RECent
Center for Economic Research

WORKING PAPER SERIES

Dynamic Factor Models with Innite-Dimensional Factor Space: Asymptotic Analysis

Mario Forni, Marc Hallin, Marco Lippi and Paolo Zaffaroni

Working Paper 115

September 2015

www.recent.unimore.it

RECent: c/o Dipartimento di Economia Politica, Viale Berengario 51, I-41100 Modena, ITALY
Phone +39 059 2056856, Fax +39 059 2056947, E-mail: RECent@unimore.it

Dynamic Factor Models with Infinite-Dimensional Factor Space: Asymptotic Analysis*

Mario Forni†

Università di Modena e Reggio Emilia, CEPR and RECent

Marc Hallin‡§

ECARES, Université Libre de Bruxelles and ORFE, Princeton University

Marco Lippi¶

Einaudi Institute for Economics and Finance, Roma

Paolo Zaffaroni‖

Imperial College London and Università di Roma La Sapienza

September 4, 2015

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

†Research supported by the PRIN-MIUR Grant 2010J3LZEN.

‡Research supported by an Interuniversity Attraction Pole (2012-2017) of the Belgian Science Policy Office, a Crédit aux Chercheurs of the Fonds de la Recherche Scientifique-FNRS of the Communauté française de Belgique, and the Australian Research Council grant DP150100210.

§Académie Royale de Belgique and CentER, Tilburg University.

¶Research supported by the PRIN-MIUR Grant 2010J3LZEN.

‖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 asymptotic properties (consistency and rates) of the corresponding estimators have been studied in Forni, Hallin, Lippi and Reichlin (2004). Those estimators, however, rely on Brillinger's *dynamic* principal components, and thus involve two-sided filters, which leads to rather poor forecasting performances. No such problem arises with estimators based on standard *(static)* principal components, which have been dominant in this literature. On the other hand, the consistency of those *static* estimators requires the assumption that the space spanned by the factors has finite dimension, which severely restricts the generality afforded by the GDFM. This paper derives the asymptotic properties of a semiparametric estimator of the loadings and common shocks based on one-sided filters recently proposed by Forni, Hallin, Lippi and Zaffaroni (2015). Consistency and exact rates of convergence are obtained for this estimator, under a general class of GDFMs that does not require a finite-dimensional factor space. A Monte Carlo experiment and an empirical exercise on US macroeconomic data corroborate those theoretical results and demonstrate the excellent performance of those estimators in out-of-sample forecasting.

JEL subject classification : C0, C01, E0.
Key words and phrases: High-dimensional time series. Generalized dynamic factor models. Vector processes with singular spectral density. One-sided representations of dynamic factor models. Consistency and rates.

1 Introduction

In the present paper, we provide consistency results and consistency rates for the estimators recently proposed by Forni, Hallin, Lippi and Zaffaroni (2015) (hereafter, FHLZ) for the *Generalized Dynamic Factor Model* (GDFM).

Let

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

be an observed $(n_0 \times T_0)$-dimensional panel, namely, a n_0-tuple of time series observed over

a time period of length T_0. The GDFM, as introduced in Forni *et al.* (2000) and Forni and Lippi (2001) consists in modeling that panel as a finite realization of a stochastic process of the form $\{x_{it},\ i \in \mathbb{N},\ t \in \mathbb{Z}\}$, that is, a countable number of stochastic processes $\{x_{it},\ t \in \mathbb{Z}\}$ admitting a decomposition of the form

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

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

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

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

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

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

where the matrices $\mathbf{D}_j$ are $r \times r$ while $\mathbf{K}$ is $r \times q$, $r \geq q$. Under (1.4), Bai and Ng (2007) and Amengual and Watson (2007) provide consistent criteria to determine q. We refer to estimators and predictors based on the existence of the static representation (1.3) and standard

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

principal components as the *static method*, as opposed to the method developed in FHLZ and the present paper, referred to as the *dynamic method*.

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

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

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

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

In FHLZ, we show how one-sided estimators without the finite-dimensionality assumption can be obtained, under the additional condition that the common components have a *rational spectral density*, that is, each filter $b_{if}(L)$ in (1.2) is a ratio of polynomials in L:

$$\chi_{it} \quad = \quad \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt}, \quad i \in \mathbb{N},\ f = 1, 2, \ldots, q, \tag{1.6}$$

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 and f for the sake of simplicity).

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

Denote by $\mathbf{x}_t$, $\boldsymbol{\chi}_t$, $\boldsymbol{\xi}_t$ the infinite-dimensional column vectors with components x_{it}, χ_{it}, and ξ_{it}, respectively. Elaborating upon recent results by Anderson and Deistler (2008a, b), FHLZ prove that, for generic values of the parameters $c_{if,k}$ and $d_{if,k}$ (i.e. apart from a lower-dimensional subset in the parameter space, see FHLZ for details), the infinite-dimensional idiosyncratic vector $\boldsymbol{\chi}_t = (\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{nt}\ \cdots)'$ admits *a unique autoregressive representation with block structure* of the form

$$\begin{pmatrix} \mathbf{A}^1(L) & 0 & \cdots & 0 & \cdots \\ 0 & \mathbf{A}^2(L) & \cdots & 0 & \\ & & \ddots & & \\ 0 & 0 & \cdots & \mathbf{A}^k(L) & \\ \vdots & & & & \ddots \end{pmatrix} \boldsymbol{\chi}_t = \begin{pmatrix} \mathbf{R}^1 \\ \mathbf{R}^2 \\ \vdots \\ \mathbf{R}^k \\ \vdots \end{pmatrix} \mathbf{u}_t, \tag{1.7}$$

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 (1.7), respectively, and letting $\mathbf{Z}_t = \underline{\mathbf{A}}(L)\mathbf{x}_t$, it follows that

$$\mathbf{Z}_t = \underline{\mathbf{R}}\mathbf{u}_t + \underline{\mathbf{A}}(L)\boldsymbol{\xi}_t. \tag{1.8}$$

Under the assumptions of the present paper, the term $\underline{\mathbf{A}}(L)\boldsymbol{\xi}_t$ is still idiosyncratic, so that (1.8) is a static representation of the form (1.4), with $\mathbf{D}(L) = \mathbf{I}$. That static representation can be estimated via traditional principal components, which does not require two-sided filters.

FHLZ thus obtain one-sided estimators for the common components without imposing the standard finite-dimension restriction. Moreover, the high-dimensional VAR (1.7) is obtained by piecing tothether the low-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 the procedure. 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 (1.7) and (1.8):

(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 $\hat{\boldsymbol{\Sigma}}_n^x(\theta)$.

(ii) Using the first q frequency domain principal components of $\hat{\boldsymbol{\Sigma}}_n^x(\theta)$, we construct an estimator of the spectral density of $\boldsymbol{\chi}_{nt} = (\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{nt})$, call it $\hat{\boldsymbol{\Sigma}}_n^\chi(\theta)$. Estimators

of the autocovariances of $\boldsymbol{\chi}_{nt}$ are then obtained from $\hat{\boldsymbol{\Sigma}}_n^{\chi}(\theta)$; call $\hat{\boldsymbol{\Gamma}}_{n,h}^{\chi}$ the estimator of the covariance between $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\chi}_{n,t-h}$. Those $\hat{\boldsymbol{\Gamma}}_{n,h}^{\chi}$'s are used, in a traditional, low-dimensional way, to construct the autoregressive estimators $\hat{\mathbf{A}}^k(L)$.

(iii) Blockwise estimators of the variables Z_{jt} are obtained by applying the finite-degree filters $\hat{\mathbf{A}}^k(L)$ to the observed variables x_{it}, while inverting the same $\hat{\mathbf{A}}^k(L)$'s provides estimators for the filters $b_{if}(L)$. Estimators for the shocks u_{ft} and the matrix $\mathbf{R}^k$ are obtained by using the first q traditional 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) and Liu and Wu (2010), as extended to the multivariate case by Wu and Zaffaroni (2015). Starting with the observable time series x_{it}, denoting by T the number of observations for each series and by $\hat{\sigma}_{ij}(\theta)$ a lag-window estimator of the cross-spectrum between x_{it} and x_{jt}, the $(i,\ j)$ entry of $\hat{\boldsymbol{\Sigma}}(\theta)$, 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}.

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

$$\zeta_{nT} = \max\left(\sqrt{n^{-1}}, \sqrt{T^{-1} B_T \log B_T}\right), \tag{1.9}$$

where B_T diverges as T^{δ}, with $1/3 < \delta < 1$. Establishing those rates, raises some nontrivial difficulties. Although model (1.8) is finite-dimensional, indeed, the series Z_{it} are estimated, not observed. As a consequence, the well-known results from static-factor literature (Stock and Watson, 2002a and b, Bai and Ng, 2002) do not readily apply, and proving that consistency holds with the same rates ζ_{nT} as if Z_{it} were observed requires non negligible efforts.

As pointed out in FHLZ (end of Section 4.5) despite the fact that the dynamic model studied in this paper is more general than model (1.4), when a dataset is given, with finite n

and T, the static approach might perform well even though the required finite-dimension assumptions are not satisfied. A Monte Carlo study is provided in Section 4, in which the static and dynamic methods have been applied to simulated data. A very short summary of our results is that (i) when the data are generated by infinite-dimensional models which are simple generalizations of (1.5), the estimation of impulse-response functions and predictions via the dynamic method is by far better than those obtained via the static one; (ii) even when the data are generated by (1.4), still the dynamic method performs slightly better. Though not conclusive, our Monte Carlo results strongly suggest that the FHLZ method may be uniformly competitive. A pseudo out-of-sample forecasting exercise with US quarterly macroeconomic series provides further evidence in favour of the dynamic method.

The paper is organized as follows. In Sections 2, we present and comment the main assumptions to be made throughout. Section 3 provides the main asymptotic results. Sections 4 and 5 contain a detailed description and analysis of the Monte Carlo experiments and the empirical exercise respectively. Section 6 concludes. Short proofs are given in the body of the paper, the longer ones in the Appendix.

2 Main assumptions and some preliminary results

2.1 Common and idiosyncratic components

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

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

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

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

fulfills the assumptions listed below as Assumptions 1 through 10.

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

(1) a q-dimensional stochastic zero-mean process $\mathbf{u}_t = (u_{1t}\ u_{2t}\ \cdots\ u_{q_t})'$, $t \in \mathbb{Z}$, and an infinite-dimensional stochastic process $\boldsymbol{\eta}_t = (\eta_{1t}\ \eta_{2t}\ \cdots)'$, $t \in \mathbb{Z}$;

(2) square-summable filters $b_{if}(L)$, $i \in \mathbb{N}$, $f = 1, \ldots, q$;

(3) coefficients $\beta_{ij,k}$, *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 vectors $\mathbf{S}_t = (\mathbf{u}_t' \ \boldsymbol{\eta}_t')'$, $t \in \mathbb{Z}$, *are 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)
$$\begin{aligned} \chi_{it} &= b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} \\ \xi_{it} &= \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \beta_{ij,k} \eta_{j,t-k}. \end{aligned} \tag{2.1}$$

Clearly, neither $\mathbf{u}_t$ nor the polynomials $b_{if}(L)$ are identified. Indeed, rewriting the first 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 (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 just second-order 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, there exist natural numbers* s_1, s_2 *such that* $b_{if}(L) = c_{if}(L)/d_{if}(L)$, *where*

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

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* $|c_{if,j}| \leq B^\chi$, $i \in \mathbb{N}$, $f = 1, \ldots, q$, $j = 0, \ldots, s_1$.

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

Assumption 3 *Common component spectral density eigenvalues: divergence and separation. There exist continuous functions*

$$\alpha_f^\chi(\theta), \quad f = 1, \ldots, q \quad \textit{and} \quad \beta_f^\chi(\theta), \quad f = 0, \ldots, q-1, \quad \theta \in [-\pi,\ \pi],$$

and a positive integer n^χ such that, for $n > n^\chi$,

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

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

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

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

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

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

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

An immediate consequence of (2.3) and (2.4) is that

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

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

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

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

Therefore, letting $\sigma_{ij}^{\xi}(\theta)$ be the cross-spectral density of 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). Assumption 4 thus implies that the cross-spectra $\sigma_{ij}^{\xi}(\theta)$ are bounded, in θ, uniformly in i and j. On the other hand, Assumption 2, (ii) and (iii), 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.

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

Proposition 1 *Under Assumptions 1 through 4,*
(i) there exists $B^{\xi} > 0$ such that $\lambda_{n1}^{\xi}(\theta) \leq B^{\xi}$ 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 $\theta \in [-\pi,\ \pi]$,

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

where the functions $\alpha_j^{\chi}(\theta)$ 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 row norms of $\mathbf{\Sigma}_n^{\xi}(\theta)$ are equal, and, by (2.7), satisfy

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

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

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

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

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

$$\frac{\lambda^x_{nf}(\theta)}{n} \geq \frac{\lambda^\chi_{nf}(\theta) + \lambda^\xi_{nn}(\theta)}{n} > \alpha^\chi_f(\theta),$$

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

$$\frac{\lambda^x_{nf}(\theta)}{n} \leq \frac{\lambda^\chi_{nf}(\theta) + \lambda^\xi_{n1}(\theta)}{n} \leq \frac{\lambda^\chi_{nf}(\theta)}{n} + \frac{B^\xi}{n} \leq \beta^\chi_{f-1}(\theta) + \frac{B^\xi}{n} < \alpha^\chi_{f-1}(\theta),$$

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

As for (iii), $\lambda^x_{n,q+1} \leq \lambda^\chi_{n,q+1}(\theta) + \lambda^\xi_{n1}(\theta)$. On the other hand, $\lambda^\chi_{n,q+1}(\theta) = 0$ for all θ. The result then follows from (i). □

Proposition 2 *Under Assumptions 1 through 4, the cross-spectral densities $\sigma^x_{ij}(\theta)$ 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*

$$\sum_{h=1}^{\nu} |\sigma^x_{ij}(\theta_h) - \sigma^x_{ij}(\theta_{h-1})| \leq A^x$$

for all $i, j, \nu \in \mathbb{N}$ and all partitions

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

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

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

$$|\gamma^\xi_{ij,h}| = \left|\sum_{k=0}^{\infty}\sum_{s=1}^{\infty}\beta_{is,k}\beta_{js,k+h}\right| \leq \sum_{k=0}^{\infty}\sum_{s=1}^{\infty} B_{is}B_{js}\rho^k\rho^{k+h} \leq \rho^h\sum_{k=0}^{\infty}\rho^{2k}\sum_{s=1}^{\infty}B_{is}B_{js} \leq \rho^h\frac{B^2}{1-\rho^2}, \tag{2.8}$$

by (2.6). For $h < 0$, $\gamma^\xi_{ij,h} = \gamma^\xi_{ji,-h}$, so that $|\gamma^\xi_{ij,h}| \leq \rho^{|h|}B^2/(1-\rho^2)$. This implies that

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

has derivatives of all orders. Moreover,

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

which entails 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 the fact that $\sigma^x_{ij}(\theta) = \sigma^\chi_{ij}(\theta) + \sigma^\xi_{ij}(\theta)$. □

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 ones). Moreover, we prove that, *generically,* the $(q+1)$-dimensional subvectors of $\boldsymbol{\chi}_t$ admit a *finite and unique* autoregressive representation (see, in particular, Section 4.1, Lemma 3). Following FHLZ, we use these genericity results as a motivation for assuming that each of the vectors

$$\begin{pmatrix} \chi_{1t} & \chi_{2t} & \cdots & \chi_{q+1,t} \end{pmatrix}, \quad \begin{pmatrix} \chi_{q+2,t} & \chi_{q+3,t} & \cdots & \chi_{2(q+1),t} \end{pmatrix}, \quad \ldots,$$

that is, each of the vectors

$$\boldsymbol{\chi}_t^k = \begin{pmatrix} \chi_{(k-1)(q+1)+1,t} & \cdots & \chi_{k(q+1),t} \end{pmatrix}, \quad k \in \mathbb{N},$$

and its lags spans the space spanned by the u's and has a unique finite autoregressive representation.

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

$$\mathbf{A}^k(L)\boldsymbol{\chi}_t^k = \mathbf{R}^k \mathbf{u}_t, \tag{2.9}$$

where

(i) $\mathbf{A}^k$ *is* $(q+1) \times (q+1)$*, of degree not greater than* $S = qs_1 + q^2 s_2$*, and* $\mathbf{A}^k(0) = \mathbf{I}_{q+1}$*;*

(ii) $\mathbf{R}^k$ *is* $(q+1) \times q$ *and has rank* q*;*

(iii) the representation (2.9) is unique among the autoregressive representations of order not greater than S, i.e. if $\mathbf{B}(L)\boldsymbol{\chi}_t^k = \tilde{\mathbf{R}}\mathbf{u}_t$, where the degree of $\mathbf{B}(L)$ does not exceed S and $\mathbf{B}(0) = \mathbf{I}_{q+1}$, then $\mathbf{B}(L) = \mathbf{A}^k(L)$ and $\tilde{\mathbf{R}} = \mathbf{R}^k$.

Representation (2.9) is a specification of (1.7) (the degrees of the polynomial matrices $\mathbf{A}^k(L)$ and their uniqueness).[3] Writing $\underline{\mathbf{A}}(L)$ for the (infinite) block-diagonal matrix

[3]Based on a genericity argument, FHLZ assume that (2.9) holds for any $(q+1)$-dimensional vector $(\chi_{i_1,t}\ \chi_{i_2,t}\ \cdots\ \chi_{i_{q+1},t})$, see Assumption A.3. The weaker version in Assumption 5 above is sufficient for our purposes.

with diagonal blocks $\mathbf{A}^1(L), \mathbf{A}^2(L), \ldots$, 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,

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

The following proposition is an immediate consequence of the fact that (2.10) is the difference between $\boldsymbol{\chi}_t$ and its orthogonal projection on its past values; details are left to the reader.

Proposition 3 *Let Assumptions 1 through 5 hold.*

(i) Let $\underline{\mathbf{A}}^*(L)\boldsymbol{\chi}_t = \underline{\mathbf{R}}^*\mathbf{v}_t$, where the degree of $\underline{\mathbf{A}}^*(L)$ is at most S: then, $\underline{\mathbf{A}}^*(L) = \underline{\mathbf{A}}(L)$, and there exists a $q \times q$ orthogonal matrix $\mathbf{Q}$ such that $\underline{\mathbf{R}}^* = \underline{\mathbf{R}}\mathbf{Q}'$ and $\mathbf{v}_t = \mathbf{Q}\mathbf{u}_t$.

(ii) Let $\mathbf{r} = (r_1 \ \cdots \ r_q)$ be the row of $\underline{\mathbf{R}}$ (the row of $\mathbf{R}^k$) corresponding to χ_{it}: then, $r_f = c_{if}(0) = c_{if,0}$, $f = 1, \ldots, q$, $i \in \mathbb{N}$.

Letting $\boldsymbol{\Psi}_t = \underline{\mathbf{A}}(L)\boldsymbol{\chi}_t = \underline{\mathbf{R}}\mathbf{u}_t$, denote by $\boldsymbol{\Gamma}_n^{\psi}$ the variance-covariance matrix of $\boldsymbol{\Psi}_{nt}$, with eigenvalues μ_{nj}^{ψ}, $j = 1, \ldots, n$, in decreasing order.

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

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

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 the matrices $\mathbf{A}^k(L)$ appearing in (2.9) can be constructed from the spectral density of the χ's. This construction, with the population quantities replaced by their estimates, leads to our estimator as explained in Section 3. It proceeds in two steps:

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

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

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

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

follow from that autocovariance function $\boldsymbol{\Gamma}_{kk,s}^{\chi}$. Indeed, defining

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

$$\mathbf{A}^{[k]} = \left(\mathbf{A}_1^k \ \mathbf{A}_2^k \ \cdots \ \mathbf{A}_S^k\right), \qquad \mathbf{B}_k^{\chi} = \left(\boldsymbol{\Gamma}_{kk,1}^{\chi} \ \boldsymbol{\Gamma}_{kk,2}^{\chi} \ \cdots \ \boldsymbol{\Gamma}_{kk,S}^{\chi}\right) \tag{2.14}$$

and

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

we have

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

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

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

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

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

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

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

Denote by $\mathbf{\Gamma}_n^\Phi$ the variance-covariance matrix of $\mathbf{\Phi}_{nt} = (\Phi_{1t}\ \Phi_{2t}\ \cdots\ \Phi_{nt})$ and by μ_{nj}^Φ its j-th eigenvalue: the following holds

Proposition 4 *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 $\mathbf{\Phi}_{nt}$. Let us show that there exists a constant C^Φ such that $\lambda_{n1}^\Phi(\theta) \leq C^\Phi$ for all n and θ. Because $\lambda_{n1}^\Phi(\theta)$, for all θ, is non-decreasing with n (see Forni and Lippi, 2001), we can assume without loss of generality 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 mutually orthogonal, a consequence of Assumption 1(i). In view of Assumption 6 and Proposition 4, the model (2.17) is thus a static factor model—a special case of (1.4), with $r = q$ and $\mathbf{N}(L) = \mathbf{I}_q$.

3 Estimation: asymptotics

Our estimation procedure follows the same steps as the population construction in Section 2.2, with the population spectral density of the x's replaced with an estimator $\hat{\mathbf{\Sigma}}_n^x(\theta)$ fulfilling Assumption 9 below. 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. Finally, estimators for $\mathbf{R}_n$ and $\mathbf{u}_t$ are obtained via a standard principal component analysis of $\hat{\mathbf{Z}}_{nt} = \hat{\mathbf{A}}(L)\mathbf{x}_{nt}$. Consistency with exact rate of convergence ζ_{nT}, as defined in equation (1.9), 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 $p_{if}^x(\theta)$, $f = 1, \dots, q$, $i = 1, \dots, n$, 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 observed values x_{it}, $i = 1, \dots, n$, $t = 1, \dots, T$. For simplicity, we say that they depend on n and T.

(vi) The same notational change applies to $\mathbf{\Gamma}_n^\psi$ 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 $\tilde{\mathbf{F}}$ its conjugate transpose, and by $||\mathbf{F}||$ its spectral norm (see Appendix B).

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

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

Definition 1 *Physical dependence. Let $\boldsymbol{\epsilon}_t$ be an i.i.d. stochastic vector process, possibly infinite-dimensional, and let $z_t = F(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1}, \ldots)$, where $F : [\mathbb{R} \times \mathbb{R} \times \cdots] \to \mathbb{R}$ is a measurable function; assume that z_t has finite p moment for $p > 0$. Let $\boldsymbol{\epsilon}^*$ be a stochastic vector with the same dimension and distribution as the $\boldsymbol{\epsilon}_t$'s, such that $\boldsymbol{\epsilon}^*$ and $\boldsymbol{\epsilon}_t$ are independent for all t. For $k \geq 0$ the* physical dependence $\delta_{kp}^{[z_t]}$ *is defined as*

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

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 $\rho_1 \in (0,1)$ and $A_1 \in (0,\infty)$ such that, for all $i \in \mathbb{N}$,*

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

PROOF. By the Minkovski inequality,

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

Using the Minkovski inequality again, condition (2.3) and Assumption 8, we obtain

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

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

Turning to 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 as ξ_{ik}, with η_{s0} replaced by η_s^*. The Minkovski inequality, condition (2.3) and Assumption 8 imply

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

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,

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

The conclusion follows. □

Consider now the lag-window estimator

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

of the spectral density $\mathbf{\Sigma}^x(\theta)$, where $\hat{\mathbf{\Gamma}}_k^x = \frac{1}{T}\sum_{t=|k|+1}^{T}\mathbf{x}_t\mathbf{x}_{t-|k|}'$.

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

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

(1) for some $\kappa > 0$, $|K(u) - 1| = O(|u|^\kappa)$ *as* $u \to 0$,

(2) $\int_{-\infty}^{\infty} K^2(u)du < \infty$,

(3) $\sum_{j\in\mathbb{Z}} \sup_{|s-j|\leq 1} |K(jw) - K(sw)| = O(1)$ *as* $w \to 0$*;*

(ii) For some $c_1, c_2 > 0$, δ *and* $\underline{\delta}$ *such that* $0 < \underline{\delta} < \delta < 1 < \underline{\delta}(2\kappa + 1)$, $c_1 T^{\underline{\delta}} \leq B_T \leq c_2 T^{\delta}$.

Proposition 6 *Under Assumptions 1 through 9, there exists* $C > 0$ *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, i *and* j *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 $\hat{\mathbf{\Sigma}}^{\chi}(\theta) = \hat{\mathbf{P}}^x(\theta)\hat{\mathbf{\Lambda}}^x(\theta)\tilde{\hat{\mathbf{P}}}^x(\theta_h)$.

Proposition 7 *Under Assumptions 1 through 7,*

$$\max_{|h|\leq B_T} |\hat{\sigma}_{ij}^{\chi}(\theta_h^*) - \sigma_{ij}^{\chi}(\theta_h^*)| = O_P\left(\zeta_{nT}\right),$$

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

$$P\left(\frac{\max_{h\leq B_T} |\hat{\sigma}_{ii}^{\chi}(\theta_h^*) - \sigma_{ii}^{\chi}(\theta_h^*)|}{\zeta_{nT}} \geq \eta(\epsilon)\right) < \epsilon.$$

See Appendix B for the proof.

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

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

where $\theta_h^* = \pi h/B_T$. Recalling that $\gamma_{ij,\ell}^{\chi} = \int_{-\pi}^{\pi} e^{\imath\ell\theta} \sigma_{ij}^{\chi}(\theta)d\theta$, we have

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

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

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

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

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

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

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

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

Under our assumptions, the common component admits the block-diagonal vector autoregressive representation (1.6) of finite order. If the χ_t's were observed, estimation by OLS would be appropriate. However, although we do not observe the χ_t's, we do have (consistent) estimates of their autocovariance function. This naturally leads to a Yule-Walker estimator of the autoregressive coefficients and the innovation covariance matrix. The definition of $\hat{\mathbf{A}}^{[k]}$ then is straightforward from (2.14), (2.15) and (2.16).

Proposition 9 *Under Assumptions 1 through 10,* $\|\hat{\mathbf{A}}^{[k]} - \mathbf{A}^{[k]}\| = O_P(\zeta_{nT})$ *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, for given i and f, the (i, f) entry of $\mathcal{R}$ depends on n, so that the matrices $\mathcal{R}$ are not nested; nor is $\mathbf{v}_t$ independent of n. However, the product of each row of $\mathcal{R}$ by $\mathbf{v}_t$ yields the corresponding coordinate of $\mathbf{\Psi}_{nt}$ which does not depend on 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,* $\|\hat{\mathcal{R}}_i - \mathcal{R}_i\hat{\mathbf{W}}_q\| = O_P(\zeta_{nT})$, *as* $T \to \infty$ *and* $n \to \infty$, *where* $\mathcal{R}_i$ *is the* i*-th row of* $\mathcal{R}$, *and* $\hat{\mathbf{W}}_q$ *is a* $q \times q$ *diagonal matrix, depending on* n *and* T, *whose diagonal entries are either* 1 *or* -1.

See Appendix D for the proof.

Let us point out again that the 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, those of $\mathcal{R}\mathbf{G}(\mathcal{R})$ do not.

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

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

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

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.

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

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

See Appendix E for the proof.

3.5 Estimation and cross-sectional ordering

Let us now focus on the observed $(n_0 \times T_0)$-dimensional panel (1.1) and assume for convenience that $n_0 = m_0(q+1)$. Because the ordering of the n_0 variables is arbitrary (macroeconomic datasets are standard in this literature), sensible concepts and sensible inference methods, as a rule, should be invariant under permutations. On the other hand, while the definitions of common and idiosyncratic components, dynamic eigenvalues and principal components, etc., as well as the estimation method proposed in Forni et al. (2000), clearly are insensitive to the order of the cross-sectional items, the one-sided estimation method introduced in the present paper is not.

The ordering of the panel (1.1) indeed has a crucial impact on the selection of the $(q+1)$-dimensional blocks in the autoregressive representation of Section 2.2. Thus, in principle, any permutation of the cross-sectional items—more precisely, any of the permutations that lead to distinct partitions of $\{1, 2, \ldots, n_0\}$ into m_0 subsets of size $(q+1)$—yields distinct estimators (this is confirmed in the numerical illustration in Section 4). That order-dependence of course is highly undesirable, and those estimators somehow should be aggregated into a unique one, which should improve performances while providing permutational invariance. We propose to achieve this, from a theoretical perspective, by averaging them; more precisely, we propose to average the estimated impulse-response functions (or forecasts) over the $n_0^* = \frac{n_0!}{[(q+1)!]^{m_0}}$ possible orderings of the cross-sectional items of the (n_0, T_0)-dimensional panel.

Now, computing the estimators for n_0^* permutations, even for moderately large values of n_0, is, of course, numerically infeasible. The averaging solution just proposed is thus inapplicable. Fortunately, it appears that, selecting a few permutations at random and averaging the corresponding estimators leads to rapidly stabilizing results, so that going through all n_0^* permutations is not required in order to attain the desired average, hence an order-free final result. See Section 4 for an empirical justification, practical details, and a numerical illustration.

To conclude, let us observe that the averaging procedure just described requires enhancing Assumption 5 within the panel (1.1). Precisely, Assumption 5 should hold for all $(q+1)$-

dimensional blocks of the panel

$$x_{i_1,t},\ x_{i_2,t},\ \ldots,\ x_{i_{n_0},t},$$

for all the n_0^* permutations $(i_1,\ i_2,\ \ldots,\ i_{n_0})$. Now, if we consider the n_0^* infinite sequences

$$x_{i_1,t},\ x_{i_2,t},\ \ldots,\ x_{i_{n_0},t};\quad x_{n_0+1,t},\ x_{n_0+2,t},\ \ldots,$$

that is, the original infinite sequence with reordering of the first n_0 items, all the asymptotic consistency results hold for the corresponding n_0^* estimators, and therefore for their average.

4 A simulation exercise

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

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

4.1 Data-generating processes

We consider the following data-generating processes.

Model I (no static factor model representation)

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

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

Estimation of the shocks and the impulse-response functions requires an identification rule. Our exercise is based on a 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 let $\mathbf{H}$ be the lower triangular matrix with positive diagonal entries such that $\mathbf{H}\mathbf{H}' = \mathbf{B}_q(0)\mathbf{B}_q(0)'$. Then, the "structural" shocks, denoted by $\mathbf{u}_t^*$, and the impulse-response functions, denoted by $\mathbf{b}_i^*(L)$, are $\mathbf{b}_i^*(L) = \mathbf{b}_i(L)\mathbf{B}_q(0)^{-1}\mathbf{H}$ and $\mathbf{u}_t^* = \mathbf{H}'\mathbf{B}_q(0)'\mathbf{u}_t$, respectively.

Model II (with static factor representation)

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

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 i.i.d. standard Gaussian and mutually 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 $\mathbf{b}_i^*(L)$ 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 f-th entry of $\mathbf{b}_i^*(L)$. Our target is the estimate of $b_{if,k}^*$, $i = 1, \ldots, n$, $f = 1, \ldots, q$, $k = 0, \ldots, K$ and u_{ft}^*, $f = 1, \ldots, q$, $t = 1, \ldots, T$, as well as the forecast of x_{iT+1}, $i = 1, \ldots, n$.

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

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

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

(iii) The number of lags for each $(q+1)$-dimensional VAR matrix $\mathbf{A}^k(L)$. This is determined by the BIC criterion.

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

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

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

The number of structural shocks is no longer assumed known. Rather, it is estimated by the Hallin and Liška (2007) method.[5] SW forecasts are obtained by regressing $x_{i,T+1}$ onto either the ordinary principal components at T and x_{iT}, or the principal components at T alone. The former method corresponds to the original Stock and Watson (2002a) method; the latter is motivated by the fact that in both of the models 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:

[5] 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 $n_j = \lfloor (3n/4 + jn/40) \rfloor$, $T_j = T$, $j = 1, \ldots, 10$.

the estimation error on the impulse response functions is measured by

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

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

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

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

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

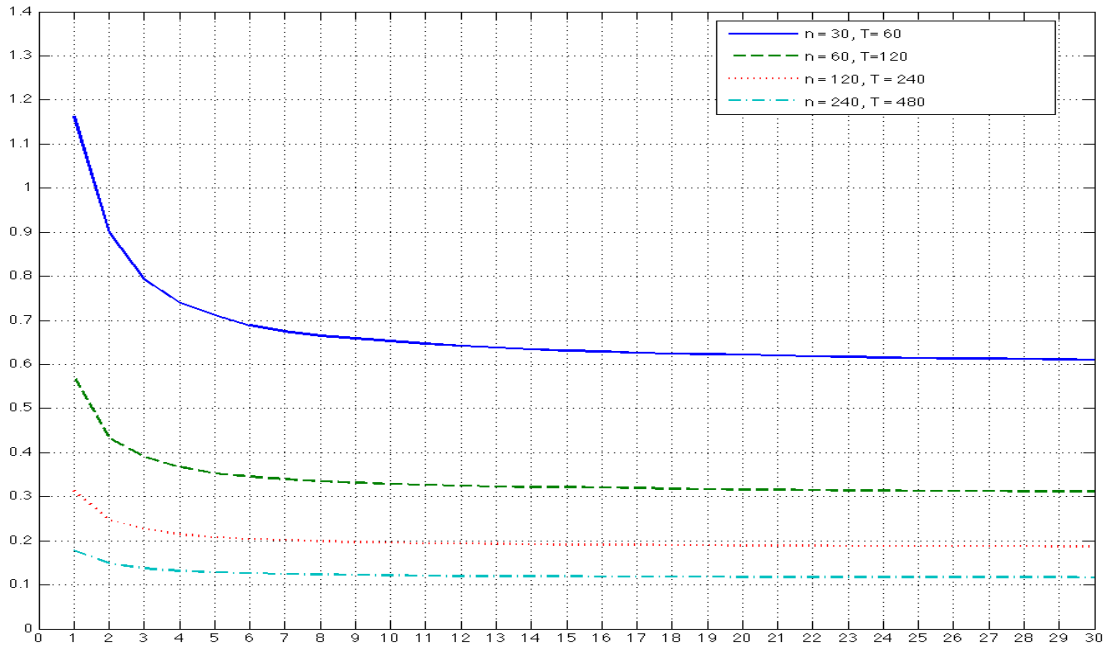
Model I is evaluated for different sample size combinations, with $n=30,60,\ 120,240$ and $T=60,120,240,480$. Model II is evaluated for a fixed sample size of $n=120$ and $T=240$, but different configurations of q and r, i.e. $r=4,6,8,12$ and $q=2,4,6$, $r>q$.[6] For each couple (n,T), Model I, and (r,q), Model II, we generated 500 data sets and computed the average MSE.

4.3 Cross-sectional permutations

As argued in Section 3.5, the estimators obtained via the FHLZ method should be averaged over different permutations of cross-sectional items. In order to study the influence of such permutations, we simulated 500 datasets from Model I and various values of n and T. For each of the resulting panels, we computed (with the Choleski identification rule described in Section 4.1) the estimated impulse response functions averaged over $\mu=1,\ldots,M$ randomly chosen permutations. For each value of μ, the MSEs (over the 500 replications) of the averaged estimators were recorded, leading to the following conclusions:

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

Figure 1: Model I. Average MSE of estimated impulse response functions over 500 experiments, as a function of the number of random reorderings of the variables used in estimation.



(i) as expected, estimates corresponding to different random permutations do differ;

(ii) averaging those estimates yields a clear improvement in the MSE;

(iii) the rate of that improvement declines steadily as the number μ of permutations increases, and rapidly stabilizes until additional permutations produce negligible effect;

(iv) as n and T increase, the improvement decreases, both in absolute and relative terms, and the number of permutations required for "stabilization" decreases: 10 for ($n = 60$, $T = 120$), only 5 for ($n = 240$, $T = 480$).

The results are reported in Figure 1. Summing up, averaging over random permutations until the resulting estimates stabilize is essentially equivalent to averaging over all possible permutations, hence restores the independence of the FHLZ method with respect to the panel ordering, while significantly improving the small-sample performance of FHLZ. Such averaging moreover does not modify the asymptotic results of Section 3.

4.4 Results

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

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

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

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

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

performance in most cases.

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

5 Empirical application

In this section we present a pseudo real-time forecast evaluation exercise with US quarterly data. We take as target variables real GDP, real private fixed investment, real consumption expenditures, the number of unemployed and the consumer price index. We compare results obtained with FHLZ, SW and a simple univariate autoregressive model. The forecasts are computed within a rolling window scheme. An extensive pseudo real-time forecasting analysis based on US monthly data is found in Forni, Giovannelli, Lippi and Soccorsi (2015).

5.1 Data and methods

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

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

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

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

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

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

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

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

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

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

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

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

(II) As for p, we tried a fixed $p = 0, 1, 4$ and a floating p determined by the BIC criterion.

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

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

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

We found that generally the inclusion of the dependent variable does not improve results.[14]

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

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

5.2 Results

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

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

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

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

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

We conclude that FHLZ outperforms standard techniques in the prediction of several macroeconomic variables, including GDP and inflation.

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

6 Conclusions

An estimate of the common-component 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 a rational spectral density for the common components, this implies that the number of parameters to estimate grows as n, not n^2.

The rational spectral density assumption also has the important consequences that $\boldsymbol{\chi}_t$ has a finite autoregressive representation and that the dynamic factor model can be transformed into the static 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 estimator 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 dynamic factor model has a static representation of the form (1.4). Sections 4 and 5 provide important empirical support to the richer dynamic structure of unrestricted GDFM's.

References

[1] Ahn, S.C., and A.R. Horenstein (2013) Eigenvalue ratio test for the number of factors, *Econometrica* **81**, 1203-1227.

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

[3] 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.

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

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

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

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

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

[9] D'Agostino, A. and D. Giannone (2012) Comparing Alternative Predictors Based on Large?Panel Factor Models, *Oxford Bulletin of Economics and Statistics* **74**, 306-326.

[10] Forni, M. and L. Gambetti (2014) Sufficient information in structural VARs, *Journal of Monetary Economics*, **66**, 124-136.

[11] Forni, M., Giovannelli, A., Lippi, M. and S. Soccorsi (2015) Dynamic Factor model with infinite dimensional factor space: forecasting, mimeo.

[12] 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.

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

[14] 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.

[15] Forni, M., M. Hallin, M. Lippi and P. Zaffaroni (2015). Dynamic factor models with infinite-dimensional factor space: one-sided representations, *Journal of Econometrics* **185**, 359-371.

[16] 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.

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

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

[19] 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.

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

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

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

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

[24] Onatski, A. (2010) Determining the number of factors from empirical distribution of eigenvalues, *The Review of Economics and Statistics* **92**, 1004-1016.

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

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

[27] 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.

[28] Stock, J.H., and M.W. Watson (2012). Generalized Shrinkage Methods for Forecasting Using Many Predictors, *Journal of Business & Economic Statistics* **30**, 481-493.

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

[30] Wu, W.B. and P. Zaffaroni (2015). On Uniform Moments Convergence of Spectral Density Estimates, arXiv 1505.03659, available at http://arxiv.org/abs/1505.03659.

Appendix

A Proof of Proposition 6

Adding and subtracting $E(\hat{\sigma}^x_{ij}(\theta^*_h))$ within the absolute value in $E\Big(\max_{|h|\leq B_T}\big|\hat{\sigma}^x_{ij}(\theta^*_h)-\sigma^x_{ij}(\theta^*_h)\big|^2\Big)$ and re-arranging gives

$$E\Big(\max_{|h|\leq B_T}\big|\hat{\sigma}^x_{ij}(\theta^*_h)-\sigma^x_{ij}(\theta^*_h)\big|^2\Big)\leq E\Big(\max_{|h|\leq B_T}\big|\hat{\sigma}^x_{ij}(\theta^*_h)-E\hat{\sigma}^x_{ij}(\theta^*_h)\big|^2\Big)+\Big(\max_{|h|\leq B_T}\big|E\hat{\sigma}^x_{ij}(\theta^*_h)-\sigma^x_{ij}(\theta^*_h)\big|^2\Big).$$

The first term (variance) on the right hand side of the above inequality satisfies

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

where C^* depends only on p (see Assumption 8), ρ_1 (see Proposition 5), δ (see Assumption 9). This is proved in Wu and Zaffaroni (2015) Lemma 10, with $\nu=2$.

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

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

Assumptions 2 and 4 imply that, for some $\phi\in(0,1)$ and some D, $|\gamma^x_{ij,l}|\leq|\gamma^\chi_{ij,l}|+|\gamma^\xi_{ij,l}|\leq D\phi^{|l|}$, for all i and j (see equation 2.8)). This inequality and Assumption 9(i) ensure that, for some F and all i, j and θ, $\mathcal{A}^T_{ij}(\theta)\leq FD\sum_{l=-\infty}^{\infty}\phi^{|l|}(|l|/B_T)^\kappa\leq[2DF\phi/(1-\phi)^2]T^{-\delta\kappa}=HT^{-\delta\kappa}$. Moreover, $\mathcal{B}^T_{ij}(\theta)\leq DT^{-1}\sum_{l=-\infty}^{\infty}\phi^{|l|}|l|=[2D\phi/(1-\phi)^2]T^{-1}=KT^{-1}$, for all i, j, and θ. Finally, $\mathcal{C}^T_{ij}(\theta)\leq D\sum_{|l|\geq T}\phi^{|l|}|l|^\kappa/T^\kappa$, since $|l|^\kappa/T^\kappa\geq1$ for $|l|\geq T$. Hence, it follows that $\mathcal{C}^T_{ij}(\theta)\leq MT^{-\kappa}\leq MT^{-\delta\kappa}$ for all i, j and θ. Thus $\mathcal{S}_{ij}(\theta)/2\pi\leq KT^{-1}+(H+M)T^{-\delta\kappa}\leq PT^{-\mu}$, where $\mu=\min(\delta\kappa,1)$, for all i, j and θ. Now, $2\delta\kappa>1-\underline{\delta}>1-\delta$, by Assumption 9(ii). Hence, $\max_{|h|\leq B_T}\left|E\hat{\sigma}^x_{ij}(\theta^*_h)-\sigma^x_{ij}(\theta^*_h)\right|^2\leq P^2T^{-2\mu}\leq C^{**}(B_T\log B_T/T)$ for all i and j. □

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 an $s\times s$ Hermitian matrix $\mathbf{A}$, and by $\|\mathbf{B}\|=\sqrt{\mu_1(\tilde{\mathbf{B}}\mathbf{B})}$

the spectral norm of an $s_1 \times s_2$ matrix $\mathbf{B}$. The norm $\|\mathbf{B}\|$ coincides with the Euclidean norm of $\mathbf{B}$ when $\mathbf{B}$ is a column matrix and is equal to $|\mu_1(\mathbf{B})|$ when $\mathbf{B}$ is square and Hermitian. Recall that, if $\mathbf{B}_1$ is $s_1 \times s_2$ and $\mathbf{B}_2$ is $s_2 \times s_3$, then

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

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

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

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

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

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

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

(i) $\max_{|h| \leq B_T} n^{-1}\|\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^x\| = O_P(\rho_T^{-1/2})$;

(ii) $\max_{|h| \leq B_T} n^{-1/2}\|\mathcal{S}_i'(\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^x)\| = O_P(\rho_T^{-1/2})$ *uniformly in* i;

(iii) $\max_{|h| \leq B_T} n^{-1}\|\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi\| = O_P(\max(n^{-1}, \rho_T^{-1/2}))$;

(iv) $\max_{|h| \leq B_T} n^{-1/2}\|\mathcal{S}_i'(\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi)\| = O_P(\max(n^{-1/2}, \rho_T^{-1/2})) = O_P(\zeta_{nT})$ *uniformly in* i.

Proof. We have

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

Using (3.4) and the Markov inquality,

$$n^{-2} \max_{|h| \leq B_T} \sum_{i=1}^n \sum_{j=1}^n |\hat{\sigma}_{ij}^x - \sigma_{ij}^x|^2 \leq n^{-2} \sum_{i=1}^n \sum_{j=1}^n \max_{|h| \leq B_T} |\hat{\sigma}_{ij}^x - \sigma_{ij}^x|^2 \leq C\rho_T^{-1}.$$

Statement (i) follows. In the same way,

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

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

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

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). □

Lemma 2 *As $T \to \infty$ and $n \to \infty$,*

(i) $\max_{|h|\leq B_T} n^{-1}\left|\hat{\lambda}_f^x - \lambda_f^\chi\right| = O_P(\max(n^{-1}, \rho_T^{-1/2}))$ *for* $f = 1, 2, \ldots, q$*;*

(ii) letting

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

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

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

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

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

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

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

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

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

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

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

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 $O_P(\max(n^{-1}, \rho_T^{-1/2}))$, uniformly in θ.

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

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

The conclusion follows. □

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

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

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

uniformly in i.

Proof. We have

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

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

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

Firstly, we show that

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

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

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

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

and

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

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

Proposition 7 now follows from

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

and

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

□

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

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

and the result of Lemma 3(iii) as

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

In the same way, we drop $\hat{\mathbf{W}}_q$ in Lemmas 6, 7, 8, though not in the conclusion of Appendix D, nor in Appendix E.

C Proof of Proposition 9

To start with, note that, as the extreme right-hand side in (3.6) contains the term

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

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

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

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

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 key result. All lemmas in this Appendix hold, and are proved under Assumptions 1 through 10.

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

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

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

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

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

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

$$\begin{aligned} P(\max_h |\hat{\sigma}^x_{ii}(\theta_h) - \sigma^x_{ii}(\theta_h)| \geq \eta) &\leq \eta^{-2} E\left(\max_{|h|\leq B_T} |\hat{\sigma}^x_{ii}(\theta^*_h) - \sigma^x_{ii}(\theta^*_h)|^2\right) \\ &\leq \eta^{-2} C(T^{-1} B_T \log B_T) \end{aligned}$$

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

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

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

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

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

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

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

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

PROOF. The left-hand side of (D.1) equals the left-hand side of (B.4) when $\mathcal{S}_i$ is replaced by $\mathbf{I}_n$. The proof goes along the same lines as that of Lemma 4. Firstly, $\|n^{1/2}\mathbf{P}^\chi\|$ is $O(n^{1/2})$.

Both $\|\mathbf{g}-\mathbf{h}\|$ and $\|\mathbf{h}-\mathbf{i}\|$ are $O_P(n^{1/2}\max(n^{-1},\rho_T^{-1/2}))$. As for $\|\mathbf{i}-\mathbf{j}\|$, the conclusion follows from Lemma 1 (iii). □

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

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

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

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

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

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

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

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

uniformly in θ.

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

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

uniformly in θ, as was to be shown. □

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

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

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

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

PROOF. We have

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

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

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

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

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

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

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

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

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

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

We are now able to state and prove the main lemma of this section. 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 $T \to \infty$,*

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

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

PROOF. Denote by $\check{\mathbf{\Gamma}}^z = \mathbf{Z}'\mathbf{Z}/T$ the empirical covariance matrix we would compute from the $\mathbf{Z}_{nt}$'s, 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).

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

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

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

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

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 and 7, and in view of 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) still 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 $\mathbf{a}^j_{s\alpha}$, $1 \leq \alpha \leq q+1$, the α-th column of $\mathbf{A}^{j'}_s$, we have

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

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

$$\begin{aligned}
&\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}((\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'\hat{\mathbf{\Gamma}}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta})^2 \\
&\qquad \leq \sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}((\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\hat{\mathbf{\Gamma}}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta} \\
&\qquad = \sum_{k=1}^{m}\sum_{\beta=1}^{q+1}\sum_{j=1}^{m}\sum_{\alpha=1}^{q+1}(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})'(\hat{\mathbf{a}}^j_{s\alpha} - \mathbf{a}^j_{s\alpha})\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\hat{\mathbf{\Gamma}}^x_{jk,s-r}\hat{\mathbf{a}}^k_{r\beta}
\end{aligned}$$

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

where

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

and

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

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 $\mathbf{g}_{s\delta}$ stand for 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,

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

$$\leq \quad 2^{3/2}(q+1)^{3/2}\big(\big[\sum_{j=1}^{m}\|\hat{\mathbf{B}}_j-\mathbf{B}_j\|^8\big]^{1/2}\big[\sum_{j=1}^{m}\|\hat{\mathbf{C}}_j^{-1}\|^8\big]^{1/2}$$
$$+\big[\sum_{j=1}^{m}\|\hat{\mathbf{C}}_j-\mathbf{C}_j\|^8\big]^{1/2}\big[\sum_{j=1}^{m}\|\hat{\mathbf{B}}_j\|^{16}\big]^{\frac{1}{4}}\big[\sum_{j=1}^{m}\|\hat{\mathbf{C}}_j^{-1}\|^{16}\|\mathbf{C}_j^{-1}\|^{16}\big]^{\frac{1}{4}}\big)^{1/2}.$$

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

$$\sum_{j=1}^{m}\|\hat{\mathbf{B}}_j-\mathbf{B}_j\|^8 \quad \leq \sum_{j=1}^{m}\big(\sum_{i=1}^{q+1}\sum_{\delta=1}^{S(q+1)}(\hat{b}_{i\delta}^j-b_{i\delta}^j)^2\big)^4$$
$$\leq (q+1)^6S^3\sum_{j=1}^{m}\sum_{i=1}^{q+1}\sum_{\delta=1}^{S(q+1)}(\hat{b}_{i\delta}^j-b_{i\delta}^j)^8=O_P(m(\max(n^{-1},\rho_T^{-1}))^4).$$

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

Collecting terms yields

$$\mathcal{A} \quad = \quad m(q+1)^{1/2}\big[\sum_{j=1}^{m}\sum_{\alpha=1}^{q+1}\big((\hat{\mathbf{a}}_{s\alpha}^j-\mathbf{a}_{s\alpha}^j)'(\hat{\mathbf{a}}_{s\alpha}^j-\mathbf{a}_{s\alpha}^j)\big)^2\big]^{1/2}$$
$$\leq \quad 2^{3/2}(q+1)^2m\big(\sum_{i=1}^{m}\|\hat{\mathbf{A}}_s^i-\mathbf{A}_s^i\|^4\big)^{1/2}=O_P\big(m^{3/2}\max(n^{-1},\rho_T^{-1})\big). \qquad \text{(D.8)}$$

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

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

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

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

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 $\|\hat{\mathbf{A}}_s\hat{\mathbf{\Gamma}}^x_{s-r}\hat{\mathbf{A}}'_r - \mathbf{A}_s\hat{\mathbf{\Gamma}}^x_{s-r}\mathbf{A}'_r\|$'s. The two terms on the right-hand side of (D.7) must be dealt with separately. In the first of those two terms, dropping one of the summations for $k = 1, \ldots, m$ and setting $k = i$,

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

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

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

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

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

$$\begin{aligned}
&\sum_{j=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}\Big((\hat{\mathbf{a}}^i_{s\alpha} - \mathbf{a}^i_{s\alpha})'\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\mathbf{a}^j_{r\beta}\Big)^2 \\
&\qquad \le m\big[\big[\sum_{\alpha=1}^{q+1}(\hat{\mathbf{a}}^i_{s\alpha} - \mathbf{a}^i_{s\alpha})'(\hat{\mathbf{a}}^i_{s\alpha} - \mathbf{a}^i_{s\alpha})\big]^2\big]^{1/2}\big[\frac{1}{m}\sum_{j=1}^{m}\sum_{\beta=1}^{q+1}(\mathbf{a}^{j\prime}_{r\beta}\hat{\mathbf{\Gamma}}^x_{ji,s-r}\hat{\mathbf{\Gamma}}^{x\prime}_{ji,s-r}\mathbf{a}^j_{r\beta})\big] \\
&\qquad = \mathcal{F}\mathcal{G}, \text{ say.}
\end{aligned}$$

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

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

Starting with Lemma 11, which plays here the same role as Proposition 6 does for the proof of Proposition 7, we can easily prove statements that replicate in this context Lemmas 1, 2, 3 and 4, using the same arguments as in Section B, with x, χ and ξ replaced by Z, Ψ and Φ, respectively. More precisely,

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

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

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

(IV) Lastly, Lemma 4 becomes

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

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

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

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

E Proof of Proposition 11

We have

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

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

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

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

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

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

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

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

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

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

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 $O_P(n^{-1/2})$. To conclude, note that the last term $\hat{\mathbf{W}}_q(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\mathbf{P}^\psi(\mathbf{\Lambda}^\psi)^{1/2}\mathbf{v}_t$ is equal to $\hat{\mathbf{W}}_q\mathbf{v}_t$. The conclusion follows. □

F Sections 4 and 5: Tables

Table 1: Model I, estimated impulse response functions and structural shocks. Average and standard deviation (in brackets) of normalized MSE across 500 data sets of 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	*FHLZ method , no averaging*							
30	0.93 (0.51)	0.54 (0.34)	0.32 (0.11)	0.22 (0.11)	0.84 (0.48)	0.59 (0.31)	0.48 (0.11)	0.45 (0.06)
60	0.68 (0.39)	0.41 (0.26)	0.25 (0.11)	0.17 (0.04)	0.55 (0.45)	0.38 (0.26)	0.30 (0.12)	0.28 (0.04)
120	0.58 (0.34)	0.34 (0.11)	0.22 (0.04)	0.15 (0.03)	0.38 (0.42)	0.23 (0.13)	0.18 (0.05)	0.16 (0.03)
240	0.55 (0.32)	0.33 (0.12)	0.21 (0.04)	0.15 (0.02)	0.31 (0.40)	0.17 (0.15)	0.11 (0.05)	0.10 (0.02)
n	*FHLZ method , with averaging*							
30	0.66 (0.31)	0.39 (0.19)	0.23 (0.06)	0.15 (0.04)	0.64 (0.30)	0.49 (0.18)	0.41 (0.07)	0.39 (0.05)
60	0.54 (0.29)	0.32 (0.15)	0.18 (0.05)	0.12 (0.03)	0.45 (0.32)	0.32 (0.16)	0.25 (0.05)	0.24 (0.03)
120	0.48 (0.24)	0.28 (0.10)	0.17 (0.04)	0.11 (0.02)	0.32 (0.29)	0.21 (0.11)	0.15 (0.04)	0.14 (0.02)
240	0.46 (0.23)	0.27 (0.09)	0.16 (0.04)	0.10 (0.02)	0.25 (0.29)	0.15 (0.10)	0.10 (0.04)	0.08 (0.02)
n	*static factor method (FGLR)*							
30	0.86 (0.45)	0.52 (0.20)	0.38 (0.07)	0.31 (0.05)	0.74 (0.45)	0.51 (0.20)	0.42 (0.08)	0.38 (0.06)
60	0.71 (0.38)	0.45 (0.13)	0.34 (0.05)	0.20 (0.03)	0.54 (0.44)	0.34 (0.15)	0.27 (0.05)	0.23 (0.04)
120	0.63 (0.32)	0.42 (0.09)	0.33 (0.04)	0.28 (0.03)	0.40 (0.38)	0.24 (0.11)	0.17 (0.05)	0.13 (0.03)
240	0.60 (0.28)	0.42 (0.10)	0.32 (0.03)	0.27 (0.02)	0.32 (0.34)	0.19 (0.13)	0.12 (0.04)	0.08 (0.02)

Table 2: Model I, one-step-ahead forecasts. Average and standard deviation (in brackets) of the normalized mean square deviation from the population forecasts across 500 data sets of different size. For the dynamic method, the number of dynamic factors is determined by Hallin and Liška's log criterion. For the static method, the number of static factors is determined by Bai and Ng's IC_{p2} criterion.

	$T = 60$	$T = 120$	$T = 240$	$T = 480$
FHLZ method, with averaging				
$n = 30$	0.97 (0.65)	0.91 (1.06)	0.73 (0.61)	0.74 (0.67)
$n = 60$	0.82 (0.32)	0.68 (0.35)	0.59 (0.95)	0.50 (0.35)
$n = 120$	0.74 (0.21)	0.58 (0.16)	0.47 (0.27)	0.39 (0.22)
$n = 240$	0.70 (0.18)	0.53 (0.14)	0.41 (0.14)	0.33 (0.14)
static factor method (SW), with lagged x's				
$n = 30$	2.58 (3.46)	1.65 (2.99)	1.12 (1.81)	0.89 (0.88)
$n = 60$	2.17 (2.22)	1.28 (1.00)	0.99 (2.31)	0.73 (0.61)
$n = 120$	1.94 (1.53)	1.16 (0.72)	0.83 (0.90)	0.64 (0.43)
$n = 240$	1.87 (1.51)	1.08 (0.62)	0.75 (0.47)	0.60 (0.35)
static factor method (SW), no lagged x's				
$n = 30$	1.90 (2.62)	1.33 (2.05)	0.94 (0.95)	0.80 (0.74)
$n = 60$	1.52 (1.54)	1.02 (0.75)	0.86 (1.84)	0.68 (0.54)
$n = 120$	1.32 (0.89)	0.89 (0.48)	0.72 (0.66)	0.61 (0.39)
$n = 240$	1.24 (0.69)	0.82 (0.41)	0.64 (0.38)	0.56 (0.33)

Table 3: Model II, estimated impulse response functions and structural shocks. Average and standard deviation (in brackets) of the normalized MSE across 500 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	*FHLZ method, with averaging*							
2	0.13 (0.05)	0.11 (0.05)	0.10 (0.05)	0.09 (0.07)	0.17 (0.08)	0.12 (0.07)	0.10 (0.06)	0.08 (0.07)
4		0.15 (0.09)	0.15 (0.11)	0.14 (0.15)		0.27 (0.15)	0.22 (0.16)	0.17 (0.17)
6			0.17 (0.09)	0.15 (0.10)			0.34 (0.13)	0.24 (0.16)
	FGLR method, r determined with IC_{p2}							
2	0.16 (0.14)	0.16 (0.13)	0.15 (0.13)	0.12 (0.14)	0.21 (0.25)	0.17 (0.26)	0.12 (0.19)	0.08 (0.13)
4		0.18 (0.15)	0.19 (0.16)	0.20 (0.23)		0.35 (0.26)	0.31 (0.28)	0.23 (0.28)
6			0.20 (0.13)	0.22 (0.15)			0.43 (0.22)	0.35 (0.25)
	FGLR method , r assumed known							
2	0.10 (0.07)	0.09 (0.07)	0.08 (0.07)	0.07 (0.06)	0.17 (0.14)	0.13 (0.13)	0.12 (0.10)	0.11 (0.06)
4		0.14 (0.12)	0.15 (0.14)	0.14 (0.19)		0.28 (0.22)	0.25 (0.23)	0.21 (0.23)
6			0.18 (0.13)	0.17 (0.14)			0.38 (0.21)	0.28 (0.21)

Table 4: Model II, one-step-ahead forecasts. Average and standard deviation (in brackets) of the normalized mean square deviations from the population forecasts, across 500 data sets with different configurations of static and dynamic factors. For the dynamic method, the number of dynamic factors is determined by Hallin and Liška's log criterion. For the static method, the number of static factors is determined by Bai and Ng's IC_{p2} criterion.

	$r = 4$	$r = 6$	$r = 8$	$r = 12$
FHLZ method, with averaging				
$q = 2$	0.79 (1.59)	0.68 (0.75)	0.59 (0.97)	0.56 (0.52)
$q = 4$		0.44 (0.36)	0.44 (0.28)	0.40 (0.20)
$q = 6$			0.40 (0.28)	0.38 (0.18)
static factor method (SW), no lagged x's				
$q = 2$	1.00 (2.10)	0.67 (1.04)	0.52 (64)	0.49 (0.66)
$q = 4$		0.61 (1.37)	0.53 (0.67)	0.43 (0.37)
$q = 6$			0.50 (0.58)	0.42 (0.34)

Table 5: Relative Mean Square Forecast Errors, $AR(4) = 1$, for five macroeconomic variables, obtained with the FHLZ method and the static factor method (with p lags of the dependent variable), with a quarterly US macroeconomic data set. Best forecasts are in boldface. The Diebold-Mariano test statistic is computed for all factor model specifications with respect to the AR(4) benchmark; its significance levels (p-values) are indicated by means of superscripts $^{+}$ (15 %), * (10 %), and ** (5 %), respectively.

Forecasting	FHLZ (with averaging)			Static Method (SW)				AR(4)
horizon	$q = 3$	$q = 3$	$q = 4$	$r = 2$	$r = 4$	$r = 6$	$r = 4$	
	$B_T = 8$	$B_T = 12$	$B_T = 8$	$p = 0$	$p = 0$	$p = 0$	$p = 1$	
Real Gross Domestic Product								
$h = 1$	0.920^{+}	0.923^{+}	$\mathbf{0.910}^{+}$	0.979	0.955	0.952	1.018	1.000
$h = 2$	0.883^{+}	0.890	**0.882**	1.021	0.990	1.060	1.012	1.000
$h = 3$	$\mathbf{0.829}^{*}$	0.835^{*}	0.836^{*}	1.026	0.901^{+}	1.000	0.979	1.000
$h = 4$	$\mathbf{0.825}^{*}$	0.830^{*}	0.839^{*}	0.993	0.881^{+}	1.023	1.040	1.000
Real Private Fixed Investment growth								
$h = 1$	0.836	0.819	0.867	0.724^{**}	0.682^{**}	0.773^{*}	$\mathbf{0.665}^{**}$	1.000
$h = 2$	0.708^{+}	0.707^{+}	0.737	0.784^{+}	0.700^{*}	0.787	$\mathbf{0.688}^{*}$	1.000
$h = 3$	**0.678**	0.683	0.709	0.806	0.721	0.818	0.774^{+}	1.000
$h = 4$	**0.689**	0.696	0.722	0.836	0.743	0.865	0.869^{+}	1.000
Real Personal Consumption Expenditures								
$h = 1$	1.109	1.099	1.136	1.188	1.239	1.223	1.244	**1.000**
$h = 2$	1.136	1.134	1.188	1.277	1.239	1.262	1.150	**1.000**
$h = 3$	1.137	1.140	1.202	1.327	1.269	1.327	1.271	**1.000**
$h = 4$	1.037	1.040	1.093	1.175	1.143	1.275	1.223	**1.000**
Number of Unemployed								
$h = 1$	0.877^{+}	0.855^{*}	0.872^{+}	0.823^{**}	0.825^{+}	0.901	$\mathbf{0.807}^{*}$	1.000
$h = 2$	0.870	**0.859**	0.870	0.980	0.925	0.997	0.901	1.000
$h = 3$	0.841^{*}	$\mathbf{0.837}^{*}$	0.845^{*}	1.032	0.906	1.032	0.936	1.000
$h = 4$	0.836^{**}	$\mathbf{0.835}^{**}$	0.839^{**}	0.985	0.852^{**}	1.001	0.942	1.000
Consumer Price Index								
$h = 1$	0.844	0.846	$\mathbf{0.823}^{+}$	0.885	0.979	1.032	1.138	1.000
$h = 2$	0.771	0.774	$\mathbf{0.754}^{+}$	0.812	0.863	0.935	0.861	1.000
$h = 3$	0.685^{+}	0.696^{+}	$\mathbf{0.691}^{+}$	0.714^{+}	0.761	0.896	0.855^{+}	1.000
$h = 4$	0.753^{+}	0.759^{+}	$\mathbf{0.746}^{+}$	0.797	0.778	0.937	0.788	1.000

Table 6: Relative Mean Square Forecast Errors, $AR(4) = 1$, for seven macroeconomic variables, and average RMSFE for NIPA variables 1-9, obtained with the FHLZ method and the static factor method (with p lags of the dependent variable), with a quarterly US macroeconomic data set. Best forecasts are in boldface. The Diebold-Mariano test statistic is computed for all factor model specifications with respect to the AR(4) benchmark (not computed for the average RMSFE of NIPA variables); its significance levels (p-values) are indicated by means of superscripts $^{+}$ (15 %), * (10 %), and ** (5 %), respectively (no superscripts means a p-value greater than 15 %).

Forecasting	FHLZ (with averaging)			Static Method (SW)				AR(4)
horizon	$q = 3$	$q = 3$	$q = 4$	$r = 2$	$r = 4$	$r = 6$	$r = 4$	
	$B_T = 8$	$B_T = 12$	$B_T = 8$	$p = 0$	$p = 0$	$p = 0$	$p = 1$	
Real Government Consumption Expenditures & Gross Investment								
$h = 1$	1.042	1.042	1.063	1.172	1.242	1.171	1.264	**1.000**
$h = 2$	1.109	1.107	1.126	1.354	1.528	1.364	1.454	**1.000**
$h = 3$	1.064	1.065	1.082	1.538	1.712	1.463	1.502	**1.000**
$h = 4$	**0.966**	0.966	0.980	1.605	1.761	1.516	1.502	1.000
Real Exports of Goods & Services								
$h = 1$	0.973	0.973	**0.948**	1.027	1.029	1.024	1.001	1.000
$h = 2$	0.916	0.923	**0.892**$^{+}$	1.124	1.057	1.067	0.972	1.000
$h = 3$	0.900	0.910	**0.885**$^{+}$	1.130	1.089	1.123	1.074	1.000
$h = 4$	0.883^{+}	0.891^{+}	**0.871***	1.078	0.972	1.024	1.031	1.000
Real Imports of Goods & Services								
$h = 1$	0.651*	0.642*	0.655*	0.561**	0.507**	0.541**	**0.499****	1.000
$h = 2$	0.553**	0.559**	0.553**	0.636*	**0.513***	0.568*	0.523**	1.000
$h = 3$	**0.595****	0.604**	0.596**	0.720*	0.628*	0.693^{+}	0.708*	1.000
$h = 4$	**0.692**$^{+}$	0.702^{+}	0.707^{+}	0.857	0.729	0.826	0.854^{+}	1.000
Nonfarm Business Sector: Output Per Hour of All Persons								
$h = 1$	0.908*	0.906*	0.898**	0.929	0.896	**0.834****	0.897	1.000
$h = 2$	0.893*	0.890*	0.892*	0.884	0.873	**0.846**$^{+}$	0.878	1.000
$h = 3$	0.832*	0.828*	0.839*	0.815	0.788	**0.728***	0.802	1.000
$h = 4$	0.832*	0.827*	0.845^{+}	0.777	**0.770**	0.793^{+}	0.813	1.000

Table 6 (continued)

Forecasting	FHLZ (with averaging)			Static Method (SW)				AR(4)
horizon	$q=3$	$q=3$	$q=4$	$r=2$	$r=4$	$r=6$	$r=4$	
	$B_T=8$	$B_T=12$	$B_T=8$	$p=0$	$p=0$	$p=0$	$p=1$	
Utilization-Adjusted Total Factor Productivity								
$h=1$	0.917^{*}	0.919^{*}	0.914^{**}	0.965	0.918	**0.894***	0.900^{*}	1.000
$h=2$	0.914^{**}	0.917^{**}	**0.913****	0.958^{+}	0.918^{**}	0.974	0.953^{*}	1.000
$h=3$	0.869^{**}	**0.865****	0.866^{**}	0.887^{**}	0.932^{*}	0.892^{*}	0.942^{+}	1.000
$h=4$	0.929^{**}	0.927^{**}	0.925^{**}	1.004	**0.905****	0.927^{+}	0.934^{+}	1.000
Industrial Production Index								
$h=1$	1.085	1.069	1.037	0.900	0.879	**0.868**	0.895	1.000
$h=2$	0.933	0.930	**0.902**	1.028	0.951	1.064	0.928	1.000
$h=3$	0.870	0.873	**0.851**	1.092	0.950	1.103	0.968	1.000
$h=4$	0.841^{+}	0.845^{+}	0.832^{+}	1.012	**0.821****	0.999	0.937	1.000
Effective Federal Funds Rate								
$h=1$	0.884	0.876	**0.850**$^{+}$	1.273	1.560	1.629	1.582	1.000
$h=2$	0.882^{+}	0.872^{*}	**0.844***	1.469	1.802	1.887	1.637	1.000
$h=3$	0.920^{*}	0.914^{*}	**0.881****	1.575	1.855	2.016	1.746	1.000
$h=4$	0.910^{**}	0.906^{**}	**0.874****	1.460	1.661	1.816	1.614	1.000
Standard & Poor's Index of 500 Common Stocks								
$h=1$	0.917^{*}	0.912^{**}	**0.891****	1.070	1.071	1.126	1.073	1.000
$h=2$	0.824^{+}	0.823^{*}	**0.810***	1.156	1.178	1.254	1.180	1.000
$h=3$	0.782^{*}	0.784^{*}	**0.781***	1.248	1.259	1.374	1.363	1.000
$h=4$	0.768^{*}	0.772^{*}	**0.770***	1.248	1.258	1.355	1.508	1.000
Average of NIPA variables 1-9								
$h=1$	0.923	0.915	0.923	**0.900**	0.905	0.915	0.919	1.000
$h=2$	**0.879**	0.880	0.887	0.970	0.944	0.975	0.925	1.000
$h=3$	**0.883**	0.887	0.895	1.056	1.004	1.052	1.023	1.000
$h=4$	**0.875**	0.879	0.889	1.066	1.008	1.091	1.084	1.000

G Data description

Transformation codes. 1 = levels; 2 = first differences of the original series; 4 = logs of the original series; 5 = first differences of the logs of the original series; 6 = second differences of the original series.

Data sources. Data on total factor productivity (series 49-50 and 57-60) are taken from Fernald's web site; the components of the consumer confidence indicator (53-56) are downloaded from the web site of the University of Michigan; series 41-47 and 61 are from Datastream; series 62-73 are taken from the Survey of Professional Forecast website; all other series are taken from the FRED data base.

Further details. Several series are provided only with monthly frequency; in such cases we computed the corresponding quarterly series by taking simple averages. For the SPF variables 62-73, we discarded the 4-quarter ahead forecasts since they have missing values until 1974. We did not include investment, consumption and CPI forecasts since they are only available from 1981:Q3. The target series for the forecasting exercise are 1, 3, 4, 26 and 36. After transformation, we corrected for outliers as in Stock and Watson (2002b).

no.series	Transf.	Mnemonic	Long Label
1	5	GDPC1	Real Gross Domestic Product
2	5	OUTNFB	Nonfarm Business Sector: Output
3	5	FPIC1	Real Private Fixed Investment
4	5	PCECC96	Real Personal Consumption Expenditures
5	5	PCDGCC96	Real Personal Consumption Expenditures: Durable Goods
6	5	FGCEC1	Real Federal Consumption Expenditures & Gross Investment
7	1	CBIC1	Real Change in Private Inventories
8	5	EXPGSC1	Real Exports of Goods & Services
9	5	IMPGSC1	Real Imports of Goods & Services
10	5	CP/GDPDEF	Corporate Profits After Tax/GDP deflator
11	5	CNCF/GDPDEF	Corporate Net Cash Flow/GDP deflator
12	5	HOANBS	Nonfarm Business Sector: Hours of All Persons
13	5	OPHNFB	Nonfarm Business Sector: Output Per Hour of All Persons
14	5	ULCNFB	Nonfarm Business Sector: Unit Labor Cost
15	5	COMPRNFB	Nonfarm Business Sector: Real Compensation Per Hour
16	5	GDPCTPI	Gross Domestic Product: Chain-type Price Index
17	5	GDPDEF	Gross Domestic Product: Implicit Price Deflator
18	5	INDPRO	Industrial Production Index
19	5	IPBUSEQ	Industrial Production: Business Equipment
20	5	IPCONGD	Industrial Production: Consumer Goods
21	5	IPMAT	Industrial Production: Materials
22	5	CE16OV	Civilian Employment
23	2	AWOTMAN	Average Weekly Hours: Overtime: Manufacturing
24	5	UNEMPLOY	Unemployed

25	2	UEMPMEAN	Average (Mean) Duration of Unemployment
26	2	UNRATE	Civilian Unemployment Rate
27	4	HOUST	Housing Starts: Total: New Privately Owned Housing Units Started
28	2	FEDFUNDS	Effective Federal Funds Rate
29	2	TB3MS	3-Month Treasury Bill: Secondary Market Rate
30	2	GS10	10-Year Treasury Constant Maturity Rate
31	2	BAA	Moody's Seasoned Baa Corporate Bond Yield
32	6	M1SL	M1 Money Stock
33	6	M2SL	M2 Money Stock
34	6	BUSLOANS	Commercial and Industrial Loans at All Commercial Banks
35	6	CONSUMER	Consumer (Individual) Loans at All Commercial Banks
36	6	CPIAUCSL	Consumer Price Index For All Urban Consumers: All Items
37	6	CPILFESL	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy
38	6	PPICPE	Producer Price Index Finished Goods: Capital Equipment
39	6	PPIFCG	Producer Price Index: Finished Consumer Goods
40	6	PPIFGS	Producer Price Index: Finished Goods
41	5	USSHRPRCF	US Dow Jones Industrial Share Price Index
42	5	US500STK	US Standard & Poor's Index Of 500 Common Stocks
43	5	USNOIDN.D	Us Manufacturers New Orders For Nondefense Capital Goods
44	5	USCNORCGD	Us New Orders Of Consumer Goods & Materials (Bci 8) Cona
45	1	USNAPMNO	Us Ism Manufacturers Survey: New Orders Index Sadj
46	5	USCYLEAD	Us The Conference Board Leading Economic Indicators Index Sadj
47	5	USECRIWLH	Us Economic Cycle Research Institute Weekly Leading Index(Monthl
48	5	GCEC1	Real Government Consumption Expenditures & Gross Investment
49	1	dtfputil	Utilization-adjusted TFP
50	1	dtfp	TFP business sector
51	5	USSHRPRCF/GDPDEF	Deflated Dow Jones Index
52	5	US500STK/GDPDEF	Deflated S&P500 Index
53	1	ECFS	Expected Change in Financial Situation
54	1	BC12M	Business conditions expected during the next 12 months
55	1	BC5Y	Business Conditions expected during the next 5 years
56	1	EI	Expected Index
57	1	dtfpI	TFP in equipment and consumer durables
58	1	dtfpC	TFP in non-equipment business output ("consumption")
59	1	dtfpIutil	Utilization-adjusted TFP in equipment and consumer durables
60	1	dtfpCutil	Utilization-adjusted TFP in non-equipment output
61	5	USI62...F	US Share Price Index
62	1	SPFGDP1	SPF GDP growth 1 quarter ahead (mean)
63	1	SPFGDP2	SPF GDP growth 2 quarters ahead (mean)
64	1	SPFGDP3	SPF GDP growth 3 quarters ahead (mean)
65	2	SPFUNRATE1	SPF unemployment rate 1 quarter ahead (mean)
66	2	SPFUNRATE2	SPF unemployment rate 2 quarters ahead (mean)
67	2	SPFUNRATE3	SPF unemployment rate 3 quarters ahead (mean)
68	2	SPFDEF1	SPF price deflator growth for GDP 1 quarter ahead (mean)
69	2	SPFDEF2	SPF price deflator growth for GDP 2 quarters ahead (mean)
70	2	SPFDEF3	SPF price deflator growth for GDP 3 quarters ahead (mean)
71	1	SPFIP1	SPF IP change 1 quarter ahead (mean)
72	1	SPFIP2	SPF IP change 2 quarters ahead (mean)
73	1	SPFIP3	SPF IP change 3 quarters ahead (mean)

RECent Working Papers Series

The 10 most RECent releases are:

No. 115 DYNAMIC FACTOR MODELS WITH INFINITE-DIMENSIONAL FACTOR SPACE: ASYMPTOTIC ANALYSIS (2015)
M. Forni, M. Hallin, M. Lippi and P. Zaffaroni

No. 114 ECONOMIC SHOCKS AND THEIR EFFECTS ON UNEMPLOYMENT IN THE EURO AREA PERIPHERY UNDER THE EMU (2015)
P. Dallari and A. Ribba

No. 113 THE EDUCATION NETWORKS OF LATIN AMERICA. EFFECTS ON TRADE DURING AND AFTER THE COLD WAR (2015)
M. Murat, M. L. Recalde and P. G. Degiovanni

No. 112 THE LONG-TERM DETERMINANTS OF FEMALE HIV INFECTION IN AFRICA: THE SLAVE TRADE, POLYGYNY, AND SEXUAL BEHAVIOR (2015)
G. Bertocchi and A. Dimico

No. 111 WHAT DRIVES US INFLATION AND UNEMPLOYMENT IN THE LONG RUN? (2015)
Antonio Ribba

No. 110 ON THE HISTORICAL ROOTS OF WOMEN'S EMPOWERMENT ACROSS ITALIAN PROVINCES: RELIGION OR FAMILY CULTURE? (2015)
M. Bozzano

No. 109 VOLATILITY RISK PREMIA AND FINANCIAL CONNECTEDNESS (2014)
A. Cipollini, I. Lo Cascio and S. Muzzioli

No. 108 SOCIAL COORDINATION WITH LOCALLY OBSERVABLE TYPES (2014)
E. Bilancini and L. Boncinelli

No. 107 SOFT, HARD OR SMART POWER? INTERNATIONAL STUDENTS AND INVESTMENTS ABROAD (2014)
M. Murat

No. 106 LIMIT PRICING AND SECRET BARRIERS TO ENTRY (2014)
L. Brighi and M. D'Amato

The full list of available working papers, together with their electronic versions, can be found on the RECent website: http://www.recent.unimore.it/workingpapers.asp