Unit Roots and Cointegrating Matrix Estimation using Subspace Methods

Alfredo G. Hiernaux Miguel Jerez* José Casals

December 13, 2005

Abstract

We propose a new procedure to detect unit roots based on subspace methods. It has three main original features. First, the same method can be applied to single or multiple time series. Second, it employs a flexible family of information criteria, which loss functions can be adapted to the statistical properties of the data. Last, it does not require the specification of a stochastic process for the series analyzed. Also, we provide a consistent estimator of the cointegrating rank and the cointegrating matrix. Simulation exercises show that the procedure has good finite sample properties. An example illustrates its application to real time series.

Keywords: State-space models, subspace methods, unit roots, cointegration
JEL Classification: C15-C32-C51-C87

*Departamento de Fundamentos del Análisis Econmico II. Facultad de Ciencias Económicas. Campus de Somosaguas. 28223 Madrid. email: mjerez@ccee.ucm.es, tel: (+34) 91 394 23 61, fax: 91 394 25 91. The authors gratefully acknowledge financial support from Ministerio de Educación y Ciencia, ref. SEJ2005-07388.

1 INTRODUCTION

This article describes a method to detect unit roots in time series. It derives from an important tradition of time series analysis in aeronautics and telecommunications engineering, which powerful instruments have not been fully assimilated by mainstream Statistics. These techniques are generically described as "system identification" and, since the seminal work of Ho and Kalman (1966), concentrate in modeling a data set using a state-space (SS) representation with no a priori restrictions. The literature usually treats this problem from a geometric or algebraic point of view, although there are also statistical approaches such as those of Akaike (1975, 1976) and Larimore (1983, 1990).

Since the 90s, system identification has been led by subspace methods. Many of these procedures build on the idea that, if the states of a system were observed, its parameter matrices could be estimated by ordinary least squares. Since the states are typically unobserved, the identification problem reduces to approximating the states from the data, with no a priori constraints on the system structure. In this context, the data is structured in past and future information matrices. From them, one may derive the state sequences by solving a reduced-rank weighted least squares problem.

In comparison with mainstream time series analysis (Box and Jenkins, 1976; Jenkins and Alavi, 1981; Tiao and Box, 1981; Tiao and Tsay, 1989) these methods have four main advantages:

1. They allow the specification of a general linear model directly from the data, without a priori knowledge of the process structure.
2. They make no distinction between the univariate and multivariate cases.
3. They are based on robust and computationally efficient tools of numerical algebra and, as a consequence,
4. iterations are not required, therefore avoiding convergence problems.

On the other hand, applying subspace methods to non-physical time series presents important challenges. For example, time series in engineering are typically non-seasonal and stationary. Consequently, current methods cannot accommodate these features so common in other contexts.

This paper deals with one of these topics: the detection of unit roots. To this end, we build on the Canonical Correlation Analysis algorithm (CCA) by Larimore (1983). This method is convenient because its choice of weighting matrices

provides, as a by-product, an efficient computation of the canonical correlation coefficients (CCC) between the past and future information subspaces. This property provides both, a statistical interpretation of the procedure and the main foundation of our method.

The basic idea is that the CCCs corresponding to unit roots converge exponentially to their true values, while the rest exhibit a much slower convergence. This property, known as superconsistency, allows distinguishing both kinds of correlations and has been used in some previous works. Poskitt (2000) used this property to detect unit roots and test for cointegration relationships. Later Bauer and Wagner (2002), hereafter BW, refined this approach by deriving an information criterion and several tests for the cointegrating rank using subspace methods. Our starting point is, therefore, similar to that of BW, with three main differences.

First, our work is focused on using information criteria, not statistical tests. Second, the information criterion suggested by BW does not take into account a relevant variable which inclusion in the criterion improves substantially its discrimination capacity. Third, BW describes the mathematical form of their information criterion as "somewhat heuristic", meaning that the weight of the different terms is arbitrary fixed. By contrast, we choose these weights empirically by means of simulation methods.

The structure of the paper is as follows. Section 2 defines the notation and summarizes the basic results that will be used later. Section 3 presents the procedure to detect unit roots and Section 4 derives consistent estimates of the cointegrating matrix. In section 5 we analyze the properties of these methods in finite samples by a set of simulation exercises. Section 6 illustrates the empirical application of the method with a real case and finally, Section 7 provides some concluding remarks and indicates how to obtain, via internet, a MATLAB toolbox for time series modeling which implements all the computational procedures required.

2 SUBSPACE METHODS AND CANONICAL CORRELATIONS

Consider a linear fixed-coefficients system that can be described by a SS model in innovations form,

$$\boldsymbol{x_{t+1}} = \boldsymbol{\Phi x_t} + \boldsymbol{E\psi_t} \quad (1)$$

$$\boldsymbol{z_t} = \boldsymbol{H x_t} + \boldsymbol{\psi_t} \quad (2)$$

where $\boldsymbol{x_t} \in \mathbb{R}^n$ is a state vector, $\boldsymbol{z_t} \in \mathbb{R}^m$ is an observable output vector and $\boldsymbol{\psi_t} \in \mathbb{R}^m$ is an innovation vector such that $iidN(\boldsymbol{0}, \boldsymbol{Q})$. Assume also that the system does not include exogenous variables. No generality is lost by these assumptions. First, this model is general in the sense that any fixed-coefficients SS model can be written in innovations form (see, Casals et al., 1999, Theorem 1). Second, any model with inputs can be decomposed into an input-related model and an error-related model, being this a rather standard result of linear system theory (see e.g., Chui and Chen, 1999).

Subspace methods derive from the representation of (1-2) in matrix form. Recursively from (1) we obtain,

$$\boldsymbol{x_t} = \boldsymbol{\Phi^t x_0} + \sum_{i=0}^{t-1} \boldsymbol{\Phi^i E \psi_{t-i-1}} \tag{3}$$

and substituting (3) into the observation equation (2),

$$\boldsymbol{z_t} = \boldsymbol{H \Phi^t x_0} + \boldsymbol{H} \sum_{i=0}^{t-1} \boldsymbol{\Phi^i E \psi_{t-i-1}} + \boldsymbol{\psi_t} \tag{4}$$

Therefore the endogenous variable, $\boldsymbol{z_t}$, depends on the initial state vector, $\boldsymbol{x_0}$, and the present and past innovation values, $\boldsymbol{\psi_t}$. Equation (4) can be written in matrix form as,

$$\boldsymbol{Z_k} = \boldsymbol{O_i X_0} + \boldsymbol{V_i \Psi_k} \tag{5}$$

where the subscript k denotes the row space of $\boldsymbol{Z_k}$. This equation requires the following matrices related to the data:

1) Block-Hankel Matrices (BHM): BHMs are partitioned in two blocks with row spaces p and f. Different choices for both, p and f, are discussed by Viberg (1995), Peternell et al. (1996) or Chui (1997) but, for convenience and simplicity, we assume here that $p = f = i$. Under these conditions, the output BHM would be given by:

$$\left(\frac{\boldsymbol{Z_p}}{\boldsymbol{Z_f}}\right) = \left(\frac{\boldsymbol{Z_{0:i-1}}}{\boldsymbol{Z_{i:2i-1}}}\right) = \left(\begin{array}{cccc} \boldsymbol{z_0} & \boldsymbol{z_1} & \cdots & \boldsymbol{z_{T-2i}} \\ \boldsymbol{z_1} & \boldsymbol{z_2} & \cdots & \boldsymbol{z_{T-2i+1}} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{z_{i-1}} & \boldsymbol{z_i} & \cdots & \boldsymbol{z_{T-i-1}} \\ \hline \boldsymbol{z_i} & \boldsymbol{z_{i+1}} & \cdots & \boldsymbol{z_{T-i}} \\ \boldsymbol{z_{i+1}} & \boldsymbol{z_{i+2}} & \cdots & \boldsymbol{z_{T-i+1}} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{z_{2i-1}} & \boldsymbol{z_{2i}} & \cdots & \boldsymbol{z_{T-1}} \end{array}\right) \tag{6}$$

Hereafter, we will consider for any BHM $\boldsymbol{A}$, that $\boldsymbol{A_p} = \boldsymbol{A_{0:i-1}}$ and $\boldsymbol{A_f} = \boldsymbol{A_{i:2i-1}}$, where the subscripts p and f denote the past and the future blocks respectively.

2) State sequences: This vector is defined as,

$$\boldsymbol{X_t} = (\boldsymbol{x_t} \quad \boldsymbol{x_{t+1}} \quad \boldsymbol{x_{t+2}} \quad \dots \quad \boldsymbol{x_{t+T-2i}}) \tag{7}$$

Taking this expression as starting point, the past and future state sequences beginning, respectively, at $t = 0$ and $t = i$, can be written as $\boldsymbol{X_p} = \boldsymbol{X_0}$ and $\boldsymbol{X_f} = \boldsymbol{X_i}$.

On the other hand, the following matrices are related to the parameters in model (1-2):

3) The Extended Observability Matrix:

$$\boldsymbol{O_i} = \begin{pmatrix} \boldsymbol{H} \\ \boldsymbol{H\Phi} \\ \boldsymbol{H\Phi^2} \\ \vdots \\ \boldsymbol{H\Phi^{i-1}} \end{pmatrix} \in \mathbb{R}^{im \times n} \tag{8}$$

4) The lower block triangular Toeplitz matrix:

$$\boldsymbol{V_i} = \begin{pmatrix} \boldsymbol{I_m} & \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{HE} & \boldsymbol{I_m} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{H\Phi E} & \boldsymbol{HE} & \boldsymbol{I_m} & \dots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{H\Phi^{i-2}E} & \boldsymbol{H\Phi^{i-3}E} & \boldsymbol{H\Phi^{i-4}E} & \dots & \boldsymbol{I_m} \end{pmatrix} \in \mathbb{R}^{im \times im} \tag{9}$$

The future state sequence can be defined as $\boldsymbol{X_f} = \boldsymbol{MZ_p}$, where $\boldsymbol{M}$ has rank n. Consequently, displacing time subscripts in (5) and substituting, we obtain,

$$\boldsymbol{Z_f} = \boldsymbol{O_i M Z_p} + \boldsymbol{V_i \Psi_f} \tag{10}$$

where $\boldsymbol{Z_f}$, $\boldsymbol{Z_p}$ and $\boldsymbol{\Psi_f}$ are as in (6), and $\boldsymbol{O_i}$ and $\boldsymbol{V_i}$ are respectively as in (8) and (9).

Subspace methods consists of estimating the matrices $\boldsymbol{O_i}, \boldsymbol{M}$ and $\boldsymbol{V_i}$ in (10) by solving a reduced-rank weighted least square problem defined over a set of subspace regressions, using the Singular Value Decomposition SVD (Eckart and Young, 1936) of the $\boldsymbol{W_1 Z_f W_2}$ matrix, where $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$ are two weighting matrices. When $\boldsymbol{W_1} = (\boldsymbol{Z_f Z_f'})^{-\frac{1}{2}}$ and $\boldsymbol{W_2} = \boldsymbol{Z_p'}(\boldsymbol{Z_p Z_p'})^{-1}\boldsymbol{Z_p}$, the Singular Values (SV) resulting from the SVD are estimates of the CCCs between $\boldsymbol{Z_f}$ and $\boldsymbol{Z_p}$. Then, the parameter matrices of (1-2) can be straightforwardly obtained from $\boldsymbol{O_i}, \boldsymbol{M}$ and $\boldsymbol{V_i}$.

3 DETECTION OF UNIT ROOTS

BW propose a fast procedure to detect unit roots which, instead of calculating the CCC between $\boldsymbol{z_t}$ and $\boldsymbol{z_{t-1}}$ (see Poskitt, 2000), uses the CCCs between the block-past information $\boldsymbol{Z_p}$ and the block-future information $\boldsymbol{Z_f}$. The authors compare a transformation of the j first SVs, $\hat{\sigma}_j$ (where $j \leq i$), with a penalty function $C(T)$ depending on the sample size:

$$1 - \hat{\sigma}_j^2 - C(T) < 0 \tag{11}$$

Any SV fulfilling this inequality can be considered as equal to one, corresponding to a unit root of the system. Note that the penalty function $C(T)$ represents a threshold such as when $1 - \hat{\sigma}_j^2$ is lesser than this level, it will be regarded equal to zero, and so the criterion will give, at least, j unit roots, since $\hat{\sigma}_1 \leq \hat{\sigma}_2 \leq ... \leq \hat{\sigma}_i \leq 0$. BW specify the following loss function:

$$C(T) = \frac{\log(T)^2}{T} \tag{12}$$

which collapses to zero as the sample size grows.

The use of $C(T)$ poses two main problems: 1) it only assures consistency when $d = 1$, being d the number of unit roots, and 2) it does not depend on the row space of the information blocks, i. The first issue is relevant, since this penalty function does not assure good results in processes with two or more non-stationary series. The second fact means that changes in i may produce important changes in the computed CCCs. Basically, larger subspace dimension implies higher CCCs, in the same way that the determination coefficient of a regression grows when the number of explanatory variables increase.

To solve both limitations, we will formulate a decision criterion similar to that of BW which, 1) it is based on Poskitt (2000) ideas and on a simulation study and 2) can be applied to systems with multiple unit roots. The procedure consists of obtaining the penalty function that distinguishes the $\hat{\sigma}_j$ considered equal to 1 and the others, using the superconsistency property. For this purpose, we impose two conditions on this function.

- *Condition 3.1:* It should depend on the sample size (T), the dimension of the information blocks (i) and the number of unit roots (d) that we wish to test.
- *Condition 3.2:* It should converge to zero when T tends to infinite, to assure the consistency of the estimated number of unit roots, $\hat{d}$. Formally, $\lim_{T\to\infty} G_l(T, i, d) = 0$, where G_l represents different loss functions.

A general representation of the criterion proposed is, therefore:

$$f(\hat{\sigma}_j) - G_l(T, i, d) \leq 0 \tag{13}$$

where $f(\hat{\sigma}_j)$ is a transformation measuring the distance between the j-th SV and the unity.

3.1 Univariate processes

To find a specification for $G_l(T, i, d)$, we analyze the empirical distribution of the $\hat{\sigma}_j$s corresponding to non-unit roots. To this end, we simulate 1000 replications of models near to non-stationarity for $T = 20, 21, \ldots, 500$. The transformation $f(\cdot)$ is applied to the first SV resultant of each simulation, summarizing all the information in a determined percentile (τ) and denoting this series by y_{1T}. Therefore, an empirical function measuring the distance to zero of $\hat{\sigma}_1$ is obtained. The question is how to find a penalty function fulfilling conditions (3.1) and (3.2) and correctly fitting y_{1T}.

For $d = 0$, the following loglinear model is considered:

$$\log(y_{1T}) = \alpha + \beta_1 \log(T) + \beta_2 \log(i) + \varepsilon_T \tag{14}$$

where ε_T corresponds to an error term and i refers to the row space of the past and future blocks.

Equation (14) generates a particular penalty function for every set of α, β_1 and β_2 values. These parameters depend on the value of the CCC which generates the simulations ($\bar{\sigma}$) and the percentile (τ), needed to define the y_{jT} variable. Table 1 shows estimated parameters of the model (14) corresponding to different $f(\cdot)$, τ and $\bar{\sigma}$.

[INSERT FIGURE 1 AND TABLE 1]

For comparison, Table 1 includes the penalty function proposed by BW. When model (14) is fitted to BW's function, its estimates are close to those obtained with our family of criteria. We denote our different loss functions by G_l, with $l = 1, 2, ..., 6$. But how do we decide which one to use? To choose the best penalty function it is reasonable to minimize the size of the decision criterion (defined as the probability of rejecting the null hypothesis when it is true) and maximize its power (the probability of rejecting the null hypothesis when it is false). Unfortunately, size and power move in opposite directions. Then we select the empirical distributions that fulfills, at least, one of the following two conditions:

- *Condition 3.3:* Minimize the size, with a power of the criterion which tends to the unity when the sample size increases.

- *Condition 3.4:* Minimize the size, requiring a minimum power for a specific sample size.

In the univariate case, in which the basic decision is if $d > 0$, we require that the minimum power in condition (3.4) is equal to 0.5 for $T = 50$. The probability of misestimating the number of unit roots in a random walk ($\nabla z_t = a_t$), as an approximation to the size of the criterion in univariate models, is used to choose the penalty functions and criteria. In the same way, we approximate the power as the probability of $\hat{d} = d$ in a stationary model with a persistent autoregressive process, $(1 - .9B)z_t = a_t$. The results for a 1000 replications simulation exercise are presented in Tables 2 and 3.

[INSERT TABLES 2-3]

Tables 2 and 3 show the performance in finite samples, both in size and power, of the criteria. Note that some of them beat the ADF-test in such properties. In view of the results, G_5 is chosen as the penalty function that fulfills the condition (3.3), hereafter called $G_a(d > 0)$. We will choose, the loss function G_2, from now on $G_b(d > 0)$, which fulfills condition (3.4).

3.2 Multivariate processes

The main advantage of the method lies in the treatment of multivariate models. First, only a few authors as Phillips and Durlauf (1986), Abuaf and Jorion (1990), Flôres et al. (1999) or BW study the identification of unit roots in vector series. Second, the ability to deal with vector of time series leads naturally to the cointegration analysis.

Estimated penalty functions not only depend on the sample size and the dimension of the information blocks, but also on the minimum number of unit roots we want to test. Therefore, each $\hat{\sigma}_j$ candidate to be equal to 1 should be compared with a different penalty function. So far, the criteria only indicate whether the series are stationary or not. In this Section, we devise criteria to sequentially evaluate if $d > 1$, $d > 2$, $d > 3$ and $d > 4$. To estimate the penalty functions, we simulate vectors of two, three, four and five series with unit and less than unit SVs. Each system contains at least one stationary, but close to unity, CCC. Then, different empirical distributions are estimated for every process, finally choosing a model that fits correctly, as we did in the univariate case.

Figure 2 compares the empirical distributions y_{1T} and y_{2T} corresponding to $(\tau = 90, \bar{\sigma} = 1.0)$ and $(\tau = 90, \bar{\sigma} = .95)$, for a bivariate system. Note that the first SV, corresponding to a unit root of the bivariate process, stays under $G_b(d > 0)$ which is the penalty function used in the criterion with biggest power to contrast $d > 0$. Indeed, if we compare a multivariate process $(\boldsymbol{A_t})$ having d unit roots, with the same system $(\boldsymbol{B_t})$ that incorporates a non-stationary new series (process $\boldsymbol{B_t}$ would have $d+1$ unit roots) then $\hat{\sigma}_j(\boldsymbol{B_t}) \geq \hat{\sigma}_j(\boldsymbol{A_t})$ for $j = 1, ..., d+1$. Hence one might use these criteria in a sequential way, to decide how many SVs are equal to 1.

[INSERT FIGURE 2 AND TABLE 4]

Table 4 presents the estimated parameters of model (14) corresponding to four alternative penalty functions to contrast $d > 1$. Figure 2 shows that the model specification, allowing to different parameter values, fits rather well the empirical distribution used to test both, $d > 0$ and $d > 1$. To discriminate between the alternative penalty functions, we will use conditions (3.3) and (3.4). Tables 5 and 6 report simulation results designed to assess size and power of the criteria formulated for $d > 1$. The exercise has been carried out using the loss function $G_b(d > 0)$ to decide if $d > 0$.

[INSERT TABLES 5-6]

We choose the penalty functions G_4 and G_1 from Table 4 and we respectively denote them by $G_a(d > 1)$ and $G_b(d > 1)$. The first minimizes the size given a weak power, while the second minimizes the size with a 0.45 power for $T = 50$. We reduce the minimum power required since the properties of the criteria, both in size and power, are lightly degraded in short samples when increasing d.

The method to obtain the penalty functions for trivariate, tetravariate and pentavariate processes is similar to that used so far. However, we re-specify the model to improve the fit to the empirical distribution by modeling two sub-samples:

$$\begin{array}{lll} T < T_j^* & y_{jT} & = \alpha_{j1} + \delta_{1j} T + \delta_{2j} T^2 + \delta_{3j} T^3 + \delta_{4j} i + \varepsilon_{1jT} \\ T \geq T_j^* & \log(y_{jT}) & = \alpha_{2j} + \beta_{1j} \log(T) + \beta_{2j} \log(i) + \varepsilon_{2jT} \end{array} \tag{15}$$

where T_j^* refers to the first observation of the second sub-sample.

Estimates of the two penalty functions fulfilling conditions (3.3) and (3.4) to test the existence of at least $j-1$ unit roots (with $j = 3, 4, 5$) and the parameter estimates of the models in (15) are depicted in Table 7. Again, we reduce the minimum power required for $T = 50$ making reference to condition (3.4): 0.4 for $d > 2, 3, 4$. Note, in Table 7, that increasing the row space of the information blocks

(i) positively affects $\hat{\sigma}_j$, and thus, negatively affects $f(\hat{\sigma}_j)$ for every j. Moreover, a bigger j implies more sensibility of y_{jT} to changes in i.

[INSERT TABLE 7]

Tables 8 and 9 describe the generating processes and simulations done to evaluate the size and the power of the criteria, using the penalty functions presented in Table 7. All the criteria show good finite sample properties when estimating d, although the results reveal a degradation when the number of unit roots increases.

[INSERT TABLES 8-9]

4 COINTEGRATION

In a multivariate context, when all the series are I(1), one may obtain the cointegrating rank (c), as $m - d$, where m is the system dimension and d is the number of unit roots. Therefore, we will devise a method to determine the cointegrating rank using the methodology of multivariate unit root detection which consists of:

1. Make sure the univariate series are I(1), applying the criterion formulated in Section 1.

2. Estimate the number of unit roots (d) of the multivariate process and,

3. obtain the cointegrating rank estimation $\hat{c}$, as the difference between the dimension system (m) and the estimated number of unit roots $(\hat{d})$.

We will now motivate why we use two penalty functions, from conditions (3.3) and (3.4), instead of only one. In the univariate analysis, it is coherent to use $C(G_a)$, which has greater size than $C(G_b)$, assuming that overdifferencing is better than underdifferencing (see Sanchez and Peña, 2001). In contrast, $C(G_b)$ will be more adequate for multivariate processes, where it can be better overestimate than underestimate the cointegrating rank. In these cases, one may apply the methodology to verify the I(0) integration order of the series computed as the product of the cointegration matrix and the original endogenous variables.

An interesting question consists of devising a procedure to estimate the cointegrating matrix. We suggest a method that provides consistent estimates, only requiring the parameter matrices of the innovation model and the computed cointegrating rank $(\hat{c})$. As we have seen, subspace methods can be directly applied to estimate the matrices in an innovations model (Favoreel et al., 2000; Knudsen, 2001). BW show that any model made up for I(1) series can be written as

an innovation model which divides the state sequences into stationary and non-stationary. Thus, the system (1-2) can be expressed, with the $\mathbf{\Phi}$ matrix in Jordan normal form, as:

$$\begin{pmatrix} \boldsymbol{X_{1,t+1}} \\ \boldsymbol{X_{2,t+1}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I_d} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\phi_{n-d}} \end{pmatrix} \begin{pmatrix} \boldsymbol{X_{1,t}} \\ \boldsymbol{X_{2,t}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{E_1} \\ \boldsymbol{E_2} \end{pmatrix} \boldsymbol{\psi_t} \tag{16}$$

$$\begin{pmatrix} \boldsymbol{Z_{1,t}} \\ \boldsymbol{Z_{2,t}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{H_{11}} & \boldsymbol{H_{12}} \\ \boldsymbol{H_{21}} & \boldsymbol{H_{22}} \end{pmatrix} \begin{pmatrix} \boldsymbol{X_{1,t}} \\ \boldsymbol{X_{2,t}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\psi_{1,t}} \\ \boldsymbol{\psi_{2,t}} \end{pmatrix} \tag{17}$$

where (16) is the state equation and (17) is the observation equation with $\boldsymbol{Z_{1,t}} \in \mathbb{R}^c, \boldsymbol{Z_{2,t}} \in \mathbb{R}^d, \boldsymbol{H_{11}} \in \mathbb{R}^{c\times d}, \boldsymbol{H_{12}} \in \mathbb{R}^{c\times(n-d)}, \boldsymbol{H_{21}} \in \mathbb{R}^{d\times d}$ and $\boldsymbol{H_{22}} \in \mathbb{R}^{d\times(n-d)}$. Also, n denotes the system order, $\boldsymbol{X_{1,t+1}}$ designates the state components of the d-dimensional non-stationary sub-system and $\boldsymbol{X_{2,t+1}}$ denotes the state components of the $(n-d)$-dimensional stationary sub-system.

To estimate the cointegrating matrix, one must find a matrix $\mathbf{\Lambda}$ that cancels $\boldsymbol{H_{11}}$, since it is the only block of $\boldsymbol{H}$ affecting the non-stationary state sequence for the cointegrated series.

Proposition 4.1 $\hat{\mathbf{\Lambda}} = (\boldsymbol{I_c} \quad -\hat{\boldsymbol{H}}_{\boldsymbol{11}}\hat{\boldsymbol{H}}_{\boldsymbol{21}}^{\boldsymbol{-1}})$ *is a consistent estimate of the cointegrating matrix.*

Proof. Premultiplying the equation (17) by $\mathbf{\Lambda}$ we obtain:

$$\mathbf{\Lambda}\begin{pmatrix} \boldsymbol{Z_{1,t}} \\ \boldsymbol{Z_{2,t}} \end{pmatrix} = \mathbf{\Lambda}\begin{pmatrix} \boldsymbol{H_{11}} & \boldsymbol{H_{12}} \\ \boldsymbol{H_{21}} & \boldsymbol{H_{22}} \end{pmatrix} \begin{pmatrix} \boldsymbol{X_{1,t}} \\ \boldsymbol{X_{2,t}} \end{pmatrix} + \mathbf{\Lambda}\begin{pmatrix} \boldsymbol{\psi_{1,t}} \\ \boldsymbol{\psi_{2,t}} \end{pmatrix} \tag{18}$$

taking $\boldsymbol{H_{11}}\boldsymbol{H_{21}^{-1}} = \boldsymbol{\lambda}$ and operating, we get:

$$\boldsymbol{Z_{1,t}} - \boldsymbol{\lambda Z_{2,t}} = (\boldsymbol{H_{12}} - \boldsymbol{\lambda H_{22}})\boldsymbol{X_{2,t}} + \boldsymbol{\psi_{1,t}} - \boldsymbol{\lambda \psi_{2,t}} \tag{19}$$

Equation (19) is a linear combination of the multivariate process $\boldsymbol{Z_t}$ that is not influenced by the non-stationary sequence $\boldsymbol{X_{1,t}}$, so it is stationary.

On the other hand, when the sample size grows (and consequently i grows), the order system estimation ($\hat{n}$) and the parameter matrices of equations (16-17) tend to their true values. Specifically, we get that:

$$\lim_{T\to\infty} \hat{\mathbf{\Lambda}} = \mathbf{\Lambda} \tag{20}$$

and so $\hat{\mathbf{\Lambda}}$ is a consistent estimate of $\mathbf{\Lambda}$.

□

5 SIMULATION EVIDENCE

To analyze the performance of the criteria proposed for cointegration analysis, we show the results obtained from several Montecarlo experiments. In the first simulation exercise, the data generating processes (DGP) considered coincide with those used by Riemers (1992), Toda (1995) and Poskitt (2000).

We start by considering a bivariate process. Table 10 shows the resulting relative frequency for different estimated cointegrating rank, $\hat{c}$. The DGP for the first system is:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} .2 & .8 \\ .4 & .6 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix}, \tag{21}$$

where

$$\begin{pmatrix} 1 & .0 \\ .0 & \nabla \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} .0 \\ \delta \end{pmatrix} + \begin{pmatrix} \kappa(B) & .0 \\ .0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix},$$

with $\kappa(B) = (1 - \alpha B)^{-1}$, and

$$\begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = N\left(\begin{pmatrix} .0 \\ .0 \end{pmatrix}, \boldsymbol{R_2} = \begin{pmatrix} 1.0 & \theta \\ \theta & 1.0 \end{pmatrix} \right)$$

Results have been obtained using different values for α, δ and θ. For each combination of parameters and sample size, we discard the 50 first $(y_{1t}, y_{2t})'$ observations to improve randomization.

[INSERT TABLE 10]

The values in Table 10 show the finite sample properties of the $C(G_a)$ and $C(G_b)$ criteria. Note that results are pretty good even for short samples. Furthermore, Table 10 reveals that results deteriorate in small samples when the process includes an autoregressive component close to the unit boundary. Obviously, the performance of the criterion using the G_a penalty function is worse in small samples than that which uses G_b since, by conditions (3.3) and (3.4), the former has more tendency to attribute the high autocorrelation to another stochastic trend, and consequently, to underestimate c.

Results shown in Table 11 are generated by a process similar to the previous one, but now the $(x_{1t}, x_{2t})'$ vector is defined as:

$$\begin{pmatrix} \nabla & .0 \\ .0 & \nabla \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} .0 \\ \delta \end{pmatrix} + \begin{pmatrix} \kappa(B) & .0 \\ .0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}, \tag{22}$$

where $\kappa(B) = 1 - \mu B$. In this case, $C(G_a)$ beats $C(G_b)$. Again, the outcomes show the consistency of both criteria. Moreover, they are rather good when μ remains relatively far from the unity. When $\mu = .9$ both criteria overestimate c because the unit root corresponding to the x_{1t} process is close to be cancelled out by the root of the moving average term, $(1 - .9B)\eta_{1t}$.

[INSERT TABLE 11]

To allow for more cointegrating relations, we also analyze the results obtained with the trivariate DGP:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} .0 & .0 & 1.0 \\ .0 & .2 & .8 \\ .1 & .3 & .6 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix}, \tag{23}$$

where

$$\begin{pmatrix} d_1(B) & .0 & .0 \\ .0 & d_2(B) & .0 \\ .0 & .0 & \nabla \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} .0 \\ .0 \\ \delta \end{pmatrix} + \begin{pmatrix} \kappa_1(B) & .0 & .0 \\ .0 & \kappa_2(B) & .0 \\ .0 & .0 & 1.0 \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{pmatrix}$$

with $(\eta_{1t}, \eta_{2t}, \eta_{3t})' = N(0, \boldsymbol{R_3})$ y, $\boldsymbol{R_3} = \begin{pmatrix} 1.0 & \theta & \theta \\ \theta & 1.0 & \theta \\ \theta & \theta & 1.0 \end{pmatrix}$

[INSERT TABLES 12-14]

Tables 12, 13 and 14 present the results obtained in this case. DGP (23) allows for different number of cointegrating relations. Thus, for the system 1: $c = 0$, $d_1(B) = d_2(B) = \nabla$ and $\kappa_1(B) = \kappa_2(B) = 1$; for the system 2: $c = 1$, $d_1(B) = 1$, $d_2(B) = \nabla$, $\kappa_1(B) = (1 - .\alpha B)^{-1}$ and $\kappa_2(B) = 1$; and last, for the system 3: $c = 2$, $d_1(B) = d_2(B) = 1$ and $\kappa_1(B) = \kappa_2(B) = (1 - .\alpha B)^{-1}$, so that the trivariant processes contain three, two and one unit roots, respectively. Table 12 show that both criteria present favorable results (better with $C(G_a)$ due to its penalty function) in middle-size samples, e.g. with 100 observations. When $\alpha = .9$, the danger to underestimate c using both criteria grows due to the high persistence of the autoregressives. In all the cases in those the data is generated with the trivariant process, $C(G_b)$ has recorded a 70% of corrected identification with $T = 300$.

To assess the influence of the system dimension on the performance of the criteria, we generate the following pentavariate process:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ y_{4t} \\ y_{5t} \end{pmatrix} = \begin{pmatrix} .0 & .0 & .0 & .0 & 1.0 \\ .0 & .0 & .2 & .0 & .8 \\ .0 & .0 & .3 & .1 & .6 \\ .0 & .1 & .4 & .1 & .4 \\ .1 & .1 & .3 & .3 & .2 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \\ x_{4t} \\ x_{5t} \end{pmatrix} \tag{24}$$

where the $\boldsymbol{X_t} = (x_{1t}, x_{2t}, ..., x_{5t})'$ vector is generated from $\boldsymbol{\Phi(B)X_t} = \boldsymbol{\Delta} + \boldsymbol{\Theta(B)\eta_t}$ with, $\boldsymbol{\Phi(B)} = diag(1, 1, 1, \nabla, \nabla)$ and $\boldsymbol{\Theta(B)} = diag((1 - \alpha B)^{-1}, (1 - \alpha B)^{-1}, (1-\alpha B)^{-1}, 1, 1)$ for the pentavariate system 1; $\boldsymbol{\Phi(B)} = diag(1, 1, \nabla, \nabla, \nabla)$ and $\boldsymbol{\Theta(B)} = diag((1-\alpha B)^{-1}, (1-\alpha B)^{-1}, (1-\mu B), 1, 1)$ for the pentavariate system 2. Also, in both processes $\boldsymbol{\Delta} = (\delta, \delta, ..., \delta)'$ and $\boldsymbol{\eta_t} = (\eta_{1t}, \eta_{2t}, ..., \eta_{5t})'$ with $\boldsymbol{\eta_t}$ $= N(0, \boldsymbol{R_5})$, where $\boldsymbol{R_5}$ denotes a diagonal matrix with ones on the main diagonal and θ elsewhere.

[INSERT TABLES 15-18]

Table 15 presents the results obtained with the pentavariate process 1. Although with $T = 100$ both criteria underestimate the cointegrating rank (in general, all the methods reveal serious difficulties to correctly estimate c) their behavior is quite good when $T = 300$. Therefore both criteria do not need a very large sample size for the cointegrating rank estimate to converge to its true value. Table 16 shows the results for a highly persistent autoregressive process. Due to the performance decline of all the techniques in these cases and following Poskitt (2000), the sample size is extended in a Fibonacci sequence to $T = 500, 800$ and 1300. When $T = 500$ and $T = 800$, $C(G_b)$ clearly dominates $C(G_a)$. Once again, when the sample grows both criteria lead to the true cointegrating rank.

The results obtained from the pentavariate process 2 are shown in Tables 17 and 18. Despite the fact that this system presents approximate cancellation of real autoregressive and moving average roots, both criteria estimate c reasonably well. When the autoregressive persistence and the sample size grow, Table 18 shows relatively good performance of both, $C(G_b)$ and $C(G_a)$. In fact, when $T = 500$, they show a 90% of corrected cointegrating rank detection, widely outperforming the other considered methods.

Finally, we extend the Monte Carlo experiment to the analysis of the cointegrating matrix estimator. Table 19 shows the probability of correctly estimating the cointegrating rank with $C(G_b)$ and, when this identification is right, the computed cointegrating vector. The process used presents a cointegration relationship

and two non-stationary common factors with cointegrating coefficients $\alpha_1 = -1.2$ and $\alpha_2 = .5$. The aim of the exercise is to: 1) analyze the performance of the cointegrating vector estimator and, 2) assess the influence of the contemporaneous correlation of the innovations.

[INSERT TABLE 19]

The results shown in Table 19 reveal the relevance of the value of the autoregressive ϕ. In fact, when $\phi = -.8$ the estimates of c, α_1 and α_2 are slightly less accurate. Besides, the contemporaneous correlation of the innovations affects on the cointegrating vector estimate, but it is not apparently clear in what way. The evidence suggests the ability of the algorithm to correctly estimate the cointegrating vector with processes which contain a cointegration relationship and two non-stationary common factors. The outcomes support the consistency, previously justified in theory, of the cointegrating rank and the cointegrating vector estimates.

6 EXAMPLE WITH REAL DATA

We will now analyze the integration order and cointegration relationships between several monthly U.S. interest rate series for August 1985-January 2003. Martín Manjón and Treadway (1997) highlight that short terms rates (of term two years or less) operate in bivariate cointegration relationships with the Federal Reserve's (Fed) Federal Funds Rate Target, while rates of longer term than two years do not. The example is divided into two parts. First, we analyze the Federal Funds Rate Target (RT_t) with the Effective Rate (ER_t) and the Treasury Bills at 3 ($TB3_t$) and 6 ($TB6_t$) month maturities (secondary market and auction averages). Secondly, we consider a trivariate system including the Federal Funds Rate Target and the T-bonds (Treasury Constant Maturity Rate) at terms 5 and 30 years, respectively $TB5_t$ and $TB30_t$.

6.1 Short term interest rates

The analysis of the series one-by-one concludes that all of them are I(1). When differencing the series, none of them generates a SV high enough for the criteria to detect a second unit root.

Table 20 summarizes the results of the integration analysis for the vector process. In this case, the criteria points out the existence of just one unit root. Therefore, three cointegration relations are found, each one corresponding to one of the three unit roots that disappeared when the series were jointly analyzed.

[INSERT TABLE 20]

As the series are I(1), we can write $\boldsymbol{Z_{1t}} = (ER_t \quad TB3_t \quad TB6_t \quad RT_t)'$ as in (16-17). This representation allows to compute an estimate of the cointegrating matrix $\boldsymbol{\Lambda}$:

$$\hat{\boldsymbol{\Lambda}} = \begin{pmatrix} 1 & 0 & 0 & -1.01 \\ 0 & 1 & 0 & -0.92 \\ 0 & 0 & 1 & -0.93 \end{pmatrix} \tag{25}$$

such that the series $\hat{\boldsymbol{\Lambda}}\boldsymbol{Z_{1t}}$ are I(0). Note that these series generated from the cointegrating matrix are roughly the spread of interest rates, characterizing a measure of the efficiency of the control of the short term interest rates.

6.2 Long term interest rates

As in Martín Manjón and Treadway (1997), we do not find any non-stationary common factor applying the detection algorithm to the process made up by $TB5_t$ and RT_t. On the contrary, we discover two unit roots. Indeed, the analysis of the CCCs reveals two coefficients greater than .965, considered enough close to 1 by both criteria. We investigate then whether there is a cointegration relationship between these two series and $TB30_t$, making up the new process $\boldsymbol{Z_{2t}} = (TB5_t \quad RT_t \quad TB30_t)'$.

The method returns only two unit roots when analyzing the process. Hence, it identifies that the T-bonds at term 5 years operates in trivariate CI(1,1) cointegration relationship with the Federal Funds Rate Target and the T-bonds at term 30 years getting the estimated cointegrating vector:

$$\hat{\boldsymbol{\Lambda}} = \begin{pmatrix} 1 & -.30 & -.68 \end{pmatrix} \tag{26}$$

such that the series $\hat{\boldsymbol{\Lambda}}\boldsymbol{Z_{2t}}$ are I(0).

7 CONCLUDING REMARKS

This paper provides a family of information criteria based on subspace methods and canonical correlation analysis, that allows to estimate the number of unit roots in both, univariate and multivariate processes. The procedure can also be used to obtain the estimation of the cointegrating rank. The simulation exercises indicate that this method has a remarkable capacity to estimate the number of unit roots and the cointegration rank in different situations. Contrary to other procedures as Johansen (1988, 1991), requiring a previous VAR specification, the methodology proposed does not need to fit a model to the data, with the resulting reduction of

error risk.

The fact that the penalty functions of the proposed criteria are fitted by simulation techniques suggests, in a natural way, the possibility to adapt the information criterion to samples with specific statistical properties. For example, it is known that high-frequency financial data often present distributions which have fatter tails than the Gaussian and, occasionally, a perceptible asymmetry. In this context, our methodology allows to design an optimized criterion for this kind of samples.

On the other hand, when the series are I(1) and the cointegrating rank is known, we provide a consistent estimator of the cointegrating matrix from a state space model in innovations form.

Finally, the procedures described in this article are implemented in a MATLAB toolbox for time series modeling called E4 that can be download at *www.ucm.es/info/icae/e4*. The source code for all the functions is freely provided under the terms of the GNU General Public License.

References

Abuaf, N. and Jorion, P. (1990). Purchasing power parity in the long run. *The Journal of Finance*, 45:157–174.

Akaike, H. (1975). Markovian representation of stochastic processes by canonical variables. *SIAM Journal of Control*, 13(1):162–173.

Akaike, H. (1976). *Canonical Correlation Analysis of Time Series and the Use of an Information Criterion*. Academic Press.

Bauer, D. and Wagner, M. (2002). Estimating cointegrated systems using subspace algorithms. *Journal of Econometrics*, 111:47–84.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, 2nd ed. edition.

Casals, J., Sotoca, S., and Jerez, M. (1999). A fast stable method to compute the likelihood of time invariant state space models. *Economics Letters*, 65(3):329–337.

Chui, C. K. and Chen, G. (1999). *Kalman Filtering with Real-Time Applications*. Berlin, Springer.

Chui, N. L. C. (1997). *Subspace Methods and Informative Experiments for System Identification*. PhD thesis, Pembroke College Cambridge.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with unit root. *Journal of the American Statistical Association*, 74:427–431.

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218.

Favoreel, W., De Moor, B., and Van Overschee, P. (2000). Subspace state space system identification for industrial processes. *Journal of Process Control*, 10:149–155.

Flôres, R. G., Jorion, P., Preumont, P. Y., and Szafarz, A. (1999). Multivariate unit roots test of ppp hypothesis. *Journal of Empirical Finance*, 6:335–353.

Ho, B. and Kalman, R. (1966). Effective construction of linear state-variable models from input-output functions. *Regelungstechnik*, 14:545–548.

Jenkins, G. M. and Alavi, A. S. (1981). Somes aspects of modelling and forecasting multivariate time series. *Journal of Time Series Analysis*, 2(1):1–47.

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

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

Knudsen, T. (2001). Consistency analysis of subspace identification methods based on linear regression approach. *Automatica*, 37:81–89.

Larimore, W. E. (1983). System identification, reduced-order filtering and modeling via canonical variate analysis. *Proceedings of the American Control Conference, ACC, San Francisco*, pages 445–451.

Larimore, W. E. (1990). Canonical variate analysis in identification, filtering and adaptive control. *Proceedings of the 29th Conference on Decision and Control, Hawaii*, pages 596–604.

Lütkepohl, H. and Poskitt, D. S. (1998). *Econometrics in Theory and Practice: Festchrift for Hans SchneeweiB*, chapter Consistent Estimation of the Number of Cointegration Relations in a Vector Autorregresive Model, pages 87–100. Springer Verlag.

Martín Manjón, R. and Treadway, A. (1997). The fed controls only one of the two interest rates in the u.s. economy. *Documento de Trabajo del ICAE*, (9716).

Peternell, K., Scherrer, W., and Deistler, M. (1996). Statistical analysis of novel subspace identification methods. *Signal Processing*, 52:161–177.

Phillips, P. B. C. and Durlauf, N. S. (1986). Multiple time series regression with integrated process. *Review of Economic Studies*, 53:473–495.

Poskitt, D. S. (2000). Strongly consistent determination of cointegrating rank via canonical correlations. *Journal of Business and Economic Statistics*, 18:77–90.

Riemers, H. E. (1992). Comparison of tests for multivariate cointegration. *Statistical Papers*, 33:335–359.

Sanchez, I. and Peña, D. (2001). Properties of predictors in overdifferenced nearly nonstationary autoregression. *Journal of Time Series Analysis*, 22(1):45–66.

Tiao, G. C. and Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, 76:802–816.

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

Toda, H. Y. (1995). Finite sample performance of likelihood ratio tests for cointegrating rank in vector autoregressions. *Econometric Theory*, 11:1015–1032.

Viberg, M. (1995). Subspace-based methods for the identification of the linear time-invariant systems. *Automatica*, 31(12):1835–1852.

Table 1: Estimated penalty functions to test $d > 0$

Empirical distribution			Estimates			Penalty
$f(\hat{\sigma}_1) = y_{1T}$	$\bar{\sigma}$	τ	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	Function
$1 - \hat{\sigma}_1^2$	0.975	95	1.09	-0.51	-0.17	$G_1(d > 0)$
$1 - \hat{\sigma}_1^2$	0.975	90	0.60	-0.50	-0.10	$G_2(d > 0)$
$1 - \hat{\sigma}_1^2$	0.95	80	0.10	-0.36	-0.10	$G_3(d > 0)$
$1 - \hat{\sigma}_1^2$	0.95	90	0.52	-0.40	-0.10	$G_4(d > 0)$
$1 - \hat{\sigma}_1$	0.95	90	0.10	-0.44	-0.05	$G_5(d > 0)$
$1 - \hat{\sigma}_1$	0.975	95	0.65	-0.56	-0.16	$G_6(d > 0)$
$\log(T)^2/T^\star$	-	-	0.98	-0.52	-0.10	$C(T)$

* Corresponding to the penalty function proposed by BW

Table 2: Size of the criteria to test $d > 0$ in the process $\phi(B)z_t = a_t$

T	ADF†	$C^\ddagger$	G_1	G_2	G_3	G_4	G_5	G_6
50	0.050	0.047	0.040	0.121	0.112	0.047	0.030	0.044
100	0.049	0.016	0.014	0.065	0.042	0.011	0.005	0.016
500	0.061	0.004	0.003	0.004	0.003	0.0	0.0	0.003

† Augmented Dickey-Fuller Test at 5% of significance level with one lag (Dickey and Fuller, 1979).

‡ Corresponding to the penalty function proposed by BW.

$\phi(B) = 1 - B$.

Table 3: Power of the criteria to test $d > 0$ in the process $\phi(B)z_t = a_t$

T	ADF†	C‡	G_1	G_2	G_3	G_4	G_5	G_6
50	0.114	0.246	0.208	0.505	0.464	0.246	0.188	0.230
100	0.341	0.451	0.420	0.762	0.628	0.376	0.290	0.451
500	1.0	1.0	1.0	1.0	1.0	0.994	0.981	1.0

† Augmented Dickey-Fuller Test at 5% of significance level with one lag (Dickey and Fuller, 1979).

‡ Corresponding to the penalty function proposed by BW.

$\phi(B) = 1 - .9B$.

Table 4: Estimated penalty functions to test $d > 1$

Empirical distribution			Estimates			Penalty
$f(\hat{\sigma}_2) = y_{2T}$	$\bar{\sigma}$	τ	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	Function
$1 - \hat{\sigma}_2^2$	0.95	75	0.43	-0.39	-0.07	G_1
$1 - \hat{\sigma}_2^2$	0.95	80	0.50	-0.39	-0.07	G_2
$1 - \hat{\sigma}_2^2$	0.95	85	0.57	-0.39	-0.07	G_3
$1 - \hat{\sigma}_2^2$	0.95	90	0.67	-0.39	-0.06	G_4

Table 5: Size of the criteria to test $d > 1$ in the process $\boldsymbol{\Phi(B)z_t = a_t}$

T	G_1	G_2	G_3	G_4
50	0.292	0.221	0.163	0.099
100	0.146	0.109	0.076	0.047
500	0.003	0.002	0.001	0.001

$\boldsymbol{\Phi(B)} = diag(\nabla, \nabla)$

Table 6: Power of the criteria to test $d > 1$ in the process $\boldsymbol{\Phi(B)z_t = a_t}$

T	G_1	G_2	G_3	G_4
50	0.452	0.373	0.293	0.204
100	0.516	0.428	0.337	0.239
500	0.993	0.989	0.968	0.947

$\boldsymbol{\Phi(B)} = diag(\nabla, (1 - .9B))$

Table 7: Estimated penalty functions to test $d > 2$, $d > 3$ and $d > 4$

Empirical distribution				$T < T^*$					$T \geq T^*$			Penalty
j	$f(\hat{\sigma}_j) = y_{jT}$	$\bar{\sigma}$	τ	$\hat{\alpha_1}$	$\hat{\delta_1}$	$\hat{\delta_2}$	$\hat{\delta_3}$	$\hat{\delta_4}$	$\hat{\alpha_0}$	$\hat{\beta_1}$	$\hat{\beta_2}$	Function
3	$1 - \hat{\sigma}_j^2$	0.95	90	-0.305	0.040	-6.5×10^{-4}	3.3×10^{-6}	-	0.786	-0.328	-0.226	$G_a(d > 2)$
3	$1 - \hat{\sigma}_j^2$	0.95	60	-0.353	0.036	-5.9×10^{-4}	3.0×10^{-6}	-	0.188	-0.285	-0.172	$G_b(d > 2)$
4	$1 - \hat{\sigma}_j^2$	0.95	90	-0.635	0.044	-6.0×10^{-4}	2.7×10^{-6}	-	1.589	-0.437	-0.365	$G_a(d > 3)$
4	$1 - \hat{\sigma}_j^2$	0.95	60	-0.621	0.041	-5.6×10^{-4}	2.5×10^{-6}	-	1.557	-0.469	-0.418	$G_b(d > 3)$
5	$1 - \hat{\sigma}_j^2$	0.95	90	-0.317	0.032	-3.2×10^{-4}	1.1×10^{-6}	-0.076	1.313	-0.383	-0.280	$G_a(d > 4)$
5	$1 - \hat{\sigma}_j^2$	0.95	60	-0.366	0.030	-3.0×10^{-4}	1.0×10^{-6}	-0.063	1.131	-0.361	-0.378	$G_b(d > 4)$

for $j = 3, 4$, $T^* = 88$;

for $j = 5$, $T^* = 121$.

Table 8: Size of the criteria to test $d > 2$, $d > 3$ and $d > 4$ in the processes $\boldsymbol{\Phi_k(B)z_t = a_t}$

	$d > 2$		$d > 3$		$d > 4$	
T	G_a	G_b	G_a	G_b	G_a	G_b
50	0.103	0.399	0.097	0.397	0.118	0.400
100	0.081	0.331	0.079	0.336	0.066	0.374
500	0.002	0.004	0.003	0.054	0.010	0.034

$\boldsymbol{\Phi_1(B)} = diag(\nabla, \nabla, \nabla)$; $\boldsymbol{\Phi_2(B)} = diag(\nabla, \nabla, \nabla, \nabla)$;
$\boldsymbol{\Phi_3(B)} = diag(\nabla, \nabla, \nabla, \nabla, \nabla)$

Table 9: Power of the criteria to test $d > 2$, $d > 3$ and $d > 4$ in the processes $\boldsymbol{\Phi_k(B)z_t = a_t}$

	$d > 2$		$d > 3$		$d > 4$	
T	G_a	G_b	G_a	G_b	G_a	G_b
50	0.166	0.402	0.161	0.398	0.110	0.408
100	0.854	0.487	0.741	0.460	0.144	0.438
500	0.940	0.951	0.920	0.994	0.458	0.844

$\boldsymbol{\Phi_1(B)} = diag(\nabla, \nabla, (1 - .9B))$; $\boldsymbol{\Phi_2(B)} = diag(\nabla, \nabla, \nabla, (1 - .9B))$;
$\boldsymbol{\Phi_3(B)} = diag(\nabla, \nabla, \nabla, \nabla, (1 - .9B))$

Table 10: Results for the bivariate process (21).⋆

T	$\hat{c}$	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{SBC}	PLR_T^{LP}
			$c = 1$, $\alpha = .7$, $\theta = .8$, $\delta = .0$				
100	0	0.0110	0.0075	0.0265	0.3250	0.0	0.1225
	1	0.9890	0.9925	0.9732	0.6280	1.0	0.8775
200	0	0.0	0.0	0.0005	0.0010	0.0	0.0050
	1	1.0	1.0	0.9995	0.9540	1.0	0.9950
300	0	0.0	0.0	0.0	0.0000	0.0	0.0
	1	1.0	1.0	1.0	0.9400	1.0	1.0
			$c = 1$, $\alpha = .8$, $\theta = .8$, $\delta = .0$				
100	0	0.2320	0.0500	0.2275	0.6250	0.0020	0.4175
	1	0.7680	0.9500	0.7725	0.3360	0.9980	0.5825
200	0	0.0060	0.0015	0.0260	0.0720	0.0	0.1855
	1	0.9940	0.9985	0.9740	0.8840	1.0	0.8145
300	0	0.0	0.0005	0.0010	0.0010	0.0	0.0465
	1	1.0	0.9995	0.9990	0.9390	1.0	0.9535
			$c = 1$, $\alpha = .9$, $\theta = .8$, $\delta = .0$				
100	0	0.7765	0.5050	0.6920	0.8150	0.0370	0.7395
	1	0.2235	0.4950	0.3080	0.1335	0.9630	0.2605
200	0	0.5645	0.2410	0.6250	0.6035	0.0020	0.8300
	1	0.4355	0.7590	0.3745	0.3630	0.9980	0.1700
300	0	0.3555	0.0800	0.4850	0.2390	0.0	0.7975
	1	0.6445	0.9195	0.5150	0.7070	1.0	0.2025

⋆The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by $\hat{c}$. $C(G_a)$ and $C(G_b)$ represent the criteria (13) with the proposed penalty functions G_a and G_b, respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion (∇_T), Johansen's (1988, 1991) likelihood ratio procedure, (LR_T), and some penalized likelihood criteria (PLR_T). These lasts require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwartz, 1978) or LP (Lütkepohl and Poskitt, 1998) superscripts. These values are taken from Poskitt (2000) where they were obtained with the same number of replications (2000), sample sizes ($T = 100, 200, 300$) and DGP. See Poskitt (2000) for more details.

Table 11: Results for the bivariate process (22).⋆

T	$\hat{c}$	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{SBC}	PLR_T^{LP}
			$c = 0$, $\mu = .3$, $\theta = .8$, $\delta = .0$				
100	0	0.7880	0.6240	0.6620	0.9510	0.1510	0.8830
	1	0.2120	0.3760	0.3380	0.0465	0.8490	0.1170
200	0	0.8815	0.7765	0.8435	0.9445	0.2140	0.9845
	1	0.1185	0.2235	0.1565	0.0520	0.7860	0.0155
300	0	0.9415	0.8690	0.9260	0.9455	0.2565	0.9980
	1	0.0585	0.1310	0.0740	0.0485	0.7435	0.0020
			$c = 0$, $\mu = .5$, $\theta = .8$, $\delta = .0$				
100	0	0.5795	0.4200	0.3185	0.9545	0.1480	0.8890
	1	0.4205	0.5800	0.6815	0.0405	0.8520	0.1110
200	0	0.7005	0.5545	0.4760	0.9465	0.2095	0.9845
	1	0.2995	0.4455	0.5240	0.0510	0.7905	0.0155
300	0	0.7870	0.6750	0.5925	0.9440	0.2600	0.9980
	1	0.2130	0.3250	0.4075	0.0505	0.7400	0.0020
			$c = 0$, $\mu = .9$, $\theta = .8$, $\delta = .0$				
100	0	0.0230	0.0095	0.0	0.7015	0.0170	0.4840
	1	0.9770	0.9875	1.0	0.2780	0.9830	0.5160
200	0	0.0285	0.0130	0.0	0.4075	0.0095	0.5540
	1	0.9715	0.9870	1.0	0.5675	0.9905	0.4460
300	0	0.0535	0.0315	0.0	0.4305	0.0255	0.7650
	1	0.9465	0.9680	1.0	0.5415	0.9745	0.2350

⋆The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by $\hat{c}$. $C(G_a)$ and $C(G_b)$ represent the criteria (13) with the proposed penalty functions G_a and G_b, respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion (∇_T), Johansen's (1988, 1991) likelihood ratio procedure, (LR_T), and some penalized likelihood criteria (PLR_T). These lasts require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwartz, 1978) or LP (Lütkepohl and Poskitt, 1998) superscripts. These values are taken from Poskitt (2000) where they were obtained with the same number of replications (2000), sample sizes ($T = 100, 200, 300$) and DGP. See Poskitt (2000) for more details.

Table 12: Results for the trivariate process (23), system 1.⋆

T	$\hat{c}$	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{SBC}	PLR_T^{LP}
			$c = 0$, $\theta = .0$, $\delta = .0$				
100	0	0.9255	0.6470	0.7560	0.9395	0.4085	0.4230
	1	0.0740	0.3480	0.2385	0.0570	0.0890	0.5585
	2	0.0005	0.0050	0.0055	0.0030	0.5026	0.0185
200	0	0.9500	0.7735	0.9385	0.9410	0.6125	0.8555
	1	0.0500	0.2260	0.0615	0.0530	0.0740	0.1455
	2	0.0	0.0005	0.0	0.0060	0.3135	0.0
300	0	0.9595	0.8270	0.9860	0.9505	0.7300	0.9690
	1	0.0405	0.1725	0.0140	0.0465	0.0695	0.0310
	2	0.0	0.0005	0.0	0.0030	0.2005	0.0

⋆The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by $\hat{c}$. $C(G_a)$ and $C(G_b)$ represent the criteria (13) with the proposed penalty functions G_a and G_b, respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion (∇_T), Johansen's (1988, 1991) likelihood ratio procedure, (LR_T), and some penalized likelihood criteria (PLR_T). These lasts require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwartz, 1978) or LP (Lütkepohl and Poskitt, 1998) superscripts. These values are taken from Poskitt (2000) where they were obtained with the same number of replications (2000), sample sizes ($T = 100, 200, 300$) and DGP. See Poskitt (2000) for more details.

Table 13: Results for the trivariate process (23), system 2.⋆

T	$\hat{c}$	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{SBC}	PLR_T^{LP}
			$c = 1$, $\alpha = .8$, $\theta = .0$, $\delta = .0$				
100	0	0.4620	0.0615	0.2015	0.7760	0.1280	0.1515
	1	0.5290	0.8750	0.7485	0.1985	0.0990	0.7850
	2	0.0090	0.0630	0.0500	0.0215	0.7730	0.0635
200	0	0.0100	0.0	0.0340	0.3090	0.0140	0.0985
	1	0.9835	0.9645	0.9530	0.6455	0.2155	0.8935
	2	0.0065	0.0355	0.0130	0.0410	0.7705	0.0080
300	0	0.0	0.0	0.0030	0.0355	0.0	0.0200
	1	0.9975	0.9850	0.9920	0.9115	0.2785	0.9780
	2	0.0025	0.0150	0.0050	0.0500	0.7215	0.0020
			$c = 1$, $\alpha = .9$, $\theta = .0$, $\delta = .0$				
100	0	0.8510	0.4180	0.5880	0.8925	0.3070	0.3310
	1	0.1470	0.5540	0.3970	0.0990	0.0740	0.6445
	2	0.0020	0.0280	0.0150	0.0070	0.6190	0.0245
200	0	0.5370	0.1450	0.6275	0.7810	0.2520	0.6090
	1	0.4605	0.8360	0.3705	0.1950	0.1200	0.3895
	2	0.0025	0.0190	0.0020	0.0210	0.6280	0.0015
300	0	0.2190	0.0175	0.5380	0.5525	0.1190	0.6300
	1	0.7810	0.9735	0.4610	0.4110	0.2245	0.3695
	2	0.0	0.0090	0.0010	0.0340	0.6565	0.0005

⋆The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by $\hat{c}$. $C(G_a)$ and $C(G_b)$ represent the criteria (13) with the proposed penalty functions G_a and G_b, respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion (∇_T), Johansen's (1988, 1991) likelihood ratio procedure, (LR_T), and some penalized likelihood criteria (PLR_T). These lasts require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwartz, 1978) or LP (Lütkepohl and Poskitt, 1998) superscripts. These values are taken from Poskitt (2000) where they were obtained with the same number of replications (2000), sample sizes ($T = 100, 200, 300$) and DGP. See Poskitt (2000) for more details.

Table 14: Results for the trivariate process (23), system 3.⋆

T	$\hat{c}$	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{SBC}	PLR_T^{LP}
			$c = 2$, $\alpha = .8$, $\theta = .0$, $\delta = .0$				
100	0	0.2190	0.0070	0.1690	0.4520	0.0195	0.0485
	1	0.4855	0.2680	0.3485	0.4060	0.0020	0.6515
	2	0.2955	0.7250	0.4825	0.1175	0.9785	0.3000
200	0	0.0	0.0	0.0340	0.0035	0.0	0.0110
	1	0.0395	0.0100	0.0710	0.2085	0.0	0.4535
	2	0.9605	0.9900	0.8950	0.7345	1.0	0.5355
300	0	0.0	0.0	0.0020	0.0	0.0	0.0000
	1	0.0	0.0005	0.0080	0.0020	0.0	0.1510
	2	1.0	0.9995	0.9900	0.9480	1.0	0.8490
			$c = 2$, $\alpha = .9$, $\theta = .0$, $\delta = .0$				
100	0	0.7500	0.2415	0.5255	0.8245	0.2100	0.2660
	1	0.2370	0.6545	0.4175	0.1525	0.0340	0.6795
	2	0.0130	0.1020	0.0570	0.0185	0.7560	0.0545
200	0	0.2495	0.0200	0.6405	0.4205	0.0525	0.4355
	1	0.6735	0.5770	0.2990	0.4175	0.0040	0.5465
	2	0.0770	0.3955	0.0650	0.1290	0.9435	0.1800
300	0	0.0415	0.0010	0.6310	0.0535	0.0015	0.4000
	1	0.7125	0.2830	0.2550	0.4565	0.0005	0.5740
	2	0.2460	0.7075	0.1140	0.4445	0.9980	0.0260

⋆The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by $\hat{c}$. $C(G_a)$ and $C(G_b)$ represent the criteria (13) with the proposed penalty functions G_a and G_b, respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion (∇_T), Johansen's (1988, 1991) likelihood ratio procedure, (LR_T), and some penalized likelihood criteria (PLR_T). These lasts require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwartz, 1978) or LP (Lütkepohl and Poskitt, 1998) superscripts. These values are taken from Poskitt (2000) where they were obtained with the same number of replications (2000), sample sizes ($T = 100, 200, 300$) and DGP. See Poskitt (2000) for more details.

Table 15: Results for the pentavariate process (24), system 1.⋆

T	$\hat{c}$	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{LP}
		$c = 3$, $\alpha = .8$, $\theta = .8$, $\delta = 1.0$				
100	0	0.0875	0.1040	0.0370	0.0850	0.0
	1	0.7750	0.4665	0.2685	0.5695	0.0830
	2	0.1370	0.3485	0.4090	0.2935	0.6705
	3	0.0005	0.0800	0.2835	0.0450	0.2435
	4	0.0	0.0005	0.0020	0.0060	0.0030
200	0	0.0005	0.0	0.0110	0.0	0.0
	1	0.0115	0.0	0.0585	0.0555	0.1035
	2	0.7060	0.1130	0.1365	0.5405	0.7075
	3	0.2820	0.8830	0.7925	0.4060	0.1890
	4	0.0	0.0040	0.0015	0.0320	0.0
300	0	0.0	0.0	0.0005	0.0	0.0
	1	0.0	0.0	0.0025	0.0	0.0925
	2	0.1390	0.0010	0.0095	0.0915	0.7310
	3	0.8610	0.9970	0.9875	0.8555	0.1765
	4	0.0	0.0020	0.0	0.0490	0.0

⋆The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by $\hat{c}$. $C(G_a)$ and $C(G_b)$ represent the criteria (13) with the proposed penalty functions G_a and G_b, respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion (∇_T), Johansen's (1988, 1991) likelihood ratio procedure, (LR_T), and some penalized likelihood criteria (PLR_T). These lasts require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwartz, 1978) or LP (Lütkepohl and Poskitt, 1998) superscripts. These values are taken from Poskitt (2000) where they were obtained with the same number of replications (2000), sample sizes ($T = 100, 200, 300$) and DGP. See Poskitt (2000) for more details.

Table 16: Results for the pentavariate process (24), system 1.⋆

T	$\hat{c}$	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{LP}
		$c = 3$, $\alpha = .9$, $\theta = .8$, $\delta = 1.0$				
500	0	0.0	0.0	0.4685	0.0	0.0
	1	0.0145	0.0	0.2880	0.0015	0.4715
	2	0.9530	0.5440	0.1180	0.2715	0.5120
	3	0.0325	0.4555	0.1255	0.6800	0.0165
	4	0.0	0.0005	0.0	0.0490	0.0
800	0	0.0	0.0	0.1625	0.0	0.0
	1	0.0	0.0	0.0825	0.0	0.3540
	2	0.5865	0.0650	0.0510	0.0005	0.6085
	3	0.4135	0.9350	0.7040	0.9535	0.0375
	4	0.0	0.0	0.0	0.0415	0.0
1300	0	0.0	0.0	0.0005	0.0	0.0
	1	0.0	0.0	0.0005	0.0	0.0645
	2	0.0290	0.0	0.0005	0.0	0.6665
	3	0.9710	1.0	0.9985	0.9475	0.2690
	4	0.0	0.0	0.0	0.0490	0.0

⋆The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by $\hat{c}$. $C(G_a)$ and $C(G_b)$ represent the criteria (13) with the proposed penalty functions G_a and G_b, respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion (∇_T), Johansen's (1988, 1991) likelihood ratio procedure, (LR_T), and some penalized likelihood criteria (PLR_T). These lasts require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwartz, 1978) or LP (Lütkepohl and Poskitt, 1998) superscripts. These values are taken from Poskitt (2000) where they were obtained with the same number of replications (2000), sample sizes ($T = 500, 800, 1300$) and DGP. See Poskitt (2000) for more details.

Table 17: Results for the pentavariate process (24), system 2.⋆

T	$\hat{c}$	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{LP}
$c = 2$, $\alpha = .8$, $\mu = .9$, $\theta = .8$, $\delta = 1.0$						
100	0	0.0120	0.0160	0.0	0.0005	0.0005
	1	0.7735	0.5045	0.2990	0.1920	0.0695
	2	0.2145	0.4715	0.7335	0.6250	0.6550
	3	0.0	0.0080	0.0375	0.1625	0.2730
	4	0.0	0.0	0.0	0.0175	0.0020
200	0	0.0	0.0	0.0	0.0	0.0
	1	0.0635	0.0035	0.0750	0.0	0.0845
	2	0.9365	0.8805	0.9025	0.3190	0.6865
	3	0.0	0.1160	0.0225	0.6390	0.2280
	4	0.0	0.0	0.0	0.0395	0.0010
300	0	0.0	0.0	0.0	0.0	0.0050
	1	0.0170	0.0090	0.0085	0.0	0.1425
	2	0.9795	0.9290	0.9145	0.0315	0.7155
	3	0.0035	0.0615	0.0770	0.9250	0.1415
	4	0.0	0.0005	0.0	0.0395	0.0

⋆The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by $\hat{c}$. $C(G_a)$ and $C(G_b)$ represent the criteria (13) with the proposed penalty functions G_a and G_b, respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion (∇_T), Johansen's (1988, 1991) likelihood ratio procedure, (LR_T), and some penalized likelihood criteria (PLR_T). These lasts require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwartz, 1978) or LP (Lütkepohl and Poskitt, 1998) superscripts. These values are taken from Poskitt (2000) where they were obtained with the same number of replications (2000), sample sizes ($T = 100, 200, 300$) and DGP. See Poskitt (2000) for more details.

Table 18: Results for the pentavariate process (24), system 2.⋆

T	$\hat{c}$	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{LP}
$c = 2$, $\alpha = .9$, $\mu = .9$, $\theta = .8$, $\delta = 1.0$						
500	0	0.0060	0.0	0.0	0.0	0.0
	1	0.0985	0.0275	0.4225	0.0	0.4025
	2	0.8955	0.9725	0.5775	0.0965	0.5615
	3	0.0	0.0	0.0	0.8525	0.0360
	4	0.0	0.0	0.0	0.0480	0.0
800	0	0.0	0.0	0.0	0.0	0.0
	1	0.0590	0.0255	0.1580	0.0	0.5130
	2	0.9410	0.9740	0.8390	0.0	0.4685
	3	0.0	0.0005	0.0030	0.9605	0.0185
	4	0.0	0.0	0.0	0.0370	0.0
1300	0	0.0	0.0	0.0	0.0	0.0
	1	0.0595	0.0055	0.0055	0.0	0.2710
	2	0.9405	0.9945	0.9210	0.0	0.7080
	3	0.0	0.0	0.0735	0.9440	0.0210
	4	0.0	0.0	0.0	0.0515	0.0

⋆The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by $\hat{c}$. $C(G_a)$ and $C(G_b)$ represent the criteria (13) with the proposed penalty functions G_a and G_b, respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion (∇_T), Johansen's (1988, 1991) likelihood ratio procedure, (LR_T), and some penalized likelihood criteria (PLR_T). These lasts require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwartz, 1978) or LP (Lütkepohl and Poskitt, 1998) superscripts. These values are taken from Poskitt (2000) where they were obtained with the same number of replications (2000), sample sizes ($T = 500, 800, 1300$) and DGP. See Poskitt (2000) for more details.

Table 19: Estimated cointegrating rank and vector for $\alpha_1 = -1.2$ and $\alpha_2 = 0.5$.*

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1+\phi B \end{pmatrix} \begin{pmatrix} \nabla & 0 & 0 \\ 0 & \nabla & 0 \\ \alpha_1 & \alpha_2 & 1 \end{pmatrix} \begin{pmatrix} z_{1t} \\ z_{2t} \\ z_{3t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}; \quad \Sigma_\varepsilon = \begin{pmatrix} 1 & \theta & \theta \\ \theta & 1 & \theta \\ \theta & \theta & 1 \end{pmatrix}$$

θ	T	$\phi = -0.8$			$\phi = 0$			$\phi = 0.8$		
		$\hat{c}=1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{c}=1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{c}=1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
	50	0.714	-1.213	0.433	0.696	-1.184	0.487	0.653	-1.197	0.505
0.2	100	0.846	-1.198	0.509	0.843	-1.203	0.497	0.807	-1.202	0.504
	300	0.984	-1.198	0.505	0.974	-1.198	0.498	0.973	-1.201	0.502
	50	0.705	-1.217	0.429	0.752	-1.193	0.520	0.695	-1.163	0.521
0.4	100	0.875	-1.163	0.544	0.874	-1.199	0.504	0.866	-1.184	0.492
	300	0.986	-1.185	0.519	0.992	-1.198	0.504	0.992	-1.198	0.502
	50	0.736	-1.084	0.446	0.817	-1.185	0.533	0.731	-1.183	0.535
0.6	100	0.876	-1.109	0.535	0.913	-1.193	0.518	0.915	-1.194	0.508
	300	0.992	-1.172	0.523	0.994	-1.196	0.508	0.994	-1.197	0.502
	50	0.725	-1.131	0.431	0.848	-1.173	0.526	0.788	-1.146	0.525
0.8	100	0.879	-1.085	0.525	0.933	-1.174	0.488	0.918	-1.183	0.501
	300	0.992	-1.172	0.523	0.991	-1.194	0.505	0.995	-1.196	0.502

* θ is the contemporaneous correlation of the innovations, $\hat{c}=1$ indicates the relative frequency of occurrence of this event. $\hat{\alpha_1}$ and $\hat{\alpha_2}$ are the average of the estimated cointegrating parameters when c is correctly estimated. This table is based on 1000 replications.

Table 20: Integration analysis of the $\boldsymbol{Z_{1t}}$ process

j	$\hat{\sigma}_j$	$f(\hat{\sigma}_j) - G_a(d > j-1)$	$f(\hat{\sigma}_j) - G_b(d > j-1)$	$\hat{d}_{G_a}$	$\hat{d}_{G_b}$
1	0.9952	$0.0048 - 0.0970 < 0$	$0.0095 - 0.1070 < 0$	1	1
2	0.8797	$0.2261 - 0.2205 > 0$	$0.2261 - 0.1797 > 0$	1	1
3	0.7324	$0.4636 - 0.2648 > 0$	$0.4636 - 0.1991 > 0$	1	1
4	0.4028	$0.8377 - 0.2642 > 0$	$0.8377 - 0.1968 > 0$	1	1

$\hat{\sigma}_j$ corresponds to the j singular value. $\hat{d}_{G_a}$ and $\hat{d}_{G_b}$ are the unit roots number found in the j firsts singular values with the provided penalty functions G_a and G_b.

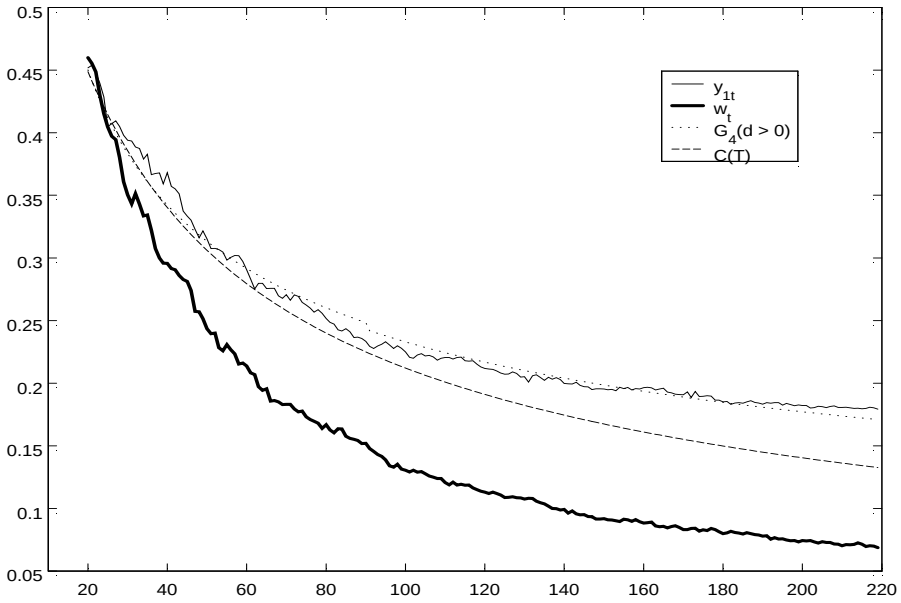

Figure 1: Simulated Singular Values (SV) and penalty functions for univariate processes. y_{1T} corresponds to a stationary process' first SV and w_T corresponds to a non-stationary process' first SV. $G_4(d > 0)$ is a loss function provided for univariate processes and $C(T)$ is the penalty function proposed by Bauer and Wagner (2002). y_{1T} and w_T show the difference between the convergence rates of the SV to their asymptotic values.

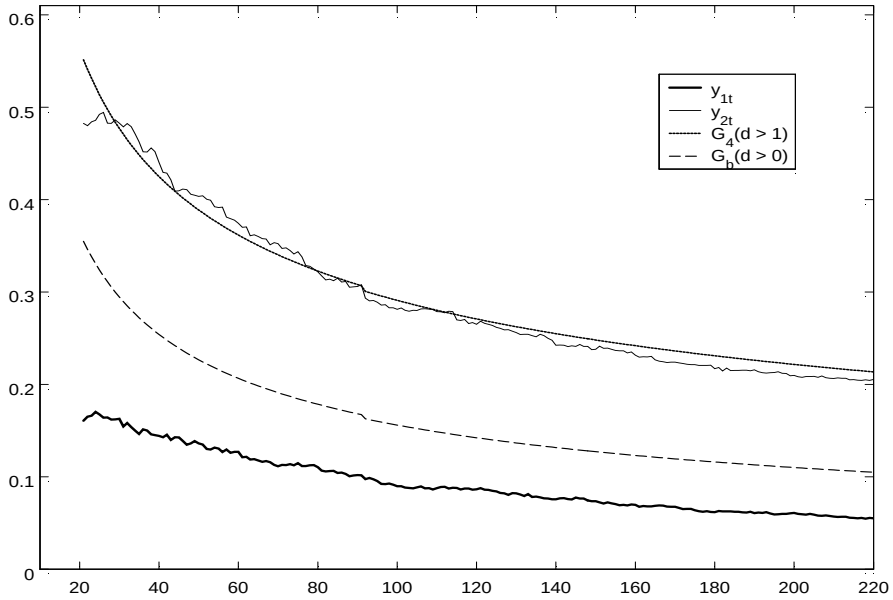

Figure 2: Simulated Singular Values (SV) and penalty functions for bivariate processes. y_{1T} corresponds to the first SV of a bivariate non-stationary system which univariate processes are I(1) and I(0). y_{2T} corresponds to the second SV of the same system, $G_4(d > 1)$ is a penalty function provided to detect (at least) two unit roots and $G_b(d > 0)$ is the penalty function used to detect (at least) one unit root.