Dynamic Factor Models with Infinite-Dimensional Factor Space: One-Sided Representations

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

December 4, 2012

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). That paper, however, relies on Brillinger's *dynamic principal components*. The corresponding estimators are two-sided filters whose performance at the end of the observation period or for forecasting purposes is rather poor. No such

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

†Académie Royale de Belgique and CentER, Tilburg University.

‡Corresponding author. Dipartimento di Scienze Statistiche, Piazzale Aldo Moro 5, 00185 Roma, Italy. Email: ml@lippi.ws

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

problem arises with estimators based on standard principal components, which have been dominant in this literature. On the other hand, those estimators require the assumption that the space spanned by the factors has finite dimension. In the present paper, we argue that such an assumption is extremely restrictive and potentially quite harmful. Elaborating upon recent results by Anderson and Deistler (2008a, b) on singular stationary processes with rational spectrum, we obtain one-sided representations for the GDFM without assuming finite dimension of the factor space. Construction of the corresponding estimators is also briefly outlined. In a companion paper, we establish consistency and rates for such estimators, and provide Monte Carlo results further motivating our approach.

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.

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 recent and 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

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

$i \in \mathbb{N}$, $t \in \mathbb{Z}$, 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). Moreover:

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

(II) 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 (II) to hold (we refer to Section 2 for a detailed presentation and discussion).

Weak covariance of the ξ_{it}'s motivates calling them *idiosyncratic*, while the χ_{it}'s, being driven by the low-dimensional vector of *common shocks* u_{ft}, $f = 1, 2 \ldots, q$, are called *common* components. The model implies that cross-covariances among the observable variables x_{it} are essentially accounted for by the common components χ_{it}.

[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.

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 finite realizations $(i = 1, \dots, n;\ t = 1, \dots, T)$ of the process $\{x_{it}\}$, as n and T both tend to infinity. 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 $\overline{\text{span}}(\ \dots\ )$ the space generated by a collection of random variables,[2] $\overline{\text{span}}(\chi_{it},\ i \in \mathbb{N})$, for given t, has finite dimension r, where $r \geq q$. Under that assumption, model (1.1) can be rewritten as

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

$i \in \mathbb{N}$, $t \in \mathbb{Z}$.[3] In this case, we say that (1.1) admits a *static representation.* If, in addition, $\mathbf{N}(L) = \mathbf{N}(0)$, so that $\mathbf{F}_t$ is a white noise vector, then (1.1) is a *static factor model.* 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

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

where the matrices $\mathbf{D}_j$ are $r \times r$ while $\mathbf{R}$ 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.

[2]More precisely, $\overline{\text{span}}(\zeta_i,\ i \in \mathbb{N})$, 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.

[3]This is fairly easy to prove under the assumptions in Section 2.2 below, see Forni *et al.* (2009).

(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 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. Even the paper by Forni *et al.* (2005), which is based on the same frequency-domain approach as Forni *et al.* (2000), adopts a finite-dimension assumption for $\overline{\text{span}}(\chi_{it},\ i \in \mathbb{N})$ to obtain one-sided estimators.[4]

The moot point is that such assumptions are far from being innocuous. For instance, (1.2) is so restrictive that even the very elementary model

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

where $q = 1$, u_t is scalar white noise, and the coefficients α_i are drawn from a uniform distribution, is ruled out. 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. Infinite dimension of $\overline{\text{span}}(\chi_{it},\ i \in \mathbb{N})$ occur *a fortiori* if the AR common component in (1.4) is replaced by more general ARMA ones.

An analysis of (1.4) or more general versions thereof, based on (1.2) or (1.3) then can be extremely misleading. In particular, criteria determining r, see (i) above, would provide meaningless results as n and T tend to infinity. This, we believe, is a strong theoretical motivation for solving the one-sidedness problem in the GDFM without the finite-dimension assumption.

[4]See also Altissimo *et al.* (2010), where the spectral-density principal-component approach is used in combination with the finite-dimensional assumption.

In the present paper we solve the problem under assumptions that include rational spctral density for the common components χ_{it} but no finite-dimension restriction.

To avoid possible misunderstandings, let us emphasize that our construction here only deals with representation issues and makes use of population covariances and spectral densities. However, as a brief outline at the end of the paper shows, obtaining the corresponding estimators is straightforward. The companion paper by Forni, Hallin, Lippi and Zaffaroni (2013) gives a detailed definition of such estimators, studies their consistency and rates, and provides Monte Carlo evidence further motivating our approach.[5]

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. More precisely, we assume the following representation for the common components:

$$\chi_{it} = \frac{c_{i1}(L)}{d_{i1}(L)}u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)}u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)}u_{qt}, \tag{1.5}$$

where

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

$f = 1, 2, \ldots, q$. The assumption that s_1 and s_2, the degrees of $c_{if}(L)$ and $d_{if}(L)$ respectively, are assumed to be independent of i is very convenient, though not necessary. As for the idiosyncratic components we do not make any parametric assumptions, nor do we restrict their cross-covariance 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

[5]Some of the results of this paper have been outlined, without proofs, in a very preliminary version in Forni and Lippi (2011).

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 each $\mathbf{A}^k(L)$ 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

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

where $\mathbf{Z}_t = \underline{\mathbf{A}}(L)\mathbf{x}_t$, and, lastly,

$$\mathbf{z}_t = \underline{\mathbf{r}}\mathbf{u}_t + \boldsymbol{\phi}_t, \tag{1.8}$$

which results from (1.7) by normalization (both sides of the i-th equation are divided by the standard deviation of Z_{it}). This is a factor model with a representation of the form (1.2) and $\mathbf{F}_t = \mathbf{u}_t$, thus, according to the definition given in Section 1.1, a static factor model.

Some comments on (1.6)-(1.8) are in order:

(i) We can rewrite (1.6) as $\mathbf{A}^k(L)\boldsymbol{\chi}_t^k = \mathbf{R}^k\mathbf{u}_t$, $k \in \mathbb{N}$, where the vectors $\boldsymbol{\chi}_t^k$ are the $(q+1)$-dimensional subvectors

$$(\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{q+1,t}),\ (\chi_{q+2,t}\ \chi_{q+3,t}\ \cdots\ \chi_{2(q+1),t}),\ \ldots$$

Thus (1.6) is made up of (a) obtaining an autoregressive representation for each of the vectors $\boldsymbol{\chi}_t^k$, and then (b) knitting together such autoregressive representations.

(ii) As regards (a), each of the subvectors has dimension $(q+1)$ and rank q (i.e. its spectral density has rank q for all $\theta \in [-\pi\ \pi]$), and is therefore singular (i.e. its dimension is greater than its rank). For singular (or reduced-rank) vectors, with rational spectral density, existence of a *finite-degree* autoregressive representation, for generic values of the parameters, has been proved in Anderson and Deistler (2008a, b). We contribute to this literature showing that when the dimension is equal to $q+1$ the minimum-lag autoregressive representation is generically unique. As regards (b), obtaining the same $\mathbf{u}_t$ for all the vectors $\boldsymbol{\chi}_t^k$ requires the additional assumption that, for each k, $\overline{\text{span}}(\boldsymbol{\chi}_{t-h}^k,\ h \geq 0) = \overline{\text{span}}(\boldsymbol{\chi}_{t-h},\ h \geq 0)$. We will motivate this restriction by a genericity argument.

(iii) The matrices $\mathbf{A}^k(L)$ and $\mathbf{R}^k$ can be obtained starting with the spectral density matrix of the observable variables x_{it}. The vector $\mathbf{z}_t$ results from the application of *one-sided filters* to the variables x_{it}, see (1.8). Lastly, $\mathbf{u}_t$ can be estimated using the first q principal components of the variables z_{it}, i.e. only *current* values of the variables z_{it}. *Our procedure thus solves the one-sidedness problem.*

(iv) Moreover, the matrices $\mathbf{A}^k(L)$ and $\mathbf{R}^k$, which are $(q+1)\times(q+1)$ and $(q+1)\times q$ respectively, result from *separate* calculations. Thus we do not incur in the difficulty known as "curse of dimensionality".

In Section 2, we state the main assumptions underlying the GDFM and review some basic results from previous literature. In Section 3, we prove some general results on stochastic vectors that are infinite-dimensional with finite rank, like $\boldsymbol{\chi}_t$, under the assumption of rational spectral density. Rational spectral density is assumed for $\boldsymbol{\chi}_t$ throughout the paper. In Section 4, we present results on autoregressive representations of singular stochastic vectors. Such results are then used to construct the blockwise autoregressive representation (1.6) for $\boldsymbol{\chi}_t$ and to transform the original variables x_{it} into another set of variables for which a static factor model holds. Lastly, we briefly outline the correspondence between our construction here and the steps of the estimation procedure that we study in the companion paper Forni *et al.* (2013). Section 5 concludes.

2 Main assumptions and background results

2.1 Notation

The GDFM (1.1) 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 counterpart. Associated with this infinite-dimensional vector notation, we also consider infinite-dimensional matrices, such as $\underline{\mathbf{A}}(L)$ or $\underline{\mathbf{R}}$ (see (1.8)), 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.

Given the infinite-dimensional process $\mathbf{y}_t = (y_{1t}\ y_{2t}\ \cdots\ y_{nt}\ \cdots)'$, we use the following notation:

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

(ii) $\mathcal{H}^y = \overline{\text{span}}(y_{it},\ i \in \mathbb{N},\ t \in \mathbb{Z})$, $\quad \mathcal{H}^{y_s} = \overline{\text{span}}(y_{it},\ 1 \le i \le s,\ t \in \mathbb{Z})$;

(iii) $\mathcal{H}^y_t = \overline{\text{span}}(y_{i\tau},\ i \in \mathbb{N},\ \tau \le t)$, $\quad \mathcal{H}^{y_s}_t = \overline{\text{span}}(y_{i\tau},\ 1 \le i \le s,\ \tau \le t)$.

If $\mathbf{y}_t$ is s-dimensional we use the notation $\mathcal{H}^y = \overline{\text{span}}(y_{it},\ i \le s,\ t \in \mathbb{Z})$, $\mathcal{H}^y_t = \overline{\text{span}}(y_{i\tau},\ i \le s, \tau \le t)$ (we never need sub-vectors of finite-dimensional vectors).

It is convenient, though not necessary, to assume throughout the paper that all white-noise vectors are orthonormal.

2.2 Basic assumptions

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

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

Assumption A.2 *There exists a positive integer q such that (i) $\bar{\lambda}_q^x(\theta)) = \infty$ for almost all θ in $[-\pi \ \pi]$, and (ii) $\bar{\lambda}_{q+1}^x(\theta)$ is essentially bounded, i.e. there exists a real B^x such that $\bar{\lambda}_{q+1}^x(\theta) \leq B^x$ almost everywhere in $[-\pi \ \pi]$.*

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, i.e. there exists a real B^ξ such that $\bar{\lambda}^\xi(\theta) \leq B^\xi$ almost everywhere in $[-\pi \ \pi]$;

(ii) $\boldsymbol{\chi}_t$ satisfies A.1 and $\bar{\lambda}_q^\chi(\theta) = \infty$ almost everywhere in θ in $[-\pi \ \pi]$ (note that $\bar{\lambda}_{q+s}^\chi(\theta) = 0$ a.e. in $[-\pi \ \pi]$ for all $s > 0$);

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

(iv) 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*. As mentioned in the Introduction, Forni *et al.* (2000) construct an estimator for the common component χ_{it}, under Assumptions A.1 and A.2, plus other technical assumptions, which is based on the first q eigenvectors of $\mathbf{\Sigma}_n^x(\theta)$ (the first q dynamic principal components of $\mathbf{x}_{nt}$).

Under the standard restriction that the dimension of $\overline{\text{span}}(\chi_{it},\ i \in \mathbb{N})$ is finite, the basic assumptions and theorem are:

Assumption B.1 *Same as A.1.*

Assumption B.2 *Let $\mathbf{\Gamma}_n^x$ be the variance-covariance matrix of $\mathbf{x}_{nt}$, μ_{nj}^x its j-th eigenvalue and $\bar{\mu}_j^x = \sup_{j\in\mathbb{N}} \mu_{nj}^x$. There exists a positive r such that (i) $\bar{\mu}_r^x = \infty$, and (ii) $\bar{\mu}_{r+1}^x < \infty$.*

Theorem B *(Chamberlain and Rothschild, 1983) Assumptions B.1 and B.2 imply that $\mathbf{x}_t$ can be represented as*

$$x_{it} = \chi_{it} + \xi_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \cdots + \lambda_{ir}F_{rt} + \xi_{it} \tag{2.2}$$

where $\mathbf{F}_t$ is a weakly stationary r-dimensional vector. Moreover,

(i) $\boldsymbol{\xi}_t$ satisfies Assumption B.1 and $\bar{\mu}_1^\xi < \infty$;

(ii) $\boldsymbol{\chi}_t$ satisfies Assumption B.1 and $\bar{\mu}_r^\chi = \infty$ (note that $\bar{\mu}_{r+s}^\chi = 0$ for all $s > 0$);

(iii) $\boldsymbol{\xi}_t$ and $\mathbf{F}_t$ are uncorrelated for all $t \in \mathbb{Z}$;

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

Conversely, if $\mathbf{x}_t$ can be represented as in (2.2) with $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ fulfilling (i), (ii) and (iii), then $\mathbf{x}_t$ satisfies Assumptions B.1 and B.2.

A detailed comparison between Assumptions and Theorems A and B is outside the scope of the present paper. Let us only observe that: (I) Assumption B.1 can be weakened, (II) Theorem B(iii) states that χ_{it} and ξ_{it} are orthogonal for all t in $\mathbb{Z}$, whereas Theorem A(iii) states that χ_{it} and ξ_{is} are orthogonal for all t and s in $\mathbb{Z}$.

Estimators of $\mathbf{F}_t$ and χ_{it} can be constructed, under Assumptions B.1 and B.2, plus other technical assumptions, by means of the first r standard principal components of $\mathbf{x}_{nt}$, see Stock and Watson (2002a,b), Bai and Ng (2002), Bai (2003).

In the next two sections we show how the variables x_{it}, which fulfill Assumptions A.1 and A.2, can, under further assumptions, be transformed into another set of variables,

z_{it}, which fulfill Assumptions B.1 and B.2 and therefore admit a representation of the form (2.2). Moreover, $\mathbf{F}_t$ is a white noise, so that the variables z_{it} evolve according to a static factor model.

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 $\boldsymbol{\chi}_t = [\underline{\mathbf{b}}(L)\mathbf{Q}][\mathbf{Q}'\mathbf{u}_t] = \underline{\mathbf{c}}(L)\mathbf{v}_t$, where $\mathbf{Q}$ is an arbitrary $q \times q$ orthogonal matrix, or, more generally, $\boldsymbol{\chi}_t = [\underline{\mathbf{b}}(L)\mathbf{Q}(L)][(\mathbf{Q}'(F)\mathbf{u}_t] = \underline{\mathbf{d}}(L)\mathbf{w}_t$, where $F = L^{-1}$ and $\mathbf{Q}(e^{-i\theta})\mathbf{Q}'(e^{i\theta}) = \mathbf{I}_q$ for almost all θ in $[-\pi\ \pi]$.

More importantly, Theorem A does not ensure that $\boldsymbol{\chi}_t$ admits a *one-sided moving-average representation*, i.e., a representation of the form $\boldsymbol{\chi}_t = \underline{\mathbf{e}}(L)\mathbf{w}_t$, where $\mathbf{w}_t$ is q-dimensional orthonormal white noise and $\underline{\mathbf{e}}(L) = \underline{\mathbf{e}}_0 + \underline{\mathbf{e}}_1 L + \cdots$. For example, if

$$\chi_{it} = u_{t+i-1}, \tag{3.1}$$

where u_t is one-dimensional white noise ($q = 1$), then statement (ii) of Theorem A holds true, 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 from Lemma 2).[6]

The existence of moving average representations, one-sided representations in particular, of infinite-dimensional stochastic vectors is analysed in Lemmas 1, 2 and 3 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 a positive integer* s *such that* $\operatorname{rank}(\boldsymbol{\Sigma}_n^y(\theta)) = q$, *for* $n \geq s$ *and almost all* θ *in* $[-\pi\ \pi]$.

[6] 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 in particular if their spectral density is singular in a positive-measure subset of $[-\pi\ \pi]$, see e.g. Pourahmadi (2001), Theorem 10.5, p. 361.

Lemma 1 *Suppose that the infinite-dimensional vector* $\mathbf{y}_t$ *fulfills A.1 and has rank* q*. Then* $\mathbf{y}_t$ *has a moving average representation,*

$$\mathbf{y}_t = \mathbf{b}(L)\mathbf{v}_t,$$

not necessarily one-sided, where $\mathbf{v}_t$ *is a* q*-dimensional orthonormal white noise and* $\mathbf{b}(L)$ *is an* $\infty \times q$ *matrix of square-summable filters.*

Proof. Let s be as in Definition 1. As a consequence, $y_{s+k,t} \in \mathcal{H}^{y_s}$. For, consider the projection

$$y_{s+k,t} = \text{proj}(y_{s+k,t} \,|\, \mathcal{H}^{y_s}) + \rho_{kt},$$

where $\rho_{kt} \perp \mathcal{H}^{y_s}$. If $\rho_{kt} \neq 0$ then the spectral density of $(y_{1t}\ y_{2t}\ \cdots\ y_{st}\ \rho_{kt})$ has rank $q+1$, against the assumption. Thus $\rho_{kt} = 0$ and $y_{s+k,t} \in \mathcal{H}^{y_s}$. Let us now recall that a stationary s-dimensional vector with a spectral density of rank q almost everywhere in $[-\pi\ \pi]$ has representations, not necessarily one-sided, as a moving average of a q dimensional orthonormal white noise (Rozanov, 1967, Chapter I, Section 9). As this is the case with $\mathbf{y}_{st}$, let

$$\mathbf{y}_{st} = \mathbf{b}_s(L)\mathbf{v}_t.$$

This implies that $\mathcal{H}^{y_s} \subseteq \mathcal{H}^v$ and therefore that $y_{s+k,t} \in \mathcal{H}^v$, so that

$$y_{s+k,t} = b_{s+k}(L)\mathbf{v}_t$$

and the lemma is proved. Q.E.D.

This result does not prove the existence of a one-sided representation for infinite-dimensional vector $\mathbf{y}_t$. A simple sufficient additional condition is given in Lemma 2. Some definitions and preliminary results are needed.

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

$$\mathbf{y}_t = \underline{\mathbf{b}}(L)\mathbf{v}_t, \tag{3.2}$$

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 (3.2) 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* $\mathcal{H}_t^v = \mathcal{H}_t^y$*.*

Now suppose that $\mathbf{y}_t$ is n-dimensional with representation

$$\mathbf{y}_t = \mathbf{b}(L)\mathbf{v}_t, \tag{3.3}$$

where $\mathbf{v}_t$ is q-dimensional orthonormal white noise and $\underline{\mathbf{b}}(L)$ is an $\infty \times q$ square summable filter. Fundamentalness of (3.3) and $\mathbf{v}_t$ are defined as in Definition 2. Moreover,

(I) If (3.3) is fundamental, then $n \geq q$. Moreover, if $\mathbf{y}_t = \mathbf{c}(L)\mathbf{w}_t$, where $\mathbf{w}_t$ is 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).

(II) If (3.3) 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 6). However,

(III) 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).

(II′) 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 2 *Suppose that the infinite-dimensional process* $\mathbf{y}_t$ *has rational spectral density and rank* q. *The following statements are equivalent:*

(i) $\mathbf{y}_t$ *has a one-sided rational moving average representation* $\mathbf{y}_t = \underline{\mathbf{b}}(L)\mathbf{v}_t$ *(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). By (III) there exists a one-sided rational fundamental representation for y_{st}, denote it by $\mathbf{y}_{st} = \mathbf{b}_s(L)\mathbf{v}_t$. We have $\mathcal{H}_t^{y_s} = \mathcal{H}_t^{v}$. By assumption $y_{s+k,t} \in \mathcal{H}_t^{y_s}$ and, therefore, $y_{s+k,t} \in \mathcal{H}_t^{v}$, so that

$$\mathbf{y}_{st} = \mathbf{b}_s(L)\mathbf{v}_t \qquad \text{and} \qquad y_{s+k,t} = b_{s+k}(L)\mathbf{v}_t. \tag{3.4}$$

The white noise $\mathbf{v}_t$ is fundamental for $\mathbf{y}_{st}$, hence also for $(\mathbf{y}'_{st}\ y_{s+k,t})'$. Thus representation (3.4) is fundamental, so that, by (III), $b_{s+k}(L)$ 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 determinants 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 a unitary $q \times q$ 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 (II′), $\mathbf{y}_{st} = \mathbf{d}_s(L)\mathbf{z}_t$ is fundamental for $\mathbf{y}_{st}$ and therefore for $\mathbf{y}_t$. Q.E.D.

Lemma 3 *Suppose that the infinite-dimensional process* $\mathbf{y}_t$ *has rational spectral density and rank* q. *Then,* *(i) If $\mathbf{y}_t$ has a one-sided rational representation $\mathbf{y}_t = \underline{\mathbf{b}}(L)\mathbf{v}_t$ then $\mathbf{y}_t$ has a fundamental (rational) 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 2. 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 2, 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 (II), 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 rational moving average representation, we obtain the existence of a rational fundamental representation for $\mathbf{y}_t$, which is unique up to multiplication by an orthogonal matrix. Moreover, for some s, the space spanned by the current and past values of $\mathbf{y}_{st}$ coincides with the space spanned by current and past values of the whole vector $\mathbf{y}_t$ (equivalently, a fundamental white noise of $\mathbf{y}_{st}$ is a fundamental white noise of $\mathbf{y}_t$).

Let us now return to the infinite-dimensional vector $\mathbf{x}_t$ and to the decomposition $\mathbf{x}_t = \boldsymbol{\chi}_t + \xi_t$. Assume that $\boldsymbol{\chi}_t$ has rational spectral density, so that either $\text{rank}(\boldsymbol{\Sigma}_n^{\chi}(\theta)) < q$ for all $\theta \in [-\pi\ \pi]$ or $\text{rank}(\boldsymbol{\Sigma}_n^{\chi}(\theta)) = q$ for almost all θ in $[-\pi\ \pi]$. On the other hand, since $\lambda_{nq}^{\chi}(\theta)$ diverges for almost all θ in $[-\pi\ \pi]$, this is Assumption A.2, there exists s such that $\text{rank}(\boldsymbol{\Sigma}_n^{\chi}(\theta)) = q$ for $n \geq s$ and almost all θ in $[-\pi\ \pi]$. Therefore $\boldsymbol{\chi}_t$ has rank q.

Adding to a rational spectral density the assumption that $\boldsymbol{\chi}_t$ has a one-sided rational representation or, equivalently, that $\mathcal{H}_t^{\chi_s} = \mathcal{H}_t^{\chi}$ for some s, so that cases like (3.1) cannot occur, Lemma 3 ensures that $\boldsymbol{\chi}_t$ has a rational fundamental representation.

More precisely, for $i \in \mathbb{N}$,

$$\chi_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt}, \tag{3.5}$$

where $c_{if}(L)$ and $d_{if}(L)$ are polynomials in L, and $\mathbf{u}_t$ is fundamental for $\boldsymbol{\chi}_t$.

However, in Assumption A.3 (see Section 4.2), we will require more than the existence of an integer s such that $\mathcal{H}_t^{\chi_s} = \mathcal{H}_t^{\chi}$. Rather, we suppose 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 (3.5) 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 in Section 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 4, of the parameter space.

4 AR representations of the vector $\boldsymbol{\chi}_t$

4.1 General results for singular stochastic vectors

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

$$y_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} v_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} v_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} v_{qt} \tag{4.1}$$

with

$$c_{if}(L) = c_{if,0} + c_{if,1} L + \cdots + c_{if,s_1} L^{s_1} \quad \text{and} \quad d_{if}(L) = 1 + d_{if,1} L + \cdots + d_{if,s_2} L^{s_2} \tag{4.2}$$

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.

We assume that for any i the filters in (4.2) are parameterized in the same set $\Pi \subset \mathbb{R}^{\nu}$, with $\nu = q(s_1 + s_2 + 1)$, where

(I) Π is the closure of an open subset of $\mathbb{R}^{\nu}$;

(II) $d_{if}(L)$ has no root of modulus smaller than or equal to unity, for $f = 1, 2, \ldots, q$.

Thus there exists a real $\phi > 1$ such that all the roots of the polynomials $d_{if}(L)$ are of modulus greater or equal to ϕ.

As a consequence, the vector $\mathbf{y}_t$ is described by a parameter vector taking values in $\Pi^n = \underbrace{\Pi \times \Pi \times \cdots \times \Pi}_{n}$, which is a subset of $\mathbb{R}^\mu$, with $\mu = n\nu$. Π^n is the closure of a non-empty open subset.

We are interested in the case $n > q$. Such "tall systems" have been studied recently by B. Anderson, M. Deistler and coauthors (see in particular, Anderson and Deistler, 2008a and b). One of their results is that when $n > q$, there exists a nowhere dense set $\mathcal{N} \subset \Pi^n$, i.e. a set whose closure has no interior points, such that if the parameter vector lies in $\Pi^n - \mathcal{N}$, $\mathbf{y}_t$ has an autoregressive representation of the form

$$\mathbf{A}(L)\mathbf{y}_t = \mathbf{R}\mathbf{v}_t, \tag{4.3}$$

where

(i) $\mathbf{R}$ is $n \times q$, $\text{rank}(\mathbf{R}) = q$;

(ii) $\mathbf{A}(L)$ is an $n \times n$ *finite-degree* matrix polynomial.

When a property holds in $\Pi^n - \mathcal{M}$ and $\mathcal{M}$ is nowhere dense in Π^n, we say that the property holds *generically* in Π^n. As $\mathbf{R}$ has generically full rank, (4.3) implies 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

$$\begin{aligned} y_{1t} &= a_1 v_t + b_1 v_{t-1} \\ y_{2t} &= a_2 v_t + b_2 v_{t-1}, \end{aligned} \tag{4.4}$$

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

$$v_t = \frac{1}{a_1 b_2 - a_2 b_1}(b_2 y_{1t} - b_1 y_{2t}). \tag{4.5}$$

Using (4.5) to get rid of v_{t-1} in (4.4), we obtain the AR(1) representation

$$\begin{aligned} y_{1t} &= d b_1 b_2 y_{1t-1} - d b_1^2 y_{2t-1} + a_1 v_t \\ y_{2t} &= d b_2^2 y_{1t-1} - d b_1 b_2 y_{2t-1} + a_2 v_t, \end{aligned} \tag{4.6}$$

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

(i) If $a_1b_2-a_2b_1=0$, no finite-degree 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_1L (which is also the root of a_2+b_2L) has modulus larger than unity.

(ii) However, as soon as $a_1b_2-a_2b_1\neq 0$, v_t is fundamental for $\mathbf{y}_t$ even if both the roots of a_i+b_iL, $i=1,2$, are smaller than unity in modulus.

(iii) Quite obviously $a_1b_2-a_2b_1\neq 0$ if and only if y_{1t-1} and y_{2t-1} are linearly independent. Therefore, generically, the projection (4.6) is unique, i.e. generically no other autoregressive representation of degree one exists.

(iv) But higher-degree autoregressive representations do exist. Rewriting (with obvious definitions of $\mathbf{A}$ and $\mathbf{a}$) (4.6) 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 (4.5) to get rid of v_{t-1}, we obtain another autoregressive representation, of degree two. Such non-uniqueness does not occur for square systems (when $n=q$).

(v) On the other hand, if $n=3$ and $y_{it}=a_iv_t+b_iv_{t-1}$, $i=1,2,3$, then, outside of the set in which $a_2b_1=a_1b_2$ and $a_3b_1=a_1b_3$, which is nowhere dense in $\mathbb{R}^2\times\mathbb{R}^2\times\mathbb{R}^2$, we have

$$v_t=\frac{1}{a_1\gamma_1+a_2\gamma_2+a_3\gamma_3}(\gamma_1y_{1t}+\gamma_2y_{2t}+\gamma_3y_{3t}),$$

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 (4.1)-(4.2), 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 $\Pi^{q+1}-\mathcal{N}$,*

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

(ii) Suppose that (i) $\mathbf{A}^*(L)$ *is a* $(q+1)\times(q+1)$ *polynomial matrix whose degree 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* $\mathbf{A}^*(L)=\mathbf{A}(L)$*,* $\mathbf{R}^*=\mathbf{RQ}$*,* $\mathbf{v}_t^*=\mathbf{Q}'\mathbf{v}_t$*, where* $\mathbf{Q}$ *is an orthogonal* $q\times q$ *matrix.*

See Appendix A for the proof.

Part (i) of Proposition 1 has already been proved in the papers by Anderson and Deistler, as we have mentioned above. However, the parameters in Anderson and Deistler's papers are the entries of the matrices in the state-space representation of the rational-spectrum vector $\mathbf{y}_t$, whereas our parameters are the coefficients of the rational functions in representation (4.1).

Note that Proposition 1 does not claim that, generically, the process $\mathbf{y}_t$ corresponding to a parameter vector 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 (4.4) 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

$$y_{it}=\left[(a_i+b_iL)\frac{1-aL}{1-a^{-1}L}\right]\left[\frac{1-a^{-1}L}{1-aL}v_t\right]=\frac{(a_i+b_iL)(1-aL)}{1-a^{-1}L}w_t, \qquad (4.7)$$

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). Thus $\mathbf{y}_t$ has the non-fundamental representation (4.7). The latter however 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 (4.1) 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*[7]. 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.

[7]Let us recall that a basis for the open sets in Π^∞ in the product topology is the family of all sets $\prod_{i=1}^\infty G_i$, where G_i is an open subset of Π and $G_i=\Pi$ but for a *finite* number of values of i.

Define the set $\mathcal{M}_m$, for $m \geq q+1$, 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 $i_1 < i_2 < \cdots < i_{q+1} \leq m$, admit a representation of the form

$$\mathbf{A}^{i_1,i_2,\ldots,i_{q+1}}(L)\mathbf{y}_t^{i_1,i_2,\ldots,i_{q+1}} = \mathbf{R}^{i_1,i_2,\ldots,i_{q+1}}\mathbf{v}_t, \tag{4.8}$$

where $\mathbf{A}^{i_1,i_2,\ldots,i_{q+1}}(L)$ is at most of degree S and unique in the sense of Proposition 1(b). From Proposition 1, we see that $\mathcal{N}_m = \Pi^\infty - \mathcal{M}_m$ is a nowhere dense subset in the product topology of Π^∞, so that the set $\mathcal{N} = \cup_{m=q+1}^{\infty}\mathcal{N}_m$, being a countable union of nowhere dense subsets of Π^∞, is a meagre subset. Thus:

Lemma 4 *Assume that $\mathbf{y}_t$ is infinite-dimensional, modeled as in (4.1) for $i \in \mathbb{N}$ and parameterized in Π^∞ . Generically in Π^∞, all the 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 $i_1 < i_2 < \cdots < i_{q+1}$, can be represented as in (4.8), where $\mathbf{A}^{i_1,i_2,\ldots,i_{q+1}}(L)$ is at most of degree than S and unique in the sense of Proposition 1(b).*

Some observations are in order. Firstly, definining 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.[8]

On the other hand, 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 m and all $n > m$, $\mathbf{c}'_n = \mathbf{c}'_m$. 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).

[8]Let us recall that: (I) because Π is a closed subset of $\mathbb{R}^\nu$, the space Π^∞ is the Cartesian product of a countable family of complete metric spaces and is therefore a complete metric space (Dunford and Schwartz (1988), p. 32, Lemma 4); (II) in complete metric spaces the complement of a meagre subset is not meagre (same reference, Baire Category Theorem, p. 20).

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 (4.1), 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 degree of $\mathbf{A}(L)$.

4.2 Existence of AR representations of $\boldsymbol{\chi}_t$

Let us now turn our attention to the common-component vector $\boldsymbol{\chi}_t$. As we have seen, assuming that $\boldsymbol{\chi}_t$ has rational spectral density and a one-sided rational representation implies, by Lemma 3, that $\boldsymbol{\chi}_t$ has a fundamental rational representation of the form (4.1). The meagreness argument developed in Section 4.1, as summarized in Lemma 4, provides a motivation for assuming more:

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

$$\mathbf{A}^{i_1,i_2,\ldots,i_{q+1}}(L)\boldsymbol{\chi}_t^{i_1,i_2,\ldots,i_{q+1}} = \mathbf{R}^{i_1,i_2,\ldots,i_{q+1}}\mathbf{u}_t, \tag{4.9}$$

where $\mathbf{A}^{i_1,i_2,\ldots,i_{q+1}}(L)$ is of degree not greater than $S = qs_1 + q^2s_2$, 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 (4.9) is unique in the sense of Proposition 1(ii).

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

$$\mathbf{A}^1(L)\begin{pmatrix}\chi_{1t}\\ \chi_{2t}\\ \vdots\\ \chi_{q+1,t}\end{pmatrix} = \mathbf{R}^1\mathbf{u}_t, \qquad \mathbf{A}^2(L)\begin{pmatrix}\chi_{q+2,t}\\ \chi_{q+3,t}\\ \vdots\\ \chi_{2(q+1),t}\end{pmatrix} = \mathbf{R}^2\mathbf{u}_t, \qquad \dots \tag{4.10}$$

where the degrees 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 degree not greater than S. Writing $\underline{\mathbf{A}}(L)$ for the (infinite) block-diagonal matrix with diagonal blocks $\mathbf{A}^1(L), \mathbf{A}^2(L), \dots$, and letting $\underline{\mathbf{R}} = (\mathbf{R}^{1\prime}, \mathbf{R}^{2\prime}, \cdots)'$, we thus have

$$\underline{\mathbf{A}}(L)\boldsymbol{\chi}_t = \underline{\mathbf{R}}\mathbf{u}_t. \tag{4.11}$$

Two comments are in order. Firstly, of course, any permutation of the variables produces a distinct $(q+1)$-blockwise autoregressive representation of ghe form (4.10). This is consistent with the observation in Section 4.1 that autoregressive representations of singular vectors are not unique, even if their degree is minimum, unless $n = q+1$, see Proposition 1.

Secondly, $\mathbf{u}_t$ and $\underline{\mathbf{R}}$ do not play any special role. By Lemma 3(ii), all the white noise vectors $\tilde{\mathbf{u}}_t$ and matrices $\tilde{\underline{\mathbf{R}}}$, corresponding to alternative representations of the form (4.11) satisfy $\tilde{\underline{\mathbf{R}}} = \mathbf{Q}\underline{\mathbf{R}}$, and $\tilde{\mathbf{u}}_t = \mathbf{Q}'\mathbf{u}_t$ where $\mathbf{Q}$ is an orthogonal $q \times q$ matrix.[9] For identification and estimation of a couple $\mathbf{u}_t^*$, $\underline{\mathbf{R}}^*$ based on economic theory, see Forni *et al.* (2009) and Forni *et al.* (2012).

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

Assumption A.3 ensures existence and uniqueness of the autoregressive representation (4.10). We now show how (4.10), 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.

[9] Of course, $\underline{\mathbf{R}}\mathbf{u}_t$, which is the one-step-ahead prediction error of $\boldsymbol{\chi}_t$, is identified.

(ii) Denote by $\boldsymbol{\chi}_t^k$ the k-th of the $(q+1)$-dimensional subvectors of $\boldsymbol{\chi}_t$ appearing in (4.10), 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 $\boldsymbol{\chi}_{t-s}^k$,

$$\boldsymbol{\Gamma}_{jk,s}^{\chi} = \mathrm{E}\left[\boldsymbol{\chi}_t^j \boldsymbol{\chi}_{t-s}^{k\,\prime}\right] = \int_{-\pi}^{\pi} e^{is\theta}\boldsymbol{\Sigma}_{jk}^{\chi}(\theta)d\theta. \tag{4.12}$$

(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

$$\boldsymbol{\Psi}_t^1 = \mathbf{A}^1(L)\boldsymbol{\chi}_t^1, \quad \boldsymbol{\Psi}_t^2 = \mathbf{A}^2(L)\boldsymbol{\chi}_t^2, \quad \ldots \tag{4.13}$$

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

$$\mathbf{A}^{[k]} = \left(\mathbf{A}_1^k\ \mathbf{A}_2^k\ \cdots\ \mathbf{A}_S^k\right), \quad \mathbf{B}_k^{\chi} = \left(\boldsymbol{\Gamma}_{kk,1}^{\chi}\ \boldsymbol{\Gamma}_{kk,2}^{\chi}\ \cdots\ \boldsymbol{\Gamma}_{kk,S}^{\chi}\right) \tag{4.14}$$

and

$$\mathbf{C}_{jk}^{\chi} = \begin{pmatrix} \boldsymbol{\Gamma}_{jk,0}^{\chi} & \boldsymbol{\Gamma}_{jk,1}^{\chi} & \cdots & \boldsymbol{\Gamma}_{jk,S-1}^{\chi} \\ \boldsymbol{\Gamma}_{jk,-1}^{\chi} & \boldsymbol{\Gamma}_{jk,0}^{\chi} & \cdots & \boldsymbol{\Gamma}_{jk,S-2}^{\chi} \\ \vdots & & & \vdots \\ \boldsymbol{\Gamma}_{jk,-S+1}^{\chi} & \boldsymbol{\Gamma}_{jk,-S+2}^{\chi} & \cdots & \boldsymbol{\Gamma}_{jk,0}^{\chi} \end{pmatrix}. \tag{4.15}$$

We have

$$\mathbf{A}^{[k]} = \mathbf{B}_k^{\chi}\left(\mathbf{C}_{kk}^{\chi}\right)^{-1} = \mathbf{B}_k^{\chi}\left(\mathbf{C}_{kk}^{\chi}\right)_{\mathrm{ad}}\det\left(\mathbf{C}_{kk}^{\chi}\right)^{-1} \quad \text{and} \quad \boldsymbol{\Gamma}_{jk}^{\psi} = \boldsymbol{\Gamma}_{jk}^{\chi} - \mathbf{A}^{[j]}\mathbf{C}_{jk}^{\chi}\mathbf{A}^{[k]}, \tag{4.16}$$

where "ad" denotes the adjoint of a square matrix. Invertibility of the matrix $\mathbf{C}_{kk}^{\chi}$ is a consequence of Assumption A.3.

(iv) The $\infty\times\infty$ matrix $\underline{\boldsymbol{\Gamma}}^{\Psi}$ obtained by piecing together the matrices $\boldsymbol{\Gamma}_{jk}^{\Psi}$ is of rank q (see Lemma 3(iii)) and can therefore be represented as $\underline{\boldsymbol{\Gamma}}^{\Psi} = \underline{\mathbf{S}}\ \underline{\mathbf{S}}'$, where $\underline{\mathbf{S}}$ is an $\infty\times q$ matrix. On the other hand, $\underline{\boldsymbol{\Gamma}}^{\Psi}$ is the covariance matrix of the right-hand side terms in (4.10), so that $\underline{\mathbf{S}} = \underline{\mathbf{R}}\mathbf{H}$, where $\mathbf{H}$ is $q\times q$ and orthogonal.

Lastly, using $\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t$, letting $\mathbf{Z}_t = \underline{\mathbf{A}}(L)\mathbf{x}_t$ and $\boldsymbol{\Phi}_t = \underline{\mathbf{A}}(L)\boldsymbol{\xi}_t$, we obtain

$$\mathbf{Z}_t = \underline{\mathbf{R}}\mathbf{u}_t + \boldsymbol{\Phi}_t. \tag{4.17}$$

In conclusion, starting with the spectral density of the χ's we obtain the filter $\underline{\mathbf{A}}(L)$, the vector $\mathbf{Z}_t$ and the model (4.17). The above construction, based on an estimate of the spectral density $\mathbf{\Sigma}_n^\chi(\theta)$, rather than $\mathbf{\Sigma}_n^\chi(\theta)$ itself, is used, step by step, in the estimation procedure studied in Forni *et al.* (2012), see Section 4.4 for a sketch.

4.4 Normalization of $\mathbf{Z}_t$

Under our assumptions, the dynamic factor model for the variables x_{it} has been transformed into model (4.17), which has the form (2.2) for the variables Z_{it}, with $r = q$ and $\mathbf{F}_t = \mathbf{u}_t$. Application of standard principal components to estimate $\mathbf{u}_t$ and $\underline{\mathbf{R}}$ requires that Assumptions B.1 and B.2 are fulfilled. The latter are equivalent to statements (i), (ii) and (iii) of Theorem B, see Section 2.2. In particular, the first eigenvalue of the variance-covariance matrix of $\mathbf{\Phi}_{nt}$ should be bounded. We show below that this is neither a consequence of our assumptions so far, nor can be a reasonable additional assumption.

To see this, let us resort again to the simple case in which $q = 1$ and the common components are MA(1),

$$x_{it} = u_t + c_i u_{t-1} + \xi_{it}.$$

Considering the 2-dimensional vectors $\boldsymbol{\chi}_t^k$, we have, see (4.6):

$$\mathbf{A}^k(L) = \mathbf{I}_2 - (c_k - c_{k-1})^{-1} \begin{pmatrix} c_{k-1}c_k & -c_{k-1}^2 \\ c_k^2 & -c_{k-1}c_k \end{pmatrix} L.$$

Assumption 3 implies that $c_k - c_{k-1} \neq 0$ for all k (and all possible groupings), but no more. In particular, it does not imply that $|c_k - c_{k-1}| \geq d$ for some $d > 0$ and all k. As a consequence, the variance of the components of $\mathbf{\Phi}_t = \underline{\mathbf{A}}(L)\boldsymbol{\xi}_t$ is not necessarily bounded, as it should be if $\mathbf{\Phi}_t$ were idiosyncratic.

On the other hand, the set of all sequences $(c_1\ c_2\ \cdots)$ such that $|c_k - c_{k-1}| \geq d$, for some $d > 0$ and all k, call it $\mathcal{S}$, is nowhere dense in Π^∞. For, as already observed in Section 4.1, given any neighborhood U of the sequence $(c_1\ c_2\ \cdots) \in \mathcal{S}$, the set U contains sequences $(c_1'\ c_2'\ \cdots)$ such that $c_s' = c_{s-1}'$ for s bigger than some s'. Of course

$|c'_k - c'_{k-1}| < d$ for $k > s'$ and any $d > 0$. Therefore, *assuming* that $|c_k - c_{k-1}| \geq d$ for some $d > 0$ and all k would be extremely restrictive.[10]

A simple and effective way out of this difficulty is the normalization of $\mathbf{Z}_t$. Define

(i) $w_i = 1$ if $\mathrm{var}(Z_{it}) = 0$, otherwise $w_i = \sqrt{\mathrm{var}(Z_{it})}$;

(ii) $\underline{\mathbf{V}}$ as the $\infty \times \infty$ diagonal matrix with w_i^{-1} in entry (i, i);

(iii) $\mathbf{z}_t = \underline{\mathbf{V}}\mathbf{Z}_t$, $\underline{\mathbf{r}} = \underline{\mathbf{V}}\ \underline{\mathbf{R}}$, $\boldsymbol{\phi}_t = \underline{\mathbf{V}}\boldsymbol{\Phi}_t$.

Equation (4.17) becomes

$$\mathbf{z}_t = \underline{\mathbf{r}}\mathbf{u}_t + \boldsymbol{\phi}_t. \tag{4.18}$$

Adding the following assumption is sufficient, though not necessary, to prove that $\boldsymbol{\phi}_t$ fulfills statement (i) of Theorem B.

Assumption A.4 *There exists a real $b^\xi > 0$ such that $\lambda^\xi_{nn}(\theta) \geq b^\xi$ for all n and θ almost everywhere in $[-\pi\ \pi]$ ($\lambda^\xi_{nn}(\theta)$ is the smallest eigenvalue of $\boldsymbol{\Sigma}^\xi_n(\theta)$, see Section 2.2).*

Proposition 2 *Let $\boldsymbol{\Gamma}^\phi_n$ the variance-covariance matrix of $\boldsymbol{\phi}_{nt}$ and μ^ϕ_{n1} its first eigenvalue. Under Assumptions A.1, A.2, A.3 and A.4, there exist a real M such that $\mu^\phi_{n1} \leq M$ for all n.*

Proof. 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. Let $\mathbf{b}$ be a $1 \times n$ vector with $|\mathbf{b}| = 1$. The following notation $\mathbf{b} = (\mathbf{b}^1\ \mathbf{b}^2\ \cdots\ \mathbf{b}^m)$ and $\mathbf{V}_m = \mathrm{diag}(\mathbf{V}^1\ \mathbf{V}^2\ \cdots\ \mathbf{V}^m)$ is used in an obvious way. We denote by $\boldsymbol{\Sigma}^{\xi k}(\theta)$ the spectral density matrix of $\boldsymbol{\xi}^k_t$ and by $\mathbf{a}^k_j(e^{-i\theta})$ the j-th row of $\mathbf{A}^k(e^{-i\theta})$, for $j = 1, 2, \ldots, q + 1$. Let $\mathbf{c} = (\mathbf{c}^1\ \mathbf{c}^2\ \cdots\ \mathbf{c}^m)$, and suppose that $\mathbf{c}^j = \mathbf{0}$ if $j \neq k$. Then $\mathbf{c}\boldsymbol{\Sigma}^\xi_n(\theta)\mathbf{c}' = \mathbf{c}^k\boldsymbol{\Sigma}^{\xi k}(\theta)\mathbf{c}^{k\prime}$. As a consequence, if $\mathbf{d}$ is $1 \times (q + 1)$, then

$$\lambda^\xi_{nn}(\theta)\mathbf{d}\mathbf{d}' \leq \mathbf{d}\boldsymbol{\Sigma}^{\xi k}(\theta)\mathbf{d}' \leq \lambda^\xi_{n1}(\theta)\mathbf{d}\mathbf{d}', \tag{4.19}$$

[10] Taking a different approach and supposing that the coefficients c_i are independently drawn from a probability distribution, we reach a similar conclusion. In this case the event $\{|c_k - c_{k-1}| \geq d$ for some $d > 0$ and all $k\}$ has probability zero.

for $k = 1, 2, \ldots, m$. Using Assumption A.4, statement (i) of Theorem A and (4.19),

$$
\begin{aligned}
&\mathbf{b}\mathbf{\Sigma}^{\phi}_{n}(\theta)\mathbf{b}' = \mathbf{b}\mathbf{V}\mathbf{A}(e^{-i\theta})\mathbf{\Sigma}^{\xi}(\theta)\mathbf{A}'(e^{i\theta})\mathbf{V}\mathbf{b}' \le \lambda^{\xi}_{n1}(\theta)\mathbf{b}\mathbf{V}\mathbf{A}(e^{-i\theta})\mathbf{A}'(e^{i\theta})\mathbf{V}\mathbf{b}' \\
&\quad = \lambda^{\xi}_{n1}(\theta)\sum_{k=1}^{m}\mathbf{b}^{k}\mathbf{V}^{k}\mathbf{A}^{k}(e^{-i\theta})\mathbf{A}^{k\prime}(e^{i\theta})\mathbf{V}^{k}\mathbf{b}^{k\prime} \\
&\quad \le \lambda^{\xi}_{n1}(\theta)\sum_{k=1}^{m}\mathbf{b}^{k}\text{trace}\left[\mathbf{V}^{k}\mathbf{A}^{k}(e^{-i\theta})\mathbf{A}^{k\prime}(e^{i\theta})\mathbf{V}^{k}\right]\mathbf{b}^{k\prime} = \lambda^{\xi}_{n1}(\theta)\sum_{k=1}^{m}\mathbf{b}^{k}\left[\sum_{j=1}^{q+1}\frac{\mathbf{a}^{k}_{j}(e^{-i\theta})\mathbf{a}^{k\prime}_{j}(e^{i\theta})}{\text{var}(Z^{k}_{it})}\right]\mathbf{b}^{k\prime} \\
&\quad \le \lambda^{\xi}_{n1}(\theta)\sum_{k=1}^{m}\mathbf{b}^{k}\left[\sum_{j=1}^{q+1}\frac{\mathbf{a}^{k}_{j}(e^{-i\theta})\mathbf{a}^{k\prime}_{j}(e^{i\theta})}{\int_{-\pi}^{\pi}\mathbf{a}^{k}_{j}(e^{-i\theta})\mathbf{\Sigma}^{\xi k}(\theta)\mathbf{a}^{k\prime}_{j}(e^{i\theta})d\theta}\right]\mathbf{b}^{k\prime} \\
&\quad \le \frac{B^{\xi}}{b^{\xi}}\sum_{k=1}^{m}\mathbf{b}^{k}\left[\sum_{j=1}^{q+1}\frac{\mathbf{a}^{k}_{j}(e^{-i\theta})\mathbf{a}^{k\prime}_{j}(e^{i\theta})}{\int_{-\pi}^{\pi}\mathbf{a}^{k}_{j}(e^{-i\theta})\mathbf{a}^{k\prime}_{j}(e^{-i\theta})d\theta}\right]\mathbf{b}^{k\prime},
\end{aligned}
$$

for θ a.e. in $[-\pi\ \pi]$. Integrating we obtain

$$
\mathbf{b}\mathbf{\Gamma}^{\phi}_{n}\mathbf{b}' = \int_{-\pi}^{\pi}\mathbf{b}\mathbf{\Sigma}^{\phi}_{n}(\theta)\mathbf{b}'d\theta \le \frac{B^{\xi}}{b^{\xi}}(q+1),
$$

which implies that $\mu^{\phi}_{n1} = \max_{|\mathbf{b}|=1}\mathbf{b}\mathbf{\Gamma}^{\phi}_{n}\mathbf{b}'$ is bounded. Q.E.D.

Let us now consider statements (ii) and (iii) of Theorem B, The definition of $\boldsymbol{\phi}_t$ and statement (i) of Theorem A imply that $\boldsymbol{\phi}_t$ and $\boldsymbol{\eta}_t = \underline{\mathbf{r}}_t\mathbf{u}_t$ fulfill statement (iii). As regards statement (ii), let again $q = 1$ and

$$
x_{it} = (c_{i0} + c_{i1}L)u_t + \xi_{it}.
$$

The corresponding representation (4.18) is

$$
z_{it} = d_{i0}u_t + \phi_{it} = \eta_{it} + \phi_{it}, \qquad d_{i0} = \frac{c_{i0}}{\sqrt{c^2_{i0} + \text{var}(\Phi_{it})}}.
$$

We have:

$$
\lambda^{\chi}_{n1}(\theta) = \frac{1}{2\pi}\sum_{i=1}^{n}|c_{i0} + c_{i1}e^{-i\theta}|^2 \quad \text{and} \quad \mu^{\eta}_{n1} = \sum_{i=1}^{n}\frac{c^2_{i0}}{c^2_{i0} + \text{var}(\Phi_{it})}.
$$

We see that divergence of $\lambda^{\chi}_{n1}(\theta)$ almost everywhere in $[-\pi\ \pi]$ does not imply divergence of μ^{η}_{n1}. However, convergence of μ^{η}_{n1} occurs only if $\text{var}(\Phi_{it})/c^2_{i0}$ diverges. Sufficient conditions for this are (1) $\text{var}(\Phi_{it}) \to \infty$ and c_{i0}^2 bounded away from zero, (2) $\text{var}(\Phi_{it})$ bounded away from zero and $c_{i0}^2 \to 0$. Regarding (1), as argued above, we cannot assume that $\text{var}(\Phi_{it})$ is bounded. However, divergence of $\text{var}(\Phi_{it})$ requires a very special sequence of coefficients $(c_{i0},\ c_{i1})$. Regarding (2), even if we do not assume a positive lower bound for c_{i0}, convergence to zero of c_{i0}^2 can be ruled out as very special. Even more far-fetched are the cases in which the ratio $\text{var}(\Phi_{it})/c_{i0}^2$ diverges though neither (1) nor (2) holds, like the ratio α_1/β_1 with

$$\alpha_i = \begin{cases} i \text{ for } i \text{ odd} \\ 1 \text{ for } i \text{ even} \end{cases} \qquad \beta_i = \begin{cases} 1 \text{ for } i \text{ odd} \\ 1/i \text{ for } i \text{ even} \end{cases}$$

Extending these considerations to $q > 1$ and more complex models for $\boldsymbol{\chi}_t$ does not seem worthwhile. We believe that the analysis of the simple example above is sufficient to motivate the following assumption on the q-th eigenvalue of the variance-covariance matrix of $\mathbf{z}_t$:

Assumption A.5 $\mu_{nq}^{\eta} \to \infty$ *as* $n \to \infty$.

Summing up, under Assumptions A.1 through A.5, the variables x_{it} can be transformed into the variables z_{it}, which evolves according to the static model (4.18). Statements (i), (ii) and (iii) of Theorem B are fulfilled or, equivalently, the variables z_{it} fulfill Assumptions B.1 and B.2.

The construction leading from the x's to the z's has a step-by-step counterpart in the estimation procedure developed in the companion paper Forni *et al.* (2013):

(I) We start with an estimate of $\mathbf{\Sigma}_n^x(\theta)$, the spectral density of the observable variables x_{it}, call $\hat{\mathbf{\Sigma}}_n^x(\theta)$ such an estimate.

(II) An estimate of the spectral density of the common components, call it $\hat{\mathbf{\Sigma}}_n^{\chi}(\theta)$, is then obtained using the first q dynamic principal components of $\hat{\mathbf{\Sigma}}_n^x(\theta)$, see Forni *et al.* (2000). An estimate of the spectral density of the idiosyncratic components is obtained as well as $\hat{\mathbf{\Sigma}}_n^{\xi}(\theta) = \hat{\mathbf{\Sigma}}_n^x(\theta) - \hat{\mathbf{\Sigma}}_n^{\chi}(\theta)$.

(III) Steps (ii), (iii) and (iv) of Section 4.3 are then reproduced starting with $\hat{\mathbf{\Sigma}}_n^{\chi}(\theta)$ instead of $\mathbf{\Sigma}_n^{\chi}(\theta)$. We thus obtain estimates $\hat{\mathbf{A}}^k(L)$, $\hat{\mathbf{R}}^k$, $\hat{\mathbf{Z}}_{nt}$, $\hat{\mathbf{r}}^k$, $\hat{\mathbf{z}}_{nt}$. Note that $\hat{\mathbf{Z}}_{nt}$ and $\hat{\mathbf{z}}_{nt}$ result from the application of *one-sided filters* to the observable variables x_{it}.

(IV) Lastly, we estimate a static representation with q factors for $\hat{\mathbf{z}}_{nt}$, obtaining an estimate $\hat{\mathbf{u}}_t$. This step employs the first q principal components of $\hat{\mathbf{z}}_{nt}$, and therefore only *current and past values* of the variables x_{it}.

5 Conclusion

We have argued that assuming a finite-dimensional factor space strongly restricts GDFM's, as even models as simple as $x_{it} = [a_i/(1 - \alpha_i L)]\, u_t + \xi_{it}$ are ruled out. On the other hand, without that assumption, only two-sided estimators have been proposed in the literature so far.

The present paper provides a solution to this problem by means of a feasible autoregressive representation of the huge-dimensional common-component vector $\boldsymbol{\chi}_{nt}$. The key result is that if a stochastic vector $\boldsymbol{\chi}_{nt}$ has dimension n and rank q, where q is fixed whereas n is huge and growing, then, under some mild assumptions, for generic values of the parameters, an autoregressive representation for $\boldsymbol{\chi}_{nt}$ can be determined piecewise. We do not need a huge, unfeasible, $n \times n$ VAR, in which each y_{it} is projected on all y_{jt-k}, $j = 1, 2, \ldots, n$. A sequence of small $(q+1) \times (q+1)$ VAR's is sufficient.

Using the autoregressive representation of $\boldsymbol{\chi}_{nt}$ we transform the original variables x_{it} into variables z_{it} that are governed by a static factor model. All the steps of our construction have a natural counterpart in an estimation procedure.

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

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

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

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

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

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

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

[10] 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.

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

[12] 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.

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

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

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

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

[17] 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.

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

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

[20] 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.

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

[22] 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.

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

[24] 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.

[25] 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.

[26] Forni, M., M. Hallin, M. Lippi and P. Zaffaroni (2013). Dynamic factor models with infinite dimensional factor space: estimation and asymptotics, EIEF Working Paper, forthcoming.

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

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

[29] 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.

[30] 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.

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

[32] 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.

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

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

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

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

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

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

[39] Quah, D. and T.J. Sargent (1993). A Dynamic Index Model for Large Cross Sections, in *Business Cycles, Indicators and Forecasting*, National Bureau of Economic Research.

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

[41] Sargent, T.J. and 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.

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

[43] 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.

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

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

Appendix

A Proof of Proposition 1

A polynomial of the form

$$p(L) = a_0 + a_1 L + \cdots + a_r L^r,$$

where the coefficients a_k are either scalar or matrices, is said to have degree not greater than r; we say that $p(L)$ has degree r if $a_r \neq 0$. We need some preliminary results.

Lemma A.1 *Assume that* $\mathbf{v}_t = (v_{1t}\ v_{2t}\ \ldots\ v_{qt})$ *is orthonormal white noise and let*

$$y_{it} = \gamma_{i1}(L)v_{1t} + \gamma_{i2}(L)v_{2t} + \cdots + \gamma_{iq}(L)v_{qt},$$

for $i = 1, 2, \ldots, n$, *where the filters* $\gamma_{if}(L)$ *are square summable. In compact form,*

$$\mathbf{y}_t = \mathbf{\Gamma}(L)\mathbf{v}_t,$$

where $\mathbf{\Gamma}(L)$ *is* $n \times q$. *For* $R \geq 1$ *consider the* nR*-dimensional stack*

$$\mathbf{Y}_t = (\mathbf{y}_t'\ \mathbf{y}_{t-1}'\ \cdots\ \mathbf{y}_{t-R+1}')'$$

and the $1 \times q$ *filter*

$$\mathbf{W}(L) = \Big(\beta_1(L) \quad \beta_2(L) \quad \cdots \quad \beta_n(L)\Big)\, \mathbf{\Gamma}(L),$$

where $\beta_i(L)$ *is a finite-degree polynomial in* L, $i = 1, 2, \ldots, n$. *The entries of* $\mathbf{Y}_t$ *are linearly dependent if and only if there exist polynomials* $\beta_i(L)$ *of degree not greater than* $R-1$, *with* $\beta_i(L) \neq 0$ *for some* i, *such that* $\mathbf{W}(L) = \mathbf{0}$. *Equivalently, the entries of* $\mathbf{Y}_t$ *are linearly independent if and only if* $\mathbf{W}(L) = \mathbf{0}$ *implies that either* $\beta_i(L) = 0$ *for all* i *or that the degree of* $\beta_i(L)$ *is greater than* $R-1$ *for some* i.

Proof. If the entries of $\mathbf{Y}_t$ are linearly dependent, there exists

$$\boldsymbol{\alpha} = (\alpha_{01} \cdots \alpha_{0,n}; \alpha_{11} \cdots \alpha_{1,n}; \cdots; \alpha_{R-1,1} \cdots \alpha_{R-1,n}) \neq \mathbf{0}$$

such that

$$\boldsymbol{\alpha}(\mathbf{y}'_t\ \mathbf{y}'_{t-1}\ \cdots\ \mathbf{y}'_{t-R+1})' = \mathbf{0}, \tag{A.1}$$

that is, setting $\boldsymbol{\alpha}_k = (\alpha_{k1}\ \cdots\ \alpha_{k,n})$,

$$\begin{aligned}\boldsymbol{\alpha}_0\boldsymbol{\Gamma}(L)\mathbf{v}_t + \boldsymbol{\alpha}_1\boldsymbol{\Gamma}(L)v_{t-1} + \cdots + \ & \boldsymbol{\alpha}_{R-1}\boldsymbol{\Gamma}(L)\mathbf{v}_{t-R+1} = \\ & (\boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 L + \cdots + \boldsymbol{\alpha}_{R-1}L^{R-1})\boldsymbol{\Gamma}(L)\mathbf{v}_t = 0.\end{aligned} \tag{A.2}$$

Because $\mathbf{v}_t$ is orthonormal white noise, (A.2) implies that

$$(\boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 L + \cdots + \boldsymbol{\alpha}_{R-1}L^{R-1})\boldsymbol{\Gamma}(L) = \mathbf{0},$$

that is, setting $\beta_i(L) = \alpha_{0i} + \alpha_{1i}L + \cdots\ + \alpha_{R-1,i}L^{R-1}$, $i = 1, 2, \ldots, n$,

$$(\beta_1(L)\ \beta_2(L)\ \cdots\ \beta_{q+1}(L))\boldsymbol{\Gamma}(L) = \mathbf{0}. \tag{A.3}$$

Since $\boldsymbol{\alpha} \neq \mathbf{0}$, $\beta_i(L) \neq 0$ for some i. Conversely, starting with (A.3), where the degree of $\beta_i(L)$ is not greater than $R-1$ and $\beta_i(L) \neq 0$ for some i, we easily obtain an $\boldsymbol{\alpha} \neq \mathbf{0}$ such that (A.1) holds. Q.E.D.

Lemma A.2 *Assume that* $\mathbf{v}_t = (v_{1t}\ v_{2t}\ \ldots\ v_{qt})$ *is orthonormal white noise and*

$$y_{it} = p_{i1}(L)v_{1t} + p_{i2}(L)v_{2t} + \cdots + p_{iq}(L)v_{qt}, \tag{A.4}$$

with

$$p_{if}(L) = p_{if,0} + p_{if,1}L + \cdots + p_{if,r}L^r,$$

for $i = 1, 2, \ldots, q+1,\ f = 1, 2, \ldots, q$. *In compact form,*

$$\mathbf{y}_t = \mathbf{P}_0\mathbf{v}_t + \mathbf{P}_1\mathbf{v}_{t-1} + \cdots + \mathbf{P}_r\mathbf{v}_{t-r} = \mathbf{P}(L)\mathbf{v}_t, \tag{A.5}$$

where the matrices $\mathbf{P}_k$ *are* $(q+1) \times q$. *Let* $R = rq$. *Assume that the entries of the stack* $\mathbf{Y}_t = (\mathbf{y}'_{t-1}\ \mathbf{y}'_{t-2}\ \cdots\ \mathbf{y}'_{t-R})'$ *are linearly independent. Then,*

$$\mathbf{y}_t = \mathbf{H}_1\mathbf{y}_{t-1} + \cdots + \mathbf{H}_R\mathbf{y}_{t-R} + \mathbf{P}_0\mathbf{v}_t, \tag{A.6}$$

for some $(q+1) \times (q+1)$ *matrices* $\mathbf{H}_k$.

PROOF. Consider the stack

$$\mathbf{Y}_{t-1} = (\mathbf{y}'_{t-1}\ \mathbf{y}'_{t-2}\ \cdots\ \mathbf{y}'_{t-R})' = \mathcal{P}_R(\mathbf{v}'_{t-1}\ \mathbf{v}'_{t-2}\ \cdots\ \mathbf{v}'_{t-R-r})'$$

where

$$\mathcal{P}_R = \begin{pmatrix} \mathbf{P}_0 & \mathbf{P}_1 & \cdots & \mathbf{P}_r & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_0 & \cdots & \mathbf{P}_{r-1} & \mathbf{P}_r & \cdots & \mathbf{0} \\ \vdots & & & & & \ddots & \\ \mathbf{0} & \mathbf{0} & \cdots & & & \cdots & \mathbf{P}_r \end{pmatrix}.$$

The matrix $\mathcal{P}_R$ is $(q+1)R \times q(R+r)$. Setting $R = rq$, $\mathcal{P}_R$ is square. By assumption, the entries $\mathbf{Y}_{t-1}$ are linearly independent. Thus the matrix $\mathcal{P}_R$ is non singular, so that $(\mathbf{v}'_{t-1}\ \mathbf{v}'_{t-2}\ \cdots\ \mathbf{v}'_{t-R-r})' = \mathcal{P}_R^{-1}(\mathbf{y}'_{t-1}\ \mathbf{y}'_{t-2}\ \cdots\ \mathbf{y}'_{t-R})'$. Substituting $\mathbf{v}_{t-1}$, $\mathbf{v}_{t-2}$ $\ldots$, $\mathbf{v}_{t-r}$ into (A.5), we get (A.6). Q.E.D.

Lemma A.3 *Rewrite (4.1) in compact form*

$$\mathbf{y}_t = \mathbf{E}(L)\mathbf{v}_t, \tag{A.7}$$

where

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

for $i = 1, 2, \ldots, q+1,\ f = 1, 2, \ldots, q$. *Let* $S = s_1 q + s_2 q^2$. *For generic values of the parameters, the entries of the stack* $(\mathbf{y}'_t\ \mathbf{y}'_{t-1}\ \cdots \mathbf{y}'_{t-S+1})'$ *are linearly independent.*

PROOF. Using the notation of Section 4, let $\mu = (q+1)q(s_1 + s_2 + 1)$. Denote by $p = (p_1\ p_2\ \cdots\ p_\mu)$ the μ-dimensional vectors of Π^{q+1} (the entries of p are the parameters c and d). In this proof, we deal with scalar polynomials in L:

$$a_0 + a_1 L + \cdots + a_r L^r,$$

where the coefficients a_m are polynomials in the parameters of the form

$$\sum_{k_1+k_2+\cdots+k_\mu \le K} \alpha_{k_1,k_2,\ldots,k_\mu} p_1^{k_1} p_2^{k_2} \cdots p_\mu^{k_\mu}. \tag{A.8}$$

Because Π^{q+1} is the closure of an open set in $\mathbb{R}^\mu$, the polynomial (A.8) is generically non zero in Π^{q+1} if and only if at least one coefficient $\alpha_{k_1,k_2,\ldots,k_\mu}$ is non zero. Note also that (A.8) can be rewritten as a polynomial in one of the variables, p_1 for example,

$$A_0 p_1^M + A_1 p_1^{M-1} + \cdots + A_M, \tag{A.9}$$

where the coefficients A_j are polynomials in $p_2, \ldots, p_\mu$, and that (A.8) is generically non zero in Π^{q+1} if and only if at least one of the coefficients A_j in (A.9) is generically non zero.

By Lemma A.1, we must prove that, for generic values in Π^{q+1}, if

$$\begin{pmatrix}\beta_1(L) & \beta_2(L) & \cdots & \beta_n(L)\end{pmatrix} \mathbf{E}(L) = \mathbf{0}, \tag{A.10}$$

where $\beta_i(L)$ is a finite-degree polynomial and $\beta_i(L) \neq 0$ for some i, then the degree of $\beta_i(L)$ is greater than $S-1$ for some i. Let $\mathbf{E}_q(L)$ be the square submatrix obtained by dropping $\mathbf{E}(L)$'s last row. We can write

$$\det(\mathbf{E}_q(L)) = h(L) / \prod_{i,f=1}^{q} d_{if}(L), \tag{A.11}$$

where numerator and denominator have degree not greater than $S_1 = qs_1 + (q^2 - q)s_2$ and $S_2 = s_2 q^2$ respectively. The coefficient of L^{S_2} in the denominator is the product $\prod_{i,f=1}^{q} d_{if,s_2}$ and is therefore generically non zero. The coefficient of L^{S_1} in the numerator contains the term

$$c_{11,s_1} c_{22,s_1} \cdots c_{qq,s_1} \prod_{\substack{i,f=1,q \\ i\neq f}} d_{if,s_2}$$

and no other term with the same exponents for the c's and the d's. Thus generically numerator and denominator in (A.11) have degrees S_1 and S_2, respectively.

Using the same argument, the (i,f) entry of the adjoint matrix of $\mathbf{E}_q(L)$ can be written as

$$h_{if}(L) / \prod_{\substack{h,k=1,\ldots q \\ h\neq f,\ k\neq i}} d_{hk}(L),$$

where generically the degrees of the numerator and the denominator are $S_3 = (q-1)s_1 + [(q-1)^2 - (q-1)]s_2$ and $S_4 = (q-1)^2 s_2$, respectively. Thus, the matrix $\mathbf{E}_q(L)$ is generically invertible, as a matrix of rational functions in L, and the entries of $[\mathbf{E}_q(L)]^{-1}$ can be written as

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

where generically the degrees of the numerator and the denominator are $S_5 = (q-1)s_1 + (q^2-(q-1))s_2$ and $S_6 = qs_1 + (q^2-q)s_2$, respectively.

Consider now 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 $\rho_k(L)$. Generically, the system has the unique solution

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

We have

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

where generically both $\nu_k(L)$ and $\delta(L)$ are polynomials of degree $S = qs_1 + q^2 s_2$. Moreover, for generic values of the parameters, $\nu_k(L)$ and $\delta(L)$ have no roots in common. To show this, recall that the polynomials

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

both of degree S, have roots in common if and only if their *resultant* vanishes. That *resultant* is a polynomial 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 $\nu_{k,S}^{S-h}$ with $0 < h \leq S$. We have

$$\nu_{k,S}^S\delta_0^S = \Big[\sum_{i=1}^{q} c_{q+1,i,s_1}\tilde{h}_{ik,g}\prod_{\substack{j=1,\dots,q\\ j\neq i}} d_{q+1,j,s_2}\Big]^S h(0)^S = c_{q+1,1,s_1}^S\left[\tilde{h}_{1k,S_5}^S\prod_{j=2,\dots,q} d_{q+1,j,s_2}^S h(0)^S\right] + \cdots, \tag{A.12}$$

where $\tilde{h}_{ik,S_5}$ is the coefficient of order S_5 of $\tilde{h}(L)$. Note that $h(L)$ and $\tilde{h}_{if}(L)$ do not contain any of the parameters $c_{q+1,i,h}$. As a consequence, all other terms in (A.12) and in the resultant of $\nu_k(L)$ and $\delta(L)$ contain powers c_{q+1,i,s_1}^{S-h}, with $0 < h \leq S$. Thus the three-term product within square brackets in the right-hand side of (A.12) is the coefficient of $c_{q+1,1,s_1}^S$ in the representation of the resultant as a polynomial in $c_{q+1,1,s_1}$. As each of the three terms is generically non zero, the coefficient is generically non zero, so that the resultant is generically non zero.

Suppose now that the polynomials $\beta_k(L)$'s are such that (A.10) holds, that is

$$(\beta_1(L)\ \beta_2(L)\ \cdots\ \beta_q(L))\mathbf{E}_q(L) = -\beta_{q+1}(L)(e_{q+1,1}(L)\ e_{q+1,2}(L)\ \cdots\ e_{q+1,q}(L)).$$

Because the matrix $\mathbf{E}_q(L)$ is generically non singular, as a matrix of rational functions, $\beta_{q+1}(L) = 0$ implies $\beta_i(L) = 0$ for all $i = 1, 2, \ldots, q+1$. Assuming that $\beta_{q+1}(L) \neq 0$, we have

$$\tau_k(L) = -\frac{\beta_k(L)}{\beta_{q+1}(L)}.$$

The results above on $\tau_k(L)$ imply that generically the degree of $\beta_{q+1}(L)$ and $\beta_k(L)$ is at least S. Q.E.D.

We now can proceed with the proof of Proposition 1. Rewrite (A.7) as

$$\begin{pmatrix} h_1(L) & 0 & \cdots & 0 \\ 0 & h_2(L) & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & h_{q+1}(L) \end{pmatrix} \mathbf{y}_t = \mathbf{G}(L)\mathbf{v}_t, \tag{A.13}$$

where

$$h_i(L) = \prod_{f=1}^{q} d_{if}(L), \quad g_{if}(L) = c_{if}(L) \prod_{\substack{f=1,\ldots,q \\ f\neq i}} d_{if}(L). \tag{A.14}$$

Let us focus on the moving average on the right-hand side. The polynomial matrix $\mathbf{G}(L)$ has degree not greater than $\tilde{S} = s_1 + s_2(q-1)$. Suppose that

$$\begin{pmatrix} \beta_1(L) & \beta_2(L) & \cdots & \beta_{q+1}(L) \end{pmatrix} \mathbf{G}(L) = \mathbf{0}. \tag{A.15}$$

where the degree of $\beta_j(L)$ is not greater than $\tilde{S}q - 1$. This implies that

$$\begin{pmatrix} \beta_1(L)h_1(L) & \beta_2(L)h_2(L) & \cdots & \beta_{q+1}(L)h_{q+1}(L) \end{pmatrix} \mathbf{E}(L) = \mathbf{0}. \tag{A.16}$$

The polynomials $\beta_j(L)h_j(L)$ have degrees not greater than $\tilde{S}q - 1 + s_2 q = s_1 q + s_2 q^2 - 1$. Lemmas A.3 and A.1 imply that generically $\beta_i(L)h_i(L) = 0$ for all $i = 1, 2, \ldots, q+1$. Because $h_i(L) \neq 0$ for all i, then generically (A.15) implies $\beta_i(L) = 0$ for all i. Using Lemma A.2, $\mathbf{G}(L)\mathbf{v}_t$ generically has an autoregressive representation of degree $s_1 q + s_2 q(q-1)$, so that, by (A.13)-(A.14), $\mathbf{y}_t$ generically has an autoregressive representation

$$\mathbf{y}_t = \mathbf{K}_1\mathbf{y}_{t-1} + \mathbf{K}_2\mathbf{y}_{t-2} + \cdots + \mathbf{K}_S\mathbf{y}_{t-S} + \mathbf{E}(0)\mathbf{v}_t \tag{A.17}$$

of degree $S = s_1 q + s_2 q^2$. Moreover, Lemma A.3 proves that generically the components of the stack

$$(\mathbf{y}'_{t-1}\ \mathbf{y}'_{t-2}\ \cdots\ \mathbf{y}'_{t-S})'$$

are independent. The uniqueness part of the proposition follows. Q.E.D.