Revista Colombiana de Estadística - Theoretical Statistics
January 2021, volume 44, issue 1, pp. 1 a 21
http://dx.doi.org/10.15446/rce.v44n1.84816

On a New Procedure for Identifying a Dynamic Common Factor Model

Sobre un nuevo procedimiento para identificar un modelo de factores comunes dinámicos

Stevenson Bolívar[1,a], Fabio H. Nieto[2,b], Daniel Peña[3,c]

[1]Department of Industrial Engineering, Pontificia Universidad Javeriana, Bogotá, Colombia

[2]Department of Statistics, Universidad Nacional de Colombia, Bogotá, Colombia

[3]Department of Statistics, Universidad Carlos III de Madrid, Madrid, España

Abstract

In the context of the exact dynamic common factor model, canonical correlations in a multivariate time series are used to identify the number of latent common factors. In this paper, we establish a relationship between canonical correlations and the autocovariance function of the factor process, in order to modify a pre-established statistical test to detect the number of common factors. In particular, the test power is increased. Additionally, we propose a procedure to identify a vector ARMA model for the factor process, which is based on the so-called simple and partial canonical autocorrelation functions. We illustrate the proposed methodology by means of some simulated examples and a real data application.

***Key words*:** Canonical correlations; Dynamic common factors; Multivariate time series.

Resumen

En el contexto del modelo exacto de factores comunes dinámicos, las correlaciones canónicas en series de tiempo multivariadas son usadas para identificar el número de factores latentes. En este artículo, establecemos la relación entre correlación canónica y la función de autocovarianza del proceso de los factores, con el fin de modificar una prueba estadística diseñada para identificar el número de factores comunes. En particular, se incrementa la potencia de la prueba. Adicionalmente, proponemos un procedimiento

[a]M.Sc. E-mail: s_bolivar@javeriana.edu.co
[b]Ph.D. E-mail: fhnietos@unal.edu.co
[c]Ph.D. E-mail: daniel.pena@uc3m.es

para identificar el modelo VARMA para el proceso de los factores, el cual está basado en lo que denominamos las funciones de autocorrelación simple y parcial. Ilustramos la metodología propuesta por medio de ejemplos simulados y una aplicación con datos reales.

***Palabras clave*:** Correlación canónica; Factores comunes dinámicos; Series de tiempo multivariadas.

1. Introduction

The dynamic common factor model is of interest when an observable multivariate stochastic process $\{y_t\}$, of dimension m, is generated by an unobservable stochastic process $\{f_t\}$, of dimension r, with $r < m$, via the equation

$$y_t = Pf_t + \varepsilon_t, \tag{1}$$

where $\{\varepsilon_t\}$ is a multivariate stationary noise process of dimension m, with mean $\mathbf{0}$ and variance Σ_ε, and P is an $m \times r$ matrix, known as the loading matrix or the weight matrix. The components of the f_t vector are the common factors of the y_t data vector, while ε_t contains their specific or idiosyncratic components. A crucial aspect in the analysis of this type of models is the identification of r, the number of common factors. Several studies have been carried out in this sense, mainly in the high-dimension scenario, in which it is assumed that m goes to infinity. For complete details, see the papers, among others, of Stock & Watson (2011), Lam & Yao (2012), Ahn & Horenstein (2013), and their associated references.

This paper is based on the low-dimension approach (fixed m) proposed by Peña & Box (1987) (from this point forward Peña-Box model), when the $\{f_t\}$ process is stationary and $\{\varepsilon_t\}$ is white noise, which is known as the Exact Dynamic Factor Model (EDFM) because its specific components are orthogonal processes. Peña & Poncela (2006) extend this paper to the case where $\{f_t\}$ is non-stationary and design a statistical test to specify r. This statistical test is a function of the canonical correlations between y_t and y_{t-k}, for some lag k, $k = 1, 2, \ldots$ Under the null hypothesis of r factors, the authors show that this statistic is distributed asymptotically as a $\chi^2_{(m-r)^2}$, for each lag k. Nevertheless, as is shown in Section 2, at each lag k the test detect the rank of the covariance matrix of f_t and f_{t-k}, but this matrix may not have full rank. Then, this alternative may detects less than r common factors, which imply that the power of the test is reduced. Indeed, let us consider this example: let $f_t = (f_{1t}, f_{2t})^\top$, where $f_{1t} = a_{1t} + \theta_1^{(1)} a_{1,t-1}$ and $f_{2t} = a_{2t} + \theta_3^{(2)} a_{2,t-3}$, with $\{a_{it}\}$, $i = 1, 2$, white noise processes. The number of factors is 2 but the rank of matrix $Cov(f_t, f_{t-k})$ is 1 if $k = 1$ or $k = 3$ and zero otherwise. This fact makes that the Peña & Poncela's (2006) test does not detect the correct number of factors at any lag $k = 1, 2, \ldots$

In this paper, we develop a modification of the Peña & Poncela's (2006) statistical test, which avoids the low-power problem quoted above. Essentially, the idea is to use in the calculation of the test statistic, the canonical correlations between y_t and a linear combination of some lagged vectors y_{t-i}, for $i \in \{1, 2, \ldots\}$.

Also, we propose a methodology to identify ARMA models for the factor processes, using the canonical and partial canonical correlation to identify the dependence orders of the factor models. We only consider the case in which $\{f_t\}$ is stationary.

It is important to highlight that in the Peña-Box model the observed data vector y_t depends on f_t contemporaneously. Then, the model is known as the DFM in static form. In the DFM in dynamic form, initially proposed by Geweke (1977) (cited by Stock & Watson (2016); Doz & Fuleky (2020)), the vector of observed time series can also depend on different lags of the common factors. Therefore, our contributions are based on the DFM in the static form.

The paper is organized as follows. Section 2 presents the basic results that relate canonical correlations to the marginal autocovariance functions of the factors processes and their use in the modification of Peña & Poncela's (2006) statistical test. In section 3, we present a procedure to identify the ARMA factor models. Section 4 includes some simulated examples and an empirical application to precipitation data in Colombia. Section 5 concludes.

2. An Extension of the Statistical Test

For each t, let $f_t = (f_{1t}, \dots, f_{rt})^\top$, where $\top$ denotes the matrix transpose operation, and let $\gamma_i(k)$ be the i-th component on the diagonal of the autocovariance matrix of $\{f_t\}$ at lag k. Following Nieto et al.'s (2016) assumptions, in particular that the marginal processes $\{f_{it}\}$ and $\{f_{jt}\}$ are orthogonal for all $i, j = 1, 2, \dots, r$, with $i \neq j$, the components outside of the diagonal are zero. In addition, the restriction $P^\top \Sigma_\varepsilon^{-1} P = I_r$ is imposed to solve the identification problem (Peña & Poncela 2006), where I_n is the $n \times n$ identity matrix. The remaining notation will be defined as new definitions are introduced.

From Peña & Poncela's (2006) paper, we consider the random matrix

$$\widehat{M}(k) = \left[\sum_{t=k+1}^{T} (y_t y_t^\top)\right]^{-1} \sum_{t=k+1}^{T} (y_t y_{t-k}^\top) \left[\sum_{t=k+1}^{T} (y_{t-k} y_{t-k}^\top)\right]^{-1} \sum_{t=k+1}^{T} (y_{t-k} y_t^\top), \quad (2)$$

where T denotes the sample size of an observed multivariate time series of the process $\{y_t\}$, and the statistic

$$S_{m-r}(k) = -(T-k) \sum_{j=1}^{m-r} log(1 - \widehat{\lambda}_j(k)), \quad (3)$$

where $\widehat{\lambda}_1(k) \leq \widehat{\lambda}_2(k) \leq \cdots \leq \widehat{\lambda}_m(k)$ are the ordered eigenvalues of the matrix $\widehat{M}(k)$, for a given lag $k = 1, 2, \dots$ The limit distributions of these statistics are obtained by Peña & Poncela (2006). In particular, they find that, when the process $\{f_t\}$ is stationary, the limit matrix of the sequence $\widehat{M}(k)$ is

$$M(k) = \left(PE[f_t f_t^\top]P^\top + \Sigma_\varepsilon\right)^{-1} PE[f_t f_{t-k}^\top]P^\top \left(PE[f_{t-k} f_{t-k}^\top]P^\top + \Sigma_\varepsilon\right)^{-1} PE[f_{t-k} f_t^\top]P^\top.$$

Before establishing our main results, it is worth noticing the following fact: in Peña & Poncela's (2006) Theorem 3, it is claimed that the limit matrix of the sequence $\{\widehat{M}(k)\}$, indexed by T, has rank r for all lags k. However, we have found that for some models and some lags k the rank is less than r. As mentioned before, let us consider this example: let $f_t = (f_{1t}, f_{2t})^\top$, where $f_{1t} = a_{1t} + \theta_1^{(1)} a_{1,t-1}$ and $f_{2t} = a_{2t} + \theta_3^{(2)} a_{2,t-3}$, with $\{a_{it}\}$, $i = 1, 2$, white noise processes. The number of factors is 2, but the rank of matrix $E[f_t f_{t-k}^\top]$ is 1 when $k = 1$ or $k = 3$ and zero otherwise. This fact makes that the rank of the limit matrix in Peña & Poncela's (2006) Theorem 3 depends on k. From now on, we denote the rank of that matrix as $r(k)$, $k = 1, 2, \dots$

To avoid this drawback of the methodology, we propose the following results.

Proposition 1. *Let $\{f_t\}$ be a stationary stochastic process, then, given the lag k, the $r(k) \leq r$ non-zero eigenvalues of $M(k)$ belong to the set*

$$\Lambda(k) := \left\{ \mu_j^2(k) \in \boldsymbol{R} : \mu_j(k) = \frac{|\gamma_j(k)|}{\gamma_j(0) + 1},\ j = 1, 2, \dots, r \right\}, \tag{4}$$

where $\mu_j(k)$ is a canonical correlation between y_t and y_{t-k}, with respective canonical variables $a_j^\top y_t$ and $b_j^\top y_{t-k}$ where

$$b_j = \frac{\Sigma_\varepsilon^{-1} P_j}{(\gamma_j(0) + 1)^{1/2}} \in \boldsymbol{R}^m \qquad \text{and} \qquad a_j = sign\{\gamma_j(k)\} b_j. \tag{5}$$

Here $\boldsymbol{R}$ and $\boldsymbol{R}^m$ denote, respectively, the set of real numbers and the m-dimensional real Euclidean space and P_j is the j-th column of the load matrix P.

Proof. See Appendix A.1. □

Note that the largest value of the set $\Lambda(k)$ is the first canonical correlation, the second largest value of $\Lambda(k)$ is the second canonical correlation and so on up to the $r(k)$-th canonical correlation (Anderson 1984). Additionally, note that $r(k) = r$ if $|\gamma_j(k)| > 0$ for all $j = 1, \dots, r$ and the $m - r$ remaining canonical correlations are equal to zero.

An important implication of Proposition 1 is the following: if for a given $k \neq 0$, $\gamma_i(k) = 0$ for some $i = 1, \dots, r$, then, at this lag, there is a maximum of $r - 1$ non-zero canonical correlations. This fact is in line with the above comment about the loss of power of the test. In order to improve the performance of Peña & Poncela's (2006) test, we propose to use the canonical correlations between y_t and $y_t^\dagger = \sum_{k \in \boldsymbol{K}} y_{t-k}$, where $\boldsymbol{K}$ is a set of lags (not necessarily consecutive). To obtain $\boldsymbol{K}$, we propose the following two-step procedure: first, we run Peña and Poncela's test for lags $k = 1, 2, \dots, k_0$, for some k_0. Then, a lag $k \in \boldsymbol{K}$ if the Peña & Poncela's (2006) test detects at least one common factor at it. In the examples below, we will give more suggestions to obtain $\boldsymbol{K}$ in practice.

Proposition 2. *Given $\boldsymbol{K}$ and the restriction $P^{\top}\Sigma_{\varepsilon}^{-1}P = I_r$, the r non-zero square canonical correlations between y_t and $y_t^{\dagger}$ belong to the set*

$$\Lambda^{\dagger}(\boldsymbol{K}) := \left\{ \mu_j^{2\dagger}(\boldsymbol{K}) \in \boldsymbol{R} : \mu_j^{\dagger}(\boldsymbol{K}) = \frac{|\gamma_j(\boldsymbol{K})|}{(\gamma_j(0)+1)^{1/2}(\gamma_{j}^{\dagger} + k^{\dagger})^{1/2}},\ j = 1, 2, \ldots, r \right\}, \quad (6)$$

and its respective canonical variables are given by $b_j^{\top} y_t$ and $a_j^{\dagger\top} y_t^{\dagger}$, with

$$b_j = \frac{\Sigma_{\varepsilon}^{-1}P_j}{(\gamma_j(0)+1)^{1/2}} \qquad \text{and} \qquad a_j^{\dagger} = sign\{\gamma_j(\boldsymbol{K})\}\frac{\Sigma_{\varepsilon}^{-1}P_j}{(\gamma_{j^{\dagger}} + k^{\dagger})^{1/2}}, \quad (7)$$

where $k^{\dagger}$ is the cardinality of $\boldsymbol{K}$, $\gamma_j(\boldsymbol{K}) = \sum_{k\in\boldsymbol{K}} \gamma_j(k)$ and $\gamma_{j}^{\dagger} = \sum_{k_1\in\boldsymbol{K}}\sum_{k_2\in\boldsymbol{K}} \gamma_j(|k_2 - k_1|)$, $j = 1, 2, \ldots, r$.

***Proof*.** See Appendix A.2. □

Now, we define

$$M^{\dagger}(\boldsymbol{K}) = \left[\sum_{t=\bar{k}+1}^{T}(y_t y_t^{\top})\right]^{-1} \sum_{t=\bar{k}+1}^{T}(y_t y_t^{\dagger\top}) \left[\sum_{t=\bar{k}+1}^{T}(y_t^{\dagger} y_t^{\dagger\top})\right]^{-1} \sum_{t=\bar{k}+1}^{T}(y_t^{\dagger} y_t^{\top}), \quad (8)$$

where $\bar{k} = max\ \boldsymbol{K}$, and obtain the following result.

Proposition 3. *Let $\Gamma_f(k) = E[f_t f_{t-k}^{\top}]$. If $\sum_{k\in\boldsymbol{K}} \Gamma_f(k)$ has rank r, then, as T goes to infinity, the matrix sequence $\{M^{\dagger}(\boldsymbol{K})\}$ converges in distribution to a constant matrix that has rank r.*

***Proof*.** Using $Q = \Sigma_{\varepsilon}^{-1}[PP_{\perp}]$, where $P_{\perp}$ is such that $P^{\top}P_{\perp} = 0$ and $P_{\perp}^{\top}P_{\perp} = I_{m-r}$, the proof follows the basic ideas in Peña & Poncela's (2006) paper (pages 1954-1955). □

Additionally, bearing in mind that the eigenvalues of $M^{\dagger}(\boldsymbol{K})$ are the square canonical correlations between y_t and $y_t^{\dagger}$, we obtain the following proposition.

Proposition 4. *Let $\widehat{\mu}_1^{2\dagger}(\boldsymbol{K}) \leq \widehat{\mu}_2^{2\dagger}(\boldsymbol{K}) \leq \cdots \leq \widehat{\mu}_m^{2\dagger}(\boldsymbol{K})$ be the ordered eigenvalues of the matrix $M^{\dagger}(\boldsymbol{K})$. Then, under the null hypothesis that the limit matrix in Proposition 3 has $m-r$ eigenvalues equal to zero, the asymptotic distribution of the statistic*

$$S_{m-r}(\boldsymbol{K}) = -(T-\bar{k})\sum_{j=1}^{m-r} log(1 - \widehat{\mu}_j^{2\dagger}(\boldsymbol{K})) \quad (9)$$

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

Proof. According to Proposition 2, there are $m-r$ canonical correlations equal to zero between y_t and $y_t^{\dagger}$. This means that there are $m-r$ linear combinations of y_t that are not correlated to $m-r$ linear combinations of $y_t^{\dagger}$; consequently, there are $m-r$ regressions of the form $a_j^{\top} y_t = W_j^{\top} y_t^{\dagger} + u_t$ where $W_j = 0_{m\times 1}$ for each $j = 1, \ldots, m-r$, where the a_j's are the canonical vectors associated to the $m-r$ canonical correlations equal to zero. The rest of the proof follows the lines of the corresponding result in Peña & Poncela's (2006) paper. We note that $T-\bar{k}$ is the number of observations that is used to compute the canonical correlations. □

Remark: To compute $\gamma_j(\boldsymbol{K})$, we sum over all the lags in $\boldsymbol{K}$ the autocovariances of the same jth factor (in the order established at the beginning of this section). It may happen that we found autocovariances of the same absolute value but different sign and thus, $\gamma_j(\boldsymbol{K})$ might be equal to zero, even if for some $k \in \boldsymbol{K}$, $\gamma_j(k) \neq 0$. Therefore, instead of using only the sum over all the lags in $\boldsymbol{K}$, an alternative is to use different linear combinations to calculate $y_t^{\dagger}$. Indeed, to define a linear combination, a lag in $\boldsymbol{K}$ is kept constant (1 is the coefficient) and the coefficients corresponding to the other lags alternate between $+1$ and -1. In this way, $2^{(k^{\dagger}-1)}$ possible linear combinations can be analyzed. The main idea behind the use of different linear combinations $y_t^{\dagger}$ is to avoid the cases, in which the number of factors is underestimated; therefore, we propose to choose r as the maximum number of factors detected with the different linear combinations. The examples below illustrate this point.

In summary, the procedure for specifying r is the following:

STEP 1 Set the maximum number of lags k_0 and run Peña and Poncela's test for lags $k = 1, 2, \ldots, k_0$.

STEP 2 Set $\boldsymbol{K}$ such as a lag $k \in \boldsymbol{K}$ if the Peña and Poncela's test detects at least one common factor at it, $k = 1, 2, \ldots, k_0$.

STEP 3 Define the $2^{(k^{\dagger}-1)}$ possible linear combinations $y_t^{\dagger}$, keeping a lag constant (1 is the coefficient) and the coefficients corresponding to the other lags alternate between $+1$ and -1.

STEP 4 Run the test (4) with each linear combination $y_t^{\dagger}$ and choose r as the maximum number of factors detected with the different linear combinations.

A simulated example

To illustrate the issues remarked above about the statistical test performance, we conduct a Monte Carlo experiment. The design of the simulation is the following: we set $m = 6$ and $r = 2$ and consider the factor model

$$y_t = \begin{bmatrix} 1.00 & 0.00 \\ 1.00 & 1.00 \\ 0.00 & 1.00 \\ 1.00 & 0.00 \\ -1.00 & 1.00 \\ 0.00 & -1.00 \end{bmatrix} f_t + \varepsilon_t, \tag{10}$$

where $f_t = (f_{1t}, f_{2t})^\top$, $\{\varepsilon_t\} \sim WN(0, I_6)$, $f_{1t} = a_{1t} + 0.8a_{1,t-1}$ and $f_{2t} = a_{2t} - 0.7a_{2,t-3}$, with $\{a_t = (a_{1t}, a_{2t})^\top\} \sim WN(0, I_2)$. We put $T = 1000$ and simulate 1000 time series from the process $\{y_t\}$. For each generated time series and fixing a significance level of 5%, Peña and Poncela's test was run sequentially as a test on the maximum number of factors and we stop and identify r factors as soon as the hypothesis is rejected, which will be denoted r' from now on.

Table 1 presents the percentage of times (in the cells) that the test identifies r' factors (columns) at each lag k (rows), $k = 1, 2, \dots, 15$. Clearly, we can see that one common factor is identified at lags $k = 1, 3$ and zero factors at the other lags.

TABLE 1: Percentage of times that the test identifies r' factors in the simulated model.

	$r' = 0$	$r' = 1$	$r' = 2$	$r' = 3$
$k = 1$	0.0	**96.0**	3.8	0.2
$k = 2$	**93.0**	6.8	0.2	0.0
$k = 3$	0.0	**93.8**	5.6	0.6
$k = 4$	**93.4**	6.2	0.4	0.0
$k = 5$	**92.4**	7.3	0.3	0.0
$k = 6$	**93.6**	6.2	0.2	0.0
$k = 7$	**92.2**	7.4	0.4	0.0
$k = 8$	**93.2**	5.9	0.9	0.0
$k = 9$	**94.0**	5.8	0.2	0.0
$k = 10$	**93.2**	6.7	0.1	0.0
$k = 11$	**94.0**	5.7	0.3	0.0
$k = 12$	**94.4**	5.6	0.0	0.0
$k = 13$	**94.2**	5.7	0.1	0.0
$k = 14$	**93.7**	6.2	0.1	0.0
$k = 15$	**92.8**	6.9	0.3	0.0

For this model, we get $\boldsymbol{K} = \{1, 3\}$ with $k^\dagger = 2$; therefore, one have $2^{2-1} = 2$ possible linear combinations. These are $y_t^\dagger = y_{t-1} + y_{t-3}$ (1) and $y_t^\dagger = y_{t-1} - y_{t-3}$ (2). In Table 2 we present the results obtained using the two linear combinations $y_t^\dagger$. We observe that $r = 2$ is clearly identified for any $y_t^\dagger$.

TABLE 2: Percentage of times that the proposed test identifies r' factors in the simulated model.

Linear combination	$r' = 0$	$r' = 1$	$r' = 2$	$r' = 3$
(1)	0.0	0.0	**96.6**	3.4
(2)	0.0	0.0	**95.3**	4.7

In order to illustrate that the test proposed in this section stays valid even when the covariance $\Gamma_f(k)$ is a full rank matrix for some lag k, we modified the factor models in this example as follow: $f_{1t} = 0.8f_{1,t-1} + a_{1t}$ and $f_{2t} = -0.7f_{2,t-3} + a_{2t}$. Table 3 presents the percentage of times that the test identifies r' factors at each lag k with this simulated AR model, $k = 1, 2, \ldots, 15$.

TABLE 3: Percentage of times that the test identifies r' factors in the simulated model.

	$r'=0$	$r'=1$	$r'=2$	$r'=3$
$k=1$	0.0	76.0	23.5	0.5
$k=2$	0.0	78.8	20.7	0.5
$k=3$	0.0	0.0	**95.5**	4.3
$k=4$	0.0	78.4	21.4	0.2
$k=5$	0.0	79.8	19.4	0.8
$k=6$	0.0	1.4	**94.5**	3.9
$k=7$	4.8	79.7	15.4	0.1
$k=8$	13.3	75.6	10.9	0.2
$k=9$	0.0	30.0	68.2	1.8
$k=10$	27.4	63.3	9.2	0.1
$k=11$	37.4	56.0	6.5	0.1
$k=12$	0.9	57.0	41.2	0.9
$k=13$	45.0	50.4	4.6	0.0
$k=14$	48.2	48.1	3.7	0.0
$k=15$	10.3	66.4	23.3	0.0

We can see that about 95% of times two common factor is identified at lags $k = 3, 6$ and between one a two common factors at lags $k = 1, 2, 4, 5$.

TABLE 4: Percentage of times that the proposed test identifies r' factors in the simulated model.

Linear combination	$r'=0$	$r'=1$	$r'=2$	$r'=3$
(1)	0.0	0.0	96.1	3.8
(2)	0.0	0.0	95.9	4.1

For this model, we use $\boldsymbol{K} = \{1, 3\}$ with $k^\dagger = 2$ with the two linear combinations (1) $y_t^\dagger = y_{t-1} + y_{t-3}$ (1) and (2) $y_t^\dagger = y_{t-1} - y_{t-3}$ (2). In Table 4 we present the results obtained using the two linear combinations $y_t^\dagger$. We observe that $r = 2$ is clearly identified for any $y_t^\dagger$.

3. A Procedure for Identifying the Common-Factors Model

Usually in practice, the factor models are identified using preliminary estimates of the factor processes (see, among others, Peña & Poncela's (2006) and Nieto et al.'s (2016) papers). However, if there is much uncertainty in these estimates, the model identification process might lead to wrong models. To avoid this problem, we propose another alternative to the factors model identification, which consists in using the relationship between canonical correlations and the factor-processes autocovariance function established in Proposition 1.

The main idea is to plot the canonical correlations that are related to the same factor against lag values. Because of Proposition 1, we plot at each lag the absolute value of the autocovariance of factor f_{jt} divided by the constant $\gamma_j(0)+1$, for each $j = 1, \ldots, r$. Loosely speaking, we plot a function that is proportional to the absolute value of the autocorrelation function of the jth factor. For each $j = 1, \ldots, r$, we call the function that corresponds to the jth factor as the jth *canonical autocorrelation function* (CACF).

It is worth noticing that, when the canonical correlations are ordered in descending way, given that each canonical correlation is proportional to the autocovariance of a particular factor (see Proposition 1), a specific order is defined for the components of f_t at each lag, but this ordering on the components of f_t is not necessarily the same at all lags. Then, to avoid this drawback, there are two main goals to achieve. The first one is to define a unique order for the components of f_t and the second, ordering at each lag the canonical correlations based on this particular components order.

From now on, we denote $\mu^2_{j_{i:k}} \in \Lambda(k)$ the ith squared canonical correlation for a given lag k, in descending order of magnitude, and $b_{j_{i:k}}$ denote its related eigenvector, where $i = 1, 2, \ldots, r(k)$ and $j_{i:k} \in \{1, 2, \ldots, r\}$.

Our proposal to define the same order for all lags is based on the following lemmas.

Lemma 1. *For any lag k, the canonical vector related to the factor f_{jt} is given by $b_j := (\gamma_j(0)+1)^{-1/2}\Sigma_\varepsilon^{-1}P_j$, $j = 1, 2, \ldots, r$.*

Proof. Let $B = [b_1, b_2, \ldots, b_r] = \Sigma_\varepsilon^{-1}P(\Gamma_f(0)+I_r)^{-1/2}$, then

$$B^\top y_t = B^\top(Pf_t + \varepsilon_t) = (\Gamma_f(0)+I_r)^{-1/2}(f_t+\nu_t),$$

where $\nu_t = P^\top\Sigma_\varepsilon^{-1}\varepsilon_t$. Notice that $\{\nu_t\}$ is a white-noise process with mean $0_{(r\times 1)}$ and variance matrix I_r; therefore, the canonical variables $B^\top y_t$ have covariance matrix I_r and their associated canonical correlations are the absolute value of the diagonal components of the matrix

$$E[B^\top y_t, y_{t-k}^\top B] = (\Gamma_f(0)+I_r)^{-1/2}\Gamma_f(k)(\Gamma_f(0)+I_r)^{-1/2},$$

as was shown in Proposition 1. □

Lemma 2. *Given the set of lags $\boldsymbol{K}$, if b_j is an eigenvector of $M(k)$ for all $k \in \boldsymbol{K}$ then b_j is an eigenvector of $MS(\boldsymbol{K})$[1] $= \sum_{k\in\boldsymbol{K}} M(k)$, with eigenvalue $\sum_{k\in\boldsymbol{K}}\mu_j^2(k)$,.*

Proof. Multiplying $MS(\boldsymbol{K})$ by b_j, we obtain

$$\left[\sum_{k\in\boldsymbol{K}} M(k)\right] b_j = \left[\sum_{k\in\boldsymbol{K}}\mu_j^2(k)\right] b_j. \tag{11}$$

□

[1] A similar idea is applied in Lam & Yao's (2012) paper to detect the number of factors.

Based on Lemma 2, we propose to use as specific order of the components of f_t, the one specified by the eigenvalues of $MS(\boldsymbol{K})$ ordered in descending way. We denote as $b_{j_{1:\boldsymbol{K}}}, b_{j_{2:\boldsymbol{K}}}, \ldots, b_{j_{r:\boldsymbol{K}}}$ their related eigenvectors, which are the column vectors $b_1, b_2, \ldots, b_r$ defined in Proposition 1, Lemma 1 and Lemma 2, but rearranged according to the defined order. Then, at each lag, we order the canonical correlations by matching each one to the eigenvector of $MS(\boldsymbol{K})$ that is collinear to its canonical vector. That association is possible by means of the cosine similarity between the eigenvectors of $MS(\boldsymbol{K})$ with the eigenvectors of $M(k)$, which are the normalized canonical vectors. Hence, by Lemma 1, all the canonical correlations that match with a particular eigenvector of $MS(\boldsymbol{K})$ are related to the same factor.

This proposal, as stated in Proposition 5, is based on the fact that the sequence of squared cosines of the angle between an eigenvector of $\widehat{M}(k)$ and an eigenvector of $\widehat{MS}(\boldsymbol{K})$, indexed by T, converges in probability to 1 if they are related to the same factor, otherwise it converges to 0.

Proposition 5. *For any lag $k \in \boldsymbol{K}$, let $\widehat{b}_{j_{i:\boldsymbol{K}}}$, $b_{j_{i:\boldsymbol{K}}}$, $\widehat{b}_{j_{i:k}}$ and $b_{j_{i:k}}$ be the eigenvectors associated to the ith non-zero largest eigenvalue of $\widehat{MS}(\boldsymbol{K})$, $MS(\boldsymbol{K})$, $\widehat{M}(k)$ and $M(k)$, respectively. If $M(k)$ has $r(k) \leq r$ eigenvalues different of zero, then, as $T \to \infty$, for $i = 1, \ldots, r$ and $i' = 1, \ldots, r(k)$,*

$$\left(\frac{\widehat{b}^{\top}_{j_{i':k}} \widehat{b}_{j_{i:\boldsymbol{K}}}}{||\widehat{b}_{j_{i':k}}|| \; ||\widehat{b}_{j_{i:\boldsymbol{K}}}||} \right)^2 \xrightarrow{p} \left(\frac{b^{\top}_{j_{i':k}} b_{j_{i:\boldsymbol{K}}}}{||b_{j_{i':k}}|| \; ||b_{j_{i:\boldsymbol{K}}}||} \right)^2 = cos^2\left(\theta_{j_{i':k} j_{i:\boldsymbol{K}}}\right), \tag{12}$$

where $\theta_{j_{i':k} j_{i:\boldsymbol{K}}}$ is the angle between $b_{j_{i':k}}$ and $b_{j_{i:\boldsymbol{K}}}$.

Proof. It follows using the consistency property of the estimators $\widehat{b}_{j_{i':k}}$ and $\widehat{b}_{j_{i:\boldsymbol{K}}}$, and the continuous mapping theorem. □

Note that $cos^2\left(\theta_{j_{i':k} j_{i:\boldsymbol{K}}}\right)$ takes its maximum value if $b_{j_{i':k}}$ and $b_{j_{i:\boldsymbol{K}}}$ are collinear to the same column vector of $\Sigma_{\varepsilon}^{-1} P$; hence, by Lemma 1, they are related to the same particular factor. In this case, $j_{i':k} = j_{i:\boldsymbol{K}}$.

In our proposed methodology, we also define the partial canonical autocorrelation function (PCACF), which, jointly with the CACF, let us identify ARMA models for the factors, in a similar way to the Box-Jenkins methodology. The main idea is to plot the partial canonical correlations that are related to the same factor against lag values, based on the same order defined on the CACF, given that we face the same issues. We call the function that corresponds to the jth factor as the jth *partial canonical autocorrelation function* (PCACF).

To establish such definition, in Proposition 6 we show the relation between partial canonical correlation and a modified partial autocorrelation of the latent factors, in which we use Reinsel's (1997) concept of partial canonical correlations between two random vectors. To fix ideas, the partial canonical correlations between y_t and y_{t-k} are the canonical correlations between y_t and y_{t-k} given y_{t-1}, y_{t-2}, $\cdots$, y_{t-k+1} and can be calculated as the non trivial canonical correlations between $[y_t^{\top}, y_{t-1}^{\top}, \ldots, y_{t-k+1}^{\top}]^{\top}$ and $[y_{t-1}^{\top}, y_{t-2}^{\top}, \ldots, y_{t-k}^{\top}]^{\top}$.

Based on these ideas and being $Y_{t;k} = [y_t^\top, y_{t-1}^\top, \dots, y_{t-k+1}^\top]^\top$, we obtain the following result.

Proposition 6. *Given a lag k, the $r_p(k) \leq r$ partial canonical correlations different from zero between y_t and y_{t-k} are equal to the absolute values of the partial autocorrelation between $f^*_{j,t}$ and $f^*_{j,t-k}$, $j = 1, \dots, r_p(k)$, with $f^*_t = f_t + \nu_t$ and $\{\nu_t\} \sim WN(0, I_r)$. Furthermore, the respective canonical variables are*

$$g_j(k)^\top Y_{t;k} \qquad \text{and} \qquad h_j(k)^\top Y_{t-1;k}, \quad j = 1, 2, \dots, r_p(k),$$

with $h_j(k) = \beta_j(k) \otimes \Sigma_\varepsilon^{-1} P_j \in \boldsymbol{R}^{km}$ and $g_j(k) = \alpha_j(k) \otimes \Sigma_\varepsilon^{-1} P_j \in \boldsymbol{R}^{km}$, where $\alpha_j(k)$ and $\beta_j(k)$ are the canonical vectors associated to the non trivial canonical correlations between $[f^*_{j,t}, f^*_{j,t-1}, \dots, f^*_{j,t-k+1}]^\top$ and $[f^*_{j,t-1}, f^*_{j,t-2}, \dots, f^*_{j,t-k}]^\top$. Here, $\otimes$ denotes the Kronecker product.

Proof. See Appendix A.3. □

Because Proposition 6, on the PCACF we plot the absolute value of the partial autocorrelations of the processes f^*_{jt}, for each $j = 1, \dots, r$. Loosely speaking, we plot a function that shows the MA process behavior of the factors plus a noise process in absolute value; therefore, if the variance of the added noise process is negligible with respect to the variance of the factor we get the PACF of the $MA(q_j)$ process, otherwise an $MA(p_j + q_j)$ process is observed (Peña 2010).

It is worth noticing that, when the partial canonical correlations are ordered in descending way, as in the CACF, a specific order is defined for the components of f_t at each lag, but this ordering on the components of f_t is not necessarily the same at all lags. Then, using the same order define for the CACF, we order at each lag the partial canonical correlations based on this particular components order, as presented below.

In what follows, let $\eta_{j_{i:k}}(k)$ be the ith partial canonical correlation for a given lag k, in descending order of magnitude, and let $h_{j_{i:k}}$ be its related eigenvector, where $i = 1, 2, \dots, r_p(k)$ and $j_{i:k} \in \{1, 2, \dots, r\}$.

Now, from Proposition 6, we note that for a given i, $i = 1, 2, \dots, r_p(k)$, the coefficients of y_t in the respective canonical variable form the vector $h_{j_{i:k}}[1:m]$, the m first elements of the canonical vectors $h_{j_{i:k}}$, which is collinear to one of the columns of the matrix $\Sigma_\varepsilon^{-1} P$. Therefore, a similar result to Proposition 5 is obtained with the sequence of squared cosines of the angles between $\widehat{h}_{j_{i':k}}[1:m]$, a consistent estimator of $h_{j_{i':k}}[1:m]$, and the previously defined vector $\widehat{b}_{j_{i:\boldsymbol{K}}}$, $i = 1, 2, \dots, r$. Hence, as in a similar way that in the CACF, we match all the partial canonical correlations related to the same factor.

Additionally, to test the number of $r_p(k)$ *active* factors on the partial canonical correlations at lag k, we use the statistic

$$C_{m-r_p(k)}(k) = -(T-k-1) \sum_{i=r_p(k)+1}^{m} log(1 - \widehat{\eta}^2_{j_{i:k}}(k)), \tag{13}$$

which is asymptotically a $\chi^2_{(m-r_p(k))^2}$, where the square partial canonical correlations, $\widehat{\eta}^2_{j_{1:k}}(k) \geq \widehat{\eta}^2_{j_{2:k}}(k) \geq \cdots \geq \widehat{\eta}^2_{j_{m:k}}(k)$, are the m smallest eigenvalues of the matrix

$$\widehat{M}_p(k) = \left[\sum_{k+1}^{T}(Y_{t;k}Y_{t;k}^{\top})\right]^{-1} \sum_{k+1}^{T}(Y_{t;k}Y_{t-1;k}^{\top}) \left[\sum_{k+1}^{T}(Y_{t-1;k}Y_{t-1;k}^{\top})\right]^{-1} \sum_{k+1}^{T}(Y_{t-1;k}Y_{t;k}^{\top}).$$

For more details of the asymptotic distribution of this statistic see Reinsel (1997), on the identification stage in the scalar component models (SCM) of Tiao & Tsay (1989).

To use these ideas in practice, we propose the following methodology, which is illustrated in the first example of Section 4.

STEP 1 First, define the set $\boldsymbol{K}$ as the set of all lags $k \leq k_0$, for a fixed $k_0 \geq 1$, such that the Peña-Poncela's test identify at least one factor and get the r eigenvectors of $\widehat{MS}(\boldsymbol{K}) := \sum_{k\in\boldsymbol{K}} \widehat{M}(k)$, related to the r maximum eigenvalues. Notice that the descending order of these eigenvalues define an specific order for the eigenvectors $\widehat{b}_{j_{1:\boldsymbol{K}}}, \widehat{b}_{j_{2:\boldsymbol{K}}}, \ldots, \widehat{b}_{j_{r:\boldsymbol{K}}}$ and by Lemma 1 for the factors, which we propose to use as the unique order for all lags.

STEP 2 At each lag k, associate each canonical correlation to a particular factor via the association to a particular eigenvector of matrix $\widehat{MS}(\boldsymbol{K})$, according to Lemma 2. For this purpose, start associating the estimated largest canonical correlation $\widehat{\mu}_{j_{1:k}}(k)$ to one of the eigenvectors $\widehat{b}_{j_{1:\boldsymbol{K}}}, \widehat{b}_{j_{2:\boldsymbol{K}}}, \ldots, \widehat{b}_{j_{r:\boldsymbol{K}}}$, by selecting the one having the maximum cosine similarity with its eigenvector $\widehat{b}_{j_{1:k}}$, that is, the largest value of the squared cosine of the angle between both vectors (maximum correlation between two random canonical vectors). Similarly, to the next estimated canonical correlations $\widehat{\mu}_{j_{2:k}}(k) \geq \widehat{\mu}_{j_{3:k}}(k) \geq \cdots \geq \widehat{\mu}_{j_{r(k):k}}(k)$ assign one of the eigenvectors $\widehat{b}_{j_{1:\boldsymbol{K}}}, \widehat{b}_{j_{2:\boldsymbol{K}}}, \ldots, \widehat{b}_{j_{r:\boldsymbol{K}}}$, but excluding eigenvectors that were already assigned to higher canonical correlations.

STEP 3 At each lag k, associate each partial canonical correlation to a particular factor via the association to a particular eigenvector of matrix $\widehat{MS}(\boldsymbol{K})$. Follow the same ideas of the STEP 2, using the partial canonical correlations $\widehat{\eta}_{j_{1:k}}(k) \geq \widehat{\eta}_{j_{2:k}}(k) \geq \cdots \geq \widehat{\eta}_{j_{r_p(k):k}}(k)$ and theirs respective eigenvectors $\widehat{h}_{j_{1:k}}[1:m], \widehat{h}_{j_{2:k}}[1:m], \ldots, \widehat{h}_{j_{r_p(k):k}}[1:m]$.

STEP 4 For each $i = 1, 2, \ldots, r$, plot the ith sample CACF defined by $\widehat{\vartheta}_i(k, \boldsymbol{K}) = \widehat{\mu}_{j_{i:\boldsymbol{K}}}(k), 0 < k < T$, where $\widehat{\mu}^2_{j_{i:\boldsymbol{K}}}(k)$ is the ith eigenvalue of the matrix $\widehat{M}(k)$, ordering on STEP 2.

STEP 5 For each $i = 1, 2, \ldots, r$, the ith sample PCACF defined by $\widehat{\varphi}_i(k, \boldsymbol{K}) := \widehat{\eta}_{j_{i:\boldsymbol{K}}}(k)$, $0 < k < T$, where $\widehat{\eta}^2_{j_{i:\boldsymbol{K}}}(k)$ is the ith of the m smallest eigenvalues of the matrix $\widehat{M}_p(k)$, ordering on STEP 3.

STEP 6 Based on the CACF and the PCACF plots, identify ARMA models for the factor processes, in a similar way to the Box-Jenkins methodology, bearing in mind that the ith CACF is proportional to the ACF of the ith factor process, see Proposition 1 and the ith PCACF shows the PACF of f^*_{it}, see Proposition 6.

For the last step, it is important to highlight that, in terms of absolute values, the CACFs show the AR behavior of each factor process because the CACFs are proportional to theirs ACFs (see Proposition 1). In the same way, the PCACFs show the MA behavior of each factor process plus a noise process (see Proposition 6). In summary, in absolute values, the ith canonical and partial canonical correlations show the $ARMA(p_j, q_j)$ behavior of the associated factor if the variance of the added noise process is negligible with respect to the variance of the factor, otherwise an $ARMA(p_j, p_j + q_j)$ behavior is observed (Peña 2010), where p_j and q_j are, respectively, the autoregressive and moving average order of the factor related with the jth partial canonical autocorrelation correlation function.

4. Some Examples

4.1. A Simulated Model

We simulate again model (11), using as sample size $T = 1000$. As was found previously, we obtain $\boldsymbol{K} = \{1, 3\}$ after setting $k_0 = 13$. We recall that using the two possible linear combinations we identify 2 common factors. With the simulated data, we get $\widehat{\mu}^2_{j_{1:\boldsymbol{K}}}(\boldsymbol{K}) = 0.22$ and $\widehat{\mu}^2_{j_{2:\boldsymbol{K}}}(\boldsymbol{K}) = 0.18$, the ordered eigenvalues of the matrix $\widehat{MS}(\boldsymbol{K})$, with eigenvectors $\widehat{b}_{j_{1:\boldsymbol{K}}} = (-0.56, -0.58, -0.31, -0.38, 0.27, 0.20)^\top$ and $\widehat{b}_{j_{2:\boldsymbol{K}}} = (-0.24, 0.18, 0.40, -0.09, 0.63, -0.59)^\top$, respectively.

The next step is to use this specific order at all lags. Then, for each eigenvector $\widehat{b}_{j_{1:\boldsymbol{K}}}$ and $\widehat{b}_{j_{2:\boldsymbol{K}}}$, we match a canonical correlation and a partial canonical correlation at each lag $k = 1, 2, \ldots, k_0$, based on the methodology mentioned before. As an illustration of our proposed ordering methodology, we calculate the cosine similarity to order the canonical correlations at lags $k = 1$ and $k = 3$. For lag $k = 1$, $\widehat{\mu}_{j_{1:1}}(1) = 0.46$ with eigenvector $\widehat{b}_{j_{1:1}} = (0.61, 0.48, 0.17, 0.37, -0.47, 0.04)^\top$ and $\widehat{\mu}_{j_{2:1}}(1) = 0.10$ with eigenvector $\widehat{b}_{j_{2:1}} = (0.62, -0.47, -0.52, -0.12, -0.03, 0.31)^\top$. The cosine similarity of $\widehat{b}_{j_{1:1}}$ with the vectors $\widehat{b}_{j_{1:\boldsymbol{K}}}$ and $\widehat{b}_{j_{2:\boldsymbol{K}}}$ are 0.87 and 0.12, respectively. Then, we relate $\widehat{\mu}_{j_{1:1}}(1)$ to $\widehat{b}_{j_{1:\boldsymbol{K}}}$ and $\widehat{\mu}_{j_{2:1}}(1)$ to $\widehat{b}_{j_{2:\boldsymbol{K}}}$. In the same way, at lag $k = 3$, we got 0.16 and 0.85 as the cosine similarity of the vector $\widehat{b}_{1:3}$ with $\widehat{b}_{j_{1:\boldsymbol{K}}}$ and $\widehat{b}_{j_{2:\boldsymbol{K}}}$, respectively; hence, we relate $\widehat{\mu}_{j_{1:3}}(3) = 0.42$ to $\widehat{b}_{j_{2:\boldsymbol{K}}}$ and $\widehat{\mu}_{j_{2:3}}(3) = 0.07$ to $\widehat{b}_{j_{1:\boldsymbol{K}}}$.

In other words, $\widehat{\vartheta}_1(1, \boldsymbol{K}) = 0.46$, $\widehat{\vartheta}_2(1, \boldsymbol{K}) = 0.07$, $\widehat{\vartheta}_1(3, \boldsymbol{K}) = 0.10$ and $\widehat{\vartheta}_2(3, \boldsymbol{K}) = 0.42$ as is shown in the Figure 1, where we plot the CACF and the PCACF according to the order specified by $\widehat{b}_{j_{1:\boldsymbol{K}}}$ and $\widehat{b}_{j_{2:\boldsymbol{K}}}$. We use gray bars in both graphics to indicate the canonical correlations and the partial canonical correlations that are statistically different from zero, according to the tests (3) and (13) mentioned before.

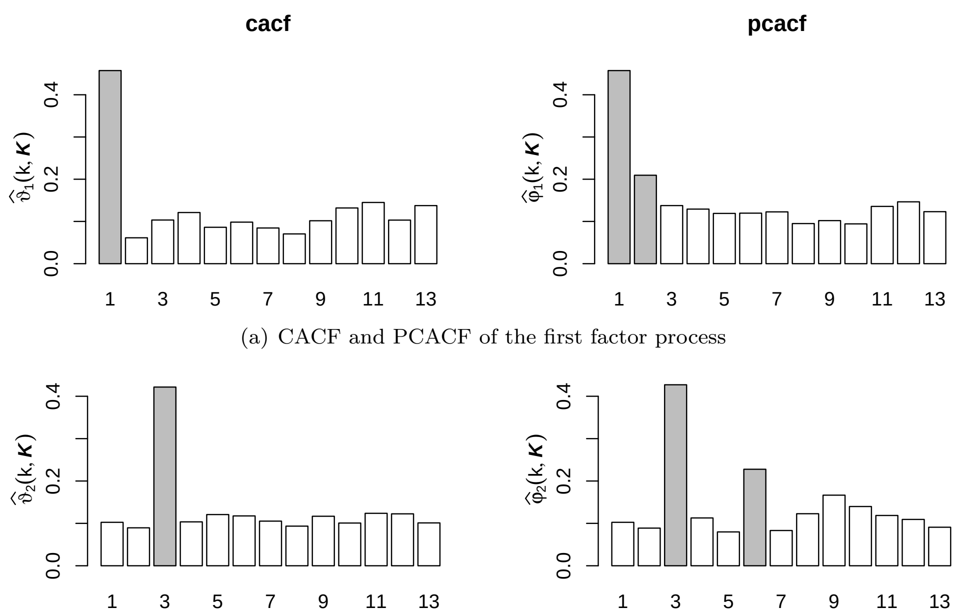


Figure 1: Plot of CACF and PCACF of the simulated data. Gray bars indicate the canonical correlations and the partial canonical correlations that are statistically different from zero, according to the tests (3) and (13) mentioned before.

Notice that the first CACF and PCACF show an $MA(1)$ behavior; hence they are related to the factor f_{1t} and the second ones, an $MA(3)$, as it is factor f_{2t}. In this example, the CACF and PCACF show exactly the same expected behavior of the ACF and PACF proposed by Box & Jenkins (1970), because an $MA(q)$ process plus a white noise process still being an $MA(q)$ process (Peña 2010).

4.2. A Real Data Application

This example is taken from Nieto et al.'s (2016) paper, where the total monthly rainfall time series were used. The rainfalls were measured in meteorological stations located at the airports of six cities in Colombia: Bucaramanga(y_1), Cúcuta(y_2), Ibagué(y_3), Medellín(y_4), Manizales(y_5) and Bogotá(y_6). Figure 2 presents the deseasonalized time series and, with the Peña & Poncela's (2006) test at the lags $1, 2, \ldots, 13$, we get that the lags $1, 4$ and 6 present at leat one factor. With $\mathbf{K} = \{1, 4, 6\}$ we detect two common factors using the test proposed in Section 2.

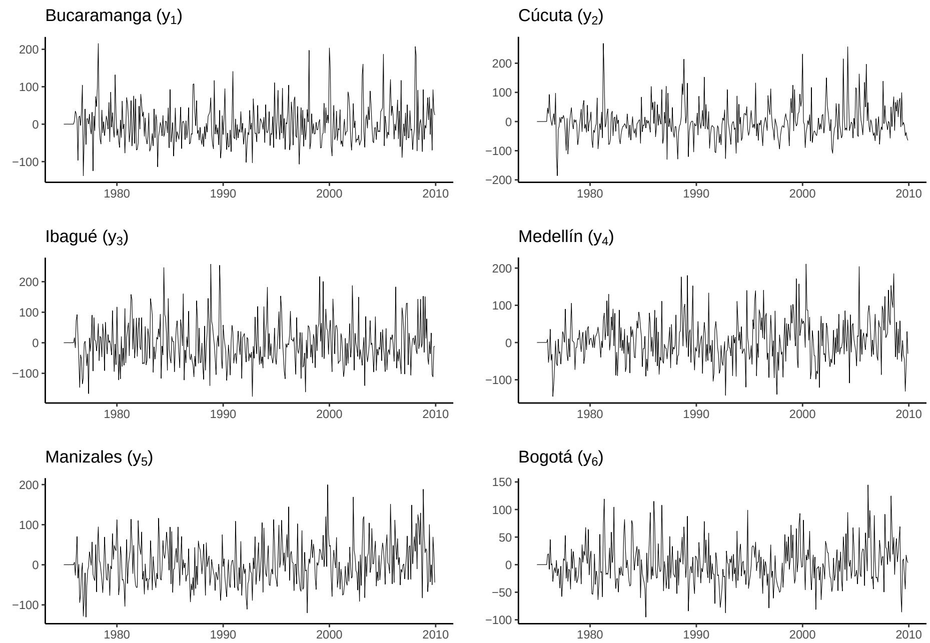
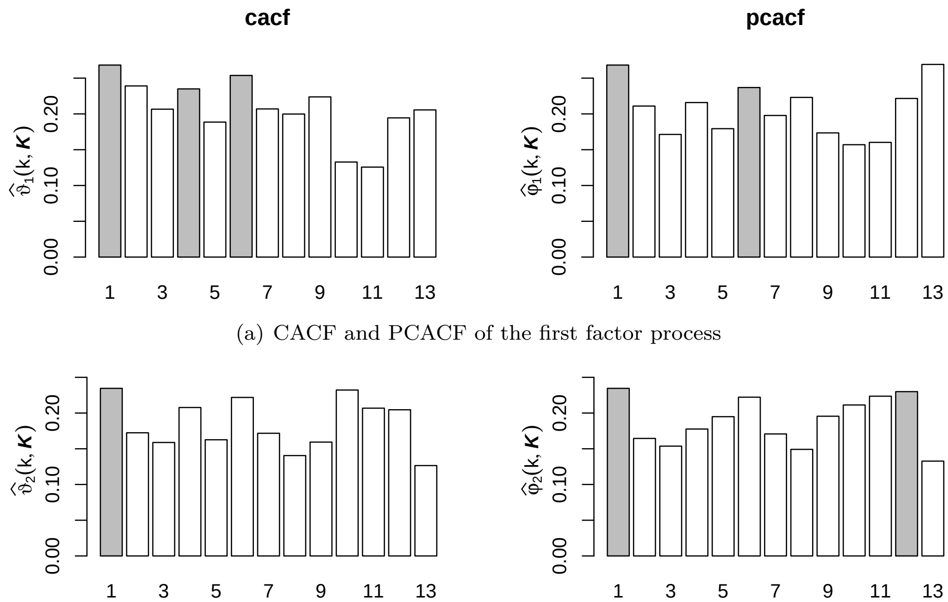


FIGURE 2: Rainfall time series.

Figure 3 presents the CACF and the PCACF of the deseasonalized time series following our proposed methodology. To the first factor, Figure 3(a), in the CACF non-zero correlations are observed at lags 1, 4 and 6 (according to Peña and Poncela's test) and in the PCACF at lags 1 and 6 (according to the test 13). Also, a possible decreasing behavior in the CACF from the first correlation, that suggests an $MA(1)$ process. For the second factor, Figure 3(b), in the CACF and the PCACF a non-zero correlations are observed at lag 1 (according to the test (3) and (13)) and a possible decreasing behavior is observed in the CACF, that suggests an $AR(1)$ process.

To the estimation of the parameters, we maximize the likelihood function using the EM algorithm. On step E, we use the Kalman filter and the smoothing algorithm. On step M, we use the space-state representation proposed by Metaxoglou & Smith (2007) for VARMA models, intending to simplify the maximization process from step M, as the authors mention in their paper. Also, to solve the identification problem of the model, restriction $P^{\top}\Sigma_{\varepsilon}^{-1}P = I_r$ was imposed, and using Jungbacker & Koopman (2015) ideas we transform the data $y_t^* = A_L y_t$ in step E, with $A_L = P^{\top}\Sigma_{\varepsilon}^{-1}$, obtaining the transformed model $y_t^* = A_L y_t = f_t + A_L \varepsilon_t$, where $A_L P = I_r$ and $var[A_L \varepsilon_t] = I_r$.



(a) CACF and PCACF of the first factor process

(b) CACF and PCACF of the second factor process

Figure 3: Plot of CACF and PCACF of Colombian rainfalls

Based on the above specification, the following model with two common factors is estimated:

$$Y_t = \begin{bmatrix} 7.52 & -18.19 \\ 11.14 & -32.96 \\ 19.73 & 8.14 \\ 22.60 & 14.33 \\ 16.52 & -6.73 \\ 12.46 & -3.03 \end{bmatrix} f_t + \varepsilon_t,$$

where $f_{1t} = 0.12 f_{1,t-6} + 0.117 a_{1,t-1} + 0.122 a_{1,t-4} + a_{1t}$, $f_{2t} = 0.36 f_{2,t-1} + a_{2t}$, $\Sigma_a = diag(4.01, 0.69)$ and $\Sigma_\varepsilon = diag(2226.9, 1853.7, 3463.2, 1279.2, 1566.2, 712.7)$.

The structure of the factors can be seen by columns of the P matrix. The first one has a positive effect on the time series, with a minor scale on the Bucaramanga rainfall (y_{1t}). The second separates Ibagué (y_{3t}) and Medellin (y_{4t}) from the others cities precipitations.

5. Conclusions

In this paper, (1) we establish the relationship between canonical correlations and the autocovariance function of the factor process. Based on this relation, we modify Peña & Poncela's (2006) test to detect the number of common factors, increasing the test power. Additionally, (2) we establish the relationship between partial canonical correlations and the partial autocorrelation function of the factor process. Finally, (3) we propose to use the canonical vectors to link the canonical and partial canonical correlations at each lag to a specific factor process. These three findings allow us to propose a procedure to identify a vector ARMA model for the factor process, which is based on the so-called simple and partial canonical autocorrelation functions.

[Received: March 2019 — Accepted: September 2020]

References

Ahn, S. C. & Horenstein, A. R. (2013), 'Eigenvalue Ratio Test for the Number of Factors', *Econometrica* **81**(3), 1203–1227.

Anderson, T. (1984), *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Series, Wiley.

Box, G. E. P. & Jenkins, G. M. (1970), *Time series analysis: forecasting and control*, San Francisco, CA: Holden-Day.

Doz, C. & Fuleky, P. (2020), Dynamic factor models, *in* 'Macroeconomic Forecasting in the Era of Big Data', Springer, pp. 27–64.

Geweke, J. (1977), 'The dynamic factor analysis of economic timeseries models', *Latent Variables in Socio-Economic Models* pp. 365–383.

Jungbacker, B. & Koopman, S. J. (2015), 'Likelihood-Based Dynamic Factor Analysis for Measurement and Forecasting', *The Econometrics Journal* **18**(2), C1–C21.

Lam, C. & Yao, Q. (2012), 'Factor modeling for high-dimensional time series: Inference for the number of factors1', *Annals of Statistics* **40**(2), 694–726.

Metaxoglou, K. & Smith, A. (2007), 'Maximun likelihood estimation of VARMA models using a state-space EM algorithm', *Journal of Time Series Analysis* **28**(5), 666–685.

Nieto, F. H., Peña, D. & Saboyá, D. (2016), 'Common seasonality in multivariate time series', *Statistica Sinica* **26**, 1389–1410.

Peña, D. (2010), *Análisis de series temporales*, El Libro Universitario - Manuales, Alianza Editorial.

Peña, D. & Box, G. E. P. (1987), 'Identifying a Simplifying Structure in Time Series', *Journal of the American Statistical Association* **82**(399), 836–843.

Peña, D. & Poncela, P. (2006), 'Nonstationary dynamic factor analysis', *Journal of Statistical Planning and Inference* **136**(4), 1237–1257.

Reinsel, G. C. (1997), *Elements of Multivariate Time Series Analysis*, 2 edn, Springer-Verlag.

Stock, J. H. & Watson, M. (2011), Dynamic factor models, *in* P. C. Michael & D. F. Hendry, eds, 'Oxford Handbook on Economic Forecasting', Oxford University Press, Oxford.

Stock, J. H. & Watson, M. W. (2016), Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics, *in* J. B. Taylor & H. Uhlig, eds, 'Handbook of macroeconomics', Vol. 2 of *Handbook of Macroeconomics*, Elsevier, chapter 8, pp. 415–525.

Tiao, G. C. & Tsay, R. S. (1989), 'Model specification in multivariate time series', *Journal of the Royal Statistical Society. Series B (Methodological)* **51**(2), 157–213.

Appendix A. Proofs of the Propositions

Appendix A.1. Proposition 1

Proof. Calling $\boldsymbol{\Gamma}_y(k) = E(\mathbf{y}_t\mathbf{y}_{t-k}^\top)$ to the lag covariance matrices of the data and $\boldsymbol{\Gamma}_f(k) = E(\mathbf{f}_t\mathbf{f}_{t-k}^\top)$ the diagonal covariance matrices for the factors, we have that $\boldsymbol{\Gamma}_y(0) = \mathbf{P}\boldsymbol{\Gamma}_f(0)\mathbf{P}^\top + \boldsymbol{\Sigma}_\varepsilon$ and for $k > 0$,

$$\boldsymbol{\Gamma}_y(k) = \mathbf{P}\boldsymbol{\Gamma}_f(k)\mathbf{P}^\top. \tag{A.1}$$

Let us first assume that $\boldsymbol{\Sigma}_\varepsilon = \mathbf{I}_m$ and consider now lineal combinations $\mathbf{a}^\top\mathbf{y}_t$ and $\mathbf{b}^\top\mathbf{y}_{t-k}$ of unit variance that has maximum squared correlation. It is easy to see that the squared correlation is the largest eigenvalue and $\mathbf{a}$ the corresponding eigenvector of the matrix

$$\mathbf{H}_k = \boldsymbol{\Gamma}_y(0)^{-1}\boldsymbol{\Gamma}_y(k)\boldsymbol{\Gamma}_y(0)^{-1}\boldsymbol{\Gamma}_y(k) = \mathbf{A}\mathbf{A}$$

where

$$\mathbf{A} = \boldsymbol{\Gamma}_y(0)^{-1}\boldsymbol{\Gamma}_y(k).$$

Therefore as the eigenvalues of a non singular matrix and its inverse are the inverse and with the same eigenvectors, $\boldsymbol{\Gamma}_y(0) = \mathbf{P}\boldsymbol{\Gamma}_f(0)\mathbf{P}^\top + \mathbf{I}_m$ have eigenvectors $\mathbf{P}$ and eigenvalues $\gamma_j(0) + 1$. Then $\boldsymbol{\Gamma}_y^{-1}(0)$ has eigenvectors $\mathbf{P}$ and eigenvalues $(\gamma_j(0) + 1)^{-1}$. Therefore, $\mathbf{A}$ has $\mathbf{P}$ eigenvectors and eigenvalues $(\gamma_j(0)+1)^{-1}\gamma_j(k)$. Matrix $\mathbf{H}_k$ has the same P eigenvectors and but now its eigenvalues are $(1+\gamma_j(0))^{-2}\gamma_j(k)^2$.

In summary, the canonical variables have the form $\mathbf{a}^\top\mathbf{y}_t = \alpha_j\mathbf{p}_j^\top\mathbf{y}_t$ where $\mathbf{p}_j$ is one of the columns of P and in order to have variables with unit variance as $E(\alpha_j^2\mathbf{p}_j^\top\mathbf{y}_t\mathbf{y}_t^\top\mathbf{p}_j) = \alpha_j^2\mathbf{p}_j^\top\ \boldsymbol{\Gamma}_y(0)\mathbf{p}_j = \alpha_j^2(\gamma_j(0) + 1)$ then

$$\mathbf{a} = \alpha_j\mathbf{p}_j = \frac{1}{(\gamma_j(0) + 1)^{1/2}}\mathbf{p}_j.$$

In the general case, calling $\mathbf{y}_t^* = \mathbf{\Sigma}_\varepsilon^{-1/2}\mathbf{y}_t$ the model can be written as in (1) with $\mathbf{P}^* = \mathbf{\Sigma}_\varepsilon^{-1/2}\mathbf{P}$ and $\mathbf{\Sigma}_\varepsilon^* = \mathbf{I}_m$. Assuming $\mathbf{P}^{*\top}\mathbf{P}^* = \mathbf{I}_r = \mathbf{P}^\top\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{P}$

$$\mathbf{\Gamma}_y^*(0) = \mathbf{\Sigma}_\varepsilon^{-1/2}\mathbf{\Gamma}_y(0)\mathbf{\Sigma}_\varepsilon^{-1/2} = \mathbf{\Sigma}_\varepsilon^{-1/2}\mathbf{P}\mathbf{\Gamma}_f(0)\mathbf{P}^\top\mathbf{\Sigma}_\varepsilon^{-1/2} + \mathbf{I}_m$$

and the eigenvectors of $\mathbf{\Gamma}_y^*(0)$ are $\mathbf{\Sigma}_\varepsilon^{-1/2}\mathbf{p}_j$ and the eigenvalues $\gamma_j(0)+1$. □

Appendix A.2. Proposition 2

Proof. Assuming that $\mathbf{\Sigma}_\varepsilon = \mathbf{I}_m$ and taking the lineal combinations $\mathbf{a}^\top\mathbf{y}_t$ and $\mathbf{b}^\top\mathbf{y}_t^\dagger$ of unit variance that has maximum squared correlation, it is easy to see that the squared correlation is the largest eigenvalue and $\mathbf{a}$ the corresponding eigenvector of the matrix

$$H_{\mathbb{K}} = \mathbf{\Gamma}_{y^\dagger}(0)^{-1}\mathbf{\Gamma}_y(\mathbb{K})\mathbf{\Gamma}_y(0)^{-1}\mathbf{\Gamma}_y(\mathbb{K})$$

where

$$\begin{aligned}
\mathbf{\Gamma}_y(\mathbb{K}) &:= E\left[\mathbf{y}_t\mathbf{y}_t^{\dagger\top}\right] = \mathbf{P}\mathbf{\Gamma}_f(\mathbb{K})\mathbf{P}^\top,\\
\text{y} \qquad \mathbf{\Gamma}_y(0) &:= E\left[\mathbf{y}_t\mathbf{y}_t^\top\right] = \mathbf{P}\mathbf{\Gamma}_f(0)\mathbf{P}^\top + \mathbf{I}_m\\
\mathbf{\Gamma}_{y^\dagger}(0) &:= E\left[\mathbf{y}_t^\dagger\mathbf{y}_t^{\dagger\top}\right] = \mathbf{P}\mathbf{\Gamma}_{f^\dagger}(0)\mathbf{P}^\top + k^\dagger\mathbf{I}_m,
\end{aligned} \tag{A.2}$$

with $\mathbf{\Gamma}_{f^\dagger}(0) := \sum_{k_1\in\mathbb{K}}\sum_{k_2\in\mathbb{K}} E\left[\mathbf{f}_{t-k_1}\mathbf{f}_{t-k_2}^\top\right]$, $\mathbf{\Gamma}_f(\mathbb{K}) := \sum_{k\in\mathbb{K}} E\left[\mathbf{f}_t\mathbf{f}_{t-k}^\top\right]$ and $k^\dagger$ the cardinality of $\mathbb{K}$.

Therefore, following the same ideas of the demonstration in Proposition 1 the result of Proposition 2 is obtained, with $\gamma_j(\mathbb{K})$ and $\gamma_j^\dagger$ the j-th elements of the diagonal matrices $\mathbf{\Gamma}_f(\mathbb{K})$ and $\mathbf{\Gamma}_{f^\dagger}(0)$, respectively.

□

Appendix A.3. Proposition 6

Proof. Let $H = [P, P^\perp]$, such that $H^T\Sigma_\varepsilon^{-1}H = I_m$ and $z_t = H^\top\Sigma_\varepsilon^{-1}y_t$, then

$$z_t = H^\top\Sigma_\varepsilon^{-1}Pf_t + H^\top\Sigma_\varepsilon^{-1}\varepsilon_t \tag{A.3}$$

$$z_t = \begin{bmatrix} f_t \\ 0_{(m-r)\times 1}\end{bmatrix} + \nu_t, \tag{A.4}$$

where ν_t is a Gaussian noise process with $E[\nu_t] = 0_{m\times 1}$ and $E[\nu_t\nu_t^\top] = I_m$. Therefore, the components of z_t are pairwise orthogonal, which means $cov[z_{j,t}, z_{j',s}] = 0$, for any $j, j' = 1,\ldots,m$, $j \neq j'$, $t, s \in \mathbb{Z}$. Additionally,

$$var[z_t] = \begin{bmatrix} \Sigma_f(0) & 0_{r\times(m-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(m-r)}\end{bmatrix} + I_m \tag{A.5}$$

$$cov[z_t, z_{t-k}] = \begin{bmatrix} \Sigma_f(k) & 0_{r\times(m-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(m-r)}\end{bmatrix}. \tag{A.6}$$

Given that $H^\top \Sigma_\varepsilon^{-1}$ is an invertible matrix, the canonical correlations between y_t and y_{t-k} are equal to the canonical correlations between z_t and z_{t-k}, aditionally, the canonical vectors of y_t are equal to the canonical vectors of z_t multiplyied by $H^\top \Sigma_\varepsilon^{-1}$ (Anderson 1984). Therefore, in the rest of the proof we find the canonical correlation between z_t and z_{t-k} and their respective canonical vectors.

To get the partial canonical correlations of the process $\{z_t\}$, note that, it is pairwise orthogonal, then the partial canonical correlations between z_t and z_{t-k} are equal to the k-th partial autocorrelation of $z_{j,t}$, the j-th component of z_t. This partial autocorrelation can be calculate as the k-th component of the vector (Reinsel 1997)

$$\begin{bmatrix}\phi^j_{1k} & \phi^j_{2k} & \cdots & \phi^j_{kk}\end{bmatrix}^\top = \left(E[Z_{j,t-1;k} Z_{j,t-1;k}^\top]\right)^{-1} E[Z_{j,t-1;k}, z_{jt}], \tag{A.7}$$

where $Z_{j,t;k} = [z_{j,t}^\top, z_{j,t-1}^\top, \ldots, z_{j,t-k+1}^\top]^\top$. Using the properties of partitioned matrices

$$\phi^j_{kk} = \frac{\gamma_j(k) - \varrho_{j,k-1}}{(\gamma_j(0)+1) - \varrho_{j,k-1}},$$

with $\varrho_{j,k-1} = \widetilde{\Gamma}^\top_{j;(k-1)} \Gamma^{-1}_{j;k-1} \Gamma_{j;(k-1)}$ if $k > 1$ and $\varrho_{j,k-1} = 0$ if $k = 1$, where

$$\begin{aligned} \Gamma_{j;k} &= E[Z_{j,t;k} Z_{j,t;k}^\top], \\ \text{and} \qquad \Gamma_{j;(k)} &= [\gamma_{z_j}(1), \gamma_{z_j}(2), \ldots, \gamma_{z_j}(m)] \\ \widetilde{\Gamma}_{j;(k)} &= [\gamma_{z_j}(m), \gamma_{z_j}(m-1), \ldots, \gamma_{z_j}(1)]. \end{aligned}$$

To get the canonical vectors related to the non trivial canonical correlations of $Z_{t;k} = [z_t^\top, z_{t-1}^\top, \ldots, z_{t-k+1}^\top]^\top$ and $Z_{t-1;k} = [z_{t-1}^\top, z_{t-2}^\top, \ldots, z_{t-k}^\top]^\top$, say $g_j^*(k)$ and $h_j^*(k)$, respectively, we use the canonical vectors of $Z_{j,t;k}$ and $Z_{j,t-1,k}$, say $\alpha_j(k)$ and $\beta_j(k)$ (dimesion $k \times 1$). Hence, given that the process $\{z_t\}$ is pairwise orthogonal these canonical vectors can be expressed as

$$g_j^*(k) = \alpha_j(k) \otimes e_j \qquad y \qquad h_j^*(k) = \beta_j(k) \otimes e_j \qquad j = 1, \ldots, r_p(k), \tag{A.8}$$

where e_j is the vector of dimension $m \times 1$ with the value one in the position j-th and zero otherwise. This is easy to verify by multiplying $g_j^*(k)^\top$ by $Z_{t;k}$ and $h_j^*(k)^\top$ by $Z_{t-1;k}$. Based on this result, it is concluded that the canonical vectors regarding to $[y_t^\top, y_{t-1}^\top, \ldots, y_{t-k+1}^\top]^\top$ and $[y_{t-1}^\top, y_{t-2}^\top, \ldots, y_{t-k}^\top]^\top$ are, respectively,

$$g_j(k) = \alpha_j(k) \otimes \Sigma_\varepsilon^{-1} P_j \quad \textit{and} \quad h_j(k) = \beta_j(k) \otimes \Sigma_\varepsilon^{-1} P_j \quad j = 1, \ldots, r_p(k), \tag{A.9}$$

beacuse, $g_j^*(k)^\top Z_{t;k} = (\alpha_j(k)^\top \otimes e_j) Z_{t;k} = (\alpha_j(k)^\top \otimes e_j)(I_k \otimes P_j^\top \Sigma_\varepsilon^{-1}) Y_{t;k}$, hence the canonical variables related to $Y_{t;k}$ and $Y_{t-1;k}$ are $(\alpha_j(k)^\top \otimes P_j^\top \Sigma_\varepsilon^{-1}) Y_{t;k}$ and $(\beta_j(k)^\top \otimes P_j^\top \Sigma_\varepsilon^{-1}) Y_{t-1;k}$.

To get $\alpha_j(k)$ and $\beta_j(k)$ we use that

$$E\left[g_j^*(k)^\top Z_{t;k} Z_{t-1;k}^\top h_j^*(k)\right] = \phi^j_{kk},$$

and hence

$$\alpha_j(k) = \frac{1}{(\gamma_j(0)+1) - \varrho_{j,k-1}} \begin{bmatrix} -1 \\ \Gamma^{-1}_{j;k-1}\Gamma_{j;(k-1)]} \end{bmatrix}, \tag{A.10}$$

$$\beta_j(k) = \frac{1}{(\gamma_j(0)+1) - \varrho_{j,k-1}} \begin{bmatrix} \Gamma^{-1}_{j;k-1}\widetilde{\Gamma}_{j,(k-1)} \\ -1 \end{bmatrix} \tag{A.11}$$

By notation, on Proposition 6, we set the first r components of z_t as $f^*_{j,t}$, $j = 1, 2, \ldots, r$. $\square$