Revista Colombiana de Estadística
ISSN: 0120-1751
revcoles_fcbog@unal.edu.co
Universidad Nacional de Colombia
Colombia

González, Mariano; Nave, Juan M.
Identification of Common Factors in Multivariate Time Series Modeling
Revista Colombiana de Estadística, vol. 38, núm. 1, 2015, pp. 219-238
Universidad Nacional de Colombia
Bogotá, Colombia

Available in: http://www.redalyc.org/articulo.oa?id=89938627013

How to cite
Complete issue
More information about this article
Journal's homepage in redalyc.org

Scientific Information System
Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal
Non-profit academic project, developed under the open access initiative

Revista Colombiana de Estadística
January 2015, Volume 38, Issue 1, pp. 219 to 238
DOI: http://dx.doi.org/10.15446/rce.v38n1.48812

Identification of Common Factors in Multivariate Time Series Modeling

Identificación de factores comunes en la modelización multivariante de series temporales

Mariano González[1,a], Juan M. Nave[2,b]

[1]Departamento de Economía y Empresa, Facultad de Derecho y Empresa, Universidad CEU Cardenal Herrera, Valencia, Spain

[2]Departamento de Análisis Económico y Finanzas, Facultad de Ciencias Sociales, Universidad de Castilla La Mancha, Cuenca, Spain

Abstract

For multivariate time series modelling, it is essential to know the number of common factors that define the behaviour. The traditional approach to this problem is investigating the number of cointegration relations among the data by determining the trace and the maximum eigenvalue and obtaining the number of stationary long-run relations. Alternatively, this problem can be analyzed using dynamic factor models, which involves estimating the number of common factors, both stationary and not, that describe the behaviour of the data. In this context, we empirically analyze the power of such alternative approaches by applying them to time series that are simulated using known factorial models and to financial market data. The results show that when there are stationary common factors, when the number of observations is reduced and/or when the variables are part of more than one cointegration relation, the common factors test is more powerful than the usually applied cointegration tests. These results, together with the greater flexibility to identify the loading matrix of the data generating process, render dynamic factor models more suitable for use in multivariate time series analysis.

***Key words*:** Cointegration, Factor Analysis, Stationarity.

[a]Tenure Professor. E-mail: mariano.gonzalez@uch.ceu.es
[b]Professor. E-mail: juan.nave@uclm.es

Resumen

Para la modelización multivariante de series temporales no estacionarias es imprescindible conocer el número de factores comunes que definen el comportamiento de las series. La forma tradicional de abordar este problema es el estudio de las relaciones de cointegración entre los datos a través de las pruebas de la traza y el máximo valor propio, obteniendo el número de relaciones de largo plazo estacionarias. Como alternativa, se pueden emplear modelos factoriales dinámicos que estiman el número de factores comunes, estacionarios o no, que describen el comportamiento de los datos. En este contexto, analizamos empíricamente el resultado de aplicar tales métodos a series simuladas mediante modelos factoriales conocidos, y a datos reales de los mercados financieros. Los resultados muestran que cuando hay factores comunes estacionarios, cuando el número de observaciones se reduce y/o cuando las variables participan en más de una relación de cointegración, la prueba de factores comunes es más potente que las pruebas habituales de cointegración. Estos resultados, junto con la mayor flexibilidad para identificar la matriz de cargas del proceso generador de datos, hacen que los modelos de factores dinámicos sean más adecuados para su utilización en el análisis multivariante.

***Palabras clave*:** cointegración, estacionariedad, factores comunes, modelo factorial dinámico.

1. Introduction

The identification of the common factors of a set of variables and their reduced representation is an open research area within the social sciences (Zhang 2009). The usual multivariate techniques used for this include Common Factor Analysis (CFA) and Principal Component Analysis (PCA). CFA is used to obtain a reduced-dimensional representation of variance shared among a set of observed variables, whereas PCA is used to obtain a reduced-dimensional representation of the total variance of the observed variables. As Widaman (1993) noted, the choice between CFA and PCA as the model for representing the dependence among a set of variables involves different loading matrices. Furthermore, Independent Component Analysis (ICA) complements PCA and extracts independent factors from the non-correlated factors of PCA for non-Gaussian variables (González & Nave 2010).

An important issue in CFA and PCA is the number of factors necessary to represent the set of observed variables. Lorenzo-Seva, Timmerman & Kiers (2011) reviewed the procedures to select this number of factors in a cross-section framework and proposed the use of the Hull method, but this heuristic methodology does not inform us about the number of possible factors, its results depend on the function used as a measure of error and cannot be used to obtain the loading matrix. Furthermore, this method has not been extended to time series analysis.

Additionally, when some of the observed variables are not stationary, PCA can extract factors with similar loadings for all variables. This lack of sparseness makes the interpretation of the results more difficult. Lansangan & Barrios (2009) suggest using Sparse Principal Component Analysis, but it requires finding a sparse

approximation of the loading vectors. Thus, when the set of observed variables contains non-stationary variables, the following two are the approaches usually used in dynamic multivariate analysis: cointegration analysis (CA) and dynamic factorial analysis (DFA).

CA identifies the long-run relationships among variables, while DFA determines the number of uncorrelated and unobservable common dynamic factors that describe the behaviour of the variables. In the long run, the solutions of both problems are complementary because for a set of variables, the number of common factors is the difference between the number of variables and the number of cointegration relationships. However, in Economics papers, CA is more widely used in the dynamic multivariate context, perhaps because it is more popular and is included in standard econometrics software packages (PcGive, Gretl and JMulti, among others). Beyond focus on the long run, other shortcomings may arise when we apply CA, especially when the model errors are serially dependent and/or are non-Gaussian (Gonzalo 1994, Gonzalo & Lee 1998). Thus, alternative approaches become necessary (Li, Pan & Yao 2009).

Most of the research on cointegration has focused on tests to estimate the cointegration relationships, including different issues such as the sample size (Banerjee, Dolado & Mestre 1998, Pesavento 2007, Bayer & Hanck 2013), structural breaks (Cavaliere & Taylor 2006, Westerlund & Egderton 2006, Jing & Junsoo 2010), decomposition matrix (Doornik & O'Brien 2002), common cycles (Cubadda 2007) and fractional cointegration (Dittmann 2000, Trenkler, Saikkonen & Lütkepohl 2007, Davidson & Monticini 2007). However, none of these studies are concerned with estimating the parameters that define cointegration relationships, and in most cases, they only consider a single relationship between two variables, where the relationship is well defined. In this case, normalizing based on one of the variables yields the parameters. In others cases, as that described by Park, Ahn & Cho (2011), an alternative estimation method is proposed. When the number of variables involved increases, to achieve an identified system, we must use a set of restrictions based on relationships among variables known a priori, which are not always available. Finally, when we analyze finite samples, the distribution of the cointegration rank test is not approximated well by the limiting distribution (Ahlgren & Antell 2008).

In this context, DFA becomes increasingly useful. Peña & Box (1987) showed how to identify common factors for time series and how to build a simple transformation to recover the factors as linear combinations of the original series. Stock & Watson (1988) were the first to connect CA and CFA. They showed that multiple cointegrated time series must have at least one common trend or factor. Then, Escribano & Peña (1994) proposed a new normalization technique, which builds common trend representations with moving-average polynomials and, under certain circumstances, with uncorrelated shocks, that allow us to see the connection between CA and CFA. Hu & Chou (2003) studied DFA to simplify multiple time series, and Hu & Chou (2004) derived a procedure to identify the Peña & Box (1987) model. Subsequently, Kapetanios & Marcellino (2009) showed, in Monte Carlo experiments, the advantages of using a parametric estimation of dynamic factor models. Correal & Peña (2008) introduced a threshold dynamic factor

model for the analysis of vector time series, that includes the non-linear threshold-type behaviour. More recently, Lopes, Gamerman & Salazar (2011) proposed the generalized spatial dynamic factor models. Some papers on CA, such as that by Miller (2010), suggest using DFA to estimate the parameters when there are more than two time series because the cointegration representation is not unique in this case.

The main objective of this paper is to analyze these two (a priori) alternative approaches to show whether they can yield the true number of hidden common factors in the data-generating processes of a set of variables. To do so, we apply to series of simulated and real data, the test developed in the DFA framework by Peña & Poncela (2006), and a traditional cointegration test, i.e., the trace test. We choose the Peña & Poncela (2006) test as the representative test among a set of similar tests (Forni, Hallin, Lippi & Reichlin 2005, Hallin & Liska 2007, Chen, Huang & Tu 2010) developed in the DFA framework due to its superior robustness (Park, Mammen, Hardle & Borak 2009) and lower computational cost compared with other tests, such as that described in Pan & Yao (2010).

The rest of the paper is structured as follows. In section 2, we formulate the problem and describe the Peña & Poncela (2006) methodology. In section 3, we perform a comparative analysis of the two methodologies, CA and DFA, using simulated time series. In section 4, we apply the methods to real data from financial markets. Finally, we summarize the main conclusions.

2. Cointegration vs. Peña and Poncela Factorial Model

In multivariate cases, when the observed variables are non-stationary, by using a stationary transformation of the series, we may not be using all of the information about the relationships among the variables (Peña & Sanchez 2007). To compensate for this, we can use two complementary perspectives as follows:

- Considering the long-run stable and stationary relationships among variables and using different techniques related to cointegration. Thus, if $\mathbf{X}$ is the vector of N non-stationary variables and $\mathbf{\Delta X}$ is its stationary transformation, the Vector Error Correction Model or VECM (Engle & Granger 1987) is usually used in the literature:

$$\mathbf{\Delta X_t} = \mathbf{\Pi} \cdot \mathbf{X_{t-1}} + \sum_{i=1}^{P} \gamma_i \cdot \mathbf{\Delta X}_{t-i} + \epsilon_t \tag{1}$$

  where $t = 1, \ldots, M$; P is the number of lags required to correct the autocorrelation and other characteristics of the data series ($\mathbf{\Delta X}$), usually determined using the Akaike Information Criterion (AIC); ϵ is an $(N \times 1)$ vector of stationary errors; and Π denotes an $(N \times m)$ matrix of the cointegration relationship. Through the well-known cointegration tests, i.e., the maximum eigenvalue and the trace test, we can estimate the cointegration rank, or rank

of the matrix Π in (1). Thus, if there are N variables and m cointegration relationships, we obtain $r = N - m$ non-stationary common factors.

- Directly obtaining the number of common factors (Escribano & Peña 1994, Gonzalo & Granger 1995). In this case, the formulation depends on the factors ($\mathbf{f}$) and their loading matrix (λ):

$$\begin{aligned} \mathbf{X}_t &= \lambda \cdot \mathbf{f}_t + \omega \cdot \epsilon_t \\ \mathbf{f}_t &= \rho \cdot \mathbf{f}_{t-1} + \mathbf{u}_t \end{aligned} \tag{2}$$

  where λ denotes an $(N \times r)$ loading matrix; $\mathbf{f}$ is an $(r \times 1)$ vector of factors; ω shows the structure of the covariance matrix; and ϵ and u are stationary residuals. We note that if $|\rho|$ is less than 1, the factor is $I(0)$ or stationary; it is only non-stationary if $|\rho|$ is equal to 1.

The relationship between these approaches can be explained using the decomposition of the matrix $\mathbf{\Pi}$,

$$\mathbf{\Pi} = \alpha \cdot \beta^T \tag{3}$$

where β is the weight matrix of the cointegration relationships for each variable and α is the coefficient matrix of the cointegration relations.

Using (3), the factorial model,(2) can be expressed as

$$\mathbf{X}_t = \beta_\perp \cdot (\alpha_\perp^T \cdot \beta_\perp) \cdot \alpha_\perp^T \cdot \mathbf{X}_t + \alpha \cdot (\beta^T \cdot \alpha) \cdot \beta^T \cdot X_t \tag{4}$$

Escribano & Peña (1994) showed the relationship between the two approaches using expressions (1) and (4), both in terms of α and β.

We can solve the multivariate problem using CA with the relevant restrictions, which, as we note above, becomes intractable as the number of variables increase[1]; alternatively, we can analyze the dynamic common factors involved. This is where the Peña & Poncela (2006) test can be used to further investigate the problem. Peña & Poncela (2006) express the factorial model (2) as follows:

$$\mathbf{X}_t = \mu + [\lambda_1 \quad \lambda_2] \cdot [\mathbf{f}_{1,t} \quad \mathbf{f}_{2,t}]^T + \epsilon_t \tag{5}$$

That is, there are N observed variables with the corresponding mean vector μ. The common factors may belong to the subset $\mathbf{f}_1$ of non-stationary factors or to the subset $\mathbf{f}_2$ of stationary common factors, and λ_1 and λ_2 are their respective loading matrices. Thus, if the first group is made up of r_1 factors and the second of r_2, then there must be $N - (r_1 + r_2)$ cointegration relationships.

The advantage of this approach is that regardless of whether the observed series are stationary, we can identify stationary and non-stationary common factors. It

[1] Additionally, from the work of Cheung & Lay (1993), we know that the log-likelihood ratio for cointegration (trace test and maximal eigenvalue) are robust for autoregressive processes because standard lag selection criteria such as the AIC and SIC can be useful for choosing the right lag order for these tests. In contrast, when the processes show moving-average dependence, these information criteria perform poorly in selecting the proper lag of $\mathbf{\Delta X}$. $LR - tests$ for cointegration are rather sensitive to under-parameterization in the lag length, though not to over-parameterization.

also allows us to test whether the cointegration test correctly discriminates between cointegration relations and stationary factors. The test is implemented as follows:

$$\begin{aligned}
\mathbf{Y}_t &= \mathbf{X}_t - \mu \\
\hat{C}_k &= [\sum_{t=k+1}^{M} (\mathbf{Y}_t \cdot \mathbf{Y}_{t-k}^T)]^{-1} \cdot [\sum_{t=k+1}^{M} (\mathbf{Y}_t \cdot \mathbf{Y}_{t-k}^T)] \\
&\times [\sum_{t=k+1}^{M} (\mathbf{Y}_{t-k} \cdot \mathbf{Y}_{t-k}^T)]^{-1} \cdot [\sum_{t=k+1}^{M} (\mathbf{Y}_{t-k} \cdot \mathbf{Y}_t^T)]
\end{aligned} \tag{6}$$

According to Theorem 3 in Peña & Poncela (2006), the matrix $\hat{C}_k$ has $N - (r_1 + r_2)$ eigenvalues that converge in probability to zero as the sample size (M) tends to infinity and the number of lags, k, increases from 0 to K, such that $\frac{K}{M}$ tends to zero. In this way, the test sorts the eigenvalues (h_j) of the matrix $\hat{C}_k$ and obtains their sum, which tends asymptotically to a $\chi_{(N-r)^2}$ distribution, as follows:

$$S_{r=r_1+r_2} = (N-k) \cdot \sum_{j=1}^{M-r} \log(1-h_j) \tag{7}$$

The cumulative explanatory power (cep) of the factors with k-lags can be expressed as

$$cep_k = \frac{\sum_{j=1}^{M-r-k} h_j}{\sum_{i=1}^{M-r} h_i} \tag{8}$$

Once we have estimated the number of common factors, based on Peña & Poncela (2006), we can use an EM algorithm to determine the loading matrix of the expression (5), using as initial values the eigenvectors (V_j) of the following matrix:

$$C_1 = N^{-(2d+d^*)} \cdot \sum_{t=2}^{N} (\mathbf{Y}_{t-1} \cdot \mathbf{Y}t^T) \tag{9}$$

As shown, (9) operates with one lag, d is the order of integration of the observed series, and d^* is equal to 0 if d is greater than zero or 1 if d is equal to zero. Thus, once we order the eigenvalues in decreasing order, the initial value of the loading matrix would be the eigenvectors associated with the first r non-zero eigenvalues. Similarly, the initial values of the factors can be obtained from these eigenvectors and the observed variables as follows:

$$\begin{aligned}
\hat{\lambda}_0 &= \mathbf{V}_r \\
\hat{\mathbf{f}}_0 &= \mathbf{V}_r^T \cdot \mathbf{Y}
\end{aligned} \tag{10}$$

Once we recover the initial values of the factors ($\hat{\mathbf{f}}_0$), then Peña & Poncela (2006) use a contrast of stationarity (e.g., Augmented Dickey-Fuller test or ADF) to determine which factors are stationary or not.

From the initial values, the final results are estimated by applying the Kalman filter on the following state space representation because, as Bauer & Wagner (2009) show, this method yields better cointegration parameters:

$$\forall t = 1, \ldots, M$$

Measurement Equation:

$$\begin{bmatrix} x_{1,t} \\ \vdots \\ x_{N,t} \end{bmatrix} = \begin{bmatrix} \beta_{1,1} & \ldots & \beta_{1,K} \\ \vdots & \ddots & \vdots \\ \beta_{N,1} & \ldots & \beta_{N,K} \end{bmatrix} \cdot \begin{bmatrix} f_{1,t} \\ \vdots \\ f_{K,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \vdots \\ \epsilon_{N,t} \end{bmatrix}$$

$$\begin{bmatrix} \epsilon_{1,t} \\ \vdots \\ \epsilon_{N,t} \end{bmatrix} \sim \mathbf{N}(0, \Sigma) \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_N^2 \end{bmatrix} \tag{11}$$

Transition Equation:

$$\begin{bmatrix} f_{1,t+1} \\ \vdots \\ f_{K,t+1} \end{bmatrix} = \begin{bmatrix} \rho_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \rho_K \end{bmatrix} \cdot \begin{bmatrix} f_{1,t} \\ \vdots \\ f_{K,t} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ \vdots \\ u_{K,t} \end{bmatrix}$$

$$\begin{bmatrix} u_{1,t} \\ \vdots \\ u_{K,t} \end{bmatrix} \sim \mathbf{N}(0, \mathbf{I})$$

where $\beta_{i,j}$ is the load of factor j on variable i; σ_i^2 is the residual variance of variable i; and the parameter ρ is 1 when the stationary test shows that the factors are non-stationary.

3. Comparative analysis of the Cointegration and Factorial Dynamic Models: Experimental Study

3.1. Design of the Experimental Study

To measure the power of the Peña & Poncela (2006) test related to CA, we apply the tests in four different scenarios with a maximum of 100 time series ($\max(N) = 100$) with maximum sizes of 5,000 data points each ($\max(M) = 5,000$), computed from simulated factors and known loading matrices. Then, we compare the test results with the actual values used in the simulated data generating processes. We use the trace test from CA because it is more robust to moving average and non-Gaussian innovations than the maximal eigenvalue test due to its greater robustness to the skewness and excess kurtosis of the innovations, as Cheung & Lay (1993) showed. We use Ox packages to simulate factors and build non-stationary variables as follows:

- Scenario I: one common factor $I(1)$, integrated of order one, for each set of two variables, i.e., there are 50 common factors in 100 time series.

- Scenario II: one common factor $I(1)$ for each set of four variables, i.e., there are 25 common factors in 100 time series.

- Scenario III: one factor $I(1)$ and another $I(0)$, or stationary, for each set of four variables, i.e., there are 50 common factors in 100 time series.

- Scenario IV: two factors $I(1)$ for each set of four variables, i.e., there are 50 common factors in 100 time series.

The processes simulated from the loading matrices, i.e., the variables grouped by common factors and scenarios, are as follows:

- Scenario I:

$$\begin{aligned} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} &= \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \cdot f_t + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \\ f_t &= \sum_{j=1}^{t} u_j \\ (\epsilon_{1,t}, \epsilon_{2,t}, u_j) &\sim \mathbf{N}(0,1) i.i.d. \end{aligned} \tag{12}$$

- Scenario II:

$$\begin{aligned} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \\ Z_{4,t} \end{bmatrix} &= \begin{bmatrix} 1 \\ 0.25 \\ 0.50 \\ 0.75 \end{bmatrix} \cdot f_t + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \end{bmatrix} \\ f_t &= \sum_{j=1}^{t} u_j \\ (\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t}, \epsilon_{4,t}, u_j) &\sim \mathbf{N}(0,1) i.i.d. \end{aligned} \tag{13}$$

- Scenario III:

$$\begin{aligned} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \\ Z_{4,t} \end{bmatrix} &= \begin{bmatrix} 1 & 0.25 \\ 0.75 & 0.50 \\ 0.50 & 0.75 \\ 0.25 & 1 \end{bmatrix} \cdot \begin{bmatrix} f_{1,t} \\ f_{2,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \end{bmatrix} \\ f_{1,t} &= f_{1,t-1} + u_{1,t} \\ f_{2,t} &= 0.8 \cdot f_{2,t-1} + u_{2,t} \\ (\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t}, \epsilon_{4,t}, u_{1,t}, u_{2,t}) &\sim \mathbf{N}(0,1) i.i.d. \end{aligned} \tag{14}$$

- Scenario IV:

$$\begin{aligned}
\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \\ Z_{4,t} \end{bmatrix} &= \begin{bmatrix} 1 & 0.25 \\ 0.75 & 0.50 \\ 0.50 & 0.75 \\ 0.25 & 1 \end{bmatrix} \cdot \begin{bmatrix} f_{1,t} \\ f_{2,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \end{bmatrix} \\
f_{1,t} &= f_{1,t-1} + u_{1,t} \\
f_{2,t} &= f_{2,t-1} + u_{2,t} \\
(\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t}, \epsilon_{4,t}, u_{1,t}, u_{2,t}) &\sim \mathbf{N}(0,1) i.i.d.
\end{aligned} \tag{15}$$

In Table 1, we show a statistical summary of the simulated values for each of the four scenarios. As we can see, the variables and the factors are not neither normal nor stationary, except factor $I(0)$ of Scenario-III. It also highlights that the variables have a minimum in the ADF test for stationarity nearest to being accepted; that is, though the generating factors are $I(1)$, in some cases, the variables constructed from these may seem $I(0)$, at least when 95 percent confidence intervals are considered.

From the results shown in Table 2, when we use the trace test, we can see that if there is one common factor and sufficient degrees of freedom, the trace tests determine the correct number of factors without errors. However, when the series are two-cointegrated as in Scenario-IV, the performance of this test declines. Thus, this result suggests that if the variables have more than one common factor, this test does not discriminate them correctly. Moreover, if one of them is $I(0)$, as in Scenario-III, this test does not detect it. Another drawback is that this technique does not permit estimation of the loading matrix associated with each common factor.

However, from the results shown in Table 2, when we use the Peña & Poncela (2006) test, we can see that when the number of observations for each variable (M) is higher than the number of variables (N), the test converges quickly to the correct number of factors. Therefore, the test is valid for samples where the number of observations by variables is greater than the number of variables. Additionally, in Scenario-III, there is one common factor $I(0)$ that is not detected by the trace test, but the Peña & Poncela (2006) test detects it.

3.2. Results

We apply the trace test and Peña & Poncela (2006) test on different samples, i.e., we analyze the results of the tests in samples with different numbers of variables (N) and different numbers of observations by variable (M) in each of the four scenarios described. The sample data range from 8 observations of 25 variables to 5,000 observations of 100 variables. In Table 2, we show the errors of these tests, i.e., the difference between the number of real factors and the number of factors estimated.

Table 1: Statistical Summary of simulated data.

Scenario-I						
Data	Factor $I(1)$			Variables		
Statistics	min.	mean	max.	min.	mean	max.
Mean	-81.33	-3.82	83.76	-81.33	-2.69	83.76
Std. Dev.	12.84	23.80	48.53	6.49	18.39	48.56
Skewness	-1.16	-0.01	1.09	-1.16	-0.01	1.09
Excess Kurtosis	-1.39	-0.75	1.07	-1.39	-0.73	1.06
Min.	-143.74	-48.38	1.78	-144.02	-38.87	2.69
Max.	2.67	41.98	157.12	3.19	30.23	156.95
Normality	139.04	560.37	4,824.90	100.69	438.16	2,997.01
ADF	-2.74	-1.69	0.12	-3.19(*)	-1.76	-0.01

Scenario-II						
Data	Factor I(1)			Variables		
Statistics	min.	mean	max.	min.	mean	max.
Mean	-111.85	9.02	67.44	-111.83	3.34	67.43
Std. Dev.	11.298	18.66	50.98	2.98	12.53	50.99
Skewness	-1.52	0.22	1.25	-1.51	0.21	1.24
Excess Kurtosis	-1.45	-0.53	2.81	-1.45	-0.52	2.77
Min.	-183.62	-40.41	-0.03	-184.87	-20.51	0.21
Max.	3.57	52.86	159.79	2.29	28.81	160.15
Normality	42.77	295.19	4,706.60	34.39	279.60	4,679.50
ADF	-2.53	-2.00	-0.34	-3.09(*)	-1.93	-0.29

Scenario-III									
Data	Factor $I(1)$			Factor $I(0)$			Variables		
Statistics	min.	mean	max.	min.	mean	max.	min.	mean	max.
Mean	-81.32	-3.82	83.76	-0.03	-0.01	0.06	-81.33	-2.69	83.75
Std. Dev.	12.84	23.80	48.53	0.98	0.99	1.02	6.56	18.41	48.56
Skewness	-1.16	-0.01	1.09	-0.06	0.01	0.07	-1.16	-0.01	1.09
Excess Kurtosis	-1.40	-0.75	1.07	-0.26	-0.03	0.12	-1.39	-0.72	1.06
Min.	-143.74	-48.38	1.78	-4.55	-3.54	-3.10	-144.15	-40.04	1.98
Max.	2.67	41.98	157.12	3.09	3.56	4.49	3.28	31.49	157.15
Normality	52.19	379.68	4,824.90	0.02(**)	1.12(**)	16.31(*)	39.21	366.54	4,786.30
ADF	-2.67	-1.78	0.11	-71.76(**)	-29.53(**)	-24.06(**)	-3.39(*)	-1.80	0.04

Scenario-IV									
Data	Factor $I(1)$			Factor $I(1)$			Variables		
Statistics	min.	mean	max.	min.	mean	max.	min.	mean	max.
Mean	-111.85	9.02	67.44	-68.99	-8.12	98.83	-123.94	1.65	114.54
Std. Dev.	11.30	18.66	50.98	11.81	22.94	50.62	11.72	23.47	53.69
Skewness	-1.52	0.22	1.24	-0.99	-0.18	0.86	-1.27	-0.06	1.25
Excess Kurtosis	-1.45	-0.53	2.81	-1.18	-0.51	0.86	-1.63	-0.71	1.47
Min.	-183.62	-40.41	-0.03	-164.28	-54.02	-0.71	-194.71	-46.84	0.44
Max.	3.57	52.86	159.79	-0.77	27.20	162.75	-1.05	48.08	189.18
Normality	42.77	295.19	4706.60	33.11	375.76	2,948.40	18.74	320.73	4,943.80
ADF	-2.43	-2.00	-0.34	-2.56	-1.76	0.36	-3.09(*)	-1.71	0.24

Note: Normality is tested using the Jarque-Bera test with a Chi^2_2 distribution. The ADF is the Augmented Dickey-Fuller test on unit roots with a constant and without a trend. Significance at 95 and 99 per cent levels are indicated by (*) and (**), respectively.

Now, we analyze the impact on the results of the two tests of the level of confidence required and the number of lags used. We apply the tests to the same samples used in Table 2 but vary, in increments of one, the lag number from 2 to 100 and the confidence level from 0.5 to 0.9999 in increments of 0.1 to 0.9 and then in increments of 0.01 to 0.99, and so on. The results[2], which confirm in all cases those in Table 2, can be summarized as follows.

[2]Available upon request to the authors

Table 2: Errors in the number of factors extracted by cointegration and factorial tests.

Errors cointegration test						Errors factorial test					
Scenario-I											
$\frac{M}{N}$	8	20	40	80	100	$\frac{M}{N}$	8	20	40	80	100
25	-3	*ndf*	*ndf*	*ndf*	*ndf*	25	2	-7	1	23	41
50	0	-1	*ndf*	*ndf*	*ndf*	50	1	0	-12	0	3
100	0	1	-6	*ndf*	*ndf*	100	0	0	2	-25	-29
250	0	0	-2	-6	2	250	0	0	1	3	1
500	0	0	0	-1	0	500	0	0	0	1	1
1,000	0	0	0	0	0	1,000	0	0	0	0	1
2,500	0	0	0	0	0	2,500	0	0	0	0	0
5,000	0	0	0	0	0	5,000	0	0	0	0	0
Scenario-II											
$\frac{M}{N}$	8	20	40	80	100	$\frac{M}{N}$	8	20	40	80	100
25	-1	*ndf*	*ndf*	*ndf*	*ndf*	25	1	-1	0	3	16
50	0	1	*ndf*	*ndf*	*ndf*	50	0	-1	0	0	-1
100	0	0	-15	25	*ndf*	100	0	-1	0	0	-1
250	0	0	0	0	-8	250	0	0	-1	-5	-13
500	0	0	0	-1	3	500	0	0	0	0	0
1,000	0	0	0	0	0	1,000	0	0	0	0	0
2,500	0	0	0	0	0	2,500	0	0	0	0	0
5,000	0	0	0	0	0	5,000	0	0	0	0	0
Scenario-III											
$\frac{M}{N}$	8	20	40	80	100	$\frac{M}{N}$	8	20	40	80	100
25	0	*ndf*	*ndf*	*ndf*	*ndf*	25	3	-7	4	23	41
50	1	-2	*ndf*	*ndf*	*ndf*	50	2	3	1	0	3
100	1	-2	-3	*ndf*	*ndf*	100	1	2	4	0	2
250	2	0	-1	0	-8	250	0	1	2	6	3
500	2	5	5	0	-3	500	0	0	1	3	6
1,000	2	5	10	15	14	1,000	0	0	0	0	1
2,500	2	5	10	20	25	2,500	0	0	0	0	0
5,000	2	5	10	20	25	5,000	0	0	0	0	0
Scenario-IV											
$\frac{M}{N}$	8	20	40	80	100	$\frac{M}{N}$	8	20	40	80	100
25	1	*ndf*	*ndf*	*ndf*	*ndf*	25	3	-6	1	23	41
50	1	-3	*ndf*	*ndf*	*ndf*	50	1	3	-11	0	3
100	0	-2	-3	*ndf*	*ndf*	100	0	0	2	0	-1
250	0	0	2	-3	-8	250	0	0	1	4	2
500	0	0	0	0	-4	500	0	0	0	0	4
1,000	0	0	0	1	2	1,000	0	0	0	0	0
2,500	0	0	0	0	2	2,500	0	0	0	0	0
5,000	0	0	1	0	0	5,000	0	0	0	0	0

Note: *ndf* denotes that there were not sufficient degrees of freedom.

For the trace test, the results primarily depend on the degrees of freedom: by increasing the number of lags, the cointegrating rank converges to the right value, provided that the number of observations and variables are large enough. In contrast, when the number of observations decreases, the best results are achieved with lower lags, as this way, the degrees of freedom are higher. Therefore, the

trace test results depend on the lag number used, which in turn depends on the number of observations.

For the Peña & Poncela (2006) test, the results show that as M approaches N, we can reduce the error by increasing the number of lags. Nevertheless, if N is greater than M, we achieve the correct value only when the number of variables N is not much higher than M. Again, this is due to the number of degrees of freedom. However, the Peña & Poncela (2006) test is less sensitive to the degrees of freedom than the trace test, as the errors are smaller.

Finally, we analyze whether the methodology proposed by Peña & Poncela (2006) allows us recover the true common factors, by determining if they are stationary or not and by determining the corresponding loading matrices. To do so, we apply the two tests to six variables (Z), two of Scenario-I and four of Scenario-IV. In Table 3, we present the results of the trace test, which, as expected based on the previous results, correctly identify the number of common factors but do not inform us about the two groups of variables. Because the cointegration rank is three, we could assume that there are three sets of two cointegrated variables.

Table 3: Cointegration test in the subset Z.

Variable	Lag	Fact	Trace test	Prob.	Max. eigen. test	Prob.
6	1	3	34.12	0.064	25.81	0.013(*)
6	2	3	27.18	0.284	20.18	0.097
6	3	3	28.89	0.121	22.12	0.051

Similarly, the Peña & Poncela (2006) test identifies three factors with one lag, a t-value of 11.62 and a p-value of 0.236. In addition, the ADF test of the 3 factors recovered by the product of the matrix transpose of the eigenvectors associated with the three largest eigenvalues and the observed variables resulted in three factors being $I(1)$, as we can see in Table 4. Thus, both methods yield the number of factors, but it is interesting to note how the methodology proposed by Peña & Poncela (2006) recovers and groups the factors.

Table 4: Testing ADF on common factors.

Statistics	f_1	f_2	f_3
ADF	-1.506	-1.913	-2.763
Lags	2	2	2
AIC	0.0000	0.0000	0.0343

Note: The Augmented Dickey Fuller model is estimated with a constant. The null hypothesis is that the variable is integrated with order 1. This hypothesis is rejected at a value of -2.86 and -3.43 at 5 and 1 per cent significance levels, respectively.

The parameters σ and β of the loading matrix are shown in Table 5.

The Start, the EM algorithm estimate and the True factors are shown in Figure 1, Figure 2 and Figure 3, respectively, for each factor. As we can see, the factors estimated after the maximization are quite similar to the true values.

Finally, we show, for the Start factors and estimated factors, as suggested by Peña & Poncela (2006), the mean absolute error (MAE) and root mean square error (RMSE) in Table 6 to measure the efficiency of the EM algorithm.

Table 5: Estimated parameters and statistics.

Concept	Var_1	Var_2	Var_3	Var_4	Var_5	Var_6
σ	1.5222	0.8391	1.4767	1.4590	1.3841	1.3459
t-value	9.67(**)	2.97(*)	13.71(**)	12.99(**)	11.49(**)	10.58(**)
$\beta(\cdot,1)$	-0.035	-0.006	0.348	0.546	0.739	0.942
t-value	-0.32	-0.05	3.47(**)	2.84(*)	5.84(**)	5.10(**)
$\beta(\cdot,2)$	0.053	0.035	0.886	0.638	0.385	0.130
t-value	0.31	0.27	5.71(**)	4.49(**)	2.38(*)	1.95
$\beta(\cdot,3)$	0.888	0.476	0.074	0.062	0.026	0.004
t-value	6.35(**)	3.22(**)	0.60	0.50	0.16	0.03
Start f_1	-0.328	-0.165	0.068	0.375	0.684	1
Start f_2	0.076	0.038	1	0.555	0.112	-0.339
Start f_3	1	0.506	-0.052	0.063	0.165	0.268
True f_1	0	0	0.25	0.5	0.75	1
True f_2	0	0	1	0.75	0.5	0.25
True f_3	1	0.5	0	0	0	0

Note: To facilitate the comparison, we present the initial weights (resulting from the eigenvectors of the autocovariance matrix with one lag) or START, the estimates after the EM algorithm or Loading, with the TRUE weights that generated the series. $(**)$ and $(*)$ show significance at the 99 and 95 percent confidence intervals, respectively. The average computation time for the test was 1:03 (min:sec), while the average EM estimation time was 6:17 (min:sec). We used a DELL Precision M 6500 mobile workstation with 32 GB (RAM), Intel Core i7 (processor) and two HDDs of 465 GB each.

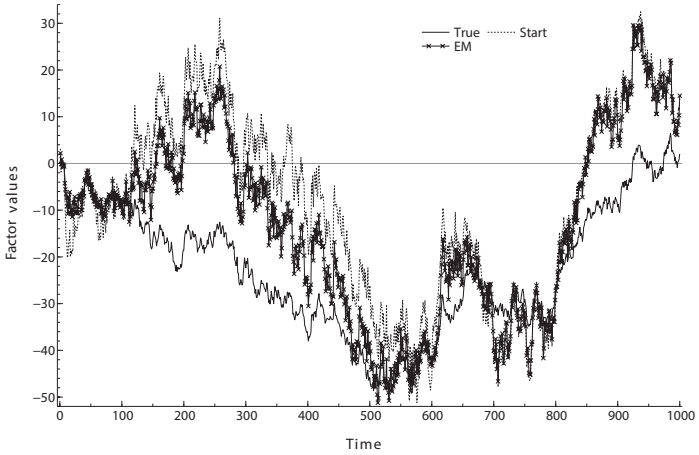

Figure 1: Start, estimated EM and true factor-1.

Table 6: Mean Absolute Error and Root Mean Square Error of the estimates.

Factors	f_1		f_2		f_3	
Estimations	MAE	RMSE	MAE	RMSE	MAE	RMSE
Start	16.1086	19.3675	11.1100	12.9020	6.8305	7.9816
EM algorithm	14.6399	16.2678	4.3636	5.4261	1.143	1.4386

In summary, the Peña & Poncela (2006) test is a perfect complement to CA because it is more consistent when the variables are involved in more than one cointegration relationship and/or when there are common stationary factors. Ad-

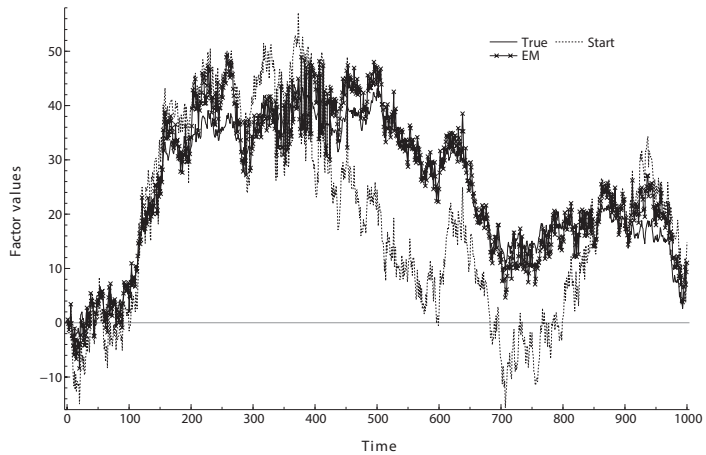

Figure 2: Start, estimated EM and true factor-2.

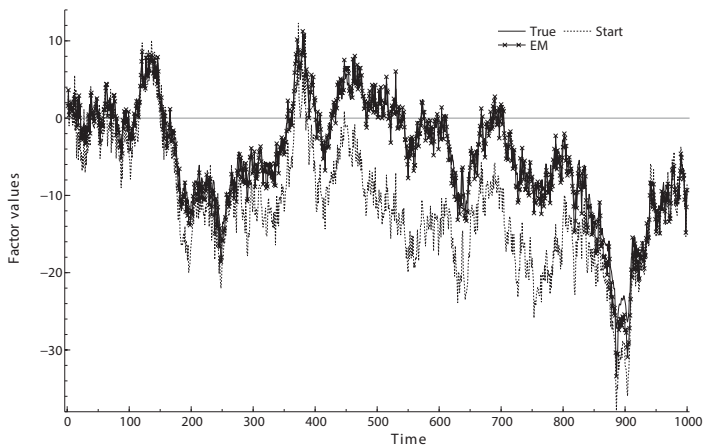

Figure 3: Start, estimated EM and true factor-3.

ditionally, the Peña & Poncela (2006) test is less sensitive to a low number of degrees of freedom and allows us to estimate the loading matrix associated with each common factor.

4. Financial Market Data Analysis

Several works, such as those by Baillie & Bollerslev (1994) and Diebold, Gardeazabal & Yilmaz (1994), have tested the market efficiency hypothesis for the spot exchange rate through the triangular arbitrage relation using the CA. The cross exchange rate, or exchange rate of triangular arbitrage, is defined as follows:

$$c_{i,t} = s_{j,t} \cdot s_{h,t} \tag{16}$$

In (16), for day t and three currencies (i, j and h), c is the cross exchange rate and s is the spot exchange market rate. In addition, each exchange market rate has an offer price (*bid*) and a demand price (*ask*). The difference between these two prices (*spread*) indicates the transaction costs. Taking logarithms in (16) to obtain a linear triangular arbitrage relationship and including the bid-ask spread, we obtain:

$$\begin{aligned} \ln c_{i,t,\psi} &= \ln s_{j,t,\psi} + \ln s_{h,t,\psi} \\ \ln s_{i,t,b} &= \delta_i + \ln s_{j,t,a} \\ \psi &= \{a, b\} \end{aligned} \tag{17}$$

Where a is the ask price, b is the bid price and δ_i is the transaction cost for the currency i. When we operate with three currencies and their respective bid and ask prices, there are six possible triangular relations among them: three relationships of the ask type and three of the bid type. Thus, the problem is to determine the true number of relations of long-term equilibrium among the three currencies. For this, the general expression for arbitrage strategies (17) becomes the following econometric model:

$$\begin{gathered} \ln s_{i,t,\psi} = \delta_{i,\psi} + \delta_{i,j,\psi} \cdot \ln s_{j,t,\psi} + \delta_{i,h,\psi} \cdot \ln s_{h,t,\psi} + u_{i,t,\psi} \\ i \neq j \neq h \quad \psi = \{a, b\} \quad u_{i,t,\psi} \sim i.i.d.(0, \varpi^2) \end{gathered} \tag{18}$$

where $\delta_{i,\psi}$ is the long-run result of triangular arbitrage for currency i with price ψ and the parameters $\delta_{i,j,\psi}$ and $\delta_{i,h,\psi}$ are the weights in an arbitrage portfolio of currencies j and h, respectively, whose expected values according to (17) are $|\delta_{i,\bullet,\psi}| = 1$. In this context, we could estimate long-run relationships with the CA or the DFA, but from a financial perspective, only two relationships are possible.

For our analysis, we use daily data on the Bid and Ask close prices of the exchange rate of the JPY (Japan yen) against USD (US dollar) and EUR (EMU euro), from 1-july-2002 to 30-december-2011: 2,478 observations from the Reuters platform of inter-dealer markets. We first test univariate evidence of unit roots in the sample (logarithms of the exchange rate and first difference of the log-exchange rate) and reject the stationary hypothesis for all exchange rates in the data. Table7 shows that the logarithms of the exchange rates are non-stationary, while the first difference of log (returns) is stationary; therefore, all the series, in logarithms, have a unit root.

Then, we estimate the number of long-run relationships using the trace test and Peña & Poncela (2006) test. The results are presented in Table-8. The CA results show four long-run relationships among the six spot exchange market rates (bid and ask price of JPY against USD and EUR), while the DFA indicates only two cointegration relations (or equivalently, four common factors). Therefore, to test whether triangular arbitrage opportunities exist, while the factorial analysis result is correct, the trace test is inconsistent with two market cointegration relations, one for the bid price and the other for the ask price, here due to the selection criteria used performing poorly in selecting the proper lag, as Cheung & Lay (1993) note.

Table 7: Stationarity test of the spot exchange rate.

ADF test of LN(exchange rate)				
Currency	Ask_{USD}	Bid_{USD}	Ask_{EUR}	Bid_{EUR}
EUR	-2.5583 [0]	-2.5601 [0]	–	–
JPY	-0.1143 [16]	-0.1145 [16]	-0.6834 [30]	-0.6844 [30]
ADF test first difference of LN(exchange rate)				
Currency	Ask_{USD}	Bid_{USD}	Ask_{EUR}	Bid_{EUR}
EUR	-50.0916 [0]	-50.1425 [0]	–	–
JPY	-13.0692 [15]	-13.0736 [15]	-9.4456 [29]	-9.4431 [29]

Note: The ADF test is estimated with a constant, and the [number of lags] is estimated using the AIC. Critical values of rejection of non-stationarity are -2.86 and -3.44 at significance levels of 5 and 1 per cent, respectively.

Table 8: Cointegration rank with ask and bid prices.

Currency	AIC lags	rank	trace test	p-value	common factors	test value	p-value
JPY	11	4	10.549	[0.593]	4	8.459	[0.489]

Note: Columns 2-5 show the results (using AIC criteria to select lags) of traditional cointegration analyses (trace test). Columns 6-8 show the results of the dynamic factorial test, that is, the number of variables (six exchange rates) minus the number of common dynamic factors indicates the cointegration rank.

5. Conclusions

When, in a univariate time series framework, the variable involved turns out to be non-stationary, we transform it. However, when the problem is multivariate, doing so may involve a loss of relevant information for modelling the data. In this case, the cointegration analysis and the determination of the cointegration relations allow us to model the relationships of long-run equilibrium among the variables involved.

However, when variables are involved in more than one cointegration relationship, estimating the parameters of the cointegration relations is conditioned upon known constrains that define the problem. Furthermore, estimating the correct number of non-correlated factors when one of them is stationary is also difficult because the tests that are commonly used cannot distinguish between relationships of cointegration and stationary factors. In this paper, we have analyzed simulated series with known data-generating processes, the power of the cointegration test of the trace and the common factors test proposed by Peña & Poncela (2006).

The results for both methodologies are similar, with high numbers of degrees of freedom, except when there are stationary common factors, where the cointegration tests do not identify the correct number of factors. Moreover, when the data sample sizes decrease, and/or the cointegration relationships involve more than two variables, the Peña & Poncela (2006) methodology results overcome the cointegration results. An added advantage of the Peña & Poncela (2006) methodology compared with the cointegration analysis is that it allows for a better approximation to the data-generating process. In short, whereas the trace test presented drawbacks for the stationary factors, a low number of observations and a high-order autocorrelation of the series, Peña & Poncela (2006) test results showed

more consistency. These conclusions were also confirmed when real market data on spot exchange rates were used.

Acknowledgement

This work was supported by research projects funded by the Spanish Goverment (MEYC ECO 2012-36685) and the Generalitat Valenciana (PROMETEO-II/ 2013/ 015).

[Received: December 2013 — Accepted: November 2014]

References

Ahlgren, N. & Antell, J. (2008), 'Bootstrap and fast double bootstrap tests of cointegration rank with financial time series', *Computational Statistics and Data Analysis* **52**(10), 4754–4767.

Baillie, R. T. & Bollerslev, T. (1994), 'Cointegration, fractional cointegration, and exchange rate dynamics', *Journal of Finance* **49**(2), 737–745.

Banerjee, A., Dolado, J. J. & Mestre, R. (1998), 'Error correction mechanism tests for cointegration in a single-equation framework', *Journal of Time Series Analysis* **19**(3), 267–283.

Bauer, D. & Wagner, M. (2009), 'Using subspace algorithm cointegration analysis: Simulation performance and application to term structure', *Computational Statistics and Data Analysis* **53**(6), 1954–1973.

Bayer, C. & Hanck, C. (2013), 'Combining non-cointegration tests', *Journal of Time Series Analysis* **34**(1), 83–95.

Cavaliere, G. & Taylor, A. M. (2006), 'Testing the null of co-integration in the presence of variance breaks', *Journal of Time Series Analysis* **27**(4), 613–636.

Chen, Y. P., Huang, H. C. & Tu, I. P. (2010), 'A new approach for selecting the number of factors', *Computational Statistics and Data Analysis* **54**(12), 2990–2998.

Cheung, Y. W. & Lay, K. S. (1993), 'Finite-sample sizes of Johansen's likelihood ratio tests for cointegration', *Oxford Bulletin of Economics* **55**(3), 313–328.

Correal, M. E. & Peña, D. (2008), 'Thresold dynamic factor model', *Revista Colombiana de Estadística* **31**(2), 183–192.

Cubadda, G. (2007), 'A unifying framework for analysing common cyclical features in cointegrated time series', *Computational Statistics and Data Analysis* **52**(2), 896–906.

Davidson, J. & Monticini, A. (2007), 'Test for cointegration with structural breaks base on subsamples', *Computational Statistics and Data Analysis* **54**(11), 2498–2511.

Diebold, F. X., Gardeazabal, J. & Yilmaz, K. (1994), 'On cointegration and exchange rate dynamics', *Journal of Finance* **49**(2), 727–735.

Dittmann, I. (2000), 'Residual-based tests for fractional cointegration: A Monte Carlo study', *Journal of Time Series Analysis* **21**(6), 615–647.

Doornik, J. A. & O'Brien, R. J. (2002), 'Numerically stable cointegration analysis', *Computational Statistics and Data Analysis* **41**(1), 185–193.

Engle, R. F. & Granger, C. W. J. (1987), 'Cointegration and error correction: Representation, estimation and testing', *Econometrica* **55**(2), 251–276.

Escribano, A. & Peña, D. (1994), 'Cointegration and common factors', *Journal of Time Series Analysis* **15**(6), 577–586.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2005), 'The generalized dynamic-factor model: One-sided estimation and forecasting', *Journal of the American Statistical Association* **100**(471), 577–586.

González, M. & Nave, J. M. (2010), 'Portfolio immunization using Independent Component Analysis', *The Spanish Review of Financial Economics* **21**(1), 37–46.

Gonzalo, J. (1994), 'Five alternative methods of estimating long-run equilibrium relationshisps', *Journal of Econometrics* **60**(1-2), 203–233.

Gonzalo, J. & Granger, C. W. J. (1995), 'Estimation of common long-memory components in cointegrated systems', *Journal of Business and Economic Statistics* **13**(1), 27–35.

Gonzalo, J. & Lee, T. H. (1998), 'Pitfalls in testing for long-run relationshisps', *Journal of Econometrics* **86**(1), 129–154.

Hallin, M. & Liska, R. (2007), 'Determining the number of factors in the general dynamic factor model', *Journal of the American Statistical Association* **102**(478), 603–617.

Hu, Y. P. & Chou, R. J. (2003), 'A dynamic factor model', *Journal of Time Series Analysis* **24**(5), 529–538.

Hu, Y. P. & Chou, R. J. (2004), 'On the Peña-Box model', *Journal of Time Series Analysis* **25**(6), 811–830.

Jing, L. & Junsoo, L. (2010), 'ADL tests for threshold cointegration', *Journal of Time Series Analysis* **31**(4), 241–254.

Kapetanios, G. & Marcellino, M. (2009), 'A parametric estimation method for dynamic factor models of large dimensions', *Journal of Time Series Analysis* **30**(2), 208–238.

Lansangan, J. R. & Barrios, E. B. (2009), 'Principal components analysis of non-stationary time series data', *Statistics and Computing* **19**(2), 173–187.

Li, Q., Pan, J. & Yao, Q. (2009), 'On determination of cointegration ranks', *Statistics and Its Interface* **2**(1), 45–56.

Lopes, H. F., Gamerman, D. & Salazar, E. (2011), 'Generalized spatial dynamic factor models', *Computational Statistics and Data Analysis* **55**(3), 1319–1330.

Lorenzo-Seva, U., Timmerman, M. E. & Kiers, H. A. (2011), 'The Hull method for selecting the number of common factors', *Multivariate Behavioral Research* **46**(2), 340–364.

Miller, J. I. (2010), 'Cointegrating regressions with messy regressors and an application to mixed-frecuency series', *Journal of Time Series Analysis* **31**(4), 255–277.

Pan, J. & Yao, Q. (2010), 'Modelling multiple time series via common factors', *Biometrika* **95**(2), 365–379.

Park, B., Mammen, E., Hardle, W. & Borak, S. (2009), 'Modelling dynamic semiparametric factor models', *Journal of the American Statistical Association* **104**(485), 284–298.

Park, S., Ahn, S. K. & Cho, S. (2011), 'Modelling dynamic semiparametric factor models', *Computational Statistics and Data Analysis* **55**(9), 2065–2618.

Peña, D. & Box, G. E. P. (1987), 'Identifying a simplifying structure in time series', *Journal of the American Statistical Association* **82**(399), 863–843.

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

Peña, D. & Sanchez, I. (2007), 'Measuring the advantages of multivariate versus univariate forecasts', *Journal of Time Series Analysis* **28**(6), 886–909.

Pesavento, E. (2007), 'Residuals-based tests for the null of no-cointegration: An analytical comparision', *Journal of Time Series Analysis* **28**(1), 111–137.

Stock, J. H. & Watson, M. W. (1988), 'Testing for common trends', *Journal of the American Statistical Association* **83**(404), 1097–1107.

Trenkler, C., Saikkonen, P. & Lütkepohl, H. (2007), 'Testing for the cointegrating rank of a VAR process with level shift and trend break', *Journal of Time Series Analysis* **29**(2), 331–358.

Westerlund, J. & Egderton, D. L. (2006), 'New improved tests for cointegration with structural breaks', *Journal of Time Series Analysis* **28**(2), 188–224.

Widaman, K. F. (1993), 'Common Factors Analysis versus Principal Component Analysis: Differential bias in representing model', *Multivariate Behavioral Research* **28**(3), 263–311.

Zhang, H. (2009), 'Comparación entre dos métodos de reducción de dimensionalidad en series de tiempo', *Revista Colombiana de Estadística* **32**(2), 189–212.