SFB
823

One-sided representations of generalized dynamic factor models

Mario Forni, Marc Hallin, Marco Lippi,
Paolo Zaffaroni

Discussion Paper

Nr. 55/2011

SFB
823

One-Sided Representations of Generalized Dynamic Factor Models

Mario Forni

Università di Modena e Reggio Emilia, CEPR and RECent

Marc Hallin*†

ECARES, Université Libre de Bruxelles and ORFE, Princeton University

Marco Lippi

Università di Roma La Sapienza and EIEF

Paolo Zaffaroni‡

Imperial College London and Università di Roma La Sapienza

July 28, 2011.

Abstract. Factor model methods recently have become extremely popular in the theory and practice of large panels of time series data. Those methods rely on various *factor models* which all are particular cases of the *Generalized Dynamic Factor Model* (GDFM) introduced in Forni, Hallin, Lippi and Reichlin (2000). In that paper, however, estimation relies on Brillinger's concept of *dynamic principal components*, which produces filters that are in general two-sided and therefore yield poor performances at the end of the observation period and hardly can be used for forecasting purposes. In the present paper, we remedy this problem, and show how, based on recent results on singular stationary processes with rational spectra, one-sided estimators are possible for the parameters and the common shocks in the GDFM. Consistency is obtained, along with rates. An empirical section, based on US macroeconomic

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

†Académie Royale de Belgique, ECORE, and CentER, Tilburg University.

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

time series, compares estimates based on our model with those based on the usual static-representation restriction, and provide convincing evidence that the assumptions underlying the latter are not supported by the data.

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

1 Introduction

1.1 Dynamic factor models

Large-dimensional factor model methods can be traced back to two seminal papers by Chamberlain (1983) and Chamberlain and Rothschild (1983). The fastly growing literature on the subject, however, is starting with the contributions by Forni *et al.* (2000), Forni and Lippi (2001), Stock and Watson (2002a,b), Bai and Ng (2002) and Bai (2003). Fostered by their success in applications, factor model methods since then have attracted considerable attention. The recent literature in the area is so abundant that a complete review is impossible here, and we restrict ourselves to a short and unavoidably somewhat subjective selection of "representative" references. Applications include (a) forecasting (Stock and Watson 2002a and b, Forni *et al.* 2005, Boivin and Ng, 2006), (b) business cycle indicators and nowcasting (Cristadoro *et al.*, 2005, Giannone *et al.*, 2008, Altissimo *et al.* 2010), (c) structural macroeconomic analysis and monetary policy (Bernanke and Boivin, 2003, Bernanke *et al.* 2005, Stock and Watson, 2005, Giannone *et al.*, 2005, Favero *et al.*, 2005, Eickmeier, 2007, Forni *et al.*, 2009, Boivin *et al.*, 2009, Forni and Gambetti, 2010b), (d) the analysis of financial markets (Corielli and Marcellino, 2006, Ludvigson and Ng (2007 and 2009), Hallin *et al.*, 2011), to quote only a few.

Apart for some minor features, most factor models considered in the literature are particular cases of the so-called *Generalized Dynamic Factor Model* (GDFM) introduced in Forni *et al.* (2000). Consider a countable set $\{x_{it}\}$, $i \in \mathbb{N}$ of observable stationary stochastic processes: the GDFM relies on a decomposition of the form

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

where $\mathbf{u}_t = (u_{1t}\ u_{2t}\ \cdots\ u_{qt})'$ is a q-dimensional orthonormal unobservable white noise vector and $b_{if}(L), i \in \mathbb{N},\ f = 1, \ldots, q$ are square-summable filters (L, as usual, stands for the lag operator). The basic assumptions are

(A1) $\mathbf{u}_t$ is orthogonal to $\xi_{i,t-k}$ for all $i \in \mathbb{N}$ and $k \in \mathbb{Z}$;

(A2) cross-covariances among the ξ_{it}'s are "weak".

By "weak", we mean that, while some cross-covariance among the ξ's is allowed, all sequences of weighted cross-sectional averages of the form $\sum_{i=1}^n w_{ni}\xi_{it}$ such that $\lim_{n\to\infty}\sum_{i=1}^n w_{ni}^2 = 0$ tend to zero in mean square as $n \to \infty$ (the sequence of arithmetic averages $n^{-1}\sum_{i=1}^n \xi_{it}$ being a particular case).[1] Note that $\mathrm{E}(\xi_{it}^2) \leq M$ for all i and $\mathrm{E}(\xi_{it}\xi_{jt}) = 0$ for all $i \neq j$, is sufficient, but not necessary for (A2) to hold (we refer to Section 2 for a detailed presentation and discussion). Being mildly cross-correlated, the ξ_{it}'s are called *idiosyncratic*, while the χ_{it}'s are called *common*. The model implies that that cross-covariances among the observable variables x_{it}, essentially, is accounted for by the *common components* χ_{it}, the latter being driven by the small-dimensional vector of *common shocks* u_{ft}, $f = 1, 2 \ldots, q$.

The problem consists in recovering the unobserved common and idiosyncratic components χ_{it} and ξ_{it}, the common shocks $\mathbf{u}_t$ and the filters $b_{if}(L)$, from a finite realization ($i = 1, \ldots, n;\ t = 1, \ldots, T$) of the process $\{x_{it}\}$. The main tool so far has been a *principal component analysis* (PC) of the variables x_{it}, either standard or in the frequency domain (Brillinger's concept of *dynamic principal components*), depending on the assumptions made. The results obtained can be summarized as follows.

(i) Most authors assume that, denoting by $\overline{\mathrm{span}}(\ \ldots\ )$ the space generated by a collection of random variables,[2] $\overline{\mathrm{span}}(\chi_{it},\ i \in \mathbb{N})$, for given t, has finite dimension r,

[1] *Weak* cross-covariance among the ξ's, as opposed to cross-sectional orthogonality (that is, the much stronger assumption of no cross-covariances at all), is the reason for using the term "generalized" in the denomination of the GDFM. It constitutes a major difference with respect to the dynamic factor models studied in Sargent and Sims (1977), Geweke (1977), Quah and Sargent (1993), which, being based on a finite number n of equations of the form (1.1), require strict cross-sectional orthogonality.

[2] More precisely, $\overline{\mathrm{span}}(\zeta_i,\ i \in \mathbb{N})$, where ζ_i belongs to the Hilbert space of square-summable random variables defined over some probability space, equipped with the corresponding L^2 norm, is the closed Hilbert space of all mean-square convergent linear combinations of the ζ_i's and limits of convergent sequences thereof.

where $r \geq q$. Under that assumption, model (1.1) can be rewritten as

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

In this case, we say that the GDFM admits a *static representation.* Criteria to determine r consistently are given in Bai and Ng (2002) (see also Alessi et al. 2010). The vectors $\mathbf{F}_t$ and the loadings λ_{ij} can be estimated consistently using the first r standard principal components, see Stock and Watson (2002a,b), Bai and Ng (2002). Moreover, the second equation in (1.2) is usually specified as a singular VAR, so that (1.2) becomes

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

where the matrices $\mathbf{D}_j$ are $r \times r$ while $\mathbf{K}$ is $r \times q$. Under (1.3), Bai and Ng (2007) and Amengual and Watson (2007) provide consistent criteria to determine q. VAR estimation, and therefore, up to multiplication by an orthogonal matrix, estimation of $\mathbf{u}_t$ in (1.3) is standard.

(ii) Using the frequency-domain principal components (Brillinger 1981), and without any finite-dimensional assumption of the form (1.2), Forni et al. (2000) obtain an estimator of the spectral density of the common components χ_{it} and show how to consistently recover the common components themselves. Criteria to determine q without assuming (1.2) or (1.3) are obtained in Hallin and Liška (2007) and Onatski (2009). Unfortunately, frequency-domain principal components produce estimators of the χ_{it}'s that are based two-sided filters, which hence cannot be used at the end of the sample or for prediction.

Due to that unpleasant two-sidedness feature, the GDFM is seldom considered in practice, and finite-dimensional structure assumptions like (1.2) or (1.3) are made with almost no exception.[3] The moot point is that such assumptions are far from being innocuous, and, in many cases, are not supported by the data. For instance, (1.2) is so restrictive that even the very elementary model

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

[3]Some papers (see e.g. Forni et al., 2005, Altissimo et al., 2010) use the spectral density principal component approach in combination with finite-dimensional assumptions as in (1.2) or (1.3).

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

An analysis based on (1.2) or (1.3) then can be extremely misleading (spuriously identifying an infinite number of factors, etc.). This is a strong motivation for solving the one-sidedness issue in the GDFM. This is the objective of the present paper.

1.2 Outline of the paper

Instead of finite-dimensional assumptions of the form (1.2) or (1.3), we impose the much milder condition that the common components have a *rational spectral density*, that is, each filter $b_{if}(L)$ in (1.1) is a ratio of polynomials in L. [4] Under that assumption, we construct one-sided estimators for the common components χ_{it}, the common shocks $\mathbf{u}_t$, and the corresponding filters $b_{if}(L)$. Such estimators are then applied in an empirical investigation based on US quarterly macroeconomic data. We find that our method outperforms the standard PC estimator, which is based on assumption (1.2), both for the matrices $\mathbf{A}^k(L)$, the common components and the common shocks $\mathbf{u}_t$. Thus, assumption (1.2) is not supported by the US macroeconomic dataset we use. We believe that this provides strong empirical motivation for the present research. Let us give a detailed description of the construction leading to our estimator.

(A) Population results. Our assumption that the common components have rational spectral density implies, for the common components χ_{it}, the representation

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

where

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

(the degrees s_1 and s_2 of the polynomials are assumed to be independent of i for the sake of simplicity, but this is a minor point). As for the idiosyncratic components we do not make any parametric assumptions, nor do we restrict their cross-covariance

[4] Under that assumption, the dimension of $\overline{\text{span}}(\chi_{it}\ i \in \mathbb{N})$ is infinite, apart from a set of values of the coefficients of the polynomials defining the rational filters lying in negligible subsets (subsets that are, roughly speaking, lower-dimensional; see Section 2 for a formal definition).

structure—except of course for "weakness", as described above. Our model, in that sense, is a semiparametric one, with a huge nuisance; in particular, the autocorrelation structures of idiosyncratic components remain completely unspecified. We show that for generic values of the parameters $c_{if,k}$ and $d_{if,k}$ (i.e. apart from a subset that is negligible, in a sense to be specified in Section 2), the infinite-dimensional idiosyncratic vector $\boldsymbol{\chi}_t = (\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{nt}\ \cdots)'$ has an autoregressive representation with block structure, of the form

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

where $\mathbf{A}^k(L)$ is a $(q+1)\times(q+1)$ polynomial matrix *with finite degree* and $\mathbf{R}^k$ is $(q+1)\times q$. Denoting by $\underline{\mathbf{A}}(L)$ and $\underline{\mathbf{R}}$ the (infinite) matrices on the left- and right-hand sides of (1.6), and defining $\mathbf{x}_t$ and $\boldsymbol{\xi}_t$ in analogy with $\boldsymbol{\chi}_t$, we obtain

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

which is a factor model for $\underline{\mathbf{A}}(L)\mathbf{x}_t$, with a static representation of the form (1.2), playing a crucial role in the estimation of $\mathbf{u}_t$. Some features of (1.6) deserve some further comments:

(i) Because the infinite-dimensional vector $\boldsymbol{\chi}_t$ is driven by the q-dimensional white noise $\mathbf{u}_t$, for generic values of the parameters we can invert the infinite-dimensional moving average representation (1.5) *piecewise*, by partitioning $\boldsymbol{\chi}_t$ into the $(q+1)$-dimensional subvectors $(\chi_{1t}\ \chi_{2t}\ \dots\ \chi_{q+1,t})$, $(\chi_{q+2,t}\ \chi_{q+3,t}\ \cdots\ \chi_{2(q+1),t})$, $\cdots$ (see Forni and Lippi, 2010).

(ii) For generic values of the parameters, each of the subvectors, whose dimension and rank are $(q+1)$ and q, respectively, has a *finite-order* autoregressive representation. This is an application of a general result obtained in Anderson and Deistler (2008a and b) for rational-spectrum vector stochastic processes that are singular (i.e. with reduced-rank spectral density for all $\theta \in [-\pi\ \pi]$). We contribute to this literature showing that when the dimension is equal to $q+1$ the minimum-lag autoregressive representation is generically unique.

(iii) Under the assumption that the degrees of the VAR matrices $\mathbf{A}^k(L)$ are bounded, the number of VAR coefficients grows at rate n, not n^2. Moreover, each matrix $\mathbf{A}^k(L)$ can be estimated independently of the others.

(B) Estimation results. The spectral density of the common components can be consistently estimated by using the first q frequency-domain principal components (see Forni et al., 2000). Using such spectral density, we obtain consistent estimators of the autocovariance functions of the common components, which in turn are used to estimate the matrices $\mathbf{A}^k(L)$ and $\mathbf{R}^k$. Lastly, once the matrices $\mathbf{A}^k(L)$ have been estimated, we use equation (1.7) to estimate $\mathbf{u}_t$. As already observed, (1.7) has a static factor representation, so that standard principal components are the appropriate tool. However, the matrices $\mathbf{A}^k(L)$ must be replaced by their estimates, this implying considerable complications in the proof of consistency (see Section 3.3). For the entries of the matrices $\mathbf{A}^k(L)$ and $\mathbf{R}^k$, and the components of $\mathbf{u}_t$, we obtain the consistency rate $\max(n^{-1/2}, \rho_T^{-1/2})$, where ρ_T is any sequence with divergence slower than $T^{2/3}/\log T$ (as both n and T go to infinity. That ρ_T is the toll to be paid for using non-parametric spectral estimation. However, our empirical exercise provides evidence that the general dynamic nature of our model can offset a lower speed of consistency, as compared to the rate $T^{1/2}$ that can be obtained with model (1.3). Alternative estimators for the matrices $\mathbf{A}^k(L)$ and $\mathbf{R}^k$ are briefly discussed in Section 3.5.

The body of the paper contains detailed discussion and motivation of the main assumptions. Longer proofs are collected in the Appendix. The population and estimation results are derived in Sections 2 and 3, respectively. The empirical results are presented and discussed in Section 4. Section 5 concludes.

2 Main assumptions and population results

2.1 Notation

The GDFM (1.1) throughout can be thought of as (i) a double-indexed stochastic process $\{x_{it},\ i \in \mathbb{N},\ t \in \mathbb{Z}\}$, (ii) a family of stationary processes $\{x_{it},\ t \in \mathbb{Z}\}$ indexed by $i \in \mathbb{N}$, or (iii) a family of cross-sections $\{x_{it},\ i \in \mathbb{N}\}$ indexed by $t \in \mathbb{Z}$, i.e. a process of infinite-dimensional stochastic vectors. We find the third option convenient, and accordingly write $\mathbf{x}_t$ for $(x_{1t}\ x_{2t}\ \cdots\ x_{nt}\ \cdots)'$. The notation $\boldsymbol{\chi}_t$, $\boldsymbol{\xi}_t$ and $\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t$

is used in similar way, with obvious componentwise meaning. Associated with this infinite-dimensional vector notation, we also consider infinite-dimensional matrices, such as $\underline{\mathbf{A}}(L)$ or $\mathbf{R}$ (see (1.7)), which are $\infty \times \infty$ and $\infty \times q$, respectively. Also, defining $\underline{\mathbf{b}}(L)$ as the $\infty \times q$ matrix with (i, f)-entry $b_{if}(L)$, (1.1) is rewritten as $\mathbf{x}_t = \underline{\mathbf{b}}(L)\mathbf{u}_t + \boldsymbol{\xi}_t$. The reader will easily check that we never produce infinite sums of products, so that our infinite-dimensional matrices are no more than a notational convenience. All infinite-dimensional matrices are underlined, while their finite-dimensional submatrices are not. In particular, $\mathbf{A}_s(L)$ denotes the $s \times s$ upper left submatrix of $\underline{\mathbf{A}}(L)$, $\mathbf{b}_s(L)$ and $\mathbf{R}_s$ the $s \times q$ upper submatrices of $\underline{\mathbf{b}}(L)$ and $\underline{\mathbf{R}}$, respectively.

In Section 3, explicit reference to s in $\mathbf{A}_s(L)$, $\mathbf{b}_s(L)$, $\mathbf{R}_s$, etc., is no longer necessary, and we switch to a somewhat different and more convenient notation. Given the infinite-dimensional process $\mathbf{y}_t = (y_{1t}\ y_{2t}\ \cdots\ y_{nt}\ \cdots)'$, we use the following notation:

(1) $\mathbf{y}_{st}$ is the s-dimensional process $(y_{1t}\ y_{2t}\ \cdots\ y_{st})'$;

(2) $\mathcal{H}^y = \overline{\text{span}}(y_{it},\ i \in \mathbb{N},\ t \in \mathbb{Z})$, $\quad \mathcal{H}^{y_s} = \overline{\text{span}}(y_{it},\ i \le s,\ t \in \mathbb{Z})$;

(3) $\mathcal{H}^y_t = \overline{\text{span}}(y_{i\tau},\ i \in \mathbb{N},\ \tau \le t)$, $\quad \mathcal{H}^{y_s}_t = \overline{\text{span}}(y_{i\tau},\ i \le s,\ \tau \le t)$.

The same notation $\mathcal{H}^{y_s}$, $\mathcal{H}^{y_s}_t$, etc. is used when $\mathbf{y}_t$ is finite-dimensional.

2.2 Basic assumptions

All the stochastic variables x_{it}, χ_{it} and ξ_{it} below have mean zero and finite variance.

Assumption A.1 *For all $n \in \mathbb{N}$, the vector $\mathbf{x}_{nt}$ is weakly stationary and has a spectral density (an absolutely continuous spectral measure).*

Denote by $\mathbf{\Sigma}^x_n(\theta)$, with entries $\sigma^x_{ij}(\theta)$, $i, j \in \mathbb{N}$, $\theta \in [-\pi\ \pi]$, the nested spectral density matrices of the vectors $\mathbf{x}_{nt} = (x_{1t}\ x_{2t}\ \cdots\ x_{nt})'$. The matrix $\mathbf{\Sigma}^x_n(\theta)$ is Hermitian, non-negative definite and has therefore non-negative real eigenvalues for all $\theta \in [-\pi\ \pi]$. Denote by $\lambda^x_{nj}(\theta)$ the j-th eigenvalue, in decreasing order, of $\mathbf{\Sigma}^x_n(\theta)$, and let $\bar{\lambda}^x_f(\theta) = \sup_{n \in \mathbb{N}} \lambda^x_{nf}(\theta)$. The notation $\mathbf{\Sigma}^\chi_n(\theta)$, $\sigma^\chi_{ij}(\theta)$, $\lambda^\chi_{nj}(\theta)$, $\bar{\lambda}^\chi_f(\theta)$, $\mathbf{\Sigma}^\xi_n(\theta)$, $\sigma^\xi_{ij}(\theta)$, $\lambda^\xi_{nj}(\theta)$, and $\bar{\lambda}^\xi_f(\theta)$ is used in a similar way. Our second assumption is

Assumption A.2 *There exists a positive integer q such that (i) $\bar{\lambda}^x_q(\theta)) = \infty$ for almost all θ in $[-\pi\ \pi]$, and (ii) $\bar{\lambda}^x_{q+1}(\theta)$ is essentially bounded.*

Forni and Lippi (2001) prove that

Theorem A *Assumptions A.1 and A.2 imply that* $\mathbf{x}_t$ *can be represented as in (1.1), i.e.*

$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t = \underline{\mathbf{b}}(L)\mathbf{u}_t + \boldsymbol{\xi}_t, \tag{2.1}$$

where $\underline{\mathbf{b}}(L)$ *is an* $\infty \times q$ *matrix of square summable filters,* $\mathbf{u}_t$ *is a* q*-dimensional orthonormal white noise. Moreover,*

(i) $\boldsymbol{\xi}_{nt}$ *satisfies Assumption A.1, and* $\bar{\lambda}_1^{\xi}(\theta)$ *is essentially bounded;*

(ii) $\boldsymbol{\chi}_t$ *(which obviously also satisfies A.1) is such that* $\bar{\lambda}_q^{\chi}(\theta) = \infty$ *for almost all* θ *in* $[\pi\ \pi]$ *(note that* $\bar{\lambda}_{q+1}^{\chi}(\theta) = 0$ *a.e. in* $[\pi\ \pi]$*);*

(iii) $\boldsymbol{\xi}_t$ *and* $\mathbf{u}_{t-k}$ *are uncorrelated for all* $t \in \mathbb{Z}$ *and* $k \in \mathbb{Z}$*;*

(iv) the integer q *and the components* χ_{it} *and* ξ_{it} *are unique.*

Conversely, if $\mathbf{x}_t$ *can be represented as in (2.1) with* $\boldsymbol{\chi}_t$ *and* $\boldsymbol{\xi}_t$ *fulfilling (i), (ii) and (iii), then* $\mathbf{x}_t$ *satisfies Assumptions A.1 and A.2.*

An infinite-dimensional vector fulfilling (i) is called an *idiosyncratic vector*. Divergence of the first q eigenvectors of $\boldsymbol{\chi}_{nt}$ ensures that a representation of $\boldsymbol{\chi}_t$ as a moving average involving lower-dimensional white noise is not possible.

2.3 Infinite-dimensional processes with finite rank

Of course uniqueness of $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ in (2.1) does not imply that $\mathbf{u}_t$ or $\underline{\mathbf{b}}(L)$ are unique. Alternative representations are $\boldsymbol{\chi}_t = [\underline{\mathbf{b}}(L)\mathbf{B}][\mathbf{B}'\mathbf{u}_t] = \underline{\mathbf{c}}(L)\mathbf{v}_t$, where $\mathbf{B}$ is an arbitrary $q \times q$ orthogonal matrix, or, more generally, $\boldsymbol{\chi}_t = [\underline{\mathbf{b}}(L)\mathbf{C}(L)][(\mathbf{C}'(F)\mathbf{u}_t] = \underline{\mathbf{d}}(L)\mathbf{w}_t$, where $F = L^{-1}$ and $\mathbf{C}(L)\mathbf{C}'(F) = \mathbf{I}_q$ for almost all θ in $[-\pi\ \pi]$.

More importantly, Theorem A does not ensure that $\boldsymbol{\chi}_t$ admits a one-sided representation, i.e., a representation of the form $\boldsymbol{\chi}_t = \underline{\mathbf{e}}(L)\mathbf{z}_t$ such that $\underline{\mathbf{e}}(L) = \underline{\mathbf{e}}_0 + \underline{\mathbf{e}}_1 L + \cdots$ for some $\infty \times q$ matrices $\underline{\mathbf{e}}_j$ and some q-dimensional white noise $\mathbf{z}_t$. For example, if

$$\chi_{it} = u_{t+i-1}, \quad i \in \mathbb{N},\ t \in \mathbb{Z}, \tag{2.2}$$

where $(q = 1)$ u_t is one-dimensional white noise, then statement (ii) of Theorem A is fulfilled, so that $\boldsymbol{\chi}_t$ is the common component of some process $\mathbf{x}_t$ satisfying A.1 and A.2, but $\boldsymbol{\chi}_t$ has no one-sided representations (this is quite obvious, see Lemma 1).[5]

Simple conditions for the existence of one-sided representations of infinite-dimensional stochastic vectors are given in Lemmas 1 and 2 below.

Definition 1 *Consider the infinite-dimensional process* $\mathbf{y}_t = (y_{1t}\ y_{2t}\ \cdots\ y_{nt}\ \cdots)'$. *Assume that* $\mathbf{y}_t$ *fulfills Assumption A.1. We say that* $\mathbf{y}_t$ *has rank* q *if there exists* s *such that* $\text{rank}(\boldsymbol{\Sigma}_n^y(\theta)) = q$, *for* $n \geq s$ *and almost all* θ *in* $[-\pi\ \pi]$.

Definition 2 *Let* $\mathbf{y}_t$ *denote an infinite-dimensional stationary stochastic vector, which has a moving average representation*

$$\mathbf{y}_t = \underline{\mathbf{b}}(L)\mathbf{v}_t, \quad t \in \mathbb{Z} \tag{2.3}$$

where $\mathbf{v}_t$ *is* q*-dimensional orthonormal white noise and* $\underline{\mathbf{b}}(L)$ *is an* $\infty \times q$ *square summable filter. We say that (2.3) is a fundamental representation if (1)* $\underline{\mathbf{b}}(L)$ *is one-sided, and (2)* $\mathbf{v}_t$ *belongs to* $\mathcal{H}_t^y$. *In that case, we also say that the white noise* $\mathbf{v}_t$ *is fundamental for* $\mathbf{y}_t$.

Note that if $\mathbf{v}_t$ is fundamental for $\mathbf{y}_t$, then $\mathcal{H}_t^{v_q} = \mathcal{H}_t^y$. The same definition can be given, *mutatis mutandis*, when $\mathbf{y}_t$ is n-dimensional. In that case $q \leq n$. (Orthonormality of $\mathbf{v}_t$ is convenient but not necessary.)

Now suppose that $\mathbf{y}_t$ is n-dimensional: the following properties hold.

(A) If (2.3) is fundamental and $\mathbf{y}_t = \mathbf{c}(L)\mathbf{w}_t$, with $\mathbf{w}_t$ orthonormal, is another fundamental representation, then $\mathbf{w}_t$ has dimension q, $\mathbf{c}(L) = \mathbf{b}(L)\mathbf{Q}$ and $\mathbf{w}_t = \mathbf{Q}'\mathbf{v}_t$, where $\mathbf{Q}$ is a $q \times q$ orthogonal matrix (Rozanov 1967, pp. 56-57).

(B) If (2.3) is fundamental, then $\text{rank}(\mathbf{b}(z)) = q$ for all complex z such that $|z| < 1$ (Rozanov 1967, p. 63, Remark 3). In particular, $\text{rank}(b_0) = \text{rank}(\mathbf{b}(0)) = q$.

A finite-dimensional stationary process with a spectral density does not necessarily possess a fundamental representation. For example, if the spectral density of $\mathbf{y}_t$ is singular on a positive-measure subset of $[-\pi\ \pi]$, then $\mathbf{y}_t$ has no fundamental representations (indeed, it has no one-sided representations, see footnote 5). However,

[5] The possibility that $\boldsymbol{\chi}_t$ has no one-sided representations arises here from infinite dimension. This bears no relationship with the possible non-existence of one-sided representations for finite-dimensional processes, which occurs if their spectral density is singular in a positive-measure subset of $[-\pi\ \pi]$, see e.g. Pourahmadi (2001), Theorem 10.5, p. 361.

(C) If $\mathbf{y}_t$ has rational spectral density, then it has fundamental representations. If $\mathbf{y}_t = \mathbf{b}(L)\mathbf{v}_t$ is one of them, $\mathbf{v}_t$ being q-dimensional orthonormal white noise, then the entries of $\mathbf{b}(L)$ are rational functions of L (Rozanov 1967, Chapter I, Section 10; Hannan 1970, pp. 62-67).

(B′) Suppose that $\mathbf{y}_t$ has rational spectral density, that $\mathbf{y}_t = \mathbf{b}(L)\mathbf{v}_t$, where $\mathbf{b}(L)$ is $n \times q$, rational, square summable and one-sided, $\mathbf{v}_t$ is q-dimensional orthonormal white noise, and that $\text{rank}(\mathbf{b}(z)) = q$ for all z such that $|z| < 1$. Then, $\mathbf{y}_t = \mathbf{b}(L)\mathbf{v}_t$ is fundamental (Hannan, 1970, pp. 62-67).

We say that the infinite-dimensional process $\mathbf{y}_t$ has rational spectral density if $\mathbf{y}_{nt}$ has rational spectral density for all n.

Lemma 1 *Suppose that the infinite-dimensional process $\mathbf{y}_t$ fulfills A.1, has rational spectral density and rank q. The following statements are equivalent:*

(i) $\mathbf{y}_t$ has a one-sided rational moving average representation $\mathbf{y}_t = \underline{\mathbf{b}}(L)\mathbf{v}_t$ (i.e. the entries of $\underline{\mathbf{b}}(L)$ are rational functions of L), where $\mathbf{v}_t$ is q-dimensional orthonormal white noise.

(ii) There exists a positive integer s such that $\mathcal{H}_t^{y_s} = \mathcal{H}_t^{y}$.

Proof. Assume (ii) and let $\mathbf{y}_{st} = \mathbf{b}_s(L)\mathbf{v}_t$ be rational, one-sided and fundamental, so that $\mathcal{H}_t^{y_s} = \mathcal{H}_t^{v_q}$. By assumption $y_{s+k,t} \in \mathcal{H}_t^{y_s}$ and, therefore, $y_{s+k,t} \in \mathcal{H}_t^{v_q}$, so that

$$\mathbf{y}_{st} = \mathbf{b}_s(L)\mathbf{v}_t \qquad \text{and} \qquad y_{s+k,t} = b_{s+k}(L)\mathbf{v}_t. \tag{2.4}$$

The white noise $\mathbf{v}_t$ is fundamental for $\mathbf{y}_{st}$, hence also for $(\mathbf{y}_{st}\ y_{s+k,t})$. Thus representation (2.4) is fundamental, so that, by (C), $b_{s+k}(L)$ must be rational. The conclusion follows. Assume now that (i) holds. We say that β is a zero of $\underline{\mathbf{b}}(L)$ if the determinant of all the $q \times q$ submatrices of $\underline{\mathbf{b}}(\beta)$ vanish. Assume that α is a zero of $\underline{\mathbf{b}}(L)$ and that $|\alpha| < 1$. There exists an orthogonal $q \times q$ matrix $\mathbf{B}_\alpha$ such that all the entries of the first column of $\underline{\mathbf{b}}(L)\mathbf{B}_\alpha$ vanish at α. Defining $\boldsymbol{\gamma}_\alpha(L)$ as the $q \times q$ diagonal matrix with diagonal entries $((1-\alpha L)(L-\alpha)^{-1}\ \ 1\ \cdots\ 1)$, we have

$$\mathbf{y}_t = [\underline{\mathbf{b}}(L)\mathbf{B}_\alpha\boldsymbol{\gamma}_\alpha(L)]\left[\boldsymbol{\gamma}_{\tilde{\alpha}}(L^{-1})\tilde{\mathbf{B}}_\alpha\mathbf{v}_t\right] = \underline{\mathbf{c}}(L)\mathbf{w}_t,$$

where a tilde denotes transposition and conjugation. This is an alternative one-sided rational representation in which the multiplicity of α as a zero of the matrix polynomial has decreased by one unit. Because a zero of $\underline{\mathbf{b}}(L)$ is a zero of $\mathbf{b}_q(L)$, with a finite number of iterations we obtain a rational representation, $\mathbf{y}_t = \underline{\mathbf{d}}(L)\mathbf{z}_t$, say, such that $\underline{\mathbf{d}}(L)$ has no zeros of modulus less than unity. For the same reason, there exists an integer s such that $\mathbf{d}_s(L)$ has no zeros of modulus less than unity. By (B′), $\mathbf{y}_{st} = \mathbf{d}_s(L)\mathbf{z}_t$ is fundamental for $\mathbf{y}_{st}$ and therefore for $\mathbf{y}_t$. Q.E.D.

Lemma 2 *Suppose that the infinite-dimensional process $\mathbf{y}_t$ fulfills A.1, has rational spectral density and rank q. Then,*

(i) $\mathbf{y}_t$ has a fundamental rational representation $\mathbf{y}_t = \underline{\mathbf{b}}(L)\mathbf{v}_t$ if and only if it has a one-sided representation;

(ii) if $\mathbf{y}_t = \underline{\mathbf{b}}(L)\mathbf{v}_t$ and $\mathbf{y}_t = \underline{\mathbf{c}}(L)\mathbf{w}_t$ are fundamental, with $\mathbf{v}_t$ and $\mathbf{w}_t$ q-dimensional and orthonormal, then $\underline{\mathbf{c}}(L) = \underline{\mathbf{b}}(L)\mathbf{Q}$ and $\mathbf{w}_t = \mathbf{Q}'\mathbf{v}_t$, where $\mathbf{Q}$ is some $q \times q$ orthogonal matrix;

(iii) if $\mathbf{y}_t = \underline{\mathbf{b}}(L)\mathbf{v}_t = \underline{\mathbf{b}}_0\mathbf{v}_t + \underline{\mathbf{b}}_1\mathbf{v}_{t-1} + \cdots$ is fundamental, then $\underline{\mathbf{b}}_0$ has rank q.

Proof. Statement (i) is part of the proof of Lemma 1. As for (ii), suppose that $\mathbf{y}_t = \underline{\mathbf{b}}(L)\mathbf{v}_t$ and $\mathbf{y}_t = \underline{\mathbf{c}}(L)\mathbf{w}_t$ both are fundamental. By Lemma 1, there exists s such that $\mathcal{H}_t^{y_s} = \mathcal{H}_t^{y}$. As a consequence, both $\mathbf{v}_t$ and $\mathbf{w}_t$ belong to $\mathcal{H}_t^{y_s}$, and therefore are fundamental for $\mathbf{y}_{st}$. This implies that $\mathbf{w}_t = \mathbf{Q}'\mathbf{v}_t$, where $\mathbf{Q}$ is orthogonal. Thus $\mathbf{y}_t = \underline{\mathbf{c}}(L)\mathbf{w}_t = [\underline{\mathbf{c}}(L)\mathbf{Q}']\mathbf{v}_t = \underline{\mathbf{b}}(L)\mathbf{v}_t$. As $\mathbf{v}_t$ is orthonormal white noise, we have $\underline{\mathbf{c}}(L) = \underline{\mathbf{b}}(L)\mathbf{Q}$. Because $\mathbf{v}_t$ is fundamental for $\mathbf{y}_{st}$, $\mathbf{b}_s(0)$ has rank q, see (B), so that $\underline{\mathbf{b}}(0) = \underline{\mathbf{b}}_0$ has rank q. Q.E.D.

Summing up, given the infinite-dimensional vector $\mathbf{y}_t$, assuming A.1, finite rank, rational spectral density, and the existence of a one-sided moving average representation, thus ruling out cases like (2.2), we obtain the existence of a rational fundamental representation for $\mathbf{y}_t$, which is unique up to an orthogonal matrix.

Let us now return to the infinite-dimensional vector $\boldsymbol{\chi}_t$. As we have seen, $\boldsymbol{\chi}_t$ fulfills A.1. Assume that $\boldsymbol{\chi}_t$ has rational spectral density, so that either $\mathrm{rank}(\boldsymbol{\Sigma}_n^{\chi}(\theta)) < q$ for all $\theta \in [-\pi\ \pi]$ or $\mathrm{rank}(\boldsymbol{\Sigma}_n^{\chi}(\theta)) = q$ for almost all θ in $[-\pi\ \pi]$. On the other hand, since $\lambda_{nq}^{\chi}(\theta)$ diverges θ-almost everywhere in $[-\pi\ \pi]$, there exists s such that $\mathrm{rank}(\boldsymbol{\Sigma}_n^{\chi}(\theta)) = q$ for $n \geq s$ and almost all θ in $[-\pi\ \pi]$. Therefore $\boldsymbol{\chi}_t$ has rank q.

Assuming that $\boldsymbol{\chi}_t$ has rational spectral density and that $\mathcal{H}_t^{\chi_s} = \mathcal{H}_t^{\chi}$ for some s, so that cases like (2.2) cannot occur, Lemma 2 ensures that $\boldsymbol{\chi}_t$ has a rational fundamental representation. More precisely, for $i \in \mathbb{N}$,

$$\chi_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt}, \tag{2.5}$$

where $c_{if}(L)$ and $d_{if}(L)$ are polynomials in L, and $\mathbf{u}_t$ is fundamental for $\boldsymbol{\chi}_t$. Representation (2.5) is unique up to an orthogonal matrix.

However, in Assumption A.3 (see Section 2.5), we require more than the existence of an integer s such that $\mathcal{H}_t^{\chi_s} = \mathcal{H}_t^{\chi}$, and rather assume that the space spanned by $\chi_{i_1\tau}, \chi_{i_2\tau}, \ldots, \chi_{i_{q+1},\tau}$, $\tau \leq t$, coincides with $\mathcal{H}_t^{\chi}$ for all $(q+1)$-tuples $i_1 < i_2 < \cdots < i_{q+1}$. Thus, $\mathbf{u}_t$ in (2.5) is fundamental for any $(q+1)$-dimensional subvector of $\boldsymbol{\chi}_t$, not only for the subvector $\boldsymbol{\chi}_{st}$ associated with some s. This stronger requirement is motivated by the main result of Section 2.4. We prove that, under a quite general parameterization, the stronger condition holds generically, i.e. outside of a negligible subset, as defined in Section 2.4, of the parameter space.

2.4 AR representations of singular stochastic vectors

Consider a n-dimensional vector $\mathbf{y}_t$ such that

$$y_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} v_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} v_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} v_{qt} \tag{2.6}$$

with

$$c_{if}(L) = c_{if,0} + c_{if,1}L + \cdots + c_{if,s_1}L^{s_1} \quad \text{and} \quad d_{if}(L) = 1 + d_{if,1}L + \cdots + d_{if,s_2}L^{s_2} \tag{2.7}$$

for $i = 1, 2, \ldots, n$, $f = 1, 2, \ldots, q$, where $\mathbf{v}_t = (v_{1t}\ v_{2t}\ \cdots\ v_{qt})$ is orthonormal white noise. For each value of i, the filters in (2.7) are parameterized by a ν-dimensional real parameter, with $\nu = q(s_1 + s_2 + 1)$. More precisely, for each i, the parameter space for (2.7) is the set $\Pi \subset \mathbb{R}^{\nu}$ such that all the roots of the polynomial $d_{if}(L)$ are of modulus greater than unity. Thus the vector $\mathbf{y}_t$ is described by a parameter taking values in $\Pi^n = \underbrace{\Pi \times \Pi \times \cdots \times \Pi}_{n}$, which is an open subset of $\mathbb{R}^{\mu}$, with $\mu = n\nu$.

We are interested in the case $n > q$. Such "tall systems" have been studied recently in Anderson and Deistler (2008a and b). One of their results is that (if $n > q$), then, for *generic* values of the parameters, $\mathbf{y}_t$ has an autoregressive representation of the form

$$\mathbf{A}(L)\mathbf{y}_t = \mathbf{R}\mathbf{v}_t, \tag{2.8}$$

where $\mathbf{R}$ is $n \times q$, $\text{rank}(\mathbf{R}) = q$, and $\mathbf{A}(L)$ is an $n \times n$ matrix polynomial with *finite* degree. Precisely, there exists a nowhere dense set $\mathcal{N} \subset \Pi^n$, i.e. a set whose closure has no interior points, such that, for all parameter vectors in $\Pi^n - \mathcal{N}$, $\mathbf{y}_t$ has the finite-order autoregressive representation (2.8). As $\mathbf{R}$ has full rank, (2.6) and (2.8) imply that, generically, $\mathbf{v}_t$ is fundamental for $\mathbf{y}_t$.

To provide an intuition for this result and Proposition 1 below, let us consider the following elementary example, in which $n = 2$, $q = 1$, and

$$y_{1t} = a_1 v_t + b_1 v_{t-1} \qquad y_{2t} = a_2 v_t + b_2 v_{t-1}, \quad t \in \mathbb{Z} \tag{2.9}$$

with parameter (a_1, b_1, a_2, b_2) in $\mathbb{R}^2 \times \mathbb{R}^2$. Outside of the nowhere dense subset in which $a_1 b_2 - a_2 b_1 = 0$, we obtain

$$v_t = \frac{1}{a_1 b_2 - a_2 b_1}(b_2 y_{1t} - b_1 y_{2t}). \tag{2.10}$$

Using (2.10) to get rid of v_{t-1} in (2.9), we obtain the AR(1) representation

$$y_{1t} = d b_1 b_2 y_{1t-1} - d b_1^2 y_{2t-1} + a_1 v_t \quad y_{2t} = d b_2^2 y_{1t-1} - d b_1 b_2 y_{2t-1} + a_2 v_t, \quad t \in \mathbb{Z} \tag{2.11}$$

where $d = 1/(a_1 b_2 - a_2 b_1)$. Note that

(i) If $a_1 b_2 - a_2 b_1 = 0$, no finite-order autoregressive representation exists, unless $b_1 = b_2 = 0$. Moreover, fundamentalness of v_t for $\mathbf{y}_t$ requires that the root of $a_1 + b_1 L$ (which is also the root of $a_2 + b_2 L$) has modulus larger than one.

(ii) As soon as $a_1 b_2 - a_2 b_1 \neq 0$, however, the position of the root of $a_i + b_i L$ does not play any role in the fundamentalness of v_t for $\mathbf{y}_t$.

(iii) Quite obviously $a_1 b_2 - a_2 b_1 \neq 0$ if and only if χ_{1t-1} and χ_{2t-1} are linearly independent. Therefore, generically, the projection (2.11) is unique, i.e. generically no other autoregressive representation of order one exists.

(iv) But other autoregressive representations do exist. Rewriting (with obvious definitions of $\mathbf{A}$ and $\mathbf{a}$) (2.11) as $\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{a} v_t$, we get $\mathbf{y}_t = \mathbf{A}^2 \mathbf{y}_{t-2} + \mathbf{A}\mathbf{a} v_{t-1} + \mathbf{a} v_t$. Using (2.10) to get rid of v_{t-1}, we obtain another autoregressive representation, of order two. Such non-uniqueness does not occur for square systems (when $n = q$).

(v) On the other hand, if $n = 3$ and $y_{it} = a_i v_t + b_i v_{t-1}$, $i = 1, 2, 3$, then, outside of the set in which $a_2 b_1 = a_1 b_2$ and $a_3 b_1 = a_1 b_3$, which is nowhere dense in $\mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2$, we have

$$v_t = \frac{1}{a_1 \gamma_1 + a_2 \gamma_2 + a_3 \gamma_3}(\gamma_1 y_{1t} + \gamma_2 y_{2t} + \gamma_3 y_{3t}),$$

where $b_1\gamma_1+b_2\gamma_2+b_3\gamma_3=0$. This can be used to get rid of v_{t-1}, in the same way as we did in the $n=2$ case. Thus, generically, $\mathbf{y}_t$ has an AR(1) representation. However, the variables y_{it-1}, $i=1,2,3$, are not linearly independent, so that such minimum-lag autoregressive representation is not unique.

Let us show that remark (iii) can be generalized. Precisely, if $n=q+1$, then, generically, there exists only one minimal-lag autoregressive representation.

Proposition 1 *Consider an n-dimensional vector $\mathbf{y}_t$ with representation (2.6), and assume that $n=q+1$. There exists a set $\mathcal{N}\subset\Pi^{q+1}$, nowhere dense in Π^{q+1}, such that, if the parameter vector lies in $\Pi^{q+1}-\mathcal{N}$,*

(a) $\mathbf{y}_t$ has a finite-order AR representation $\mathbf{A}(L)\mathbf{y}_t=\mathbf{R}\mathbf{v}_t$, where $\mathbf{R}=(R_{if})$ is $(q+1)\times q$, $R_{if}=c_{if}(0)$, $\operatorname{rank}(\mathbf{R})=q$, $\mathbf{A}(L)$ is $(q+1)\times(q+1)$ and has order not exceeding $S=qs_1+q^2s_2$. This implies that $\mathbf{v}_t$ is fundamental for $\mathbf{y}_t$.

(b) If (i) $\mathbf{A}^*(L)$ is a $(q+1)\times(q+1)$ polynomial matrix whose order does not exceed S, with $\mathbf{A}^*(0)=\mathbf{I}$, (ii) $\mathbf{R}^*$ is $(q+1)\times q$, (iii) $\mathbf{v}_t^*$ is a q-dimensional orthonormal white noise orthogonal to $\mathbf{y}_{t-k}$, $k\geq 1$, (iv) $\mathbf{A}^*(L)\mathbf{y}_t=\mathbf{R}^*\mathbf{v}_t^*$, then $\mathbf{A}^*(L)=\mathbf{A}(L)$, $\mathbf{R}^*=\mathbf{R}\mathbf{B}$, $\mathbf{v}_t^*=\mathbf{B}'\mathbf{v}_t$, where $\mathbf{B}$ is an orthogonal $q\times q$ matrix.

See Appendix A for the proof.

Note that Proposition 1 does not claim that (generically) the process $\mathbf{y}_t$ corresponding to a parameter value in Π^{q+1} has no non-fundamental representations. What it claims is that (generically) such non-fundamental representations are not parameterized in Π^{q+1}. For example, representation (2.9) is generically fundamental in $\mathbb{R}^2\times\mathbb{R}^2$. On the other hand, given any a with $|a|>1$, the process $\mathbf{y}_t$ also has the representation

$$y_{it}=\left[(a_i+b_iL)\frac{1+aL}{1+a^{-1}L}\right]\left[\frac{1-a^{-1}L}{1-aL}v_t\right]=\frac{(a_i+b_iL)(1-aL)}{1-a^{-1}L}w_t, \qquad (2.12)$$

for $i=1,2$, where

$$w_t=\frac{1-a^{-1}L}{1-aL}v_t=-a^{-1}F\frac{1-a^{-1}L}{1-a^{-1}F}v_t$$

is white noise (this is easily proved by showing that its spectral density is constant). This is a non-fundamental representation for $\mathbf{y}_t$. However, (2.12) is parameterized in $\mathbb{R}^2\times\mathbb{R}^2\times\mathbb{R}$, not $\mathbb{R}^2\times\mathbb{R}^2$.

Now assume that $\mathbf{y}_t$ is infinite-dimensional with y_{it} modeled as in (2.6) for $i\in\mathbb{N}$. The vector $\mathbf{y}_t$ is parameterized in $\Pi^\infty=\Pi\times\Pi\times\cdots$. We define negligible sets and genericity in Π^∞ with respect to the product topology. We say that a subset of Π^∞ is negligible if it is *meagre*, i.e. the union of a countable set of nowhere dense subsets, and that a property holds generically in Π^∞ if the subset where it does not hold is meagre.

Define the set $\mathcal{M}_s$, for $s \geq q+1$, as the set of points in Π^∞ such that all vectors $\mathbf{y}_t^{i_1,i_2,\ldots,i_{q+1}} = (y_{i_1t}\ y_{i_2t}\ \cdots\ y_{i_{q+1}t})$, with $i_1 < i_2 < \cdots < i_{q+1} \leq s$, admit a representation of the form

$$\mathbf{A}^{i_1,i_2,\ldots,i_{q+1}}(L)\mathbf{y}_t^{i_1,i_2,\ldots,i_{q+1}} = \mathbf{R}^{i_1,i_2,\ldots,i_{q+1}}\mathbf{v}_t, \tag{2.13}$$

where $\mathbf{A}^{i_1,i_2,\ldots,i_{q+1}}(L)$ is of order not greater than S and unique in the sense of Proposition 1(b). From Proposition 1, we see that $\mathcal{N}_s = \Pi^\infty - \mathcal{M}_s$ is a nowhere dense subset in the product topology of Π^∞, so that the set $\mathcal{N} = \cup_{s=q+1}^\infty \mathcal{N}_s$, being a countable union of nowhere dense subsets of Π^∞, is a meagre subset. We can conclude that, in $\Pi^\infty - \mathcal{N}$, thus generically in Π^∞, all vectors of the form $\mathbf{y}_t^{i_1,i_2,\ldots,i_{q+1}} = (y_{i_1t}\, y_{i_2t}\ \cdots\ y_{i_{q+1}t})$, with $i_1 < i_2 < \cdots < i_{q+1}$ (no upper limit for i_{q+1}), can be represented as in (2.13), where $\mathbf{A}^{i_1,i_2,\ldots,i_{q+1}}(L)$ is of order not greater than S and unique in the sense of Proposition 1(b).[6]

Some observations are in order. Firstly, defininig negligible subsets of Π^∞ as meagre subsets has a good motivation in the fact that (i) the complement of a meagre subset of Π^∞ is not meagre, (ii) if a subset of Π^∞ is not meagre, obtaining it as the union of a family of nowhere dense subsets requires an uncountable family.[7]

Secondly, the family of meagre subsets of Π^∞ is strictly broader than the family of nowhere dense subsets. In particular, the set $\mathcal{N}$ is not nowhere dense. To see this, consider again the MA(1) example $y_{it} = a_i v_t + b_i v_{t-1}$, with $i \in \mathbb{N}$. Denote by $\mathbf{c} = (\mathbf{c}_1\ \mathbf{c}_2\ \cdots\ \mathbf{c}_n\ \cdots)$, where $\mathbf{c}_i = (a_i\ b_i)$, a point in Π^∞. A well-known feature of the product topology is that any neighborhood G of $\mathbf{c}$ contains points $\mathbf{c}'$ such that, for some s and all $n > s$, $\mathbf{c}'_n = \mathbf{c}'_s$. Such points obviously belong to $\mathcal{N}$. Thus $\mathcal{N}$ is meagre but dense in Π^∞ (in the same way as the rational numbers are a meagre but dense subset of the real numbers).

[6]An analogous genericity result can be obtained if negligible subsets of $\Pi^\infty$are defined as subsets of zero measure with respect to the product measure.

[7]Denote by $\overline{\Pi}$ the closure of Π. Then, (1) the space $\overline{\Pi}^\infty$, being the Cartesian product of a countable family of complete metric spaces, is a complete metric space; (2) in complete metric spaces the complement of a meagre subset is not meagre (Baire Category Theorem; see Dunford and Schwartz (1988), p. 32, Lemma 4, and p. 20, Theorem 9 (Baire Theorem), respectively. It is easily seen that the Baire Theorem also applies to Π^∞, which is an open dense subset of $\overline{\Pi}^\infty$).

Lastly, assuming that the parameter space indexing the polynomials $c_{ij}(L)$ and $d_{ij}(L)$ does not depend on i, as we do in (2.6), is convenient but not necessary. With the dimension of the parameter space depending on i, a more general version of Proposition 1 holds as well as the meagreness result for infinite-dimensional vectors $\mathbf{y}_t$. However, the gain in generality does not seem to justify the substantial additional complications in the proof of Proposition 1 and the determination of the order of $\mathbf{A}(L)$.

2.5 Autoregressive representations for the vector $\boldsymbol{\chi}_t$

Let us now turn our attention to the vector $\boldsymbol{\chi}_t$ of common components. As we have seen, assuming that $\boldsymbol{\chi}_t$ has rational spectral density and that $\mathcal{H}_t^{\chi_s} = \mathcal{H}_t^{\chi}$ for some s implies, by Lemmas 1 and 2 that $\boldsymbol{\chi}_t$ has a fundamental rational representation of the form (2.6). The meagreness argument above motivates assuming that statements (a) and (b) hold for all $(q+1)$-dimensional subvectors of $\boldsymbol{\chi}_t$. More precisely,

Assumption A.3 *The vector $\boldsymbol{\chi}_t$ has a representation*

$$\chi_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt},$$

where

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

for all $i \in \mathbb{N}$ and $f = 1, 2, \ldots, q$. Moreover,

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

$$\mathbf{A}^{i_1,i_2,\ldots,i_{q+1}}(L)\boldsymbol{\chi}_t^{i_1,i_2,\ldots,i_{q+1}} = \mathbf{R}^{i_1,i_2,\ldots,i_{q+1}}\mathbf{u}_t, \tag{2.14}$$

where $\mathbf{A}^{i_1,i_2,\ldots,i_{q+1}}(L)$ is of order not greater than $S = qs_1 + q^2 s_2$, and $\mathbf{R}^{i_1,i_2,\ldots,i_{q+1}}$ has rank q. This implies that $\mathbf{u}_t$ is fundamental for all $(q+1)$-dimensional subvectors of $\boldsymbol{\chi}_t$.

(ii) Representation (2.14) is unique in the sense of Proposition 1(b).

An immediate consequence of Assumption A.3 is that $\boldsymbol{\chi}_t$ can be represented as in (1.6), that is,

$$\mathbf{A}^1(L)\begin{pmatrix}\chi_{1t}\\ \chi_{2t}\\ \vdots\\ \chi_{q+1,t}\end{pmatrix} = \mathbf{R}^1\mathbf{u}_t, \quad \mathbf{A}^2(L)\begin{pmatrix}\chi_{q+2,t}\\ \chi_{q+3,t}\\ \vdots\\ \chi_{2(q+1),t}\end{pmatrix} = \mathbf{R}^2\mathbf{u}_t, \quad \ldots \tag{2.15}$$

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

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

Of course, any permutation of the variables produces a distinct $(q+1)$-blockwise autoregressive representation. Precisely, let $\tilde{\chi}_{it} = \chi_{g(i),t}$ with $g : \mathbb{N} \to \mathbb{N}$ a one-to-one mapping. Assumptions A.1, A.2 and A.3 imply that $\tilde{\boldsymbol{\chi}}_t$ has a representation of the form (2.15), with matrices $\tilde{\mathbf{A}}^k(L)$, $\tilde{\mathbf{R}}^k$ and a white noise vector $\tilde{\mathbf{u}}_t$. Assumption A.3 implies that $\tilde{\mathbf{u}}_t = \mathbf{H}\mathbf{u}_t$, with $\mathbf{H}$ orthogonal.

It must be pointed out that neither $\mathbf{u}_t$ nor $\underline{\mathbf{R}}$ play any special role. Assumption A.3 states that there exists $\mathbf{u}_t$ such that (2.14) holds. All the white noise vectors and matrices corresponding to alternative representations are linked to $\mathbf{u}_t$ and $\underline{\mathbf{R}}$ by orthogonal transformations. For identification and estimation of a couple $\mathbf{u}_t^*$, $\underline{\mathbf{R}}^*$ based on economic theory, see Section 3.3.

2.6 Construction of the autoregressive representations of $\boldsymbol{\chi}_t$

Assumption A.3 ensures the existence of the autoregressive representation (2.15). We now show how (2.15), i.e. the matrices $\mathbf{A}^k(L)$ and (up to multiplication by an orthogonal matrix) $\mathbf{R}^k$, can be constructed from the spectral density of the χ's.

(i) Assume that the population spectral density of the vector $\boldsymbol{\chi}_t$ is known, i.e. that the nested spectral density matrices $\boldsymbol{\Sigma}_n^{\chi}(\theta)$, $n \in \mathbb{N}$, are known.

(ii) Denote by $\boldsymbol{\chi}_t^k$ the k-th of the $(q+1)$-dimensional subvectors of $\boldsymbol{\chi}_t$ (which is unobservable) appearing in (2.15), and call $\boldsymbol{\Sigma}_{jk}^{\chi}(\theta)$ the $(q+1) \times (q+1)$ cross-spectral density between $\boldsymbol{\chi}_t^j$ and $\boldsymbol{\chi}_t^k$. Then, denoting by $\boldsymbol{\Gamma}_{jk,s}^{\chi}$ the covariance between $\boldsymbol{\chi}_t^j$ and $\boldsymbol{\chi}_{t-s}^k$,

$$\boldsymbol{\Gamma}_{jk,s}^{\chi} = \mathrm{E}\left[\boldsymbol{\chi}_t^j \boldsymbol{\chi}_{t-s}^{k\ \prime}\right] = \int_{-\pi}^{\pi} e^{\imath s\theta} \boldsymbol{\Sigma}_{jk}^{\chi}(\theta) d\theta, \tag{2.16}$$

where $\imath$ stands for the imaginary unit.

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

$$\boldsymbol{\psi}_t^1 = \mathbf{A}^1(L)\boldsymbol{\chi}_t^1, \quad \boldsymbol{\psi}_t^2 = \mathbf{A}^2(L)\boldsymbol{\chi}_t^2, \quad \ldots \tag{2.17}$$

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

$$\mathbf{A}^{[k]} = \left(\mathbf{A}^k_1 \; \mathbf{A}^k_2 \; \cdots \; \mathbf{A}^k_S\right), \quad \mathbf{B}^\chi_k = \left(\mathbf{\Gamma}^\chi_{kk,1} \; \mathbf{\Gamma}^\chi_{kk,2} \; \cdots \; \mathbf{\Gamma}^\chi_{kk,S}\right) \tag{2.18}$$

and

$$\mathbf{C}^\chi_{jk} = \begin{pmatrix} \mathbf{\Gamma}^\chi_{jk,0} & \mathbf{\Gamma}^\chi_{jk,1} & \cdots & \mathbf{\Gamma}^\chi_{jk,S-1} \\ \mathbf{\Gamma}^\chi_{jk,-1} & \mathbf{\Gamma}^\chi_{jk,0} & \cdots & \mathbf{\Gamma}^\chi_{jk,S-2} \\ \vdots & & & \vdots \\ \mathbf{\Gamma}^\chi_{jk,-S+1} & \mathbf{\Gamma}^\chi_{jk,-S+2} & \cdots & \mathbf{\Gamma}^\chi_{jk,0} \end{pmatrix}. \tag{2.19}$$

We have

$$\mathbf{A}^{[k]} = \mathbf{B}^\chi_k \left(\mathbf{C}^\chi_{kk}\right)^{-1} = \mathbf{B}^\chi_k \left(\mathbf{C}^\chi_{kk}\right)_{\mathrm{ad}} \det\left(\mathbf{C}^\chi_{kk}\right)^{-1} \text{ and } \mathbf{\Gamma}^\psi_{jk} = \mathbf{\Gamma}^\chi_{jk} - \mathbf{A}^{[j]}\mathbf{C}^\chi_{jk}\mathbf{A}^{[k]}, \tag{2.20}$$

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

(iv) The $\infty \times \infty$ matrix $\underline{\mathbf{\Gamma}}^\psi$ obtained by piecing together the matrices $\mathbf{\Gamma}^\psi_{jk}$ is of rank q (see Lemma 2(iii)) and can therefore be represented as $\underline{\mathbf{\Gamma}}^\psi = \underline{\mathbf{S}}\;\underline{\mathbf{S}}'$, where $\underline{\mathbf{S}}$ is an $\infty \times q$ matrix. On the other hand, $\underline{\mathbf{\Gamma}}^\psi$ is the covariance matrix of the right-hand side terms in (2.15), so that $\underline{\mathbf{S}} = \underline{\mathbf{R}}\mathbf{H}$, where $\mathbf{H}$ is $q \times q$ and orthogonal. In conclusion, using $\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t$, we obtain, with $\mathbf{z}_t = \underline{\mathbf{A}}(L)\mathbf{x}_t$ and $\boldsymbol{\phi}_t = \underline{\mathbf{A}}(L)\boldsymbol{\xi}_t$,

$$\mathbf{z}_t = \underline{\mathbf{R}}\mathbf{u}_t + \boldsymbol{\phi}_t \qquad t \in \mathbb{Z}. \tag{2.21}$$

The above construction, based on an estimate of the spectral densites $\mathbf{\Sigma}^\chi_n(\theta)$ rather than $\mathbf{\Sigma}^\chi_n(\theta)$ itself, is used, step by step, in our estimation procedure, see Section 3.

2.7 Assumptions on the representation $\mathbf{z}_t = \underline{\mathbf{R}}\mathbf{u}_t + \boldsymbol{\phi}_t$

Equation (2.21) looks like a *static* representation of the form (1.2) for $\mathbf{z}_t = \underline{\mathbf{A}}(L)\mathbf{x}_t$, with $r = q$ and $\mathbf{N}(L) = \mathbf{I}_q$. However, the requirements that the q-th eigenvalue of the spectral density of $\mathbf{R}_n\mathbf{u}_t$ diverges (such eigenvalue is constant as a function of θ), whereas the first eigenvalue of $\mathbf{A}_n(L)\boldsymbol{\xi}_t$ is bounded, calls for some further assumptions.

It is convenient here to assume, without loss of generality, that n, the number of variables, increases by blocks of size $q+1$. Thus $n = m(q+1)$, where m is the number of blocks, so that n and m grow at the same pace. Moreover, consistently with Section 2.1, denote by $\mathbf{A}_n(L)$ the $n \times n$ upper-left submatrix of $\underline{\mathbf{A}}(L)$, i.e. the block-diagonal matrix with diagonal blocks $\mathbf{A}^s(L)$, $s = 1, \ldots, m$.

Denote by λ^{ψ}_{nj} the j-th eigenvalue of the spectral density matrix of $\boldsymbol{\psi}_{nt} = \mathbf{R}_n\mathbf{u}_t = \mathbf{A}_n(L)\boldsymbol{\chi}_{nt}$ (a constant spectral density). Our assumption that $\lambda^{x}_{nq}(\theta)$, hence $\lambda^{\chi}_{nq}(\theta)$, diverges a.e. in $[-\pi\ \pi]$ does not imply that λ^{ψ}_{nq} diverges. This is easily seen using the MA(1) example $\chi_{it} = a_i u_t + b_i u_{t-1}$, for which

$$\lambda^{\chi}_{n1}(\theta) = \sum_{i=1}^{n} |a_i + b_i e^{-\imath\theta}|^2 \quad \text{and} \quad \lambda^{\psi}_{n1} = \sum_{i=1}^{n} a_i^2.$$

Clearly, $\lambda^{\chi}_{n1}(\theta)$ can diverge for almost all θ in $[-\pi\ \pi]$ even though λ^{ψ}_{n1} does not. An additional assumption is therefore necessary, see Assumption A.8 in Section 3.1.

Next, we need an assumption implying that $\boldsymbol{\phi}_t$ is an idiosyncratic vector, i.e. that the first eigenvalue of its spectral density is bounded. Still assuming that $n = m(q+1)$, the spectral density of $\mathbf{A}_n(L)\boldsymbol{\xi}_{nt}$ is $\mathbf{A}_n(e^{-\imath\theta})\boldsymbol{\Sigma}^{\xi}_n(\theta)\mathbf{A}'_n(e^{\imath\theta})$. If $\mathbf{a}$ is an n-dimensional column unit vector, we have

$$\mathbf{a}'\mathbf{A}_n(e^{-\imath\theta})\boldsymbol{\Sigma}^{\xi}_n(\theta)\mathbf{A}'_n(e^{\imath\theta})\mathbf{a} \le \lambda^{\xi}_{n1}(\theta)\left[\mathbf{a}'\mathbf{A}_n(e^{-\imath\theta})\mathbf{A}'_n(e^{\imath\theta})\mathbf{a}\right] \le \lambda^{\xi}_{n1}(\theta)\lambda^{A_n}_1(\theta),$$

where $\lambda^{A_n}_1(\theta)$ is the first eigenvalue of $\mathbf{A}_n(e^{-\imath\theta})\mathbf{A}'_n(e^{\imath\theta})$, which is Hermitian, non-negative definite. By Theorem A, $\sup_n \lambda^{\xi}_{n1}(\theta)$ is essentially bounded. Thus we have to discuss conditions under which $\sup_n \lambda^{A_n}_1(\theta)$ or, equivalently, $\sup_{k\in\mathbb{N}} \lambda^{A^k}_1(\theta)$, where $\lambda^{A^k}_1(\theta)$ now is $\mathbf{A}^k(e^{-\imath\theta})\mathbf{A}^{k\prime}(e^{\imath\theta})$'s first eigenvalue, is essentially bounded. Inspection of (2.20) shows that the next two assumptions are sufficient to ensure that there exists a common upper bound for the moduli of the entries of $\mathbf{A}^{[k]}$, and therefore that $\lambda^{A^k}_1(\theta)$, as a function of k and θ, is bounded.

Assumption A.4 *There exist a real d such that*

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

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

Assumption A.5 *There exists a real number G such that $|\sigma^{\chi}_{ij}(\theta)| \le G$ for all $i \in \mathbb{N}$, $j \in \mathbb{N}$ and $\theta \in [-\pi\ \pi]$.*

Under Assumptions A.4 and A.5 the first eigenvalue of the spectral density of $\mathbf{A}_n(L)\boldsymbol{\xi}_{nt}$ is bounded, i.e. $\mathbf{A}(L)\boldsymbol{\xi}_t$ is an idiosyncratic component. Assumption A.8, see Section 3.1, ensures that the q-th eigenvalue of $\mathbf{R}_n\mathbf{u}_t$ diverges. Using Theorem A (its converse part), (2.21) is a GDFM representation for $\mathbf{z}_t$ with q factors—as noted above, a static one, of the form (1.2), with $r = q$.

3 Estimation

Our estimation procedure follows the same steps as the population construction in Section 2.6, with the population spectral density replaced with an estimator. Firstly, via the frequency-domain principal components of $\mathbf{x}_{nt}$ (see Forni, Hallin, Lippi and Reichlin 2000), we obtain (step (ii)) an estimator of $\mathbf{\Sigma}_n^{\chi}(\theta)$ from a non-parametric estimator of (step (i)) the spectral density $\mathbf{\Sigma}_n^x(\theta)$ of $\mathbf{x}_{nt}$. Then (step (iii)), $\hat{\mathbf{A}}_n(L)$ and $\hat{\mathbf{\Gamma}}_n^{\psi}$ are computed as a natural counterparts of their population versions in Section 2.6. Lastly (step (iv)), estimators for $\mathbf{R}_n$ and $\mathbf{u}_t$ are obtained via a standard principal component analysis of $\hat{\mathbf{z}}_{nt} = \hat{\mathbf{A}}(L)\mathbf{x}_{nt}$, see Sections 3.3 and 3.4. consistency and consistency rates for all the above estimators are provided in Propositions 2 through 6.

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

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 elments $\lambda_f^x(\theta)$;

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

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

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

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

(precise definitions are provided below) depend both on n and the sample x_{it}, $i = 1, \dots, n$, $t = 1, \dots, T$. For simplicity, we say that they depend on n and T;

(vi) the same notational change applies to $\mathbf{\Gamma}_n^{\psi}$ and related eigenvalues and eigenvectors;

(vii) $\mathbf{A}(L)$ and $\mathbf{R}$, denoting the upper left $n \times n$ and $n \times q$ submatrices of $\underline{\mathbf{A}}(L)$ and $\underline{\mathbf{R}}$, respectively, are used instead of $\mathbf{A}_n(L)$ and $\mathbf{R}_n$; $\hat{\mathbf{A}}(L)$ and $\hat{\mathbf{R}}$ stand for their estimated counterparts;

(viii) to avoid confusion, however, we keep explicit reference to n in $\mathbf{x}_{nt}$, $\boldsymbol{\chi}_{nt}$, $\mathbf{z}_{nt}$ etc., with estimated counterparts of the form $\hat{\boldsymbol{\chi}}_{nt}$, $\hat{\mathbf{z}}_{nt}$, etc.; thus, we write, for instance, $\mathbf{z}_{nt} = \mathbf{A}(L)\mathbf{x}_{nt} = \mathbf{S}\mathbf{v}_t + \boldsymbol{\phi}_{nt}$.

Our estimation procedure relies on an estimator $\hat{\boldsymbol{\Sigma}}^x(\theta)$ of the spectral density $\boldsymbol{\Sigma}^x(\theta)$. Consistency results quite naturally require some regularity assumptions on $\boldsymbol{\Sigma}^x(\theta)$ and the asymptotic behavior, as n and T tend to infinity, of $\hat{\boldsymbol{\Sigma}}^x(\theta)$. We assume (i) continuity of the spectral densities $\sigma^x_{ij}(\theta)$ and (ii) linear divergence for the first q eigenvalues of $\boldsymbol{\Sigma}^x(\theta)$. The latter condition is also assumed for the first q eigenvalues of $\boldsymbol{\Gamma}^\psi$.

Assumption A.6 *The functions $\theta \mapsto \sigma^x_{ij}(\theta)$ are continuous for all $i, j \in \mathbb{N}$.*

Assumption A.7 *There exist real numbers $a_1 > b_1 > a_2 > b_2 > \cdots > a_q > b_q > 0$, and an integer $\bar{n}$ such that, for $n > \bar{n}$, $a_s \geq \lambda^x_s(\theta)/n \geq b_s$, for $s = 1, \ldots, q$, and all $\theta \in [-\pi\ \pi]$.*

Assumption A.8 *There exist real numbers $h_1 > k_1 > h_2 > k_2 > \cdots > h_q > k_q > 0$, and an integer $\bar{n}$ such that, for $n > \bar{n}$, $h_s \geq \lambda^\psi_s/n \geq k_s$ for $s = 1, \ldots, q$.*

A consequence of Assumptions A.3 and A.6 is that $\boldsymbol{\xi}_t$ has a continuous spectral density as well, so that $\theta \mapsto \lambda^\xi_{n1}(\theta)$ also is continuous. It is easily seen that essential boundedness of $\bar{\lambda}^\xi_1(\theta) = \sup_n \lambda^\xi_{n1}(\theta)$, see Theorem A(i), then implies strict boundedness, i.e. the existence of a constant $\lambda > 0$ such that $\lambda^\xi_{n1}(\theta) \leq \lambda$ for all n and θ.

Turning to $\hat{\boldsymbol{\Sigma}}^x(\theta)$, let B_T be a sequence of positive integers,

$$\theta_h = \pi h {B_T}^{-1}, \qquad h = -B_T, -B_T + 1, \ldots,\ B_T - 1, B_T,$$

and denote by $\hat{\sigma}_{ij}(\theta_h)$ the Bartlett lag-window estimate

$$\hat{\sigma}^x_{ij}(\theta_h) = \sum_{k=-B_T}^{B_T} \hat{\gamma}^x_{ij,k} w_k e^{-\imath k\theta_h}, \qquad w_k = 1 - \frac{|k|}{B_T + 1}, \tag{3.1}$$

where $\hat{\gamma}^x_{ij,k}$ is the usual estimator of the cross-covariance $\gamma^x_{ij,k} = \mathrm{E}(x_{it}x_{j,t-k})$. We will need more than the usual consistency of the non-parametric estimator $\hat{\sigma}^x_{ij}$, namely that consistency be uniform with respect to the frequency θ. Very few papers are addressing that uniformity problem. Robinson (1991), for instance, gives conditions for uniform consistency but provides no rates. Uniformity with rates is obtained in Benktus (1985). Denote by f the spectral density of a Gaussian process $\{x_t\}$, and assume that $\theta \mapsto f(\theta)$ has r derivatives in $[-\pi, \pi]$. Let $\hat{f}^T(\theta)$ be a lag-window estimator of f based on an observed realization of length T, with bandwidth B_T of the order $(T/\log T)^{1/2\beta+1}$ as $T \to \infty$. Then, see Benktus (1985), Theorem 2.2 (statement 2),

$$\mathrm{E}\left(\sup_{\theta \in [-\pi \ \pi]} |\hat{f}^x(\theta) - f(\theta)| \right) \leq K(T/\log T)^{-\beta/2\beta+1}, \tag{3.2}$$

where $K > 0$, $\beta = \alpha + r$, $0 < \alpha \leq 1$. Recently, under fairly general conditions, Liu and Wu (2010) show that, provided that B_T is of the order of T^γ as $T \to \infty$ for some γ such that $1/3 < \gamma < 1$,

$$\max_{|h| \leq B_T} |\hat{f}(\theta_h) - f(\theta_h)|^2 = O_{\mathrm{P}}(B_T T^{-1} \log T) \quad \text{as} \quad T \to \infty \tag{3.3}$$

(see their Theorems 3, 4, 5 and Remark 5). Note that for small values of r the rates in (3.2) and (3.3), as functions of α and γ respectively, have a very similar range, whereas for $r \to \infty$ the rate in (3.2) tends to the standard parametric rate $T^{1/2}$.

A discussion of the conditions imposed in Benktus or Liu and Wu is of course outside the scope of the present paper. Rather, we directly suppose, in the spirit of their results, that:

Assumption A.9 *Let B_T be of the order of T^γ, with $1/3 < \gamma < 1$. Then*

$$\mathrm{E}\left(\max_{|h| \leq B_T} |\hat{\sigma}^x_{ij}(\theta_h) - \sigma^x_{ij}(\theta_h)|^2 \right) \leq K\left[B_T T^{-1} \log T \right]. \tag{3.4}$$

with K independent of i and j.

Note that: (i) we take it for granted that (3.2) or (3.3) can be extended to cross-spectra; (ii) we assume the same rate as in (3.3), but enhance it by bounding the moments, as in (3.2).

Using the assumption that K is independent of i and j in A.9 and the Markov Inequality, it is easily seen that

$$\begin{aligned} n^{-2} \sum_{i=1}^n \sum_{j=1}^n \max_{|h| \leq B_T} |\hat{\sigma}^x_{ij}(\theta_h) - \sigma^x_{ij}(\theta_h)|^2 &= O_P(B_T T^{-1} \log T) \\ n^{-1} \sum_{i=1}^n \max_{|h| \leq B_T} |\hat{\sigma}^x_{ij}(\theta_h) - \sigma^x_{ij}(\theta_h)|^2 &= O_P(B_T T^{-1} \log T), \end{aligned} \tag{3.5}$$

for all i, as $T \to \infty$ and $n \to \infty$. As the proof of our consistency results is based on (3.5), Assumption A.9 might be replaced by this weaker condition. Assumptions in the form of uniform bounds, like K in A.9, or bounds for cross-sectional averages, like in (3.5), play a crucial role in the literature on dynamic factor models, see e.g. Assumptions A, B, C and D (C1 in particular) in Bai and Ng (2002) and Bai (2003), or Assumptions F1 and M1 (F1c and M1c in particular) in Stock and Watson (2002b).

Based on $\hat{\mathbf{\Sigma}}^x(\theta)$, our estimate of the spectral density of $\boldsymbol{\chi}_{nt}$ is (see Forni et al. 2000)

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

where $\tilde{\mathbf{F}}$ denotes the transposed conjugate of $\mathbf{F}$. We then have the following result. Let $\zeta_{Tn} = \max(1/\sqrt{n},\ 1/\sqrt{T/B_T \log T})$.

Proposition 2 *Under Assumptions A.1 through A.9,*

$$\max_{|h|\le B_T} |\hat{\sigma}^\chi_{ij}(\theta_h) - \sigma^\chi_{ij}(\theta_h)| = O_{\mathrm{P}}\left(\zeta_{Tn}\right),$$

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

See Appendix B for a proof.

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

$$\hat{\gamma}^\chi_{ij,\ell} = \frac{\pi}{B_T}\sum_{s=-B_T+1}^{B_T} e^{\imath\ell\theta_s}\hat{\sigma}^\chi_{ij}(\theta_s). \tag{3.6}$$

Recalling that

$$\gamma^\chi_{ij,\ell} = \int_{-\pi}^{\pi} e^{\imath\ell\theta}\sigma^\chi_{ij}(\theta)d\theta,$$

we have

$$\begin{aligned}
|\hat{\gamma}^\chi_{ij,\ell} - \gamma^\chi_{ij,\ell}| &\le \frac{\pi}{B_T}\sum_{s=-B_T+1}^{B_T} |e^{\imath\ell\theta_s}\hat{\sigma}^\chi_{ij}(\theta_s) - e^{\imath\ell\theta_s}\sigma^\chi_{ij}(\theta_s)| \\
&\qquad + \left|\frac{\pi}{B_T}\sum_{s=-B_T+1}^{B_T} e^{\imath\ell\theta_s}\sigma^\chi_{ij}(\theta_s) - \int_{-\pi}^{\pi} e^{\imath\ell\theta}\sigma^\chi_{ij}(\theta)d\theta\right| \\
&\le \frac{\pi}{B_T}\sum_{s=-B_T+1}^{B_T} |\hat{\sigma}^\chi_{ij}(\theta_s) - \sigma^\chi_{ij}(\theta_s)| \\
&\qquad + \frac{\pi}{B_T}\sum_{s=-B_T+1}^{B_T} \max_{\theta_{s-1}\le\theta\le\theta_s} |e^{\imath\ell\theta_s}\sigma^\chi_{ij}(\theta_s) - e^{\imath\ell\theta}\sigma^\chi_{ij}(\theta)| \\
&\le 2\pi \max_{|s|\le B_T} |\hat{\sigma}^\chi_{ij}(\theta_s) - \sigma^\chi_{ij}(\theta_s)| + \frac{\pi G}{B_T}\sum_{s=-B_T+1}^{B_T} \max_{\theta_{s-1}\le\theta\le\theta_s} |e^{\imath\ell\theta_s} - e^{\imath\ell\theta}| \\
&\qquad + \frac{\pi}{B_T}\sum_{s=-B_T+1}^{B_T} \max_{\theta_{s-1}\le\theta\le\theta_s} |\sigma^\chi_{ij}(\theta_s) - \sigma^\chi_{ij}(\theta)| \\
&\le 2\pi \max_{|s|\le B_T} |\hat{\sigma}^\chi_{ij}(\theta_s) - \sigma^\chi_{ij}(\theta_s)| \\
&\quad + \frac{\pi G}{B_T}\sum_{s=-B_T+1}^{B_T} \Big(|e^{\imath\ell\theta_{s-1}} - e^{\imath\ell\theta^*_{s-1}}| + |e^{\imath\ell\theta^*_{s-1}} - e^{\imath\ell\theta_{s-1}}|\Big) \\
&\quad + \frac{\pi}{B_T}\sum_{s=-B_T+1}^{B_T} \Big(|\sigma^\chi_{ij}(\theta_{s-1}) - \sigma^\chi_{ij}(\theta^{**}_{s-1})| + |\sigma^\chi_{ij}(\theta^{**}_{s-1}) - \sigma^\chi_{ij}(\theta_s)|\Big),
\end{aligned} \tag{3.7}$$

where (i) G is the bound in Assumption A.5, (ii) θ^{*}_{s-1} and θ^{**}_{s-1} are points in the interval $[\theta_{s-1}\ \theta_s]$ where $|e^{i\ell\theta_s} - e^{i\ell\theta}|$ and $\sigma_{ij}(\theta_s) - \sigma_{ij}(\theta)|$, respectively, attain a maximum. Of course, the function $e^{\imath\ell\theta}$ is of bounded variation. Under Assumption A.3, the functions $\sigma^{\chi}_{ij}(\theta)$ are of bounded variation as well. This is sufficient for Propositions 3 and 4. However, the proof of Lemma 12 (see Appendix D), which is part of the proof of Proposition 5, requires that the functions σ_{ij} are of bounded variation uniformly in i and j. Precisely:

Assumption A.10 *There exists M such that for all i, j and w in $\mathbb{N}$, and all w-tuple $-\pi = \theta_0 < \theta_1 < \cdots < \theta_w = \pi$, we have*

$$\sum_{k=1}^{w} |\sigma^{\chi}_{ij}(\theta_k) - \sigma^{\chi}_{ij}(\theta_{k-1})| \leq M.$$

Using Proposition 2 and Assumption A.10, we obtain that $|\hat{\gamma}^{\chi}_{ij,\ell} - \gamma^{\chi}_{ij,\ell}|$ is of the form $O_{\mathrm{P}}(\zeta_{Tn}) + O(1/B_T)$. Because B_T is of the order of T^{γ} with $1/3 < \gamma < 1$, and $\zeta_{Tn} = \max(1/\sqrt{n},\ 1/\sqrt{T/B_T \log T})$, we have the following consistency result.

Proposition 3 *Under Assumptions A.1 through A.10, for each $\ell \geq 0$,*

$$|\hat{\gamma}^{\chi}_{ij,\ell} - \gamma^{\chi}_{ij,\ell}| = O_{\mathrm{P}}(\zeta_{Tn}), \tag{3.8}$$

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

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

The definition of the estimates $\hat{\mathbf{A}}^{[k]}$and $\hat{\mathbf{\Gamma}}^{\psi}_{jk}$ is straightforward from (2.18), (2.19) and (2.20). Denote by $\|\mathbf{F}\|$ the *spectral norm* of an $s_1 \times s_2$ matrix $\mathbf{F}$ (see Appendix B for details).

Proposition 4 *Under Assumptions A.1 through A.10,*

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

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

See Appendix C for the proof.

3.3 Estimation of $\underline{\mathbf{R}}$

We now turn to the crucial point of estimating representation (2.21). For $n \geq q$, we have $\mathbf{z}_{nt} = \boldsymbol{\psi}_{nt} + \boldsymbol{\phi}_{nt} = \mathbf{R}\mathbf{u}_t + \boldsymbol{\phi}_{nt}$, where $\boldsymbol{\psi}_{nt}$ has covariance matrix $\mathbf{R}\mathbf{R}' = \mathbf{P}^{\psi}\boldsymbol{\Lambda}^{\psi}\mathbf{P}^{\psi\prime} = \mathbf{P}^{\psi}(\boldsymbol{\Lambda}^{\psi})^{1/2}(\boldsymbol{\Lambda}^{\psi})^{1/2}\mathbf{P}^{\psi\prime}$; the columns of the $n \times q$ matrix $\mathbf{P}^{\psi}$ are the eigenvectors of $\mathbf{R}\mathbf{R}'$, and $\boldsymbol{\Lambda}^{\psi}$ is $q \times q$ with the corresponding eigenvalues on the main diagonal. Thus we have the representation

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

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

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

Proposition 5 *Under Assumptions A.1 through A.10,*

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

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

See Appendix D for a proof.

Let us point out again that the i-th row of $\mathcal{R}$ depends on n. Therefore, Proposition 5 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 $\mathbf{z}_t = \underline{\mathbf{R}}^{*}\mathbf{u}_t^{*} + \boldsymbol{\phi}_t$, or $\mathbf{z}_{nt} = \mathbf{R}^{*}\mathbf{u}_t^{*} + \boldsymbol{\phi}_t$, where the $n \times q$ matrices $\mathbf{R}^{*}$ are nested. In particular, starting with $\mathbf{x}_{nt} = \mathcal{R}\mathbf{v}_t + \boldsymbol{\phi}_{nt}$, there exists exactly one orthogonal matrix $\mathbf{G}(\mathcal{R})$ (actually $\mathbf{G}(\mathcal{R})$ only depends on the $q \times q$ upper submatrix of $\mathcal{R}$) such that $\mathbf{R}^* = \mathcal{R}\mathbf{G}(\mathcal{R})$. Thus, while the entries of $\mathcal{R}$ depend on n, the entries of $\mathcal{R}\mathbf{G}(\mathcal{R})$ do not.

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

Lastly, define the population *impulse-response functions* as the entries of the $n \times q$ matrix $\mathbf{B}(L) = \mathbf{A}(L)^{-1}\mathbf{R}^*$ and their estimators as those of $\hat{\mathbf{B}}(L) = \hat{\mathbf{A}}(L)^{-1}\hat{\mathbf{R}}^*$. Denoting by $B_{if}(L) = B_{if,0} + B_{if,1}L + \cdots$ and $\hat{B}_{if}(L) = \hat{B}_{if,0} + \hat{B}_{if,1}L + \cdots$, respectively, such entries, Propositions 4 and 5 imply that $|\hat{B}_{if,k} - B_{if,k}| = O_P(\zeta_{Tn})$ 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.[8]

3.4 Estimation of $\mathbf{v}_t$

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 6 *Under Assumptions A.1 through A.10,*

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

as $T \to \infty$ and $n \to \infty$, where $\hat{\mathbf{W}}^z$ is a $q \times q$ diagonal matrix, depending on n and T, whose diagonal entries equal either 1 or -1.

3.5 Alternative estimators for $\mathbf{A}^k(L)$ and $\mathbf{R}^k$

Under further assumptions on the idiosyncratic components, $\mathbf{A}^k(L)$ and $\mathbf{R}^k$ can be estimated consistently, as $T \to \infty$, by fully parametric methods. Let us return to the k-th $(q+1)$-dimensional block $\mathbf{A}^k(L)\boldsymbol{\chi}_t^k = \mathbf{R}^k\mathbf{u}_t$ and the corresponding equation for $\mathbf{x}_t^k$:

$$\mathbf{A}^k(L)\mathbf{x}_t^k = \mathbf{R}^k\mathbf{u}_t + \mathbf{A}^k(L)\boldsymbol{\xi}_t^k. \tag{3.9}$$

[8] All just-identifying rules considered in the SVAR literature can be dealt with along the same lines, see Forni, Giannone, Lippi and Reichlin, 2009.

Under the assumption that $\boldsymbol{\xi}_t^k$ is white noise, (3.9) is a VARMA with equal AR and MA orders, allowing direct estimation of $\mathbf{A}^k(L)$. Direct estimation of $\mathbf{A}^k(L)$ is also possible if $\boldsymbol{\xi}_t^k$ is a vector moving average. Assuming that $\boldsymbol{\xi}_t^k$ has a VARMA structure, $\mathbf{A}^k(L)$ and $\mathbf{R}^k$ in (3.9) can be estimated using unobserved components model's techniques. However, VARMA and unobserved components models consistently estimate $\mathbf{A}^k(L)$, $\mathbf{A}^k(L)$ and $\mathbf{R}^k$, respectively, as $T \to \infty$, but not $\mathbf{u}_t$. Consistent estimation of $\mathbf{u}_t$ requires that both T and n diverge.

Altissimo et al. (2009) estimate the α_i's and u_t in model (1.4) by means of an iterative procedure which starts by the estimation of α_i, equation by equation, from

$$(1 - \alpha_i L)x_{it} = a_i u_t + (1 - \alpha_i L)\xi_{it}.$$

In this particular case estimating q-dimensional instead of $(q+1)$-dimensional blocks is correct because in (1.4) u_t is fundamental for χ_{it} for all i. However, if $\chi_{it} = [c_i(L)/d_i(L)]u_t$, u_t is generically fundamental for 2-dimensional but not 1-dimensional blocks, nor has $c_i(L)/d_i(L)$ a finite inverse.

4 An empirical exercise

In the present section we compare the estimation performance of the methods studied in the present paper, referred to as FHLZ, and model (1.3), whose consistency has been studied in Forni et al. (2009), referred to as FGLR. Let us recall that both models assume rational spectral density for the common components, but FGLR also assumes finite dimension for the space spanned by the variables χ_{it} for any given t and $i \in \mathbb{N}$. Using a Monte Carlo simulation based on actual US macroeconomic data, we compare (i) impulse response functions and (ii) structural shocks, estimated using FHLZ and FGLR, respectively, under the same iterative identification scheme.

4.1 Simulation design

Let us firstly illustrate the simulation design. We use two macroeconomic panels. The first is the one used in Forni and Gambetti (2010a), with 101 US quarterly series, covering the period 1959 I - 2007 IV. The second is the one used in Forni and Gambetti (2010b), which is essentially an updating of the panel used in Stock and Watson (2002a, 2002b) and Bernanke, Boivin and Eliasz (2005). It includes 112 US monthly series between March 1973 and November 2007. Details on both panels and their treatment are reported in Appendix F.

We run the FGLR and FHLZ methods on both the quartely and the monthly panels. Here are some of the major features of this empirical study.

(1) Based on Hallin and Liška (2007) and Onatski (2009), we identify $q = 4$ for both panels.

(2) In FGLR, using the Bai and Ng (2002) IC2 criterion, we obtain, for the dimension of $\mathbf{F}_t$, $r = 12$ in the quarterly panel and $r = 16$ in the monthly panel. Moreover, identification via the BIC criterion of the number of lags in the VAR, see the second equation in (1.3), yields $p = 2$ for both panels.

(3) In FHLZ, the number of lags in each $(q+1)$-dimensional VAR is chosen by the BIC criterion for each VAR. In the estimation of the spectral density of $\mathbf{x}_{nt}$, we set the Bartlett lag-window size to 12 for quarterly data and 30 for monthly data, which is large enough to retain the most important cyclical auto- and cross-correlations.

(4) Our simulations require an estimate of the idiosyncratic components. For FGLR, we take the residuals of the projection of the x's on the first r principal components. For FHLZ, we take the residuals of the projection of x_{it} onto present, past and future values of the first q dynamic principal components, i.e. we apply the two-sided filters described in Forni et al. (2000) to get an estimate of the common components and take the deviations from the x's. The resulting poor end-of-sample estimation, see the Introduction, is harmless for the present analysis.

(5) Impulse-response functions and shocks are estimated by FHLZ and FGLR applying the same identification scheme. Precisely, we use a recursive identification scheme on 4 selected variables (see Section 3.2). With the quarterly data we use GDP, the GDP deflator, the federal funds rate and the Standard & Poor index of 500 Common Stocks; for the monthly data, we use Industrial Production, the CPI, the federal funds rate and the NAPM commodity price index. Similar schemes are widely used in the VAR literature on monetary policy (see e.g. Christiano, Eichenbaum and Evans, 1999).[9]

[9] We take the variables in the order specified above. The ordering however is irrelevant in the present context, since our measure of the estimation error (described below) is invariant to the application of

(6) Regarding FHLZ, whereas the population impulse-response functions and the shocks, given the identification scheme, do not depend on the particular grouping of the variables, some dependence occurs in the estimates. This is due to the estimation errors contained in the covariances used in VAR estimation, and possibly to incorrect specification of the number of lags. We deal with this problem by averaging the impulse-response functions obtained with a number of random permutations of the variables in the panel. We find that 30 random permutations are sufficient to stabilize the averages.

Denote by (IRF, $\xi)_{\text{FGLR}}$ and (IRF, $\xi)_{\text{FHLZ}}$ the impulse-response functions and idiosyncratic components estimated via FGLR and FHLZ, respectively. Based on (IRF, $\xi)_{\text{FGLR}}$, we generate 500 artificial quarterly panels and 200 artificial monthly panels as follows. Firstly, we produce 4 random independent standard normal shocks, filter them with the impulse-response functions, and add the resulting series to get the common components (the impulse-response functions are truncated at 20 lags for quarterly data and 48 lags for monthly data). Then we add artificial idiosyncratic components obtained by block bootstrapping (without overlapping) the idiosyncratic components estimated as in (4) above. We take blocks of 19 periods for quarterly data and 51 periods for monthly data. Block bootstrapping is intended to randomize the idiosyncratic components while preserving the idiosyncratic auto- and cross-correlation structure of macroeconomic time series. The same procedure is applied to obtain the 500 quarterly and 200 monthly panels based on (IRF,$\xi)_{\text{FHLZ}}$.

Lastly, the impulse-response functions are estimated for each artificial panel using the two competing methods, with the recursive identification scheme described under (5) above:

(a) The true number of structural shocks is assumed to be known, i.e. equal to 4, for both methods.

(b) For each artificial panel we set the parameters, r, p when using FGLR, the lag-window size and the length of the $(q+1)$-dimensional VAR's when using FHLZ, according to the criteria specified under (2) and (3) above. In particular, for FHLZ, the BIC criterion is applied separately to each of the of the $(q+1)$-dimensional VAR's, the (Bartlett) lag window size is 12 for quarterly data and 30

the same orthonormal transformations to both the target and the estimator.

for monthly data (we do not report results for different values of the lag-window size since they are fairly stable within the range 2-4 years). The number of random permutations of the variables in the panel, see (6) above, was set to 30 for all experiments. Results obtained by using the HQC criterion to determine the length of the VAR's, in both methods, are also reported. On the other hand, both AIC and the Bai and Ng (2002) IC1 criterion produce poor results, which are not reported.

(c) We also report results obtained using a grid of values for some parameters. For FGLR estimation, we report results for $r = 6, 12$, $p = 1, 2, 3, 6$ (quarterly data), $r = 8, 16$, $p = 2, 4, 6, 12$ (monthly data). For FHLZ, we report results obtained by setting the length of all the $(q + 1)$-dimensional VAR's to $1, 2, 3, 6$ (quarterly data), $2, 4, 6, 12$ (monthly data).

The estimation error for the impulse-response functions is defined as the normalized sum of the squared deviations of the estimated from the "true" impulse response coefficients respectively. Precisely, let $b_{if,h}$ and $\hat{b}_{if,h}$ be the true and estimated impulse-response coefficients, respectively, of variable i, shock f, lag h. Then the estimation error of the impulse response functions is measured by

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

The truncation lag H was set to 20 for quarterly data, to 48 for monthly data.

4.2 Results

Tables 1-4 report the average and the corresponding standard deviation (slanted figures) of the estimation error across the artificial panels. Boldface figures indicate the best result.

The first four columns report the results obtained using preassigned values for the lag in the VAR's, see (c) above, the last two columns are reporting the results obtained using the BIC or HQC criterion, see (b) above. The second and third row report results for FHLR when r is set to preassigned values, see (c) above, the fourth row reports the same when r is chosen according to $IC2$, see (b).

Comparing the best results provides an idea of the potential performance of the competing methods, independently of model selection techniques. We see that the

FHLZ outperforms FGLR when the DGP is FHLZ and in one of the cases in which the DGP is FGLR. However, the relevant comparison is in the last two columns, first and fourth rows. Underscored figures in the first row report results from FHLZ, with the VAR length chosen according to the BIC or the HQC criteria. Results from FGLR, with the VAR length chosen with the same criteria and r by $IC2$, are reported in the fourth row. The outcome does not change: FHLZ outperforms FGLR in three out of four cases. The standard deviations show that the performance of FHLZ is almost always less volatile than FGLR, often to a large extent.

Table 1: Average and standard deviation (slanted) of MSE across 500 artificial data set. Data generating Process (DGP): FGLR. Quarterly data.

	$p = 1$	$p = 2$	$p = 3$	$p = 6$	$p = BIC$	$p = HQC$
FHLZ	0.2494	0.1915	**0.1857**	0.2447	<u>0.2040</u>	<u>0.1928</u>
	0.0256	*0.0274*	*0.0281*	*0.0388*	*0.0266*	*0.0281*
FGLR r = 6	0.2468	0.2030	0.2276	0.2937	0.2288	0.2070
	0.0490	*0.0628*	*0.0714*	*0.0699*	*0.0604*	*0.0604*
FGLR r = 12	0.2137	**0.1862**	0.2163	0.3302	0.2137	0.1959
	0.0298	*0.0321*	*0.0349*	*0.0445*	*0.0298*	*0.0360*
FGLR r = IC2	0.2305	0.1931	0.2190	0.3095	<u>0.2226</u>	<u>0.2160</u>
	0.0369	*0.0476*	*0.0518*	*0.0764*	*0.0663*	*0.0824*

Table 2: Average and standard deviation (slanted) of MSE across 500 artificial data set. DGP: FHLZ. Quarterly data.

	$p = 1$	$p = 2$	$p = 3$	$p = 6$	$p = BIC$	$p = HQC$
FHLZ	0.1401	0.1186	0.1287	0.1740	**<u>0.1184</u>**	<u>0.1280</u>
	0.0179	*0.0182*	*0.0184*	*0.0232*	*0.0178*	*0.0193*
FGLR r = 6	0.1651	0.1665	0.1894	0.2659	0.1651	0.1660
	0.0204	*0.0232*	*0.0261*	*0.0325*	*0.0204*	*0.0210*
FGLR r = 12	0.1494	0.1631	0.1951	0.3149	**0.1494**	0.1546
	0.0205	*0.0239*	*0.0271*	*0.0344*	*0.0205*	*0.0350*
FGLR r = IC2	0.1585	0.1657	0.1932	0.2914	<u>0.1764</u>	<u>0.1862</u>
	0.0200	*0.0235*	*0.0300*	*0.0624*	*0.0732*	*0.0858*

Table 3: Average and standard deviation (slanted) of MSE across 200 artificial data set. DGP: FGLR. Monthly data.

	$p=2$	$p=4$	$p=6$	$p=12$	$p=BIC$	$p=HQC$
FHLZ	0.3003	0.2768	0.2797	0.3133	0.3012	**0.2760**
	0.0435	*0.0383*	*0.0386*	*0.0338*	*0.0400*	*0.0364*
FGLR r = 8	0.2435	0.2417	0.2606	0.3274	0.2603	0.2408
	0.0919	*0.0882*	*0.0914*	*0.0916*	*0.0955*	*0.0954*
FGLR r = 16	**0.2156**	0.2325	0.2649	0.3962	0.2562	0.2320
	0.0799	*0.0837*	*0.0833*	*0.0795*	*0.0905*	*0.0893*
FGLR r = IC2	0.2273	0.2286	0.2523	0.3412	0.2632	0.2417
	0.0820	*0.0723*	*0.0765*	*0.0893*	*0.0947*	*0.1006*

Table 4: Average and standard deviation (slanted) of MSE across 200 artificial data set. DGP: FHLZ. Monthly data.

	$p=2$	$p=4$	$p=6$	$p=12$	$p=BIC$	$p=HQC$
FHLZ	0.1226	0.1292	0.1394	0.1832	0.1228	**0.1220**
	0.0250	*0.0243*	*0.0237*	*0.0219*	*0.0236*	*0.0214*
FGLR r = 8	0.2890	0.3131	0.3423	0.4040	0.3018	0.2949
	0.1064	*0.0980*	*0.0978*	*0.0887*	*0.1129*	*0.1104*
FGLR r = 16	**0.2564**	0.2780	0.3148	0.4448	0.2619	0.2679
	0.0951	*0.0842*	*0.0823*	*0.0718*	*0.1067*	*0.1134*
FGLR r = IC2	0.2652	0.2872	0.3208	0.4134	0.2826	0.2928
	0.0906	*0.0912*	*0.0923*	*0.0849*	*0.0981*	*0.1178*

The results obtained when the artificial panels are generated by $(\text{IRF}, \xi)_{\text{FHLZ}}$, i.e. a fairly large advantage of FHLZ over FGLR, Tables 2 and 4, provide evidence that IRF_{FHLZ}, the impulse-response functions estimated by FHLZ with the actual panels, do not fulfill a finite-dimension restriction, neither strictly nor approximately. They moreover demonstrate the danger of a *static* approach when such finite-dimension restriction is not supported by the data at hand.

When the artificial panels are generated using $(\text{IRF}, \xi)_{\text{FGLR}}$, FHLZ is, apart from very special cases, a less parsimonious representation as compared to FGLR. In particular, if $r \geq 10$ in (1.3), the length of the $(q+1)$-dimensional VAR's in the corresponding

FHLZ representation is larger than 10 (this can be seen by elementary algebraic manipulation). On the other hand, *as a consequence of the value of T in the datasets under study,* the length of the $(q+1)$-dimensional VAR's determined by the BIC or the HQC criterion, when estimating with FHLZ the artificial panels generated by $(\text{IRF}, \xi)_{\text{FGLR}}$, is either 1 or 2. Thus, a quite parsimonious FHLZ specification is sufficient to account for the dynamics in the panel, which is a possible explanation for the unexpected better performance of FHLZ even in Table 1.

5 Conclusions

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

The rational spectral density assumption has also the important consequences that $\boldsymbol{\chi}_t$ has a finite autoregressive representation and that the dynamic factor model can be transformed into the static model $\mathbf{z}_t = \mathbf{R}\mathbf{v}_t + \boldsymbol{\phi}_t$, where $\mathbf{z}_t = \mathbf{A}(L)\mathbf{x}_t$. We construct estimators for $\mathbf{A}(L)$, $\mathbf{R}$ and $\mathbf{v}_t$ starting with a standard non-parametric 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 (1.3). Section 4 provides important empirical support to the richer dynamic structure of unrestricted GDFM's.

References

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

[2] Altissimo, F., R. Cristadoro, M. Forni, M. Lippi and G. Veronese, (2010). New Eurocoin: tracking economic growth in real time, *The Review of Economics and Statistics,* **92**, 1024-1034.

[3] Altissimo F., Mojon B. and P. Zaffaroni, (2009). Can aggregation explain the persistence of inflation?, *Journal of Monetary Economics*, **56**, 231-241.

[4] 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.

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

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

[7] Bai, J. (2003). Inferential theory for factor models of large dimensions, *Econometrica* **71**, 135-171.

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

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

[10] Benktus, R. (1985). Rate of uniform convergence of statistical estimators of spectral density in spaces of differentiable functions, *Lietuvos Matematikos Rinkinys*, **25**, 17-31 (English translation: Plenum Publishing Corporation, 1986).

[11] Bernanke, B. S., and J. Boivin (2003). Monetary policy in a data-rich environment, *Journal of Monetary Economics* **50**, 525546.

[12] Bernanke, B. S., J. Boivin and P. Eliasz (2005). Measuring monetary policy: a factor augmented autoregressive (FAVAR) approach, *The Quarterly Journal of Economics* **120**, 387-422.

[13] Boivin, J. and S. Ng (2006). Are more data always better for factor analysis?, *Journal of Econometrics* **132**, 169-194.

[14] Boivin, J., Giannoni, M.P. and I. Mihov (2009). Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data, *American Economic Review* **99**, 350-384.

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

[16] Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models, *Econometrica* **51**, 1281-1304.

[17] Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets, *Econometrica* **51**, 1305-1324.

[18] Christiano, L., Eichenbaum, M. and C.L. Evans (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of Macroeconomics*, Chapter 02, 65-148, J. B. Taylor & M. Woodford (ed.).

[19] Corielli, F. and M. Marcellino (2006). Factor based index tracking, *Journal of Banking & Finance* **30**, 2215-2233.

[20] Cristadoro, R., Forni, M., Reichlin, L. and G. Veronese (2005). A Core Inflation Indicator for the Euro Area, *Journal of Money, Credit and Banking*, 539-560.

[21] Dunford, N. and J. T. Schwartz, (1988). *Linear Operators*, Part I. New York: Wiley.

[22] Eickmeier, S. (2007). Business cycle transmission from the US to Germany–A structural factor approach, *European Economic Review* **51** 521-551.

[23] Favero, C., M. Marcellino, and F. Neglia (2005). Principal components at work: empirical analysis of monetary policy with large data sets, *Journal of Applied Econometrics* **20**, 603-620.

[24] Forni, M. and L. Gambetti, (2010a). Macroeconomic shocks and the business cycle: a structural factor model approach, CEPR Discussion Paper Series no. 7692.

[25] Forni, M. and L. Gambetti (2010b) The dynamic effects of monetary policy: A structural factor model approach, *Journal of Monetary Economics* **57**, 203-216.

[26] 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.

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

[28] 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.

[29] 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.

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

[31] Forni, M. and M. Lippi (2011). The generalized dynamic factor model: one-sided representation results, *Journal of Econometrics* **163**, 23-28.

[32] Geweke, J. (1977). The dynamic factor analysis of economic time series, in D.J. Aigner and A.S. Goldberger, Eds., *Latent Variables in Socio-Economic Models.* Amsterdam: North Holland.

[33] Giannone, D., L. Reichlin and L. Sala (2005). Monetary policy in real time, in *NBER Macroeconomic Annual 2004*, eds. M. Gertler and K. Rogoff, Cambridge, MA: MIT Press, 161200.

[34] Giannone, D., Reichlin, L. and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data, *Journal of Monetary Economics* **55**, 665-676.

[35] 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.

[36] Hallin M., Ch. Mathias, H. Pirotte and D. Veredas (2011). Market liquidity as dynamic factors, *Journal of Econometrics* **163**, 42-50.

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

[38] Ludvigson, S.C. and S. Ng (2007). The empirical risk-return relation: a factor analysis approach, *Journal of Financial Economics* **83**, 171-222.

[39] Ludvigson, S.C. and S. Ng (2009). Macro factors in bond risk premia, *The Review of Financial Studies* **22**, 5027-5067.

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

[41] Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory*, New York: Wiley.

[42] Robinson, P. M. (1991). Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models, *Econometrica* **59**, 1329-1363.

[43] Rozanov, Yu. (1967). *Stationary Random processes*, San Francisco: Holden Day.

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

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

[46] 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.

[47] Stock, J.H., and M.W. Watson (2005). Implications of dynamic factor models for VAR analysis, NBER Working Papers no. 11467.

[48] van der Waerden, B.L. (1953). *Modern Algebra*, Vol. I. New York: Frederick Ungar.

Appendix

A Proof of Proposition 1

Consider first the case $s_2 = 0$, so that $\mathbf{y}_t$ is a moving average. Setting $s = s_1$, we have

$$\mathbf{y}_t = \mathbf{C}_0\mathbf{v}_t + \mathbf{C}_1\mathbf{v}_{t-1} + \cdots + \mathbf{C}_s\mathbf{v}_{t-s} = \mathbf{C}(L)\mathbf{v}_t, \tag{A.1}$$

Consider the stack

$$(\mathbf{y}_{t-1}\ \mathbf{y}_{t-2}\ \cdots\ \mathbf{y}_{t-S})' = \mathcal{C}_S(\mathbf{v}_{t-1}\ \mathbf{v}_{t-2}\ \cdots\ \mathbf{v}_{t-S-s})' \tag{A.2}$$

where

$$\mathcal{C}_S = \begin{pmatrix} \mathbf{C}_0 & \mathbf{C}_1 & \cdots & \mathbf{C}_s & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_0 & \cdots & \mathbf{C}_{s-1} & \mathbf{C}_s & \cdots & \mathbf{0} \\ \vdots & & & & & \ddots & \\ \mathbf{0} & \mathbf{0} & \cdots & & & \cdots & \mathbf{C}_s \end{pmatrix}. \tag{A.3}$$

is $(q+1)S \times q(S+s)$. Setting $S = sq$, $\mathcal{C}_S$ is square. If it is non singular, then $(\mathbf{v}_{t-1}\ \mathbf{v}_{t-2}\ \cdots\ \mathbf{v}_{t-S-s})' = \mathcal{C}_S^{-1}(\mathbf{y}_{t-1}\ \mathbf{y}_{t-2}\ \cdots\ \mathbf{y}_{t-S})'$. Substituting into (A.1),

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_S\mathbf{y}_{t-S} + \mathbf{C}_0\mathbf{v}_t. \tag{A.4}$$

Note that (2.6) implies that $\mathbf{v}_t$ is orthogonal to $\mathbf{y}_{t-k}$ for $k > 0$. Thus, (A.4) is the orthogonal projection of $\mathbf{y}_t$ on its past values. Moreover, non-singularity of $\mathcal{C}_S$ implies that the entries of the left hand side of (A.2) are linearly independent. As a consequence, (A.4) is the unique AR representation of $\mathbf{y}_t$ of order less than or equal to S, up to an orthogonal transformation of $\mathbf{v}_t$ and $\mathbf{C}_0$.

It remains to prove that $\mathcal{C}_S$ is non singular for generic values of $\mathbf{C}_j$. Note that the determinant of $\mathcal{C}_S$ is a polynomial in the parameters, and is therefore either zero for all parameters' values or generically non zero. Thus, if we find a particular value of the parameters for which $\det \mathcal{C}_S \neq 0$ we can conclude that $\mathcal{C}_S$ is generically non singular.

Suppose that, for some $\boldsymbol{\alpha} = (\alpha_{01} \cdots \alpha_{0,q+1}; \alpha_{11} \cdots \alpha_{1,q+1}; \cdots; \alpha_{S-1,1} \cdots \alpha_{S-1,q+1}) \neq \mathbf{0}$,

$$\boldsymbol{\alpha}\mathcal{C}_S = 0. \tag{A.5}$$

Setting $\boldsymbol{\alpha}_i = (\alpha_{i1} \ \cdots \ \alpha_{i,q+1})$, (A.5) can be rewritten as

$$\begin{aligned}\boldsymbol{\alpha}_0(\mathbf{C}_0 + \mathbf{C}_1 L + \cdots + \mathbf{C}_s L^s) &+ \boldsymbol{\alpha}_1 L(\mathbf{C}_0 + \mathbf{C}_1 L + \cdots + \mathbf{C}_s L^s) \\ &+ \cdots + \boldsymbol{\alpha}_{S-1} L^{S-1}(\mathbf{C}_0 + \mathbf{C}_1 L + \cdots + \mathbf{C}_s L^s) \\ &= (\boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 L + \cdots + \boldsymbol{\alpha}_{S-1} L^{S-1})(\mathbf{C}_0 + \mathbf{C}_1 L + \cdots + \mathbf{C}_s L^s) = \mathbf{0},\end{aligned}$$

that is,

$$(\beta_1(L) \ \beta_2(L) \ \cdots \ \beta_{q+1}(L))\mathbf{C}(L) = \mathbf{0}, \tag{A.6}$$

where $\beta_j(L) = \alpha_{0j} + \alpha_{1j} L + \cdots + \alpha_{S-1,j} L^{S-1}$. Thus, (A.5) is equivalent to the existence of $q+1$ scalar polynomials $\beta_j(L)$, of degree $S-1$, such that $\beta_j(L) \neq 0$ for some j and (A.6) holds.

Let us now construct a point in the parameter space as follows. Let $d_i(L)$, for $i = 1, \ldots, q$, denote polynomials of degree s such that $d_i(L)$ and $d_j(L)$ have no common roots for $i \neq j$. Put

$$\mathbf{C}(L) = \begin{pmatrix} d_1(L) & 0 & \cdots & 0 \\ 0 & d_2(L) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_q(L) \\ 1 & 1 & \cdots & 1 \end{pmatrix}. \tag{A.7}$$

If there exist $\beta_i(L)$, $i = 1, \ldots, q+1$, not all zero, such that (A.6) holds, then, for $i \leq q$, $d_i(L)\beta_i(L) = -\beta_{q+1}(L)$. Therefore, the set of roots of $\beta_{q+1}(L)$ includes those of the polynomials $d_i(L)$ for $i \leq q$. But then, given our assumption on the roots of the $d_i(L)$'s, either $\beta_i(L) = 0$ for all $i = 1, \ldots, q+1$, or the degree of $\beta_{q+1}(L)$ is at least $qs = S$. Thus, $\det \mathcal{C}_S \neq 0$ in this case, hence also generically.

Let us now turn to the general rational case

$$\mathbf{y}_t = \mathbf{E}(L)\mathbf{v}_t, \qquad t \in \mathbb{Z} \tag{A.8}$$

where

$$e_{if}(L) = \frac{c_{if}(L)}{d_{if}(L)} = \frac{c_{if,0} + c_{if,1} L + \cdots + c_{if,s_1} L^{s_1}}{1 + d_{if,1} L + \cdots + d_{if,s_2} L^{s_2}}.$$

Rewrite (A.8) as

$$\begin{pmatrix} h_1(L) & 0 & \cdots & 0 \\ 0 & h_2(L) & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & h_{q+1}(L) \end{pmatrix} \mathbf{y}_t = \mathbf{G}(L)\mathbf{v}_t,$$

where $h_i(L) = \prod_{f=1}^{q} d_{if}(L)$ and $g_{if}(L) = c_{if}(L)h_i(L)/d_{if}(L)$. For generic parameter values, the degrees of $h_j(L)$ and $g_{ik}(L)$ are qs_2 and $s_1 + (q-1)s_2$, respectively. Now consider the moving average $\mathbf{G}(L)\mathbf{v}_t$. Select a point in the parameter space such that

(i) $c_{if}(L) = 0$ if $i \neq f$ and $i \leq q$, which implies $g_{if}(L) = 0$ if $i \neq f$ and $i \leq q$,

(ii) $g_{ii}(L)$ has degree $s_1 + (q-1)s_2$ for $i \leq q$,

(iii) $g_{ii}(L)$ and $g_{jj}(L)$ have no roots in common for $i \neq j$,

(iv) $g_{q+1,f}(L) = 1$ for $f = 1, \dots, q$.

This reproduces the situation in (A.7). As the entries of $\mathbf{G}(L)$ are polynomial functions of $p \in \Pi^{q+1}$, for generic parameter values, $\mathbf{G}(L)\mathbf{v}_t$ has an autoregressive representation of order not greater than $[s_1 + (q-1)s_2]q$. This implies that, for generic values of the parameters, $\mathbf{y}_t$ has an autoregressive representation of order not greater than

$$S = qs_2 + [s_1 + (q-1)s_2]q = qs_1 + q^2 s_2.$$

To prove uniqueness we now show that the entries of $\mathbf{y}_{t-1}, \ \mathbf{y}_{t-2}, \ \dots, \ \mathbf{y}_{t-S}$, still generically, are linearly independent, which is equivalent to proving that, if $\beta_j(L)$, $j = 1, \dots, q+1$, are polynomials in L and

$$(\beta_1(L) \ \beta_2(L) \ \cdots \ \beta_{q+1}(L))\mathbf{E}(L) = 0, \tag{A.9}$$

then, for generic values of the parameters, either $\beta_j(L) = 0$ for all j, or the degree of $\beta_j(L)$ is greater than $S - 1$ for some j. The sequel of the proof is subdivided into seven steps, numbered (I) through (VII).

(I) Let $\mathbf{E}_q(L)$ be the square submatrix obtained by dropping $\mathbf{E}(L)$'s last row. Rationality of the entries of $\mathbf{E}(L)$ implies that $\det(\mathbf{E}_q(z)) = 0$ either for all $z \in \mathbb{C}$ or for a finite subset of $\mathbb{C}$. On the other hand, if $\det(\mathbf{E}_q(z)) = 0$ for all $z \in \mathbb{C}$, then the parameters fulfill a set of polynomial equations. Obviously, there exist parameter values such that

$$\det(\mathbf{E}_q(z)) = 0 \text{ for a finite subset of } \mathbb{C},$$

i.e. such that $\mathbf{E}_q(z)$ is a non-zero rational function. Thus, for the parameter belonging to $\Pi^{q+1} - \mathcal{M}_1$, where $\mathcal{M}_1$ is a nowhere dense subset of Π^{q+1}, $\mathbf{E}_q(z)$ is a non-zero rational function, and hence $[\mathbf{E}_q(L)]^{-1}$ is well defined. As a consequence, still for parameter values in $\Pi^{q+1} - \mathcal{M}_1$, the system of equations

$$(\rho_1(L) \ \rho_2(L) \ \cdots \ \rho_q(L))\mathbf{E}_q(L) = (e_{q+1,1}(L) \ e_{q+1,2}(L) \ \cdots \ e_{q+1,q}(L))$$

in the unknown rational functions $\rho_j(L)$ has the unique solution

$$(\tau_1(L)\ \tau_2(L)\ \cdots\ \tau_q(L)) = (e_{q+1,1}(L)\ e_{q+1,2}(L)\ \cdots\ e_{q+1,q}(L))[\mathbf{E}_q(L)]^{-1}.$$

(II) There exists a nowhere dense in Π^{q+1} set $\mathcal{M}_2$ such that, for parameter values in $\Pi^{q+1} - \mathcal{M}_2$, $\det(\mathbf{E}_q(L)) = h(L)/\prod_{i,j=1}^q d_{ij}(L)$, where $h(L)$ has degree $qs_1 + (q^2 - q)s_2$.

(III) There exists a nowhere dense in Π^{q+1} set $\mathcal{M}_3$ such that, for parameter values in $\Pi^{q+1} - \mathcal{M}_3$, the (i, j) entry of the adjoint matrix of $\mathbf{E}_q(L)$ can be written as

$$h_{ij}(L)/\prod_{\substack{h,k=1,\dots q\\ h\neq j,\ k\neq i}} d_{hk}(L),$$

the degrees of numerator and denominator being $(q-1)s_1 + [(q-1)^2 - (q-1)]s_2$ and $(q-1)^2 s_2$, respectively.

(IV) There exists a nowhere dense in Π^{q+1} set $\mathcal{M}_4$ such that, for parameter values in $\Pi^{q+1} - \mathcal{M}_4$, the entries of $[\mathbf{E}_q(L)]^{-1}$ can be written as

$$h_{ij}(L)\prod_{\substack{h,j=1,\dots,q\\ h=j\ \text{or}\ k=i}} d_{hk}(L)/h(L) = \tilde{h}_{ij}(L)/h(L),$$

where the degrees of the numerator and denominator are $(q-1)s_1 + (q^2 - (q-1))s_2$ and $qs_1 + (q^2 - q)s_2$, respectively.

(V) There exists a nowhere dense in Π^{q+1} set $\mathcal{M}_5$ such that, for parameter values in $\Pi^{q+1} - \mathcal{M}_5$,

$$\tau_k(L) = \sum_{i=1}^q \frac{c_{q+1,i}(L)\tilde{h}_{ik}(L)}{d_{q+1,i}(L)h(L)} = \frac{\sum_{i=1}^q c_{q+1,i}(L)\tilde{h}_{ik}(L)\prod_{\substack{j=1,\dots,q\\ j\neq i}} d_{q+1,j}(L)}{h(L)\prod_{i=1}^q d_{q+1,i}(L)} = \nu_k(L)\delta(L),$$

where both $\nu_k(L)$ and $\delta(L)$ are polynomials of degree $S = qs_1 + q^2 s_2$.

(VI) Moreover, for generic values of the parameters, $\nu_k(L)$ and $\delta(L)$ have no roots in common. To show this, recall that

$$\nu_k(z) = \nu_{k,S}z^S + \nu_{k,S-1}z^{S-1} + \cdots + \nu_{k,0} \quad \text{and} \quad \delta(z) = \delta_S z^S + \delta_{S-1}z^{S-1} + \cdots + \delta_0,$$

both of degree S, have roots in common if and only if their *resultant* vanishes. That *resultant* is a homogeneous polynomial of degree S in the coefficients $\nu_{k,j}$ and δ_j, involving the term $\nu_{k,S}^S\delta_0^S$ (see van der Waerden 1953, pp. 83-5). All other terms contain powers $\nu_{k,S}^{S-h}$ of $\nu_{k,S}$, with $0 < h \leq S$. We have

$$\nu_{k,S}^S\delta_0^S = \left[\sum_{i=1}^q c_{q+1,i,s_1}\tilde{h}_{ik,g}\prod_{\substack{j=1,\dots,q\\ j\neq i}} d_{q+1,j,s_2}\right]^S h(0)^S = c_{q+1,1,s_1}^S\left[\tilde{h}_{1k,g}^S\prod_{j=2,\dots,q} d_{q+1,j,s_2}^S h(0)^S\right] + \dots, \tag{A.10}$$

where $\tilde{h}_{ik,g}$ is the coefficient of order g of $\tilde{h}(z)$ and $g = (q-1)s_1 + (q^2 - (q-1))s_2$. Note that $h(z)$ and $\tilde{h}_{ik}(z)$ do not contain any of the parameters $c_{q+1,i,h}$. As a consequence, all other terms in (A.10) and in the resultant of $\nu_k(L)$ and $\delta(L)$ contain powers c_{q+1,i,s_1}^{S-h} of c_{q+1,i,s_1}, with $0 < h \leq S$. Moreover, the coefficient of $c_{q+1,1,s_1}^{S}$ in (A.10) is generically non zero. Thus, the resultant of $\nu_k(z)$ and $\delta(z)$ does not vanish everywhere in Π^{q+1}, and therefore vanishes only on a nowhere dense subset. Summing up, there exists a nowhere dense set $\mathcal{M}_6$ in Π^{q+1}, such that, for parameter values in $\Pi^{q+1} - \mathcal{M}_6$, none of the resultants of $\delta(z)$ and $\nu_k(z)$, $k = 1, 2, \ldots, q$ vanishes.

(VII) Lastly, if the $\beta_k(L)$'s are such that (A.9) holds, then $\tau_k(L) = -\beta_k(L)/\beta_{q+1}(L)$. The results above imply that the degree of the polynomials $\beta_j(L)$ is at least S for parameter values in $\Pi^{q+1} - \cup_{k=1}^{6}\mathcal{M}_k$. Q.E.D.

B Proof of Proposition 2

The proof below closely follows Forni et al. (2009). Denote by $\mu_j(\mathbf{A})$, $j = 1, 2, \ldots, s$, the (real) eigenvalues, in decreasing order, of a complex $s \times s$ Hermitian matrix $\mathbf{A}$, and by $\|\mathbf{B}\| = \sqrt{\mu_1(\tilde{\mathbf{B}}\mathbf{B})}$ the spectral norm of an $s_1 \times s_2$ matrix $\mathbf{B}$, which coincides with the Euclidean norm of $\mathbf{B}$ in case $\mathbf{B}$ is a row matrix. 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 also will use of the following inequality: for any two $s \times s$ Hermitian matrices $\mathbf{A}_1$ and $\mathbf{A}_2$,

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

This result, also known as Weyl's inequality, is an obvious consequence of Lancaster and Tismenetsky (1985), p. 301 (see also Forni and Lippi 2001, Fact M and Forni et al. 2009, Appendix). In particular, as $\mathbf{\Sigma}^x(\theta) = \mathbf{\Sigma}^\chi(\theta) + \mathbf{\Sigma}^x i(\theta)$, we have

$$|\mu_1(\mathbf{\Sigma}^x(\theta)) - \mu_1(\mathbf{\Sigma}^x(\theta))| \leq \mu_1(\mathbf{\Sigma}^\xi(\theta)). \tag{B.3}$$

The proof of Proposition 2 is divided into several intermediate propositions. Let $a_1 < a_2 < \cdots < a_s$ be integers, and put $\mathbf{M} = \{a_1, a_2, \cdots, a_s\}$. Denote by $\mathcal{S}_\mathbf{M}$ the $n \times s$ matrix with 1 in entries (a_j, j) and zero elsewhere, and define $\rho_T = T/B_T \log T$. As most of the arguments below depend on equalities and inequalities that hold for all

$\theta \in [-\pi\ \pi]$, the notation has been simplified by dropping θ. Moreover, properties holding for $\max_{|h|\le B_T} F(\theta_h)$, where F is some function of θ, are often phrased as holding for F *uniformly in* θ. Lastly, all lemmas in this Appendix hold, and are proved under Assumptions 1 through 9.

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

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

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

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

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

Proof. Since

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

Because

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

statement (i) follows from (3.5). Similarly, since

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

statement (ii) also follows from (3.5). As for (iii), orthogonality of common and idiosyncratic components at all leads and lags implies $\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi = \hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^x + \mathbf{\Sigma}^\xi$, so that, by the triangle inequality for matrix norm, $\|\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi\| \le \|\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^x\| + \|\mathbf{\Sigma}^\xi\|$. The statement follows from (i) and the fact that $\|\mathbf{\Sigma}^\xi\| = \lambda_1^\xi$ is bounded. Statement (iv) is obtained in a similar way, using (ii) instead of (i). QED

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

(i) $\max_{|h|\le B_T} n^{-1}\left|\hat{\lambda}^x_f - \lambda^\chi_f\right| = O_{\mathrm{P}}(\max(n^{-1}, \rho_T^{-1/2}))$ *for* $f = 1, 2, \ldots, q$;

(ii) Letting

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

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

(iii) $\max_{|h|\le B_T}\left\|n^{-1}\hat{\mathbf{\Lambda}}^x\hat{\mathbf{G}}^x - \mathbf{I}_q\right\| = O_{\mathrm{P}}(\max(n^{-1}, \rho_T^{-1/2}))$.

PROOF. Setting $\mathbf{A}_1 = \mathbf{\Sigma}^\chi$ and $\mathbf{A}_2 = \hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi$, (B.2) yields $|\hat{\lambda}^x_f - \lambda^\chi_f| \leq \|\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi\|$; hence, statement (i) follows from Lemma 3 (iii). Boundedness of $n^{-1}\|\mathbf{\Lambda}^\chi\|$ and $\|\mathbf{G}^\chi\|$, uniformly in θ, is a consequence of Assumption A.7. Boundedness in probability of $n^{-1}\|\hat{\mathbf{\Lambda}}^x\|$ and $\|\hat{\mathbf{G}}^x\|$, uniformly in θ, follows from statement (i). Statement (iii) thus follows from (i) and Assumption A.7. Q.E.D.

Lemma 5 *As $T \to \infty$ and $n \to \infty$,*

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

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

(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* $\max_{|h|\leq B_T} \|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^\chi - \hat{\mathbf{W}}_q\| = O_{\rm P}(\max(n^{-1}, \rho_T^{-1/2}))$.

PROOF. By (B.1), $\|\tilde{\mathbf{P}}^\chi\hat{\mathbf{P}}^x\hat{\mathbf{\Lambda}}^x - \mathbf{\Lambda}^\chi\tilde{\mathbf{P}}^\chi\hat{\mathbf{P}}^x\| = \|\tilde{\mathbf{P}}^\chi(\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi)\hat{\mathbf{P}}^x\| \leq \|\hat{\mathbf{\Sigma}}^x - \mathbf{\Sigma}^\chi\|$. Statement (i) thus follows from Lemma 3 (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 $\|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^\chi\tilde{\mathbf{P}}^\chi\hat{\mathbf{P}}^x - \mathbf{I}_q\| \leq \|\mathbf{a}-\mathbf{b}\| + \|\mathbf{b}-\mathbf{c}\| + \|\mathbf{c}-\mathbf{d}\| + \|\mathbf{d}-\ \mathbf{f}\|$. Using Lemma 4, statement (i), 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 that inequality can be shown to be $O_{\rm P}(\max(n^{-1}, \rho_T^{-1/2}))$, uniformly in θ. As for (iii), note that, from statement (i), $n^{-1}\tilde{\hat{\mathbf{P}}}^x_h\mathbf{P}^\chi_k(\lambda^\chi_k - \hat{\lambda}^x_h) = O_{\rm P}(\max(n^{-1}, \rho_T{}^{-1/2}))$. Assumption A.7 implies that, for $h \neq k$, $\tilde{\hat{\mathbf{P}}}^x_h\mathbf{P}^\chi_k = O_{\rm P}(\max(n^{-1}, \rho_T{}^{-1/2}))$. This and the fact that, in view of statement (ii), $\sum_{f=1}^q |\tilde{\hat{\mathbf{P}}}^x_h\mathbf{P}^\chi_f|^2 - 1 = O_{\rm P}(\max(n^{-1}, \rho_T{}^{-1/2}))$ implies that

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

The conclusion follows. Q.E.D.

Note that Lemma 5 clearly also holds for $n^{-1}\|\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^\chi\mathbf{\Lambda}^\chi - \hat{\mathbf{\Lambda}}^x\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^\chi\|$, $\|\tilde{\mathbf{P}}^\chi\hat{\mathbf{P}}^x\tilde{\hat{\mathbf{P}}}^x\mathbf{P}^\chi - \mathbf{I}_q\|$ and $\|\tilde{\mathbf{P}}^\chi\hat{\mathbf{P}}^x - \tilde{\hat{\mathbf{W}}}_q\|$.

Lemma 6 *Given* $\mathbf{M}$*, as $T \to \infty$ and $n \to \infty$,*

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

PROOF. We have

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

By Lemma 4 (i), thus, we only need to prove that

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

Firstly, we show that, uniformly in θ,

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

Assumption A.5 implies that $\sigma^{\chi}_{ii} = \sum_{f=1}^{q}\lambda^{\chi}_{f}|p^{\chi}_{if}|^2 = O(1)$, uniformly in θ. As all the terms in the sum are positive, $\lambda^{\chi}_{f}|p^{\chi}_{if}|^2 = (\lambda^{\chi}_{f}/n)n|p^{\chi}_{if}|^2$ also is $O(1)$, uniformly in θ. Assumption A.7, Theorem A(i) and (B.3) imply that λ^{χ}_{f}/n is bounded away from zero uniformly in θ, so that $n|p^{\chi}_{if}|^2$ must be $O(1)$, uniformly in θ. Hence, the eigenvalues of $n\mathcal{S}'_{\mathbf{M}}\mathbf{P}^{\chi}\tilde{\mathbf{P}}^{\chi}\mathcal{S}_{\mathbf{M}}$ are $O(1)$ uniformly in θ; (B.5) follows. Next, define

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

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

and

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

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

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 6 as

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

In the same way, the result of Lemma 5(iii) will be referred to as

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

Proposition 3 now follows from the fact that $\hat{\mathbf{\Sigma}}^{x} = \hat{\mathbf{P}}^{x}\hat{\mathbf{\Lambda}}^{x}\tilde{\hat{\mathbf{P}}}^{x}$ and $\mathbf{\Sigma}^{\chi} = \mathbf{P}^{\chi}\mathbf{\Lambda}^{\chi}\tilde{\mathbf{P}}^{\chi}$.

C Proof of Proposition 4

Firstly, note that, as the last term in (3.7) contains

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

convergence in (3.8) is not uniform with respect to ℓ. However, estimation of the matrices $\mathbf{B}_k^{\chi}$ and $\mathbf{C}_{jk}^{\chi}$ only requires the covariances $\hat{\gamma}_{ij,\ell}$ with $\ell \le S$, where S is finite. Therefore, Proposition 3 implies that $\|\hat{\mathbf{B}}_k^{\chi} - \mathbf{B}_k^{\chi}\|$ and $\|\hat{\mathbf{C}}_{jk}^{\chi} - \mathbf{C}_{jk}^{\chi}\|$ are $O_{\mathrm{P}}(\max(n^{-1/2}, \rho_T^{-1/2}))$. From (2.20), applying (B.1),

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

By Assumption A.5, $\|\mathbf{B}_k^{\chi}\| \le W$ for some $W > 0$, so that $\|\hat{\mathbf{B}}_k^{\chi}\|$ is bounded in probability. By Assumptions A.4 and A.5, $\| (\mathbf{C}_{kk}^{\chi})^{-1} \| \le W_1$ for some $W_1 > 0$. Observing that the entries of $(\mathbf{C}_{kk}^{\chi})^{-1}$ are rational functions of the entries of $\mathbf{C}_{kk}^{\chi}$, and that $\det(\mathbf{C}_{kk}^{\chi}) > 0$ by Assumption A.4, Proposition 3 implies that $\|(\hat{\mathbf{C}}_{kk}^{\chi})^{-1} - (\mathbf{C}_{kk}^{\chi})^{-1}\|$ is $O_{\mathrm{P}}(\max(n^{-1/2}, \rho_T^{-1/2}))$. Thus $\|\hat{\mathbf{A}}^{[k]} - \mathbf{A}^{[k]}\|$ is $O_{\mathrm{P}}(\max(n^{-1/2}, \rho_T^{-1/2}))$. As regards $\hat{\mathbf{\Gamma}}_{jk}^{\psi}$, using (B.1),

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

The conclusion follows.

D Proof of Proposition 5

Consider the static model $\mathbf{z}_{nt} = \mathcal{R}\mathbf{v}_t + \boldsymbol{\phi}_{nt}$. If $\mathbf{z}_{nt} = \mathbf{A}(L)\mathbf{x}_{nt}$ were observed, i.e. if the matrices $\mathbf{A}(L)$ were known, then Proposition 5, with an estimator of $\mathcal{R}$ based on the empirical covariance $\mathbf{\Gamma}^z$ of the $\mathbf{z}_{nt}$'s, 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 13, which establishes the asymptotic behavior of the difference $\mathbf{\Gamma}^z - \hat{\mathbf{\Gamma}}^z$; Lemmas 7 through 12 are but a preparation for that crucial result. All lemmas in this Appendix hold, and are proved under Assumptions 1 through 10.

Lemma 7 *As $T \to \infty$ and $n \to \infty$,*

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

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

Proof. The first part of (i) already has been taken care of in the proof of Lemma 6. Lemma 6 and Assumption A.5 jointly imply that $\hat{\sigma}^{x}_{ii} = \sum_{f=1}^{q} \hat{\lambda}^{x}_{f} |\hat{p}^{x}_{if}|^2 = O_{\mathrm{P}}(1)$, uniformly in θ. As all the terms in the sum are positive, $\hat{\lambda}^{x}_{f} |\hat{p}^{x}_{if}|^2 = (\hat{\lambda}^{x}_{f}/n) n |\hat{p}^{x}_{if}|^2$ is $O_{\mathrm{P}}(1)$ as well, uniformly in θ. Lemma 4 and Assumption A.7 imply that $\hat{\lambda}^{x}_{f}/n$ is $O_{\mathrm{P}}(1)$ and bounded away from zero in probability uniformly in θ. The conclusion follows.

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

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

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

Lemma 8 *As $T \to \infty$ and $n \to \infty$,*

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

Proof. The left-hand side of (D.1) equals the left-hand side of (B.4) when $\mathcal{S}_{\mathbf{M}}$ is replaced by $\mathbf{I}_n$. The proof goes along the same lines as that of Lemma 6. Firstly, $\|n^{1/2}\mathbf{P}^{\chi}\|$ is $O(n^{1/2})$. Both $\|\mathbf{g} - \mathbf{h}\|$ and $\|\mathbf{h} - \mathbf{i}\|$ are $O_{\mathrm{P}}(n^{-1/2} \max(n^{-1}, \rho_T^{-1/2}))$. As for $\|\mathbf{i} - \mathbf{j}\|$, the conclusion follows from Lemma 3 (iii). Q.E.D.

Lemma 9 *As $T \to \infty$ and $n \to \infty$,*

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

(ii) $n^{-1} \sum_{i=1}^{n} |p^{\chi}_{if} - \hat{p}^{x}_{if}| = O_{\mathrm{P}}(n^{-1/2} \max(n^{-1}, \rho_T^{-1/2}))$, uniformly in θ.

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

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

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

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

by Lemma 4(i), Assumption A.7 and Lemma 9(i) above. Thus, $(\hat{\lambda}^{x}_{f})^{1/2}\left(p^{\chi}_{if} - \hat{p}^{x}_{if}\right)$ is $O_{\mathrm{P}}(\max(n^{-1/2}, \rho_T^{1/2}))$. By Assumption A.7, $n^{-1/2}(\hat{\lambda}^{x}_{f})^{1/2}$ is bounded away from zero. The conclusion follows.

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

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

Regarding the first term on the right-hand side, by Jensen's inequality and Lemma 8:

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

Lemma 4(i)-(ii) and Lemma 7(ii) yield bounds for the second term. Q.E.D.

Lemma 10 *For any integer $d \in \mathbb{N}$, as $T \to \infty$ and $n \to \infty$,*

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

uniformly in θ.

PROOF. By induction. Lemma 9(ii) implies that $n^{-1}\sum_{i=1}^{n} | \hat{p}^{\chi}_{if} - p^{\chi}_{ij} |$ is $O_{\mathrm{P}}((n^{-1}\max(n^{-1}, \rho_T^{-1}))^{1/2})$. In fact, to avoid unnecessary complications, we consider here a slightly looser bound than the one provided by Lemma 9. Assume now that $d \geq 2$ and that the result holds for $d - 1$. Using summation by parts,

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

since without loss of generality we can assume $| p^{\chi}_{i+1,f} - \hat{p}^{x}_{i+1,f} | \geq | p^{\chi}_{if} - \hat{p}^{x}_{if} |$. The result follows from Lemma 9(i). Q.E.D.

Lemma 11 *For $n \to \infty$ and $T \to \infty$, uniformly in θ,*

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

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

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

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 4, 7 and 10,

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

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

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

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

Q.E.D.

Lemma 12 *For $n \to \infty$ and $T \to \infty$,*

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

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

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

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

The first term on the right-hand side is bounded using Lemma 11. Because ℓ takes only a finite number of values, the second term is $O(B_T^{-d})$ (see the proof of Proposition 4). Because the functions σ_{ij} are of bounded variation uniformly in i and j, see Assumption A.10, the third term is $O(B_T^{-d})$. The same argument used to obtain Proposition 3 applies. The second statement is proved in the same way. Q.E.D.

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

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

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

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

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

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

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

for $s > 0$ and $\mathbf{A}_0 = \mathbf{I}_n$, we obtain

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

which is a sum of $(S+1)^2$ terms. Inspection of the right-hand side of (D.6) shows that (D.4) holds, with $\|\hat{\boldsymbol{\Gamma}}^z - \boldsymbol{\Gamma}^z\|$ replaced with $\|\check{\boldsymbol{\Gamma}}^z - \boldsymbol{\Gamma}^z\|$, under assumptions A.4, A.5, A.9, and A.10.

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

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

where $\hat{\boldsymbol{\Gamma}}^x_{jk,s-r}$ is the (j,k)-block of $\hat{\boldsymbol{\Gamma}}^x_{s-r} = T^{-1}\sum_{t=1}^{T}\mathbf{x}_{t-r}\mathbf{x}'_{t-s}$, and the second inequality follows from applying the C_r inequality to each term of the form

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

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 the Cauchy-Schwartz, C_r and Jensen inequalities, we obtain

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

where

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

First consider $\mathcal{A}$. Letting $\mathbf{a}^{j\prime}_{s\alpha} = (a^j_{s\alpha,1}\ a^j_{s\alpha,2}\ \cdots\ a^j_{s\alpha,q+1})$, note that $\mathbf{a}^j_{s\alpha,\delta} = \mathbf{e}'_\alpha\mathbf{A}^{[j]}\mathbf{g}_{s\delta}$, where $\mathbf{e}_\alpha$ and $\mathbf{g}_{s\delta}$ are the α-th and $(s-1)(q+1)+\delta$-th unit vectors in the $(q+1)$- and

$(q+1)S$-dimensional canonical bases, respectively. Writing, for the sake of simplicity, $\mathbf{B}_j$ and $\mathbf{C}_j$ instead of $\mathbf{B}_j^{\chi}$ and $\mathbf{C}_{jj}^{\chi}$, as defined in (2.18) and (2.19), we obtain, from (B.1), the C_r, the triangular and the Cauchy-Schwartz inequalities,

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

Denoting by b_{il}^j the entries of $\mathbf{B}_j$, $i=1,\dots,q+1$, $l=1,\dots,S(q+1)$, the C_r inequality and Lemma 12 entail

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

In a similar way, one can prove that $\sum_{j=1}^m\|\hat{\mathbf{C}}_j-\mathbf{C}_j\|^8$ is $O_{\mathrm{P}}(m(\max(n^{-1},\rho_T^{-1}))^4)$. Moreover, Assumptions A.4 and A.5 imply that $\sum_{j=1}^m\|\hat{\mathbf{B}}_j\|^{16}$ and $\sum_{j=1}^m\|\mathbf{C}_j^{-1}\|^{16}$, as well as $\sum_{j=1}^m\|\hat{\mathbf{C}}_j^{-1}\|^{8}$ and $\sum_{j=1}^m\|\hat{\mathbf{C}}_j^{-1}\|^{16}$, are $O_P(m)$.

Collecting terms:

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

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

$$\mathcal{C} \leq ((q+1)/m)^{1/2}\Big\{\Big[\sum_{k=1}^{m}\Big(\sum_{\beta=1}^{q+1}(\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\mathbf{a}}^{k}_{r\beta})^2\Big)^2\Big]^{1/2}\Big[\sum_{j=1}^{m}\Big(\sum_{k=1}^{m}\Big(\text{trace}[\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\hat{\mathbf{\Gamma}}^{x}_{jk,s-r}]\Big)^4\Big)^{1/2}\Big]\Big\}^{1/2}$$

$$\leq ((q+1)/m)^{1/2}\Big\{\Big[(q+1)\sum_{k=1}^{m}\sum_{\beta=1}^{q+1}(\hat{\mathbf{a}}^{k\prime}_{r\beta}\hat{\mathbf{a}}^{k}_{r\beta})^4\Big]^{1/2}\Big[\sum_{j=1}^{m}\Big(\sum_{k=1}^{m}\Big(\text{trace}[\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\hat{\mathbf{\Gamma}}^{x}_{jk,s-r}]\Big)^4\Big)^{1/2}\Big]\Big\}^{1/2}$$

$$\leq (q+1)^{1/2}\Big[(q+1)^4\sum_{k=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}(\hat{a}^{k}_{r,\alpha\beta})^8\Big]^{1/4}\Big[m^{-1}\sum_{j=1}^{m}\sum_{k=1}^{m}\Big(\text{trace}[\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\hat{\mathbf{\Gamma}}^{x}_{jk,s-r}]\Big)^4\Big]^{1/4}$$

$$\leq (q+1)^{3/2}\Big[\sum_{k=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}(\hat{a}^{k}_{r,\alpha\beta})^8\Big]^{1/4}\Big[((q+1)^6/m)\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}(\hat{\gamma}^{x}_{jk,\alpha\beta}(s-r))^8\Big]^{\frac{1}{4}}$$

$$= O_{\mathrm{P}}(m^{1/2}),$$

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

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

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. Without loss of generality, we can assume that the number s of elements selected by $\mathcal{S}_{\mathbf{M}}$ is of the form $s=s^*(q+1)$ for some integer s^*. The two terms on the right-hand side of (D.7) must be dealt with separately, since there is only one summation ranging from 1 to n. In the first of those two terms, substituting the summation $\sum_{k=1}^{s^*}$ for $\sum_{k=1}^{m}$ gives

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

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

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

and

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

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

As for the second term, using arguments similar to those used in the first part of the proof, we obtain

$$\sum_{k=1}^{s^*}\sum_{j=1}^{m}\sum_{\alpha=1}^{q+1}\sum_{\beta=1}^{q+1}\Big((\hat{\mathbf{a}}^{k}_{s\alpha}-\mathbf{a}^{k}_{s\alpha})'\hat{\mathbf{\Gamma}}^{x\prime}_{jk,s-r}\mathbf{a}^{j}_{r\beta}\Big)^2$$

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

$$= \mathcal{FG}, \text{ say.}$$

It easily follows from Proposition 4 that $\mathcal{F} = O_{\mathrm{P}}(m\,\zeta_{TN}^2)$, while $\mathcal{G} = O_{\mathrm{P}}(1)$ can be obtained using the arguments used to bound $\mathcal{C}$ in the proof of the first statement. Collecting terms, we obtain, as desired,

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

Starting with (D.4), which plays here the same role as (3.5) does for the proof of Proposition 2, we can easily prove statements that replicate in this context Lemmas 3, 4, 5 and 6, using the same arguments used in Section B, with x, χ and ξ replaced by z, ψ and ϕ respectively. Precisely:

(I) In the results corresponding to Lemma 3 we obtain the rate ζ_{Tn} 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 3. For, (iii) has $O_{\mathrm{P}}(\zeta_{Tn}) + O(1/n) = O_{\mathrm{P}}(\zeta_{Tn})$, while (iv) has $O_{\mathrm{P}}(\zeta_{Tn}) + O(1/\sqrt{n}) = O_{\mathrm{P}}(\zeta_{Tn})$.

(II) The same rate ζ_{Tn} is obtained for the results of Lemma 4.

(III) The same holds for Lemma 5. 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 6 becomes

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

Going over the proof of Lemma 6, we see that $\|c-d\|$ has the worst rate, whereas here $\|a-b\|$, $\|b-c\|$ and $\|c-d\|$ all have rate $O_{\mathrm{P}}(\zeta_{Tn})$.

(V) Moreover, in the same way as the proof of Lemma 6 can be replicated to obtain (D.9), the proof of Lemma 8, see below, can be replicated to obtain:

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

E Proof of Proposition 6

Let

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

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

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

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

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

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

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

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

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

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

boundedness of $\lambda_1(\hat{\mathbf{A}}_r'\hat{\mathbf{A}}_r)$ being a consequence of Assumptions A.4 and A.5. As for the third term on the right hand side of (E.11), $(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\mathbf{A}(L)\xi_t$ is $O_{\mathrm{P}}(n^{-1/2})$. To conclude, note that $\hat{\mathbf{W}}^z(\mathbf{\Lambda}^\psi)^{-1/2}\mathbf{P}^{\psi\prime}\mathbf{P}^\psi(\mathbf{\Lambda}^\psi)^{1/2}\mathbf{v}_t = \hat{\mathbf{W}}^z\mathbf{v}_t$. Q.E.D.

F Data description

Quarterly data. Most series are taken from the FRED data base. A few stock market and leading indicators are taken from Data Stream. Some series have been constructed as transformations of the original FRED series. Monthly data have been temporally aggregated to get quarterly figures. Outliers are treated as in Stock and Watson (2002b). Transformations: 1 = levels, 2 = first differences of the original series, 5 = first differences of logs of the original series, 6 = second differences of logs of the original series.

no.series	Transf.	Mnemonic	Long Label
1	5	GDPC1	Real Gross Domestic Product, 1 Decimal
2	5	GNPC96	Real Gross National Product
3	5	NICUR/GDPDEF	National Income/GDPDEF
4	5	DPIC96	Real Disposable Personal Income
5	5	OUTNFB	Nonfarm Business Sector: Output
6	5	FINSLC1	Real Final Sales of Domestic Product, 1 Decimal
7	5	FPIC1	Real Private Fixed Investment, 1 Decimal
8	5	PRFIC1	Real Private Residential Fixed Investment, 1 Decimal
9	5	PNFIC1	Real Private Nonresidential Fixed Investment, 1 Decimal
10	5	GPDIC1	Real Gross Private Domestic Investment, 1 Decimal
11	5	PCECC96	Real Personal Consumption Expenditures
12	5	PCNDGC96	Real Personal Consumption Expenditures: Nondurable Goods
13	5	PCDGCC96	Real Personal Consumption Expenditures: Durable Goods
14	5	PCESVC96	Real Personal Consumption Expenditures: Services
15	5	GPSAVE/GDPDEF	Gross Private Saving/GDP Deflator
16	5	FGCEC1	Real Federal Consumption Expenditures & Gross Investment, 1 Decimal
17	5	FGEXPND/GDPDEF	Federal Government: Current Expenditures/ GDP deflator
18	5	FGRECPT/GDPDEF	Federal Government Current Receipts/ GDP deflator
19	2	FGDEF	Federal Real Expend-Real Receipts
20	1	CBIC1	Real Change in Private Inventories, 1 Decimal
21	5	EXPGSC1	Real Exports of Goods & Services, 1 Decimal
22	5	IMPGSC1	Real Imports of Goods & Services, 1 Decimal
23	5	CP/GDPDEF	Corporate Profits After Tax/GDP deflator
24	5	NFCPATAX/GDPDEF	Nonfinancial Corporate Business: Profits After Tax/GDP deflator
25	5	CNCF/GDPDEF	Corporate Net Cash Flow/GDP deflator
26	5	DIVIDEND/GDPDEF	Net Corporate Dividends/GDP deflator
27	5	HOANBS	Nonfarm Business Sector: Hours of All Persons
28	5	OPHNFB	Nonfarm Business Sector: Output Per Hour of All Persons
29	5	UNLPNBS	Nonfarm Business Sector: Unit Nonlabor Payments
30	5	ULCNFB	Nonfarm Business Sector: Unit Labor Cost
31	5	WASCUR/CPI	Compensation of Employees: Wages & Salary Accruals/CPI
32	6	COMPNFB	Nonfarm Business Sector: Compensation Per Hour
33	5	COMPRNFB	Nonfarm Business Sector: Real Compensation Per Hour
34	6	GDPCTPI	Gross Domestic Product: Chain-type Price Index
35	6	GNPCTPI	Gross National Product: Chain-type Price Index
36	6	GDPDEF	Gross Domestic Product: Implicit Price Deflator
37	6	GNPDEF	Gross National Product: Implicit Price Deflator
38	5	INDPRO	Industrial Production Index
39	5	IPBUSEQ	Industrial Production: Business Equipment
40	5	IPCONGD	Industrial Production: Consumer Goods
41	5	IPDCONGD	Industrial Production: Durable Consumer Goods
42	5	IPFINAL	Industrial Production: Final Products (Market Group)
43	5	IPMAT	Industrial Production: Materials
44	5	IPNCONGD	Industrial Production: Nondurable Consumer Goods
45	2	AWHMAN	Average Weekly Hours: Manufacturing
46	2	AWOTMAN	Average Weekly Hours: Overtime: Manufacturing
47	2	CIVPART	Civilian Participation Rate
48	5	CLF16OV	Civilian Labor Force
49	5	CE16OV	Civilian Employment
50	5	USPRIV	All Employees: Total Private Industries
51	5	USGOOD	All Employees: Goods-Producing Industries
52	5	SRVPRD	All Employees: Service-Providing Industries
53	5	UNEMPLOY	Unemployed
54	5	UEMPMEAN	Average (Mean) Duration of Unemployment
55	2	UNRATE	Civilian Unemployment Rate
56	5	HOUST	Housing Starts: Total: New Privately Owned Housing Units Started

57	2	FEDFUNDS	Effective Federal Funds Rate
58	2	TB3MS	3-Month Treasury Bill: Secondary Market Rate
59	2	GS1	1-Year Treasury Constant Maturity Rate
60	2	GS10	10-Year Treasury Constant Maturity Rate
61	2	AAA	Moody's Seasoned Aaa Corporate Bond Yield
62	2	BAA	Moody's Seasoned Baa Corporate Bond Yield
63	2	MPRIME	Bank Prime Loan Rate
64	6	BOGNONBR	Non-Borrowed Reserves of Depository Institutions
65	6	TRARR	Board of Governors Total Reserves, Adjusted for Changes in Reserve
66	6	BOGAMBSL	Board of Governors Monetary Base, Adjusted for Changes in Reserve
67	6	M1SL	M1 Money Stock
68	6	M2MSL	M2 Minus
69	6	M2SL	M2 Money Stock
70	6	BUSLOANS	Commercial and Industrial Loans at All Commercial Banks
71	6	CONSUMER	Consumer (Individual) Loans at All Commercial Banks
72	6	LOANINV	Total Loans and Investments at All Commercial Banks
73	6	REALLN	Real Estate Loans at All Commercial Banks
74	6	TOTALSL	Total Consumer Credit Outstanding
75	6	CPIAUCSL	Consumer Price Index For All Urban Consumers: All Items
76	6	CPIULFSL	Consumer Price Index for All Urban Consumers: All Items Less Food
77	6	CPILEGSL	Consumer Price Index for All Urban Consumers: All Items Less Energy
78	6	CPILFESL	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy
79	6	CPIENGSL	Consumer Price Index for All Urban Consumers: Energy
80	6	CPIUFDSL	Consumer Price Index for All Urban Consumers: Food
81	6	PPICPE	Producer Price Index Finished Goods: Capital Equipment
82	6	PPICRM	Producer Price Index: Crude Materials for Further Processing
83	6	PPIFCG	Producer Price Index: Finished Consumer Goods
84	6	PPIFGS	Producer Price Index: Finished Goods
85	6	OILPRICE	Spot Oil Price: West Texas Intermediate
86	5	USSHRPRCF	US Dow Jones Industrials Share Price Index (EP) NADJ
87	5	US500STK	US Standard & Poor's Index of 500 Common Stocks
88	5	USI62...F	US Share Price Index NADJ
89	5	USNOIDN.D	US Manufacturers New Orders for Non Defense Capital Goods (BCI 27)
90	5	USCNORCGD	US New Orders of Consumer Goods & Materials (BCI 8) CONA
91	1	USNAPMNO	US ISM Manufacturers Survey: New Orders Index SADJ
92	5	USVACTOTO	US Index of Help Wanted Advertising VOLA
93	5	USCYLEAD	US The Conference Board Leading Economic Indicators Index SADJ
94	5	USECRIWLH	US Economic Cycle Research Institute Weekly Leading Index
95	2	GS10-FEDFUNDS	
96	2	GS1-FEDFUNDS	
97	2	BAA-FEDFUNDS	
98	5	GEXPND/GDPDEF	Government Current Expenditures/ GDP deflator
99	5	GRECPT/GDPDEF	Government Current Receipts/ GDP deflator
100	2	GDEF	Governnent Real Expend-Real Receipts
101	5	GCEC1	Real Government Consumption Expenditures & Gross Investment, 1 Decimal

Monthly data. Most series are those of the Stock-Watson data set used in Bernanke *et al.* (2005). A few real exchange rates and short-term interest rate spreads between US and some foreign countries are added, and some discontinued series are eliminated. The basic source is the FRED data base; some series have been constructed as transformations of the original series. Outliers are treated as in Stock and Watson (2002b). Transformations: 1 = levels, 4 = logs, 5 = first differences of logs of the original series.

no.series	Mnemonic	Long Label	Transformation
1	DSPIC96	Real Disposable Personal Income	5
2	A0M051	Personal Income Less Transfer Payments	5
3	PCEC96	Real Personal Consumption Expenditures	5
4	A0M059	Sales, Orders, And Deliveries, Sales, Retail Stores	5
5	IPS10	Industrial Production Index - Total Index	5
6	IPS11	Industrial Production Index - Products, Total	5
7	IPS12	Industrial Production Index - Consumer Goods	5
8	IPS13	Industrial Production Index - Durable Consumer Goods	5
9	IPS18	Industrial Production Index - Nondurable Consumer Goods	5
10	IPS25	Industrial Production Index - Business Equipment	5
11	IPS299	Industrial Production Index - Final Products	5
12	IPS306	Industrial Production Index - Fuels	5
13	IPS307	Industrial Production Index - Residential Utilities	5
14	IPS32	Industrial Production Index - Materials	5
15	IPS34	Industrial Production Index - Durable Goods Materials	5
16	IPS38	Industrial Production Index - Nondurable Goods Materials	5
17	IPS43	Industrial Production Index - Manufacturing (SIC)	5
18	PMP	NAPM Production Index (Percent)	1
19	MCUMFN	Capacity Utilization: Manufacturing (NAICS)	1
20	LHEL	Index Of Help-Wanted Advertising In Newspapers	5
21	LHELX	Employment: Ratio; Help-Wanted	4
22	LHEM	Civilian Labor Force: Employed, Total	5
23	LHNAG	Civilian Labor Force: Employed, Nonagric.Industries	5
24	LHU14	Unemploy.By Duration: Persons Unempl.5 To 14 Wks	1
25	LHU15	Unemploy.By Duration: Persons Unempl.15 Wks +	1
26	LHU26	Unemploy.By Duration: Persons Unempl.15 To 26 Wks	1
27	LHU27	Unemploy.By Duration: Persons Unempl.27 Wks +	1
28	LHU5	Unemploy.By Duration: Persons Unempl.Less Than 5 Wks	1
29	LHU680	Unemploy.By Duration: Average(Mean)Duration In Weeks	1
30	LHUR	Unemployment Rate: All Workers, 16 Years & Over (%,SA)	1
31	CES002	Employees On Nonfarm Payrolls - Total Private	5
32	CES003	Employees On Nonfarm Payrolls - Goods-Producing	5
33	CES006	Employees On Nonfarm Payrolls - Mining, Thousands	5
34	CES011	Employees On Nonfarm Payrolls - Construction	5
35	CES015	Employees On Nonfarm Payrolls - Manufacturing	5
36	CES017	Employees On Nonfarm Payrolls - Durable Goods	5
37	CES033	Employees On Nonfarm Payrolls - Nondurable Goods	5
38	CES046	Employees On Nonfarm Payrolls - Service-Providing	5
39	CES048	Employees On Nonfarm Payrolls - Trade, Transp., Utilities	5
40	CES049	Employees On Nonfarm Payrolls - Wholesale Trade	5
41	CES053	Employees On Nonfarm Payrolls - Retail Trade	5
42	CES088	Employees On Nonfarm Payrolls - Financial Activities	5
43	CES140	Employees On Nonfarm Payrolls - Government	5
44	AWHI	Aggregate Weekly Hours Index: Total Private Industries	5
45	CES151	Average Weekly Hours Goods-Producing	1
46	CES155	Average Weekly Hours Manufacturing Overtime Hours	1
47	AWHMAN	Average Weekly Hours: Manufacturing	1
48	PMEMP	Napm Employment Index (Percent)	1
49	HSBMW	Houses Authorized By Build. Permits:Midwest	4
50	HSBNE	Houses Authorized By Build. Permits:Northeast	4
51	HSBR	Housing Authorized: Total New Priv Housing Units	4
52	HSBSOU	Houses Authorized By Build. Permits:South	4
53	HSBWST	Houses Authorized By Build. Permits:West	4
54	HSFR	Housing Starts:Nonfarm (1947-58);Total Farm&Nonfarm(1959-)	4
55	HSMW	Housing Starts:Midwest	4

no.series	Mnemonic	Long Label	Transformation
56	HSNE	Housing Starts:Northeast	4
57	HSSOU	Housing Starts:South	4
58	HSWST	Housing Starts:West	4
59	PMDEL	Napm Vendor Deliveries Index	1
60	PMI	Purchasing Managers' Index	1
61	PMNO	Napm New Orders Index	1
62	PMNV	Napm Inventories Index	1
63	A0M007	New Orders, Durable Goods Industries	5
64	A0M027	New Orders, Capital Goods Industries, Nondefense	5
65	A1M092	Manufacturers' Unfilled Orders, Durable Goods Industries	5
66	FM1	Money Stock: M1	5
67	FM2	Money Stock:M2	5
68	FMFBA	Monetary Base, Adj For Reserve Requirement Changes	5
69	FMRNBA	Depository Inst Reserves:Nonborrowed,Adj Res Req Chgs	5
70	FMRRA	Depository Inst Reserves:Total,Adj For Reserve Req Chgs	5
71	FCLBMC	Wkly Rp Lg Com'L Banks:Net Change Com'L & Indus Loans	1
72	CCINRV	Consumer Credit Outstanding - Nonrevolving(G19)	5
73	FSPCOM	S&P'S Common Stock Price Index: Composite	5
74	FSPIN	S&P'S Common Stock Price Index: Industrials	5
75	FYFF	Interest Rate: Federal Funds (Effective)	1
76	FYGM3	Interest Rate: U.S.Treasury Bills,Sec Mkt,3-Mo.	1
77	FYGM6	Interest Rate: U.S.Treasury Bills,Sec Mkt,6-Mo.0	1
78	FYGT1	Interest Rate: U.S.Treasury Const Maturities,1-Yr.	1
79	FYGT10	Interest Rate: U.S.Treasury Const Maturities,10-Yr.	1
80	FYGT5	Interest Rate: U.S.Treasury Const Maturities,5-Yr.	1
81	FYAAAC	Bond Yield: Moody'S Aaa Corporate	1
82	FYBAAC	Bond Yield: Moody'S Baa Corporate	1
83	EXRUS	United States;Effective Exchange Rate (MERM)	5
84	EXRCAN	Foreign Exchange Rate: Canada (Canadian $ Per U.S.$)	5
85	EXRJAN	Foreign Exchange Rate: Japan (Yen Per U.S.$)	5
86	EXRSW	Foreign Exchange Rate: Switzerland (Swiss Franc Per U.S.$)	5
87	EXRUK	Foreign Exchange Rate: United Kingdom (Cents Per Pound)	5
88	PWFCSA	Producer Price Index:Finished Consumer Goods	5
89	PWFSA	Producer Price Index: Finished Goods	5
90	PWCMSA	Producer Price Index:Crude Materials	5
91	PWIMSA	Producer Price Index:Intermed Mat.Supplies & Components	5
92	PMCP	Napm Commodity Prices Index	1
93	PU83	CPI-U: Apparel & Upkeep	5
94	PU84	CPI-U: Transportation	5
95	PU85	CPI-U: Medical Care	5
96	PUNEW	CPI-U: All Items	5
97	PUC	CPI-U: Commodities	5
98	PUCD	CPI-U: Durables	5
99	PUS	CPI-U: Services	5
100	PUXF	CPI-U: All Items Less Food	5
101	PUXHS	CPI-U: All Items Less Shelter	5
102	PUXM	CPI-U: All Items Less Medical Care	5
103	CES277	Average Hourly Earnings - Construction	5
104	CES278	Average Hourly Earnings - Manufacturing	5
105	CES275	Average Hourly Earnings Goods-Producing	5
106		Real Foreign Exchange Rate: Swiss	4
107		Real Foreign Exchange Rate: Japan	4
108		Real Foreign Exchange Rate: Uk	4
109		Real Foreign Exchange Rate: Canada	4
110		Us - Canada Interest Rates Spread	1
111		Us - Japan Interest Rates Spread	1
112		Us - Uk Interest Rates Spread	1