

# Unit roots and cointegration modeling through a family of flexible information criteria

Alfredo García-Hiernaux\*    Miguel Jerez    José Casals

September 17, 2008

## Abstract

We propose a fast and consistent procedure to detect unit roots based on subspace methods. It has three distinctive features. First, the same method can be applied to single or multiple time series. Second, it employs a flexible family of information criteria, whose loss functions can be adapted to the statistical properties of the data. Last, it does not require the specification of a model for the analyzed series. In addition, we provide a subspace-based consistent estimator for the cointegrating rank and the cointegrating matrix. Simulation exercises show that these procedures have good finite sample properties.

**Keywords:** State-space models, subspace methods, unit roots, cointegration

**MSC:** 91B84-62M10

---

\*Statistics Department, Universidad Carlos III de Madrid. E-mail: aghierna@econ.uc3m.es, tel: (+34) 91 624 58 52, fax: 91 624 98 49. The authors gratefully acknowledge financial support from Ministerio de Educación y Ciencia, ref. ECO2008-02588/ECON.

# 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 engineering, whose 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. Since the 90s, system identification has been led by the so-called subspace methods.

In comparison with mainstream time series analysis (Box and Jenkins, 1976; Tiao and Tsay, 1989) the main advantages of the subspace methods are: a) they allow one to specify a general linear model directly from the data, without a priori knowledge of the process structure, b) the univariate and multivariate cases are treated in the same way, c) they are based on robust and computationally efficient algebraic tools and, as a consequence, d) iterations are not required, avoiding convergence problems.

On the other hand, applying subspace methods to the data generated by non-physical systems presents important challenges as they cannot currently accommodate trends or seasonality, being these features very common in other contexts.

This paper deals with the detection of unit roots and the cointegrating relations. The time series literature provides many tools to tackle these features which, however, have some shortcomings such as: a) they are consistent but often perform poorly in finite samples, b) they usually assume the Gaussian distribution, c) methods as those provided by Dickey and Fuller (1981) or Johansen (1988, 1991) require the choice of lag length (see, Emerson, 2007), and d) for most cointegration tests, they often do not provide the empirical support that might have been expected for seemingly innocuous long-run relationships, as it has been shown by Haug (1996) or Bewley and Yang (1998).

We address these questions building on the Canonical Correlation Analysis algorithm by Larimore (1983). This is a convenient starting point as its choice of weighting matrices yields, as a by-product, an efficient computation of the canonical correlation coefficients (CCCs) between the past and future information subspaces (see Katayama, 2005). Thus, our method is based on the CCCs as Poskitt's (2000) but ours takes into account the short-run dynamics. However, this is done in a different way than in Johansen (1991) or Bewley and Yang (1995).

The basic idea is that the CCCs corresponding to unit roots converge to their true values faster than the rest. This property, known as *superconsistency*, allows distinguishing both kinds of correlations. Bauer and Wagner (2002), hereafter BW, adapted this approach to the subspace methods providing two tests to estimate the cointegrating rank based on the unit roots detection of multivariate processes. The first procedure, called *eigenvalue test sequence*, depends critically on the correct estimation of the system order in a previous step. The second one, the *combined test sequence*, shows a better behavior but strongly depends on an initial null hypothesis.

The contribution of this paper is threefold. We provide: a) an information criterion to detect the number of unit roots in univariate and multivariate processes, b) a simulation-based methodology to fit its penalty functions to the characteristics of the data and c) a consistent estimator for the cointegrating rank and the cointegrating matrix. Our proposal: a) leads to consistent estimates of the number of unit roots, b) exhibits a satisfactory performance in small samples, c) can be used as a good initial null hypothesis for the *combined test sequence* or to obtain a first and good insight into the number of possible trends in univariate and multivariate cases, and d) can be adapted for non-Gaussian distributions.

The structure of the paper is as follows. Section 2 introduces the subspace methods and the notation. Section 3 presents the procedure to detect unit roots. Section 4 extends these results to consistently estimate the order of cointegration

and presents a subspace-based method to estimate the cointegrating matrix. Section 5 analyzes the methods proposed through a set of simulation exercises and, finally, Section 6 summarizes the main conclusions.

## 2 SUBSPACE METHODS AND CANONICAL CORRELATIONS

Consider a linear fixed-coefficients system that is described by the SS model:

$$\mathbf{x}_{t+1} = \mathbf{\Phi}\mathbf{x}_t + \mathbf{E}\boldsymbol{\psi}_t \quad (1)$$

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \boldsymbol{\psi}_t \quad (2)$$

where  $\mathbf{x}_t \in \mathbb{R}^n$  is the state vector,  $\mathbf{z}_t \in \mathbb{R}^m$  is an observable output vector and  $\boldsymbol{\psi}_t \in \mathbb{R}^m$  is an unobservable noise.  $\mathbf{\Phi} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{E} \in \mathbb{R}^{n \times m}$  and  $\mathbf{H} \in \mathbb{R}^{m \times n}$  are the parametric matrices of this system, known as “innovations form”. No generality is lost when choosing this particular representation because any fixed-coefficients SS model can be written in innovations form under weak assumptions (see, Casals et al., 1999, Theorem 1).

**Assumptions:** We assume the system to be minimal and strictly minimum phase. Minimality means that there is no a SS model that can realize  $\mathbf{z}_t$  with less than  $n$  states, while the minimum-phase condition requires all the eigenvalues of  $(\mathbf{\Phi} - \mathbf{E}\mathbf{H})$  to lie inside the unit circle. We also assume that the eigenvalues of  $\mathbf{\Phi}$  lie on or inside the unit circle. Note that stationary and integrated systems with different cointegrating ranks fulfill these assumptions. Finally,  $\boldsymbol{\psi}_t$  is a vector of white noise processes such that  $E(\boldsymbol{\psi}_t) = \mathbf{0}$  and  $E(\boldsymbol{\psi}_t\boldsymbol{\psi}_t') = \mathbf{Q}$ .

Subspace methods derive from the representation of (1-2) in a specific form.

By substituting (2) into (1) in  $\boldsymbol{\psi}_t$  and solving by recursion we have:

$$\boldsymbol{x}_p = (\boldsymbol{\Phi} - \boldsymbol{E}\boldsymbol{H})^p \boldsymbol{x}_0 + \sum_{j=1}^p (\boldsymbol{\Phi} - \boldsymbol{E}\boldsymbol{H})^{p-j} \boldsymbol{E} \boldsymbol{z}_{j-1} \quad (3)$$

which means that the states, in a particular time  $p$ , depend on the initial state and past values of the output. We will use this equation afterward. On the other hand, by recursive substitution in (1) we obtain:

$$\boldsymbol{x}_p = \boldsymbol{\Phi}^p \boldsymbol{x}_0 + \sum_{j=1}^p \boldsymbol{\Phi}^{p-j} \boldsymbol{E} \boldsymbol{\psi}_{j-1} \quad (4)$$

and substituting (4) into the observation equation (2), we get:

$$\boldsymbol{z}_p = \boldsymbol{H}\boldsymbol{\Phi}^p \boldsymbol{x}_0 + \boldsymbol{H} \sum_{j=1}^p \boldsymbol{\Phi}^{p-j} \boldsymbol{E} \boldsymbol{\psi}_{j-1} + \boldsymbol{\psi}_p \quad (5)$$

so the endogenous variable,  $\boldsymbol{z}_p$ , depends on the initial state vector,  $\boldsymbol{x}_0$ , and the present and past innovation values,  $\boldsymbol{\psi}_p$ . Equation (5) can be written in matrix form as,

$$\boldsymbol{Z}_p = \boldsymbol{O}\boldsymbol{X}_0 + \boldsymbol{V}\boldsymbol{\Psi}_p \quad (6)$$

where the subscript  $p$  is deterministically chosen by the user. In this work, we will determine  $p = \max\{4, T^*\}$ , being  $T^*$  the nearest integer to the log of the sample size. This fulfills, in all the cases studied in this paper, conditions over  $p$  given by Bauer (2005) which assure the consistency of the estimators in the integrated case. However, our methodology can be used with any other suitable definition of  $p$ . Equation (6) and the subspace representation require the following matrices related to the data:

- 1) The Block-Hankel Matrices which organize the information of the output

and the (unobservable) input:

$$\mathbf{Z}_p = \begin{pmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_{T-2i+1} \\ \mathbf{z}_2 & \mathbf{z}_3 & \dots & \mathbf{z}_{T-2i+2} \\ \vdots & \vdots & & \vdots \\ \mathbf{z}_i & \mathbf{z}_{i+1} & \dots & \mathbf{z}_{T-i} \end{pmatrix}; \quad \mathbf{Z}_f = \begin{pmatrix} \mathbf{z}_{i+1} & \mathbf{z}_{i+2} & \dots & \mathbf{z}_{T-i+1} \\ \mathbf{z}_{i+2} & \mathbf{z}_{i+3} & \dots & \mathbf{z}_{T-i+2} \\ \vdots & \vdots & & \vdots \\ \mathbf{z}_{2i} & \mathbf{z}_{2i+1} & \dots & \mathbf{z}_T \end{pmatrix} \quad (7)$$

where  $\mathbf{z}_1$  is a vector of  $m$  components. In (6),  $\Psi_p$  is as  $\mathbf{Z}_p$  but with  $\boldsymbol{\psi}_t$  instead of  $\mathbf{z}_t$ . Note that for simplicity, in (7), we have assumed that  $p = f = i$ , that is, the dimension of the past and future information sets is the same.

2) The state sequence which is defined as  $\mathbf{X}_t = (\mathbf{x}_t \quad \mathbf{x}_{t+1} \quad \mathbf{x}_{t+2} \quad \dots \quad \mathbf{x}_{t+T-2i})$ . Specially, we will use the past and future state sequences, defined, respectively, as  $\mathbf{X}_p = \mathbf{X}_0$  and  $\mathbf{X}_f = \mathbf{X}_i$ .

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

3) The Extended Observability Matrix:

$$\mathbf{O} = \left( \mathbf{H}' \quad (\mathbf{H}\Phi)' \quad (\mathbf{H}\Phi^2)' \quad \dots \quad (\mathbf{H}\Phi^{i-1})' \right)' \in \mathbb{R}^{im \times n} \quad (8)$$

4) The lower block triangular Toeplitz matrix:

$$\mathbf{V} = \begin{pmatrix} \mathbf{I}_m & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\mathbf{E} & \mathbf{I}_m & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\Phi\mathbf{E} & \mathbf{H}\mathbf{E} & \mathbf{I}_m & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{H}\Phi^{i-2}\mathbf{E} & \mathbf{H}\Phi^{i-3}\mathbf{E} & \mathbf{H}\Phi^{i-4}\mathbf{E} & \dots & \mathbf{I}_m \end{pmatrix} \in \mathbb{R}^{im \times im} \quad (9)$$

From equation (3), for proper choices of  $p$  (see Bauer, 2005, integrated case) and the Assumptions given previously, the term including the initial state is negligible and the future state sequence,  $\mathbf{X}_f$ , can be approximated by a linear combination

of the past of the output,  $\mathbf{M}\mathbf{Z}_p$ . Therefore, shifting time subscripts in (6) and substituting  $\mathbf{X}_f$  by  $\mathbf{M}\mathbf{Z}_p$ , we obtain:

$$\mathbf{Z}_f = \mathbf{O}\mathbf{M}\mathbf{Z}_p + \mathbf{V}\Psi_f \quad (10)$$

where  $\mathbf{Z}_f$ ,  $\mathbf{Z}_p$  and  $\Psi_f$  are as in (7), and  $\mathbf{O}$  and  $\mathbf{V}$ , respectively, as in (8) and (9).

Subspace methods estimate the matrices  $\mathbf{O}$ ,  $\mathbf{M}$  and  $\mathbf{V}$  in (10) by solving a reduced-rank weighted least square problem, using the Singular Value Decomposition of the product  $\mathbf{W}_1\mathbf{Z}_f\mathbf{W}_2$ , being  $\mathbf{W}_1$  and  $\mathbf{W}_2$  two weighting matrices (see, e.g., Katayama, 2005). Then, the parameter matrices of (1-2) can be straightforwardly obtained from  $\mathbf{O}$ ,  $\mathbf{M}$  and  $\mathbf{V}$ . When  $\mathbf{W}_1 = (\mathbf{Z}_f\mathbf{Z}'_f)^{-\frac{1}{2}}$  and  $\mathbf{W}_2 = \mathbf{Z}'_p(\mathbf{Z}_p\mathbf{Z}'_p)^{-1}\mathbf{Z}_p$ , the singular values resulting from the SVD agree with the CCCs between  $\mathbf{Z}_f$  and  $\mathbf{Z}_p$ .

### 3 DETECTION OF UNIT ROOTS

BW propose a fast procedure to detect unit roots which, instead of calculating the CCC between  $z_t$  and  $z_{t-1}$ , uses the CCCs between the past and future information sets,  $\mathbf{Z}_p$  and  $\mathbf{Z}_f$  respectively, including therefore, the short-run dynamics. The authors compare any singular value  $\hat{\sigma}_j$ , where  $j \leq i$ , with a penalty function  $C(T)$ , so that any  $\hat{\sigma}_j$  fulfilling the inequality  $1 - \hat{\sigma}_j^2 - C(T) < 0$  can be considered as equal to one, corresponding to a unit root. BW specify  $C(T) = \log(T)^2/T$  which collapses to zero as the sample size grows. However, this loss function poses two main problems as it only assures consistency when the number of unit roots is  $d = 1$  (see BW) and does not depend on the nuisance parameter  $i$ , which is the dimension of  $\mathbf{Z}_p$  (and  $\mathbf{Z}_f$ ). 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 should also be taken into account because 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 increases.

To solve both limitations we formulate a decision criterion, based on Poskitt's (2000) and on a simulation study, that can be applied to systems with multiple unit roots. The key idea consists of obtaining a family of loss functions, denoted by  $G_l$ , that distinguishes the  $\hat{\sigma}_j$  considered equal to 1 from the others, using the superconsistency property. For this purpose, we suggest a penalty function,  $G_l(T, i, \bar{d})$ , that depends on the sample size ( $T$ ), the dimension of  $\mathbf{Z}_p$  ( $i$ ) and the number of unit roots we wish to evaluate ( $\bar{d}$ ). Therefore, a general representation of the *Unit Roots Criterion* would be:

$$URC(G_l) = f(\hat{\sigma}_j) - G_l(T, i, \bar{d}) \leq 0 \quad (11)$$

Even if *URC* is not, strictly speaking, an information criterion, by analogy we adopt this terminology throughout the paper, as  $f(\hat{\sigma}_j)$  is a measure of the distance from the exact non-stationarity, i.e. the unity, balanced with a penalty term  $G_l(T, i, \bar{d})$ .

**Proposition 3.1** *Let  $d$  denote the number of  $I(1)$  series of a process and let  $\hat{\sigma}_{j,T}^2 \rightarrow \sigma_j^2$  at rate  $T$  for  $j = 1, 2, \dots, d$  and at rate  $\sqrt{T}$  for  $j = d + 1, \dots, im$ . Consider, also, that  $G_l(T, i, \bar{d})$  is a decreasing function in  $T$ , such that  $G_l \rightarrow 0$  at rate  $T^r$ , being  $\frac{1}{2} < r < 1$ . Then the estimated number of unit roots,  $\hat{d}$ , obtained by *URC* as defined in (11), converges in probability to  $d$ , i.e.,  $\hat{d} \xrightarrow{p} d$ .*

The proof is given in Appendix A.

Note that the conditions stated in Proposition 3.1 can be satisfied by a wide family of penalty functions. In the next subsections we will illustrate how to find particular specifications for  $G_l(T, i, \bar{d})$  in univariate (Section 3.1) and multivariate (Section 3.2) processes by assuming a Gaussian distribution. Obviously, this procedure could be repeated with any other distribution according to the characteristics of the data, which is one of the advantages of the procedure.

### 3.1 Univariate processes

To find a particular function  $G_l(T, i, \bar{d})$  with good performance in finite samples, we build on the proposal of Bengtsson and Cavanaugh (2006), who estimate penalty functions via Monte Carlo simulations, to improve the performance of the AIC in small samples. We apply this idea to analyze the empirical distribution of the  $\hat{\sigma}_j$  corresponding to non-unit roots with the following procedure:

1. Fix the value of a CCC, denoted by  $\bar{\sigma}$ , near to non-stationarity. This parameter plays the role of the confidence level. In fact, the closer  $\bar{\sigma}$  is to the unity, the less permissive will be the procedure to detect unit roots. Here we use  $\bar{\sigma} = .95$  and we recommend that  $\bar{\sigma} \in [.90, .99]$ .
2. Simulate a large number of replications of models characterized by  $\bar{\sigma}$ , increasing the sample size one-by-one, obtaining for each  $T$  and replication, a sequence of CCCs such that:  $\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \dots \geq \hat{\sigma}_i \geq 0$ . That is, an empirical distribution of  $\hat{\sigma}_j$ ,  $j = 1, \dots, i$  for each  $T$ .
3. Select a function,  $f(\cdot)$ , to measure the distance between the  $j$ -th singular value and the unity. In this work we use  $1 - \hat{\sigma}_j^2$ , but other could be possible.
4. Fix a percentile ( $\tau$ ) that summarizes the information of the empirical distribution of  $f(\hat{\sigma}_j)$  in single series denoted by  $y_{j,t}$ . We recommend that  $\tau \in [75, 99]$ . As it happens with  $\bar{\sigma}$ , the higher is  $\tau$ , the less permissive will be the procedure to detect unit roots.
5. Specify and estimate an equation that depends on  $T, i$  and  $\bar{d}$ , correctly fitting  $y_{1,t}$  and fulfilling the assumptions of Proposition 3.1.

Following these five steps yields a concrete specification for  $G_l(T, i, \bar{d})$ . In this illustrative example, the stochastic processes simulated in step 2 have been specify assuming a zero-mean Gaussian distribution. This assumption has been adopted to facilitate the comparison with other detecting procedures. Note, however, that the method can accommodate any other distribution. Note, also, that the specific form of the data generating processes assumed is not very relevant in this

procedure, because many of them imply the same theoretical  $\sigma_1$ . The important features here are the probability distribution assumed and the value of  $\bar{\sigma}$ . Table 1 shows these values for the univariate case.

For  $\bar{d} = 0$ , the following penalty function is considered:

$$G_l(T, i, 0) = \hat{\alpha}_{1,0} T^{\hat{\beta}_{1,0}} i^{\hat{\beta}_{2,0}} \quad (12)$$

since it fulfills the assumptions of Proposition 3.1 for some  $(\hat{\alpha}_{1,0}, \hat{\beta}_{1,0}, \hat{\beta}_{2,0})$  and models like  $\log(y_{1,t}) = \alpha_{1,0} + \beta_{1,0} \log(T) + \beta_{2,0} \log(i) + \varepsilon_t$ , where  $\varepsilon_t$  is an error term, present adequate goodness-of-fit values ( $R^2 > .9$ ). Obviously, the estimated parameters in (12) depend on  $\bar{\sigma}$  and  $\tau$ , needed to define  $y_{1,t}$ . Table 1 also shows the estimated parameters in (12) corresponding to different  $\tau$  and  $\bar{\sigma}$ .

[INSERT FIGURE 1 AND TABLE 1]

For comparison, Table 1 includes the penalty function proposed by BW. When model (12) is fitted to BW's function, its estimates are close to those obtained with our penalty terms. We denote our different loss functions by  $G_l$ , with  $l = 1, 2, \dots$ . As all of them are consistent, since they fulfill Proposition 3.1, we will choose the penalty function on the basis of their small sample properties. To this end, following Bewley and Yang (1998), we will use a Gaussian random walk,  $\nabla z_t = a_t$ , as an approximation of the size of the criterion in univariate models and a stationary model with a persistent autoregressive process,  $(1 - .9B)z_t = a_t$ , to approximate the size-unadjusted power. Simulation results for 1000 replications are presented in Figure 2 which shows the performance in finite samples, both in empirical size and power, of the criteria. Note that some of them beat the ADF test in such properties for these models.

To finally obtain a concrete specification of the penalty terms, we choose those which fulfill (at least one of) the following conditions:

- *Condition 3.1:* The empirical size is 5% for  $n = 50$ .

- *Condition 3.2*: The empirical power is 50% for  $n = 50$ .

The reason we suggest two conditions instead of only one will be argued in Section 4. On the basis of the results in Figure 2, we select  $G_1$ , hereafter called  $G(T, i, 0)$ , as it fulfills both, Conditions 3.1 and 3.2.

[INSERT FIGURE 2]

### 3.2 Multivariate processes

Another advantage of our method lies in its ability to accommodate simple and multiple time series. This is relevant because of two reasons. First, few authors as Phillips and Durlauf (1986), 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 cointegration analysis.

Estimated penalty functions not only depend on  $T$  and  $i$ , but also on the minimum number of unit roots we want to test,  $\bar{d}$ . Therefore, each  $\hat{\sigma}_j$  candidate to be equal to 1 should be compared with a different penalty function. So far, the criteria only indicates whether the series are stationary or not. In this Section, we devise criteria to sequentially check if  $d > k$ ,  $k = 1, 2, 3$ . To estimate the corresponding sequence of penalties, we simulate vectors of two, three and four series with unit and less than unit CCCs. Each system contains at least one CCC stationary but close to unity. Then, different empirical distributions are estimated for every process, to finally choose a model that fits well the data, as we did in the univariate case. One could use these criteria in a sequential way as, if we compare a multivariate process  $(\mathbf{A}_t)$  having  $d$  unit roots, with the same system  $(\mathbf{B}_t)$  that incorporates a non-stationary new series (process  $\mathbf{B}_t$  would have  $d + 1$  unit roots) then  $\hat{\sigma}_j(\mathbf{B}_t) \geq \hat{\sigma}_j(\mathbf{A}_t)$  for  $j = 1, \dots, d + 1$ .

[INSERT TABLE 2]

Table 2 presents the estimated parameters corresponding to two penalty functions of the type  $G_l(T, i, 1) = \hat{\alpha}_{1,1} T^{\hat{\beta}_{1,1}} i^{\hat{\beta}_{2,1}}$ . The first one, which fulfills condition

3.1, is denoted by  $G_a(T, i, 1)$ . The second one, which fulfills condition 3.2, is denoted by  $G_b(T, i, 1)$ .

The method to estimate penalty functions for multivariate processes is similar to the one described in the previous section. However, since the features of the simulated data have changed, we re-specify the model to improve the fit to the empirical distribution by modeling two sub-samples:

$$\begin{aligned} T < T_{\bar{d}}^* & \quad y_{\bar{d}+1,t} &= \alpha_{2,\bar{d}} + \delta_{1,\bar{d}}T + \delta_{2,\bar{d}}T^2 + \delta_{3,\bar{d}}T^3 + \varepsilon'_{\bar{d},t} \\ T \geq T_{\bar{d}}^* & \quad \log(y_{\bar{d}+1,t}) &= \alpha_{1,\bar{d}} + \beta_{1,\bar{d}}\log(T) + \beta_{2,\bar{d}}\log(i) + \varepsilon''_{\bar{d},t} \end{aligned} \quad (13)$$

where  $T_{\bar{d}}^*$  refers to the first observation of the second sub-sample and  $\bar{d} = 2, 3$ . Estimates of the two penalty functions fulfilling conditions 3.1 and 3.2 to test the existence of at least  $j - 1$  unit roots (with  $j = 2, 3, 4$ ) are depicted in Table 2.

Figure 3 describes the generating processes and simulations done to assess the empirical size and power of the criteria, using the penalty functions presented in Table 2. All the criteria show good finite sample properties when estimating  $d$ , although the results reveal a degradation when the number of unit roots increases. We do not estimate loss functions to test  $d > 4$  for two reasons: a) previous criteria can be applied to processes with large dimensions but four or less than four unit roots and b) the degradation of the criteria performance as  $d$  increases and the complexity of the analysis suggest that this research, though feasible, would not be productive.

[INSERT FIGURE 3]

In short, we recommend the following loss functions to be used with the criterion in (11):

$$G_l(T, i, \bar{d}) = \begin{cases} \text{for } \bar{d} = 0, 1 & \hat{\alpha}_{1,\bar{d}}T^{\hat{\beta}_{1,\bar{d}}}i^{\hat{\beta}_{2,\bar{d}}} \quad \forall T, i \\ \text{for } \bar{d} = 2, 3, 4 & \begin{cases} \hat{\alpha}_{2,\bar{d}} + \hat{\delta}_{1,\bar{d}}T + \hat{\delta}_{2,\bar{d}}T^2 + \hat{\delta}_{3,\bar{d}}T^3 & T < T_{\bar{d}}^* \\ \hat{\alpha}_{1,\bar{d}}T^{\hat{\beta}_{1,\bar{d}}}i^{\hat{\beta}_{2,\bar{d}}} & T \geq T_{\bar{d}}^* \end{cases} \end{cases} \quad (14)$$

with  $l = a, b$  and where the choice between them will be motivated in the next section. Values for all the parameters for  $\bar{d} = 0, 1, 2, 3$  are presented in Tables 1 and 2.

## 4 COINTEGRATION

When working with multiple I(1) series, the cointegrating rank is  $c = m - d$ , being  $m$  the system dimension and  $d$  the number of unit roots. From this definition, we will devise a method to determine the cointegrating rank using the multivariate unit root detection procedure which consists of:

1. Check that the series are I(1), applying the criterion formulated in Section 3 to every single series.
2. Estimate the number of unit roots of the multivariate process ( $\hat{d}$ ).
3. Estimate the cointegrating rank as the difference between the system dimension and the number of unit roots estimated in step 2, i.e.,  $\hat{c} = m - \hat{d}$ .

We will now motivate why we proposed two penalty functions corresponding to conditions 3.1 and 3.2. In the univariate analysis, it would be coherent to use  $URC(G_a)$ , which has greater size than  $URC(G_b)$ , assuming that overdifferencing is better than underdifferencing (see Sanchez and Peña, 2001). In contrast,  $URC(G_b)$  will be more adequate for multivariate processes, where overestimating the cointegrating rank may be better than underestimating it. This is coherent with the widespread use of increasing the size of tests to 10% (see, e.g., Haug, 1996; Ho and Sørensen, 1996). Additionally, one may apply the method to check if the deviations from the assumed cointegration relationship are effectively I(0). Doing this requires a procedure to estimate the cointegrating matrix. With this aim, we suggest a method that provides consistent estimates, only requiring the innovation model and the computed cointegrating rank,  $\hat{c}$ .

**Proposition 4.1** *Let  $\mathbf{Z}_t$  be made up of I(1) cointegrated series generated by (1-2) and fulfilling the Assumptions given in Section 2. Let  $\hat{n}$  and  $\hat{\mathbf{H}}$  denote, respectively,*

a consistent estimate of the system order and the matrix  $\mathbf{H}$  in (2). Then it is possible to compute a matrix,  $\hat{\mathbf{\Lambda}}$ , which is a consistent estimate of the cointegrating matrix of  $\mathbf{Z}_t$ .

The proof is given in Appendix A.

## 5 SIMULATION EVIDENCE

In the first exercise, the data generating processes (DGP) are standard in performance analyses of cointegration tests, such as those of Riemers (1992), Toda (1995) or Poskitt (2000). Here we present the results for some trivariate and pentivariate DGPs although a more exhaustive simulation exercise is available under request to the authors. We start by considering the trivariate process given by:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} .0 & .0 & 1.0 \\ .0 & .2 & .8 \\ .1 & .3 & .6 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix}, \quad (15)$$

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  $(\eta_{1t}, \eta_{2t}, \eta_{3t})' \sim N(\mathbf{0}, \mathbf{R}_3)$ , where  $\mathbf{R}_3$  is a diagonal matrix with ones on the main diagonal and  $\theta$  elsewhere.

Tables 3, 4 and 5 present the results obtained with different versions of the DGP (15). 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 trivariate processes contain three, two and one unit roots, respectively. The figures presented are the resulting relative frequency

for different estimated cointegrating rank,  $\hat{c}$ . Results have been obtained using different values for  $\alpha$ ,  $\delta$  and  $\theta$ . For each combination of parameters and sample size, we discard the 50 first output ( $y_{;t}$ ) observations to improve randomization.

[INSERT TABLES 3-5]

Table 3 shows that  $URC(G_a)$  outperforms the rest of the methods in system 1, which is a favorable scenario as there is no risk of cancelations. The outcomes reported in Tables 4 and 5 deteriorate in small samples as the processes include one or two autorregressive components close to the unit boundary. Naturally, in those cases, the criterion based on  $G_b$  beats the one based on  $G_a$  since, by Conditions 3.1 and 3.2,  $G_a$  tends to attribute the high autocorrelation to another stochastic trend and, consequently, to underestimate  $c$ . In Tables 3-5, the performance of  $URC(G_a)$  uniformly dominates Johansen's procedure while  $URC(G_b)$  proves to be the most reliable procedure as it is often the best or the second-best option (10 times out of 15). Overall,  $URC(G_a)$  correctly selects  $c$  about 81% of the time, followed by  $URC(G_b)$  and  $\nabla_T$  (61%),  $PLR_T^{SBC}$  and  $PLR_T^{LP}$  (56%), and finally  $LR_T$  (51%).

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

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ y_{4t} \\ y_{5t} \end{pmatrix} = \begin{pmatrix} .0 & .0 & .0 & .0 & 1.0 \\ .0 & .0 & .2 & .0 & .8 \\ .0 & .0 & .3 & .1 & .6 \\ .0 & .1 & .4 & .1 & .4 \\ .1 & .1 & .3 & .3 & .2 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \\ x_{4t} \\ x_{5t} \end{pmatrix} \quad (16)$$

where the  $\mathbf{X}_t = (x_{1t}, x_{2t}, \dots, x_{5t})'$  vector is generated from  $\Phi(B)\mathbf{X}_t = \Delta + \Theta(B)\boldsymbol{\eta}_t$  with,  $\Phi(B) = \text{diag}(1, 1, 1, \nabla, \nabla)$  and  $\Theta(B) = \text{diag}((1 - \alpha B)^{-1}, (1 - \alpha B)^{-1}, (1 - \alpha B)^{-1}, 1, 1)$  for the pentavariate system 1;  $\Phi(B) = \text{diag}(1, 1, \nabla, \nabla, \nabla)$  and  $\Theta(B) = \text{diag}((1 - \alpha B)^{-1}, (1 - \alpha B)^{-1}, (1 - \mu B), 1, 1)$  for the system 2. Also, in both processes  $\Delta = (.0, .0, .0, \delta, \delta)'$  and  $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t}, \dots, \eta_{5t})'$  with  $\boldsymbol{\eta}_t = N(\mathbf{0}, \mathbf{R}_5)$ , where  $\mathbf{R}_5$

denotes a diagonal matrix with ones on the main diagonal and  $\theta$  elsewhere.

[INSERT TABLES 6-9]

Table 6 presents the results obtained with process 1, where  $URC(G_a)$ ,  $URC(G_b)$  and  $\nabla_T$  have pretty much achieved the asymptotic position by the time  $T = 300$ . Table 7 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  $1,300$ . When  $T = 500$ ,  $URC(G_b)$  clearly dominates the rest of the methods, whereas by the time  $T = 1,300$ , only  $PLR_T^{LP}$  still exhibits a strong negative bias.

The results pertaining to the second five-dimensional process are shown in Tables 8 and 9. Despite the fact that this system presents approximate cancellation of real autoregressive and moving average roots, our two criteria estimate  $c$  reasonably well. When the autoregressive persistence and the sample size grow, Table 9 shows a very good performance of  $URC(G_b)$ . In fact, from  $T = 500$ , it shows a 96% of correct cointegrating rank detection, widely outperforming the second best method, which is  $\nabla_T$  with 58%. As a summary of the results obtained with the pentivariate processes,  $URC(G_b)$  correctly chooses  $c$  just over 92% of the occasions, whereas  $\nabla_T$ ,  $URC(G_b)$ ,  $LR_T$  and  $PLR_T^{LP}$  estimate  $\hat{c} = c$  around 73%, 60%, 41% and 39%, respectively. Again,  $URC(G_b)$  is the most reliable procedure as it is the best option in 9 out of 12 processes and sample sizes.

Finally, we extend the Monte Carlo experiment to the analysis of the cointegrating matrix estimator. Table 10 shows the probability of correctly estimating the cointegrating rank with  $URC(G_b)$  and, when this identification is right, the computed cointegrating vector. The DGP presents a cointegration relationship and two non-stationary common factors with cointegrating coefficients  $\alpha_1$  and  $\alpha_2$ .

[INSERT TABLE 10]

The results shown in Table 10 indicate that the value of the autoregressive  $\phi$  is relevant as when  $\phi = -.8$  the estimates of  $c$ ,  $\alpha_1$  and  $\alpha_2$  are slightly less

accurate. The contemporaneous correlation of the innovations also affects on the cointegrating vector estimate, but it is not apparently clear in what way. This evidence suggests that the algorithm is able to estimate correctly the cointegrating vector with DGP containing a cointegration relationship and two non-stationary common factors. The outcomes support the consistency of the cointegrating rank and the cointegrating vector estimates.

## 6 CONCLUSIONS

This paper provides a family of subspace-based information criteria, useful to estimate the number of unit roots. We also propose a methodology to specify some particular penalty functions within this set of criteria. This is done by using simulation techniques, while bearing in mind the distributional properties of the data. The criteria can also be used to estimate the cointegrating rank in processes made up of  $I(1)$  single series. In that case, a consistent estimator of the cointegrating matrix is provided.

The procedure presents several advantages, as it: a) does not need to fit a model to the data, which usually causes several difficulties (see Emerson, 2007), b) can be applied to single and multiple time series, and c) can be adapted to any distribution.

Simulation exercises indicate that the methodology has a remarkable capacity to estimate the number of unit roots and the cointegration rank in different situations. Specifically, it works better than the alternatives in vector processes of relatively high dimension.

Concerning the flexibility of the procedure, 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 enables to design a criterion optimized for samples with non-standard characteristics

such as these. This idea will be explored in future works.

Finally, the procedures described in this article are implemented in a MATLAB toolbox for time series modeling called (this information has been deliberately omitted). The source code for all the functions is freely provided under the terms of the GNU General Public License.

## References

- Bauer, D. (2005). Estimating linear dynamical systems using subspace methods. *Econometric Theory*, 21:181–211.
- Bauer, D. and Wagner, M. (2002). Estimating cointegrated systems using subspace algorithms. *Journal of Econometrics*, 111:47–84.
- Bengtsson, T. and Cavanaugh, J. (2006). An improved Akaike information criterion for state-space model selection. *Computational Statistics and Data Analysis*, 50(10):2635–2654.
- Bewley, R. and Yang, M. (1995). Test for cointegration based on canonical correlation analysis. *Journal of the American Statistical Association*, 90:990–996.
- Bewley, R. and Yang, M. (1998). On the size and power of system tests for cointegration. *The Review of Economics and Statistics*, 80(4):675–679.
- Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, 2nd ed. edition.
- Casals, J., Jerez, M., and Sotoca, S. (2002). An exact multivariate model-based structural decomposition. *Journal of the American Statistical Association*, 97(458):553–564.
- Casals, J., Sotoca, S., and Jerez, M. (1999). A fast and stable method to compute the likelihood of time invariant state space models. *Economics Letters*, 65(3):329–337.

- Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49:1057–72.
- Emerson, J. (2007). Cointegration analysis and the choice of lag length. *Applied Economics Letters*, 14:881–885.
- 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.
- García-Hiernaux, A., Jerez, M., and Casals, J. (2009). Fast estimation methods for time series models in state-space form. *Journal of Statistical Computation and Simulation*, DOI:10.1080/00949650701617249.
- Haug, A. (1996). Tests for cointegration: A Monte Carlo comparison. *Journal of Econometrics*, 71:89–115.
- Ho, B. and Kalman, R. (1966). Effective construction of linear state-variable models from input-output functions. *Regelungstechnik*, 14:545–548.
- Ho, M. and Sørensen, B. (1996). Finding cointegration rank in high dimensional systems using the johansen test: An illustration using data based Monte Carlo simulations. *Review of Economics and Statistics*, 78:726–732.
- 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.
- Katayama, T. (2005). *Subspace Methods for System Identification*. Springer Verlag, London.
- 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.

- Lütkepohl, H. and Poskitt, D. S. (1998). *Econometrics in Theory and Practice: Festschrift for Hans Schneeweiß*, chapter Consistent Estimation of the Number of Cointegration Relations in a Vector Autorregressive Model, pages 87–100. Springer Verlag.
- 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.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464.
- 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.

## Appendix A. Proofs

**Proof of Proposition 3.1.** For simplicity we avoid the more appropriate nomenclature  $\hat{\sigma}_{j,T}$  and  $G_l(T, i, \bar{d})$  using  $\hat{\sigma}_j$  and  $G_l$  instead. From Bauer and Wagner (2002) we have that  $\hat{\sigma}_k^2 \rightarrow 1$  at rate  $T$ ,  $\forall k \leq d$ , whereas  $\hat{\sigma}_k^2 \rightarrow \sigma_k^2 < 1$  at rate  $\sqrt{T}$ ,  $\forall k > d$ . Consider the continuous function  $f(\hat{\sigma}_k)$  such that  $f(\hat{\sigma}_k) \rightarrow f(\sigma_k) = 0$  at rate  $T$ ,  $\forall k \leq d$ , and  $f(\hat{\sigma}_k) \rightarrow f(\sigma_k) > 0$  at rate  $\sqrt{T}$ ,  $\forall k > d$ . Then, as  $T \rightarrow \infty$ :

(i)  $\forall k \leq d$ ,  $P[f(\hat{\sigma}_k) - G_l \leq 0] \rightarrow 1$ , since  $f(\hat{\sigma}_k) \rightarrow 0$  at rate  $T > T^r$  and then faster than  $G_l \rightarrow 0$ ;

(ii)  $\forall k > d$ ,  $P[f(\hat{\sigma}_k) - G_l \leq 0] \rightarrow 0$ , as  $f(\hat{\sigma}_k) \rightarrow f(\sigma_k) > 0$  at rate  $\sqrt{T} < T^r$  and then slower than  $G_l \rightarrow 0$ .

Therefore, for every  $\epsilon > 0$ ,  $P[|d - \hat{d}| < \epsilon] \rightarrow 1$  as  $T \rightarrow \infty$  and, consequently,  $\hat{d} \rightarrow d$  in probability.

□

**Proof of Proposition 4.1.** Any model made up of I(1) series can be written as an innovation model, which divides the state sequences into stationary and non-stationary (see Casals et al., 2002). Thus, system (1-2) can be expressed, with the  $\Phi$  matrix in Jordan normal form, as:

$$\begin{pmatrix} \mathbf{X}_{1,t+1} \\ \mathbf{X}_{2,t+1} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_d & \mathbf{0} \\ \mathbf{0} & \phi_{n-d} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{1,t} \\ \mathbf{X}_{2,t} \end{pmatrix} + \begin{pmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{pmatrix} \psi_t \quad (17)$$

$$\begin{pmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{1,t} \\ \mathbf{X}_{2,t} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\psi}_{1,t} \\ \boldsymbol{\psi}_{2,t} \end{pmatrix} \quad (18)$$

where  $\mathbf{Z}_{1,t} \in \mathbb{R}^c$ ,  $\mathbf{Z}_{2,t} \in \mathbb{R}^d$ ,  $\mathbf{H}_{11} \in \mathbb{R}^{c \times d}$ ,  $\mathbf{H}_{12} \in \mathbb{R}^{c \times (n-d)}$ ,  $\mathbf{H}_{21} \in \mathbb{R}^{d \times d}$  and  $\mathbf{H}_{22} \in \mathbb{R}^{d \times (n-d)}$ . Also,  $n$  denotes the system order,  $\mathbf{X}_{1,t+1}$  designates the state components of the  $d$ -dimensional non-stationary sub-system and  $\mathbf{X}_{2,t+1}$  denotes the state components of the  $(n-d)$ -dimensional stationary sub-system.

Building from this representation, we look for a matrix,  $\mathbf{\Lambda}$ , that cancels  $\mathbf{H}_{11}$  which is the only block of  $\mathbf{H}$  affecting the non-stationary state sequence for the cointegrated series. As a normalization must be chosen, we define  $\mathbf{\Lambda} = (\mathbf{I}_c \quad -\boldsymbol{\lambda})$ . Thus, premultiplying equation (18) by  $\mathbf{\Lambda}$  we obtain:

$$\mathbf{\Lambda} \begin{pmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \end{pmatrix} = \mathbf{\Lambda} \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{1,t} \\ \mathbf{X}_{2,t} \end{pmatrix} + \mathbf{\Lambda} \begin{pmatrix} \boldsymbol{\psi}_{1,t} \\ \boldsymbol{\psi}_{2,t} \end{pmatrix} \quad (19)$$

defining  $\boldsymbol{\lambda} = \mathbf{H}_{11}\mathbf{H}_{21}^{-1}$  and operating, we get:

$$\mathbf{Z}_{1,t} - \boldsymbol{\lambda}\mathbf{Z}_{2,t} = (\mathbf{H}_{12} - \boldsymbol{\lambda}\mathbf{H}_{22})\mathbf{X}_{2,t} + \boldsymbol{\psi}_{1,t} - \boldsymbol{\lambda}\boldsymbol{\psi}_{2,t} \quad (20)$$

Equation (20) is a linear combination of the multivariate process  $\mathbf{Z}_t$  that is not influenced by the non-stationary sequence  $\mathbf{X}_{1,t}$ , so it is stationary.

On the other hand, if  $n$  and  $\mathbf{H}$  are estimated consistently, for instance, with Schwarz (1978) and García-Hiernaux et al. (2009), respectively, then  $\hat{\mathbf{\Lambda}} \xrightarrow{p} \mathbf{\Lambda}$ .

□

## Tables and Figures

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

Empirical distribution		Estimates			Penalty	
$f(\hat{\sigma}_1) = y_{1T}$	$\bar{\sigma}$	$\tau$	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	Function
$1 - \hat{\sigma}_1^2$	0.95	90	0.61	-0.51	-0.10	$G_1(T, i, 0)$
$1 - \hat{\sigma}_1^2$	0.95	80	0.35	-0.51	-0.05	$G_2(T, i, 0)$
$1 - \hat{\sigma}_1^2$	0.95	95	0.5	-0.51	-0.01	$G_3(T, i, 0)$
$\log(T)^2/T^*$	-	-	0.98	-0.52	-0.10	$C(T)$

\* Corresponding to the penalty function proposed by BW

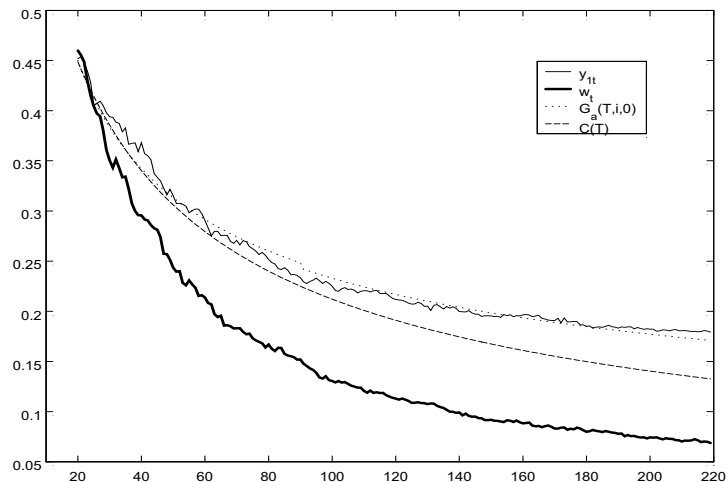


Figure 1: Simulated Singular Values (SV) and penalty functions for univariate processes.  $y_{1T}$  corresponds to a stationary process' first SV and  $w_T$  corresponds to a non-stationary process' first SV.  $G_a(T, i, 0)$  is a loss function provided for univariate processes and  $C(T)$  is the penalty 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.

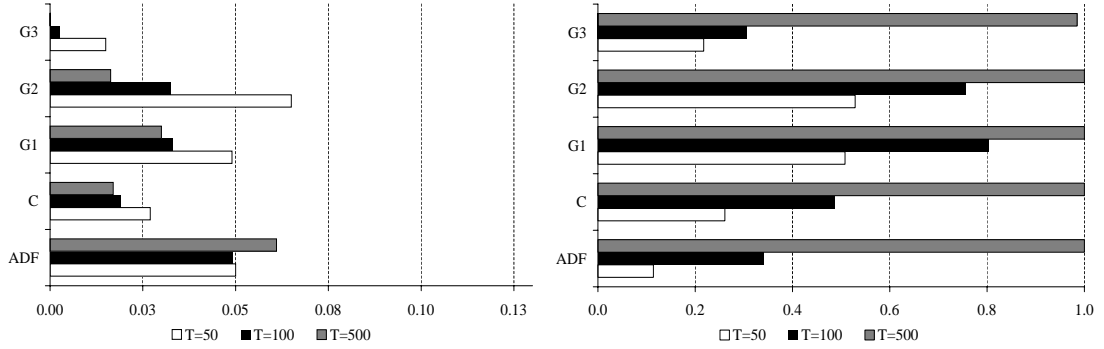


Figure 2: Empirical size and power of the criteria to test  $d > 0$  in the process  $\phi(B)z_t = a_t$ .  $C$  corresponds to the penalty function proposed by BW.  $ADF$  is the Augmented Dickey-Fuller test at 5% of significance level with one lag. Left plot: Empirical size when  $\phi(B) = 1 - B$ . Right plot: Empirical power when  $\phi(B) = 1 - .9B$

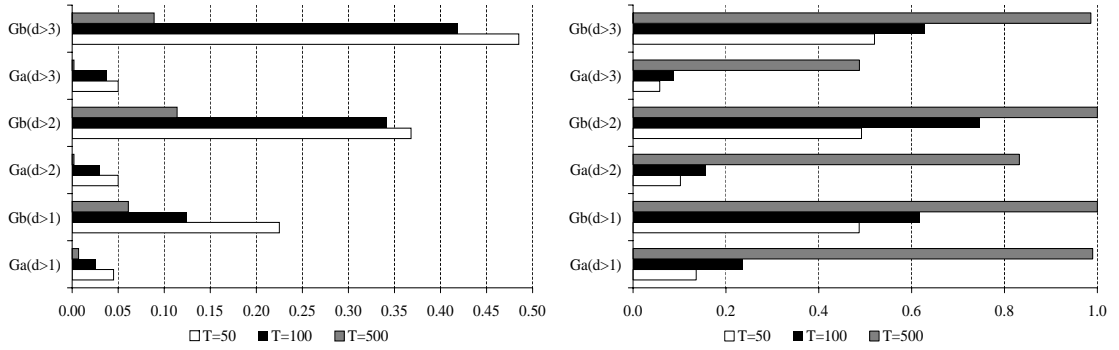


Figure 3: Empirical size and power of the criteria to test  $d > k$ ,  $k = 1, 2, 3$  in the processes  $\Phi_k(B)z_t = a_t$ . Left plot: Empirical size when  $\Phi_1(B) = \text{diag}[\nabla, \nabla]$ ;  $\Phi_2(B) = \text{diag}[\nabla, \nabla, \nabla]$ ;  $\Phi_3(B) = \text{diag}[\nabla, \nabla, \nabla, \nabla]$ . Right plot: Empirical power when  $\Phi_1(B) = \text{diag}[\nabla, (1 - .9B)]$ ;  $\Phi_2(B) = \text{diag}[\nabla, \nabla, (1 - .9B)]$ ;  $\Phi_3(B) = \text{diag}[\nabla, \nabla, \nabla, (1 - .9B)]$ .

Table 2: Estimated penalty functions to test  $d > 1$ ,  $d > 2$  and  $d > 3$

Empirical distribution		$T < T^*$			$T \geq T^*$			Penalty Function			
$j$	$f(\hat{\sigma}_j) = y_{jT}$	$\bar{\sigma}$	$\tau$	$\hat{\alpha}_1$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\alpha}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	
2	$1 - \hat{\sigma}_j^2$	0.95	90	-	-	-	-	1.100	-0.505	-0.012	$G_a(T, i, 1)$
2	$1 - \hat{\sigma}_j^2$	0.95	75	-	-	-	-	0.750	-0.505	-0.014	$G_b(T, i, 1)$
3	$1 - \hat{\sigma}_j^2$	0.95	90	-0.145	0.040	$-6.5 \times 10^{-4}$	$3.3 \times 10^{-6}$	1.600	-0.505	-0.101	$G_a(T, i, 2)$
3	$1 - \hat{\sigma}_j^2$	0.95	60	-0.253	0.036	$-5.9 \times 10^{-4}$	$3.0 \times 10^{-6}$	1.010	-0.505	-0.052	$G_b(T, i, 2)$
4	$1 - \hat{\sigma}_j^2$	0.95	90	-0.334	0.044	$-6.0 \times 10^{-4}$	$2.7 \times 10^{-6}$	2.250	-0.505	-0.201	$G_a(T, i, 3)$
4	$1 - \hat{\sigma}_j^2$	0.95	60	-0.441	0.041	$-5.5 \times 10^{-4}$	$2.5 \times 10^{-6}$	1.681	-0.505	-0.250	$G_b(T, i, 3)$

for  $j = 2, T^* = 0$ ;

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

Table 3: Results for the trivariate process (15), system 1.\*

T	$\hat{c}$	$URC(G_a)$	$URC(G_b)$	$\nabla_T$	$LR_T$	$PLR_T^{SBC}$	$PLR_T^{LP}$
$c = 0, \theta = .0, \delta = .0$							
100	0	<i>0.9355</i>	0.5070	0.7560	0.9345	0.4085	0.4230
	1	0.0635	0.4585	0.2385	0.0625	0.0890	0.5585
	2	0.0010	0.0345	0.0055	0.0030	0.5025	0.0185
200	0	<i>0.9855</i>	0.7570	0.9385	0.9410	0.6125	0.8555
	1	0.0145	0.2370	0.0615	0.0530	0.0740	0.1455
	2	0.0	0.0060	0.0	0.0060	0.3135	0.0
300	0	<i>0.9940</i>	0.8600	0.9860	0.9505	0.7300	0.9690
	1	0.0060	0.1390	0.0140	0.0465	0.0695	0.0310
	2	0.0	0.0010	0.0	0.0030	0.2005	0.0

\*The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by  $\hat{c}$ . The higher frequency where  $\hat{c} = c$  is italicized for each sample.  $URC(G_a)$  and  $URC(G_b)$  represent the criterion (11) with the proposed penalty functions  $G_a$  and  $G_b$ , respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion ( $\nabla_T$ ), Johansen's (1988, 1991) likelihood ratio procedure, ( $LR_T$ ), and some penalized likelihood criteria ( $PLR_T$ ). These last require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwarz, 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 ( $T = 100, 200$  and  $300$ ) and DGP. See Poskitt (2000) for more details.

Table 4: Results for the trivariate process (15), system 2.\*

T	$\hat{c}$	$URC(G_a)$	$URC(G_b)$	$\nabla_T$	$LR_T$	$PLR_T^{SBC}$	$PLR_T^{LP}$
$c = 1, \alpha = .8, \theta = .0, \delta = .0$							
100	0	0.4445	0.0145	0.2015	0.7760	0.1280	0.1515
	1	0.5275	<i>0.7895</i>	0.7485	0.1985	0.0990	0.7850
	2	0.0275	0.1945	0.0500	0.0215	0.7730	0.0635
200	0	0.0670	0.0	0.0340	0.3090	0.0140	0.0985
	1	0.9180	0.8990	<i>0.9530</i>	0.6455	0.2155	0.8935
	2	0.0150	0.1015	0.0130	0.0410	0.7705	0.0080
300	0	0.0010	0.0	0.0030	0.0355	0.0	0.0200
	1	<i>0.9920</i>	0.9425	<i>0.9920</i>	0.9115	0.2785	0.9780
	2	0.0070	0.0575	0.0050	0.0500	0.7215	0.0020
$c = 1, \alpha = .9, \theta = .0, \delta = .0$							
100	0	0.8445	0.2390	0.5880	0.8925	0.3070	0.3310
	1	0.1495	<i>0.6545</i>	0.3970	0.0990	0.0740	0.6445
	2	0.0055	0.1060	0.0150	0.0070	0.6190	0.0245
200	0	0.7805	0.1155	0.6275	0.7810	0.2520	0.6090
	1	0.2155	<i>0.8360</i>	0.3705	0.1950	0.1200	0.3895
	2	0.0040	0.0740	0.0020	0.0210	0.6280	0.0015
300	0	0.5940	0.0245	0.5380	0.5525	0.1190	0.6300
	1	0.4060	<i>0.9400</i>	0.4610	0.4010	0.2245	0.3695
	2	0.0	0.0355	0.0010	0.0465	0.6565	0.0005

\*The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by  $\hat{c}$ . The higher frequency where  $\hat{c} = c$  is italicized for each sample.  $URC(G_a)$  and  $URC(G_b)$  represent the criterion (11) with the proposed penalty functions  $G_a$  and  $G_b$ , respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion ( $\nabla_T$ ), Johansen's (1988, 1991) likelihood ratio procedure, ( $LR_T$ ), and some penalized likelihood criteria ( $PLR_T$ ). These last require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwarz, 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 ( $T = 100, 200$  and  $300$ ) and DGP. See Poskitt (2000) for more details.

Table 5: Results for the trivariate process (15), system 3.\*

T	$\hat{c}$	$URC(G_a)$	$URC(G_b)$	$\nabla_T$	$LR_T$	$PLR_T^{SBC}$	$PLR_T^{LP}$
$c = 2, \alpha = .8, \theta = .0, \delta = .0$							
100	0	0.1,300	0.0	0.1690	0.4520	0.0195	0.0485
	1	0.3115	0.0310	0.3485	0.4060	0.0020	0.6515
	2	0.5190	0.9630	0.4825	0.1175	<i>0.9785</i>	0.3000
200	0	0.0	0.0	0.0340	0.0035	0.0	0.0110
	1	0.0110	0.0010	0.0710	0.2085	0.0	0.4535
	2	0.9630	0.9715	0.8950	0.7345	<i>1.0</i>	0.5355
300	0	0.0	0.0	0.0020	0.0	0.0	0.0000
	1	0.0	0.0	0.0080	0.0020	0.0	0.1510
	2	<i>1.0</i>	<i>1.0</i>	0.9900	0.9480	<i>1.0</i>	0.8490
$c = 2, \alpha = .9, \theta = .0, \delta = .0$							
100	0	0.7210	0.0935	0.5255	0.8245	0.2100	0.2660
	1	0.2420	0.5360	0.4175	0.1525	0.0340	0.6795
	2	0.0350	0.3590	0.0570	0.0185	<i>0.7560</i>	0.0545
200	0	0.4990	0.0145	0.6405	0.4205	0.0525	0.4355
	1	0.3365	0.2255	0.2990	0.4175	0.0040	0.5465
	2	0.1620	0.7430	0.0650	0.1290	<i>0.9435</i>	0.1800
300	0	0.2100	0.0	0.6310	0.0535	0.0015	0.4000
	1	0.3680	0.0430	0.2550	0.4565	0.0005	0.5740
	2	0.4170	0.9490	0.1140	0.4445	<i>0.9980</i>	0.0260

\*The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by  $\hat{c}$ . The higher frequency where  $\hat{c} = c$  is italicized for each sample.  $URC(G_a)$  and  $URC(G_b)$  represent the criterion (11) with the proposed penalty functions  $G_a$  and  $G_b$ , respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion ( $\nabla_T$ ), Johansen's (1988, 1991) likelihood ratio procedure, ( $LR_T$ ), and some penalized likelihood criteria ( $PLR_T$ ). These last require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwarz, 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 ( $T = 100, 200$  and  $300$ ) and DGP. See Poskitt (2000) for more details.

Table 6: Results for the pentivariate process (16), system 1.\*

T	$\hat{c}$	$URC(G_a)$	$URC(G_b)$	$\nabla_T$	$LR_T$	$PLR_T^{LP}$
$c = 3, \alpha = .8, \theta = .8, \delta = 1.0$						
100	0	0.1580	0.0	0.0370	0.0850	0.0
	1	0.6840	0.0490	0.2685	0.5695	0.0830
	2	0.1280	0.2030	0.4090	0.2935	0.6705
	3	0.0300	<i>0.6590</i>	0.2835	0.0450	0.2435
	4	0.0	0.0890	0.0020	0.0060	0.0030
200	0	0.0020	0.0	0.0110	0.0	0.0
	1	0.1460	0.0	0.0585	0.0555	0.1035
	2	0.3250	0.0010	0.1365	0.5405	0.7075
	3	0.5190	<i>0.9110</i>	0.7925	0.4060	0.1890
	4	0.0080	0.0880	0.0015	0.0320	0.0
300	0	0.0	0.0	0.0005	0.0	0.0
	1	0.0020	0.0	0.0025	0.0	0.0925
	2	0.0250	0.00	0.0095	0.0915	0.7310
	3	0.9690	0.9510	<i>0.9875</i>	0.8555	0.1765
	4	0.0040	0.0490	0.0	0.0490	0.0

\*The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by  $\hat{c}$ . The higher frequency where  $\hat{c} = c$  is italicized for each sample.  $URC(G_a)$  and  $URC(G_b)$  represent the criterion (11) with the proposed penalty functions  $G_a$  and  $G_b$ , respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion ( $\nabla_T$ ), Johansen's (1988, 1991) likelihood ratio procedure, ( $LR_T$ ), and some penalized likelihood criteria ( $PLR_T$ ). These last require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwarz, 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 ( $T = 100, 200$  and  $300$ ) and DGP. See Poskitt (2000) for more details.

Table 7: Results for the pentivariate process (16), system 1.\*

T	$\hat{c}$	$URC(G_a)$	$URC(G_b)$	$\nabla_T$	$LR_T$	$PLR_T^{LP}$
$c = 3, \alpha = .9, \theta = .8, \delta = 1.0$						
500	0	0.0	0.0	0.4685	0.0	0.0
	1	0.5230	0.0	0.2880	0.0015	0.4715
	2	0.2940	0.0060	0.1180	0.2715	0.5120
	3	0.1400	<i>0.9820</i>	0.1255	0.6800	0.0165
	4	0.0	0.0120	0.0	0.0490	0.0
800	0	0.0	0.0	0.1625	0.0	0.0
	1	0.0290	0.0	0.0825	0.0	0.3540
	2	0.1350	0.0	0.0510	0.0005	0.6085
	3	0.8360	<i>0.9950</i>	0.7040	0.9535	0.0375
	4	0.0	0.0050	0.0	0.0415	0.0
1,300	0	0.0	0.0	0.0005	0.0	0.0
	1	0.0	0.0	0.0005	0.0	0.0645
	2	0.0	0.0	0.0005	0.0	0.6665
	3	<i>1.0</i>	<i>1.0</i>	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  $\hat{c}$ . The higher frequency where  $\hat{c} = c$  is italicized for each sample.  $URC(G_a)$  and  $URC(G_b)$  represent the criterion (11) with the proposed penalty functions  $G_a$  and  $G_b$ , respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion ( $\nabla_T$ ), Johansen's (1988, 1991) likelihood ratio procedure, ( $LR_T$ ), and some penalized likelihood criteria ( $PLR_T$ ). These last require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwarz, 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 ( $T = 500, 800$  and  $1,300$ ) and DGP. See Poskitt (2000) for more details.

Table 8: Results for the pentavariate process (16), system 2.\*

T	$\hat{c}$	$URC(G_a)$	$URC(G_b)$	$\nabla_T$	$LR_T$	$PLR_T^{LP}$
$c = 2, \alpha = .8, \mu = .9, \theta = .8, \delta = 1.0$						
100	0	0.0	0.	0.0	0.0005	0.0005
	1	0.7830	0.2350	0.2990	0.1920	0.0695
	2	0.0855	0.6940	0.7335	0.6250	0.6550
	3	0.0	0.0710	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	0.3865	0.0035	0.0750	0.0	0.0845
	2	0.6125	0.9575	0.9025	0.3190	0.6865
	3	0.0010	0.0390	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	0.0595	0.0	0.0085	0.0	0.1425
	2	0.9405	0.9720	0.9145	0.0315	0.7155
	3	0.0	0.0270	0.0770	0.9250	0.1415
	4	0.0	0.0010	0.0	0.0395	0.0

\*The table shows the relative frequency of obtaining the estimated cointegrating rank denoted by  $\hat{c}$ . The higher frequency where  $\hat{c} = c$  is italicized for each sample.  $URC(G_a)$  and  $URC(G_b)$  represent the criterion (11) with the proposed penalty functions  $G_a$  and  $G_b$ , respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion ( $\nabla_T$ ), Johansen's (1988, 1991) likelihood ratio procedure, ( $LR_T$ ), and some penalized likelihood criteria ( $PLR_T$ ). These last require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwarz, 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 ( $T = 100, 200$  and  $300$ ) and DGP. See Poskitt (2000) for more details.

Table 9: Results for the pentivariate process (16), system 2.\*

T	$\hat{c}$	$URC(G_a)$	$URC(G_b)$	$\nabla_T$	$LR_T$	$PLR_T^{LP}$
$c = 2, \alpha = .9, \mu = .9, \theta = .8, \delta = 1.0$						
500	0	0.0	0.0	0.0	0.0	0.0
	1	0.6990	0.0325	0.4225	0.0	0.4025
	2	0.3010	<i>0.9635</i>	0.5775	0.0965	0.5615
	3	0.0	0.0040	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	0.2300	0.0	0.1580	0.0	0.5130
	2	0.7700	<i>0.9905</i>	0.8390	0.0	0.4685
	3	0.0	0.0095	0.0030	0.9605	0.0185
	4	0.0	0.0	0.0	0.0370	0.0
1,300	0	0.0	0.0	0.0	0.0	0.0
	1	0.0020	0.0	0.0055	0.0	0.2710
	2	<i>0.9980</i>	0.9945	0.9210	0.0	0.7080
	3	0.0	0.0055	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  $\hat{c}$ . The higher frequency where  $\hat{c} = c$  is italicized for each sample.  $URC(G_a)$  and  $URC(G_b)$  represent the criterion (11) with the proposed penalty functions  $G_a$  and  $G_b$ , respectively. Columns 5-8 show the results achieved by Poskitt's (2000) criterion ( $\nabla_T$ ), Johansen's (1988, 1991) likelihood ratio procedure, ( $LR_T$ ), and some penalized likelihood criteria ( $PLR_T$ ). These last require an order of the autoregressive model fitted to the data specification denoted by SBC (Schwarz, 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 ( $T = 500, 800$  and  $1,300$ ) and DGP. See Poskitt (2000) for more details.

Table 10: Estimated cointegrating rank and vector for  $\alpha_1 = -1.2$  and  $\alpha_2 = 0.5$ .\*

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

$\theta$	T	$\phi = -0.8$			$\phi = 0$			$\phi = 0.8$		
		$\hat{c} = 1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{c} = 1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{c} = 1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
0.2	50	0.714	-1.213	0.433	0.696	-1.184	0.487	0.653	-1.197	0.505
	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
0.4	50	0.705	-1.217	0.429	0.752	-1.193	0.520	0.695	-1.163	0.521
	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
0.6	50	0.736	-1.084	0.446	0.817	-1.185	0.533	0.731	-1.183	0.535
	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
0.8	50	0.725	-1.131	0.431	0.848	-1.173	0.526	0.788	-1.146	0.525
	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

\*  $\theta$  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 2000 replications.