



ELSEVIER

Journal of Econometrics 86 (1998) 129–154

JOURNAL OF
Econometrics

Pitfalls in testing for long run relationships

Jesús Gonzalo^{a,*}, Tae-Hwy Lee^b

^a*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Spain*

^b*Department of Economics, University of California, Riverside, CA 92521, USA*

Received 1 June 1995; received in revised form 1 July 1997; accepted 22 September 1997

Abstract

This paper analyzes the robustness of the two most commonly used cointegration tests: the single equation based test of Engle and Granger (EG) and the system based test of Johansen. We show analytically and numerically several important situations where the Johansen LR tests tend to find spurious cointegration with probability approaching one asymptotically. The situations investigated are of two types. The first one corresponds to variables that have long-memory properties and a trending behavior, but they are not pure $I(1)$ processes although they are difficult to tell from $I(1)$ with standard unit root tests. The second corresponds to $I(1)$ variables whose VAR representation has a singular or near-singular error covariance matrix. In most of the situations investigated in this paper, EG test is more robust than Johansen LR tests. This paper shows that a proper use of the LR test in applied cointegration analysis requires a deeper data analysis than the standard unit root test. We conclude by recommending to use both tests (EG and Johansen) to test for cointegration in order to avoid or to discover a pitfall. © 1998 Elsevier Science S.A. All rights reserved.

JEL classification: C32; C22

Keywords: Cointegration; Engle–Granger test; Johansen test; Explosive roots; Stochastic roots; $I(2)$; Fractional unit roots; Deterministic trends; Singular covariance matrix

1. Introduction

During the past decade a considerable amount of research has focused on the issue of stochastic trends in economic variables and subsequently on the

* Corresponding author. E-mail: jgonzalo@elrond.uc3m.es

eventual presence of cointegrating relationships among these variables. Many papers have analyzed the properties of different methods to study cointegration. A very complete list of the literature can be found in a survey by Watson (1995). To the best of our knowledge, none of the papers investigate the consequences that misspecification of the trend components can have on the cointegration testing results. They do not either emphasize enough the fact that different methods are designed to accomplish different objectives. This paper analyzes the robustness of the cointegration test against those types of misspecifications and shows that the main reason why different misspecifications hurt some methods more than others is because they have different loss functions.

In cointegration analysis, and in general in econometrics and statistics, most methods are based on the minimization of certain variances or the maximization of certain correlations. For instance, single equation methods, like the one suggested by Engle and Granger (1987), look for the linear combination of $(y_t \ x_t)'$ with minimum variance and therefore belong to the first category. On the contrary, reduced rank system based methods, like the approaches suggested by Johansen (1995) or Ahn and Reinsel (1990) belong to the second category, because they look for the linear combination of $(y_t \ x_t)'$ with maximum correlation with a linear combination of $(\Delta y_t \ \Delta x_t)'$, where $\Delta = 1 - B$ and B is the backshift operator.

This paper investigates the robustness of the two most commonly used cointegration tests: the single equation based test of Engle and Granger (EG hereafter) and the system based tests of Johansen. The situations under investigation are of two types. The first one corresponds to variables that have long-memory properties and a trend behavior, but they are not pure $I(1)$ processes although they can not be distinguished from $I(1)$ with standard unit root tests. The second corresponds to $I(1)$ variables whose VAR representation has a singular or near-singular error covariance matrix. The effects that these two types of misspecifications have on testing for cointegration are analyzed in relation with the different loss functions the EG and Johansen methods have. The paper will focus more on the first type because it is more relevant for empirical research. We show analytically and numerically that Johansen LR tests tend to find spurious cointegration with probability approaching one asymptotically. We define this fact (size going to one asymptotically) as a pitfall. Note that a pitfall is more serious than either a simple finite sample problem or a size distortion issue. We find that EG is more robust than the LR tests in most of the situations investigated in the paper; mainly because misspecifications of the long-memory components of the variables affect their correlation structure more than their variances. Nevertheless, we learn from using both tests. It is well known that in other dimensions the Johansen procedure is superior to the EG method. The paper recommends that when both tests give different outcomes, instead of simply attributing it to finite sample problems, the existence of a pitfall should be considered (similar to Hausman's (1978) idea on specification

testing) and therefore a deeper pre-cointegration analysis than a simple unit root testing should be conducted, especially testing for models that imply a different trend and long memory behavior than the one implied by a pure $I(1)$ model.

To place our results in the context of the literature on consequences in estimation and inference of model misspecification, note that in this paper we only show the consequences on testing for cointegration and not on estimation. The misspecifications we analyze are the following: (i) random versus constant unit root coefficient models (Section 4.2); (ii) different order of integration than one (Sections 4.3 and 4.5); (iii) omission of some deterministic components (Section 4.4); (iv) singularity of the error covariance matrix (Section 5) and (v) dynamic misspecification jointly with some of the previous misspecifications (Sections 4.3, 4.5 and 5.2). To the best of our knowledge, only some cases of (iii) and (v) have been treated in the literature (see Perron and Campbell, 1994; Toda, 1995) but for different reasons than they are here. In a more global sense, this paper can be seen as an extension of the research done in spurious regression (see Granger and Newbold, 1974) where the misspecification is also in the long memory component. The consequences of our misspecifications are even more serious than the ones of standard spurious regression, because in the latter case any cointegration test will detect the spurious regression by rejecting cointegration. This does not occur with the misspecifications investigated in this paper.

The paper is organized as follows. Section 2 of this paper reviews EG and the Johansen ML approaches. Section 3 defines the concept of pitfalls and classifies them in two major types. Sections 4 and 5 present, theoretically and with a Monte Carlo experiment, these two types of pitfalls where the LR tests tend to find spurious cointegration with probability approaching one asymptotically. Section 6 concludes. All proofs are gathered in the appendix.

2. Testing for cointegration

The components of an $n \times 1$ vector X_t are said to be cointegrated if all components of X_t are $I(d)$ and there exists a vector α ($\neq 0$) such that $\alpha'X_t$ is $I(d - b)$, $b > 0$. For the sake of brevity in presentation only bivariate systems ($X_t = (y_t \ x_t)'$) will be considered in this paper, but our results are valid for $n > 2$.

EG test for cointegration is based on the augmented Dickey–Fuller (DF) statistic (see Dickey and Fuller, 1979) of order k , $ADF(k)$, the t -value for $\hat{\rho}$ in the OLS regression

$$\Delta z_t = \rho z_{t-1} + \rho_1 \Delta z_{t-1} + \dots + \Delta z_{t-k} + \text{error}, \quad (1)$$

where z_t are the residuals from the OLS cointegration regression $y_t = \hat{a}_0 + \hat{a}_1 x_t + z_t$.

Reduced rank regression methods, like the Johansen approach, exploit the fact that $I(1)$ and $I(0)$ variables are asymptotically uncorrelated and look for

a vector α that maximizes the correlation between $\alpha'X_t$ and a linear combination of ΔX_t . More formally, what Johansen method does is to test for the rank of the long run impact matrix Π in the following VECM:

$$\Delta X_t = \mu + \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \dots + \Gamma_k \Delta X_{t-k} + \varepsilon_t, \tag{2}$$

where

$$\varepsilon_t = \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega \equiv \begin{pmatrix} \sigma_1^2 & \theta \sigma_1 \sigma_2 \\ \theta \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \right]. \tag{3}$$

Throughout the paper it will be assumed that $\sigma_1 = \sigma_2 = 1$ to simulate the series, and $\theta = 0$ in Section 4. If X_t is cointegrated, Π can be written as

$$\Pi = \gamma \alpha', \tag{4}$$

where α and γ are 2×1 matrices. Testing for cointegration is therefore equivalent to testing the rank of Π equals to one. The Johansen LR statistics for testing the null hypothesis of no cointegration $H_0 : r = 0$ are

$$Q_1 = -T \ln[(1 - \hat{\lambda}_1)(1 - \hat{\lambda}_2)], \tag{5}$$

$$Q_2 = -T \ln(1 - \hat{\lambda}_1), \tag{6}$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2$ are the eigenvalues of $S_{11}^{-1} S_{10} S_{00}^{-1} S_{01}$, and

$$S_{ij} = T^{-1} \sum_{t=1}^T R_{it} R'_{jt} \quad (i, j = 0, 1) \tag{7}$$

are the product moment matrices of the residuals, R_{0t} and R_{1t} , from the regressions of ΔX_t and X_{t-1} on the lagged differences, respectively. The ML estimator of the cointegrating vector is the eigenvector of $S_{11}^{-1} S_{10} S_{00}^{-1} S_{01}$ corresponding to $\hat{\lambda}_1$. The next proposition shows an easy way of obtaining the eigenvalues, $\hat{\lambda}_1$ and $\hat{\lambda}_2$, by estimating by OLS the VECM (2) equation by equation.

Proposition 1. The eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2$ of $S_{11}^{-1} S_{10} S_{00}^{-1} S_{01}$ are the same as those of $\hat{M} = I - S_{00}^{-1} \hat{\Omega}$, where $\hat{\Omega}$ is the covariance matrix of the residuals of the VECM, calculated under no rank constraint.

According to this proposition it is clear that the behavior of Johansen method depends only on the behavior of the matrices S_{00} and $\hat{\Omega}$. Next section characterizes the pitfalls of the cointegration tests on the basis of the behavior of these two matrices.

3. Pitfalls

We define a pitfall of a test for cointegration in the following way.

Definition 1. a. A pitfall exists if the size of a test for the null hypothesis of no cointegration approaches one asymptotically.

b. A pitfall exists if the power of a test for the null hypothesis of no cointegration tends to zero under the alternative of cointegration.

A pitfall of the Johansen LR tests occurs if Q_1 or Q_2 diverges when there is cointegration (size pitfall), or if Q_1 or Q_2 converges to zero when there is no cointegration (power pitfall). In this paper we will mainly focus on size pitfalls. The size pitfall may be called spurious cointegration instead of spurious regression, because the latter defined in Granger and Newbold (1974) and Yule (1926) can be detected by any cointegration test. In this paper, however, we show that some cointegration tests can reject the null hypothesis of no cointegration when in fact there is no cointegration. We emphasize that the issues are asymptotic, and that the problems are more serious as $T \rightarrow \infty$, unlike some finite sample characteristics of the unit root tests such as their low power and their size distortions (so well documented in the unit root literature). There are many situations that can produce pitfalls in the LR tests. Before examining some of them in more detail, we characterize and classify them in two types.

Proposition 2. A pitfall of Q_1 or Q_2 exists if $\hat{\lambda}_1$ does not converge to zero in probability or if $\hat{\lambda}_1 \xrightarrow{p} 0$ at a rate slower than T as $T \rightarrow \infty$, when there is no cointegration. Sufficient conditions for $\hat{\lambda}_1$ not to converge to zero in probability are

- a. $\det(S_{00}) = O_p(T^\delta), \delta > 0$, while $\det(\hat{Q})$ is bounded in probability, or
- b. $\det(\hat{Q}) \xrightarrow{p} 0$ as $T \rightarrow \infty$, while $\det(S_{00})$ does not converge to zero in probability.

Definition 2. a. A Type-A pitfall is a pitfall with $\det(S_{00}) = O_p(T^\delta), \delta > 0$.

b. A Type-B pitfall is a pitfall with $\det(\hat{Q}) \xrightarrow{p} 0$ as $T \rightarrow \infty$.

4. Type-A pitfalls

In this section we consider several alternative trending and long-memory processes. These processes are different from an $I(1)$ process but the null hypothesis of a unit root is very unlikely to be rejected by a standard DF test. The processes investigated are characterized by a diverging S_{00} . This makes Q_1 and Q_2 diverge too as $T \rightarrow \infty$ and therefore the null hypothesis of no cointegration is wrongly rejected.

4.1. Processes with AR roots close to unity

This section examines the robustness of cointegration tests when the largest autoregressive root of at least one of the variables in the system $X_t = (y_t \ x_t)'$ is

larger than one. In related papers, Cavanagh et al. (1995) and Elliott (1995) study the near unit root cases (for values below unity) focusing on how this affects the inference on the cointegrating vector. This section could be seen as an extension in one direction of these papers in the sense that we show how important it is to know the right order of integration in order to test for cointegration. Elliott (1995) also examines the Johansen tests when the roots are greater than one, but he analyzes the size of the tests for $r = 1$ versus $r = 2$ in a cointegrated system. In our case, we concern about testing the existence of cointegration itself ($r = 0$ versus $r = 1$). Consider the following simple DGP (data generating process):

$$y_t = ay_{t-1} + e_{1t}, \tag{8}$$

$$x_t = x_{t-1} + e_{2t}. \tag{9}$$

What are the values of a most likely in real economic data? Stock (1991) provides asymptotic confidence intervals for the autoregressive root of a time series when the root is close to one. When applied to the Nelson and Plosser (1982) data set, the main conclusion is that the confidence intervals are typically wide, so that the conventional emphasis on unit root tests may fail to convey the correct information. For example, the 90% confidence interval for bond yield in the US with 71 annual observations is (1.033, 1.078). The 90% confidence interval for nominal GNP with 62 annual observations is (0.847, 1.074).

To analyze the performance of the cointegration tests for model (8)–(9), we write that DGP in a VECM form

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t, \tag{10}$$

where

$$\Pi = \begin{pmatrix} a - 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a - 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix}. \tag{11}$$

Note $\alpha' = (1 \ 0)$, where α is defined from Eq. (4). If $a \neq 1$, then we have $\lambda_1 \neq 0$ and $\lambda_2 = 0$, since $r = 1$.

Proposition 3. Let X_t be generated from model (10).

- a. If $|a| = 1$, then $\hat{\lambda}_1 = O_p(T^{-1})$.
- b. If $|a| \leq 1$, then $\hat{\lambda}_1 \xrightarrow{p} (1 - a)/[(a + 1)\theta^2 - 2]$ as $T \rightarrow \infty$.
- c. If $|a| > 1$, then $\hat{\lambda}_1 \xrightarrow{p} 1$ as $T \rightarrow \infty$.

Proposition 3 shows that if $a > 1$, then $\hat{\lambda}_1$ converges to one and Q_1, Q_2 diverge as $T \rightarrow \infty$, rejecting with probability approaching one the null hypothesis of no cointegration, when there is no cointegration at all. Table 1 presents the size of the Johansen tests along with that of EG for $a > 1$ and $T = 100$. Q_1 and Q_2 find

Table 1

A root close to unity. DGP: $y_t = ay_{t-1} + e_{1t}$ and $x_t = x_{t-1} + e_{2t}$

	EG(y/x)	EG(x/y)	Q_1	Q_2
$a = 1.00$	0.050	0.049	0.048	0.048
$a = 1.01$	0.046	0.047	0.119	0.107
$a = 1.02$	0.031	0.049	0.435	0.422
$a = 1.03$	0.015	0.040	0.736	0.721
$a = 1.05$	0.000	0.027	0.965	0.964

$T = 100$. The frequency of rejecting the null hypothesis in 1000 replications is reported at the 5% level. The critical values are simulated from 90,000 replications using the DGP with $a = 1$. EG denotes the test of Engle and Granger (1987). y/x denotes the regression of the dependent variable y on the explanatory variable x . Similarly for x/y . Q_1 and Q_2 are Johansen's trace statistic and maximum eigenvalue statistic, respectively.

too much cointegration incorrectly and have a pitfall when $a > 1$. EG is clearly more robust than the LR tests.

Note that in general it will not be possible to detect this pitfall by examining the adjustment coefficient γ (see Gonzalo and Lee, 1996a).

It is known that when $a < 1$, the DF test diverges to $-\infty$ as $T \rightarrow \infty$ and the power of the test tends to one as $T \rightarrow \infty$. It is also known that when a is close to one the power of the DF test is very small in finite samples (see Gonzalo and Lee, 1996b; Stock, 1995, among others). When $a > 1$, it can be shown that the DF test diverges to $+\infty$ as $T \rightarrow \infty$ and therefore it has no power asymptotically. From this pitfall, it is learned that we should test for a unit root hypothesis against the right tail of the DF distribution as well as against the left tail of the distribution, especially when the unit root test is a pre-step for a cointegration test.

It does not seem very plausible to think any economic variable could be explosive forever. But it is not so unlikely to find variables that may behave like explosives for certain periods, for instance, the number of connections to internet in the last three years. It could be more likely to find variables that sometimes are stationary and sometimes explosive. This is the case examined in the next section.

4.2. Processes with stochastic AR roots

Granger and Swanson (GS henceforth, 1996) introduce a class of 'near' $I(1)$ processes, having an AR root that is not constant, but stochastic, and varying around unity. In this way, the process can be stationary for some periods ($a < 1$ in model (8)), and mildly explosive for others ($a > 1$). These stochastic unit roots (STUR) arise naturally in economics. If a consumer maximizes $E_0 \sum_{t=0}^{\infty} \beta^t u(c_t)$, $0 < \beta < 1$, subject to a budget constraint, the Euler equation is $u'(c_t) =$

$E_t \beta R_t u'(c_{t+1})$, where c_t is consumption and R_t is the real gross rate of return. Assuming a quadratic utility, we have $c_t = (\beta R_{t-1})^{-1} c_{t-1} + e_t$, and thus the process $\{c_t\}$ has a stochastic AR root equal to $(\beta R_{t-1})^{-1}$.

Consider a series y_t generated by

$$y_t = a_t y_{t-1} + e_{1t}, \tag{12}$$

where $e_{1t} \sim N(0, \sigma_1^2)$ and a_t follows a stochastic process. This is an example of a doubly stochastic process as considered by Tjostheim (1986). GS consider the exponential form

$$a_t = \exp(\alpha_t) \tag{13}$$

with α_t a Gaussian stationary series having mean m and variance σ_α^2 . In particular, GS consider the following AR(1) process for α_t

$$\alpha_t = \mu + \rho \alpha_{t-1} + \eta_t, \quad |\rho| < 1, \tag{14}$$

where $\eta_t \sim N(0, \sigma_\eta^2)$ and is independent of e_{1t} . Then $m = \mu / (1 - \rho)$ and $\sigma_\alpha^2 = \sigma_\eta^2 / (1 - \rho^2)$. Unlike an exact unit root process, a stochastic unit root process has an increasing variance even after first-differencing.

Proposition 4. Suppose $(y_t, x_t)'$ are STUR processes generated from Eqs. (12)–(14). Then

- a. $var(\Delta x_t)$ and $var(\Delta y_t)$ increase as t gets larger, and
- b. $\hat{\lambda}_1 \xrightarrow{P} 1$ as $T \rightarrow \infty$.

GS show that standard unit root tests, such as the DF test, can not easily distinguish between exact unit roots and stochastic unit roots. Processes against which unit root tests have little power are called ‘generic unit root’ processes in Granger (1993). GS find that STUR models can often forecast better than various other models including exact unit root models when considering a variety of macroeconomic time series. GS also develop a test for STUR against a unit root, but the new test for STUR indicates that it is difficult to tell between these two types of processes. Therefore, it is possible that economic variables may be generated with STUR but we think of them as exact unit root processes, and have a pitfall when we use the LR tests.

In the Monte Carlo experiment, both variables are generated with the same a_t . In Table 2, we examine the performances of the cointegration tests under various choices of μ , ρ , and σ_η^2 . As the choice of the parameters in this experiment is somehow arbitrary we use parameter values estimated in GS’s Table 3 for which ten monthly US macroeconomic series are used. Since GS show that standard tests can not easily distinguish between $I(1)$ and STUR, we also simulate the power of the DF test statistics for the processes used for Table 2. According to the estimated DF statistics, most of the processes are

Table 2
 STUR. DGP: $y_t = a_1 y_{t-1} + e_{1t}$ and $x_t = a_2 x_{t-1} + e_{2t}$, where $a_t = \exp(\alpha_t)$, $\alpha_t = \mu + \rho \alpha_{t-1} + \eta_t$

	T = 100			T = 1,000			T = 10,000					
	DF	EG	Q ₁	Q ₂	DF	EG	Q ₁	Q ₂	DF	EG	Q ₁	Q ₂
FM2	0.062	0.053	0.048	0.053	0.045	0.055	0.052	0.050	0.030	0.054	0.261	0.229
SPI	0.060	0.052	0.051	0.054	0.033	0.056	0.094	0.087	0.000	0.127	1.000	1.000
RL	0.063	0.053	0.046	0.050	0.077	0.062	0.041	0.040	0.689	0.282	0.494	0.214
RS	0.074	0.056	0.046	0.047	0.283	0.111	0.151	0.081	1.000	1.000	1.000	1.000
RM	0.071	0.056	0.047	0.048	0.201	0.089	0.101	0.063	1.000	1.000	1.000	1.000
IP	0.062	0.053	0.048	0.053	0.045	0.055	0.052	0.049	0.031	0.054	0.251	0.221
NI	0.058	0.052	0.053	0.054	0.025	0.057	0.329	0.290	0.000	0.438	1.000	1.000
UR	0.064	0.053	0.046	0.050	0.069	0.059	0.040	0.039	0.471	0.186	0.281	0.132
CPI	0.060	0.052	0.053	0.054	0.029	0.057	0.206	0.186	0.000	0.284	1.000	1.000
PPI	0.060	0.052	0.051	0.054	0.033	0.055	0.092	0.081	0.002	0.123	1.000	1.000

The frequency of rejecting the null hypothesis in 1,000 replications is reported at the 5% level. The critical values for each T are simulated from 90,000 replications using the DGP with $a_t = 1$. We use estimated parameter values in GS (Table 3) for 10 monthly macro and financial series in US. The first row is the results using the STUR parameters estimated for the actual economic series FM2, and similarly for the other rows. FM2 is money stock (M2) in 1982 dollars. SPI is S & P common stock price index. RL is the interest rate on US Treasury Notes (Composite – over 10 years). RS is three month US T-bill interest rate. RM is the interest rate on one year US Treasury Notes. IP is industrial production. NI is nominal inventories in manufacturing and trade sectors. UR is unemployment rate. CPI is consumer price index. PPI is producer price index. DF denotes the power of DF-t test statistics for a unit root in each univariate series.

generic unit root processes, except the processes generated using the RL, RS, RM, and UR parameterizations for large T . Depending on the values of m and σ_α^2 (or of $\mu, \rho,$ and σ_η^2), the sample size T may have to be very large to see a pitfall in the LR tests. Note that $Ea_t = 1$ if $m + \sigma_\alpha^2/2 = 0$ since $Ea_t = \exp(m + \sigma_\alpha^2/2)$. The estimated values of $(m + \sigma_\alpha^2/2)$ in GS's Table 3 are 0.000151 (for FM2), 0.001693 (SPI), -0.000735 (RL), -0.004093 (RS), -0.003050 (RM), 0.000144 (IP), 0.002027 (NI), -0.000642 (UR), 0.000973 (CPI), and 0.000589 (PPI). Table 2 shows that EG is more robust than the LR tests.

4.3. *I(2) processes*

There are some economic variables that can be considered to be $I(2)$. For instance, prices (p) are $I(2)$ if inflation rate is $I(1)$. At the same time, money supply (m) could be $I(2)$ if we consider real money balance ($m - p$) to be $I(1)$. Recent papers by Stock and Watson (1993), Haldrup (1994), Paruolo (1996), and Johansen (1995), p. 137) consider that variables like money stock and prices could be $I(2)$. In many European and east Asian countries, many economic series are in the border line between $I(2)$ and an $I(1)$ with drift, something almost impossible to distinguish with econometric tools. The purpose of this section is to show what happens when the true DGP is $I(2)$, but we mistakenly assume it as an $I(1)$ with drift, and we estimate a VECM with a constant term to compute the LR statistics.

The DGP in Table 3 is

$$\Delta^2 y_t = (1 - bB)e_{1t}, \tag{15}$$

$$\Delta^2 x_t = (1 - bB)e_{2t}, \tag{16}$$

with $b \neq 1$, and its VECM representation is

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \end{pmatrix} + \begin{pmatrix} (1 - bB)e_{1t} \\ (1 - bB)e_{2t} \end{pmatrix}. \tag{17}$$

Proposition 5. Suppose $(y_t, x_t)'$ are $I(2)$ processes generated from Eqs. (15) and (16).

- a. *If $k = 0, \hat{\lambda}_1$ does not converge to zero in probability.*
- b. *If $k \geq 1, T\hat{\lambda}_1 = O_p(1)$, i.e., $\hat{\lambda}_1 \xrightarrow{p} 0$ as $T \rightarrow \infty$.*

Proposition 5 shows that when the estimated VECM does not contain any lags, the LR tests will have a pitfall if at least one of the variables is $I(2)$. A way to avoid that is to introduce more lags in the VECM. Doing that the Johansen LR tests will only have a standard size distortion due to the use of the wrong critical values.

It is well known in time series that it is very difficult to tell between a trend stationary series and a difference stationary ($I(1)$) series, as well as between an

Table 3
 $I(2)$ DGP: $\Delta^2 y_t = (1 - bB)e_{1t}$ and $\Delta^2 x_t = (1 - bB)e_{2t}$

Panel A. $T = 100$		$b = 0$		$b = 0.7$		$b = 0.9$			
	EG	Q_1	Q_2	EG	Q_1	Q_2	EG	Q_1	Q_2
$k = 0$	0.042	0.980	0.985	0.027	0.929	0.937	0.009	0.509	0.548
$k = 1$	0.057	0.182	0.198	0.000	0.712	0.734	0.010	0.412	0.448
$k = 2$	0.066	0.194	0.222	0.002	0.440	0.487	0.007	0.346	0.378
$k = 5$	0.072	0.261	0.274	0.014	0.209	0.231	0.011	0.185	0.212
Panel B. $T = 1000$		$b = 0$		$b = 0.7$		$b = 0.9$			
	EG	Q_1	Q_2	EG	Q_1	Q_2	EG	Q_1	Q_2
$k = 0$	0.206	0.999	0.999	0.204	0.999	0.999	0.159	0.999	0.999
$k = 1$	0.055	0.184	0.186	0.018	0.943	0.948	0.083	0.991	0.993
$k = 2$	0.052	0.189	0.190	0.006	0.788	0.802	0.043	0.976	0.979
$k = 5$	0.057	0.184	0.189	0.010	0.268	0.287	0.010	0.876	0.885
$k = 10$	0.052	0.182	0.210	0.040	0.184	0.200	0.002	0.556	0.570

The frequency of rejecting the null hypothesis in 1,000 replications is reported at the 5% level. The critical values are simulated from 90,000 replications using two independent random walks, $\Delta y_t = e_{1t}$ and $\Delta x_t = e_{2t}$.

$I(2)$ series and an $I(1)$ series with drift. In the latter case, the distinction becomes even harder when we have an MA component in the VECM that cancels a part of the second unit root. The purpose of this section is to show what happens when the true DGP is $I(2)$, but we mistakenly assume it is an $I(1)$ with a drift, and therefore we estimate a VECM with a constant term to compute the Johansen statistics. If we know the variables X_t are $I(2)$ then we could use Paruolo (1996) method or at least include an extra lag ($k = 1$). However, when there is an MA(1) error in the VECM ($b \neq 0$), Table 3 shows that although the pitfall theoretically does not exist with $k = 1$, in finite samples we will have a huge size problem even for $k = 1$. Table 3 also shows that EG is not only more robust to this kind of misspecification than the LR tests, but it suffers almost no size problem at all. This is different from the standard MA size problem shown in Schwert (1989) for the case when variables are $I(1)$ with MA errors. It should be noted that when the variables are $I(1)$ with MA errors in the VECM, both EG and LR tests suffer a huge size problem as is expected from Schwert (1989).

The case $k = 0$ is very relevant in practice, because as shown by Gonzalo and Pitarakis (1996), in large systems when the errors are MA(1), information criteria such as the BIC very often chooses $k = 0$.

This pitfall shows how important it is to know the exact degree of integration of the variables in order to use Johansen LR tests properly. In a different situation, a recent paper by Dolado and Marmol (1996) show that a mistake in the order of integration affects seriously the estimation and inferences derived from the FM-OLS method of Phillips and Hansen (1990). Thus, a unit root test on the levels of the variables is not enough, in the sense that the null hypothesis of a unit root will usually not be rejected.

4.4. $I(1)$ processes with deterministic components

Perron and Campbell (1994) raised some issues that arise when Johansen’s framework is used to analyze cointegrating relationships among variables with deterministic linear trends. This section continues their idea but focusing on non-cointegrated systems that contain deterministic trend components. The DGP is

$$\Delta y_t = a_1 + b_1 t + e_{1t}, \tag{18}$$

$$\Delta x_t = a_2 + b_2 t + e_{2t}. \tag{19}$$

Proposition 6. Suppose $(y_t, x_t)'$ are two $I(1)$ processes with deterministic components generated from Eqs. (18) and (19).

a. *If $b_1 \neq 0$ or $b_2 \neq 0$, and if the VECM does not include any deterministic trends, then $\hat{\lambda}_1$ does not converge to zero in probability.*

b. *If $b_1 = b_2 = 0$, $a_1 \neq 0$ or $a_2 \neq 0$, and if the VECM does not include any deterministic components, then $\hat{\lambda}_1$ does not converge to zero in probability.*

This proposition and Table 4 (Panels A and B) show that both cases (a) and (b) constitute a pitfall for the LR tests. In order to properly use the Johansen LR tests, the deterministic components have to be introduced in the VECM correctly. This result matches the conclusion of Perron and Campbell (1994). Note that EG is not sensitive to the wrong modeling of the deterministic trend components, producing an empirical size very close to the nominal size.

Moreover, to make the case more realistic and interesting, we consider a VECM with X_{t-1} augmented with 1 and/or t as considered in Johansen (1995, p. 80) by using $X_{t-1}^* = (X'_{t-1} \ 1)'$ or $X_{t-1}^* = (X'_{t-1} \ 1 \ t)'$. This model was developed to capture a cointegration relationship around a common deterministic trend (i.e., stochastic cointegration as well as deterministic cointegration). It shows that estimating the VECM with X_{t-1}^* when there is no (stochastic) cointegration will produce a pitfall. This happens because Eqs. (18) and (19) can be written as

$$\Delta X_t = \begin{pmatrix} 0 & 0 & a_1 & b_1 \\ 0 & 0 & a_2 & b_2 \end{pmatrix} (X'_{t-1} \ 1 \ t)' + \varepsilon_t \equiv \Pi^* X_{t-1}^* + \varepsilon_t$$

where the rank of Π^* is not zero, and therefore we may mistakenly interpret the result of the rank test as a rejection of the null hypothesis of no cointegration.

Table 4

$I(1)$ with deterministic components. DGP: $\Delta y_t = a_1 + b_1 t + e_{1t}$ and $\Delta x_t = a_2 + b_2 t + e_{2t}$

	EG	Q_1	Q_2
Panel A. When $a_1 = a_2 = 0$ and $b_1 = b_2 = 0.01$:			
SM ₁ : $\Delta X_t = \Pi_1 X_{t-1} + \varepsilon_t$	0.049	1.000	1.000
SM ₂ : $\Delta X_t = \mu + \Pi_2 X_{t-1} + \varepsilon_t$	0.067	0.622	0.582
SM ₃ : $\Delta X_t = \mu + \phi t + \Pi_3 X_{t-1} + \varepsilon_t$	0.051	0.018	0.040
SM ₄ : $\Delta X_t = \Pi_4 (X'_{t-1} \ 1) + \varepsilon_t$		1.000	1.000
SM ₅ : $\Delta X_t = \Pi_5 (X'_{t-1} \ 1 \ t) + \varepsilon_t$		1.000	1.000
Panel B. When $a_1 = a_2 = 1$ and $b_1 = b_2 = 0$:			
SM ₁ : $\Delta X_t = \Pi_1 X_{t-1} + \varepsilon_t$	0.074	1.000	1.000
SM ₂ : $\Delta X_t = \mu + \Pi_2 X_{t-1} + \varepsilon_t$	0.082	0.020	0.044
SM ₃ : $\Delta X_t = \mu + \phi t + \Pi_3 X_{t-1} + \varepsilon_t$	0.068	0.062	0.063
SM ₄ : $\Delta X_t = \Pi_4 (X'_{t-1} \ 1) + \varepsilon_t$		1.000	1.000
SM ₅ : $\Delta X_t = \Pi_5 (X'_{t-1} \ 1 \ t) + \varepsilon_t$		1.000	1.000

$T = 100$. SM_{*i*} ($i = 1, \dots, 5$) are the statistical models used for estimating the Johansen statistics, Q_1 and Q_2 . μ, ϕ are 2×1 ; Π_1, Π_2, Π_3 are 2×2 ; Π_4 is 2×3 ; and Π_5 is 2×4 . EG is computed from the cointegrating OLS regression without a constant (SM₁), with a constant (SM₂), and with both a constant and a trend (SM₃). The frequency of rejecting the null hypothesis in 1,000 replications is reported at the 5% level. The critical values for each of three regression models are simulated from 90,000 replications using the DGP with $a_1 = a_2 = 0$ and $b_1 = b_2 = 0$.

Therefore it is crucial to know whether the deterministic components should be modelled inside or outside the cointegrating relationship.

This may be a pitfall apparently easy to be detected, and once detected, easy to be eliminated. Nevertheless, we want to emphasize how sensitive the LR tests are to a possible mistake in the way how deterministic components are treated. The main point of this pitfall is double, first to show how important it is to deal correctly with deterministic components, especially to know whether these components should be modelled outside the long run equilibrium or inside, and second to note that EG is more robust to this kind of misspecification than the LR tests.

4.5. Processes with fractional unit roots

The literature on fractionally integrated processes has grown rapidly since the early contributions of Granger and Joyeux (1980) and Hosking (1981). The model has been used in applied econometric work where flexible characterization of low frequency dynamics is important, and has yielded new empirical regularities and insights into understanding behavior of many economic time series. There is considerable evidence on long memory properties in macroeconomic and financial time series data. Examples includes real output dynamics (Sowell, 1992), disposable income (Diebold and Rudebusch, 1991), stock prices (Lo, 1991), interest rates (Shea, 1991), exchange rates (Cheung, 1993), and the purchasing power parity hypothesis (Cheung and Lai, 1993).

Suppose $X_t = (y_t, x_t)'$ are generated from

$$(1 - B)^d y_t = e_{1t}, \tag{20}$$

$$(1 - B)^d x_t = e_{2t}. \tag{21}$$

When d is not an integer, the series is said to be fractionally integrated. An $I(d)$ process y_t generated from Eq. (20) has the following properties: (a) If $d < 0.5$, y_t is covariance stationary. (b) If $d > -0.5$, y_t has an invertible moving average representation. (c) If $d < 1$, y_t is mean-reverting. (d) If $d > 0$, y_t has long memory, the autocovariances of y_t are not absolutely summable, and the power spectrum of y_t is unbounded for frequencies approaching zero. (e) If $d > 0.5$, y_t has an infinite variance. (f) The DF- t statistic diverges to $-\infty$ if $d < 1$, and diverges to $+\infty$ if $d > 1$ as $T \rightarrow \infty$ (Sowell, 1990).

The fractional difference operator $(1 - B)^d$ defined by its Maclaurin series is

$$(1 - B)^d = \sum_{j=0}^{\infty} \frac{\Gamma(-d + j)}{\Gamma(-d) \Gamma(j + 1)} B^j = \sum_{j=0}^{\infty} d_j B^j,$$

$$d_j = \frac{j - 1 - d}{j} d_{j-1}, \quad d_0 = 1, \tag{22}$$

where $\Gamma(\cdot)$ is the gamma function, and d_j are squared summable if $d > -0.5$. Since $(1 - B)^{d-1}\Delta y_t = e_{1t}$, Δy_t can be written as

$$\Delta y_t = \sum_{j=1}^{\infty} a_j \Delta y_{t-j} + e_{1t}, \tag{23}$$

where $a_j = -\Gamma(j - d + 1)/[\Gamma(j + 1)\Gamma(-d + 1)]$. The VECM is then

$$\begin{aligned} \begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} \\ &+ \sum_{j=1}^{\infty} \begin{pmatrix} a_j & 0 \\ 0 & a_j \end{pmatrix} \begin{pmatrix} \Delta y_{t-j} \\ \Delta x_{t-j} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}. \end{aligned} \tag{24}$$

For simplicity, in the following proposition, we consider only the case with $k = 0$ in the VECM to compute the LR statistics. Any finite k that is not sufficiently large enough to make the error ε_t a vector white noise will lead to the same results.

Proposition 7. Suppose $(y_t, x_t)'$ are $I(d)$ processes generated from Eqs. (20) and (21), and we estimate a VECM with $k = 0$.

- a. *If $d \geq 1.5$, then $\hat{\lambda}_1$ does not converge to zero in probability as $T \rightarrow \infty$.*
- b. *If $1 < d < 1.5$, then $T^{(3-2d)}\hat{\lambda}_1 = O_p(1)$.*
- c. *If $d = 1$, then $T\hat{\lambda}_1 = O_p(1)$.*
- d. *If $0.5 < d < 1$, then $T^{(2d-1)}\hat{\lambda}_1 = O_p(1)$.*

If $d = 1$, $T\hat{\lambda}_1 = O_p(1)$. If $d \neq 1$, $T\hat{\lambda}_1 \xrightarrow{p} \infty$ as $T \rightarrow \infty$. Although the case with $1 < d < 1.5$ produces a pitfall, only the case when $d \geq 1.5$ yields a pitfall of type A. It is true that a sufficiently large k such that the residuals are white noise will solve the pitfall. But there are many situations in macroeconomics where it is not possible in practice to try a large k . For instance, in large systems like the Summers and Heston (1991) data set it is often practically impossible to estimate a system with k larger than one. As remarked already, Proposition 7 will hold not only for $k = 0$ but also for any $k > 0$ not sufficiently large to make the error a vector white noise.

In Table 5, the size of the cointegration tests is simulated for various values of d in the DGP. To simulate the series it is assumed that $d_j = 0$ for $j > 1,000$ and the first 2,000 observations are discarded.

When $d < 1$, the size is large for both EG and the Johansen tests. Although Sowell (1990) shows that the DF t statistic diverges to $-\infty$ if $d < 1$ as $T \rightarrow \infty$, it is well known (see Gonzalo and Lee, 1996a) that the power of the DF test is very small against fractional alternatives in finite samples. Therefore if the variables are fractionally integrated with $d < 1$, it is likely that we proceed

Table 5
 Fractionally integrated processes. DGP: $(1 - B)^d y_t = e_{1t}$ and $(1 - B)^d x_t = e_{2t}$

d	$k = 0$				$k = 3$				$k = 9$			
	EG	Q_1	Q_2		EG	Q_1	Q_2		EG	Q_1	Q_2	
	0.5	0.986	1.000	0.997		0.361	0.651	0.432		0.083	0.233	0.171
0.6	0.900	0.991	0.947		0.223	0.406	0.262		0.060	0.180	0.141	
0.7	0.636	0.832	0.652		0.146	0.224	0.167		0.043	0.152	0.132	
0.8	0.341	0.447	0.340		0.094	0.142	0.105		0.042	0.129	0.129	
0.9	0.143	0.147	0.132		0.065	0.087	0.076		0.038	0.132	0.122	
1.0	0.047	0.048	0.052		0.048	0.068	0.072		0.032	0.133	0.130	
1.1	0.016	0.058	0.069		0.037	0.077	0.078		0.034	0.153	0.149	
1.2	0.010	0.152	0.166		0.028	0.090	0.085		0.036	0.163	0.165	
1.3	0.031	0.337	0.351		0.021	0.105	0.113		0.030	0.171	0.171	
1.4	0.053	0.563	0.592		0.024	0.110	0.128		0.027	0.188	0.184	
1.5	0.076	0.753	0.774		0.019	0.135	0.144		0.032	0.221	0.206	
1.6	0.100	0.860	0.873		0.016	0.165	0.185		0.037	0.236	0.230	
1.7	0.111	0.921	0.931		0.012	0.185	0.196		0.048	0.274	0.267	
1.8	0.135	0.952	0.957		0.016	0.191	0.213		0.045	0.312	0.294	
1.9	0.150	0.969	0.973		0.029	0.231	0.243		0.053	0.357	0.336	
2.0	0.176	0.982	0.981		0.039	0.271	0.254		0.054	0.400	0.385	

$T = 100$. The frequency of rejecting the null hypothesis in 1,000 replications is reported at the 5% level. The critical values are simulated from 90,000 replications using the DGP with $d = 1$.

assuming the series are $I(1)$ and get the incorrect conclusion that the system has a long-run relationship.

When $d > 1$, EG is more robust than the LR tests. Sowell, 1990) shows that if $d > 1$, the DF test has zero power asymptotically to test for a unit root against a stationary root. Thus if the variables are fractionally integrated with $d > 1$, it is likely that we proceed assuming they are $I(1)$ (see Gonzalo and Lee, 1996a) and have a pitfall.

Since the true number of lagged differences in the VECM (24) is infinity, one may want to increase k with T , in the way similar to Berk (1974) for stationary and ergodic processes. However, we are not aware of any result in the literature on how to do this for nonstationary and non-ergodic processes. We suspect the problem must be complicated because a fractionally integrated process ($d > 0.5$) is not ergodic. The sum of absolute correlations is not bounded, therefore any finite k will produce inconsistent estimates. In any case, what we have done in the Monte Carlo exercise is to increase k up to 9. We did the same experiment as in Table 5 for $T = 1,000$ (not reported but available upon request), and it is found that we still keep getting a pitfall. Based on these Monte Carlo experiments we have to agree with Brockwell and Davis (1991), (p. 520) when they say “While a long memory process can always be approximated by an ARMA(p,q) the orders p and q required to achieve a reasonably good approximation may be so large as to make parameter estimation extremely difficult.”

How often do we have $d > 1$ in practice? Examples of values of estimates of d reported in the literature are: $d = 1.17$ for annual disposable income (Diebold and Rudebusch, 1991), $d = 1.29$ for quarterly real GNP (Sowell, 1992), and d is ranged from 1.04 to 1.36 for various nominal spot exchange rates (Cheung, 1993). Also d is estimated about 0.6 for money growth rates (Tieslau, 1991) and is ranged from 0.40 to 0.57 for inflation rates in several developed countries (Hassler and Wolters, 1995), indicating money stock and price series may have d greater than 1.5.

Another important and related question is to see how precisely d can be estimated with the sample sizes used in applied studies. Simulation evidence on the finite sample performance of the standard estimators of d (Agiakloglou et al., 1992; Cheung and Diebold, 1994; Chung and Baillie, 1993) shows severe biases of these estimators. This difficulty of estimating d gives even more relevance of a pitfall in the fractional difference case.

4.6. Summary

Section 4 analyzes five situations corresponding to variables with different trending behavior than the one of $I(1)$ processes. These variables are not cointegrated but nevertheless the Johansen LR tests produce spurious cointegration with probability approaching one if the VECM (2) is estimated.

The main reason is that the matrices S_{00} and $\hat{\Omega}$ do not converge to the same limit in probability as they do when the variables are $I(1)$ and not cointegrated. At the same time the EG test shows a more robust behavior than the LR tests.

Both tests behave very similar under the null of no pitfall but totally different under the alternative of a pitfall. Therefore a pitfall could be tested with a Hausman type test. Because this new test will have a new non-standard distribution, we just recommend a simpler approach consisting of applying both tests. If they give different outcomes, then the variables should be tested for different trending processes and the VECM (2) should be modified accordingly if possible.

In addition to these five situations investigated, a pitfall may occur to variables with other trending behavior like variables with exponential trends, higher order polynomial deterministic trends, or with structural breaks in mean or/and trend, which make it difficult to tell from an $I(1)$ process. These issues are under current investigation.

5. Type-B pitfalls

This section investigates how cointegration tests are affected by the singularity or near singularity of the covariance matrix of the VAR model. It shows when any of the two analyzed tests have a size or a power pitfall.

5.1. Perfect and near singularity of Ω

A large number of econometric studies have focused on the estimation of parameters in singular systems (Barten, 1969; Berndt and Savin, 1975; Anderson and Blundell, 1982; Dhrymes, 1994; Adda and Ng, 1995). Systems with singular covariance matrices arise in models explaining the allocation of consumer expenditure amongst its components. They also arise in the context of derived systems of equations from production functions, market share analysis, demand for financial assets, and in many other models. In this literature the standard way to avoid singularity is to eliminate one of the equations and estimate the $(n - 1)$ (non-singular) system. According to Dhrymes (1994) the results are invariant to the equation deleted only when all the explanatory variables appear in every equation, and when there are no autoregressive dynamics and no-cross equation restrictions. These are conditions that do not hold in a cointegrated system like (2).

In general there does not exist satisfactory solution to the problem of singularity in the literature. In this paper, in order to test for the number of cointegration vectors in a singular system, we propose generalized versions of Johansen LR tests. There are other alternatives like the use of a ridge-type adjustment for singularity but we have not investigated them.

Definition 3. The generalized versions of the Johansen LR (GLR) tests for the null hypothesis $H_0:r = 0$ are

$$Q_1^* = -T \ln(1 - \hat{\lambda}_1^*)(1 - \hat{\lambda}_2^*), \tag{25}$$

$$Q_2^* = -T \ln(1 - \hat{\lambda}_1^*), \tag{26}$$

where $\hat{\lambda}_1^* \geq \hat{\lambda}_2^*$ are the eigenvalues of $S_{00}^+S_{01}S_{11}^+S_{10}$ with S_{ii}^+ being the Moore–Penrose generalized inverse of S_{ii} , $i = 0,1$.

Note that the distributions of Q_1^* and Q_2^* will be different than those of Q_1 and Q_2 but it will not be shown here because it is not relevant for our analysis. The generalized eigenvalues $\hat{\lambda}_1^*$ and $\hat{\lambda}_2^*$ are easily obtained by estimating by OLS the system (2) equation by equation as the next proposition shows.

Proposition 8. Let $\hat{M}^* = S_{00}^+S_{01}S_{11}^+S_{10}$. Then $\hat{M}^* = S_{00}^+S_{00} - S_{00}^+\hat{\Omega}^*$, where $\hat{\Omega}^* = S_{00} - S_{01}S_{11}^+S_{10}$.

The singularity of the covariance matrix implies the existence of a linear combination of X_t that is completely predictable from past values of X_t . We name that exact cointegration.

Definition 4. A set of $I(1)$ variables X_t is exactly cointegrated when the covariance matrix of the error terms of the MA representation of Δx_t is singular.

In this section we focus on the performance of the cointegration tests when there is no standard cointegration, because it can be shown that in a cointegrated system $(y_t \ x_t)'$, the singularity of Ω does not change the cointegrating structure unless $\gamma_{\perp}\Omega = 0$ where $\gamma_{\perp}\gamma = 0$ and γ is defined in Eq. (4).

Let $X_t = (y_t \ x_t)'$ be generated from $\Delta y_t = e_{1t}$ and $\Delta x_t = e_{2t}$. Suppose $\theta = \text{corr}(e_{1t}, e_{2t}) = 1$. Then $e_{1t} = e_{2t}$, $y_t = x_t$, $\det(\Omega) = 0$, and $(y_t \ x_t)'$ are exact-cointegrated. However, in this case, the GLR tests are not able to capture this kind of cointegration. This happens because $S_{00} - \hat{\Omega}^* \xrightarrow{P} 0$ as $T \rightarrow \infty$. Therefore $\hat{M}^* \xrightarrow{P} 0$ and both $\hat{\lambda}_1^*, \hat{\lambda}_2^* \xrightarrow{P} 0$. The GLR test can not reject the null of no cointegration, and has a power pitfall. Singularity in a bivariate system may be a trivial case. But in some applied research, singularity is likely to happen in large systems, and it could be difficult to detect it.

When $\theta \neq 1$, both tests have a correct size as it is shown in Table 6 and this is due to the invariant property that both tests have to the covariance parameters. This will change if we introduce some noise like in the next section.

5.2. Near singular Ω with misspecified dynamics

Near singularity affects dramatically the outcome of the cointegration tests when it goes together with some dynamic misspecification. To show that, we

analyze the following DGP:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \phi \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \tag{27}$$

where $|\phi| < 1$ and $e_{1t} = \theta e_{2t} + \sqrt{1 - \theta^2} v_t$. Then

$$y_t - \theta(1 - \phi)x_t = \sqrt{1 - \theta^2} \Delta^{-1} v_t + \frac{\theta\phi}{1 - \phi B} e_{2t} \tag{28}$$

and therefore $(y_t \ x_t)'$ is cointegrated only if $\theta = 1$.

In Table 6 when $\phi = 0$, we find that the size of tests does not change with θ at all, and the empirical size is close to the nominal size no matter how close θ is to one if not one. Now, let us examine what happens if $\phi \neq 0$ but we mistakenly specify the VECM (27) with $k = 0$. In Table 7, we report the empirical size from

Table 6
Near singularity

	EG	Q_1	Q_2
$\theta = 0$	0.053	0.046	0.053
$\theta = 0.9$	0.053	0.046	0.053
$\theta = 0.99$	0.053	0.046	0.053
$\theta = 0.999$	0.053	0.046	0.053
$\theta = 0.99999$	0.053	0.046	0.053
$\theta = 1.0$	1.000	0.000	0.000

$T = 100$. The frequency of rejecting the null hypothesis in 1,000 replications is reported at the 5% level. The critical values are simulated from 90,000 replications using the DGP with $\theta = 0$.

Table 7
Near singularity with dynamic misspecification. DGP:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \phi \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

	$k = 0$			$k = 1$		
	EG	Q_1	Q_2	EG	Q_1	Q_2
$\theta = 0$	0.077	0.071	0.070	0.047	0.046	0.044
$\theta = 0.9$	0.153	0.179	0.188	0.049	0.047	0.046
$\theta = 0.99$	0.598	0.700	0.702	0.197	0.047	0.046
$\theta = 0.999$	0.979	0.990	0.990	0.808	0.048	0.050
$\theta = 1.0$	1.000	1.000	1.000	1.000	0.000	0.000

$\phi = -0.2$. $T = 100$. The frequency of rejecting the null hypothesis in 1,000 replications is reported at the 5% level. The critical values are simulated from 90,000 replications using the DGP with $\phi = 0$.

simulations with $\phi = -0.2$. When $k = 0$ (dynamic misspecification) both tests show very serious size distortions. These distortions get worse as θ gets larger. When $k = 1$ (no dynamic misspecification) the behavior of Johansen tests is similar to the one in Table 6 (power pitfall but no size distortions). However, EG does not have the correct size. This is because EG is based on the unit root test on the residual from the static cointegrating regression and the residual has far longer memory than one lag ($k = 1$) as it can be seen from Eq. (28).

The main point of this subsection is to show that near singularity makes the testing results extremely sensitive to the number of lags of the VECM. When this happens in real practice we suggest to analyze the eigenvalues of the covariance matrix of the residuals, something that it is not normally done in empirical research.

6. Conclusions

Standard cointegration analysis consists mainly of two econometric steps: first, univariate unit root tests and second, an LR test for cointegration. These two steps are done almost automatically by many of the menu driven statistical packages available these days. This paper studies the robustness of cointegration tests to departures from $I(1)$, and shows that pre-testing for individual unit roots is not enough. We have to be sure that the variables do not have any other trending or long-memory behavior different from that of a unit root process. We also need to be sure that the only singularity in the VAR comes from the long-run impact matrix and not from the error covariance matrix. When any of the these two conditions are not satisfied, the paper shows that Johansen LR tests for cointegration have a pitfall tending to give spurious cointegration with probability approaching one asymptotically. On the contrary, in most of the cases investigated in this paper, the EG test is more robust.

The paper concludes recommending the use of both tests, Johansen LR test and EG test, in order to detect the possibility of a pitfall and therefore increasing the chance of avoiding it by modeling the VECM (2) in a different way if it is possible.

Acknowledgements

We thank two anonymous referees, an associate editor, Graham Elliott, Rob Engle, Gloria Gonzalez-Rivera, Clive Granger, Cheng Hsiao, Jim Stock, Norm Swanson, Hal White, and participants of seminars at Boston University, Pennsylvania State University, Universidad Carlos III de Madrid, Universidad Complutense de Madrid, Universidad del Pais Vasco, Groningen University, University of Montreal, UCLA, UC Riverside, UCSD, USC, the US Bureau of

Labor Statistics, the 1995 Winter Econometric Society Meeting at Washington DC, and the 7th World Congress of Econometric Society at Tokyo for comments and suggestions. All errors are our own. The first author gratefully acknowledges financial support from the Spanish Secretary of Education (PB 950298).

Appendix A.

Proof of Proposition 1. The eigenvalues of $S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$ are the same as those of $S_{00}^{-1}S_{01}S_{11}^{-1}S_{10}$. The result follows from $\hat{M} = S_{00}^{-1}S_{01}S_{11}^{-1}S_{10}$ and $\hat{\Omega} = S_{00} - S_{01}S_{11}^{-1}S_{10}$. \square

Proof of Proposition 2. If $\hat{\lambda}_1$ does not converge to zero, then $T\hat{\lambda}_1$ diverges. Since $T\hat{\lambda}_1 \leq Q_2 \leq Q_1, Q_1$ and Q_2 diverge. From Proposition 1, $\det(I - \hat{M}) = \det(\hat{\Omega})/\det(S_{00})$. Under each condition, $\det(I - \hat{M}) \xrightarrow{p} 0$. Then \hat{M} does not converge to 0. \square

Proof of Proposition 3. a. $S_{11} = O_p(T), S_{10} = O_p(1),$ and $S_{00} = O_p(1)$.

b. $M \equiv I - \Sigma_{00}^{-1}\Omega,$ where Ω is defined in (3).

$$\Sigma_{00} = \text{var}(\Delta X_t) = \begin{pmatrix} \text{var}(\Delta y_t) & \theta\sigma_1\sigma_2 \\ \theta\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

with

$$\text{var}(\Delta y_t) = \frac{a - 1}{a + 1}(a^{2t} - 1)\sigma_1^2 + \sigma_1^2.$$

Therefore

$$\lambda_1 = \frac{a - 1}{(a + 1)\theta^2 - 2} \quad \text{if } |a| \leq 1.$$

c. Since $\Delta y_t = (a - 1)y_{t-1} + e_{1t}$ and $a - 1 \neq 0, \det(S_{00}) = O_p(T^\delta), \delta > 0$. Using the results of Anderson (1959), it can be shown that $\hat{\Omega} \xrightarrow{p} \Omega$. From Proposition 1, $\det(I - \hat{M}) = \det(\hat{\Omega})/\det(S_{00}),$ and the result follows. \square

Proof of Proposition 4. Eq. (12) can be solved as

$$y_t = e_{1t} + a_t e_{1,t-1} + a_t a_{t-1} e_{1,t-2} + \dots + (a_t a_{t-1} \dots a_{t-k+2}) e_{1,t-k+1} + a_{t-k+1} y_{t-k}$$

or

$$y_t = e_{1t} + \pi_{t,1} e_{1,t-1} + \pi_{t,2} e_{1,t-2} + \dots + \pi_{t,k-1} e_{1,t-k+1} + \pi_{t,k} y_{t-k}$$

for any integer k , $0 \leq k \leq t$, where $\pi_{t,j} \equiv \exp(S_t(j))$, $S_t(j) \equiv \sum_{i=0}^{j-1} \alpha_{t-i}$, and $S_t(0) \equiv 0$. Then

$$\begin{aligned} \Delta y_t &= e_{1t} + \sum_{j=1}^{k-1} (\pi_{t,j} - \pi_{t-1,j-1})e_{1,t-j} - \pi_{t-1,k-1}e_{1,t-k} \\ &\quad + \pi_{t,k}y_{t-k} - \pi_{t-1,k}y_{t-k-1}. \end{aligned} \tag{A.1}$$

As $S_t(j)$ follows a log normal distribution, $E(\pi_{t,k}) \simeq \exp(k\varphi)$, for large k , where $\varphi \equiv m + 0.5f_\alpha(0)$. Taking expectation of Eq. (A.1) with $k = t - 1$ gives

$$E(\Delta y_t) = (y_1 - y_0)\exp((t - 1)\varphi)$$

and because ε_t is i.i.d.,

$$\begin{aligned} \text{var}(\Delta y_t) &= \sigma_1^2 + \sigma_1^2 \sum_{j=1}^{t-2} E(\pi_{t,j} - \pi_{t-1,j-1})^2 + \sigma_1^2 E\pi_{t-1,k-1}^2 \\ &\quad + E[\pi_{t,t-1}y_1 - \pi_{t-1,t-1}y_0 - E(\Delta y_t)]^2, \end{aligned}$$

where

$$\begin{aligned} \sum_{j=1}^{t-2} E(\pi_{t,j} - \pi_{t-1,j-1})^2 &= E(a_t - 1)^2 \sum_{j=1}^{t-2} \pi_{t-1,j-1}^2 \\ &= 2E(a_t - 1)^2 \sum_{j=1}^{t-2} \exp(S_{t-1}(j - 1)). \end{aligned}$$

The last expression increases with t because $\exp(S_{t-1}(j - 1)) > 0$ for all j . Thus $\text{var}(\Delta y_t)$ also increases with t . Hence, $\det(S_{00}) = O_p(T^\delta)$, $\delta > 0$. The second part follows from noticing that $\hat{\Omega}$ is bounded and $\det(I - \hat{M}) = \det(\hat{\Omega})/\det(S_{00})$. \square

Proof of Proposition 5. a. Following *Gourieroux et al. (1989)*, *Johansen (1995)*, and *Haldrup (1994)*, when $k = 0$, $S_{11} = O_p(T^3)$, $S_{00} = O_p(T)$, and $S_{10} = O_p(T^2)$. Therefore $\hat{M} = S_{00}^{-1}S_{01}S_{11}^{-1}S_{10} = O_p(1)$, and since M is not degenerate, the result follows from noting that $\hat{\lambda}_1$ is a continuous function of \hat{M} .

b. When $k \geq 1$, $S_{11} = O_p(T^3)$, $S_{00} = O_p(1)$, and $S_{10} = O_p(T)$. Therefore $\hat{M} = O_p(T^{-1})$ and the result follows. \square

Proof of Proposition 6. For simplicity let us assume that in both cases the ECM does not include any deterministic element. The next results follow from a simple application of $\sum_{t=1}^T t^r = O(T^{r+1})$, $r = 0, 1, 2, \dots$ (see *Anderson, 1971*, p. 83).

a. It can be shown that $T^{-4}S_{11}$, $T^{-2}S_{00}$, and $T^{-3}S_{10}$ converge in probability to constant matrices. Therefore $\hat{M} = S_{00}^{-1}S_{01}S_{11}^{-1}S_{10}$ converges to a non-zero constant matrix too, and the result follows.

b. It can be shown that $T^{-2}S_{11}$, S_{00} , and $T^{-1}S_{10}$ converge in probability to constant matrices. Therefore \hat{M} converges to a non-zero constant matrix too, and the result follows. \square

Proof of Proposition 7. a. Following Gourieroux et al. (1989), for $d \geq 1.5$, $S_{11} = O_p(T^{2d-1})$, $S_{00} = O_p(T^{2d-3})$, and $S_{10} = O_p(T^{2d-2})$. Therefore $\hat{M} = O_p(1)$ and the result follows.

b. For $1 < d < 1.5$, $S_{11} = O_p(T^{2d-1})$, $S_{00} = O_p(1)$, and $S_{10} = O_p(T^{2d-2})$. Therefore $\hat{M} = O_p(T^{2d-3})$ and the result follows.

c. For $d = 1$, $S_{11} = O_p(T)$, $S_{00} = O_p(1)$, and $S_{10} = O_p(1)$. Therefore $\hat{M} = O_p(T^{-1})$ and the result follows.

d. For $0.5 < d < 1$, $S_{11} = O_p(T^{2d-1})$, $S_{00} = O_p(1)$, and $S_{10} = O_p(1)$. Therefore $\hat{M} = O_p(T^{1-2d})$ and the result follows. \square

Proof of Proposition 8. Similar to Proof of Proposition 1. \square

References

- Adda, J., Ng, S., 1995. Demand systems in growing economies. Boston University.
- Agiakloglou, C., Newbold, P., Wohar, M., 1992. Bias in an estimator of the fractional difference parameter. *Journal of Time Series Analysis* 14, 235–246.
- Ahn, S.K., Reinsel, G.C., 1990. Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association* 85, 813–823.
- Anderson, G.J., Blundell, R.W., 1982. Estimation and hypothesis testing in dynamic singular equation systems. *Econometrica* 50, 1559–1571.
- Anderson, T.W., 1959. On the asymptotic distribution of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30, 676–687.
- Anderson, T.W., 1971. *The Statistical Analysis of Time Series*. Wiley, New York.
- Barten, A.P., 1969. Maximum likelihood estimation of a complete system of demand equations. *European Economic Review* 1, 7–73.
- Berk, K., 1974. Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489–502.
- Berndt, E.R., Savin, E.N., 1975. Estimation and hypothesis testing in singular equation systems with autoregressive disturbances. *Econometrica* 43, 937–957.
- Brockwell, P., Davis, R.A., 1991. *Time Series: Theory and Methods*, 2nd ed. Springer, Berlin.
- Cavanagh, C.L., Elliott, G., Stock, J.H., 1995. Inference in models with nearly integrated regressors. *Econometric Theory* 11, 1131–1147.
- Cheung, Y.W., 1993. Long memory in foreign exchange rates. *Journal of Business and Economic Statistics* 11, 93–101.
- Cheung, Y.W., Diebold, F.X., 1994. On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. *Journal of Econometrics* 62, 301–316.
- Cheung, Y.W., Lai, K.S., 1993. A fractional cointegration analysis of purchasing power parity. *Journal of Business and Economic Statistics* 11, 103–112.
- Chung, C.F., Baillie, R.T., 1993. Small sample bias in conditional sum of squares estimators of fractionally integrated ARMA models. *Empirical Economics* 18, 791–806.
- Dhrymes, P.J., 1994. Autoregressive errors in singular systems of equations. *Econometric Theory* 10, 254–285.
- Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of American Statistical Association* 74, 427–431.

- Diebold, F.X., Rudebusch, G.D., 1991. Is consumption too smooth? Long memory and the Deaton paradox. *Review of Economics and Statistics* 71, 1–9.
- Dolado, J.J., Marmol, F., 1996. Efficient estimation of cointegrating relationships among higher order and fractionally integrated processes. *Journal of Econometrics*, forthcoming.
- Elliott, G., 1995. On the robustness of cointegration methods when regressors almost have unit roots. UCSD.
- Engle, R.F., Granger, C.W.J., 1987. Cointegration and error correction: representation, estimation, and testing. *Econometrica* 55, 251–276.
- Gonzalo, J., Lee, T.H., 1996a. Pitfalls for testing in long run relationships. Universidad Carlos III de Madrid.
- Gonzalo, J., Lee, T.H., 1996b. Relative power of t type tests for stationary and unit root processes. *Journal of Time Series Analysis* 17, 37–47.
- Gonzalo, J., Pitarakis, J.Y., 1996. Lag length estimation in large dimensional systems. Universidad Carlos III de Madrid.
- Gourieroux, C., Maurel, F., Monfort, A., 1989. Least squares and fractionally integrated regressors. INSEE, Working Paper No. 8913.
- Granger, C.W.J., 1993. What are we learning about the long-run?. *Economic Journal* 103, 307–317.
- Granger, C.W.J., Joyeux, R., 1980. An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–29.
- Granger, C.W.J., Newbold, P., 1974. Spurious regressions in econometrics. *Journal of Econometrics* 2, 111–120.
- Granger, C.W.J., Swanson, N.R., 1996. An introduction to stochastic unit root processes. *Journal of Econometrics*, forthcoming.
- Haldrup, N., 1994. The asymptotics of single equation cointegration regressions with $I(1)$ and $I(2)$ variables. *Journal of Econometrics* 63, 153–181.
- Hassler, U., Wolters, J., 1995. Long memory in inflation rates: international evidence. *Journal of Business and Economic Statistics* 13, 37–45.
- Hausman, J.A., 1978. Specification tests in econometrics. *Econometrica* 46, 1251–1272.
- Hosking, J.R.M., 1981. Fractional differencing. *Biometrika* 68, 165–176.
- Johansen, S., 1995. Likelihood-based Inference in Cointegrated Vector Auto-regressive Models. Oxford University Press, Oxford.
- Lo, A.W., 1991. Long term memory in stock market prices. *Econometrica* 59, 1279–1313.
- Nelson, C.R., Plosser, C.I., 1982. Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics* 10, 129–162.
- Paruolo, P., 1996. On the determination of integration indices in $I(2)$ systems. *Journal of Econometrics* 72, 313–356.
- Perron, P., Campbell, J.Y., 1994. A note on Johansen's cointegration procedure when trends are present. In: Dufour, J.M., Raj, B. (Eds.), *New Developments in Time Series Econometrics*. Physica-Verlag, Berlin.
- Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with $I(1)$ processes. *Review of Economic Studies* 57, 282–306.
- Schwert, G.W., 1989. Tests for unit root: a Monte Carlo investigation. *Journal of Business and Economic Statistics* 7, 147–159.
- Shea, G.S., 1991. Uncertainty and implied variance bounds in long memory models of the interest rate term structure. *Empirical Economics* 16, 287–312.
- Sowell, F.B., 1990. The fractional unit root distribution. *Econometrica* 58, 495–505.
- Sowell, F.B., 1992. Modeling long run behavior with the fractional ARIMA model. *Journal of Monetary Economics* 29, 277–302.
- Stock, J.H., 1991. Confidence intervals for the largest autoregressive root in US macroeconomic time series. *Journal of Monetary Economics* 28, 435–459.

- Stock, J.H., 1995. Unit roots, structural breaks and trends. In: Engle, R.F., McFadden, D. (Eds.), *Handbook of Econometrics*, vol. 4. North-Holland, Amsterdam.
- Stock, J.H., Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.
- Summers, R., Heston, A., 1991. The Penn world table (Mark 5): an expanded set of international comparisons, 1950–1988. *Quarterly Journal of Economics*, 327–368.
- Tieslau, M.A., 1991. Long memory models and macroeconomic time series. Michigan State University, Working Paper, No. 9005.
- Tjøstheim, D., 1986. Some doubly stochastic time series models. *Journal of Time Series Analysis* 7, 51–72.
- Toda, H.Y., 1995. Finite sample performance of likelihood tests for cointegrating ranks in vector autoregression. *Econometric Theory* 11, 984–1014.
- Watson, M.W., 1995. Vector autoregressions and cointegration. In: Engle, R.F., McFadden, D. (Eds.), *Handbook of Econometrics*, vol. 4. North-Holland, Amsterdam.
- Yule, G.U., 1926. Why do we sometimes get nonsense correlations between time series? A study in sampling and the nature of time series?. *Journal of the Royal Statistical Society* 89, 1–64.