Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Statistical Planning and Inference 136 (2006) 1237 – 1257

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Nonstationary dynamic factor analysis

Daniel Peña[a], Pilar Poncela[b,*]

[a]*Department of Statistics and Econometrics, Universidad Carlos III de Madrid C/ Madrid, 126, 28903 Getafe (Madrid) Spain*

[b]*Departament of Análisis Económico: Economía Cuantitativa, Facultad de Económicas, E-III, 306, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain*

Received 24 July 2003; accepted 30 August 2004
Available online 27 October 2004

Abstract

In this paper, we present a procedure to build a dynamic factor model for a vector of time series. We assume a model in which the common dynamic structure of the time series vector is explained through a set of common factors, which may be nonstationary, as in the case of common trends. Identification of the nonstationary $I(d)$ factors is made through the common eigenstructure of the generalized covariance matrices, properly normalized. The number of common nonstationary factors is the number of nonzero eigenvalues of the above matrices. A chi-square statistic is proposed to test for the number of factors, stationary or not. The estimation of the model is carried out in state space form. This proposal is illustrated through several simulations and a real data set.

MSC: 62H25; 62H20; 62M10

Keywords: Canonical correlation; Cointegration and common factors; Eigenvectors and eigenvalues; Generalized covariance matrices; Vector time series; Wiener processes

1. Introduction

Factor models are of great importance to achieve dimensionality reduction. When data are dynamic, the so-called curse of dimensionality is an important problem since for vector autoregressive moving average (VARMA) models the number of parameters grows with

* Corresponding author. Tel.:+34 91 4975521; fax: +34 91 4974091.
E-mail addresses: dpena@est-econ.uc3m.es (D. Peña), pilar.poncela@uam.es (P. Poncela).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2004.08.020

the squared of the number of series considered. Some works related to this topic are, for instance, Anderson (1963), Priestley et al. (1974), Box and Tiao (1977), Geweke and Singleton (1981), Brillinger (1981), Velu et al. (1986), Peña and Box (1987), Stock and Watson (1988, 2002), Tiao and Tsay (1989), Reinsel (1993), Ahn (1997) and Forni et al. (2000), among others.

A key problem in factor models is to identify the number of factors. In the stationary case Peña and Box (1987) develop a procedure to identify this number by looking at the eigenvalues of lagged covariance matrices, while Bai and Ng (2002) propose several information criteria to determine the number of factors in approximate factor models, that is, when the factors are approximated by principal components. In the nonstationary case finding the number of nonstationary factors is related to finding the cointegration rank in the econometrics field (whose vast literature we do not pretend to review here), since the number of cointegration relations among the components of a vector of time series is the dimension of the vector minus the number of nonstationary common factors (see Escribano and Peña, 1994). King et al. (1994) estimate factors until the specific variance of the last factor included in the model is zero in a conditional heteroskedastic model.

Dynamic factor models can be estimated in state space form by the Kalman filter. The estimation of the parameters can be done by maximum likelihood through the Expectation-Maximization (EM) or the scoring algorithms. The EM algorithm of Dempster et al. (1977) was first introduced for this kind of models by Shumway and Stoffer (1982) and Watson and Engle (1983), who also compared it to the score algorithm.

In this article we propose a methodology for building dynamic factor models for nonstationary time series. It is organized as follows. Section 2 presents the generalized dynamic factor model and studies its properties. Section 3 generalizes the definition of covariances matrices to the nonstationary case and presents two theorems that characterized the asymptotic behavior of these matrices. Section 4 proposes a chi-square statistic to test for the number of factors. Section 5 summarizes the methodology to build nonstationary factor models. Finally, Section 6 illustrates this proposal through a real data set.

2. The factor model

Let $\mathbf{y}_t$ be an m-dimensional vector of observed time series, generated by a set of $r < m$ nonobserved common factors. We assume that each component of the vector of observed series, $\mathbf{y}_t$, can be written as a linear combination of common factors plus noise

$$\mathbf{y}_t = \mathbf{P}\mathbf{f}_t + \mathbf{e}_t, \tag{1}$$

where $\mathbf{f}_t$ is the r-dimensional vector of common factors, $\mathbf{P}$ is a $m \times r$ factor loading matrix, and the sequence of vectors $\mathbf{e}_t$ are normally distributed, have zero mean and full rank diagonal covariance matrix $\mathbf{\Sigma}_e$. Thus, all the common dynamic structure comes through the common factors, $\mathbf{f}_t$. We suppose that the vector of common factors follows a vector autoregressive moving average, VARMA, (p, q) model

$$\mathbf{\Phi}(B)\mathbf{f}_t = \mathbf{d} + \mathbf{\Theta}(B)\mathbf{a}_t, \tag{2}$$

where $\mathbf{\Phi}(B)=\mathbf{I}-\mathbf{\Phi}_1 B-\cdots-\mathbf{\Phi}_p B^p$ and $\mathbf{\Theta}(B)=\mathbf{I}-\mathbf{\Theta}_1 B-\cdots-\mathbf{\Theta}_q B^q$ are polynomial matrices $r\times r$, B is the backshift operator, such that $B\mathbf{y}_t=\mathbf{y}_{t-1}$, the roots of the determinantal equation $|\mathbf{\Phi}(B)|=0$ can be on or outside the unit circle, $\mathbf{d}$ is a $r\times 1$ vector of constants and $\mathbf{a}_t\sim N_r(\mathbf{0},\mathbf{\Sigma}_a)$ with $\mathbf{\Sigma}_a$ a full rank variance–covariance matrix, is serially uncorrelated, $E(\mathbf{a}_t\mathbf{a}'_{t-h})=\mathbf{0}$, $h\neq 0$. The components of the vector of common factors, $\mathbf{f}_t$, can be either stationary or nonstationary, and we assume the usual conditions for the invertibility of the VARMA model (see, for instance, Lütkepohl, 1993, p. 222). It will be useful to write (2) as $\mathbf{f}_t=\mathbf{U}(B)\mathbf{a}_t$, where $\mathbf{\Phi}(B)\mathbf{U}(B)=\mathbf{\Theta}(B)$, being $\mathbf{U}(B)=\sum_i \mathbf{U}_i B^i$ with $\mathbf{U}_0=\mathbf{I}$.

We assume that both noises appearing in the model are uncorrelated for all lags, $\mathrm{E}(\mathbf{a}_t\mathbf{e}'_{t-h})=\mathbf{0}$, for all $h=0,\pm1,\pm2,\ldots$ When the factors are stationary, model (1) and (2) is the factor model studied by Peña and Box (1987).

The model as stated is not identified because for any $r\times r$ nonsingular matrix $\mathbf{H}$ the observed series $\mathbf{y}_t$ can be expressed in terms of a new set of factors, $\mathbf{y}_t=\mathbf{P}^*\mathbf{f}^*_t+\mathbf{e}_t$, where $\mathbf{\Phi}^*(B)\mathbf{f}^*_t=\mathbf{\Theta}^*(B)\mathbf{a}^*_t$. In this case $\mathbf{P}^{*\prime}\mathbf{P}^*=(\mathbf{H}'^{-1})\mathbf{P}'\mathbf{P}\mathbf{H}^{-1}$, $\mathbf{f}^*_t=\mathbf{H}\mathbf{f}_t$, $\mathbf{a}^*_t=\mathbf{H}\mathbf{a}_t$, $\mathbf{\Phi}^*(B)=\mathbf{H}\mathbf{\Phi}(B)=\mathbf{H}^{-1}$, $\mathbf{\Theta}^*(B)=\mathbf{H}\mathbf{\Theta}(B)\mathbf{H}^{-1}$ and $\mathbf{\Sigma}^*_a=\mathbf{H}\mathbf{\Sigma}_a\mathbf{H}'$. To solve this identification problem, we can always choose either $\mathbf{\Sigma}_a=\mathbf{I}$ or $\mathbf{P}'\mathbf{P}=\mathbf{I}$, but it is easy to see that the model is not yet identified under rotations. The standard restriction in static factor analysis is that $\mathbf{P}'\mathbf{\Sigma}_e^{-1}\mathbf{P}$ should be diagonal. Harvey (1989) and Aguilar and West (2000) impose that $p_{ij}=0$, for $j>i$, where $\mathbf{P}=[p_{ij}]$. This condition is not restrictive, since the factor model can be rotated for a better interpretation when needed (see Harvey, 1989, for a brief discussion about it).

Notice that the model can include also the case where lagged factors are present in Eq. (1). For instance, assume the presence of lagged factors on the observation equation, such as $\mathbf{y}_t=\mathbf{P}\mathbf{v}(B)\mathbf{F}_t+\mathbf{e}_t$ where $\mathbf{v}(B)=\mathbf{I}+\mathbf{v}_1 B+\cdots+\mathbf{v}_l B^l$, $l<\infty$ and $\mathbf{F}_t$ follows a VARMA model $\mathbf{F}_t=\boldsymbol{\beta}(B)\mathbf{a}_t$, $\boldsymbol{\beta}_0=\mathbf{I}$. This model can be rewritten as in Eqs. (1) and (2) with $\mathbf{f}_t=\mathbf{v}(B)\mathbf{F}_t$ following the VARMA model $\mathbf{f}_t=\widetilde{\boldsymbol{\beta}}(B)\mathbf{a}_t$ where $\widetilde{\boldsymbol{\beta}}(B)=\sum_{i=1}^{\infty}\widetilde{\boldsymbol{\beta}}_i B^i=\mathbf{v}(B)\boldsymbol{\beta}(B)$.

3. Properties of stationary and nonstationary factors

Assume that $\mathbf{y}_t$ is $I(d)$. We define the generalized sample covariance matrices $\mathbf{C}_y(k)$ as

$$\mathbf{C}_y(k)=\frac{1}{T^{2d+d\prime}}\sum_{t=k+1}^{T}(\mathbf{y_{t-k}}-\bar{\mathbf{y}})(\mathbf{y_t}-\bar{\mathbf{y}})', \tag{3}$$

where $\bar{\mathbf{y}}=\frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t$ and d' can be either 0 or 1. We will see that for nonstationary factor identification these matrices play the same role as the sample covariance matrices in the stationary case. Suppose that there are r_1 common nonstationary factors $\mathbf{f}_{1,t}$, and r_2 common zero mean stationary factors $\mathbf{f}_{2,t}$. Let us divide the vectors of common factors and their noise as $\mathbf{f}'_t=(\mathbf{f}'_{1,t},\mathbf{f}'_{2,t})$ and $\mathbf{a}'_t=(\mathbf{a}'_{1,t},\mathbf{a}'_{2,t})$, respectively. The block diagonal variance matrix for $\mathbf{a}_t$ can also be partitioned as $\mathbf{\Sigma}_a=\begin{bmatrix}\mathbf{\Sigma}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_2\end{bmatrix}$ and the factor loading matrix as $\mathbf{P}=[\mathbf{P}_1\mathbf{P}_2]$. Then:

Assumption 1. The r_1 common nonstationary factors are generated by

$$\begin{aligned}(1-B)^d \mathbf{f}_{1,t} &= \mathbf{d}_1 + \mathbf{u}_t,\\ \mathbf{u}_t &= \mathbf{\Psi}(B)\mathbf{a}_{1,t},\end{aligned} \tag{4}$$

where d is a positive integer, $\mathbf{d}_1$ is a $r_1 \times 1$ vector of constants or drifts, $\mathrm{E}(\mathbf{a}_{1,t}) = \mathbf{0}$, $\mathrm{var}(\mathbf{a}_{1,t}) = \mathbf{\Sigma}_1 > 0$, $\mathbf{f}_{1,-(d-1)} = \mathbf{f}_{1,-(d-2)} = \cdots = \mathbf{f}_{1,0} = \mathbf{0}$, $\sum i \|\mathbf{\Psi}_i\| < \infty$ and $\|\mathbf{M}\| = [\mathrm{tr}(\mathbf{M}'\mathbf{M})]^{1/2}$ for any matrix or vector $\mathbf{M}$. Define $\mathbf{\Psi}(1) = \sum_{i=0}^{\infty} \mathbf{\Psi}_i$ where $\mathrm{rank}(\mathbf{\Psi}(1)) = r_1$.

Theorem 1. *For the nonstationary factor model given by* (1), (2) *and* (4) *with* $\mathbf{d}_1 = \mathbf{0}$ *and defining* $\mathbf{C}_y(k)$ *as in* (3) *with* $d' = 0$, *for* $k = 0, 1, \ldots, K$, *where K is small enough compared to the sample size T, so that when* $T \to \infty$, $\frac{K}{T} \to 0$:

(i) *The generalized sample covariance matrices*, $\mathbf{C}_y(k)$, *converge weakly to a random matrix* $\mathbf{\Gamma}_y$, *for* $k = 0, 1, \ldots, K$, *where limits are taken as T goes to infinity and* $\mathbf{\Gamma}_y$ *is defined as*

$$\mathbf{\Gamma}_y = \mathbf{P}_1 \mathbf{\Psi}(1) \mathbf{\Sigma}_1^{1/2} \left(\int_0^1 \mathbf{V}_{d-1}(\tau) \mathbf{V}_{d-1}(\tau)' \, \mathrm{d}\tau \right) (\mathbf{\Sigma}_1^{1/2})' \mathbf{\Psi}(1)' \mathbf{P}_1', \tag{5}$$

where $\mathbf{V}_d(\tau) = \mathbf{F}_d(\tau) - \int_0^1 \mathbf{F}_d(\tau)\,\mathrm{d}\tau$, $\mathbf{F}_d(\tau)$ *is the d* times integrated Brownian motion, *and it is defined recursively by* $\mathbf{F}_d(\tau) = \int_0^\tau \mathbf{F}_{d-1}(s)\,\mathrm{d}s$, *for* $d = 1, 2, \ldots$ *with* $\mathbf{F}_0(\tau) = \mathbf{W}(\tau)$, *the* r_1*-dimensional standard Brownian motion.*

(ii) $\mathbf{\Gamma}_y$ *has* r_1 *eigenvalues greater than zero almost sure and* $m - r_1$ *equal to zero.*

(iii) *The eigenvectors corresponding to the* r_1 *eigenvalues of* $\mathbf{\Gamma}_y$ *greater than zero are a basis of the space spanned by the columns of the loading submatrix* $\mathbf{P}_1$.

Proof. See the appendix. □

Remarks. (1) Empirically, the number of common nonstationary factors could be found as the number of eigenvalues of $\mathbf{C}_y(k)$ that converge weakly to the r_1 nonzero eigenvalues of $\mathbf{\Gamma}_y$, since $\mathbf{C}_y(k) \Rightarrow \mathbf{\Gamma}_y$ and the ordered eigenvalues are continuous functions of the coefficient matrix (Lemma 2 of Anderson et al. (1983)), and applying the continuous mapping theorem, the ordered eigenvalues of $\mathbf{C}_y(k)$ converge weakly to those of $\mathbf{\Gamma}_y$. We will give a standard chi-square test for the number of common factors, stationary or not, in the next section.

(2) Similar results are found if we use generalized sample second moment matrices, $\mathbf{A}_y(k) = \frac{1}{T^{2d}} \sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t'$, instead of generalized covariance matrices. In this case $\mathbf{A}_y(k) \Rightarrow \mathbf{P}_1 \mathbf{\Psi}(1) \mathbf{\Sigma}_1^{1/2} \int \mathbf{F}_{d-1}(r) \mathbf{F}_{d-1}(r)' \,\mathrm{d}r (\mathbf{\Sigma}_1^{1/2})' \mathbf{\Psi}(1)' \mathbf{P}_1'$.

(3) The results in Phillips and Durlauf (1986) and Chan and Wei (1988) apply to processes that satisfy more general assumptions of the innovations than the ones needed here. In particular, these results can be generalized to the case where the innovations present certain degree of heteroskedasticity.

(4) The theorem can be extended in a straightforward way to other models. For example, consider the model $y_t^j = f_t + e_t^j$ $(j = 1, 2, \ldots, m)$ in which f_t is a random walk with, for simplicity, unit residual variance ($f_t = f_{t-1} + a_t$; $\mathrm{var}(a_t) = 1$) and the variance of

the matrix of the serially uncorrelated process $\mathbf{e}_t = (e_t^1, e_t^2, \ldots, e_t^m)'$ is *nondiagonal*. It can be easily shown (in a similar way to the demonstration of the theorem given) that $\frac{1}{T^2}\sum_{t=k+1}^{T}(\mathbf{y}_{t-k}-\overline{\mathbf{y}})(\mathbf{y}_t-\overline{\mathbf{y}})' \Rightarrow \mathbf{1}\left(\int_0^1 V_0^2(r)\,\mathrm{d}r\right)\mathbf{1}'$, where $V_0(s) = W(s) - \int_0^1 W(\tau)\,\mathrm{d}\tau$, $W(s)$ is the scalar Brownian motion and $\mathbf{1}=(1, 1, \ldots, 1)'$, since all the stationary terms and the cross products between a stationary and a nonstationary term converge in probability to zero. The only relevant part at this point are the nonstationary terms. For the same reason, this result can be extended to the case where e_t^j $(j = 1, 2, \ldots, m)$ exhibits stationary serial correlation.

(5) We have assumed the same order of integration for all the series. This assumption can be easily relaxed. For instance, consider the following trivariate model, that we will call *example* 1, $y_t^j = f_t^1 + e_t^j\,(j = 1, 2)$ where $(1-B)^2 f_t^1 = a_t^1$ (assume also for simplicity that $\mathrm{var}(a_t^1) = 1$) and $y_t^3 = f_t^2 + e_t^3$ where f_t^2 is a random walk also with unit residual variance $(f_t^2 = f_{t-1}^2 + a_t^2;\ \mathrm{var}(a_t^2) = 1)$. Obviously the first column of the factor loading matrix is the vector $\mathbf{p}_1 = (1, 1, 0)'$. It can be shown in a similar way as it is done to prove Theorem 1 that the generalized covariance matrices of $\mathbf{y}_t$, that is, $\frac{1}{T^4}\sum_{t=k+1}^{T}(\mathbf{y}_{t-k}-\overline{\mathbf{y}})(\mathbf{y}_t-\overline{\mathbf{y}})' \Rightarrow \mathbf{p}_1\left(\int_0^1 V_1^2(\tau)\,\mathrm{d}\tau\right)\mathbf{p}_1'$ with $V_1(\tau)$ now a scalar process.

The next theorem describes the convergence results when the stochastic process of the common factors has drifts.

Theorem 2. *For the nonstationary factor model given by* (1), (2) *and* (4) *with* $\mathbf{d}_1 \neq \mathbf{0}$ *and defining* $\mathbf{C}_y(k)$ *as in* (3) *with* $d' = 1$, *for* $k = 0, 1, \ldots, K$, *such that when* $T \to \infty$, $\frac{K}{T} \to 0$,

$$\mathbf{C}_y(k) \xrightarrow{\mathrm{p}} q\mathbf{P}_1\mathbf{d}_1\mathbf{d}_1'\mathbf{P}_1' \tag{6}$$

where q is a constant that depends on d.

Proof. See the appendix. □

Remark. Theorems 1and 2 prove that the asymptotic behavior of the generalized sample second moment matrices depend on whether or not the stochastic process of the common factors has a drift. In this last case, these matrices are $\mathrm{O_p}(T^{2d})$ and converge weakly to matrices of rank r_1. With a drift different from zero, the deterministic part dominates the convergence results, these matrices are $\mathrm{O}(T^{2d+1})$ and converge to matrices of rank 1.

4. A chi-square tests for the number of factors

We have seen in Section 3 that when $\mathbf{d}_1 = \mathbf{0}$, the sample second moment matrices, conveniently normalized, converge to random matrices of rank r_1 in the nonstationary case. From Peña and Box (1987) we already know that in the stationary case the population autocovariance matrices are of rank $r = r_2$. So the next step will be to test the rank of these matrices. This task has been accomplished previously in the literature. In fact, Tiao and Tsay (1989) also check the rank of some moment matrices for both the stationary and nonstationary cases using a canonical correlation analysis and a chi-square statistic.

We will use canonical correlation analysis to detect the rank of the sample lagged second moment matrices. To fix ideas, consider first the case of zero mean stationary time series and let $\mathbf{\Gamma}_y(k) = E(\mathbf{y}_{t-k}\mathbf{y}'_t)$, $m \times m$, and $\mathbf{\Gamma}_f(k) = E(\mathbf{f}_{t-k}\mathbf{f}'_t)$, $r \times r$, be the lagged k second moment matrices of $\mathbf{y}_t$ and $\mathbf{f}_t$. The relation between them is

$$\mathbf{\Gamma}_y(k) = \mathbf{P}\mathbf{\Gamma}_f(k)\mathbf{P}' \tag{7}$$

for all $k \neq 0$. Throughout this section we will use the identification restriction $\mathbf{P}'\mathbf{P} = \mathbf{I}$, since it will simplify all the proofs. Therefore, $\operatorname{rank}(\mathbf{\Gamma}_y(k)) = \operatorname{rank}(\mathbf{\Gamma}_f(k)) = r$. Since relation (7) is true for all $k \neq 0$, there exists a $m \times (m-r)$ matrix $\mathbf{P}_\perp$, such that for all $k \neq 0$,

$$\mathbf{\Gamma}_y(k)\mathbf{P}_\perp = \mathbf{P}\mathbf{\Gamma}_f(k)\mathbf{P}'\mathbf{P}_\perp = 0. \tag{8}$$

For instance, $\mathbf{P}_\perp$ can be built by using the $m-r$ linearly independent eigenvectors associated to the $m-r$ zero eigenvalues of $\mathbf{\Gamma}_y(k)$. The condition in (8) also implies that the $m-r$ independent linear combinations of the observed series given by $\mathbf{P}'_\perp\mathbf{y}_t$ are cross and serially uncorrelated for all lags $k \neq 0$. Also these linear combinations will be uncorrelated to $\mathbf{P}'_\perp\mathbf{y}_{t-k}$. Consider now the $m \times m$ matrix

$$\mathbf{M}(k) = [E(\mathbf{y}_t\mathbf{y}'_t)]^{-1}E(\mathbf{y}_t\mathbf{y}'_{t-k})[E(\mathbf{y}_{t-k}\mathbf{y}'_{t-k})]^{-1}E(\mathbf{y}_{t-k}\mathbf{y}'_t). \tag{9}$$

The number of zero canonical correlations between $\mathbf{y}_{t-k}$ and $\mathbf{y}_t$ is given by the number of zero eigenvalues of the matrix defined in (9) and since $\operatorname{rank}(\mathbf{M}(k)) = \operatorname{rank}(\mathbf{\Gamma}_y(k)) = r$, this number is $m-r$. Thus, the number of common factors, r, is equivalent to the number of nonzero canonical correlations between $\mathbf{y}_{t-k}$ and $\mathbf{y}_t$.

Consider now the finite sample case in which T observations are available. The squared sample canonical correlations between $\mathbf{y}_{t-k}$ and $\mathbf{y}_t$ are the eigenvalues of

$$\widehat{\mathbf{M}}_1(k) = \left[\sum_{t=k+1}^{T}(\mathbf{y}_t\mathbf{y}'_t)\right]^{-1}\sum_{t=k+1}^{T}(\mathbf{y}_t\mathbf{y}'_{t-k})\left[\sum_{t=k+1}^{T}(\mathbf{y}_{t-k}\mathbf{y}'_{t-k})\right]^{-1}\sum_{t=k+1}^{T}(\mathbf{y}_{t-k}\mathbf{y}'_t). \tag{10}$$

We will see that under certain conditions, given $\mathbf{y}_t$ stationary or not, the limit of $\widehat{\mathbf{M}}_1(k)$ exists as the sample size T goes to infinity and that $\widehat{\mathbf{M}}_1(k)$ has $m-r$ eigenvalues that converge in probability to zero.

Theorem 3. *For the r common factors model presented in Section* 2, *with r_1 nonstationary common factors given by Assumption* 1, *let $\widehat{\mathbf{M}}_1(k)$ be defined as in* (10), *for $k = 0, 1, \ldots, K$, such that $K < T$. Then, the limit of $\widehat{\mathbf{M}}_1(k)$ exists as the sample size T goes to infinity for all $k = 0, 1, \ldots, K$, such that $K/T \to 0$. Moreover $\widehat{\mathbf{M}}_1(k)$ has $m-r$ eigenvalues that converge in probability to zero.*

Proof. See the appendix. □

Remarks. As for Theorem 1, the order of integration of the nonstationary factors does not need to be the same. Assume that the r_1 common nonstationary factors have orders of integration d_i, $i = 1, \ldots, r_1$, where d_i is a positive integer. Then using the matrix $\mathbf{D} =$

$\mathrm{diag}(\frac{1}{T^{d_1}}, \ldots, \frac{1}{T^{d_{r_1}}}, \frac{1}{T^{1/2}}, \ldots, \frac{1}{T^{1/2}})$ in the proof, it can be shown in a similar way that $\widehat{\mathbf{M}}_1(k)$ has $m-r$ eigenvalues that converge in probability to zero.

Based on this result we propose the following chi-square test for the number of factors.

Lemma 1. *Consider the factor model given by* (1) *and* (2) *with r common factors, that can be nonstationary, let* $\widehat{\lambda}_1 \leqslant \widehat{\lambda}_2 \leqslant \cdots \leqslant \widehat{\lambda}_m$ *be the ordered eigenvalues of the matrix* $\widehat{\mathbf{M}}_1(k)$ *given by* (10). *Then, the statistic*

$$S_{m-r} = -(T-k)\sum_{j=1}^{m-r} \log(1-\widehat{\lambda}_j) \tag{11}$$

is asymptotically a $\chi^2_{(m-r)^2}$.

Proof. See the appendix. □

Remarks. (1) The limiting distribution is in agreement to what is found for the case of i.i.d observations. The result of this lemma is also in the line of Robinson (1973) to test for zero canonical correlation of stationary time series. This result was modified for Tiao and Tsay (1989) to test for scalar component models (SCM) by dividing each eigenvalue by the maximum possible variance that the sample cross correlation might have in the case of SCM. In our case, the variance of the cross correlation associated to white noise canonical variates is correctly specified as $1/(T-k)$ and we do not need to standardize all the eigenvalues. Johansen (1991) for the cointegration rank of a vector autoregressive (VAR) model and Reinsel and Ahn (1992) for the number of unit roots in reduced rank regression models use a similar statistic but with nonstandard asymptotic theory to test for zero canonical correlations between the levels of the variables and their first differences corrected both of them for serial correlation.

(2) The lag k used to perform the test is important. For instance, if the r common factors follow a moving average process of order q, MA(q), then the observed series follow a MA(q) process as well. Therefore, if we use $k=q+1$, we will find that the number of zero canonical correlations is greater than $m-r$. To overcome this difficulty one can perform the test for increasing $k=1, 2, \ldots$

The performance of the test is shown through the following set of simulations. From now on, we will denote by (m, r, p, q) a model for m observed series generated by r common factors which follow a vector ARMA(p, q) model. The first generating process we use is a factor model with three series, one common nonstationary factor $I(1)$ and no specific components. The system matrices are $\mathbf{P}' = (1, 1, 1)'$, $\mathbf{\Sigma}_e = \mathbf{I}_3$, the factor is a random walk and $\sigma_a^2 = 1$. We denote this model as (3,1,1,0). The second generating process has also three series and one common factor, but now the factor model is I(2), $(1-B)^2 f_t = a_t$.This model is denoted as (3,1,2,0). The third generating process consists of six series and three common nonstationary factors $I(1)$ so that the series are $I(1)$. The system matrices are $\mathbf{P} = [\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3]$ where $\mathbf{p}_1' = (1, 1, 0, 1, -1, 0)$, $\mathbf{p}_2' = (0, 1, 1, 0, 1, -1)$, $\mathbf{p}_3' = (1, 0, 0, 0, 1, 1)$; $\mathbf{\Sigma}_e = \mathbf{I}_6$ and the factors are autoregressive of order 1, AR(1), with $\mathbf{\Phi}_1 = \mathbf{I}_3$ and $\mathbf{\Sigma}_a = \mathbf{I}_3$. We denote this

Table 1
Number of times in which the hypothesis of r common factors is rejected for models (3,1,1,0) and (3,1,2,0)

Model	r	lag k				
		1	2	3	4	5
(3,1,1,0)	0	1000	1000	1000	1000	1000
	1	47	55	48	53	37
	2	6	5	4	1	6
(3,1,2,0)	0	1000	1000	1000	1000	1000
	1	47	61	47	44	50
	2	2	3	4	3	0

Table 2
Number of times the hypothesis of a maximum of r common factors is rejected for models (6,3,1,0) and (6,3,2,0)

Model	r	lag k				
		1	2	3	4	5
(6,3,1,0)	0	1000	1000	1000	1000	1000
	1	1000	1000	1000	1000	1000
	2	1000	1000	1000	999	998
	3	67	44	31	39	27
	4	6	0	0	0	1
	5	0	0	0	0	0
(6,3,2,0)	0	1000	1000	1000	1000	1000
	1	1000	1000	1000	1000	1000
	2	1000	1000	1000	999	993
	3	55	53	50	35	23
	4	1	2	2	0	3
	5	1	0	0	1	0

model as (6,3,1,0). Finally, the fourth generating process is a slight modification of the third one, being the second common factor the I(2) process $(1-B)^2 f_t = a_t$. We denote this model as (6,3,2,0). The number of replications in each simulation was 1000. In all three simulations, the sample size was $T=200$; we generated 1000 observations and discarded the first 800 to lessen the dependence of the initial conditions.

The performance of the proposed test for testing for r factors is shown in Table 1 for the first two models and in Table 2 for the last two ones. Testing that the series have r common factors is equivalent to testing the assumption that there are $m-r$ zero canonical correlations between $\mathbf{y}_t$ and $\mathbf{y}_{t-k}$, $i=1,\dots,m$. Note that if we accept that $m-r$ canonical correlation coefficients are equal to zero (r factors) we will always accept with this test that $m-r-1$ will be equal to zero ($r+1$ factors). Thus, the test should be applied sequentially as a test on the maximum number of factors and stop as soon as the hypothesis is rejected. In column 1 we describe the model used to generate the data by (m, r, p, q), column 2 includes the maximum number of factors tested, and the rest of the columns indicate the lag k used to perform the test. By rows, from the third column onwards, we give the number of times that the hypothesis of a maximum of r common factors was rejected at the 5% significance level using the chi-squared statistic of Lemma 1.

Table 1 shows that the hypothesis of zero common factors is always rejected, indicating a power of the test equal to 1 for both models. The proportion of times the hypothesis of one common factor is rejected goes from 3.7% to 5.5% for model (3,1,1,0), and from 4.4% to 6.1% from model (3,1,2,0), indicating that the empirical size of the test seems to be appropriate. The hypothesis of a maximum of two factors is almost always accepted, as expected.

Table 2 contains the same information for models (6,3,1,0) and (6,3,2,0). The table is read as Table 1.

The test seems to be very powerful to detect the number of factors and the size of the test obtained when $r = 3$ is close to the nominal 5%. Also, notice that the test can handle factors with different orders of integration without any modification since we seek for stationary linear combinations.

5. A proposed methodology

5.1. Identification of the number of factors

The procedure we propose will give us the number of stationary and nonstationary common factors, as well as a first estimation of the factor loading matrix and the time series of the common factors. We assume that the common factors do not have drifts for the reason given in Peña (1995). If that was not the case, one could subtract the deterministic part from the observed series and then apply the following methodology.

Step 1: Build the matrix $\widehat{\mathbf{M}}_1(k)$ in (10) for the variables $\mathbf{y}_t$ and $\mathbf{y}_{t-k}$, $k = 1, 2, \ldots, K$. Compute their eigenvalues and sort them in ascending order. Compute the test statistic defined in (11) for increasing r and compare it with a $\chi^2_{(m-r)^2}$ for a certain level of significance α. Reject a maximum of r common factors if the statistic is greater than the $\chi^2_{(m-r)^2}$ for the level of significance α.

Step 2: Compute the generalized covariance matrices for $\mathbf{y}_t$, $\mathbf{C}_y(k)$, $k = 1, 2, \ldots K$, and their eigenvalues and eigenvectors. Sort them in descending order. The number of common stable eigenvectors should be r. An initial estimation of the factor loading matrix $\widehat{\mathbf{P}}^0$ (the final estimation will be obtained by maximum likelihood) could be the r first eigenvectors of the generalized covariance matrix for lag $k = 1$ and the initial estimation of the common factors is given by $\widehat{\mathbf{f}}^0 = (\widehat{\mathbf{P}}^0)'\mathbf{y}_t$.

Step 3: Analyze the univariate time series of the common factors $\widehat{\mathbf{f}}^0$ to decide how many of them are nonstationary and their order of integration.

Remarks. (1) An alternative initial estimator of $\mathbf{P}$ could be the mean of the r eigenvectors associated to the r largest eigenvalues of the first K generalized covariance matrices. However, we have not observed a clear advantage with respect to the simplest method of taking the first r eigenvectors of the generalized covariance matrices for $k=1$.

(2) Sometimes, the common eigenvectors associated to the stationary factors may not appear clearly because they are associated to zero eigenvalues and they might not be extracted always in the same order. A more complicated alternative procedure to solve this problem is as follows. Let $r_2 = r - r_1$ the common eigenvectors of the matrices $\mathbf{\Gamma}_y(k)$, now associated to the common stationary factors and to zero eigenvalues of these matrices. The remaining $m-r$ eigenvectors are not common. Let $\mathbf{P}_1^{\perp}$ be a matrix of the null space of $\mathbf{P}_1$. The $m \times r_2$ matrix $\mathbf{P}_2$ is formed by r_2 columns of $\mathbf{P}_1^{\perp}$. To find out which ones, first we project $\mathbf{y}_t$ over $\mathbf{P}_1^{\perp\prime}\mathbf{P}_1^{\perp}$. This is the same as estimating the nonstationary factors through $\mathbf{P}_1'\mathbf{y}_t$ and subtracting them from the observed series. (If $\mathbf{P}'\mathbf{P} \neq \mathbf{I}$, this should be equivalent to estimating the nonstationary factors through $(\mathbf{P}_1'\mathbf{P}_1)^{-1}\mathbf{P}_1'\mathbf{y}_t$ and subtracting them from the observed series). Define a new set of variables $\mathbf{y}_t^*$ as

$$\mathbf{y}_t^* = \mathbf{y}_t - \mathbf{P}_1\mathbf{P}_1'\mathbf{y}_t. \tag{12}$$

The auxiliary variables $\mathbf{y}_t^*$ are also equal to $\mathbf{P}_2\mathbf{f}_{2,t} + \mathbf{e}_t^*$, where $\mathbf{e}_t^* = \mathbf{P}_1^{\perp\prime}\mathbf{P}_1^{\perp}\mathbf{e}_t$. Let $\mathbf{\Gamma}_{\mathbf{y}}^*(k)$, $m \times m$, be the lagged k covariance matrices of the new variables $\mathbf{y}_t^*$, and $\mathbf{\Gamma}_{f2}(k)$, $r_2 \times r_2$, the lagged k covariance matrix of the common stationary factors. The relation between them is

$$\mathbf{\Gamma}_{\mathbf{y}}^*(k) = \mathbf{P}_2\mathbf{\Gamma}_{f2}(k)\mathbf{P}_2' \quad \text{if } k \neq 0. \tag{13}$$

Extract the principal components of $\mathbf{\Gamma}_{\mathbf{y}}^*(k)$. We give a first estimation of $\mathbf{P}_2$, denoted by $\widehat{\mathbf{P}}_2^0$, as the first r_2 eigenvectors of the lagged covariance matrix for $k=1$. Finally, we give an estimation of the common stationary factors as $(\widehat{\mathbf{P}}_2^0)'\mathbf{y}_t$.

(3) We have assumed that the order of integration of all the series was the same. This assumption can be relaxed. If the series have different orders of integration, and we want to proceed as in Remark (2), first, we have to identify the number of nonstationary factors with the highest order of integration; then, project the series over the subspace spanned by these factors and subtract it from the original series defining a new set of variables which will be of a lower order of integration than the original ones; identify the new order of integration and apply the same technique. Proceed in a similar way until all the common factors have been extracted.

5.2. Estimation

Model estimation is carried out by maximum likelihood, writing the model in state space form. The vector of observable time series $\mathbf{y}_t$, is assumed to be generated by the measurement equation,

$$\mathbf{y}_t = \tilde{\mathbf{P}}\mathbf{z}_t + \boldsymbol{\epsilon},$$

where the state space vector has dimension s and $\mathrm{E}(\boldsymbol{\epsilon}_t)=0$, $\mathrm{E}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')=\mathbf{\Sigma}_\varepsilon$ and $\mathrm{E}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_\tau')=0$ if $t \neq \tau$. The state space vector $\mathbf{z}_t$ is driven by the transition equation,

$$\mathbf{z}_t = \mathbf{G}\mathbf{z}_{t-1} + \mathbf{u}_t$$

with $\mathrm{E}(\mathbf{u}_t) = 0$, $\mathrm{E}(\mathbf{u}_t\mathbf{u}_t') = \mathbf{\Sigma}_u$ and $\mathrm{E}(\mathbf{u}_t\mathbf{u}_\tau') = \mathbf{0}$ if $t \neq \tau$. Both noises, $\boldsymbol{\epsilon}_t$ and $\mathbf{u}_t$, are also uncorrelated for all lags, $\mathrm{E}(\boldsymbol{\epsilon}_t\mathbf{u}_\tau') = \mathbf{0}$ for all t and τ. To write an ARMA(p, q) model in state space form, a state vector of dimension $\max(p, q + 1)$ (e.g. Akaike, 1974; Gardner et al., 1980; Ansley and Kohn, 1983) gives a minimal representation with uncorrelated errors in the transition and measurement equations. The dimension of the state vector can be reduced if we allow for correlation between the error terms in the transition and measurement equations. In our case, the state vector, $\mathbf{z}_t$, will contain lagged and contemporaneous factors and will have dimension $R_1 + R_2$, where R_1 is the number of elements of the state vector associated to the common nonstationary factors, and R_2 are the components of the state vector linked to the stationary common factors. The matrix $\tilde{\mathbf{P}}$ typically contains the elements of $\mathbf{P}$ and zeros. For example if there is an AR(2) stationary factor $f_{2,t}$ that loads into the series through the $m \times 1$ vector $\mathbf{p}$, the state vector has two components associated to this factor ($f_{2,t}$ and $f_{2,t-1}$) and the columns of the $\tilde{\mathbf{P}}$ matrix are $\mathbf{p}$ and $\mathbf{0}_{m\times 1}$.

Estimation of the model by maximum likelihood can be done by the EM algorithm of Dempster et al. (1977) or through the scoring algorithm. Watson and Engle (1983) compared both algorithms for estimation of state space models. The estimation through the scoring algorithm is more tedious, but one can obtain as a by-product the standard error of all the estimates of the parameters in the model and it permits joint estimation in the case of specific components. The EM algorithm was also proposed for state space models by Shumway and Stoffer (1982) and Wu et al. (1996), among others. See also Koopman (1993), Harvey (1989), Shumway and Stoffer (2000) and Durbin and Koopman (2001) for application details of the EM algorithm. Further studies show how the precision of the parameter estimates can be assessed. For instance, Stoffer and Wall (1991) propose to use bootstrap techniques while Cavanaugh and Shumway (1996) give a recursive algorithm to compute the expected Fisher information matrix.

To run either of the algorithms we first proceed as follows: (i) Set the number of nonstationary and stationary common factors and $\mathbf{P}$ as it was discussed previously. (ii) Set $\mathbf{G}$ by writing the ARMA model for the factors as the transition equation of the state space model. (iii) Set the initial condition for the state vector as $\mathbf{z}_1 = \tilde{\mathbf{P}}^{-}\mathbf{y}_1$ where $\tilde{\mathbf{P}}^{-}$ is a generalized inverse of $\tilde{\mathbf{P}}$ and its covariance matrix as $\kappa\mathbf{I}$ with κ great enough. (See, for instance, Harvey, 1989 for details about the selection of κ). (iv) Set $\Sigma_\varepsilon = \mathbf{I}$ or any diagonal matrix. (v) In this paper, we will impose for identification purposes at the estimation stage that $\mathbf{\Sigma}_a = \mathbf{I}$ (this restriction excludes the case where a common factor is just a constant which is not analyzed here) and $p_{ij} = 0$, for $j > i$.

6. A real data example

Data consists of 238 observations, from June 1982 until March 2002, of four time series of Spanish interbank interest rates: one day r_1, three months r_{90}, six months r_{180}, and one year r_{365}, interest rates. Fig. 1 shows a graph of the series. We will identify the number of common factors and obtain an initial estimate for the factor loading matrix and the common factors using Steps 1 through 3 of the proposed automatic procedure.

Step 1: We apply the test of Section 4 for $k = 1$–5 and obtain the results shown in Table 3. The entries of the table are the values of the S statistic for each lag k. These values are

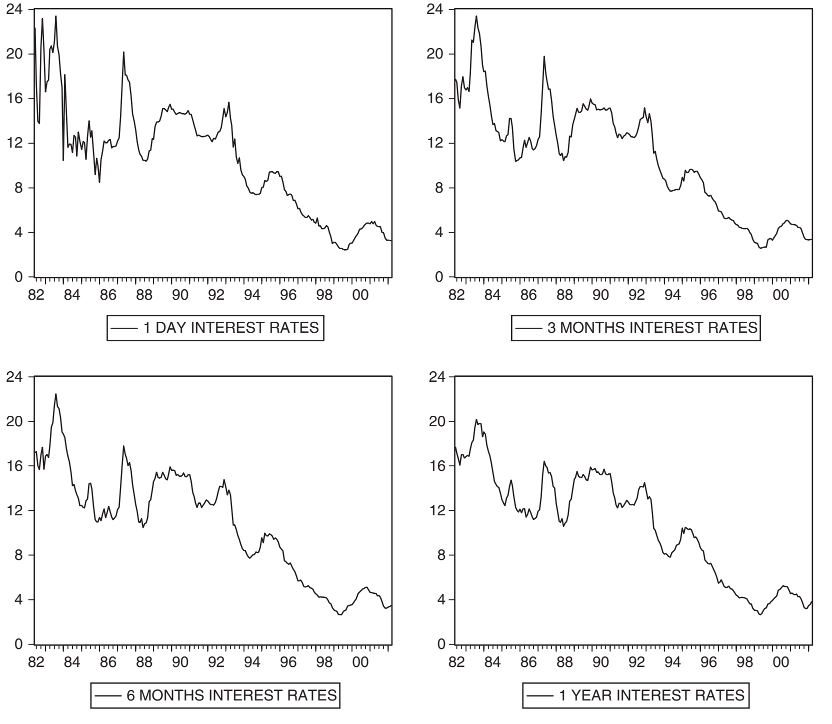

Fig. 1. Graphs of the four series of Spanish interbank interest rates.

Table 3
Outcome of the test of Section 4 for the number of factors

r	lag k				
	1	2	3	4	5
0	79.2*	61.2*	52.3*	47.5*	43.3*
1	29.8*	16.6*	11.3	9.7	7.8
2	9.13	4.9	4.9	7.0	4.4
3	0.3	0.2	0.3	0.7	0.2

An asterisk indicates that the null of a maximum of r common factors was rejected at the 5% significance level.

used to test a maximum of r common factors and should be compared to a χ^2 with $(m-r)^2$ degrees of freedom. We perform the test sequentially for increasing r, for $k = 1$–5. For any lag k we reject a maximum of zero common factors, therefore there is at least one common factor very persistent (probably nonstationary); a second factor appears for the first two lags indicating the possibility of a second common stationary factor (its autocorrelation dies faster than for the first factor). Therefore, the number of common factors is two, with the first one possible nonstationary and the second one possible stationary.

Table 4
First and second eigenvalues and associated eigenvectors for the first six generalized covariance matrices associated to the interest rates data

1st eigenvalue and eigenvector							2nd eigenvalue and eigenvector						
k	0	1	2	3	4	5	k	0	1	2	3	4	5
$10^2\widehat{\lambda}$	0.52	0.51	0.50	0.49	0.48	0.47	$10^4\widehat{\lambda}$	0.36	0.27	0.23	0.22	0.20	0.17
v	0.50	0.50	0.50	0.50	0.50	0.50	v	−0.78	−0.75	−0.77	−0.78	−0.80	−0.78
e	0.50	0.50	0.50	0.50	0.50	0.50	e	−0.03	−0.09	−0.09	−0.05	−0.04	−0.00
c	0.50	0.50	0.50	0.50	0.50	0.50	c	0.27	0.26	0.26	0.27	0.29	0.31
t	0.49	0.49	0.50	0.50	0.50	0.50	t	0.55	0.60	0.60	0.58	0.54	0.51

Step 2: To provide an initial estimator of the factor loading matrix and the common factors, we build the generalized covariance matrices. The first two eigenvalues and associated eigenvectors for the generalized lagged covariance matrices $\mathbf{C}_y(k)$, for $k = 0, 1, \ldots, 5$ and $d = 1$ are shown in Table 4. The second row of each matrix shows the lag considered, the third row shows the eigenvalue and rows 4–7 show each component of the corresponding eigenvector.

As it was pointed out in Section 5, we use as pre-estimation of the loading matrix, the eigenvectors associated to lag $k=1$. Our initial estimate of the factor loading matrix is $\widehat{\mathbf{P}}^0 = [\widehat{\mathbf{p}}_1^0 \widehat{\mathbf{p}}_2^0]$ where $(\widehat{\mathbf{p}}_1^0)' = (0.50, 0.50, 0.50, 0.49)$ and $(\widehat{\mathbf{p}}_2^0)' = (-0.75, -0.09, 0.26, 0.60)$. With this matrix we calculate a first estimation of the common factors as $\widehat{\mathbf{f}}^0 = (\widehat{\mathbf{P}}^0)'\widehat{\mathbf{y}}_t$.

Step 3: The analysis of the plots and the correlation structure of the factors time series shows that clearly the first one is nonstationary and there are some doubts about the second one. The outcome of the augmented Dickey–Fuller (ADF) unit root test for the pre-estimated common factors were -1.10 and -3.44, respectively. Compared with a critical value of -2.88 a unit root cannot be rejected at the usual 5% significance level for the first common factor.

We conclude that there is one common nonstationary factor and one common stationary factor. Then the EM algorithm is applied assuming as initial condition $\widehat{\mathbf{P}}^0$ for the factor loading matrix. For identification purposes we interchange series one and two and on the original matrix we restrict $p_{22} = 0$. The estimated model is given by $\mathbf{y}_t = \mathbf{P}\mathbf{f}_t + \mathbf{e}_t$ and $\mathbf{f}_t = \mathbf{\Phi}\mathbf{f}_{t-1} + \mathbf{a}_t$ with the following estimates for the system matrices $\widehat{\mathbf{P}} = [\widehat{\mathbf{p}}_1 \widehat{\mathbf{p}}_2]$, $\widehat{\mathbf{p}}_1 = (0.52, 0.52, 0.51, 0.48)$ and $\widehat{\mathbf{p}}_2 = (-0.16, 0, 0.14, 0.31)$; $\widehat{\mathbf{\Phi}} = \mathrm{diag}(1, 0.925)$, $\mathbf{\Sigma}_a = \mathbf{I}$ and $\widehat{\mathbf{\Sigma}}_e = \mathrm{diag}(0.9547, 0.0042, 0.0155, 0.0051)$.

Our results are similar to those given in Engle and Granger (1987), Stock and Watson (1988), Hall et al. (1992) and Reinsel and Ahn (1992) for USA interest rates who found one common nonstationary factor. Zhang (1993) found three common nonstationary factors in a bigger set of series. In this application, we find two common factors, one related to the general level of the series which is roughly the average of the four and is nonstationary, and a second one which measures the difference between the short and the long run and is stationary.

Appendix

Proof of Theorem 1. (From now on $\mathbf{0}$ is a vector or a matrix of appropriate dimensions.)

(i) Replacing $\mathbf{y}_t$, expressed as in (1), in Eq. (3) with $d'=0$, we have

$$\begin{aligned}\mathbf{C}_{\mathbf{y}}(k) &= \frac{1}{T^{2d}}\sum_{t=k+1}^{T}(\mathbf{y}_{t-k}-\bar{\mathbf{y}})(\mathbf{y}_t-\bar{\mathbf{y}})' \\ &= \mathbf{P}\left(\frac{1}{T^{2d}}\sum_{t=k+1}^{T}(\mathbf{f}_{t-k}-\bar{\mathbf{f}})(\mathbf{f}_t-\bar{\mathbf{f}})'\right)\mathbf{P}' + \mathbf{P}\frac{1}{T^{2d}}\sum_{t=k+1}^{T}(\mathbf{f}_{t-k}-\bar{\mathbf{f}})\mathbf{e}_t' \\ &\quad + \left(\frac{1}{T^{2d}}\sum_{t=k+1}^{T}\mathbf{e}_{t-k}(\mathbf{f}_t-\bar{\mathbf{f}})'\right)\mathbf{P}' + \frac{1}{T^{2d}}\sum_{t=k+1}^{T}\mathbf{e}_{t-k}\mathbf{e}_t' \\ &= \mathbf{P}_1\left(\frac{1}{T^{2d}}\sum_{t=k+1}^{T}(\mathbf{f}_{1,t-k}-\bar{\mathbf{f}}_1)(\mathbf{f}_{1,t}-\bar{\mathbf{f}}_1)'\right)\mathbf{P}_1' + \mathrm{o_p}(1), \end{aligned} \tag{A.1}$$

where $\bar{\mathbf{f}}_1 = 1/T\sum_{t=1}^{T}\mathbf{f}_{1,t}$. Since $\mathbf{e}_t$ is white noise, $\frac{1}{T^{2d}}\sum_{t=k+1}^{T}\mathbf{e}_{t-k}\mathbf{e}_t' \xrightarrow{\mathrm{p}} \mathbf{0}_{m\times m}$. In fact, all the cross products of stationary terms are $\mathrm{o_p}(1)$ since $d\geqslant 1$. For the nonstationary factors, it will be shown that $\frac{1}{T^{2d}}\sum_{t=k+1}^{T}(\mathbf{f}_{1,t-k}-\bar{\mathbf{f}}_1)\mathbf{e}_t' \xrightarrow{\mathrm{p}} \mathbf{0}$. Let $\sigma_{i,j} = \frac{1}{T^{2d}}\sum_{t=k+1}^{T}(f_{1,t-k}^i - \bar{f}_1^i)e_t^j$, for $i=1,\ldots,r_1$, $j=1,\ldots,m$, where $f_{j,t}^i$ stands for the ith component of vector $\mathbf{f}_{j,t}$, $j=1,2$, and e_t^i for the ith component of $\mathbf{e}_t$. It will be shown that $\sigma_{i,j}\xrightarrow{\mathrm{p}} 0$, for all $i=1,\ldots,r_1$ and $j=1,\ldots,m$. Each $\sigma_{i,j}$ is bounded from above since, calling $M\geqslant \max_{1\leqslant t\leqslant T}|e_t^j|$,

$$\begin{aligned}\frac{1}{T^{2d}}\sum_{t=k+1}^{T}(f_{1,t-k}^i-\bar{f}_1^i)\,e_t^j &\leqslant \frac{1}{T^{2d}}\sum_{t=k+1}^{T}(f_{1,t-k}^i-\bar{f}_1^i)|e_t^j| \\ &\leqslant \frac{1}{T^{2d}}\sum_{t=k+1}^{T}(f_{1,t-k}^i-\bar{f}_1^i)\max_{1\leqslant t\leqslant T}|e_t^j| \\ &\leqslant \frac{M}{T^{2d}}\sum_{t=k+1}^{T}(f_{1,t-k}^i-\bar{f}_1^i) \\ &= \frac{M}{T^{d-1/2}}\,\frac{1}{T^{d+1/2}}\,\frac{k}{T}\sum_{1}^{T}f_{1,t}^i. \end{aligned}$$

From Tanaka (1996), $\frac{1}{T^{d+1/2}}\sum_{t=1}^{T}f_{1,t}^i$ is $\mathrm{O_p}(1)$, therefore $\sigma_{i,j}\xrightarrow{\mathrm{p}} 0$, and $\frac{1}{T^{2d}}\sum(\mathbf{f}_{1,t-k}-\bar{\mathbf{f}}_1)\mathbf{e}_t \xrightarrow{\mathrm{p}} \mathbf{0}$. The proof of $\frac{1}{T^{2d}}\sum_{t=k+1}^{T}(\mathbf{f}_{1,t-k}-\bar{\mathbf{f}}_1)(\mathbf{f}_{2,t}-\bar{\mathbf{f}}_2)'\xrightarrow{\mathrm{p}}\mathbf{0}_{r_1\times r_2}$ goes as before, with $\mathbf{f}_{2,t}-\bar{\mathbf{f}}_2$ instead of $\mathbf{e}_t$ and $M\geqslant \max_{1\leqslant t\leqslant T}|f_{2,t}^j-\bar{f}_2^j|$.

From (4), and following the notation in Tanaka (1996, p. 99), the $I(d)$ factors, $\mathbf{f}_{1,t}$ can also be expressed as $\mathbf{f}_{1,t}=\mathbf{f}_{1,t}^{(d)}=\mathbf{f}_{1,t}^{(d-1)}+\mathbf{f}_{1,t-1}^{(d)}=\sum_{j=1}^{t}\mathbf{f}_{1,j}^{(d-1)}$, where $\{\mathbf{f}_{1,t}^{(d-1)}\}_{t=1}^{T}$

is an $I(d-1)$ process recursively defined in the same way, with $\mathbf{f}_{1,t}^{(0)}=\mathbf{u}_{1,t}$. With this notation,$\mathbf{f}_{1,t-k}=\mathbf{f}_{1,t}-\sum_{i=0}^{k-1}\mathbf{f}_{1,t-i}^{(d-1)}$, so

$$\frac{1}{T^{2d}}\sum_{t=k+1}^{T}(\mathbf{f}_{1,t-k}-\bar{\mathbf{f}}_1)(\mathbf{f}_{1,t}-\bar{\mathbf{f}}_1)'$$
$$=\frac{1}{T^{2d}}\sum_{t=k+1}^{T}(\mathbf{f}_{1,t}-\bar{\mathbf{f}}_1)(\mathbf{f}_{1,t}-\bar{\mathbf{f}}_1)'-\frac{1}{T^{2d}}\sum_{t=k+1}^{T}\left(\sum_{i=0}^{k-1}\mathbf{f}_{1,t-i}^{(d-1)}\right)(\mathbf{f}_{1,t}-\bar{\mathbf{f}}_1)'. \quad \text{(A.2)}$$

From Chan and Wei (1988) and Tanaka (1996, Sections 3.8 and 3.10), $\sum\mathbf{f}_{1,t-i}^{(d-1)}\mathbf{f}_{1,t}'$ is $\mathrm{O_p}(T^{2d-1})$ for finite i and i small enough compared to T; also

$$\frac{1}{T^{d+1/2}}\sum_{t=1}^{T}\mathbf{f}_{1,t}\Rightarrow\mathbf{\Psi}(1)\mathbf{\Sigma}_1^{1/2}\int_0^1\mathbf{F}_{d-1}(\tau)\,\mathrm{d}\tau,$$

$$\frac{1}{T^{2d}}\sum_{t=1}^{T}\mathbf{f}_{1,t}\mathbf{f}_{1,t}'\Rightarrow\mathbf{\Psi}(1)\mathbf{\Sigma}_1^{1/2}\int_0^1\mathbf{F}_{d-1}(\tau)\mathbf{F}_{d-1}(\tau)'\,\mathrm{d}\tau(\mathbf{\Sigma}_1^{1/2})'\mathbf{\Psi}(1)',$$

where $\mathbf{F}_d(\tau)$ is defined as in (i). By the continuous mapping theorem (Billingsley, 1968)

$$\frac{1}{T^{2d}}\sum_{t=1}^{T}(\mathbf{f}_{1,t}-\bar{\mathbf{f}}_1)(\mathbf{f}_{1,t}-\bar{\mathbf{f}}_1)'$$
$$\Rightarrow\mathbf{\Psi}(1)\mathbf{\Sigma}_1^{1/2}\int_0^1\mathbf{V}_{d-1}(\tau)\mathbf{V}_{d-1}(\tau)'\,\mathrm{d}\tau(\mathbf{\Sigma}_1^{1/2})'\mathbf{\Psi}(1)' \quad \text{(A.3)}$$

with $\mathbf{V}_d(\tau)=\mathbf{F}_d(\tau)-\int_0^1\mathbf{F}_d(\tau)d\tau$. Partitioning $\mathbf{P}$ as $[\mathbf{P}_1\mathbf{P}_2]$, where $\mathbf{P}_1$ is the $m\times r_1$ submatrix associated to the common nonstationary factors, and $\mathbf{P}_2$ is the $m\times r_2$ submatrix associated to the common stationary ones, and by the continuous mapping theorem (Billingsley, 1968)

$$\mathbf{C}_y(k)\Rightarrow\mathbf{P}_1\mathbf{\Psi}(1)\mathbf{\Sigma}_1^{1/2}\int_0^1\mathbf{V}_{d-1}(\tau)\mathbf{V}_{d-1}(\tau)'\,\mathrm{d}\tau(\mathbf{\Sigma}_1^{1/2})'\mathbf{\Psi}(1)'\mathbf{P}_1'=\mathbf{P}_1\mathbf{\Gamma}_{f_1}\mathbf{P}_1'=\mathbf{\Gamma}_y,$$

where $\mathbf{\Gamma}_{f_1}=\mathbf{\Psi}(1)\mathbf{\Sigma}_1^{1/2}\int_0^1\mathbf{V}_{d-1}(\tau)\mathbf{V}_{d-1}(\tau)'\,\mathrm{d}\tau(\mathbf{\Sigma}_1^{1/2})'\mathbf{\Psi}(1)'$. Notice that matrix $\mathbf{L}=\mathbf{\Sigma}_1^{1/2}(\int_0^1\mathbf{V}(r)\mathbf{V}(r)'\,\mathrm{d}r)(\mathbf{\Sigma}_1^{1/2})'$ is a nondiagonal matrix and that all generalized covariance matrices (for lag 0, as well as, for lag k, finite) have the same limiting distribution.

To prove part (ii), notice that (see, for instance, Hamilton, 1994, p. 546) for the nonstationary variables $\mathbf{f}_{1,t}$ defined in (4), the matrix $\mathbf{S}_T=\frac{1}{T^{2d}}\sum(\mathbf{f}_{1,t}-\bar{\mathbf{f}}_1)(\mathbf{f}_{1,t}-\bar{\mathbf{f}}_1)'\Rightarrow\mathbf{\Psi}(1)\mathbf{\Sigma}_1^{1/2}\int\mathbf{V}_{d-1}(r)\mathbf{V}_{d-1}(r)'\,\mathrm{d}r(\mathbf{\Sigma}_1^{1/2})'\mathbf{\Psi}(1)'=\mathbf{\Psi}(1)\mathbf{L}\mathbf{\Psi}(1)'$. The eigenvalues of the limiting sequence $\mathbf{S}_T$, are all greater than zero, since this is always a positive definite symmetric matrix as it can be easily seen using equality (A.4) given in Bellman (1960, p. 49), for $s=1,2,\ldots,r_1$, which it is used to show that all principal minors of $\mathbf{S}_T$ have determinants strictly greater than zero, so $\mathbf{S}_T$ is positive defined. Let x^i, $i=1,2,\ldots,s$ be the set of T-dimensional vectors, $T\geqslant s$, given by $x^i=\mathbf{f}_1^i-\bar{f}_1^i\mathbf{1}$, where $\mathbf{f}_1^i=(f_{1,1}^i,f_{1,2}^i,\ldots,f_{1,T}^i)'$

is the ith sample common nonstationary factor, of dimensions $T \times 1$, $\bar{f}_1^i = 1/T \sum_{j=1}^{T} f_{1,j}^i$ and $\mathbf{1}' = (1, \ldots, 1)$ is the $T \times 1$ vector of ones. Then,

$$|(x^i, x^j)|_{i,j=1,2,\ldots,s} = \frac{1}{s!} \sum_{\{i_s\}} \begin{vmatrix} x_{i_1}^1 & x_{i_2}^1 & \cdots & x_{i_s}^1 \\ x_{i_1}^2 & x_{i_2}^2 & \cdots & x_{i_s}^2 \\ \vdots & \vdots & & \vdots \\ x_{i_1}^s & x_{i_2}^s & \cdots & x_{i_s}^s \end{vmatrix}^2, \tag{A.4}$$

where $(x^i, x^j) = \sum_{k=1}^{T} x_k^i x_k^j$ is the i, j element of the left-hand side matrix of the above equation,whose determinant we are calculating and the sum goes over the whole set of integers, $\{i_s\}$, con $1 \leqslant i_1 \leqslant i_2 \leqslant \cdots \leqslant i_s \leqslant T$.

Now, we are going to show that $\mathbf{N} = \int_0^1 \mathbf{V}_{d-1}(\tau)\mathbf{V}_{d-1}(\tau)' \, \mathrm{d}\tau$ is nonsingular almost sure. Let

$$\Omega = \left\{ \begin{array}{c} w\colon \mathbf{W}(w, t) \text{ is the standard } r_1 \text{ dimensional Brownian motion} \\ \text{with continuos and nondifferentiable paths for } 0 \leqslant t \leqslant 1 \end{array} \right\}.$$

Since $\mathbf{W}(w, t)$ is the standard Brownian motion, $\nexists w \in \Omega$ and $\mathbf{a} = (a_1, \ldots, a_{r_1})' \neq \mathbf{0}$ such that $\mathbf{a}'\mathbf{W} = 0$, (recall that the variance–covariance matrix of $\mathbf{W}(w, t)$ is the identity). Also $P(\Omega) = 1$. Let $V_g^j(w, \tau)$ be the jth component of the process $\mathbf{V}_g(w, \tau)$, for $g = 0, 1, \ldots, d-1$, we will show that $\mathbf{N} = \mathbf{N}(w)$ is nonsingular for any $w \in \Omega$. If not, there exists $\mathbf{c} = (c_1, \ldots, c_{r_1})' \neq \mathbf{0}$ and $w \in \Omega$ such that $\mathbf{c}'\mathbf{N}\mathbf{c} = 0$, or $\int_0^1 \left(\sum_{j=1}^{r_1} c_j V_{d-1}^j(w, \tau) \right)^2 \mathrm{d}\tau$. Since $\sum_{j=1}^{r_1} c_j V_{d-1}^j(w, \tau)$ is a continuous function in τ it must happen for the previous integral to be equal to zero that $\sum_{j=1}^{r_1} c_j V_{d-1}^j(w, \tau) = 0$, for $0 \leqslant \tau \leqslant 1$. Since $\mathbf{c} \neq \mathbf{0}$, $\exists$ $c_i \neq 0$, such that $V_{d-1}^i(w, \tau) = \frac{1}{c_i} \sum_{j=1, j \neq i}^{r_1} c_j V_{d-1}^j(w, \tau)$. But for each realization of $\mathbf{V}_{d-1}(w, \tau) = \mathbf{F}_{d-1}(w, \tau) - \int_0^1 \mathbf{F}_{d-1}(w, \tau) \, \mathrm{d}\tau$; this means that

$$F_{d-1}^i(w, \tau) - \int_0^1 F_{d-1}^i(w, \tau) \, \mathrm{d}\tau = \frac{1}{c_i} \times \sum_{j=1, j \neq i}^{r_1} c_j \left(F_{d-1}^j(w, \tau) - \int_0^1 F_{d-1}^j(w, \tau) d\tau \right)$$

or

$$F_{d-1}^i(w, \tau) = \frac{1}{c_i} \sum_{j=1, j \neq i}^{r_1} c_j F_{d-1}^j(w, \tau) - H,$$

where $H = 1/c_i \sum_{j=1, j \neq i}^{r_1} c_j \int_0^1 F_{d-1}^j(w, \tau) \, \mathrm{d}\tau - \int_0^1 F_{d-1}^i(w, \tau) \, \mathrm{d}\tau$. Now the technique used in the proof of Lemma 3.1.1 of Chan and Wei (1988) can be employed to prove the result. Notice that differentiating $d-1$ times H is equal to zero. Therefore $\mathbf{N}$ is nonsingular almost sure and by (5) has r_1 strictly positive eigenvalues almost sure and $m - r_1$ equal zero.

(iii) The spectral decomposition of matrix $\mathbf{L}$ for each realization, leads to $\mathbf{L}=\mathbf{B}\mathbf{\Lambda}\mathbf{B}'$, where $\mathbf{B}$ is orthogonal. Then $\mathbf{\Gamma}_y = \mathbf{A}\mathbf{\Lambda}\mathbf{A}'$, where $\mathbf{A} = \mathbf{P}_1\mathbf{B}$, $\mathbf{A}'\mathbf{A} = \mathbf{I}$ and $\mathbf{\Lambda}$ has its r_1 eigenvalues different from zero. Therefore, the number of zero eigenvalues of $\mathbf{\Gamma}_y$ is $m - r_1$. The columns of $\mathbf{A}$ are eigenvectors of $\mathbf{\Gamma}_y$ and they span the same linear subspace as the columns of $\mathbf{P}_1$. □

Proof of Theorem 2. Again, following the notation in Tanaka (1996, p. 99), we can write

$$\mathbf{f}_{1,t} = \mathbf{f}_{1,t}^{(d)} = \mathbf{f}_{1,0}^{(d)} + \sum_{h=1}^{d-1} \mathbf{f}_{1,0}^{(d-h)} \sum_{\tau=1}^{t} \tau^{h-1} + \mathbf{c}_1 \sum_{\tau=1}^{t} \tau^{d-1} + \sum_{i_1=1}^{t-1} \sum_{i_2=1}^{t-i_1-1} \cdots \sum_{i_d=1}^{t-i_1-i_2-\cdots-i_{d-1}-1} \mathbf{f}_{1,t-i_1-i_2-\cdots-i_d}^{(0)}. \tag{A.5}$$

Therefore $\mathbf{f}_{1,t} = \mathrm{O_p}(\sum_{\tau=1}^{t} \tau^{d-1})$. Notice that $\sum_{\tau=1}^{t} \tau^{d-1} = \mathrm{O}(t^d)$. In a similar way to (A.1) we can prove that

$$\mathbf{C}_{\mathbf{y}}(k) = \mathbf{P}_1 \left(\frac{1}{T^{2d+1}} \sum_{t=k+1}^{T} (\mathbf{f}_{1,t-k} - \bar{\mathbf{f}}_1)(\mathbf{f}_{1,t} - \bar{\mathbf{f}}_1)' \right) \mathbf{P}_1' + \mathrm{o_p}(1). \tag{A.6}$$

And analogously to (A.2)

$$\begin{aligned} &\frac{1}{T^{2d+1}} \sum_{t=k+1}^{T} (\mathbf{f}_{1,t-k} - \bar{\mathbf{f}}_1)(\mathbf{f}_{1,t} - \bar{\mathbf{f}}_1)' \\ &\quad = \frac{1}{T^{2d+1}} \sum_{t=k+1}^{T} (\mathbf{f}_{1,t} - \bar{\mathbf{f}}_1)(\mathbf{f}_{1,t} - \bar{\mathbf{f}}_1)' \\ &\qquad - \frac{1}{T^{2d+1}} \sum_{t=k+1}^{T} \left(\sum_{i=0}^{k-1} \mathbf{f}_{1,t-i}^{(d-1)} \right) (\mathbf{f}_{1,t} - \bar{\mathbf{f}}_1)'. \end{aligned} \tag{A.7}$$

Notice that the stochastic part in (A.7) is $\mathrm{O_p}(T^{2d})$ by Theorem 1; and by (A.5) $\bar{\mathbf{f}}_1 = \frac{1}{T}\sum_{t=1}^{T} \mathbf{f}_{1,t} = \mathrm{O_p}(T^d)$ and $\mathbf{f}_{1,t-i}^{(d-1)}$ is $\mathrm{O_p}(T^{d-1})$ for finite i and i small enough compared to T. Therefore $\sum_{t=k+1}^{T} (\mathbf{f}_{1,t} - \bar{\mathbf{f}}_1)(\mathbf{f}_{1,t} - \bar{\mathbf{f}}_1)'$ is $\mathrm{O_p}(T^{2d+1})$. Moreover, by (A.5) and (A.7)

$$\begin{aligned} \mathbf{C}_{\mathbf{y}}(k) &= \mathbf{P}_1 \left(\frac{1}{T^{2d+1}} \sum_{t=k+1}^{T} (\mathbf{f}_{1,t-k} - \bar{\mathbf{f}}_1)(\mathbf{f}_{1,t} - \bar{\mathbf{f}}_1)' \right) \mathbf{P}_1' + \mathrm{o_p}(1) \\ &\overset{\mathrm{p}}{\to} q\mathbf{P}_1\mathbf{d}_1\mathbf{d}_1'\mathbf{P}_1', \end{aligned}$$

where q is a constant that depends on d. □

Proof of Theorem 3. Define $\widehat{\phi}^{+}(k) = [\sum_{t=k+1}^{T}(\mathbf{y}_t\mathbf{y}_t')]^{-1} \sum_{t=k+1}^{T}(\mathbf{y}_t\mathbf{y}_{t-k}')$ and $\widehat{\phi}(k) = [\sum_{t=k+1}^{T}(\mathbf{y}_{t-k}\mathbf{y}_{t-k}')]^{-1} \sum_{t=k+1}^{T}(\mathbf{y}_{t-k}\mathbf{y}_t')$, and denote their limits (if they exist) when T goes to infinity as $\phi^+(k)$ and $\phi(k)$, respectively. Therefore $\widehat{\mathbf{M}}_1(k) = \widehat{\phi}^+(k)\widehat{\phi}(k)$. Define also $\mathbf{z}_t = \mathbf{D}\mathbf{Q}'\mathbf{y}_t$, where $\mathbf{Q} = [\mathbf{P}\ \mathbf{P}_\perp]$ is such that $\mathbf{P}'\mathbf{P}_\perp = \mathbf{0}$ and $\mathbf{P}'_\perp\mathbf{P}_\perp = \mathbf{I}_{m-r}$, and $\mathbf{D}$ as

$$\mathbf{D} = \begin{bmatrix} \frac{1}{T^d}\mathbf{I}_{r_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{T^{1/2}}\mathbf{I}_{r_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{T^{1/2}}\mathbf{I}_{m-r} \end{bmatrix}. \tag{A.8}$$

Notice that $\mathbf{Q}' = \mathbf{Q}^{-1}$ and that $\widehat{\mathbf{M}}_1(k)$ can also be written as

$$\widehat{\mathbf{M}}_1(k) = \mathbf{Q}\mathbf{D}\left[\sum_{t=k+1}^{T}(\mathbf{z}_t\mathbf{z}'_t)\right]^{-1}\sum_{t=k+1}^{T}(\mathbf{z}_t\mathbf{z}'_{t-k})\left[\sum_{t=k+1}^{T}(\mathbf{z}_{t-k}\mathbf{z}'_{t-k})\right]^{-1} \times \sum_{t=k+1}^{T}(\mathbf{z}_{t-k}\mathbf{z}'_t)\mathbf{D}^{-1}\mathbf{Q}'.$$

In this appendix, it is shown that for $\mathbf{d}_1 = \mathbf{0}$, $\widehat{\mathbf{M}}_1(k)$ converges to a matrix that has $m-r$ eigenvalues equal to zero. First, we will find the limits of $\sum_{t=k+1}^{T}(\mathbf{z}_t\mathbf{z}'_t)$, $\sum_{t=k+1}^{T}(\mathbf{z}_{t-k}\mathbf{z}'_{t-k})$, $\sum_{t=k+1}^{T}(\mathbf{z}_t\mathbf{z}'_{t-k})$ and $\sum_{t=k+1}^{T}(\mathbf{z}_{t-k}\mathbf{z}'_t)$. Notice that $\mathbf{z}_t = \mathbf{D}\mathbf{Q}'\mathbf{y}_t$ and taking into account Eq. (1), then

$$\mathbf{z}_t = \begin{bmatrix} \frac{1}{T^d}(\mathbf{f}_{1,t} + \mathbf{P}'_1\mathbf{e}_t) \\ \frac{1}{T^{1/2}}(\mathbf{f}_{2,t} + \mathbf{P}'_2\mathbf{e}_t) \\ \frac{1}{T^{1/2}}\mathbf{P}'_\perp\mathbf{e}_t \end{bmatrix}$$

and therefore

$$\sum_{t=1}^{T}(\mathbf{z}_t\mathbf{z}'_t) = \begin{bmatrix} \frac{1}{T^{2d}}\sum(\mathbf{f}_{1,t}\mathbf{f}'_{1,t} + \mathbf{P}'_1\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_1) & \frac{1}{T^{d+1/2}}\sum(\mathbf{f}_{1,t}\mathbf{f}'_{2,t} + \mathbf{P}'_1\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_2) & \frac{1}{T^{d+1/2}}\sum(\mathbf{f}_{1,t}\mathbf{e}'_t\mathbf{P}_\perp + \mathbf{P}'_1\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_\perp) \\ \frac{1}{T^{d+1/2}}\sum(\mathbf{f}_{2,t}\mathbf{f}'_{1,t} + \mathbf{P}'_2\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_1) & \frac{1}{T}\sum(\mathbf{f}_{2,t}\mathbf{f}'_{2,t} + \mathbf{P}'_2\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_2) & \frac{1}{T}\sum(\mathbf{f}_{2,t}\mathbf{e}'_t\mathbf{P}_\perp + \mathbf{P}'_2\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_\perp) \\ \frac{1}{T^{d+1/2}}\sum(\mathbf{P}'_\perp\mathbf{e}_t\mathbf{f}'_{1t} + \mathbf{P}'_\perp\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_1) & \frac{1}{T}\sum(\mathbf{P}'_\perp\mathbf{e}_t\mathbf{f}'_{2t} + \mathbf{P}'_\perp\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_2) & \frac{1}{T}\sum\mathbf{P}'_\perp\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_\perp \end{bmatrix},$$

where all the summations in the previous matrix go from $t = 1$ to T. Now, it can be easily proven, as it was made in Theorem 1, from standard asymptotic results for stationary variables and the asymptotic results of Chan and Wei (1988), Phillips and Durlauf (1986), Tsay and Tiao (1990) and Tanaka (1996), that $\frac{1}{T}\sum_{t=1}^{T}\mathbf{P}'_\perp\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_\perp \xrightarrow{\text{p}} \mathbf{P}'_\perp\mathbf{\Sigma}_e\mathbf{P}_\perp$; $\frac{1}{T}\sum_{t=1}^{T}(\mathbf{f}_{2,t}\mathbf{f}'_{2,t} + \mathbf{P}'_2\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_2) \xrightarrow{\text{p}} E(\mathbf{f}_{2t}\mathbf{f}'_{2t}) + \mathbf{P}_2\mathbf{\Sigma}_e\mathbf{P}'_2$; $\frac{1}{T}\sum_{t=1}^{T}(\mathbf{P}'_\perp\mathbf{e}_t\mathbf{f}'_{2t} + \mathbf{P}'_\perp\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_2) \xrightarrow{\text{p}} \mathbf{P}'_\perp\mathbf{\Sigma}_e\mathbf{P}_2$; $\frac{1}{T^{2d}}\sum_{t=1}^{T}(\mathbf{f}_{1,t}\mathbf{f}'_{1,t} + \mathbf{P}'_1\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_1) \Rightarrow \mathbf{\Psi}(1)\mathbf{\Sigma}_1^{1/2}\int\mathbf{F}_{d-1}(r)\mathbf{F}_{d-1}(r)'\,\mathrm{d}r(\mathbf{\Sigma}_1^{1/2})'\mathbf{\Psi}(1)'$; $\frac{1}{T^{d+1/2}}\sum_{t=1}^{T}(\mathbf{f}_{1,t}\mathbf{f}'_{2,t} + \mathbf{P}'_1\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_2) \xrightarrow{\text{p}} \mathbf{0}$; and $\frac{1}{T^{d+1/2}}\sum_{t=1}^{T}(\mathbf{f}_{1,t}\mathbf{e}'_t\mathbf{P}_\perp + \mathbf{P}'_1\mathbf{e}_t\mathbf{e}'_t\mathbf{P}_\perp) \xrightarrow{\text{p}} \mathbf{0}$.

In a similar way, it can be proven that $\sum_{t=k+1}^{T}(\mathbf{z}_{t-k}\mathbf{z}'_{t-k})$ has the same limiting distribution as $\sum_{t=k+1}^{T}(\mathbf{z}_t\mathbf{z}'_t)$ and that

$$\sum_{t=k+1}^{T}(\mathbf{z}_t\mathbf{z}'_{t-k}) \Rightarrow \begin{bmatrix} \mathbf{\Psi}(1)\mathbf{\Sigma}_1^{1/2}\int\mathbf{F}_{d-1}(r)\mathbf{F}_{d-1}(r)'\,\mathrm{d}r(\mathbf{\Sigma}_1^{1/2})'\mathbf{\Psi}(1)' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & E(\mathbf{f}_{2,t}\mathbf{f}'_{2,t-k}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{(m-r)\times(m-r)} \end{bmatrix}$$

and

$$\sum_{t=k+1}^{T}(\mathbf{z}_{t-k}\mathbf{z}_t') \Rightarrow \begin{bmatrix} \mathbf{\Psi}(1)\mathbf{\Sigma}_1^{1/2}\int \mathbf{F}_{d-1}(r)\mathbf{F}_{d-1}(r)'\,\mathrm{d}r(\mathbf{\Sigma}_1^{1/2})'\mathbf{\Psi}(1)' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & E(\mathbf{f}_{2,t-k}\mathbf{f}_{2,t}') & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{(m-r)\times(m-r)} \end{bmatrix}.$$

Therefore from the continuous mapping theorem $\widehat{\mathbf{M}}_1(k)$ converges weakly to a random matrix that has $m-r$ eigenvalues equal to zero. □

Proof of Lemma 1. Since there exist $m-r$ linear combinations of the observed series that are white noise, it means that there are at least $m-r$ zero canonical correlations between $\mathbf{y}_{t-k}$ and $\mathbf{y}_t$. These canonical variates can be estimated as $\mathbf{v}_j'\mathbf{y}_t$ and $\mathbf{w}_j'\mathbf{y}_{t-k}$ where $\mathbf{v}_j$ and $\mathbf{w}_j$, $j=1,\ldots,m-r$, are the eigenvectors associated to the smallest eigenvalues of the matrices $\widehat{\mathbf{M}}_1(k)$ and

$$\widehat{\mathbf{M}}_2(k)=\left[\sum\nolimits_{t=k+1}^{T}(\mathbf{y}_{t-k}\mathbf{y}_{t-k}')\right]^{-1}\sum_{t=k+1}^{T}(\mathbf{y}_{t-k}\mathbf{y}_t')\left[\sum\nolimits_{t=k+1}^{T}(\mathbf{y}_t\mathbf{y}_t')\right]^{-1}\sum\nolimits_{t=k+1}^{T}(\mathbf{y}_t\mathbf{y}_{t-k}'),$$

respectively.

Notice that $\log(1-\widehat{\lambda}_j)\simeq-\widehat{\lambda}_j$ for small and positive $\widehat{\lambda}_j$ and $\widehat{\lambda}_j$ is the squared of the sample cross correlation between two canonical variates. The limit distribution of the sample cross correlations is jointly normal; each cross correlation has asymptotic variance $(T-k)^{-1}$; therefore $(T-k)\widehat{\lambda}_j$ is asymptotically a chi square where $T-k$ is the number of observations used to compute the cross correlation between the two canonical variates.

The number of degrees of freedom is computed as follows. Denote by $\mathbf{V}_j$ and $\mathbf{W}_j$, $j=1,\ldots,m-r$, the population vectors estimated by the sample vectors $\mathbf{v}_j$ and $\mathbf{w}_j$, respectively. Testing that 0 is an eigenvalue of multiplicity $m-r$ is equivalent to testing that there are $m-r$ linear combinations of $\mathbf{y}_t$ uncorrelated with $m-r$ linear combinations of $\mathbf{y}_{t-k}$. That is, there are $m-r$ regressions such that $\mathbf{W}_j=\mathbf{0}_{m\times 1}$ for $j=1,\ldots,m-r$ in the equation

$$\mathbf{V}_j'\mathbf{y}_t=\mathbf{W}_j'\mathbf{y}_{t-k}+\mathbf{u}_t, \tag{A.9}$$

which gives a total of $m(m-r)$ zero restrictions. Notice now that $\mathbf{V}_j$ has to be estimated and that the subspace associated to the eigenvectors linked to the zero eigenvalues of $\mathbf{M}_1(k)$ is of dimension $m-r$ by hypothesis. This means that in each of the vectors $\mathbf{V}_i$, $i=1,\ldots,m-r$ there are r restrictions among its components. Since they are estimated from cross products of the data, there are $(m-r)r$ restrictions in the cross products and therefore the number of degrees of freedom of the chi-squared is $m(m-r)-(m-r)r=(m-r)^2$. □

Acknowledgements

We are very thankful to Professors Ruey S. Tsay and Juan Romo for very helpful comments. We also acknowledge financial support from Cátedra BBVA de Calidad and DGES Projects BEC2000-0167 and BEC2002-00081.

References

Aguilar, O., West, M., 2000. Bayesian dynamic factor models and portfolio allocation. J. Business Econ. Statist. 18, 338–356.

Ahn, S.K., 1997. Inference of vector autoregressive models with cointegration and scalar components. J. Amer. Statist. Assoc. 437, 350–356.

Akaike, H., 1974. Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. Ann. Inst. Statist. Math. 26, 363–387.

Anderson, T.W., 1963. The use of factor analysis in the statistical analysis of multiple time series. Psychometrika 28, 1–25.

Anderson, S.A., Bröns, H.K., Jensen, S.T., 1983. Distribution of eigenvalues in Multivariate Statistical Analysis. Ann. Statist. 11, 392–415.

Ansley, S.K., Kohn, R., 1983. Exact likelihood of vector autoregressive-moving average process with missing or aggregated data. Biometrika 70, 275–278.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–222.

Bellman, R., 1960. Introduction to Matrix Analysis, McGraw-Hill, New York.

Billingsley, P., 1968. Convergence of Probability Measures, Wiley, New York.

Box, G., Tiao, G., 1977. A canonical analysis of multiple time series. Biometrika 64, 355–365.

Brillinger, D.R., 1981. Time series Data Analysis and Theory, expanded ed. Holden-Day, San Francisco.

Cavanaugh, J.E., Shumway, R.H., 1996. On computing the expected Fisher information matrix for state-space model parameters. Statist. Probab. Lett. 26, 347–355.

Chan, N.H., Wei, C.Z., 1988. Limiting distributions of least squares estimates of unstable autoregressive processes. Ann. Statist. 16, 367.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. B 39, 1–38.

Durbin, J., Koopman, S.J., 2001. Time series analysis by state space methods, Oxford University Press, Oxford.

Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: Representation, estimation and testing. Econometrica 55, 251–276.

Escribano, A., Peña, D., 1994. Cointegration and common factors. J. Time Ser. Anal. 15, 577–586.

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

Gardner, G., Harvey, A.C., Phillips, G.D.A., 1980. An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. Appl. Statist. 29, 311–322.

Geweke, J.F., Singleton, K.J., 1981. Maximum likelihood confirmatory analysis of economic time series. Internat. Econ. Rev. 22, 37–54.

Hall, A.D., Anderson, H.M., Granger, C.W.J., 1992. A cointegration analysis of treasury bill yields. Rev. Econ. Statist. 74, 117–126.

Hamilton, J., 1994. Time Series Analysis, Princeton University Press, Princeton.

Harvey, A., 1989. Forecasting Structural Time Series Models and the Kalman Filter, second ed. Cambridge University Press, Cambridge.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59, 1551–1580.

King, M., Sentana, E., Wadhwani, S., 1994. Volatility and links between national stock markets. Econometrica 62, 901–933.

Koopman, S.J., 1993. Disturbance moother for state space models. Biometrika 80, 117–126.

Lütkepohl, H., 1993. Introduction to Multiple Time Series Analysis, Springer, Berlin.

Peña, D., 1995. Forecasting growth with time series models. J. Forecasting 14, 97–105.

Peña, D., Box, G., 1987. Identifying a simplifying structure in time series. J. Amer. Statist. Assoc. 82, 836–843.

Phillips, P.C.B., Durlauf, S.N., 1986. Multiple Time Series Regressions with Integrated Processes. Rev. Econ. Stud. 53, 473–496.

Priestley, M.B., Rao, T.S., Tong, J., 1974. Applications of principal component analysis and factor analysis in the identification of multivariable systems. IEEE Trans. Automat. Control 19, 703–704.

Reinsel, G.C., 1993. Elements of Multivariate Time Series Analysis, Springer, New York.

Reinsel, G.C., Ahn, S.K., 1992. Vector autoregressive models with unit roots and reduced rank structure: estimation, likelihood ratio test, and forecasting. J. Time Ser. Anal. 13, 353–375.
Robinson, P.M., 1973. Generalized canonical analysis for time series. J. Multivariate Anal. 3, 141–160.
Shumway, R.H., Stoffer, D.S., 1982. An approach to time series smoothing and forecasting using the EM algorithm. J. Time Ser. Anal. 3, 253–264.
Shumway, R.H., Stoffer, D.S., 2000. Time Series Analysis and its Applications, Springer, New York.
Stock, J.H., Watson, M.W., 1988. Testing for common trends. J. Amer. Statist. Assoc. 83, 1097–1107.
Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. J. Amer. Statist. Assoc. 97, 1167–1179.
Stoffer, D.S., Wall, K.D., 1991. Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. J. Amer. Statist. Assoc. 86, 1024–1033.
Tanaka, K., 1996. Time series analysis. Nonstationary and Noninvertible Distribution Theory, Wiley, New York.
Tiao, G.C., Tsay, R.S., 1989. Model specification in multivariate time series. J. Roy. Statist. Soc. B 51, 157–213.
Tsay, R.S., Tiao, G.C., 1990. Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. Ann. Statist. 18, 220–250.
Velu, R.P., Reinsel, G.C., Wichern, D.W., 1986. Reduced rank models for multiple time series. Biometrika 73, 105 –118.
Watson, M.W., Engle, R.F., 1983. Alternative algorithms for the estimation of dynamic, mimic and varying coefficient regression models. J. Econometr. 23, 385–400.
Wu, L.S., Pai, J.S., Hosking, J.R.M., 1996. An algorithm for estimating parameters of state-space models. Statist. Probab. Lett. 28, 99–106.
Zhang, H., 1993. Treasury yield curves and cointegration. Appl. Econ. 25, 361–367.