# 28

# Dimension Reduction in Multivariate Time Series

**Daniel Peña and Pilar Poncela**

***Universidad Carlos III de Madrid, Madrid, Spain***
***Universidad Autónoma de Madrid, Madrid, Spain***

**Abstract:** This article compares models for dimension reduction in time series and tests of the dimension of the dynamic structure. We consider both stationary and nonstationary time series and discuss principal components, canonical analysis, scalar component models, reduced rank models and factor models. The unifying view of canonical correlation analysis between the present and past values of the series is emphasized. Then, we review some of the tests based on canonical correlation analysis to find the dimension of the dynamic relationship among the time series. Finally, the procedures are compared through a real data example.

**Keywords and phrases:** Canonical correlation analysis, dimension reduction, vector time series

## 28.1 Introduction

Dimension reduction is very important in vector time series because the number of parameters in a model grows very fast with the dimension $m$ of the vector of time series $\boldsymbol{y}_t$. Linear models usually have a number of parameters which grows with $m^2$ and, for instance, a VARMA$(p, q)$ model contains $m^2(p+q)$ parameters. This problem can be even more important in nonlinear vector time series and, for instance, in a bilinear vector model or a threshold AR vector the number of parameters can easily be very large. The same problem appears in models with changing conditional variance as multivariate ARCH or GARCH models. Finding simplifying structures or factors in these models is important to reduce the number of parameters required to apply them to real data. In this article, we will consider linear time series models and we will concentrate in the time domain approach. See Brillinger (1981) and Shumway and Stoffer (2000) for analysis in the frequency domain. The first approach for reducing the dimension of a dynamic linear system is, by analogy with standard multivariate statistical analysis, finding linear combinations of the time series variables with simple properties. In stationary time series, we would be interested in finding linear combinations which are white noise because then the dynamics of the vector time series can be expressed by a number of components smaller than its dimension, $m$. In nonstationary series, we also would be interested in finding linear combinations which are stationary, reducing the dimension of the nonstationary space. This has been an important topic of research in the econometric literature under the name of cointegration; see, for instance, Engle and Granger (1987), Banarjee *et al.* (1993), and Johansen (1995). For VARMA models, dimension reduction was already analyzed in the pioneering work of Quenouille (1968). Some seminal contributions to this problem are the canonical analysis [Box and Tiao (1977)], the scalar component models, SCM, [Tiao and Tsay (1989)] and the reduced-rank models [Velu *et al.* (1986), Ahn and Reinsel (1990), Ahn (1997) and Reinsel and Velu (1998)]. A second approach for dimension reduction is by using Dynamic Factor models; see Anderson (1963), Priestly *et al.* (1974), Geweke and Singleton (1981), Brillinger (1981), Peña and Box (1987), Stock and Watson (1988), Molenaar *et al.* (1992), Forni *et al.* (2000) and Peña and Poncela (2004, 2006), among others. Factor models are very related to cointegration as it can be shown that 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 [Escribano and Peña (1994)].

In the state space approach [see Durbin and Koopman (2001)] dimension reduction appears in a natural way in defining the dimension of the state. Akaike (1974) in a seminal work introduced canonical correlation between the present and the future to determine the dimension of the state variables. Aoki (1987) made also important contributions. The dynamic factor model in state space form has been considered by Harvey (1989). State space models for multivariate time series have two advantages over the VARMA representation. First, the number of parameters in the model depends on the dimension of the state vector and when the series can be represented by a low dimension state vector the number of parameters is automatically reduced. Second, the state space representation provides a direct interpretation of the time series vector in components such as trend, cycle, seasonal and disturbance terms. In this way, we have the additional flexibility of searching for dimension reduction in the components, instead of trying a simplifying structure of the whole vector of time series.

One of the main tools for building tests for the dimension of a linear system is canonical correlation analysis. It can be shown that both linear combinations which are white noise and linear combinations which are stationary or nonstationary can be obtained from this approach. Also, it provides dimension tests which are invariant to affine transformations of the time series variables. The test proposed by Tiao and Tsay (1989) for SCM, the one used by Ahn and Reinsel (1988) and Reinsel and Ahn (1992) for the reduced rank autoregressive model, the cointegration test by Johansen (1988, 1991), and the tests proposed by Hu and Chou (2004) and Peña and Poncela (2006) for dynamic factor models are all based on canonical correlation analysis between the vector of time series or some of its differences and its lags. Related tests are the principal component test of Stock and Watson (1988) and Harris (1997).

This article is organized as follows. Section 28.2 presents different approaches for finding simplifying linear combinations in time series. Section 28.3 discusses tests for finding the dimension of the system based on canonical correlation analysis. Section 28.4 applies the procedures to an example and Section 28.5 includes some final remarks.

## 28.2 Models for Dimension Reduction

Suppose a $m \times 1$ vector $\boldsymbol{y}_t$ follows a linear time series process. We are interested in finding linear combinations $x_{1t} = \boldsymbol{m}'\boldsymbol{y}_t$ of the vector of time series with useful properties for model simplification and dimension reduction. Also, we will consider dynamic factor models in which the factors are not necessarily linear combinations of the observed time series.

### 28.2.1 Principal components

Let $\boldsymbol{y}_t$ be a stationary process with mean $\boldsymbol{\mu}$. Define the covariance matrices by

$$\boldsymbol{\Gamma}_y(k) = E\left\{(\boldsymbol{y}_{t-k} - \boldsymbol{\mu})(\boldsymbol{y}_t - \boldsymbol{\mu})'\right\},$$

and suppose that we are interested in linear combinations, $x_{1t} = \boldsymbol{m}'\boldsymbol{y}_t$, with maximum variance. Let $x_{it} = \psi(B)u_t$ be the model for the linear combination $x_{1t}$; then, as $\mathrm{Var}(x_{it}) = \sigma_u^2 \sum \psi_i^2$, linear combinations which are white noise will be associated to a small variance, and linear combinations close to nonstationary will be associated to a large variance. This association suggests looking for linear combinations of large or small variance, and it is well known that they will be given by the eigenvectors $\boldsymbol{m}_i$ in

$$\boldsymbol{\Gamma}_y(0)\boldsymbol{m}_i = \lambda_i \boldsymbol{m}_i$$

and the corresponding eigenvalues, $\lambda_i$, will be the variances of the linear combinations. In the particular case in which exact dimension reduction can be obtained, because one of the series is a linear combination of the others, this fact will be revealed by a zero eigenvalue in this covariance matrix $\boldsymbol{\Gamma}_y(0)$, and the linear combination will be given by the corresponding eigenvector. This approach can be extended to the nonstationary case. Suppose $\boldsymbol{y}_t$ is nonstationary $I(d)$. Then, following Peña and Poncela (2006), we define the generalized covariance matrices by

$$\boldsymbol{C}(k) = \frac{1}{T^{2d}} \sum (\boldsymbol{y}_{t-k} - \overline{\boldsymbol{y}})(\boldsymbol{y}_t - \overline{\boldsymbol{y}})',$$

where $\overline{\boldsymbol{y}} = T^{-1} \sum \boldsymbol{y}_t$. The solutions of

$$\boldsymbol{C}(0)\boldsymbol{m}_i = \lambda_i \boldsymbol{m}_i$$

will provide the interesting linear combinations: those link to large eigenvalues may define the nonstationary components, and those link to small eigenvalues may define the stationary components. However, note that principal components are not invariant under scale transformation of the variables and we may, by changing the scale, make the variance of a stationary component much larger than the one of a nonstationary one. For this reason, principal components in time series could be useful when all the series have a common scale of measurement, but are less justified otherwise.

### 28.2.2 The Box and Tiao canonical analysis

Box and Tiao (1977) proposed to find linear combinations of a stationary time series with maximum predictability, and called the procedure canonical analysis. We will refer to this procedure as BT analysis. Let $x_{1t} = \widehat{x}_{1t-1}(1) + u_t$, where $\widehat{x}_{1t-1}(1)$ is the one step ahead prediction and $u_t$ the forecast error. Let $\sigma_x^2$ be the variance of $x_{1t}$, and $\sigma_u^2$ the variance of $u$. These authors define the predictability by

$$q = \frac{(\sigma_x^2 - \sigma_u^2)}{\sigma_x^2} = 1 - \sigma_x^{-2}\sigma_u^2. \tag{28.1}$$

Thus, a white noise series has a predictability equal to zero and a nonstationary process has a predictability close to one. For instance, an AR(1) has $\sigma_x^2 = \sigma_u^2/(1-\phi^2)$ and $q = \phi^2$. If $\phi \to 1$, then $q \to 1$. This measure can be interpreted as a generalized determination coefficient. A vector time series model implies a decomposition of the form

$$\boldsymbol{y}_t = \widehat{\boldsymbol{y}}_{t-1}(1) + \boldsymbol{\varepsilon}_t,$$

where $\widehat{\boldsymbol{y}}_{t-1}(1)$ is now the vector of one step ahead predictions and $\boldsymbol{\varepsilon}_t$ the forecast error. As these terms are uncorrelated, we can also split the covariance matrix, $\boldsymbol{\Gamma}_y(0)$, as

$$\boldsymbol{\Gamma}_y(0) = \boldsymbol{F}_y(0) + \boldsymbol{\Sigma},$$

where $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \boldsymbol{\Sigma}$ and $E\left[(\widehat{\boldsymbol{y}}_{t-1}(1)-\boldsymbol{\mu})(\widehat{\boldsymbol{y}}_{t-1}(1)-\boldsymbol{\mu})'\right] = \boldsymbol{F}_y(0)$. It can be shown that the linear combinations of maximum predictability are defined by the largest eigenvectors of the predictability matrix

$$\boldsymbol{Q} = \boldsymbol{I} - \boldsymbol{\Gamma}_y(0)^{-1}\boldsymbol{\Sigma}, \tag{28.2}$$

and that the eigenvalues give the predictability of these linear combinations. Note that (28.2) reduces to (28.1) for scalar time series. If $h$ linear combinations are white noise, this matrix will have $h$ eigenvalues equal to zero and if $r$ linear combinations approach the nonstationary case, $\boldsymbol{Q}$ will have $r$ eigenvalues close to one. This analysis can be seen as (1) a generalized principal components approach for time series, and (2) a canonical correlation analysis between the vector of variables $\boldsymbol{y}_t$ and its lags. To illustrate the first interpretation we will use that, as the eigenvectors $\boldsymbol{m}_i$ of $\boldsymbol{Q}$ must satisfy $\boldsymbol{Q}\boldsymbol{m}_i = (\boldsymbol{I} - \boldsymbol{\Gamma}_y(0)^{-1}\boldsymbol{\Sigma})\boldsymbol{m}_i$ $= \lambda_i\boldsymbol{m}$, then $\boldsymbol{\Gamma}_y(0)^{-1}\boldsymbol{\Sigma}\boldsymbol{m}_i = (1-\lambda_i)\boldsymbol{m}_i$, and also

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_y(0)\boldsymbol{m}_i = \alpha_i\boldsymbol{m}_i, \tag{28.3}$$

where $\alpha_i = (1-\lambda_i)^{-1}$. Note that in the matrix $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_y(0)$ the eigenvectors link to eigenvalues equal to one define white noise components and those link to a large eigenvalue define nonstationary components. In the particular case $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$, that is, the noises are uncorrelated with the same variance, the BT analysis is a principal component analysis of the vector time series. For instance, the linear combination of maximum predictability is the first principal component of the data. In the general case where $\boldsymbol{\Sigma}$ is a positive definite covariance matrix, calling $\boldsymbol{\Sigma} = \boldsymbol{ADA'}$ to the spectral decomposition of the noise covariance matrix, from (28.3) we have

$$(\boldsymbol{D}^{-1/2}\boldsymbol{A}'\boldsymbol{\Gamma}_y(0)\boldsymbol{A}\boldsymbol{D}^{-1/2})(\boldsymbol{D}^{1/2}\boldsymbol{A}'\boldsymbol{m}_i) = \alpha_i(\boldsymbol{D}^{1/2}\boldsymbol{A}'\boldsymbol{m}_i),$$

and the BT analysis can be interpreted as: (a) transforming the vector of time series by $\boldsymbol{s}_t = \boldsymbol{D}^{-1/2}\boldsymbol{A}'\boldsymbol{y}_t$, so that the noise covariance of the transformed time series is the identity; (b) computing the principal components of $\boldsymbol{s}_t$, let us call them $\boldsymbol{v}_i$; and (c) transforming back the principal components by $\boldsymbol{m}_i = \boldsymbol{A}\boldsymbol{D}^{-1/2}$ $\boldsymbol{v}_i$. To obtain the canonical correlation analysis interpretation note that the canonical correlation coefficients between $\boldsymbol{y}_t$ and $\boldsymbol{y}_t^* = (\boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-k}')'$ are given by the non-zero eigenvalues of the matrix

$$\boldsymbol{M} = \boldsymbol{\Gamma}_y^{-1}(0)\boldsymbol{\Gamma}_{yy^*}(k)\boldsymbol{\Gamma}_{y^*}^{-1}(0)\boldsymbol{\Gamma}_{yy^*}'(k) \tag{28.4}$$

where, assuming to simplify that $E(\boldsymbol{y}_t) = 0$, we have $\boldsymbol{\Gamma}_y(0) = E(\boldsymbol{y}_t\boldsymbol{y}_t')$, $\boldsymbol{\Gamma}_{y^*}(0) = E(\boldsymbol{y}_t^*\boldsymbol{y}_t^{*\prime})$ and $\boldsymbol{\Gamma}_{yy^*}(k) = E(\boldsymbol{y}_t\boldsymbol{y}_t^{*\prime})$. Let

$$\boldsymbol{\Gamma}_{y|y^*} = E\left[(\boldsymbol{y}_t - \widehat{\boldsymbol{\beta}}\boldsymbol{y}_t^*)(\boldsymbol{y}_t - \widehat{\boldsymbol{\beta}}\boldsymbol{y}_t^*)'\right]$$

be the residual covariance matrix of a multivariate regression equation between $\boldsymbol{y}_t$ and $\boldsymbol{y}_t^* = (\boldsymbol{y}_{t-1}', ..., \boldsymbol{y}_{t-k}')'$. As $\widehat{\boldsymbol{\beta}} = \boldsymbol{\Gamma}_{yy^*}\ (k)\boldsymbol{\Gamma}_{y^*}^{-1}(0)$, we have

$$\boldsymbol{\Gamma}_{y|y^*} = \boldsymbol{\Gamma}_y(0) - \boldsymbol{\Gamma}_{yy^*}(k)\boldsymbol{\Gamma}_{y^*}^{-1}(0)\boldsymbol{\Gamma}_{yy^*}'(k) \tag{28.5}$$

and inserting $\boldsymbol{\Gamma}_{yy^*}(k)\boldsymbol{\Gamma}_{y^*}^{-1}(0)\boldsymbol{\Gamma}_{yy^*}'(k) = \boldsymbol{\Gamma}_y(0) - \boldsymbol{\Gamma}_{y|y^*}$ in (28.4), the $\boldsymbol{M}$ matrix can be written as

$$\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{\Gamma}_y^{-1}(\mathbf{0})\boldsymbol{\Gamma}_{y|y^*}$$

which is equivalent to the predictability matrix $\boldsymbol{Q}$ defined in (28.2). Thus, the linear combinations of maximum predictability are equivalent to the linear combinations of maximum correlation between the present and the past.

As an illustration, consider the VAR(1) model

$$\boldsymbol{y}_t = \boldsymbol{\Phi}\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{28.6}$$

Then $\boldsymbol{\Gamma}_y(0) = \boldsymbol{\Phi}\boldsymbol{\Gamma}_y(0)\boldsymbol{\Phi}' + \boldsymbol{\Sigma}$ , $\boldsymbol{\Gamma}_y'(1) = \boldsymbol{\Phi}\boldsymbol{\Gamma}_y(0)$ and the matrix $\boldsymbol{Q}$ given by (28.2) can also be written as $\boldsymbol{Q} = \boldsymbol{I} - \boldsymbol{\Gamma}_y^{-1}(0)(\boldsymbol{\Gamma}_y(0) - \boldsymbol{\Phi}\boldsymbol{\Gamma}_y(0)\boldsymbol{\Phi}')$, or

$$\boldsymbol{Q} = \boldsymbol{\Gamma}_y^{-1}(0)\boldsymbol{\Phi}\boldsymbol{\Gamma}_y(0)\boldsymbol{\Phi}'$$

which implies

$$\boldsymbol{Q} = \boldsymbol{\Gamma}_y^{-1}(0)\boldsymbol{\Gamma}_y'(1)\boldsymbol{\Gamma}_y^{-1}(0)\boldsymbol{\Gamma}_y(1).$$

This matrix is the standard canonical correlation matrix whose eigenvalues are the canonical correlations between $\boldsymbol{y}_t$ and $\boldsymbol{y}_{t-1}$. A zero canonical correlation defines a linear combination which is white noise and a close to one canonical correlation defines a close to nonstationary component.

### 28.2.3 Reduced rank models

An alternative procedure for finding linear combinations with useful properties for model simplification are the reduced rank models; see Robinson (1973), Ahn and Reinsel (1990), Reinsel and Ahn (1992), and Reinsel and Velu (1998). Suppose for simplicity that a vector of time series is fitted by the VAR(1) model (28.6) and suppose that $\boldsymbol{\Phi} = \boldsymbol{A}_r\boldsymbol{B}_r$, where $\boldsymbol{A}_r$ is a full rank matrix of dimension $m \times r$, $(m > r)$, and $\boldsymbol{B}_r$ is also full rank with dimension $r \times m$. Denoting $\boldsymbol{z}_{t-1} = \boldsymbol{B}_r\boldsymbol{y}_{t-1}$, the model for the series can be written as

$$\boldsymbol{y}_t = \boldsymbol{A}_r\boldsymbol{z}_{t-1} + \boldsymbol{a}_t \tag{28.7}$$

and also, as $\boldsymbol{B}_r\boldsymbol{y}_t = \boldsymbol{B}_r\boldsymbol{A}_r\boldsymbol{z}_{t-1} + \boldsymbol{B}_r\boldsymbol{a}_t$, we have

$$\boldsymbol{z}_t = \boldsymbol{C}\boldsymbol{z}_{t-1} + \boldsymbol{u}_t \tag{28.8}$$

where $\boldsymbol{C} = \boldsymbol{B}_r\boldsymbol{A}_r$ is a $r \times r$ matrix and $\boldsymbol{u}_t = \boldsymbol{B}_r\boldsymbol{a}_t$. This is like a factor model with $r$ factors $\boldsymbol{z}_{t-1}$ which follow an AR(1) model. An important implication from this model is that there exist $m-r$ linear combinations which are white noise. Denoting $\boldsymbol{A}_{m-r,\perp}$ for the orthogonal complement of $\boldsymbol{A}_r$, defined as the $m \times (m-r)$ matrix such that

$$\boldsymbol{A}'_{m-r,\perp}\boldsymbol{A}_r = \boldsymbol{0},$$

the $m-r$ linear combinations $\boldsymbol{A}'_{m-r,\perp}\boldsymbol{y}_t$ are white noise, or, in other words, there must be $m-r$ zero canonical correlations between $\boldsymbol{y}_t$ and $\boldsymbol{y}_{t-1}$. These ideas can be generalized to general VAR(p) models. We can write

$$\boldsymbol{y}_t = \boldsymbol{F}\boldsymbol{y}_t^* + \boldsymbol{a}_t,$$

where $\boldsymbol{y}_t^* = (\boldsymbol{y}'_{t-1}, ..., \boldsymbol{y}'_{t-k})'$ and $\boldsymbol{F} = (\Phi_1, ..., \Phi_k)$. Then, as before, if $\boldsymbol{F}$ has reduced rank, $\boldsymbol{F} = \boldsymbol{A}_r\boldsymbol{B}_r$, we have

$$\boldsymbol{y}_t = \boldsymbol{A}_r\boldsymbol{z}_t + \boldsymbol{a}_t,$$

where $\boldsymbol{z}_t = \boldsymbol{B}_r\ \boldsymbol{y}_t$. The implication of this model is that the canonical correlations between $\boldsymbol{y}_t$ and $\boldsymbol{y}_t^*$ will have as many zero canonical correlations as white noise combinations. Also, it can be shown that the number of canonical correlations equal to one is the number of nonstationary linear combinations of the vector.

### 28.2.4 The scalar component models

Tiao and Tsay (1989) presented the concept of scalar component models as simplifying tools in VARMA models. A scalar component model is a linear combination of the vector time series which follows a simpler structure than the vector itself. These authors define SCM as follows. Assume that we can write $\boldsymbol{y}_t = \sum \boldsymbol{\Psi}_i\boldsymbol{a}_{t-i}$, where $\boldsymbol{a}_t$ is white noise. We will say that $x_t = \boldsymbol{v}'_0\boldsymbol{y}_t$ follows a SCM($p_1$,$q_1$) if there exist $p_1$ vectors $m \times 1$, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{p_1}$ such that (i) $\boldsymbol{v}_{p_1}$ is non-zero when $p_1 > 0$, and (ii) the linear combination of $\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_{t-p_1}$ given by $m_t = \boldsymbol{v}'_0\boldsymbol{y}_t + \sum_{l=1}^{p_1}\boldsymbol{v}'_l\boldsymbol{y}_{t-l}$ satisfies

$$E(\boldsymbol{a}_{t-j}m_t)\begin{cases} \neq 0 & \text{if } j = q_1 \\ = 0 & \text{if } j > q_1 \end{cases}.$$

The above definition implies the following restriction among the autocovariance matrices of $\boldsymbol{y}_t$:

$$\boldsymbol{\Gamma}_y(k)\boldsymbol{v}_0 + \boldsymbol{\Gamma}_y(k-1)\boldsymbol{v}_1 + \cdots + \boldsymbol{\Gamma}_y(k-p_1)\boldsymbol{v}_{p_1} = \boldsymbol{0}, \text{ for } l > q_1. \tag{28.9}$$

Of particular interest are SCM(0,0) which are white noise and SCM(1,0), which can define a particular type of common trends. See Peña, Tiao and Tsay (2001) for a simple introduction to the use of SCM for model simplification. To find out the number of scalar component models, Tiao and Tsay (1989) proposed a chi-square test based on canonical correlation ideas for the rank of extended second moment matrices, which will be discussed in Section 28.3.

### 28.2.5 Dynamic factor models

A generalization of the idea of linear combinations with useful properties is the Dynamic Factor model. In this model, the $m$-dimensional vector of observed time series is generated by a set of $r$ non-observed common factors and $m$ specific components as follows:

$$\underset{m\times 1}{\boldsymbol{y}_t} = \underset{m\times r}{\boldsymbol{P}} \; \underset{r\times 1}{\boldsymbol{f}_t} + \underset{m\times 1}{\boldsymbol{n}_t}, \tag{28.10}$$

where $\boldsymbol{f}_t$ is the $r$-dimensional vector of common factors, $\boldsymbol{P}$ is the factor loading matrix, and $\boldsymbol{n}_t$ is the vector of specific components. Thus, all the common dynamic structure comes through the common factors, $\boldsymbol{f}_t$, whereas the vector $\boldsymbol{n}_t$ explains the specific dynamics for each component. If there is no specific dynamic structure, $\boldsymbol{n}_t$ is reduced to white noise. We assume linear time series models for the latent variable $\boldsymbol{f}_t$ and the noise $\boldsymbol{n}_t$. In particular, using the VARIMA$(p,d,q)$ representation, the latent variable will be given by

$$\underset{r\times r}{\boldsymbol{\Phi}(B)} \; \underset{r\times 1}{\boldsymbol{f}_t} = \underset{r\times r}{\boldsymbol{\Theta}(B)} \; \underset{r\times 1}{\boldsymbol{a}_t}, \tag{28.11}$$

where $B$ is the backshift operator, such that $B\boldsymbol{y}_t = \boldsymbol{y}_{t-1}$, and (i) the $r\times r$ matrix $\boldsymbol{\Phi}(B) = \boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$ has the roots of the determinantal equation$|\boldsymbol{\Phi}(B)| = 0$ on or outside the unit circle; (ii) the $r\times r$ matrix $\boldsymbol{\Theta}(B) = \boldsymbol{I} - \boldsymbol{\Theta}_1 B - \cdots - \boldsymbol{\Theta}_q B^q$ has the roots of the determinantal equation$|\boldsymbol{\Theta}(B)| = 0$ outside the unit circle; and (iii) $\boldsymbol{a}_t \sim N_r(\boldsymbol{0}, \boldsymbol{\Sigma}_a)$ is serially uncorrelated, $\mathrm{E}(\boldsymbol{a}_t\boldsymbol{a}'_{t-h}) = \boldsymbol{0}$, $h \neq 0$. The noise, $\boldsymbol{n}_t$, also follows the VARMA model

$$\boldsymbol{\Phi}_n(B)\boldsymbol{n}_t = \boldsymbol{\Theta}_n(B)\boldsymbol{\varepsilon}_t, \tag{28.12}$$

where $\boldsymbol{\Phi}_n(B)$ and $\boldsymbol{\Theta}_n(B)$ are $m\times m$ diagonal matrices with $\boldsymbol{\Phi}_n(B) = \boldsymbol{I} - \boldsymbol{\Phi}_{n1}B - \cdots - \boldsymbol{\Phi}_{np}B^p$ and $\boldsymbol{\Theta}_n(B) = \boldsymbol{I} - \boldsymbol{\Theta}_{n1}B - \cdots - \boldsymbol{\Theta}_{nq}B^q$. The most interesting case is when the specific component is stationary so that the possible nonstationary dynamic structure in the vector of time series is due to the common factors. In this case the roots of the determinantal equations $|\boldsymbol{\Phi}_n(B)| = 0$ and $|\boldsymbol{\Theta}_n(B)| = 0$ are outside the unit circle. Therefore, each component follows a univariate ARMA$(p_i, q_i)$, $i = 1, 2, \cdots, m$, being $p$=max$(p_i)$ and $q$=max$(q_i)$, $i = 1, 2, \cdots, m$. The sequence of vectors $\boldsymbol{\varepsilon}_t$ are normally distributed, with zero mean and diagonal covariance matrix $\boldsymbol{\Sigma}_\varepsilon$. We assume that the noises from the common factors and specific components are also uncorrelated for all lags, that is, $\forall h$ $\mathrm{E}(\boldsymbol{a}_t\boldsymbol{\varepsilon}'_{t-h}) = \boldsymbol{0}$. When $\boldsymbol{n}_t$ is white noise and the factors are stationary, models (28.10) and (28.11) are the factor model studied by Peña and Box (1987). The model as stated is not identified and we can choose either $\boldsymbol{\Sigma}_{\boldsymbol{a}} = \boldsymbol{I}$ or $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}$, although the model is not yet identified under rotations. Harvey (1989) imposes the additional condition that $p_{ij} = 0$ for $j > i$, where $\boldsymbol{P} = [p_{ij}]$.

Note that this factor model is very general and includes other formulations presented in the literature. For instance, Molenaar *et al.* (1992) have proposed a model of the form

$$\boldsymbol{y}_t = \sum_{i=0}^{s} \boldsymbol{P}_i \boldsymbol{f}_{t-i} + \boldsymbol{n}_t.$$

Letting $\boldsymbol{f}_t^* = (\boldsymbol{I} + \boldsymbol{P}_0^{-1}\boldsymbol{P}_1 B + \cdots + \boldsymbol{P}_0^{-1}\boldsymbol{P}_s B^s)\boldsymbol{f}_t = \varphi(B)\boldsymbol{f}_t$, we can write this model as $\boldsymbol{y}_t = \boldsymbol{P}_0\boldsymbol{f}_t^* + \boldsymbol{\varepsilon}_t$ where the new factors follow a different VARMA model. The factor model has an interesting implication in terms of canonical correlation. Suppose that there is no specific components so that the model is

$$\boldsymbol{y}_t = \boldsymbol{P}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t;$$

then, denoting $\boldsymbol{P}_\perp'$ for the $(m-r)\times m$ matrix which defines the null space of $\boldsymbol{P}$, such that $\boldsymbol{P}_\perp'\boldsymbol{P} = \boldsymbol{0},$ we have

$$\boldsymbol{P}_\perp'\boldsymbol{y}_t = \boldsymbol{P}_\perp'\boldsymbol{\varepsilon}_t$$

and there must be $m-r$ zero canonical correlations between $\boldsymbol{y}_t$ and $\boldsymbol{y}_t^*$.

### 28.2.6 State space models

State space models have been studied by Akaike (1974), Aoki (1987), Hannan and Deistler (1988), Harvey (1989), and Durbin and Koopman (2001), among others. They are defined by a measurement equation

$$\boldsymbol{y}_t = \boldsymbol{C}\boldsymbol{z}_t + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{C}$ is $m\times\ s$, $\boldsymbol{z}_t$ is the $s\times 1$ state vector and $\boldsymbol{\varepsilon}_t$, $m\times 1$, is the innovation vector with $\mathrm{E}(\boldsymbol{\varepsilon}_t) = 0$, $\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \boldsymbol{\Sigma}_\epsilon$ $\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_\tau') = \boldsymbol{0}$ if $t \neq \tau$. The transition equation is

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

with $\mathrm{E}(\boldsymbol{u}_t) = \boldsymbol{0}$, $\mathrm{E}(\boldsymbol{u}_t\boldsymbol{u}_t') = \boldsymbol{\Sigma}_u$ and $\mathrm{E}(\boldsymbol{u}_t\boldsymbol{u}_\tau') = \boldsymbol{0}$ if $t \neq \tau$. Although any VARMA model can be written in the state space form and we can always obtain the VARMA form of a state space representation, the state space formulation has the advantage of being defined in terms of the state vector which is the key component for dimension reduction. In fact, Akaike (1974) introduced canonical correlation in time series in order to find the dimension of the state space vector. For instance, we may have a dynamic factor model by

$$\boldsymbol{y}_t = \boldsymbol{C}\boldsymbol{z}_t + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{C}$ is $m\times r$ and

$$\boldsymbol{z}_t = \boldsymbol{z}_{t-1} + \boldsymbol{\beta} + \boldsymbol{u}_t.$$

This is the common trends model because the vector of dimension $m$ is generated by $r$ factors which follows a random walk with drift model. Note that the state vector coincides with the factor. The VARMA form of this model is

$$\nabla \boldsymbol{y}_t = \boldsymbol{C}(\boldsymbol{\beta} + \boldsymbol{u}_t) + \nabla \boldsymbol{\varepsilon}_t = \boldsymbol{c} + (\boldsymbol{I} - \boldsymbol{\Theta} B)\boldsymbol{a}_t$$

and the observed series will follow a VARIMA(1,1). However, in this formulation, the factor is completely lost and, as shown by Peña and Box (1987), fitting an ARIMA model to an observed time series generated from this model may be a difficult task because of lack of identification of the parameter matrices. An additional advantage of the state space approach is that it allows for dimension reduction in some of the time series components and not in the others. See Casals *et al.* (2002) for useful structural decompositions in the state space approach. Suppose that the state vector is written as including the trend and the cycle of the time series as

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{T}_t + \boldsymbol{B}\boldsymbol{s}_t + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{A}$ is $m \times r$ and $\boldsymbol{B}$ is $m \times c$ where $r \leq m$ and $c \leq m$. Then, if $\boldsymbol{A}'_{m-r,\perp}\boldsymbol{A} = \boldsymbol{0}$ and $\boldsymbol{B}'_{m-c,\perp}\boldsymbol{B} = \boldsymbol{0}$**,** we have

$$\boldsymbol{A}'_{m-r,\perp}\boldsymbol{y}_t = \boldsymbol{A}'_{m-r,\perp}\boldsymbol{B}\boldsymbol{s}_t + \boldsymbol{A}'_{m-r,\perp}\boldsymbol{\varepsilon}_t$$

and

$$\boldsymbol{B}'_{m-c,\perp}\boldsymbol{y}_t = \boldsymbol{B}'_{m-c,\perp}\boldsymbol{A}\boldsymbol{T}_t + \boldsymbol{B}'_{m-c,\perp}\boldsymbol{\varepsilon}_t,$$

and we may have some linear combinations free from the trend and others free from the cycle. It could be that some of them are white noise if there are common vectors in the null space of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$**.**

### 28.2.7 Some conclusions

We have seen that canonical analysis plays a key role in all of the dimension reduction procedures for time series. If $h \geq 1$ linear combinations are white noise, there is only dynamics in $m - h$ dimensions and this implies $h$ zero canonical correlations between $\boldsymbol{y}_t$ and $\boldsymbol{y}_t^*$. Also, for integrated processes, an important simplification tool is finding linear combinations which are stationary. If there is cointegration and $h \geq 1$ linear combinations are stationary, then $m - h$ canonical correlations between $\boldsymbol{y}_t$ and $\boldsymbol{y}_t^*$ will be equal to one. It is interesting to understand the relationship between canonical analysis and principal components in time series. We have shown that when $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$**,** principal components and canonical analysis leads to similar conclusions. This is similar to the relationship between factor analysis and principal components in the static case. If the specific innovations of all the time series have the same variance and are uncorrelated, then a zero eigenvalue in the canonical correlation analysis of the series and its past values will be equivalent to an eigenvalue equal to one in the standardized principal components (SPC) of the matrix $\sigma^{-2}\mathbf{\Gamma}_y(0)$. Also a canonical correlation close to one will be equivalent to a large eigenvalue in the SPC. In practice, with nonstationary time series, the elements of $\mathbf{\Gamma}_y(0)$ will be much larger than those of $\mathbf{\Sigma}$, and this matrix often has diagonal elements of similar sizes and larger than the off-diagonal elements. In this case, the principal component matrix $\mathbf{\Gamma}_y(0)$ will be similar to the canonical correlation matrix $\mathbf{\Sigma}^{-1}\mathbf{\Gamma}_y(0)$, and both approaches will lead to similar results when applied for finding the cointegration or the factor space.

## 28.3 Dimension Reduction Tests

We present in this section tests for dimension reduction based on canonical correlation coefficients. Other related tests are the principal components tests by Stock and Watson (1988) and Harris (1997). An alternative way to decide about the dimension of the system is by using model selection criteria, such as AIC, BIC and others. The relative advantages of these two approaches require more research before a clear recommendation can be made.

### 28.3.1 A test for zero canonical correlation coefficients

Let $\boldsymbol{y}_t^* = (\boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-k})'$ be a $km \times 1$ vector of lag values of the series. We want to test if there exist linear combinations of $\boldsymbol{y}_t$ which are uncorrelated to linear combinations of $\boldsymbol{y}_t^*$ or, in other words, if there are zero canonical correlation coefficients between the two sets of variables. This test will allow to find the least predictable components in the canonical analysis of Box-Tiao, the rank $r$ in the reduced rank model, and can also be used to test for the number of factors in the dynamic factor model. Suppose that the null hypothesis is that there are $h$ zero canonical coefficients. Note that if we accept the presence of $h$ zero coefficients we must accept the presence of $h-1$. Thus, the test must be done sequentially starting with $h = 0$ and increasing $h$ until $m-1$. The alternative hypothesis will be that there are less than $h$ zero canonical correlation coefficients, and the test is: $H_0 : h \ (h = 0, 1, ..., m-1)$ zero correlation coefficients versus $H_1$ : less than $h$ zero correlation coefficients. The standard multivariate test for $h$ zero canonical correlation coefficients is

$$L = -\left\{(T - mk) + g(m,k)\right\} \sum_{j=1}^{h} \log(1 - \widehat{\lambda}_j), \tag{28.13}$$

where $g(m,k) = (mk-m-1)/2$ is a correction factor to improve the asymptotic distribution of the test statistic and $\widehat{\lambda}_1 \leq \widehat{\lambda}_2 \leq \cdots \leq \widehat{\lambda}_j \leq \cdots \leq \widehat{\lambda}_m$ are the ordered eigenvalues of

$$\widehat{\boldsymbol{M}}_k = \boldsymbol{C}_y^{-1} \boldsymbol{C}_{yy^*} \boldsymbol{C}_{y^*}^{-1} \boldsymbol{C}_{y^* y}, \tag{28.14}$$

where

$$\begin{aligned}
\boldsymbol{C}_y &= T^{-1} \sum_{t=2}^{T} (\boldsymbol{y}_t \boldsymbol{y}_t'), \\
\boldsymbol{C}_{yy^*} &= T^{-1} \sum_{t=k}^{T} (\boldsymbol{y}_t \boldsymbol{y}_t^{*\prime}), \\
\boldsymbol{C}_{y^*} &= T^{-1} \sum_{t=k}^{T} (\boldsymbol{y}_t^* \boldsymbol{y}_t^{*\prime}),
\end{aligned}$$

and $L$ it is distributed asymptotically as $\chi^2_{h(mk-(m-h))}$. This test can be derived as a likelihood ratio test [see, for instance, Rechner (1995)]. It has a simple interpretation as a Box-Pierce test on the canonical correlation coefficients as under the null

$$L \simeq T \sum_{j=1}^{m-h} \widehat{\lambda}_j = T \sum_{j=1}^{m-h} \widehat{\rho}_j^2$$

where $\widehat{\rho}_j^2$ are the canonical correlations. This test has been used in reduced rank models [see Reinsel and Velu (1998)] to test for the dimension of the reduced rank matrix.

A modification of the previous test was proposed by Tiao and Tsay (1989) in order to test for SCM. Let $\boldsymbol{Y}^*_{h,t-j-1} = (\boldsymbol{y}'_{t-j-1}, \ldots, \boldsymbol{y}'_{t-h-j-1})'$ and $\boldsymbol{Y}_{k,t} = (\boldsymbol{y}'_t, \ldots, \boldsymbol{y}'_{t-k})'$be $(h+1)m \times 1$ and $(k+1)m \times 1$, respectively, vectors of lag values of the series for $h \geq k \geq j \geq 0$. The purpose is to test the number of zero eigenvalues or zero canonical correlations between $\boldsymbol{Y}^*_{h,t-j-1}$ and $\boldsymbol{Y}_{k,t}$ which is determined by the rank of the lag second moment matrices $\boldsymbol{y}_t$ and the Yule-Walker equations of the overall process for $\boldsymbol{y}_t$. The test statistic is

$$TT = -(T-h-j) \sum_{j=i}^{s} \log \left(1 - \frac{\widehat{\lambda}_j}{d_j}\right), \tag{28.15}$$

where $s$ is the number of zero canonical correlations between $\boldsymbol{Y}^*_{h,t-j-1}$ and $\boldsymbol{Y}_{k,t}$ and $d_j/(T-h-j)$ is the sample variance of the two canonical variates whose sample canonical correlation is given by $\widehat{\lambda}_j$. Under the null hypothesis of $s$ zero canonical correlations, the test statistic follows a chi-squared with $s((h-k) \times m + s)$ degrees of freedom.

### 28.3.2 A non-standard test for canonical correlations

Suppose that in an $I(1)$ process we are interested in finding the number of nonstationary dimensions $r$ or the number of independent linear combinations which are stationary, $m-r$, which is the cointegration dimension. We say that the components of a nonstationary $I(d)$ time series vector $\boldsymbol{y}_t$ are cointegrated if there exists a linear combination of them which is $I(d-b)$, where $b>0$, $d\geq b$ and $d$ and $b$ belong to the set of the natural numbers. The most interesting case is when the series are $I(1)$ but some linear combinations are $I(0)$ or stationary. A cointegration test in this case tries to determine how many independent linear combinations of the series can be considered as stationary. To simplify the exposition, suppose the VAR(1) given by (28.6). If all the roots of $|\boldsymbol{I}-\boldsymbol{\Phi}B|=0$ are equal to one, all the eigenvalues of the matrix $\boldsymbol{\Phi}$ are equal to one and all the eigenvalues of the matrix

$$\boldsymbol{\Pi}=\boldsymbol{\Phi}-\boldsymbol{I}$$

are equal to zero. Note that this does not imply that $\boldsymbol{\Pi}$ is a zero matrix because it may not be symmetric. If the series are stationary, all the roots of $|\boldsymbol{I}-\boldsymbol{\Phi}B|=0$ are inside the unit circle and the matrix $\boldsymbol{\Pi}$ is a full rank matrix. Cointegration represents the intermediate situation in which the series are nonstationary but some linear combinations are stationary. Suppose that the matrix $\boldsymbol{\Phi}$ has $r$ eigenvalues equal to one, or, equivalently, the matrix $\boldsymbol{\Pi}$ has $r$ eigenvalues equal to zero. These properties can be applied to the error correction formulation of the VAR(1) obtained subtracting $\boldsymbol{y}_{t-1}$ from both sides of (28.6). Then

$$\nabla\boldsymbol{y}_t=\boldsymbol{\Pi}\boldsymbol{y}_{t-1}+\boldsymbol{\varepsilon}_t. \tag{28.16}$$

If all the series are nonstationary, but there is no cointegration, $\boldsymbol{\Pi}$ is a null rank matrix; if all of them are stationary, $\boldsymbol{\Pi}$ is a full rank matrix and if there is cointegration the matrix $\boldsymbol{\Pi}$ must be rank deficient. Then, if $rank(\boldsymbol{\Pi})=m-r$, we can write

$$\boldsymbol{\Pi}=\boldsymbol{A}_{m-r}\boldsymbol{B}_{m-r}$$

and the $r$ linear combinations

$$\boldsymbol{A}'_{r,\perp}\nabla\boldsymbol{y}_t=\boldsymbol{A}'_{r,\perp}\boldsymbol{\varepsilon}_t \tag{28.17}$$

must be white noise. Note that the cointegration relations are given by $\boldsymbol{z}_t=\boldsymbol{B}_{m-r}\boldsymbol{y}_t$. To see this, multiplying (28.16) by $\boldsymbol{B}_{m-r}$, we have

$$\nabla\boldsymbol{z}_t=\boldsymbol{B}_{m-r}\boldsymbol{A}_{m-r}\boldsymbol{z}_{t-1}+\boldsymbol{B}_{m-r}\boldsymbol{\varepsilon}_t$$

and as $\boldsymbol{B}_{m-r}\boldsymbol{A}_{m-r}$ is a squared full rank matrix of dimension $m-r$, $\boldsymbol{z}_t$ must be stationary. We may build a test of cointegration by searching for zero canonical correlations between $\nabla\boldsymbol{y}_t$ and $\boldsymbol{y}_{t-1}$. Let $0\leq\widehat{\lambda}_1\leq\cdots\leq\widehat{\lambda}_m\leq 1$ be the eigenvalues of the matrix

$$\boldsymbol{M}_2 = \boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01},$$

where

$$\boldsymbol{S}_{11} = T^{-1}\sum_{t=1}^{T} \nabla \boldsymbol{y}_t \nabla \boldsymbol{y}_t',$$

$$\boldsymbol{S}_{10} = T^{-1}\sum_{t=1}^{T} \nabla \boldsymbol{y}_t \boldsymbol{y}_{t-1}',$$

$$\boldsymbol{S}_{00} = T^{-1}\sum_{t=1}^{T} \boldsymbol{y}_{t-1} \boldsymbol{y}_{t-1}'.$$

Then the statistic for testing that there are $r$ zero canonical correlations, or $m-r$ cointegration relations, is

$$L_{m-r} = -T\sum_{j=1}^{r} \log(1-\widehat{\lambda}_j). \tag{28.18}$$

This is the cointegration test for $I(1)$ variables developed by Johansen (1991, 1995) for VAR processes, which has become very popular in econometrics. The distribution of the test is non-standard because although the linear combinations $\boldsymbol{A}_{r,\perp}'\nabla\boldsymbol{y}_t$ are white noise and uncorrelated to $\boldsymbol{z}_{t-1} = \boldsymbol{B}_{m-r}\boldsymbol{y}_{t-1}$, these linear combinations are not white noise. The percentiles of the distribution have been tabulated by simulation. Note that we could also test for zero canonical correlations between $\nabla\boldsymbol{y}_t$ and $\nabla\boldsymbol{y}_{t-1}$, $\nabla\boldsymbol{y}_{t-2}, \ldots$ since by (28.17) there are $r$ linear combinations of $\nabla\boldsymbol{y}_t$ that are white noise. For instance, if we want to search for zero canonical correlations between $\nabla\boldsymbol{y}_t$ and its first lag $\nabla\boldsymbol{y}_{t-1}$, we will find zero canonical correlations between each sample and also within each sample of the variables due to (28.17). In this particular case, the asymptotic distribution of the test statistic is $\chi^2$ since under the null hypothesis the smallest $r$ canonical variates are white noise.

The generalization of the test for VAR($p$) is straightforward. Suppose

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_1\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0},\boldsymbol{\Sigma})$. The process is nonstationary if some of the roots of the determinantal equation $|\boldsymbol{\Phi}(\boldsymbol{B})| = 0$ are on the unit circle, which implies that the matrix $\boldsymbol{I} - \sum_{i=1}^{p}\boldsymbol{\Phi}_i = -\boldsymbol{\Pi}$ is rank deficient. In order to use this property, we write the VAR model in the error correction form

$$\nabla\boldsymbol{y}_t = \boldsymbol{\Pi}\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1}\boldsymbol{\Gamma}_i\nabla\boldsymbol{y}_{t-i} + \boldsymbol{\varepsilon}_t, \tag{28.19}$$

where

$$\boldsymbol{\Pi} = \sum_{i=1}^{p} \boldsymbol{\Phi}_i - \boldsymbol{I}, \text{ and } \boldsymbol{\Gamma}_i = \sum_{j=i+1}^{p} \boldsymbol{\Phi}_j. \tag{28.20}$$

Then, if rank($\boldsymbol{\Pi}$)$= m-r$, this matrix can be written as $\boldsymbol{\Pi} = \boldsymbol{A}_{m-r}\boldsymbol{B}_{m-r}$ and there will be $m-r$ cointegration relationships and $r$ zero canonical correlations between $\nabla \boldsymbol{y}_t^* = \nabla \boldsymbol{y}_t - \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \nabla \boldsymbol{y}_{t-i}$ and $\boldsymbol{y}_{t-1}^* = \boldsymbol{y}_{t-1} - \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \nabla \boldsymbol{y}_{t-i}$. Note that, by (28.19), the $r$ linear combinations

$$\nabla \boldsymbol{A}'_{r,\perp}(\boldsymbol{y}_t - \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \nabla \boldsymbol{y}_{t-i}) = \boldsymbol{A}'_{r\perp}\boldsymbol{\varepsilon}_t$$

are white noise, where $\boldsymbol{A}_{r,\perp}$ is the orthogonal complement of $\boldsymbol{A}_{m-r}$, that is $\boldsymbol{A}'_{r,\perp}\boldsymbol{A}_{m-r} = \boldsymbol{0}$. The $m-r$ linear combinations given by $\boldsymbol{B}_{m-r}\boldsymbol{y}_t$ are I(0). Thus, the test uses the residuals of a regression of $\nabla \boldsymbol{y}_t$ and $\boldsymbol{y}_{t-1}$ on the lags of the first differences and then looks at the canonical correlation between these two sets or residuals. As before, the test is done sequentially assuming 0 cointegration relations at the initial stage and going up to $m-1$ cointegration relations. The (nonstandard) critical values can be taken from Johansen (1995). Reinsel and Ahn (1992) have proposed a similar test for the number of unit roots in reduced rank autoregression models.

It is interesting to analyze this test when is applied to the dynamic factor model. Assuming that the factors are integrated with $d = 1$, and follow the model

$$\underset{r \times r}{(1-B)\boldsymbol{\Phi}^*(B)} \; \underset{r \times 1}{\boldsymbol{f}_t} = \underset{r \times r}{\boldsymbol{\Theta}(B)} \; \underset{r \times 1}{\boldsymbol{a}_t}, \tag{28.21}$$

with $\boldsymbol{\Phi}^*(B)$ having all its roots outside the unit circle. Then

$$\boldsymbol{f}_t = \boldsymbol{f}_{t-1} + (\boldsymbol{\Phi}^*(B))^{-1}\boldsymbol{\Theta}(B) \quad \boldsymbol{a}_t. \tag{28.22}$$

From (28.10), we obtain

$$\boldsymbol{f}_t = \boldsymbol{P}^+(\boldsymbol{y}_t - \boldsymbol{n}_t), \tag{28.23}$$

where $\boldsymbol{P}^+ = (\boldsymbol{P}'\boldsymbol{P})^{-1}\boldsymbol{P}$, $r \times m$, is the Moore-Penrose inverse matrix of $\boldsymbol{P}$, and from (28.10), (28.23) and (28.22) we can write

$$\boldsymbol{y}_t = \boldsymbol{P}\boldsymbol{P}^+(\boldsymbol{y}_{t-1} - \boldsymbol{n}_{t-1}) + \boldsymbol{P}\left(\boldsymbol{\Phi}^*(B)\right)^{-1}\boldsymbol{\Theta}(B)\boldsymbol{a}_t + \boldsymbol{n}_t$$

and subtracting $\boldsymbol{y}_{t-1}$, we have

$$(1-B)\boldsymbol{y}_t = -(\boldsymbol{I} - \boldsymbol{P}\boldsymbol{P}^+)\boldsymbol{y}_{t-1} + \boldsymbol{P}\left(\boldsymbol{\Phi}^*(B)\right)^{-1}\boldsymbol{\Theta}(B)\boldsymbol{a}_t + \boldsymbol{n}_t - \boldsymbol{P}\boldsymbol{P}^+\boldsymbol{n}_{t-1}. \tag{28.24}$$

This is the error correction form implied by the factor model. Notice now that $\boldsymbol{P}\boldsymbol{P}^+ = \boldsymbol{P}(\boldsymbol{P}'\boldsymbol{P})^{-1}\boldsymbol{P}'$ is a projection matrix, such that rank($\boldsymbol{P}\boldsymbol{P}^+$) $= r$ and it has all its eigenvalues equal one or zero since it is an idempotent matrix. Therefore, rank $(\boldsymbol{I}-\boldsymbol{P}\boldsymbol{P}^{+}) = m-r$. The matrix $(\boldsymbol{I}-\boldsymbol{P}\boldsymbol{P}^{+})$ plays the role of the $\boldsymbol{\Pi} = \boldsymbol{A}\boldsymbol{B}$ matrix in the cointegration analysis and the test of $r$ common factors is equivalent to the test of $m-r$ cointegration relations. However, in order to use Johansen's cointegration test, we have to assume that the process followed by $\boldsymbol{y}_t$ can be approximated by an unrestricted VAR. When the true model is the dynamic factor model usually we also have MA structure.

### 28.3.3 A canonical correlation test for factor models

Canonical correlation tests for factor models have been proposed by Hu and Chou (2004) and Peña and Poncela (2006). In this subsection, we review the latest one. Suppose the factor model without specific components is

$$\boldsymbol{y}_t = \boldsymbol{P}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t. \tag{28.25}$$

Then, as shown by Peña and Box (1987), denoting $\boldsymbol{\Gamma}_f(k)$ for the covariance matrix of order $k$ of the factors and assuming stationarity we have, for $k \neq 0$

$$\boldsymbol{\Gamma}_y(k) = \boldsymbol{P}\boldsymbol{\Gamma}_f(k)\boldsymbol{P}' \tag{28.26}$$

and rank($\boldsymbol{\Gamma}_y(k)$)=rank($\boldsymbol{\Gamma}_f(k)$). Since (28.26) is true for all $k \neq 0$, there exists a $m \times (m-r)$ matrix $\boldsymbol{P}_\perp$, such that for all $k \neq 0$,

$$\boldsymbol{\Gamma}_y(k)\boldsymbol{P}_\perp = \boldsymbol{P}\boldsymbol{\Gamma}_f(k)\boldsymbol{P}'\boldsymbol{P}_\perp = \boldsymbol{0}. \tag{28.27}$$

The condition in (28.27) also implies that the $m-r$ independent linear combinations of the observed series given by $\boldsymbol{P}'_\perp\boldsymbol{y}_t$ are cross and serially uncorrelated for all lags $k \neq 0$. Therefore, the number of zero canonical correlations between $\boldsymbol{y}_{t-k}$ and $\boldsymbol{y}_t$ is given by the number of zero eigenvalues of the matrix $\boldsymbol{M}(k)$ defined as

$$\boldsymbol{M}(k) = \left[E(\boldsymbol{y}_t\boldsymbol{y}'_t)\right]^{-1} E(\boldsymbol{y}_t\boldsymbol{y}'_{t-k}) \left[E(\boldsymbol{y}_{t-k}\boldsymbol{y}'_{t-k})\right]^{-1} E(\boldsymbol{y}_{t-k}\boldsymbol{y}'_t) \tag{28.28}$$

and since rank($\boldsymbol{M}(k)$)=rank($\boldsymbol{\Gamma}_y(k)) = r$, this number is $m-r$. Thus, the number of common factors, $r$, is equivalent to the number of non-zero canonical correlations between $\boldsymbol{y}_{t-k}$ and $\boldsymbol{y}_t$.

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

$$\widehat{\boldsymbol{M}}_1(k) = \left[\sum_{t=k+1}^{T}(\boldsymbol{y}_t\boldsymbol{y}'_t)\right]^{-1} \sum_{t=k+1}^{T}(\boldsymbol{y}_t\boldsymbol{y}'_{t-k}) \left[\sum_{t=k+1}^{T}(\boldsymbol{y}_{t-k}\boldsymbol{y}'_{t-k})\right]^{-1} \sum_{t=k+1}^{T}(\boldsymbol{y}_{t-k}\boldsymbol{y}'_t). \tag{28.29}$$

In Peña and Poncela (2006), it has been shown that, given $\widehat{\lambda}_1 \leq \widehat{\lambda}_2 \leq \cdots \leq \widehat{\lambda}_m$, the ordered eigenvalues of the matrix $\widehat{\boldsymbol{M}}_1(k)$ given by (28.29), the statistic

$$S_{m-r} = -(T-k) \sum_{j=1}^{m-r} \log(1 - \widehat{\lambda}_j) \tag{28.30}$$

is asymptotically a $\chi^2_{(m-r)^2}$, both for stationary and nonstationary series. Note that we obtain standard distribution because: (1) $\boldsymbol{P}'_\perp \boldsymbol{y}_t$ and $\boldsymbol{P}'_\perp \boldsymbol{y}_{t-1}$ are uncorrelated, and (2) both $\boldsymbol{P}'_\perp \boldsymbol{y}_t$ and $\boldsymbol{P}'_\perp \boldsymbol{y}_{t-1}$ are white noise.

The result of this lemma is in the line of Robinson (1973) to test for zero canonical correlation of stationary time series. This result was modified by Tiao and Tsay (1989) to test for SCM, 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 $1/(T-k)$. Hu and Chou (2004) proposed a similar test using several second moment matrices simultaneously but instead of using canonical correlation between $\boldsymbol{y}_t$ and its past and future in order to check the rank of the second moment matrices, they use canonical correlation twice: once between $\boldsymbol{y}_t$ and its past and future in order to define past and future canonical variates and a second time between $\boldsymbol{y}_t$ and the canonical variates define in the previous step. This means that while we are interested in the rank of the matrices defined in (28.26), they test for the rank of the matrices defined by $\boldsymbol{Q} = \boldsymbol{M}\boldsymbol{\Gamma}_y(k)\boldsymbol{M}'$ which have $m-r$ eigenvalues equal to zero if $\boldsymbol{M} = [\boldsymbol{P}'_\perp \ \boldsymbol{P}]$. Note that the test presented in this section leads to standard distribution in contrast to the ones presented in 28.3.2, as Johansen (1988) test for the cointegration rank of a VAR model and Reinsel and Ahn (1992) for the number of unit roots in reduced rank regression models.

## 28.4 Real Data Analysis

We study seven monthly stock indexes from November 1990 until April 2000. The indexes are (by alphabetical order as they are collected in the vector of time $\boldsymbol{y}_t$) DAX-30 from Germany, Dow Jones Composite (DJCOM) form the USA, FTSE from United Kingdom, NASDAQ, New York Stock Exchange (NYSE), Standard and Poor 500 (SP500) from USA, and the Canadian TSE. In order to correct for heteroskedasticity, we take the natural log of all the indexes. Plots of the logs of these indices are shown in Figure 28.1.

We apply the common factors canonical correlation test of Section 28.3.3 and obtain the results shown in Table 28.1. We have used up to 18 lags to show that the number of identified factors does not depend on the upper bound used

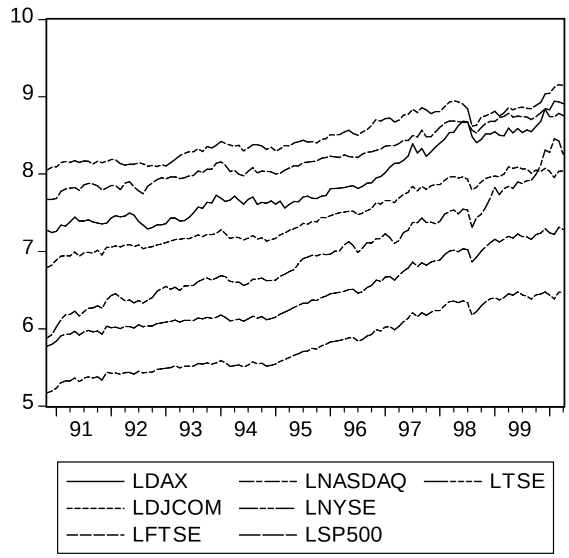


Figure 28.1: Logs of the monthly stock indexes

for the number of lags. The statistics have already been divided by their critical values, so that a number greater than 1 means that we reject the null hypothesis of a maximum of $r$ common factors at the usual 5% significance level, while a number smaller than 1 means that we cannot reject the null hypothesis of a maximum of $r$ common factors. We present the results after lag two because some small correlations are found for lags 1 and 2. The outcome of the test indicates that a maximum of 6 common factors cannot be rejected.

To obtain the factors we build the generalized covariance matrices for lags one to five and extract the eigenvectors associated to the first common six eigenvalues of each matrix.

The first common factor is a weighted mean of all the indexes, and it can be interpreted as the general level of the world stock indexes. The second factor differentiates the behavior of the Nasdaq, the NYSE and the SP500 from the Canadian TSE and the British FTSE. The third factor separates the NASDAQ and British FTSE from the others. The fourth and sixth common factors are mainly assigned to a single index to characterize its differential performance (the fourth common factor to the German DAX and the sixth to the Chicago's SP500). Finally, the fifth common factor differentiates the British FTSE from the TSE.

In order to obtain the dynamics of the factors, we can perform univariate analysis over the linear combinations of the stock indexes given by the common eigenvectors (we have chosen the eigenvector associated to the generalized co-

Table 28.1: Outcome of the test of Section 28.3.3 for the number of factors. The statistics have already been divided by their critical value, so that an outcome greater than 1 means that the null of a maximum of $r$ common factors is rejected at the 5% significance level, while an outcome smaller than 1 means that the null of a maximum of $r$ common factors cannot be rejected at the 5% significance level

| | lag $k$ | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $r$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 0 | 32.6 | 30.5 | 29.5 | 29.0 | 28.0 | 27.0 | 26.3 | 25.7 | 25.8 | 25.4 | 25.1 | 24.6 | 24.2 | 24.9 | 25.0 | 24.7 |
| 1 | 18.8 | 16.9 | 16.1 | 15.9 | 15.0 | 13.8 | 13.0 | 12.4 | 12.8 | 12.6 | 12.2 | 11.6 | 11.2 | 11.9 | 12.2 | 12.0 |
| 2 | 12.8 | 10.7 | 10.0 | 9.9 | 9.0 | 7.8 | 6.9 | 6.4 | 6.6 | 6.7 | 6.5 | 6.4 | 6.7 | 7.4 | 7.3 | 6.9 |
| 3 | 10.0 | 7.8 | 6.9 | 7.2 | 7.0 | 6.3 | 5.4 | 4.5 | 4.1 | 4.1 | 4.4 | 4.9 | 5.3 | 5.5 | 5.6 | 5.4 |
| 4 | 6.2 | 3.9 | 3.4 | 4.5 | 4.2 | 3.3 | 2.4 | 1.8 | 1.7 | 1.7 | 1.8 | 2.1 | 2.8 | 3.1 | 3.3 | 3.7 |
| 5 | 2.5 | 1.5 | 1.6 | 3.2 | 3.4 | 1.8 | 1.0 | 1.2 | 1.1 | 1.0 | 1.0 | 1.1 | 1.2 | 1.0 | 1.2 | 1.3 |
| 6 | 0.7 | 0.05 | 0.5 | 0.5 | 0.05 | 0.09 | 0.10 | 0.2 | 0.7 | 0.5 | 0.04 | 0.02 | 0.04 | 0.08 | 0.2 | 0.06 |

Table 28.2: Eigenvectors associated to the first and second eigenvalues for the first five generalized covariance matrices of the stock indexes data

| 1st eigenvector | | | | | 2nd eigenvector | | | | |
|---|---|---|---|---|---|---|---|---|---|
| lag $k$ | | | | | lag $k$ | | | | |
| 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 0.40 | 0.40 | 0.40 | 0.40 | 0.40 | 0.10 | 0.10 | 0.10 | 0.10 | 0.107 |
| 0.38 | 0.38 | 0.38 | 0.38 | 0.38 | -0.09 | -0.08 | -0.06 | -0.05 | -0.05 |
| 0.42 | 0.42 | 0.42 | 0.42 | 0.42 | -0.36 | -0.36 | -0.35 | -0.35 | -0.35 |
| 0.36 | 0.36 | 0.36 | 0.36 | 0.36 | 0.64 | 0.62 | 0.60 | 0.58 | 0.57 |
| 0.30 | 0.30 | 0.30 | 0.30 | 0.30 | 0.21 | 0.22 | 0.24 | 0.25 | 0.26 |
| 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.28 | 0.29 | 0.31 | 0.31 | 0.32 |
| 0.43 | 0.43 | 0.43 | 0.43 | 0.43 | -0.56 | -0.57 | -0.59 | -0.60 | -0.60 |

Table 28.3: Eigenvectors associated to the 3rd and 4th eigenvalues for the first five generalized covariance matrices of the stock indexes data

| 3rd eigenvector | | | | | 4th eigenvector | | | | |
|---|---|---|---|---|---|---|---|---|---|
| lag $k$ | | | | | lag $k$ | | | | |
| 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 0.36 | 0.62 | 0.77 | 0.78 | 0.73 | 0.88 | 0.86 | 0.87 | 0.86 | 0.84 |
| 0.20 | 0.04 | - 0.10 | - 0.13 | - 0.12 | -0.28 | -0.22 | -0.30 | -0.35 | -0.36 |
| - 0.10 | - 0.21 | - 0.28 | - 0.33 | - 0.42 | -0.21 | -0.24 | -0.25 | -0.24 | -0.27 |
| - 0.66 | - 0.62 | - 0.52 | - 0.49 | - 0.48 | - 0.29 | - 0.38 | - 0.25 | - 0.16 | - 0.12 |
| 0.39 | 0.29 | 0.17 | 0.15 | 0.20 | 0.09 | 0.00 | - 0.09 | - 0.16 | 0.20 |
| 0.30 | 0.20 | 0.08 | 0.05 | 0.06 | - 0.06 | - 0.01 | - 0.10 | - 0.15 | - 0.16 |
| - 0.37 | - 0.25 | - 0.12 | - 0.04 | - 0.05 | -0.02 | -0.05 | 0.04 | 0.09 | 0.15 |

Table 28.4: Eigenvectors associated to the 5th and 6th eigenvalues for the first five generalized covariance matrices of the stock indexes data

| 5th eigenvector | | | | | 6th eigenvector | | | | |
|---|---|---|---|---|---|---|---|---|---|
| lag $k$ | | | | | lag $k$ | | | | |
| 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 0.09 | 0.07 | 0.02 | 0.01 | 0.09 | 0.17 | 0.17 | 0.18 | 0.18 | 0.19 |
| - 0.11 | - 0.12 | - 0.11 | - 0.13 | - 0.18 | 0.41 | 0.27 | 0.23 | 0.21 | 0.20 |
| 0.76 | 0.76 | 0.76 | 0.74 | 0.67 | 0.06 | -0.04 | -0.16 | -0.22 | -0.20 |
| 0.12 | 0.13 | 0.16 | 0.21 | 0.29 | 0.11 | 0.11 | 0.11 | 0.12 | 0.13 |
| - 0.31 | - 0.35 | - 0.39 | - 0.44 | - 0.54 | 0.33 | 0.42 | 0.38 | 0.35 | 0.35 |
| - 0.13 | - 0.09 | - 0.05 | - 0.06 | - 0.14 | - 0.81 | - 0.82 | - 0.81 | - 0.81 | - 0.81 |
| - 0.52 | - 0.50 | - 0.48 | - 0.44 | - 0.34 | -0.15 | -0.18 | - 0.26 | - 0.30 | - 0.27 |

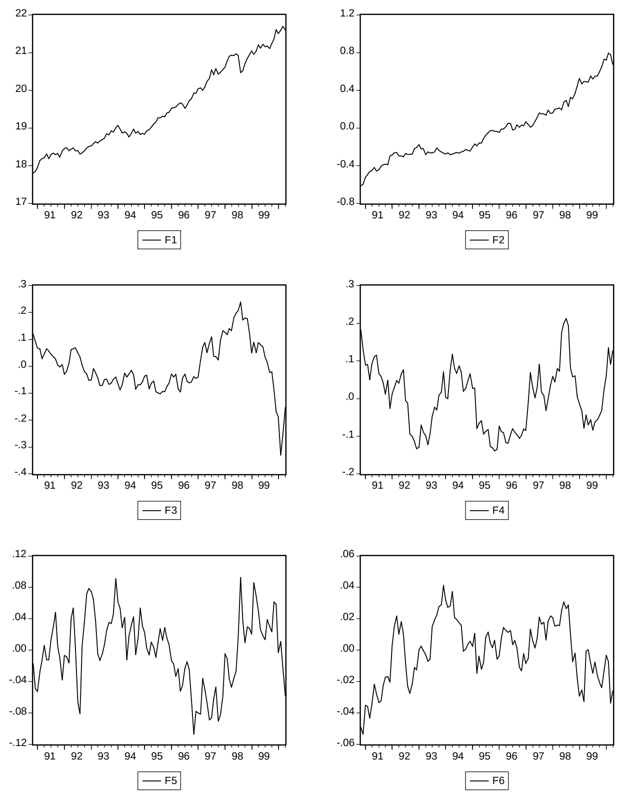


Figure 28.2: Plots of the six common factors of the stock indexes

variance matrix of lag one since it is built with more data than the remaining ones). Plots of the six common factors are shown in Figure 28.2.

From the plot, we see that the first two common factors, and possibly the third one, are nonstationary. In fact, if we apply the Augmented Dickey-Fuller unit root test with automatic lag selection to minimize the Schwarz information criterion, we cannot reject a unit root for the first three factors, the p-value of the test for the fourth factor is 0.0688 and it is clearly rejected for the fifth and sixth common factors. The first three factors are random walks. The fourth factor could be considered as an AR(1) with autoregressive parameter very close to one (it is estimated as 0.9). The fifth and sixth common factors can be modeled as stationary autoregressive processes of order 2 and 1, respectively.

This analysis shows that the dimension of the nonstationary subspace for the 7 stock indexes can be reduced to 3 or, at most, 4.

From this analysis, we expect to find three or four cointegration relations (number of series minus number of common trends) if we apply Johansen's cointegration test. All the selection criteria (Akaike, Schwartz, Hannan-Quin, maximum likelihood and forecasting prediction error) indicate the order of the VAR process in levels should be 1. With this in mind, we perform Johansen's cointegration test and for a significance level of $\alpha = 0.05$. Assuming that there is no deterministic trends in the data and using the trace statistic of Section 28.3.2, we found 4 cointegration relations, which is in agreement with the factor analysis results. It is interesting to check that if we assume a deterministic trend, which we believe is a rather unusual fact with economic data [see, Peña (1995)], the number of cointegration relationships found is zero. In order to check the robustness of this conclusion, we perform the maximum eigenvalue test of Johansen, which tests the null hypothesis of $s$ cointegrating relations against the alternative of $s+1$ cointegrating relations. This test statistic is computed as (being the eigenvalues as the same ones as in Section 28.3.2)

$$L(s|s+1) = -T \log(1 - \widehat{\lambda}_{s+1}), \tag{28.31}$$

but now we found zero cointegration relations. This result is also obtained if we assume the rather unlikely assumption of deterministic trends in the data. When computing the roots of the companion matrix of the VAR process, one root very close to 1 (estimated as 0.998) and three (a real one and a pair of complex ones) of modulus 0.97 and 0.93 are found. The remaining roots are not close to 1. This might explain why the different versions of the tests detect from 0 to 4 cointegration relations, depending on the assumptions made in order to perform the test.

## 28.5 Concluding Remarks

We have shown in this paper that canonical correlation analysis between the present and past values of the time series is a very powerful tool for dimension reduction. This approach allows a unified view of many of the procedures proposed for dimension reduction, including principal components, the canonical analysis of Box and Tiao, the reduced rank models of Reinsel *et al.*, the scalar component models of Tiao and Tsay, the Dynamic Factor model, and state space models. Canonical correlation offers also a unifying view for dimension reduction tests and will lead to similar results than principal components tests when the innovation covariance matrix of the time series is close to a scalar matrix $\sigma^2 \boldsymbol{I}$.

**Acknowledgments.** This research has been supported by Cátedra BBVA de Calidad, DGES projects SEJ2004-03303 and BEC2002-00081, and CAM project 06/HSE/0016/2004.

---

## References

1. Ahn, S. K. and Reinsel, G. C. (1988). Nested reduced-rank autoregressive models for multiple time series, *Journal of the American Statistical Association*, **83**, 849–856.

2. Ahn, S. K. and Reinsel, G. C. (1990). Estimation for partially nonstationary multivariate autoregressive models, *Journal of the American Statistical Association*, **85**, 813–823.

3. Ahn, S. K. (1997). Inference of vector autoregressive models with cointegration and scalar components, *Journal of the American Statistical Association*, **92**, 350–356.

4. Akaike, H. (1974). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes, *Annals of the Institute of Statistical Mathematics*, **26**, 363–387.

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

6. Aoki, M. (1987). *State Space Modeling of Time Series*, Springer-Verlag, Heidelberg.

7. Banarjee, A. , Dolado, J., Galbraith, J. W. and Hendry, D. (1993). *Cointegration, Error Correction and the Econometric Analysis of Non-Stationary Data*, Oxford University Press, Oxford.

8. Box, G. and Tiao, G. (1977). A canonical analysis of multiple time series, *Biometrika*, **64**, 355–365.

9. Brillinger, D. R. (1981). *Time Series Data Analysis and Theory*, Expanded edition, Holden-Day, San Francisco.

10. Casals J., Jerez, M. and Sotoca, S. (2002). An exact multivariate model-based structural decomposition, *Journal of the American Statistical Association*, **97**, 533–564.

11. Durbin, J. and Koopman, S. K. (2001). *Time Series Analysis by State Space Models*, Oxford University Press, Oxford.

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

13. Escribano, A. and Peña, D. (1994). Cointegration and common factors, *Journal of Time Series Analysis*, **15**, 577–586.

14. Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation, *The Review of Economic and Statistics*, **82**, 540–554.

15. Geweke, J. F. and Singleton K. J. (1981). Maximum likelihood confirmatory analysis of economic time series, *International Economic Review*, **22**, 37–54.

16. Hannan, E. J. and Deistler, M. (1988). *The Statistical Analysis of Linear Systems*, John Wiley & Sons, New York.

17. Harvey, A. (1989). *Forecasting Structural Time Series Models and the Kalman Filter*, Second edition, Cambridge University Press, Cambridge.

18. Harvey, A. (2001). Testing in unobserved components models, *Journal of Forecasting*, **20**, 1–19.

19. Harris, D. (1997). Principal components analysis of cointegrated time series, *Econometric Theory*, **13**, 529–557.

20. Hu, Y. and Chou, R. (2004). On the Peña-Box model, *Journal of Time Series Analysis,* **25**, 811–830.

21. Johansen, S. (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, **12**, 231–254.

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

23. Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

24. Molenaar, P. C. M., De Gooijer, J. G. and Schmitz, B. (1992). Dynamic factor analysis of nonstationary multivariate time series, *Psychometrika*, **57**, 333–349.

25. Peña, D. (1995). Forecasting growth with time series models, *Journal of Forecasting*, **14**, 97–105.

26. Peña, D. and Box, G. (1987). Identifying a simplifying structure in time series, *Journal of the American Statistical Association*, **82**, 836–843.

27. Peña, D. and Poncela, P. (2004). Forecasting with nonstationary dynamic factor models, *Journal of Econometrics*, **119**, 291–321.

28. Peña, D. and Poncela, P. (2006). Nonstationary dynamic factor analysis, *Journal of Statistical Planning and Inference*, (to appear).

29. Peña, D., Tiao. G. C. and Tsay, R. S. (2001). *A Course in Time Series Analysis*, John Wiley & Sons, New York.

30. Quenouille, M. H. (1968). *The Analysis of Multiple Time Series*, Charles Griffin, London.

31. Priestley, M. B., Rao, T. S. and Tong, J. (1974). Applications of principal component analysis and factor analysis in the identification of multivariable systems, *IEEE Transactions on Automation and Control*, **19**, 703–704.

32. Rechner, A. C. (1995). *Methods of Multivariate Analysis*, John Wiley & Sons.

33. Reinsel G. C. and Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: estimation, likelihood ratio tests and forecasting, *Journal of Time Series Analysis*, **13**, 353–375.

34. Reinsel, G. C. and Velu, R. P. (1998). *Multivariate Reduced-Rank Regression*, Springer, New York.

35. Robinson, P. M. (1973). Generalized canonical analysis for time series, *Journal of Multivariate Analysis*, **3**, 141–160.

36. Shumway, R. H. and Stoffer, D. S. (2000). *Time Series Analysis and its Applications*, Springer, New York.

37. Stock, J. H. and Watson, M. W. (1988). Testing for common trends, *Journal of the American Statistical Association*, **83**, 1097–1107.

38. Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series, *Journal of the Royal Statistical Society, Series B*, **51**, 157–213.

39. Tsay, R. S. and Tiao, G. C. (1990). Asymptotic properties of multivariate nonstationary processes with applications to autoregressions, *The Annals of Statistics*, **18**, 220–250.

40. Velu, R. P., Reinsel, G. C. and Wichern, D. W. (1986). Reduced rank models for multiple time series, *Biometrika*, **73**, 105–118.