rational spectral density for the common components, we obtain one-sided filters for the common components without the standard finite-dimension restriction. Moreover, the large-dimensional VAR (1.6) is obtained by piecing tothether the small-dimensional matrices $\mathbf{A}^k(L)$, each one depending only on the covariances of $q+1$ common components. Therefore no curse of dimensionality occurs with our procedure.

Our estimation of the common components χ_{it}, the shocks $\mathbf{u}_t$ and the filters $b_{if}(L)$ is based on the sample analogues of representations (1.6) and (1.7):

(i) We start with a lag-window estimator of the spectral density matrix of the observed vector $\mathbf{x}_{nt} = (x_{1t}\ x_{2t}\ \cdots\ x_{nt})$, call it $\hat{\boldsymbol{\Sigma}}^x_n(\theta)$.

(ii) Using the first q frequency domain principal components of $\hat{\boldsymbol{\Sigma}}^x_n(\theta)$, we construct an estimator of the spectral density of $\boldsymbol{\chi}_{nt} = (\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{nt})$, call it $\hat{\boldsymbol{\Sigma}}^\chi_n(\theta)$ (like in Forni et al., 2000) Estimators of the autocovariances of $\boldsymbol{\chi}_{nt}$ are then obtained from $\hat{\boldsymbol{\Sigma}}^\chi_n(\theta)$, call $\hat{\boldsymbol{\Gamma}}^\chi_{n,h}$ the estimator of the covariance between $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\chi}_{n,t-h}$. The estimated covariances $\hat{\boldsymbol{\Gamma}}^\chi_{n,h}$ are used to obtain estimators $\hat{\mathbf{A}}^k(L)$.

(iii) A blockwise estimator of the variables Z_{jt} is obtained by applying the finite-degree matrices $\hat{\mathbf{A}}^k(L)$ to the observed variables x_{it}. Inversion of the matrices $\hat{\mathbf{A}}^k(L)$ provides estimators for the filters $b_{if}(L)$. Estimators for the shocks u_{ft} and the entries of the matrix $\mathbf{R}$ are obtained by using the first q time-domain principal components of the variables Z_{it}.

Our consistency results for the estimators described in (ii) and (iii) above are based on recent results on lag-window spectral estimators in Shao and Wu (2007), Liu and Wu (2010) as extended to the multivariate case in Wu and Zaffaroni (2014). Starting with the observable time series x_{it}, denoting by T the number of observations for each series and $\hat{\sigma}_{ij}(\theta)$ a lag-window estimator of the cross spectrum between x_{it} and x_{jt}, the $(i,\ j)$ entry of $\hat{\mathbf{\Sigma}}(\theta)$), under quite general assumptions on the processes x_{it}, x_{jt} and the kernel, these papers prove that $\hat{\sigma}_{ij}(\theta)$ is consistent, as $T \to \infty$, *uniformly* with respect to θ with rate $\sqrt{B_T \log B_T / T}$, where B_T is the size of the lag window. As an important innovation with respect to the previous literature on spectral estimation, these results are obtained without assuming linearity or Gaussianity of the processes x_{it}.

The use of these results in our framework requires significant enhancement of the assumptions on the common shocks and the idiosyncratic components that we make in FHLZ. In particular, (1) the vector $\mathbf{u}_t$, which is a white noise in FHLZ, is i.i.d. here, (2) the idiosyncratic components depend on an infinite-dimensional i.i.d. vector. These, as well as other changes in the assumptions with respect to FHLZ will be discussed in detail in the paper. Under this enhanced set of assumptions we prove that the estimators $\hat{\mathbf{\Sigma}}^{\chi}(\theta)$, $\hat{\mathbf{\Gamma}}^{\chi}$ and $\hat{\mathbf{A}}^k(L)$ are consistent with rate

$$\zeta_{nT} = \max\left(\sqrt{n^{-1}}, \sqrt{T^{-1} B_T \log B_T}\right), \tag{1.8}$$

where B_T diverges as T^{δ}, with $\frac{1}{3} < \delta < 1$.

Lastly, though model (1.7) is finite-dimensional, the series Z_{it} are estimated, not observed. As a consequence, the well-known results in factor literature (Stock and Watson, 2002a and b, Bai and Ng, 2002) do not immediately apply and proving consistency of the standard principal components estimators for the shocks $\mathbf{u}_t$ and the loadings $\mathbf{R}^k$ implies serious technical complications. Still we are able to achieve consistency without losing in consistency rate, which is, again, ζ_{nT}.

As we have pointed out in FHLZ, end of Section 4.5, though the dynamic model studied in the present paper is more general than model (1.3), when a dataset is given, with finite n and T, the static approach might perform well even if the data were generated by a model not fulfilling the finite-dimension assumption. In the present paper the static and dynamic methods have been applied to simulated data in several Monte Carlo experiments. A very short summary of our results is that (i) when the data are generated by infinite-dimensional models which are simple generalization of (4.8), the estimation of impulse-response functions and predictions by the dynamic method is by far better than those obtained via the static method; (ii) when the data are generated by (1.3), still the dynamic method performs slightly better. Though not conclusive, our Monte Carlo results strongly suggest that the model proposed in the present paper may be a competitive specification for dynamic factor models.

In Sections 2 and 3 we present and comment the assumptions and the estimators' asymptotic properties respectively. Section 4 gives a detailed description of the Monte Carlo experiments. Section 5 concludes. Short proofs are given in the body of the paper, the longer ones in the Appendix

2 Assumptions

2.1 Common and idiosyncratic components

The Dynamic Factor Model studied in the present paper is a family of stochastic variables

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

such that $x_{it} = \chi_{it} + \xi_{it}$. The variables x_{it} and their components χ_{it} (called the *common* components) and ξ_{it} (the *idiosyncratic* components) fulfill 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 $\mathbf{u}_t = (u_{1t} \; u_{2t} \; \cdots \; u_{qt})'$, $t \in \mathbb{Z}$, 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 $\beta_{ij,k}$, for $i, j \in \mathbb{N}$, $k = 0, 1, \ldots, \infty$, where $\sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \beta_{ij,k}^2 < \infty$ for all $i \in \mathbb{N}$; such that:

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

(ii)

$$\begin{aligned} \chi_{it} &= b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} \\ \xi_{it} &= \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \beta_{ij,k} \eta_{j,t-k}. \end{aligned} \tag{2.1}$$

Let us observe that neither $\mathbf{u}_t$ nor the polynomials $b_{if}(L)$ are identified. For, rewriting the second equation in (2.1) as $\chi_{it} = \mathbf{b}_i(L)\mathbf{u}_t$, for any orthogonal matrix $\mathbf{Q}$ the common component χ_{it} has the alternative representation $\chi_{it} = \left[\mathbf{b}_i(L)\mathbf{Q}^{-1}\right] [\mathbf{Q}\mathbf{u}_t] = \mathbf{b}_i^*(L)\mathbf{u}_t^*$.

Note that (i) and the definition of ξ_{it} in (2.1) imply $\mathrm{cov}(u_{ft}, \xi_{i,t-k}) = 0$ for all f, i, k. Two differences with respect to FHLZ must be pointed out. Firstly, here $\mathbf{u}_t$ is i.i.d., not only white noise as in FHLZ. Secondly, unlike in FHLZ, the idiosyncratic components are modeled as (infinite-order) moving averages of the infinite-dimensional i.i.d. vector $\boldsymbol{\eta}_t$.

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

(i) The filters $b_{if}(L)$ are rational. More precisely, $b_{if}(L) = \dfrac{c_{if}(L)}{d_{if}(L)}$ where

$$c_{if}(L) = c_{if,0} + c_{if,1}L + \cdots + c_{if,s_1}L^{s_1} \quad \textit{and} \quad d_{if}(L) = 1 + d_{if,1}L + \cdots + d_{if,s_2}L^{s_2}, \tag{2.2}$$

for $i \in \mathbb{N}$, $f = 1, \ldots, q$.

(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 $|c_{if,j}| \le B^\chi$, $i \in \mathbb{N}$, $f = 1, \dots, q$, $j = 0, \dots, s_1$.

Under Assumption 2, the vectors $\boldsymbol{\chi}_{nt} = (\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{nt})'$ have rational spectral density.

Assumption 3 *Eigenvalues of the spectral density of the common components.*
Let $\boldsymbol{\Sigma}_n^\chi(\theta)$ be the spectral density matrix of $\boldsymbol{\chi}_{nt}$ and $\lambda_{nj}^\chi(\theta)$ be its eigenvalues in decreasing order. There exist real numbers α_f^χ, $f = 1, \dots, q$, β_f^χ, $f = 0, \dots, q-1$, and a positive integer n^χ such that for $n > n^\chi$,

$$\beta_0^\chi \ge \frac{\lambda_{n1}^\chi(\theta)}{n} \ge \alpha_1^\chi > \beta_1^\chi \ge \frac{\lambda_{n2}^\chi(\theta)}{n} \ge \alpha_2^\chi > \beta_2^\chi \ge \cdots \ge \alpha_{q-1}^\chi > \beta_{q-1}^\chi \ge \frac{\lambda_{nq}^\chi(\theta)}{n} \ge \alpha_q^\chi > 0,$$

for all $\theta \in [-\pi,\ \pi]$.

Assumption 3 is an enhancement of the standard assumption on the eiegenvalues of the common components. It will be used in our consistency proof, see in particular Lemma 3, Appendix B.

Assumption 4 *Cross and time dependence of the idiosyncratic components.*
There exists finite positive numbers B, B_{is}, $i \in \mathbb{N}$, $s \in \mathbb{N}$, and ρ, $0 \le \rho < 1$, such that

$$\sum_{s=1}^{\infty} B_{is} \le B, \quad \textit{for all } i \in \mathbb{N} \tag{2.3}$$

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

$$|\beta_{is,k}| \le B_{is}\rho^k, \quad \textit{for all } i, s \in \mathbb{N} \textit{ and } k = 0, 1, \dots \tag{2.5}$$

A consequence of (2.3) and (2.4) is that

$$\sum_{i=1}^{\infty}\sum_{s=1}^{\infty} B_{is}B_{js} \le B^2, \quad \text{for all } j \in \mathbb{N}. \tag{2.6}$$

For, the right-hand side is

$$\sum_{s=1}^{\infty}\left(B_{js}\sum_{i=1}^{\infty} B_{is}\right) \le B\sum_{s=1}^{\infty} B_{js} \le B^2.$$

Conditions (2.3) and (2.4) are quite obviously satisfied in the "pure idiosyncratic" case $\xi_{it} = \eta_{it}$ and for finite "cross-section moving averages", for example $\xi_{it} = \eta_{it} + \eta_{i+1,t}$. By condition (2.5), the time dependence of the variables ξ_{it} declines geometrically at the 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,

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

Therefore, letting $\sigma_{ij}^{\xi}(\theta)$ be the cross spectral density between ξ_{it} and ξ_{jt},

$$\begin{aligned} \sum_{i=1}^{\infty} |\sigma_{ij}^{\xi}(\theta)| &\leq \frac{1}{2\pi} \sum_{i=1}^{\infty} \sum_{s=1}^{\infty} |\beta_{is}(e^{-\imath\theta}) \overline{\beta_{js}(e^{-\imath\theta})}| \leq \frac{1}{2\pi(1-\rho)^2} \sum_{i=1}^{\infty} \sum_{s=1}^{\infty} B_{is} B_{js} \\ &\leq B^2 \frac{1}{2\pi(1-\rho)^2}, \end{aligned} \tag{2.7}$$

by (2.6). Thus Assumption 4 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), implies 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.

Define the spectral density matrices $\mathbf{\Sigma}_n^{\xi}(\theta)$, $\mathbf{\Sigma}_n^{x}(\theta)$ and their eigenvalues $\lambda_{nj}^{\xi}(\theta)$ and $\lambda_{nj}^{x}(\theta)$ in the same way as $\mathbf{\Sigma}_n^{\chi}(\theta)$ and $\lambda_{nj}^{\chi}(\theta)$.

Proposition 1 *Under Assumptions 1 through 4,*
(i) there exists $B^{\xi} > 0$ *such that*

$$\lambda_{n1}^{\xi}(\theta) \leq B^{\xi}$$

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* $\theta \in [-\pi,\ \pi]$,

$$\frac{\lambda_{n1}^{x}(\theta)}{n} > \alpha_1^{\chi} > \frac{\lambda_{n2}^{x}(\theta)}{n} > \alpha_2^{\chi} > \cdots > \frac{\lambda_{nq}^{x}(\theta)}{n} > \alpha_q^{\chi},$$

where the numbers α_j^{χ} *are defined in Assumption 3.*
(iii) There exists $B^x > 0$ *such that* $\lambda_{n,q+1}^{x}(\theta) \leq B^x$ *for all* $n \in \mathbb{N}$ *and* $\theta \in [-\pi,\ \pi]$.

PROOF. The column and the row norm of $\mathbf{\Sigma}_n^\xi(\theta)$ are

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

by (2.7). On the other hand, the product of the row and the column norms, the square of the column norm in our case, is greater or equal to the square of the spectral norm, see Lancaster and Tismenetsky, p. 366, Exercise 11. As a consequence, setting $B^\xi = B^2\frac{1}{2\pi(1-\rho)^2}$, we have $\lambda_{n1}^\xi(\theta) \le B^\xi$ for all n and θ.

Regarding (ii), $\mathbf{\Sigma}_n^x(\theta) = \mathbf{\Sigma}_n^\chi(\theta) + \mathbf{\Sigma}_n^\xi(\theta)$ implies that $\lambda_{nf}^x(\theta) \ge \lambda_{nf}^\chi(\theta) + \lambda_{nn}^\xi(\theta)$ (this is one of the Weyl's inequalities, see Franklin (2000), p. 157, Theorem 1; see also Appendix B in the present paper). By Assumption 3,

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

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

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

if $n > n^\chi$ and such that $\frac{B^\xi}{n} < \min_{f=1,2,\dots,q}(\alpha_f^\chi - \beta_f^\chi)$.

For (iii), $\lambda_{n,q+1}^x \le \lambda_{n,q+1}^\chi(\theta) + \lambda_{n1}^\xi(\theta)$ (another Weyl inequality). 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

$$-\pi = \theta_0 < \theta_1 < \dots < \theta_{\nu-1} < \theta_\nu = \pi$$

of the interval $[-\pi, \pi]$.

PROOF. Let $h \geq 0$ and denote by $\gamma^{\xi}_{ij,h}$ the covariance between ξ_{it} and $\xi_{j,t-h}$.

$$|\gamma^{\xi}_{ij,h}| = \left| \sum_{k=0}^{\infty} \sum_{s=1}^{\infty} \beta_{is,k} \beta_{js,k+h} \right| \leq \sum_{k=0}^{\infty} \sum_{s=1}^{\infty} B_{is} B_{js} \rho^k \rho^{k+h} \leq \rho^h \sum_{k=0}^{\infty} \rho^{2k} \sum_{s=1}^{\infty} B_{is} B_{js} \leq \rho^h \frac{B^2}{1-\rho^2},$$

by (2.6). If $h < 0$, $\gamma^{\xi}_{ij,h} = E(\xi_{it}\xi_{j,t-h}) = E(\xi_{jt}\xi_{i,t-(-h)}) = \gamma^{\xi}_{ji,-h}$. In conclusion

$$|\gamma^{\xi}_{ij,h}| \leq \rho^{|h|} \frac{B^2}{1-\rho^2}.$$

This implies that

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

has all derivatives. Moreover,

$$|\sigma^{\xi}_{ij}{}'(\theta)| = \frac{1}{2\pi} \left| \sum_{h=-\infty}^{\infty} (-\imath h) \gamma^{\xi}_{ij,h} e^{-\imath h\theta} \right| \leq \frac{B^2}{\pi(1-\rho^2)} \sum_{h=1}^{\infty} h\rho^h = \frac{B^2}{\pi(1-\rho^2)(1-\rho)^2}$$

This implies bounded variation of $\sigma^{\xi}_{ij}(\theta)$ uniformly in i and j. Bounded variation of $\sigma^{\chi}_{ij}(\theta)$, uniformly in i and j, is an obvious consequence of Assumption 2. The conclusion follows from $\sigma^{x}_{ij}(\theta) = \sigma^{\chi}_{ij}(\theta) + \sigma^{\xi}_{ij}(\theta)$. □

2.2 Autoregressive representation of the χ's

In FHLZ we prove that, for *generic values* of the parameters $c_{if,k}$ and $d_{if,k}$ in (2.2), the space spanned by $u_{f,t-k}$, $f = 1, 2, \ldots, q$, $k \geq 0$, is equal to the space spanned by any $(q+1)$-dimensional subvector of $\boldsymbol{\chi}_t$ and its lags. In other words, $\mathbf{u}_t$ is *fundamental* for all the $(q+1)$-dimensional subvectors of $\boldsymbol{\chi}_t$ (but not for all q-dimensional subvectors). Moreover, we prove that *generically* the $(q+1)$-dimensional subvectors of $\boldsymbol{\chi}_t$ have a *finite and unique* autoregressive representation (see Section 4.1, Lemma 3 in particular). Following FHLZ, we use these genericity results as a motivation for *assuming* that any $(q+1)$-dimensional subvector of $\boldsymbol{\chi}_t$ and its lags spans the space spanned by the u's and has a unique finite autoregressive representation.

Assumption 5 *Each vector* $\boldsymbol{\chi}_t^{i_1,i_2,\dots,i_{q+1}} = (\chi_{i_1t}\ \chi_{i_2t}\ \cdots\ \chi_{i_{q+1}t})'$, *with* $i_1 < i_2 < \cdots < i_{q+1}$, *has an autoregressive representation*

$$\mathbf{A}^{i_1,i_2,\dots,i_{q+1}}(L)\boldsymbol{\chi}_t^{i_1,i_2,\dots,i_{q+1}} = \mathbf{R}^{i_1,i_2,\dots,i_{q+1}}\mathbf{u}_t, \tag{2.8}$$

where

(i) $\mathbf{A}^{i_1,i_2,\dots,i_{q+1}}(L)$ *is of degree not greater than* $S = qs_1 + q^2s_2$, $\mathbf{A}^{i_1,i_2,\dots,i_{q+1}}(0) = \mathbf{I}_{q+1}$,

(ii) $\mathbf{R}^{i_1,i_2,\dots,i_{q+1}}$ *has rank* q,

(iii) Representation (2.8) is unique among the autoregressive representations of order not greater than S, *i.e. if* $\mathbf{B}(L)\boldsymbol{\chi}_t^{i_1,i_2,\dots,i_{q+1}} = \tilde{\mathbf{R}}\mathbf{u}_t$, *where the degree of* $\mathbf{B}(L)$ *does not exceed* S *and* $\mathbf{B}(0) = \mathbf{I}_{q+1}$, *then* $\mathbf{B}(L) = \mathbf{A}^{i_1,i_2,\dots,i_{q+1}}(L)$ *and* $\tilde{\mathbf{R}} = \mathbf{R}^{i_1,i_2,\dots,i_{q+1}}$.

An immediate consequence of Assumption 5 is that $\boldsymbol{\chi}_t$ can be represented as in (1.6), that is,

$$\mathbf{A}^1(L)\begin{pmatrix}\chi_{1t}\\ \chi_{2t}\\ \vdots\\ \chi_{q+1,t}\end{pmatrix} = \mathbf{R}^1\mathbf{u}_t, \qquad \mathbf{A}^2(L)\begin{pmatrix}\chi_{q+2,t}\\ \chi_{q+3,t}\\ \vdots\\ \chi_{2(q+1),t}\end{pmatrix} = \mathbf{R}^2\mathbf{u}_t, \qquad \dots \tag{2.9}$$

where the degrees of the polynomial matrices $\mathbf{A}^k(L)$ do not exceed S. Moreover, the matrices $\mathbf{A}^k(L)$ are unique among the autoregressive representations of degree not greater than S. Writing $\underline{\mathbf{A}}(L)$ for the (infinite) block-diagonal matrix with diagonal blocks $\mathbf{A}^1(L), \mathbf{A}^2(L), \dots$, and letting $\underline{\mathbf{R}} = (\mathbf{R}^{1\prime}, \mathbf{R}^{2\prime}, \cdots)'$, we thus have

$$\underline{\mathbf{A}}(L)\boldsymbol{\chi}_t = \underline{\mathbf{R}}\mathbf{u}_t. \tag{2.10}$$

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,

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

PZ: The block-diagonal structure of $\mathbf{A}_n(L)$ holds for every arbitrary partition of the indexes $1, 2, ..., n$ into blocks of size m. In other words, the choice of the partition is irrelevant. In practical estimation, obviously, every partition could lead to slightly different results, a small sample effect that vanishes asymptotically.

The following proposition is an immediate consequence of the fact that (2.10) is the orthogonal projection of $\boldsymbol{\chi}_t$ on its past values.

Proposition 3 *Under Assumptions 1 through 5,*

(i) Let $\underline{\mathbf{A}}^*(L)\boldsymbol{\chi}_t = \underline{\mathbf{R}}^*\mathbf{v}_t$, *where* $\text{degree}(\underline{\mathbf{A}}^*(L)) \leq S$, *then* $\underline{\mathbf{R}}^* = \underline{\mathbf{R}}\mathbf{Q}'$, $\mathbf{v}_t = \mathbf{Q}\mathbf{u}_t$, $\underline{\mathbf{A}}^*(L) = \underline{\mathbf{A}}(L)$, *where* $\mathbf{Q}$ *is a* $q \times q$ *orthogonal matrix.*

(ii) Let $\mathbf{r} = (r_1 \ \cdots \ r_q)$ *be the row of* $\underline{\mathbf{R}}$, *or of* $\mathbf{R}^{i_1,i_2,\dots,i_{q+1}}$, *corresponding to* χ_{it}. *Then* $r_f = c_{if}(0) = c_{if,0}$, $f = 1, \dots, q$.

Let $\boldsymbol{\Psi}_t = \underline{\mathbf{A}}(L)\boldsymbol{\chi}_t = \underline{\mathbf{R}}\mathbf{u}_t$, let $\boldsymbol{\Gamma}_n^{\Psi}$ be the variance-covariance matrix of $\boldsymbol{\Psi}_{nt}$, with eigenvalues μ_{nj}^{Ψ}, in decreasing order.

Assumption 6 *There exist real numbers* α_f^{Ψ}, $f = 1, \dots, q$, β_f^{Ψ}, $f = 0, \dots, q-1$, *and a positive integer* n^{ψ} *such that for* $n > n^{\Psi}$,

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

Note that the eigenvalues μ_{nf}^{Ψ} depend on the coefficients $c_{if,0}$, see Proposition 3(ii), but 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.

We now show how (2.9), i.e. the matrices $\mathbf{A}^k(L)$ and (up to multiplication by an orthogonal matrix) $\mathbf{R}^k$, can be constructed starting with the spectral density of the χ's. This procedure leads to our estimator as explained in Section 3, with the population quantities replaced by their estimates.

(i) The nested spectral density matrices $\boldsymbol{\Sigma}_n^{\chi}(\theta)$, $n \in \mathbb{N}$, are known functions of the coefficients $c_{if,s}$ and $d_{if,s}$.

(ii) Denote by $\boldsymbol{\chi}_t^k$ the k-th of the $(q+1)$-dimensional subvectors of $\boldsymbol{\chi}_t$ appearing in (2.9), and call $\boldsymbol{\Sigma}_{jk}^{\chi}(\theta)$ the $(q+1) \times (q+1)$ cross-spectral density between $\boldsymbol{\chi}_t^j$ and $\boldsymbol{\chi}_t^k$. Then, denoting by $\boldsymbol{\Gamma}_{jk,s}^{\chi}$ the covariance between $\boldsymbol{\chi}_t^j$ and $\boldsymbol{\chi}_{t-s}^k$,

$$\mathbf{\Gamma}^{\chi}_{jk,s} = \mathrm{E}\left(\boldsymbol{\chi}^j_t \boldsymbol{\chi}^{k\prime}_{t-s}\right) = \int_{-\pi}^{\pi} e^{\imath s\theta} \mathbf{\Sigma}^{\chi}_{jk}(\theta) d\theta, \tag{2.12}$$

where $\imath$ stands for the imaginary unit.

(iii) Using the autocovariance function $\mathbf{\Gamma}^{\chi}_{kk,s}$, we obtain the minimum-lag matrix polynomial $\mathbf{A}^k(L)$ and the variance-covariance function of the unobservable vectors

$$\mathbf{\Psi}^1_t = \mathbf{A}^1(L)\boldsymbol{\chi}^1_t, \quad \mathbf{\Psi}^2_t = \mathbf{A}^2(L)\boldsymbol{\chi}^2_t, \quad \ldots \tag{2.13}$$

Indeed, letting $\mathbf{A}^k(L) = \mathbf{I}_{q+1} - \mathbf{A}^k_1 L - \cdots - \mathbf{A}^k_S L^S$, define

$$\mathbf{A}^{[k]} = \left(\mathbf{A}^k_1 \; \mathbf{A}^k_2 \; \cdots \; \mathbf{A}^k_S\right), \quad \mathbf{B}^{\chi}_k = \left(\mathbf{\Gamma}^{\chi}_{kk,1} \; \mathbf{\Gamma}^{\chi}_{kk,2} \; \cdots \; \mathbf{\Gamma}^{\chi}_{kk,S}\right) \tag{2.14}$$

and

$$\mathbf{C}^{\chi}_{jk} = \begin{pmatrix} \mathbf{\Gamma}^{\chi}_{jk,0} & \mathbf{\Gamma}^{\chi}_{jk,1} & \cdots & \mathbf{\Gamma}^{\chi}_{jk,S-1} \\ \mathbf{\Gamma}^{\chi}_{jk,-1} & \mathbf{\Gamma}^{\chi}_{jk,0} & \cdots & \mathbf{\Gamma}^{\chi}_{jk,S-2} \\ \vdots & & & \vdots \\ \mathbf{\Gamma}^{\chi}_{jk,-S+1} & \mathbf{\Gamma}^{\chi}_{jk,-S+2} & \cdots & \mathbf{\Gamma}^{\chi}_{jk,0} \end{pmatrix}. \tag{2.15}$$

We have

$$\mathbf{A}^{[k]} = \mathbf{B}^{\chi}_k \left(\mathbf{C}^{\chi}_{kk}\right)^{-1} = \mathbf{B}^{\chi}_k \left(\mathbf{C}^{\chi}_{kk}\right)_{\mathrm{ad}} \det\left(\mathbf{C}^{\chi}_{kk}\right)^{-1}, \tag{2.16}$$

where $\mathbf{F}_{\mathrm{ad}}$ denotes the adjoint of the square matrix $\mathbf{F}$.

Assumption 7 *There exist a real d such that*

$$\det \mathbf{C}^{\chi}_{kk} > d > 0,$$

for all $k \in \mathbb{N}$.

PZ: Non-singularity of the $\mathbf{C}^{\chi}_{kk}$ is necessary for existence of the $\mathbf{A}^{[k]}$. However, we require a slightly stronger condition to ensure that the $\mathbf{A}^{[k]}$ are (uniformly) bounded, in norm, as n and hence k diverge to infinity.

Now define $\mathbf{Z}_t = \underline{\mathbf{A}}(L)\mathbf{x}_t$ and consider the following factor model for $\mathbf{Z}_t$:

$$\begin{aligned} \mathbf{\Psi}_t &= \underline{\mathbf{R}}\mathbf{u}_t \\ \mathbf{\Phi}_t &= \underline{\mathbf{A}}(L)\boldsymbol{\xi}_t \\ \mathbf{Z}_t &= \mathbf{\Psi}_t + \mathbf{\Phi}_t. \end{aligned} \tag{2.17}$$

Proposition 4 *Let* $\mathbf{\Gamma}_n^\Phi$ *the variance-covariance matrix of* $\mathbf{\Phi}_{nt} = (\Phi_{1t}\ \Phi_{2t}\ \cdots\ \Phi_{nt})$ *and* μ_{nj}^Φ *its* j*-th eigenvalue. Under Assumptions 1 through 7 there exists* $B^\Phi > 0$ *such that* $\mu_{n1}^\Phi \leq B^\Phi$ *for all* $n \in \mathbb{N}$.

Proof. Let $\lambda_{nj}^\Phi(\theta)$ be the j-th eigenvalue of the spectral density matrix of $\mathbf{\Phi}_{nt}$. We want to prove that there exists $C^\Phi > 0$ such that $\lambda_{n1}^\Phi(\theta) \leq C^\Phi$ for all n and θ. Because $\lambda_{n1}^\Phi(\theta)$ is non-decreasing as n increases, for all θ (see Forni and Lippi, 2001), we can assume that $n = m(q+1)$. The spectral density of $\mathbf{\Phi}_{nt}$ is

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

where $\mathbf{A}_n(L)$ (see equation (2.11)) has the matrices $\mathbf{A}^k(L)$ on the diagonal. 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

$$\mathbf{a}(\theta)'\mathbf{A}_n(e^{-\imath\theta})\mathbf{\Sigma}_n^\xi(\theta)\mathbf{A}_n'(e^{\imath\theta})\overline{\mathbf{a}(\theta)} \leq \lambda_{n1}^\xi(\theta)\left(\mathbf{a}'(\theta)\mathbf{A}_n(e^{-\imath\theta})\mathbf{A}_n'(e^{\imath\theta})\overline{\mathbf{a}(\theta)}\right) \leq \lambda_{n1}^\xi(\theta)\lambda_1^{A_n}(\theta),$$

where $\lambda_1^{A_n}(\theta)$ is the first eigenvalue of $\mathbf{A}_n(e^{-\imath\theta})\mathbf{A}_n'(e^{\imath\theta})$, which is Hermitian, non-negative definite. By Proposition 1 $\sup_n \lambda_{n1}^\xi(\theta) \leq B^\xi$. Moreover, given the diagonal structure of $\mathbf{A}_n(L)$, $\lambda_1^{A_n}(\theta) = \sup_{k=1,2,\ldots,m} \lambda_1^{A^k}(\theta) \leq \sup_{k\in\mathbb{N}} \lambda_1^{A^k}(\theta)$, where $\lambda_1^{A^k}(\theta)$ is the first eigenvalue of $\mathbf{A}^k(e^{-\imath\theta})\mathbf{A}^{k\prime}(e^{\imath\theta})$. Assumptions 2 and 7 imply that $\sup_{k\in\mathbb{N}} \lambda_1^{A^k}(\theta) \leq D^\Phi$ for some $D^\Phi > 0$ and all θ. On the other hand,

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

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

Lastly,

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

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

Note that $\mathbf{\Phi}_t$ and $\mathbf{\Psi}_t$ are orthogonal, a consequence of Assumption 1(i). In view of Assumption 6 and Proposition 4, the model (2.17) is a special case of (1.3), 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 x's replaced with an estimator $\hat{\mathbf{\Sigma}}_n^x(\theta)$ fulfilling Assumption 9. Based on Forni *et al.* (2000), we obtain the estimator $\hat{\mathbf{\Sigma}}_n^\chi(\theta)$ by means of the first q frequency-domain principal components of the x's (using the first q eigenvectors of $\hat{\mathbf{\Sigma}}_n^x(\theta)$). Then the matrices $\hat{\mathbf{\Gamma}}_{jk}^\chi$, $\hat{\mathbf{B}}_{jk}^\chi$, $\hat{\mathbf{C}}_{jk}^\chi$ and $\hat{\mathbf{A}}_n(L)$ are computed as natural counterparts of their population versions in Section 2.2. Lastly, 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 with exact rate of convergence ζ_{nT}, as defined in equation (1.8), for all the above estimators are provided in Propositions 7 through 11.

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) $\mathbf{\Sigma}^x(\theta) = \left(\sigma_{ij}^x(\theta)\right)_{i,j=1,\dots,n}$ and $\lambda_f^x(\theta)$ replace $\mathbf{\Sigma}_n^x(\theta)$ and $\lambda_{nf}^x(\theta)$, respectively.

(ii) $\mathbf{\Lambda}^x(\theta)$ denotes the $q \times q$ diagonal matrix with diagonal elements $\lambda_f^x(\theta)$.

(iii) $\mathbf{P}^x(\theta)$ denotes the $n \times q$ matrix the q columns of which are the unit-modulus eigenvectors corresponding to $\mathbf{\Sigma}^x(\theta)$'s first q eigenvalues. The columns and entries of $\mathbf{P}^x(\theta)$ are denoted by $\mathbf{P}_f^x(\theta)$ and $p_{if}^x(\theta)$, respectively.

(iv) $\mathbf{\Sigma}^\chi(\theta) = \left(\sigma_{ij}^\chi(\theta)\right)_{i,j=1,\dots,n}$, $\lambda_f^\chi(\theta)$, $\mathbf{\Lambda}^\chi(\theta)$, $\mathbf{P}^\chi(\theta)$, $\mathbf{\Sigma}^\xi(\theta)$, etc. are defined as in (i).

(v) All the above matrices and scalars depend on n; the corresponding estimators,

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

(precise definitions are provided below) depend both on n and the sample 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 $\mathbf{\Gamma}_n^\psi$ 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 $\tilde{\mathbf{F}}$ the conjugate transposed of $\mathbf{F}$ and by $||\mathbf{F}||$ the spectral norm of $\mathbf{F}$ (see Appendix B).

3.1 Estimation of $\boldsymbol{\Sigma}_n^x(\theta)$

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 infinite-dimensional i.i.d. stochastic vector process and let*

$$z_t = F(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1}, \ldots),$$

where $F : [\mathbb{R} \times \mathbb{R} \times] \cdots \to \mathbb{R}$ is a measurable function. Let $\boldsymbol{\epsilon}^*$ be an infinite-dimensional stochastic vector with the same distribution as $\boldsymbol{\epsilon}_t$, such that $\boldsymbol{\epsilon}^*$ and $\boldsymbol{\epsilon}_t$ are independent for all t. Assume that z_t has finite p moment for $p > 0$. For $k \geq 0$ define

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

Assumption 8 *There exist p, A, with $p > 4$, $0 < A < \infty$, 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 $f = 1, \ldots, q$.

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 p-th moments and the physical dependence of the x's.

Proposition 5 *Under Assumptions 1 through 8, there exist ρ_1 and A_1, $0 < \rho_1 < 1$ and $0 < A_1 < \infty$, such that*

$$E\left(|x_{it}|^p\right) \leq A_1, \quad \delta_{kp}^{[x_{it}]} \leq A_1 \rho_1^k, \tag{3.2}$$

for all $i \in \mathbb{N}$.

Proof. For the first inequality, $\left(E\left(|x_{it}|^p\right)\right)^{\frac{1}{p}} = \left(E\left(|\chi_{it} + \xi_{it}|^p\right)\right)^{\frac{1}{p}} \leq \left(E\left(|\chi_{it}|^p\right)\right)^{\frac{1}{p}} + \left(E\left(|\xi_{it}|^p\right)\right)^{\frac{1}{p}}$, by the Minkovski inequality. Then, again using the Minkovski inequality, condition (2.3) and Assumption 8,

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

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

As regards the second inequality, for $k \geq 0$,

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

where ξ_{ik}^* has the same definition of ξ_{ik} with η_{s0} replaced by η_s^*. Then, using the Minkovski inequality, condition (2.3) and Assumption 8,

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

An analogous inequality can be ontained 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:

$$\delta_{kp}^{[x_{it}]} = \left(E\left|x_{it} - x_{it}^*\right|^p\right)^{\frac{1}{p}} = \left(E\left(\left|(\chi_{it} - \chi_{it}^*) + (\xi_{it} - \xi_{it}^*)\right|^p\right)^{\frac{1}{p}}\right)$$
$$\leq \left(E\left(\left|\chi_{it} - \chi_{it}^*\right|\right)^p\right)^{\frac{1}{p}} + \left(E\left(\left|\xi_{it} - \xi_{it}^*\right|\right)^p\right)^{\frac{1}{p}}$$
$$= \delta_{kp}^{[\chi_{it}]} + \delta_{kp}^{[\xi_{it}]}.$$

The conclusion follows. □

Consider now the lag-window estimator of the spectral density $\boldsymbol{\Sigma}_n^x(\theta)$:

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

where $\hat{\boldsymbol{\Gamma}}_k = \frac{1}{T}\sum_{t=|k|+1}^{T} \mathbf{x}_t \mathbf{x}_{t-|k|}$.

Assumption 9 *(i) The kernel function K is even, bounded, with support $[-1, 1]$. Moreover, (1) for some $\kappa > 0$ it satisfies $\lim_{u\to 0} |K(u) - 1| = O(|x|^\kappa)$, (2) $\int_{-\infty}^{\infty} K^2(u)du < \infty$, (3) $\sum_{j\in\mathbb{Z}} \sup_{|s-j|\leq 1} |K(jw) - K(sw)| = O(1)$ as $w \to 0$.*
(ii) For some $c_1, c_2 > 0$, δ and $\underline{\delta}$ such that $0 < \underline{\delta} < \delta < 1 < \delta(2\kappa + 1)$,

$$c_1 T^{\underline{\delta}} \leq B_T \leq c_2 T^{\delta}.$$

Proposition 6 *Under Assumptions 1 through 9, there exists $C > 0$, independent of i and j, such that*

$$E\left(\max_{|h|\leq B_T} \left|\hat{\sigma}_{ij}^x(\theta_h^*) - \sigma_{ij}^x(\theta_h^*)\right|^2\right) \leq C\left(T^{-1} B_T \log B_T\right), \tag{3.4}$$

where $\theta_h^* = \pi h / B_T$, for all $T \in \mathbb{N}$.

See Appendix A for the proof.

3.2 Estimation of $\sigma_{ij}^\chi(\theta)$ and $\gamma_{ij,k}^\chi$

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

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

Proposition 7 *Under Assumptions 1 through 7,*

$$\max_{|h|\le B_T} |\hat{\sigma}^{\chi}_{ij}(\theta^*_h) - \sigma^{\chi}_{ij}(\theta^*_h)| = O_{\mathrm{P}}\left(\zeta_{nT}\right),$$

where $\theta^*_h = \pi h/B_T$*, as* $T \to \infty$ *and* $n \to \infty$*.*

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),

$$\hat{\gamma}^{\chi}_{ij,\ell} = \frac{\pi}{B_T} \sum_{|h|\le B_T} e^{\imath\ell\theta^*_h} \hat{\sigma}^{\chi}_{ij}(\theta^*_h), \tag{3.5}$$

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

$$\begin{aligned}
|\hat{\gamma}^{\chi}_{ij,\ell} - \gamma^{\chi}_{ij,\ell}| \;&\le\; \frac{\pi}{B_T} \sum_{|h|\le B_T} |e^{\imath\ell\theta^*_h} \hat{\sigma}^{\chi}_{ij}(\theta^*_h) - e^{\imath\ell\theta^*_h} \sigma^{\chi}_{ij}(\theta^*_h)| \\
&\qquad + \left| \frac{\pi}{B_T} \sum_{|h|\le B_T} e^{\imath\ell\theta^*_h} \sigma^{\chi}_{ij}(\theta^*_h) - \int_{-\pi}^{\pi} e^{\imath\ell\theta} \sigma^{\chi}_{ij}(\theta) d\theta \right| \\
&\le\; \frac{\pi}{B_T} \sum_{|h|\le B_T} |\hat{\sigma}^{\chi}_{ij}(\theta^*_h) - \sigma^{\chi}_{ij}(\theta^*_h)| \\
&\qquad + \frac{\pi}{B_T} \sum_{|h|\le B_T} \max_{\theta^*_{h-1}\le\theta\le\theta^*_h} |e^{\imath\ell\theta^*_h} \sigma^{\chi}_{ij}(\theta^*_h) - e^{\imath\ell\theta} \sigma^{\chi}_{ij}(\theta)| \\
&\le\; \pi \max_{|h|\le B_T} |\hat{\sigma}^{\chi}_{ij}(\theta^*_h) - \sigma^{\chi}_{ij}(\theta^*_h)| + \frac{\pi B}{B_T} \sum_{|h|\le B_T} \max_{\theta^*_{h-1}\le\theta\le\theta^*_h} |e^{\imath\ell\theta^*_h} - e^{\imath\ell\theta}| \\
&\qquad + \frac{\pi}{B_T} \sum_{|h|\le B_T} \max_{\theta^*_{h-1}\le\theta\le\theta^*_h} |\sigma^{\chi}_{ij}(\theta^*_h) - \sigma^{\chi}_{ij}(\theta)| \\
&\le \pi \max_{|h|\le B_T} |\hat{\sigma}^{\chi}_{ij}(\theta^*_h) - \sigma^{\chi}_{ij}(\theta^*_h)| \\
&\quad + \frac{\pi B}{B_T} \sum_{|h|\le B_T} \left(|e^{\imath\ell\theta^*_{h-1}} - e^{\imath\ell\tilde{\theta}^*_{h-1}}| + |e^{\imath\ell\tilde{\theta}^*_{h-1}} - e^{\imath\ell\theta^*_{h-1}}| \right) \\
&\quad + \frac{\pi}{B_T} \sum_{|h|\le B_T} \left(|\sigma^{\chi}_{ij}(\theta^*_{h-1}) - \sigma^{\chi}_{ij}(\check{\theta}^*_{h-1})| + |\sigma^{\chi}_{ij}(\check{\theta}^*_{h-1}) - \sigma^{\chi}_{ij}(\theta^*_h)| \right),
\end{aligned} \tag{3.6}$$

where: (a) B is the bound in Proposition 1(i), (b) $\tilde{\theta}^*_{h-1}$ 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 $(1/B_T)$. Using Proposition 7 we obtain that $|\hat{\gamma}^{\chi}_{ij,\ell} - \gamma^{\chi}_{ij,\ell}|$ is $O_{\mathrm{P}}(\zeta_{nT}) + O(1/B_T)$.

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

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

$$|\hat{\gamma}^{\chi}_{ij,\ell} - \gamma^{\chi}_{ij,\ell}| = O_{\mathrm{P}}(\zeta_{nT}), \tag{3.7}$$

as $T \to \infty$ and $n \to \infty$.

3.3 Estimation of $\mathbf{A}^k(L)$ and $\mathbf{\Gamma}^{\psi}_{jk}$

Under our assumptions, the common component satisfies the block-diagonal vector autoregression (1.5) of finite order. If the χ_t were observed, estimation by OLS would be appropriate. However, although we do not observe the χ_t, we do have (consistent) estimates of their autocovariance function. This leads to the Yule-Walker estimator of both the autoregressive coefficients as well as of the one-step ahead innovation covariance matrix. The definition of the estimators $\hat{\mathbf{A}}^{[k]}$and $\hat{\mathbf{\Gamma}}^{\Psi}_{jk}$ is straightforward from (2.14), (2.15) and (2.16).

Proposition 9 *Under Assumptions 1 through 10,*

$$\|\hat{\mathbf{A}}^{[k]} - \mathbf{A}^{[k]}\| = O_{\mathrm{P}}(\zeta_{Tn}) \qquad \textit{and} \qquad \|\hat{\mathbf{\Gamma}}^{\Psi}_{jk} - \mathbf{\Gamma}^{\Psi}_{jk}\| = O_{\mathrm{P}}(\zeta_{nT})$$

as $T \to \infty$ and $n \to \infty$.

See Appendix C for the proof.

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

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

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

where $\mathbf{\Lambda}^{\Psi}$ 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

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

say, where $\mathbf{v}_t = \mathbf{H}\mathbf{u}_t$, with $\mathbf{H}$ orthogonal. Note that, given i and f, the entry (i, f) 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 $\mathbf{\Psi}_{nt}$ and is therefore independent of n.

Our estimator of $\mathcal{R} = \mathbf{P}^{\psi}(\mathbf{\Lambda}^{\psi})^{1/2}$ is $\hat{\mathcal{R}} = \hat{\mathbf{P}}^{z}(\hat{\mathbf{\Lambda}}^{z})^{1/2}$, where $\hat{\mathbf{P}}^{z}$ and $\hat{\mathbf{\Lambda}}^{z}$ 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, as already observed, is the reason for the complications we have to deal with in Appendix D.

Proposition 10 *Under Assumptions 1 through 10,*

$$\|\hat{\mathcal{R}}_i - \mathcal{R}_i\hat{\mathbf{W}}^{z}\| = O_{\mathrm{P}}(\zeta_{nT}),$$

as $T \to \infty$ and $n \to \infty$, where $\mathcal{R}_i$ is the i-th row of $\mathcal{R}$, and $\hat{\mathbf{W}}^{z}$ is a $q \times q$ diagonal matrix, depending on n and T, whose diagonal entries equal either 1 *or* -1.

See Appendix D for the proof.

Let us point out again that the i-th row of $\mathcal{R}$ depends on n. Therefore, Proposition 10 only states that the difference between the estimated entries of $\hat{\mathcal{R}}$ and the entries of $\mathcal{R}$ converges 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 $\mathbf{Z}_t = \underline{\mathbf{R}}^* \mathbf{u}_t^* + \mathbf{\Phi}_t$, or $\mathbf{Z}_{nt} = \mathbf{R}^* \mathbf{u}_t^* + \mathbf{\Phi}_t$, where the $n \times q$ matrices $\mathbf{R}^*$ are nested. In particular, starting with $\mathbf{Z}_{nt} = \mathcal{R}\mathbf{v}_t + \mathbf{\Phi}_{nt}$, 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, the entries 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_{Tn}.

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.[1]

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

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

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

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

Proposition 11 *Under Assumptions 1 through 10,*

$$\|\hat{\mathbf{v}}_t - \hat{\mathbf{W}}^z \mathbf{v}_t\| = O_{\mathrm{P}}(\zeta_{nT}),$$

as $T \to \infty$ and $n \to \infty$, where $\hat{\mathbf{W}}^z$ 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.

4 A simulation exercise

In the present section we evaluate the performance of the methods studied in the present paper, referred to as FHLZ. We focus on (i) estimation of impulse response function, (ii) estimation of structural shocks and (iii) 1-step-ahead forecasts. Regarding (i) and (ii), we compare FHLZ with model (1.3), which has been studied in Forni *et al.* (2009) and is 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 a static factor representation. We generate artificial data according to two simple models: (I) a dynamic factor model with no static factor representation (so that neither FGLR nor SW are consistent) and (II) a model having a static factor representation (under which all methods are consistent).

4.1 Data generating processes

We consider the following data generating processes.

Model I (with no static factor representation)

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

We generate u_{jt}, $j = 1, 2$ and ξ_{it}, $i = 1, \ldots, n$, $t = 1, \ldots, T$ as Gaussian, unit variance, independent variables; a_{ji} as independent variables, uniformly distributed on the interval $[-1, 1]$; α_{ji} 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 Choleski identification scheme on the first q variables. Precisely, denote by $\mathbf{B}_q(0)$ the matrix with $b_{if}(0)$, $i = 1, 2, \ldots, q$, $f = 1, 2, \ldots, q$, in the (i, f) entry, and $\mathbf{H}$ be the lower triangular matrix with positive diagonal entries such that $\mathbf{HH'} = \mathbf{b}_q(0)\mathbf{b}_q(0)'$. 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 $\mathbf{u}_t^* = \mathbf{H}'\mathbf{B}_q(0)'\mathbf{u}_t$.

Model II (with static factor representation)

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

Here $\mathbf{F}_t = (F_{1t} \ \ldots \ F_{rt})'$ and $\mathbf{u}_t = (u_{1t} \ \ldots \ u_{qt})'$, $\mathbf{D}$ is $r \times r$ and $\mathbf{K}$ is $r \times q$. Again, u_{jt}, $j = 1, \ldots, q$ and ξ_{it}, $i = 1, \ldots, n$, $t = 1, \ldots, T$ are Gaussian, unit variance, independent white noises. Moreover, λ_{hi}, $h = 1, \ldots, r$, $i = 1, \ldots, n$ and 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 $\mathbf{b}_i^*(L)(L)$ 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 i, f entry of $\mathbf{b}^*_i(L)$. Our target is estimation of $b^*_{if,k}$, $i = 1, \dots, n$, $f = 1, \dots, q$, $k = 0, \dots, K$ and u^*_{ft}, $f = 1, \dots, q$, $t = 1, \dots, T$, as well as forecasting of x_{iT+1}, $i = 1, \dots, n$.

The structural impulse response functions and the structural shocks are estimated by the FHLZ method and the FGLR method. For FHLZ, the number of lags for each $q+1$-dimensional VAR is determined by the BIC criterion. The contemporaneous and lagged covariances of the common components needed to compute the VAR coefficients are estimated by the FHLR (2000) dynamic principal component method, with lag window $T^{\frac{2}{3}}$. 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. The number of structural shocks is assumed known: such condition is obviously needed when estimating the structural shocks and impulse response functions. Identification is obtained by imposing the Cholesky scheme above.

Regarding prediction, FHLZ forecasts are computed by filtering the estimated shocks with the estimated impulse response functions.[2] The number of structural shocks is no longer assumed known. Rather, it is estimated by using Hallin and Liska's (2007) method.[3] SW forecasts are obtained by regressing the x's onto either the lagged principal components and the lagged x's (just the first lag), or the lagged principal components alone. The former method correspond to the original Stock and Watson's (2002a) method; the latter is motivated by the fact that in both of the models

[2] When averaging over different re-orderings of the variables, we compute the average after filtering, rather than applying the average of the filters to the average of the shocks.

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

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, shock f, lag k: the estimation error of the impulse response functions is measured by

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

The truncation lag K is set to 60. Similarly, denoting with $\hat{u}^*_{ft}$ the estimate of u^*_{ft}, the estimation error of 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 forecasts obtained by filtering the true structural shocks with the true structural impulse response functions, i.e. $x^P_{iT+1} = \sum_{f=1}^{q}\sum_{k=1}^{T} b^*_{if,k}u^*_{fT+1-k}$. Again, we normalize by dividing by the sum of the squared targets:

$$\frac{\sum_{i=1}^{n}\left(\hat{x}_{iT+1}-x^P_{iT+1}\right)^2}{\sum_{i=1}^{n}(x^P_{iT+1})^2}.$$

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 with $n = 120$ and $T = 240$, but different configurations of q and r, i.e. $r = 4, 6, 8, 12$ and $q = 2, 4, 6$, $r > q$.[4] For each couple (n, T), Model I, and (r, q), Model II, we generate 500 data sets and compute the average MSE.

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

4.3 Results

Table 1 reports results for the impulse response functions and structural shocks, Model I. PZ: Recall that the asymptotic properties of the estimates are independent of the particular partition adopted to constuct the $(q+1)$-dimensional blocks. To mitigate a finite-sample effect that could arise, we average the results across several partitions. It turns out that the results quickly stabilize by considering a limited number of partitions. We present the results corresponding to the average across a number of partitions equal to 30. The upper panel reports results for method FHLZ without averaging over different partitions of the variables in the data set, whereas the central panel reports results for FHLZ when averaging.

The static factor estimates (lower panel), despite being theoretically inconsistent, approach the target as n and T get larger. The estimation error of both response functions and shocks reduces by over 70% when passing from the smallest to the largest panel. Indeed, the number of estimated static factors increases with n and T, so that the static model better approximates the theoretical model.[5] As expected, for large n and T the performance is comparable to FHLZ without averaging, but for small samples the error is larger, particularly for impulse response functions. FHLZ with averaging dominates the static method for all n-T configurations. For small samples, the error of the estimated impulse response functions is about 30-35% smaller.

Forecast results (Table 2) are very similar. Not surprisingly, the SW method (central and lower panels) performs better when lagged x's 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 x's, i.e. the χ's, rather than the x' themselves. FHLZ forecasts (with averaging) over-performs

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

Table 1: Model I, estimated impulse response functions and structural shocks. Average normalized MSE across 500 generated data sets with different size. For the static method, the number of static factors is determined by Bai and Ng's IC_{p2} criterion.

	Impulse response functions				*Structural shocks*			
T	60	120	240	480	60	120	240	480
n	*method FHLZ, no averaging*							
30	0.456	0.321	0.235	0.167	0.465	0.362	0.293	0.247
60	0.424	0.301	0.223	0.168	0.366	0.281	0.225	0.191
120	0.426	0.294	0.211	0.148	0.345	0.241	0.181	0.139
240	0.415	0.299	0.232	0.150	0.314	0.232	0.189	0.126
n	*method FHLZ, averaging*							
30	0.369	0.259	0.191	0.122	0.387	0.308	0.261	0.209
60	0.338	0.250	0.183	0.122	0.306	0.247	0.200	0.155
120	0.332	0.242	0.178	0.125	0.271	0.209	0.168	0.131
240	0.337	0.245	0.181	0.126	0.264	0.198	0.156	0.117
n	*static factor method (FGLR)*							
30	0.542	0.445	0.375	0.328	0.456	0.372	0.301	0.256
60	0.511	0.421	0.313	0.250	0.374	0.300	0.215	0.178
120	0.507	0.396	0.246	0.153	0.353	0.272	0.199	0.133
240	0.493	0.324	0.233	0.155	0.341	0.255	0.197	0.123

SW for all n-T configurations, with an improvement ranging from 30 to 40%.[6] Observe that here we no longer impose the correct q, but estimate it with Hallin and Likska's

[6]FHLZ without averaging, not reported here, performs better than SW , consistently with results in Table 1.

Table 2: Model I, 1-step-ahead forecasts. Normalized sum of square deviation from the population forecasts: average across 500 generated data sets with different size. For the dynamic method, the number dynamic factors is determined by Hallin and Liska's log criterion. For the static method, the number of static factors is determined by Bai and Ng's IC_{p2} criterion.

	$T = 60$	$T = 120$	$T = 240$	$T = 480$
method FHLZ				
$n = 30$	0.575	0.424	0.378	0.297
$n = 60$	0.494	0.360	0.285	0.204
$n = 120$	0.449	0.321	0.247	0.161
$n = 240$	0.430	0.301	0.222	0.141
static factor method (SW), with lagged x's				
$n = 30$	1.060	0.699	0.650	0.523
$n = 60$	0.932	0.648	0.551	0.403
$n = 120$	0.867	0.592	0.430	0.266
$n = 240$	0.871	0.545	0.363	0.226
static factor method (SW), no lagged x's				
$n = 30$	0.898	0.693	0.667	0.572
$n = 60$	0.790	0.640	0.549	0.430
$n = 120$	0.734	0.556	0.400	0.253
$n = 240$	0.716	0.460	0.322	0.211

(2007) criterion, so that both forecasts in the upper an central panels are feasible.

Table 3 shows results for Model II, estimation of the structural impulse response functions and shocks. Here both FHLZ and FGLR are consistent. FHLZ (with averaging, upper panel) outperforms FGLR for almost all r-q configurations. In the present model, Bai and Ng's criterion underestimates the number of factors.[7] Hence, we computed the (unfeasible) FGLR estimation obtained by imposing the correct r (lower panel), to see whether the above result can be ascribed to underestimation of r. For $r = 4$, FGLR performs remarkably better (and better than FHLZ) when imposing the correct number of factors; but this is no longer true for larger r, particularly when estimation of the structural shocks is concerned. For instance, with $r = 12$, underestimation of r improves estimation of the shocks rather than worsening it.

Forecasts errors, reported in Table 4, confirm the above result. FHLZ performs better than SW for most r-q configurations, with the exception of $q = 2$, $r = 8, 12$, for which results are similar. The forecast error is considerably smaller for $q = 6$ (about 25%).

5 Conclusions

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

The rational spectral density assumption has also the important consequences that $\boldsymbol{\chi}_t$ has a finite autoregressive representation and that the dynamic factor model can be transformed into the static model $\mathbf{z}_t = \mathbf{R}\mathbf{v}_t + \boldsymbol{\phi}_t$, 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 es-

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

Table 3: Model II, estimated impulse response functions and structural shocks. Average normalized MSE across 500 generated data sets with different configurations of static and dynamic factors. For the static method, the number of static factors is determined by Bai and Ng's IC_{p2} criterion.

	Impulse response functions				*Structural shocks*			
r	4	6	8	12	4	6	8	12
q	*method FHLZ*							
2	0.126	0.115	0.125	0.108	0.170	0.117	0.120	0.100
4		0.119	0.106	0.083		0.245	0.148	0.091
6			0.097	0.102			0.229	0.152
	method FGLR, r determined with IC_{p2}							
2	0.138	0.128	0.138	0.126	0.208	0.136	0.137	0.129
4		0.107	0.117	0.091		0.237	0.165	0.109
6			0.092	0.114			0.220	0.167
	method FGLR, r assumed known							
2	0.100	0.119	0.125	0.106	0.159	0.146	0.157	0.149
4		0.103	0.105	0.090		0.217	0.165	0.125
6			0.090	0.114			0.215	0.179

timator of the spectral density of the x's. 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 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

Table 4: Model II, 1-step-ahead forecasts. Normalized sum of square deviation from the population forecasts: average across 500 generated data sets with different configurations of static and dynamic factors. For the dynamic method, the number dynamic factors is determined by Hallin and Liska's log criterion. For the static method, the number of static factors is determined by Bai and Ng's IC_{p2} criterion.

	$r=4$	$r=6$	$r=8$	$r=12$
method FHLZ				
$q=2$	0.436	0.423	0.379	0.387
$q=4$		0.358	0.378	0.345
$q=6$			0.365	0.386
static factor method (SW), no lagged x's				
$q=2$	0.530	0.534	0.374	0.375
$q=4$		0.372	0.426	0.390
$q=6$			0.472	0.522

dynamic factor model has a static representation of the form (1.3). Section 4 provides important empirical support to the richer dynamic structure of unrestricted GDFM's.

References

[1] Alessi, L., Barigozzi, M. and Capasso, M. (2010) Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters* **80**, 1806-1813.

[2] Amengual, D. and M. Watson, (2007) Consistent estimation of the number of dynamic factors in a large N and T panel, *Journal of Business and Economic Statistics,* **25**, 91-96.

[3] Anderson, B. and M. Deistler (2008a). Properties of zero-free transfer function matrices, *SICE Journal of Control, Measurement and System Integration* **1**, 1-9.

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

[5] Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**, 191-221.

[6] Bai, J. and S. Ng, (2007). Determining the number of primitive shocks in factor models, *Journal of Business and Economic Statistics,* **25,** 52-60.

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

[8] Forni, M., M. Hallin, M. Lippi and L. Reichlin (2000). The generalized dynamic factor model: identification and estimation, *The Review of Economics and Statistics* **82**, 540- 554.

[9] Forni, M., M. Hallin, M. Lippi and L. Reichlin (2004). The generalized dynamic factor model : consistency and rates, *Journal of Econometrics* **119**, 231-255.

[10] Forni, M., M. Hallin, M. Lippi and L. Reichlin (2005). The generalized factor model: one-sided estimation and forecasting, *Journal of the American Statistical Association* **100**, 830-40.

[11] Forni, M., M. Hallin, M. Lippi and P.Zaffaroni (2014). Dynamic factor models with infinite-dimensional factor space: one-sided representations, *Journal of Econometrics* forthcoming.

[12] Forni, M., D. Giannone, M. Lippi and L. Reichlin (2009). Opening the black box: structural factor models with large cross-sections, *Econometric Theory* **25,** 1319-1347.

[13] Forni, M. and M. Lippi (2001). The generalized dynamic factor model: representation theory, *Econometric Theory* **17**, 1113-1341.

[14] Franklin, J.N. (2000). *Matrix Theory* , New York: Dover Publications.

[15] Hallin M. and R. Liška (2007). Determining the number of factors in the general dynamic factor model, *Journal of the American Statistical Association* **102**, 603-617.

[16] Hannan, E.J. (1970). *Multiple Time Series,* New York: Wiley.

[17] Lancaster, P. and M. Tismenetsky (1985). *The Theory of Matrices,* second edition, New York: Academic Press.

[18] Liu, W.D. and Wu, W.B. (2010). Asymptotics of spectral density estimates, *Econometric Theory*, **26** , 1218-1245.

[19] Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models, *Econometrica* **77**, 1447-1479.

[20] Shao, W. and Wu, W.B. (2007). Asymptotic spectral theory for nonlinear time series, *Annals of Statistics* **35**, 1773-1801.

[21] Stock, J.H. and M.W. Watson (2002a). Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics* **20**, 147-162.

[22] Stock, J.H. and M.W. Watson (2002b). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* **97**, 1167-1179.

[23] Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences USA*, **102**, 14150-14154.

[24] Wu, W.B. and Zaffaroni, P. (2014). On Uniform Moments Convergence of Spectral Density Estimates, Preprint Imperial College London and University of Chicago.

Appendix

A Proof of Proposition 6

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

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

The result follows by deploying Theorem 5 of Wu and Zaffaroni (2014) with $\nu^*=1$ to the first term on the right hand side. In fact the second term on the right hand side, the squared bias, turns out to be of smaller order since, by the smoothness of the $\sigma^x_{ij}(\theta)$, then by standard arguments (see for instance Theorem 4.10 of Hannah (1970)) $\max_{|h|\le B_T}\left|E\hat{\sigma}^x_{ij}(\theta^*_h)-\sigma^x_{ij}(\theta^*_h)\right|^2=O(B_T^{-2\kappa})=O(T^{-2\delta\kappa})$. This term goes to zero faster than $O(B_T log B_T/T)$ whenever $1<\delta(2\kappa+1)$. Q.E.D.

B Proof of Proposition 7

The proof below closely follows Forni et al. (2009). Denote by $\mu_j(\mathbf{A})$, $j=1,2,\ldots,s$, the (real) eigenvalues, in decreasing order, of a complex $s\times s$ 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 of the following inequality: 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 is an obvious consequence of the 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. Let $a_1 < a_2 < \cdots < a_M$ be integers, and put $\mathbf{M} = \{a_1, a_2, \cdots, a_M\}$. Denote by $\mathcal{S}_\mathbf{M}$ the $n \times M$ matrix with 1 in entries (a_j, j) and zero elsewhere, 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. Moreover, 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. Lastly, all lemmas in this Appendix hold, and are proved under Assumptions 1 through 10.

Lemma 1 *As* $T \to \infty$ *and* $n \to \infty$*,*

(i) $\max_{|h|\le B_T} n^{-1}\|\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^x\| = O_\mathrm{P}(\rho_T^{-1/2})$*;*

(ii) given $\mathbf{M}$*,* $\max_{|h|\le B_T} n^{-1/2}\|\mathcal{S}'_\mathbf{M}(\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^x)\| = O_\mathrm{P}(\rho_T^{-1/2})$*;*

(iii) $\max_{|h|\le B_T} n^{-1}\|\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi\| = O_\mathrm{P}(\max(n^{-1}, \rho_T^{-1/2}))$*;*

(iv) given $\mathbf{M}$*,* $\max_{|h|\le B_T} n^{-1/2}\|\mathcal{S}'_\mathbf{M}(\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi)\| = O_\mathrm{P}(\max(n^{-1/2}, \rho_T^{-1/2})) = O_\mathrm{P}(\zeta_{nT})$.

Proof. Since

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

Because

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

statement (i) follows from (3.4), see Proposition 6, and Markov inequality. Because

$$\mathrm{trace}\big(\mathcal{S}'_\mathbf{M}(\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^x)(\tilde{\hat{\mathbf{\Sigma}}}^x - \tilde{\mathbf{\Sigma}}^x)\mathcal{S}_\mathbf{M}\big) = \sum_{i\in\mathbf{M}} \sum_{j=1}^n |\hat{\sigma}^x_{ij} - \sigma^x_{ij}|^2,$$

statement (ii) follows in the same way. As regards (iii), $\mathbf{\Sigma}^x = \mathbf{\Sigma}^\chi + \mathbf{\Sigma}^\xi$ implies $\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi = \hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^x + \mathbf{\Sigma}^\xi$, 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 $\|\mathbf{\Sigma}^\xi\| = \lambda_1^\xi$ is bounded. Statement (iv) is obtained in a similar way, using (ii) instead of (i). $\square$

Lemma 2 *As $T \to \infty$ and $n \to \infty$,*
(i) $\max_{|h|\le B_T} n^{-1} \left|\hat{\lambda}_f^x - \lambda_f^\chi\right| = O_{\mathrm{P}}(\max(n^{-1}, \rho_T^{-1/2}))$ *for $f = 1, 2, \dots, q$;*
(ii) Letting

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

$\max_{|h|\le B_T} n^{-1}\|\mathbf{\Lambda}^\chi\|$ *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* $O_{\mathrm{P}}(1)$*;*

Proof. Setting $\mathbf{A}_1 = \mathbf{\Sigma}^\chi$ and $\mathbf{A}_2 = \hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi$, (B.2) yields $|\hat{\lambda}_f^x - \lambda_f^\chi| \le \|\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi\|$; 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, follow from statement (i). $\square$

Lemma 3 *As $T \to \infty$ and $n \to \infty$,*
(i) $\max_{|h|\le B_T} n^{-1}\|\tilde{\mathbf{P}}^\chi\hat{\mathbf{P}}^x\hat{\mathbf{\Lambda}}^x - \mathbf{\Lambda}^\chi\tilde{\mathbf{P}}^\chi\hat{\mathbf{P}}^x\| = O_{\mathrm{P}}(\max(n^{-1}, \rho_T^{-1/2}))$*;*
(ii) $\max_{|h|\le B_T} \|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^\chi\tilde{\mathbf{P}}^\chi\hat{\mathbf{P}}^x - \mathbf{I}_q\| = O_{\mathrm{P}}(\max(n^{-1}, \rho_T^{-1/2}))$*;*
(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* $\max_{|h|\le B_T} \|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^\chi - \hat{\mathbf{W}}_q\| = O_{\mathrm{P}}(\max(n^{-1}, \rho_T^{-1/2}))$.

Proof. Using inequality (B.1) and $\|\tilde{\mathbf{P}}^\chi\| = \|\hat{\mathbf{P}}^{\mathbf{x}}\| = 1$, $\|\tilde{\mathbf{P}}^\chi\hat{\mathbf{P}}^x\hat{\mathbf{\Lambda}}^x - \mathbf{\Lambda}^\chi\tilde{\mathbf{P}}^\chi\hat{\mathbf{P}}^x\| = \|\tilde{\mathbf{P}}^\chi(\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi)\hat{\mathbf{P}}^x\| \le \|\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi\|$. Statement (i) thus follows from Lemma 1 (iii). Turning to (ii), set

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

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

$$\left\|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^\chi\tilde{\mathbf{P}}^\chi\hat{\mathbf{P}}^x - \mathbf{I}_q\right\| \leq \|\mathbf{a}-\mathbf{b}\| + \|\mathbf{b}-\mathbf{c}\| + \|\mathbf{c}-\mathbf{d}\| + \|\mathbf{d}- \ \mathbf{f}\|. \qquad \text{(B.3)}$$

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

As regards (iii), note that, from statement (i), $n^{-1}\tilde{\hat{\mathbf{P}}}^x_h\mathbf{P}^\chi_k(\lambda^\chi_k - \hat{\lambda}^x_h) = O_\mathrm{P}(\max(n^{-1},{\rho_T}^{-1/2}))$. Assumption 3 (asymptotic separation of the eigenvalues $\lambda^\chi_f(\theta)$) implies that, for $h \neq k$, $\tilde{\hat{\mathbf{P}}}^x_h\mathbf{P}^\chi_k = O_\mathrm{P}(\max(n^{-1},{\rho_T}^{-1/2}))$. Moreover, from statement (ii), $\sum_{f=1}^q|\tilde{\hat{\mathbf{P}}}^x_h\mathbf{P}^\chi_f|^2 - 1 = O_\mathrm{P}(\max(n^{-1},{\rho_T}^{-1/2}))$. Therefore

$$|\tilde{\hat{\mathbf{P}}}^x_h\mathbf{P}^\chi_h|^2 - 1 = (|\tilde{\hat{\mathbf{P}}}^x_h\mathbf{P}^\chi_h| - 1)(|\tilde{\mathbf{P}}^\chi_h\hat{\mathbf{P}}^x_h| + 1) = O_\mathrm{P}(\max(n^{-1},{\rho_T}^{-1/2})).$$

The conclusion follows. □

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

Lemma 4 *Given* $\mathbf{M}$*, as* $T \to \infty$ *and* $n \to \infty$*,*

$$\max_{|h|\leq B_T}\|\mathcal{S}'_\mathbf{M}(\mathbf{P}^\chi(\mathbf{\Lambda}^\chi)^{1/2}\hat{\mathbf{W}}_q - \hat{\mathbf{P}}^x(\hat{\mathbf{\Lambda}}^x)^{1/2})\| = O_\mathrm{P}(\zeta_{nT}). \qquad \text{(B.4)}$$

Proof. We have

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

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

$$\|n^{1/2}\mathcal{S}'_\mathbf{M}\mathbf{P}^\chi\hat{\mathbf{W}}_q - n^{1/2}\mathcal{S}'_\mathbf{M}\hat{\mathbf{P}}^x\| = O_\mathrm{P}(\max(n^{-1/2},{\rho_T}^{-1/2})).$$

Firstly, we show that, uniformly in θ,

$$\|n^{1/2}\mathcal{S}'_{\mathbf{M}}\mathbf{P}^{\chi}\| = O(1). \tag{B.5}$$

Assumption 2 implies that $\sigma^{\chi}_{ii} = \sum_{f=1}^{q} \lambda^{\chi}_f |p^{\chi}_{if}|^2 = O(1)$, uniformly in θ. As all the terms in the sum are positive, $\lambda^{\chi}_f |p^{\chi}_{if}|^2 = (\lambda^{\chi}_f/n)n|p^{\chi}_{if}|^2$ also is $O(1)$, uniformly in θ. Assumption 3 implies that λ^{χ}_f/n is bounded away from zero uniformly in θ, so that $n|p^{\chi}_{if}|^2$ must be $O(1)$, uniformly in θ. Hence, the eigenvalues of $n\mathcal{S}'_{\mathbf{M}}\mathbf{P}^{\chi}\tilde{\mathbf{P}}^{\chi}\mathcal{S}_{\mathbf{M}}$ are $O(1)$ uniformly in θ; (B.5) follows. Next, define

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

$$\mathbf{i} = n^{1/2}\mathcal{S}'_{\mathbf{M}}\mathbf{P}^{\chi}[(\mathbf{\Lambda}^{\chi}/n)\tilde{\mathbf{P}}^{\chi}\hat{\mathbf{P}}^{x}](\hat{\mathbf{\Lambda}}^{x}/n)^{-1} = [n^{-1/2}\mathcal{S}'_{\mathbf{M}}\mathbf{\Sigma}^{\chi}]\hat{\mathbf{P}}^{x}(\hat{\mathbf{\Lambda}}^{x}/n)^{-1},$$

and

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

Using (B.5), Lemma 3 and Lemma 1 (iv), we obtain that $\|\mathbf{g}-\mathbf{h}\|$ and $\|\mathbf{h}-\mathbf{i}\|$ are $O_{\mathrm{P}}(\max(n^{-1}, {\rho_T}^{-1/2}))$, while $\|\mathbf{i}-\mathbf{j}\|$ is $O_{\mathrm{P}}(\max(n^{-1/2}, {\rho_T}^{-1/2}))$; the result follows. □

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

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

In the same way, the result of Lemma 3(iii) will be referred to as

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

Proposition 7 now follows from the fact that $\hat{\mathbf{\Sigma}}^{\chi} = \hat{\mathbf{P}}^{x}\hat{\mathbf{\Lambda}}^{x}\tilde{\hat{\mathbf{P}}}^{x}$ and $\mathbf{\Sigma}^{\chi} = \mathbf{P}^{\chi}\mathbf{\Lambda}^{\chi}\tilde{\mathbf{P}}^{\chi}$.

C Proof of Proposition 9

Firstly, note that, as the last term in (3.6) contains

$$\frac{\pi G}{B_T}\sum_{s=-B_T+1}^{B_T}\max_{\alpha_s\le\theta\le\beta_s}|e^{\imath\ell\theta_s}-e^{\imath\ell\theta}|,$$

convergence in (3.7) is not uniform with respect to ℓ. However, estimation of the matrices $\mathbf{B}_k^\chi$ and $\mathbf{C}_{jk}^\chi$ only requires the covariances $\hat{\gamma}_{ij,\ell}$ with $\ell \le S$, where S is finite. Therefore, Proposition 8 implies that $\|\hat{\mathbf{B}}_k^\chi - \mathbf{B}_k^\chi\|$ and $\|\hat{\mathbf{C}}_{jk}^\chi - \mathbf{C}_{jk}^\chi\|$ are $O_{\mathrm{P}}(\max(n^{-1/2}, \rho_T^{-1/2}))$. From (2.16), applying (B.1),

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

By Assumption 2, $\|\mathbf{B}_k^\chi\| \le W$ for some constant $W > 0$, so that $\|\hat{\mathbf{B}}_k^\chi\|$ is bounded in probability. By Assumptions 2 and 7, $\|(\mathbf{C}_{kk}^\chi)^{-1}\| \le W_1$ for some $W_1 > 0$. Observing that the entries of $(\mathbf{C}_{kk}^\chi)^{-1}$ are rational functions of the entries of $\mathbf{C}_{kk}^\chi$, and that $\det(\mathbf{C}_{kk}^\chi) > 0$ by Assumption 7, Proposition 8 implies that $\|(\hat{\mathbf{C}}_{kk}^\chi)^{-1} - (\mathbf{C}_{kk}^\chi)^{-1}\|$ is $O_{\mathrm{P}}(\max(n^{-1/2}, \rho_T^{-1/2}))$. Thus $\|\hat{\mathbf{A}}^{[k]} - \mathbf{A}^{[k]}\|$ is $O_{\mathrm{P}}(\max(n^{-1/2}, \rho_T^{-1/2}))$. As regards $\hat{\mathbf{\Gamma}}_{jk}^\psi$, using (B.1),

$$\begin{aligned}\|\hat{\mathbf{A}}^{[j]}\hat{\mathbf{C}}_{jk}^\chi\hat{\mathbf{A}}^{[k]} - \mathbf{A}^{[j]}\mathbf{C}_{jk}^\chi\mathbf{A}^{[k]}\| \;\le\; & \|\hat{\mathbf{A}}^{[j]}\hat{\mathbf{C}}_{jk}^\chi\|\|\hat{\mathbf{A}}^{[k]} - \mathbf{A}^{[k]}\| + \|\hat{\mathbf{A}}^{[j]}\|\|\hat{\mathbf{C}}_{jk}^\chi - \mathbf{C}_{jk}^\chi\|\|\mathbf{A}^{[k]}\| \\ & + \|\hat{\mathbf{A}}^{[j]} - \mathbf{A}^{[j]}\|\|\mathbf{C}_{jk}^\chi\mathbf{A}^{[k]}\|.\end{aligned}$$

The conclusion follows. Q.E.D.

D Proof of Proposition 10

Consider the static model $\mathbf{Z}_{nt} = \mathcal{R}\mathbf{v}_t + \mathbf{\Phi}_{nt}$. 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 $\mathbf{\Gamma}^z$ 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 $\hat{\mathbf{\Gamma}}^z$, 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 $\mathbf{\Gamma}^z - \hat{\mathbf{\Gamma}}^z$; Lemmas 5 through 10 are but a preparation for that crucial result. All lemmas in this Appendix hold, and are proved under Assumptions 1 through 10.

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

(i) $| p^{\chi}_{if} | = O(n^{-1/2})$ and $| \hat{p}^{x}_{if} | = O_{\mathrm{P}}(n^{-1/2})$, uniformly in θ;

(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 $O_{\mathrm{P}}(n^{-d/2})$, respectively, uniformly in θ.

Proof. The first part of (i) already has been taken care of in the proof of Lemma 4. Lemma 4 and Assumption 2 jointly imply that $\hat{\sigma}^x_{ii} = \sum_{f=1}^q \hat{\lambda}^x_f |\hat{p}^x_{if}|^2 = O_{\mathrm{P}}(1)$, uniformly in θ. 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 $O_{\mathrm{P}}(1)$ as well, uniformly in θ. Lemma 2 (i) and Assumption 3 imply that $\hat{\lambda}^x_f/n$ is $O_{\mathrm{P}}(1)$ and bounded away from zero in probability uniformly in θ. The conclusion follows.

Statement (ii) is proved by induction. First consider $\mathbf{P}^{\chi}_f$. When $d = 1$, $n^{-1}\sum_{i=1}^n |p^{\chi}_{if}|$ is bounded by $(n^{-1}\sum_{i=1}^n |p^{\chi}_{if}|^2)^{1/2}$, which is $O(n^{-1/2})$. Assume now that the result holds for $d-1$, $d \geq 2$. Summing by parts and using part (i) of this Lemma,

$$\begin{aligned} n^{-1}\sum_{i=1}^n | p^{\chi}_{if} |^d &= n^{-1}\sum_{i=1}^n | p^{\chi}_{if} |^{d-1} | p^{\chi}_{if} | \\ &= n^{-1} | p^{\chi}_{nf} | \sum_{i=1}^n | p^{\chi}_{if} |^{d-1} - n^{-1}\sum_{i=1}^{n-1}\sum_{s=1}^{i} | p^{\chi}_{sf} |^{d-1} (| p^{\chi}_{i+1,f} | - | p^{\chi}_{if} |) \\ &\leq |p^{\chi}_{nf}| n^{-1}\sum_{i=1}^n | p^{\chi}_{if} |^{d-1} = O(n^{-1/2}\, n^{-(d-1)/2}) = O\left(n^{-d/2}\right), \end{aligned}$$

the inequality holding because without loss of generality (reordering) we can assume $| p^{\chi}_{i+1,f} | \geq | p^{\chi}_{if} |$. The same argument applies to $\hat{\mathbf{P}}^x_f$. □

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

$$\max_{|h| \leq B_T} \left\| \mathbf{P}^{\chi} (\mathbf{\Lambda}^{\chi})^{1/2} \hat{\mathbf{W}}_q - \hat{\mathbf{P}}^{x} \left(\hat{\mathbf{\Lambda}}^{x}\right)^{1/2} \right\| = O_{\mathrm{P}}(n^{1/2} \max(n^{-1}, \rho_T^{-1/2})). \tag{D.1}$$

Proof. The left-hand side of (D.1) equals the left-hand side of (B.4) when $\mathcal{S}_{\mathbf{M}}$ 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 $O_{\mathrm{P}}(n^{-1/2} \max(n^{-1}, \rho_T^{-1/2}))$. As for $\|\mathbf{i} - \mathbf{j}\|$, the conclusion follows from Lemma 1 (iii). □

Lemma 7 *For $f = 1, \dots, q$, as $T \to \infty$ and $n \to \infty$,*
(i) $|p^{\chi}_{if} - \hat{p}^{x}_{if}| = O_{\mathrm{P}}(n^{-1/2} \max(n^{-1/2}, \rho_T^{-1/2}))$, uniformly in θ;
(ii) $n^{-1} \sum_{i=1}^{n} |p^{\chi}_{if} - \hat{p}^{x}_{if}| = O_{\mathrm{P}}(n^{-1/2} \max(n^{-1}, \rho_T^{-1/2}))$, uniformly in θ.

Proof. Starting with (i), by (B.6), $p^{\chi}_{if}(\lambda^{\chi}_f)^{1/2} - \hat{p}^{x}(\hat{\lambda}^{x}_f)^{1/2} = O_{\mathrm{P}}(\max(n^{-1/2}, \rho_T^{-1/2}))$. Now,

$$p^{\chi}_{if}(\lambda^{\chi}_f)^{1/2} - \hat{p}^{x}(\hat{\lambda}^{x}_f)^{1/2} = p^{\chi}_{if}\left((\lambda^{\chi}_f)^{1/2} - (\hat{\lambda}^{x}_f)^{1/2}\right) + (\hat{\lambda}^{x}_f)^{1/2}\left(p^{\chi}_{if} - \hat{p}^{x}_{if}\right). \tag{D.2}$$

For the first term on the right-hand side of (D.2),

$$p^{\chi}_{if}\left((\lambda^{\chi}_f)^{1/2} - (\hat{\lambda}^{x}_f)^{1/2}\right) = n^{1/2} p^{\chi}_{if} \frac{(\lambda^{\chi}_{if} - \hat{\lambda}^{x}_{if})/n}{((\lambda^{\chi}_f)^{1/2} + (\hat{\lambda}^{x}_f)^{1/2})/n^{1/2}} = O_{\mathrm{P}}(\max(n^{-1}, \rho_T^{-1/2})),$$

by Lemma 2(i), Assumption 3 and Lemma 7(i) above. Thus, $(\hat{\lambda}^{x}_f)^{1/2}\left(p^{\chi}_{if} - \hat{p}^{x}_{if}\right)$ is $O_{\mathrm{P}}(\max(n^{-1/2}, \rho_T^{1/2}))$. By Assumption 3, $n^{-1/2}(\hat{\lambda}^{x}_f)^{1/2}$ is bounded away from zero. The conclusion follows.

Regarding (ii), taking modulus and summing over $i = 1, ..., n$ in (D.2) yields

$$n^{-1/2}(\hat{\lambda}^{x}_f)^{1/2} \sum_{i=1}^{n} |p^{\chi}_{if} - \hat{p}^{x}_{if}| \leq n^{-1/2} \sum_{i=1}^{n} |p^{\chi}_{if}(\lambda^{\chi}_f)^{1/2} - \hat{p}^{x}(\hat{\lambda}^{x}_f)^{1/2}| + n^{-1/2}|(\lambda^{\chi}_f)^{1/2} - (\hat{\lambda}^{x}_f)^{1/2}| \sum_{i=1}^{n} |p^{\chi}_{if}|.$$

Regarding the first term on the right-hand side, by Jensen's inequality and Lemma 6:

$$\sum_{i=1}^{n}\left|p_{if}^{\chi}(\lambda_f^{\chi})^{1/2}-\hat{p}^{x}(\hat{\lambda}_f^{x})^{1/2}\right|\leq n^{1/2}\Big(\sum_{i=1}^{n}\left|p_{if}^{\chi}(\lambda_f^{\chi})^{1/2}-\hat{p}^{x}(\hat{\lambda}_f^{x})^{1/2}\right|^2\Big)^{1/2}=O_{\mathrm{P}}(n\max(n^{-1},\rho_T^{-1/2})).$$

Lemma 2(i)-(ii) and Lemma 5(ii) provide bounds for the second term. □

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

$$n^{-1}\sum_{i=1}^{n}\mid p_{if}^{\chi}-\hat{p}_{if}^{x}\mid^{d}=O_{\mathrm{P}}((n^{-1}\max(n^{-1},\rho_T^{-1}))^{d/2}),\tag{D.3}$$

uniformly in θ.

Proof. By induction. Lemma 7(ii) implies that $n^{-1}\sum_{i=1}^{n}\mid \hat{p}_{if}^{\chi}-p_{ij}^{\chi}\mid$ is $O_{\mathrm{P}}((n^{-1}\max(n^{-1},\rho_T^{-1}))^{1/2})$. In fact, to avoid unnecessary complications, we consider here a slightly looser bound than the one provided by Lemma 7. Assume now that $d\geq 2$ and that the result holds for $d-1$. Using summation by parts,

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n}\mid p_{if}^{\chi}-\hat{p}_{if}^{x}\mid^{d}&=n^{-1}\sum_{i=1}^{n}\mid p_{if}^{\chi}-\hat{p}_{if}^{x}\mid^{d-1}\mid p_{if}^{\chi}-\hat{p}_{if}^{x}\mid\\
&=\mid p_{nf}^{\chi}-\hat{p}_{nf}^{x}\mid n^{-1}\sum_{i=1}^{n}\mid p_{if}^{\chi}-\hat{p}_{if}^{x}\mid^{d-1}\\
&\qquad-\frac{1}{n}\sum_{i=1}^{n-1}\sum_{k=1}^{i}\mid p_{kf}^{\chi}-\hat{p}_{kf}^{x}\mid^{d-1}(\mid p_{i+1,f}^{\chi}-\hat{p}_{i+1,f}^{x}\mid-\mid p_{if}^{\chi}-\hat{p}_{if}^{x}\mid)\\
&\leq\mid p_{nf}^{\chi}-\hat{p}_{nf}^{x}\mid\frac{1}{n}\sum_{i=1}^{n}\mid p_{if}^{\chi}-\hat{p}_{if}^{x}\mid^{d-1}=\mid p_{nf}^{\chi}-\hat{p}_{nf}^{x}\mid O_{\mathrm{P}}\big((n^{-1}\max(n^{-1},\rho_T^{-1}))^{(d-1)/2}\big)
\end{aligned}$$

since without loss of generality we can assume $\mid p_{i+1,f}^{\chi}-\hat{p}_{i+1,f}^{x}\mid\geq\mid p_{if}^{\chi}-\hat{p}_{if}^{x}\mid$. The result follows from Lemma 7(i). □

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

(i) $n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mid\hat{\sigma}_{ij}^{\chi}(\theta)-\sigma_{ij}^{\chi}(\theta)|^{d}=O_{\mathrm{P}}((\max(n^{-1},\rho_T^{-1}))^{d/2})$;

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

(iii) $n^{-1}\sum_{i=1}^{n} \mid \hat{\sigma}^{\chi}_{ii}(\theta) - \sigma^{\chi}_{ii}(\theta) \mid^{d} = O_{\mathrm{P}}((\max(n^{-1}, \rho_T^{-1}))^{d/2})$.

Proof. We have

$$\begin{aligned}\hat{\sigma}^{\chi}_{ij} - \sigma^{\chi}_{ij} &= (\hat{\lambda}^x_1 - \lambda^{\chi}_1)\hat{p}^x_{i1}\bar{\hat{p}}^x_{j1} + \cdots + (\hat{\lambda}^x_q - \lambda^{\chi}_q)\hat{p}^x_{iq}\bar{\hat{p}}^x_{jq} + \lambda^{\chi}_1\hat{p}^x_{i1}(\bar{\hat{p}}^x_{j1} - \bar{p}^{\chi}_{j1}) \\ &\quad + \lambda^{\chi}_1\bar{p}^{\chi}_{j1}(\hat{p}^x_{i1} - p^{\chi}_{i1}) + \ldots + \lambda^{\chi}_q\hat{p}^x_{iq}(\bar{\hat{p}}^x_{jq} - \bar{p}^{\chi}_{jq}) + \lambda^{\chi}_q\bar{p}^{\chi}_{jq}(\hat{p}^x_{iq} - p^{\chi}_{iq}).\end{aligned}$$

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

$$\begin{aligned}& n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mid \hat{\sigma}^{\chi}_{ij} - \sigma^{\chi}_{ij} \mid^{d} \\ &\le (3q)^{d-1}\Big(\mid \lambda^{\chi}_1 - \hat{\lambda}^x_1 \mid^{d} \big(n^{-1}\sum_{i=1}^{n} \mid \hat{p}^x_{i1} \mid^{d}\big)^2 + \cdots + \mid \lambda^{\chi}_q - \hat{\lambda}^x_q \mid^{d} \big(n^{-1}\sum_{i=1}^{n} \mid \hat{p}^x_{iq} \mid^{d}\big)^2\Big) \\ &\quad + (3q)^{d-1}(\lambda^{\chi}_1)^d\Big(n^{-2}\sum_{i=1}^{n} \mid \hat{p}^x_{i1} \mid^{d} \sum_{j=1}^{n} \mid p^{\chi}_{j1} - \hat{p}^x_{j1} \mid^{d} + n^{-2}\sum_{j=1}^{n} \mid p^{\chi}_{j1} \mid^{d} \sum_{i=1}^{n} \mid p^{\chi}_{i1} - \hat{p}^x_{i1} \mid^{d}\Big) \\ &\quad + \cdots \\ &\quad + (3q)^{d-1}(\lambda^{\chi}_q)^d\Big(n^{-2}\sum_{i=1}^{n} \mid \hat{p}^x_{iq} \mid^{d} \sum_{j=1}^{n} \mid p^{\chi}_{jq} - \hat{p}^x_{jq} \mid^{d} + n^{-2}\sum_{j=1}^{n} \mid p^{\chi}_{jq} \mid^{d} \sum_{i=1}^{n} \mid p^{\chi}_{iq} - \hat{p}^x_{iq} \mid^{d}\Big) \\ &= O_{\mathrm{P}}((\max(n^{-1}, \rho_T^{-1/2}))^{d}) + O_{\mathrm{P}}((\max(n^{-1}, \rho_T^{-1}))^{d/2}) = O_{\mathrm{P}}((\max(n^{-1}, \rho_T^{-1}))^{d/2}).\end{aligned}$$

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

$$\begin{aligned}& n^{-1}\sum_{i=1}^{n} \mid \hat{\sigma}^{\chi}_{ij} - \sigma^{\chi}_{ij} \mid^{d} \\ &\le (3q)^{d-1}\big(\mid \lambda^{\chi}_1 - \hat{\lambda}^x_1|^d|\hat{p}^x_{j1}|^d\, n^{-1}\sum_{i=1}^{n} |\hat{p}^x_{i1}|^d + \cdots + \mid \lambda^{\chi}_q - \hat{\lambda}^x_q \mid^{d} |\hat{p}^x_{jq}|^d\, n^{-1}\sum_{i=1}^{n} \mid \hat{p}^x_{iq} \mid^{d}\big) \\ &\quad + (3q)^{d-1}(\lambda^{\chi}_1)^d\big(\big|p^{\chi}_{j1} - \hat{p}^x_{j1}\big|^d\, n^{-1}\sum_{i=1}^{n} \big|\hat{p}^x_{i1}\big|^d + \big|p^{\chi}_{j1}\big|^d\, n^{-1}\sum_{i=1}^{n} \mid p^{\chi}_{i1} - \hat{p}^x_{i1} \mid^{d}\big) \\ &\quad + \cdots \\ &\quad + (3q)^{d-1}(\lambda^{\chi}_q)^d\big(\big|p^{\chi}_{jq} - \hat{p}^x_{jq}\big|^d\, n^{-1}\sum_{i=1}^{n} \big|\hat{p}^x_{iq}\big|^d + \big|p^{\chi}_{jq}\big|^d\, n^{-1}\sum_{i=1}^{n} \mid p^{\chi}_{iq} - \hat{p}^x_{iq} \mid^{d}\big) \\ &= O_{\mathrm{P}}((\max(n^{-1}, \rho_T^{-1/2}))^{d}) + O_{\mathrm{P}}((\max(n^{-1}, \rho_T^{-1}))^{d/2}) = O_{\mathrm{P}}((\max(n^{-1}, \rho_T^{-1}))^{d/2}).\end{aligned}$$

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

$\square$

Lemma 10 *For $n \to \infty$ and $T \to \infty$,*

$$n^{-2}\sum_{\ell=0}^{S}\sum_{i=1}^{n}\sum_{j=1}^{n} \mid \hat{\gamma}^{\chi}_{ij,\ell} - \gamma^{\chi}_{ij,\ell}|^d \quad \text{and for any } 1 \leq j \leq n \quad n^{-1}\sum_{\ell=0}^{S}\sum_{i=1}^{n} \mid \hat{\gamma}^{\chi}_{ij,\ell} - \gamma^{\chi}_{ij,\ell} \mid^d$$

are $O_{\mathrm{P}}\big((\max(n^{-1}, \rho_T^{-1}))^{d/2}\big)$.

Proof. We have $|\hat{\gamma}^{\chi}_{ij,\ell} - \gamma^{\chi}_{ij,\ell}| \leq \mathcal{U}_{ij} + \mathcal{V}_\ell + \mathcal{W}_{ij}$, where $\mathcal{U}_{ij}$, $\mathcal{V}_\ell$ and $\mathcal{W}_{ij}$ are the terms in the last line of (3.6). Using the C_r inequality we get

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

The first term on the right-hand side is bounded using Lemma 9. Because 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^x_{ij} 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. $\square$

We are now able to state and prove the main lemma of this section. Assume, without loss of generality, that n increases by blocks of size $q+1$, so that $n = m(q+1)$.

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

$$n^{-1}\|\hat{\mathbf{\Gamma}}^z - \mathbf{\Gamma}^z\| = O_{\mathrm{P}}(\zeta_{nT}) \quad \text{and} \quad n^{-1/2}\|\mathcal{S}'_{\mathbf{M}}(\hat{\mathbf{\Gamma}}^z - \mathbf{\Gamma}^z)\| = O_{\mathrm{P}}(\zeta_{nT}), \tag{D.4}$$

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 $\mathbf{Z}_{nt}$, were the matrices $\mathbf{A}(L)$ known. We have

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

so that the lemma can be proved by showing that (D.4) holds with $\|\hat{\mathbf{\Gamma}}^z - \mathbf{\Gamma}^z\|$ replaced by any of the two terms on the right-hand side of (D.5). Consider firstly $\|\check{\mathbf{\Gamma}}^z - \mathbf{\Gamma}^z\|$. Using $\mathbf{A}(L) = \mathbf{I}_n - \mathbf{A}_1 L - \cdots - \mathbf{A}_S L^S$, where

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

$s = 1, \ldots, S$ and $\mathbf{A}_0 = \mathbf{I}_n$, we obtain

$$\|\check{\mathbf{\Gamma}}^z - \mathbf{\Gamma}^z\|^2 \leq \sum_{s=0}^{S} \sum_{r=0}^{S} \|\mathbf{A}_s \hat{\mathbf{\Gamma}}^x_{s-r} \mathbf{A}'_r - \mathbf{A}_s \mathbf{\Gamma}^x_{s-r} \mathbf{A}'_r\|^2 = \sum_{s=0}^{S} \sum_{r=0}^{S} \|\mathbf{A}_s \left(\hat{\mathbf{\Gamma}}^x_{s-r} - \mathbf{\Gamma}^x_{s-r}\right) \mathbf{A}'_r\|^2, \tag{D.6}$$

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

Turning to $\|\hat{\mathbf{\Gamma}}^z - \check{\mathbf{\Gamma}}^z\|$, since $\|\hat{\mathbf{\Gamma}}^z - \check{\mathbf{\Gamma}}^z\|^2 \leq \sum_{s=0}^{S} \sum_{r=0}^{S} \|\hat{\mathbf{A}}_s \hat{\mathbf{\Gamma}}^x_{s-r} \hat{\mathbf{A}}'_r - \mathbf{A}_s \hat{\mathbf{\Gamma}}^x_{s-r} \mathbf{A}'_r\|^2$, it is sufficient to prove that (D.4) holds with $\|\hat{\mathbf{\Gamma}}^z - \mathbf{\Gamma}^z\|$ replaced with any of the $\|\hat{\mathbf{A}}_s \hat{\mathbf{\Gamma}}^x_{s-r} \hat{\mathbf{A}}'_r - \mathbf{A}_s \hat{\mathbf{\Gamma}}^x_{s-r} \mathbf{A}'_r\|$'s. Denoting by $\mathbf{a}^j_{s\alpha}$ the α-th column, with $1 \leq \alpha \leq q+1$, of $\mathbf{A}_s^{j\prime}$, we have

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

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

$$(\hat{\mathbf{a}}^{j\prime}_{s\alpha} \hat{\mathbf{\Gamma}}^x_{jk,s-r} \hat{\mathbf{a}}^k_{r\beta} - \mathbf{a}^{j\prime}_{s\alpha} \hat{\mathbf{\Gamma}}^x_{jk,s-r} \mathbf{a}^k_{r\beta})^2 = ((\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})' \hat{\mathbf{\Gamma}}^x_{jk,s-r} \hat{\mathbf{a}}^k_{r\beta} - \mathbf{a}^{j\prime}_{s\alpha} \hat{\mathbf{\Gamma}}^x_{jk,s-r} (\hat{\mathbf{a}}^k_{r\beta} - \mathbf{a}^k_{r\beta}))^2.$$

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-Schwartz inequality, then subsequently the C_r and Jensen inequalities, we obtain

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

where

$$\begin{aligned}
\mathcal{A} &= m(q+1)^{1/2}\Big[\sum_{j=1}^{m}\sum_{\alpha=1}^{q+1}\big((\hat{\mathbf{a}}_{s\alpha}^{j}-\mathbf{a}_{s\alpha}^{j})'(\hat{\mathbf{a}}_{s\alpha}^{j}-\mathbf{a}_{s\alpha}^{j})\big)^{2}\Big]^{1/2},\\
\mathcal{B} &= \frac{1}{m}\sum_{k=1}^{m}\sum_{\beta=1}^{q+1}\Big[\sum_{j=1}^{m}\Big(\hat{\mathbf{a}}_{r\beta}^{k\prime}\hat{\mathbf{\Gamma}}_{jk,s-r}^{x\prime}\hat{\mathbf{\Gamma}}_{jk,s-r}^{x}\hat{\mathbf{a}}_{r\beta}^{k}\Big)^{2}\Big]^{1/2}\\
&\qquad\leq \Big[(q+1)/m\sum_{k=1}^{m}\sum_{\beta=1}^{q+1}\sum_{j=1}^{m}\Big(\hat{\mathbf{a}}_{r\beta}^{k\prime}\hat{\mathbf{\Gamma}}_{jk,s-r}^{x\prime}\hat{\mathbf{\Gamma}}_{jk,s-r}^{x}\hat{\mathbf{a}}_{r\beta}^{k}\Big)^{2}\Big]^{1/2} = \mathcal{C},\ \text{say.}
\end{aligned}$$

First consider $\mathcal{A}$. Letting $\mathbf{a}_{s\alpha}^{j\prime} = (a_{s\alpha,1}^{j}\ a_{s\alpha,2}^{j}\ \cdots\ a_{s\alpha,q+1}^{j})$, note that $a_{s\alpha,\delta}^{j} = \mathbf{e}_{\alpha}'\mathbf{A}^{[j]}\mathbf{g}_{s\delta}$, where $\mathbf{e}_{\alpha}$ and $\mathbf{g}_{s\delta}$ are the α-th 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 $\mathbf{B}_j^{\chi}$ and $\mathbf{C}_{jj}^{\chi}$, as defined in (2.14) and (2.15), we obtain, from (B.1), and applying subsequently the C_r, the triangular, the C_r again and then twice the Cauchy-Schwartz inequalities,

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

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

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

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

In a similar way, one can prove that $\sum_{j=1}^m\|\hat{\mathbf{C}}_j-\mathbf{C}_j\|^8$ is $O_{\mathrm{P}}(m(\max(n^{-1},\rho_T^{-1}))^4)$. 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:

$$\begin{aligned}\mathcal{A} &= m(q+1)^{1/2}\big[\sum_{j=1}^{m}\sum_{\alpha=1}^{q+1}\big((\hat{\mathbf{a}}^j_{s\alpha}-\mathbf{a}^j_{s\alpha})'(\hat{\mathbf{a}}^j_{s\alpha}-\mathbf{a}^j_{s\alpha})\big)^2\big]^{1/2}\\ &\leq 2^{3/2}(q+1)^2 m\big(\sum_{i=1}^{m}\|\hat{\mathbf{A}}^i_s-\mathbf{A}^i_s\|^4\big)^{1/2} = O_{\mathrm{P}}\big(m^{3/2}\max(n^{-1},\rho_T^{-1})\big). \qquad \text{(D.8)}\end{aligned}$$

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

$$\begin{aligned}&\mathcal{C} \leq ((q+1)/m)^{1/2}\Big\{\big[\sum_{k=1}^{m}\big(\sum_{\beta=1}^{q+1}(\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\mathbf{a}}^k_{r\beta})^2\big)^2\big]^{1/2}\big[\sum_{j=1}^{m}\big(\sum_{k=1}^{m}\big(\mathrm{trace}[\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\hat{\mathbf{\Gamma}}^{x}_{jk,s-r}]\big)^4\big)^{1/2}\big]\Big\}^{1/2}\\ &\leq ((q+1)/m)^{1/2}\Big\{\big[(q+1)\sum_{k=1}^{m}\sum_{\beta=1}^{q+1}(\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\mathbf{a}}^k_{r\beta})^4\big]^{1/2}\big[\sum_{j=1}^{m}\big(\sum_{k=1}^{m}\big(\mathrm{trace}[\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\hat{\mathbf{\Gamma}}^{x}_{jk,s-r}]\big)^4\big)^{1/2}\big]\Big\}^{1/2}\\ &\leq (q+1)^{1/2}\big[(q+1)^4\sum_{k=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}(\hat{a}^k_{r,\alpha\beta})^8\big]^{1/4}\big[m^{-1}\sum_{j=1}^{m}\sum_{k=1}^{m}\big(\mathrm{trace}[\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\hat{\mathbf{\Gamma}}^{x}_{jk,s-r}]\big)^4\big]^{1/4}\\ &\leq (q+1)^{3/2}\big[\sum_{k=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}(\hat{a}^k_{r,\alpha\beta})^8\big]^{1/4}\big[((q+1)^6/m)\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}(\hat{\gamma}^x_{jk,\alpha\beta}(s-r))^8\big]^{\frac{1}{4}}\\ &= O_{\mathrm{P}}(m^{1/2}),\end{aligned}$$

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

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

Now consider the second statement in (D.4). Again, it is sufficient to prove that it holds with $\|\hat{\mathbf{\Gamma}}^z-\mathbf{\Gamma}^z\|$ replaced with any of the $\|\hat{\mathbf{A}}_s\hat{\mathbf{\Gamma}}^x_{s-r}\hat{\mathbf{A}}'_r-\mathbf{A}_s\hat{\mathbf{\Gamma}}^x_{s-r}\mathbf{A}'_r\|$. Without loss of generality, we can assume that the number M of elements selected by $\mathcal{S}_{\mathbf{M}}$ is of the form $M=M^*(q+1)$ for some integer M^*. The two terms on the right-hand side of (D.7) must be dealt with separately. In the first of those two terms, substituting the summation $\sum_{k=1}^{M^*}$ for $\sum_{k=1}^{m}$ gives

$$\sum_{j=1}^{m}\sum_{k=1}^{M^*}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}\Big((\hat{\mathbf{a}}^j_{s\alpha}-\mathbf{a}^j_{s\alpha})'\hat{\mathbf{\Gamma}}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta}\Big)^2 = O_{\mathrm{P}}\big(m(\max(n^{-1},\rho_T^{-1}))\big).$$

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

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

and

$$\mathcal{E} = \sum_{k=1}^{M^*}\sum_{\beta=1}^{q+1}\Big(\frac{1}{m}\sum_{j=1}^{m}(\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\hat{\mathbf{\Gamma}}^{x}_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta})^2\Big)^{1/2}$$

can be bounded along the same lines as $\mathcal{A}$ and $\mathcal{B}$ are 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

$$\sum_{k=1}^{M^*}\sum_{j=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}\Big((\hat{\mathbf{a}}^k_{s\alpha} - \mathbf{a}^k_{s\alpha})'\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\mathbf{a}^j_{r\beta}\Big)^2$$
$$\leq m\Big[\sum_{k=1}^{M^*}\Big[\sum_{\alpha=1}^{q+1}(\hat{\mathbf{a}}^k_{s\alpha} - \mathbf{a}^k_{s\alpha})'(\hat{\mathbf{a}}^k_{s\alpha} - \mathbf{a}^k_{s\alpha})\Big]^2\Big]^{1/2}\Big[\frac{1}{m}\sum_{j=1}^{m}\sum_{\beta=1}^{q+1}\Big[\sum_{k=1}^{M^*}(\mathbf{a}^{j\prime}_{r\beta}\hat{\mathbf{\Gamma}}^{x}_{jk,s-r}\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\mathbf{a}^j_{r\beta})^2\Big]^{1/2}\Big]$$
$$= \mathcal{F}\mathcal{G},\ \text{say.}$$

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

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

□

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, 2, 3 and 4, using the same arguments used in Section B, with x, χ and ξ replaced by Z, Ψ and Φ respectively. 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 $O_{\mathrm{P}}(\zeta_{nT}) + O(1/n) = O_{\mathrm{P}}(\zeta_{nT})$, while (iv) has $O_{\mathrm{P}}(\zeta_{nT}) + O(1/\sqrt{n}) = O_{\mathrm{P}}(\zeta_{nT})$.

(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

$$\|\mathcal{S}_{\mathbf{M}}' \left(\hat{\mathbf{P}}^z \left(\hat{\mathbf{\Lambda}}^z \right)^{1/2} - \mathbf{P}^{\psi} \left(\mathbf{\Lambda}^{\psi} \right)^{1/2} \hat{\mathbf{W}}_q \right) \| = O_{\mathrm{P}} \left(\zeta_{nT} \right). \tag{D.9}$$

Going over the proof of Lemma 4, we see that $\|c-d\|$ has the worst rate, whereas here $\|a-b\|$, $\|b-c\|$ and $\|c-d\|$ all have rate $O_{\mathrm{P}}(\zeta_{nT})$.

(V) Moreover, in the same way as the proof of Lemma 4 can be replicated to obtain (D.9), the proof of Lemma 6, see below, can be replicated to obtain:

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

E Proof of Proposition 11

Let

$$\begin{aligned}
\hat{\mathbf{v}}_t &= ((\hat{\mathbf{\Lambda}}^z)^{1/2} \hat{\mathbf{P}}^{z\prime} \hat{\mathbf{P}}^z (\hat{\mathbf{\Lambda}}^z)^{1/2})^{-1} (\hat{\mathbf{\Lambda}}^z)^{1/2} \hat{\mathbf{P}}^{z\prime} \hat{\mathbf{Z}}_t = (\hat{\mathbf{\Lambda}}^z)^{-1/2} \hat{\mathbf{P}}^{z\prime} \hat{\mathbf{Z}}_t \\
&= (\hat{\mathbf{\Lambda}}^z)^{-1/2} \hat{\mathbf{P}}^{z\prime} (\hat{\mathbf{A}}(L) - \mathbf{A}(L)) \mathbf{x}_t + ((\hat{\mathbf{\Lambda}}^z)^{-1/2} \hat{\mathbf{P}}^{z\prime} - \hat{\mathbf{W}}^z (\mathbf{\Lambda}^{\psi})^{-1/2} \mathbf{P}^{\psi\prime}) \mathbf{A}(L) \mathbf{x}_t \\
&\quad + \hat{\mathbf{W}}^z (\mathbf{\Lambda}^{\psi})^{-1/2} \mathbf{P}^{\psi\prime} \mathbf{A}(L) \xi_t + \hat{\mathbf{W}}^z (\mathbf{\Lambda}^{\psi})^{-1/2} \mathbf{P}^{\psi\prime} \mathbf{P}^{\psi} (\mathbf{\Lambda}^{\psi})^{1/2} \mathbf{v}_t.
\end{aligned} \tag{E.11}$$

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

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

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

$$\|n^{-1/2}(\hat{\mathbf{A}}(L) - \mathbf{A}(L))\mathbf{x}_t\| \leq n^{-1/2}\sum_{r=0}^{p}\big[\sum_{i=1}^{m}\mathbf{x}_{t-r}^{i\prime}(\hat{\mathbf{A}}_r^i - \mathbf{A}_r^i)'(\hat{\mathbf{A}}_r^i - \mathbf{A}_r^i)\mathbf{x}_{t-r}^i\big]^{1/2}$$
$$\leq \sum_{r=0}^{p}\big(n^{-1}\sum_{i=1}^{m}(\mathbf{x}_{t-r}^{i\prime}\mathbf{x}_{t-r}^i)^2\big)^{1/4}\big(n^{-1}\sum_{i=1}^{m}\big(\sum_{j=1}^{q+1}\sum_{h=1}^{q+1}(\hat{a}_{r,jh}^i - a_{r,jh}^i)^2\big)^2\big)^{1/4}$$
$$\leq \sum_{r=0}^{p}\big(n^{-1}\sum_{i=1}^{m}(\mathbf{x}_{t-r}^{i\prime}\mathbf{x}_{t-r}^i)^2\big)^{1/4}\big((q+1)^3 n^{-1}\sum_{i=1}^{m}\|\hat{\mathbf{A}}_r^i - \mathbf{A}_r^i\|^4\big)^{1/4}$$
$$= O_{\mathrm{P}}\left(\zeta_{nT}\right)$$

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

Next, considering the second term on the righthand side of (E.11),

$$\begin{aligned}
&\|\left((\hat{\mathbf{\Lambda}}^z)^{-1/2}\hat{\mathbf{P}}^{z\prime} - \hat{\mathbf{W}}^z(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\right)\mathbf{A}(L)\mathbf{x}_t\| \\
&= \|(\hat{\mathbf{\Lambda}}^z/n)^{-1}\left((\hat{\mathbf{\Lambda}}^z)^{1/2}\hat{\mathbf{P}}^{z\prime} - \hat{\mathbf{W}}^z\hat{\mathbf{\Lambda}}^z(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\right)\mathbf{A}(L)\mathbf{x}_t/n\| \\
&= \|(\hat{\mathbf{\Lambda}}^z/n)^{-1}\left((\hat{\mathbf{\Lambda}}^z)^{1/2}\hat{\mathbf{P}}^{z\prime} - \hat{\mathbf{W}}^z[\hat{\mathbf{\Lambda}}^z - \mathbf{\Lambda}^\psi + \mathbf{\Lambda}^\psi](\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\right)\mathbf{A}(L)\mathbf{x}_t/n\| \\
&\leq \|(\hat{\mathbf{\Lambda}}^z/n)^{-1}\|\|\left((\hat{\mathbf{\Lambda}}^z)^{1/2}\hat{\mathbf{P}}^{z\prime} - \hat{\mathbf{W}}^z(\mathbf{\Lambda}^\psi)^{1/2}\mathbf{P}^{\psi\prime}\right)\|\|\mathbf{A}(L)\mathbf{x}_t/n\| \\
&\quad + \|(\hat{\mathbf{\Lambda}}^z/n)^{-1}\|\|\hat{\mathbf{W}}^z(\hat{\mathbf{\Lambda}}^z - \mathbf{\Lambda}^\psi)(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\|\|\mathbf{A}(L)\mathbf{x}_t/n\| = O_{\mathrm{P}}\left(\zeta_{nT}\right),
\end{aligned}$$

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

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

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), $(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\mathbf{A}(L)\xi_t$ is $O_{\mathrm{P}}(n^{-1/2})$. To conclude, note that $\hat{\mathbf{W}}^z(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\mathbf{P}^\psi(\mathbf{\Lambda}^\psi)^{1/2}\mathbf{v}_t = \hat{\mathbf{W}}^z\mathbf{v}_t$. □