

# Five alternative methods of estimating long-run equilibrium relationships

Jesus Gonzalo\*

*Boston University, Boston, MA 02215, USA*

Received November 1989, final version received July 1992

This paper compares several methods (ordinary least squares, nonlinear least squares, maximum likelihood in an error correction model, principal components, and canonical correlations) of estimating cointegrating vectors. Although all of them are superconsistent, an empirical example shows that the estimates can vary significantly. The paper examines the asymptotic distribution of the estimators resulting from these methods, and shows that maximum likelihood in a fully specified error correction model (Johansen's approach) has clearly better properties than the other estimators. A Monte Carlo study indicates that finite sample properties are consistent with the asymptotic results. This is so even when the errors are non-Gaussian or when the dynamics are unknown.

*Key words:* Canonical correlations; Cointegrating vectors; Error correction models; Maximum likelihood estimators

## 1. Introduction

Several methods to estimate cointegrating vectors (long-run equilibrium relationships) have been proposed in the literature since Granger (1983) introduced the idea of cointegration. Chronologically they are: ordinary least squares (OLS) by Engle and Granger (1987), nonlinear least squares (NLS) by Stock (1987), principal components (PC) by Stock and Watson (1988), canonical correlations (CC) by Bossaerts (1988), maximum likelihood in a fully specified

*Correspondence to:* Jesus Gonzalo, Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA.

\*I am grateful to my advisor Clive Granger and to Robert Engle and Soren Johansen for many useful conversations and suggestions. I also wish to thank Juan Dolado, Devajyoti Ghose, David Hendry, Miguel Herce, Svend Hylleberg, two anonymous referees, and the participants in the Econometric Workshops at the Universities of Boston, California (San Diego), California (Los Angeles), Harvard, Toronto, and Vanderbilt. An earlier version of this paper was presented at the NSF–NBER Seminar on Time Series, San Diego, October 1990. I thank the Bank of Spain and the Sloan Foundation for their financial support. I assume responsibility for all remaining errors.

error correction model (MLECM) by Johansen (1988),<sup>1</sup> instrumental variables (IV) by Hansen and Phillips (1990), and spectral regression (SR) by Phillips (1991).<sup>2</sup> The goal of this paper is to compare the behavior of the first five methods (the most-used in empirical research) asymptotically as well as in finite samples via a Monte Carlo study.<sup>3</sup>

The conclusions obtained agree with the theoretical results in Phillips (1991) who shows that the best way to proceed in the estimation of cointegrated systems is full system estimation by maximum likelihood, incorporating all prior knowledge about the presence of unit roots. This approach ensures that coefficient estimates are symmetrically distributed, median unbiased, asymptotically efficient, and that hypothesis tests may be conducted using standard asymptotic chi-squared tests. The simplest procedure is to estimate a fully specified error correction model (ECM) by maximum likelihood (ML). This is exactly the method proposed by Johansen and the one which performs better in our Monte Carlo experiment, even when the errors are nonnormal distributed or when the dynamics are unknown.

Since the estimators resulting from the above methods are superconsistent (the rate of convergence is  $T$  instead of  $T^{1/2}$ ), it is assumed that there should not be a big difference among their estimates with real data if the number of observations is not too small. That is not the case with the example in table 1, which presents different estimates (based on these five methods) of the cointegrating vector between short-term and long-term interest rates. Without arguing whether the cointegrating vector should be one, as the expectation theory of the term structure suggests, or greater than one, indicating an overreaction of the long-term interest rates, the fact is that the estimates of the cointegrating vector can vary significantly depending on the method we use to estimate it. For instance, using Johansen's method [MLECM(12)] with twelve lags in the ECM, we can easily reject the null hypothesis that the cointegrating vector belongs to the space spanned by  $(1, -1)$  or  $(1, -\hat{\alpha}_{ols})$  when the long-term interest rate has a maturity longer than one year. Notice also that in the methods based on dynamic regressions we get different estimates depending on the number of lags we choose. This is discussed in section 3.1.

This paper is organized as follows. Section 2 describes the five estimators and derives and compares their asymptotic distributions. The asymptotic

<sup>1</sup>The asymptotic distribution of the Gaussian MLE in a vector error correction model was independently derived by Johansen (1988) and Ahn and Reinsel (1991).

<sup>2</sup>Since I last submitted this paper, two more methods to estimate cointegrating vectors have been proposed. One is the three-step estimator by Engel and Yoo (1989) and the other a single-equation error correction model with leads and lags proposed by Phillips and Loretan (1991), Saikkonen (1991), and Stock and Watson (1988). These new estimators are asymptotically equivalent to MLECM.

<sup>3</sup>The asymptotic distribution of PC and CC are derived in this paper for the first time.

Table 1  
Different point estimates of the cointegrating vector.<sup>a, b, c</sup>

$$Y_t = \theta + \beta y_t + z_t$$

where  $Y_t$  is the long-term interest rate,  $y_t$  is the short-term interest rate, and  $z_t$  is  $I(0)$ .

	OLS	NLS (0)	NLS (12)	MLECM (0)	MLECM (12)	PC	CC
(1)	$Y_t = \text{Tb6m}, y_t = \text{Tb3m}$ (sample period 1959:1–1979:7)						
$\hat{\beta}$	0.9926	1.006	1.044	0.9868	1.013	0.9960	0.9934
(2)	$Y_t = \text{Tn1y}, y_t = \text{Tb3m}$ (sample period 1954:1–1979:7)						
$\hat{\beta}$	1.029	1.070	1.120	1.027	1.072	1.043	1.043
(3)	$Y_t = \text{Tn5y}, y_t = \text{Tb3m}$ (sample period 1954:1–1979:7)						
$\hat{\beta}$	0.8685	1.085	1.123	0.9278	1.074	0.9342	1.013
(4)	$Y_t = \text{Tn10y}, y_t = \text{Tb3m}$ (sample period 1954:1–1979:7)						
$\hat{\beta}$	0.8253	1.093	1.136	0.9940	1.120	0.9201	1.063
(5)	$Y_t = \text{Tn20y}, y_t = \text{Tb3m}$ (sample period 1954:1–1979:7)						
$\hat{\beta}$	0.8256	1.137	1.211	1.089	1.222	0.9406	1.114

<sup>a</sup> Monthly data from CITIBASE. 3- and 6-month Treasury bills (secondary market): Tb3m, Tb6m. 1-, 5-, 10-, and 20-year Treasury notes (constant maturity): Tn1y, Tn5y, Tn10y, Tn20y.

<sup>b</sup> In October 1979, the Federal Reserve changed its operating procedure.

<sup>c</sup> See table 2 for a description of these methods.

distribution of PC and CC are derived for the first time in this paper. Section 3 carries out a Monte Carlo study and reports and comments on the results. Situations where MLECM does not perform so well, where OLS is better than NLS, or where PC and CC are inappropriate are shown. The conclusions are provided in section 4.

A word on notation. We use the symbol ' $\Rightarrow$ ' to signify convergence in distribution and the symbol ' $\equiv$ ' to signify equality in distribution. Stochastic processes such as the Brownian motion  $B(r)$  on  $[0, 1]$  are frequently written as  $B$  to achieve notational economy. Similarly, we write integrals with respect to Lebesgue measure such as  $\int_0^1 B(s)ds$  more simply as  $\int B$ . Vector Brownian motion with covariance matrix  $\Omega$  is written ' $BM(\Omega)$ '. Finally, all limits given in the paper are taken as the sample size  $T \rightarrow \infty$ .

## 2. Different methods of estimating cointegrating vectors: Asymptotic results

In order to simplify the comparison proposed, the following bivariate data generating process (DGP) is used:

## DGP(1)

$$y_t = \beta x_t + z_t, \quad z_t = \rho z_{t-1} + e_{z_t}, \quad (2.1)$$

$$\Delta x_t = e_{x_t}, \quad |\rho| < 1, \quad \begin{pmatrix} e_{z_t} \\ e_{x_t} \end{pmatrix} \equiv \text{iid } N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \theta \sigma_1 \sigma_2 \\ \theta \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \right]. \quad (2.2)$$

Both series are  $I(1)$ , but there is a linear combination [the cointegrating vector  $\alpha = (1, -\beta)'$ ] of  $y_t$  and  $x_t$  that is  $I(0)$ .

All the results can be expanded in a simple way to the case with  $n$  ( $n > 2$ ) variables,  $r$  ( $r < n$ ) cointegrating vectors, and more complex dynamics.

A very useful representation of the DGP(1) is its error correction model (a VAR model where the cointegration constraint has been imposed):

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \rho - 1 \\ 0 \end{pmatrix} (y_{t-1} - \beta x_{t-1}) + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, \quad (2.3)$$

with

$$u_{1t} = \beta e_{x_t} + e_{z_t} \quad \text{and} \quad u_{2t} = e_{x_t}.$$

Since  $\Delta x_t$  is weakly exogenous with respect to  $\beta$  (for any value of  $\theta$ ), we can estimate (without any loss of information) the cointegrating vector  $\alpha = (1, -\beta)'$  from the conditional model

$$\Delta y_t = \alpha_1 \Delta x_t + (\rho - 1)(y_{t-1} - \beta x_{t-1}) + u_{1,2t}, \quad (2.4)$$

where

$$u_{1,2t} = \Delta y_t - E(\Delta y_t | \Delta x_t, z_{t-1}) = e_{z_t} - \theta \frac{\sigma_1}{\sigma_2} e_{x_t}.$$

The parameter  $\alpha_1$  ('short-run or impact multiplier') in (2.4), defined in general by

$$E\{\Delta y_t | \Delta x_t, z_{t-1}, \text{lags}(\Delta y_t, \Delta x_t)\} = \alpha_1 \Delta x_t + \alpha_2 z_{t-1} + \text{lags}(\Delta y_t, \Delta x_t), \quad (2.5)$$

plays a key role in the results of this paper, especially its relation with the long-run multiplier  $\beta$ .

A brief summary of the five alternative estimators for a general bivariate framework is in table 2. Their asymptotic distributions (a.d.) are shown in table 3. The derivations of these a.d.s are in the appendix.

In the next subsections we comment the asymptotic results.

Table 2

Summary of different methods of estimating the cointegrating vector  $\alpha' = (1, -\beta)$ .*Ordinary least squares (OLS)*Estimate of OLS  $y_t = \beta x_t + z_t$ .

$$\hat{\beta}_{ols} = \hat{\beta}.$$

*Nonlinear squares (NLS (q))*Estimate by OLS  $\Delta y_t = \pi_{11}y_{t-1} + \pi_{12}x_{t-1} + \sum_{i=1}^q \theta_i \Delta y_{t-i} + \sum_{i=1}^q \delta_i \Delta x_{t-i} + u_{1t}$ .

$$\hat{\beta}_{nls} = -\hat{\pi}_{12}/\hat{\pi}_{11}.$$

*Maximum likelihood in ECM (MLECM (q))* $\Delta H_t = -\gamma' H_{t-1} + \sum_{i=1}^q \Gamma_i \Delta H_{t-i} + u_t$  where  $H_t = (y_t, x_t)'$  and  $u_t \equiv \text{iid } N(0, \Lambda)$ .Regress  $\Delta H_t$  on  $\Delta H_{t-1}, \dots, \Delta H_{t-q}$ . Save the residuals  $R_{ot}$ .Regress  $H_{t-1}$  on  $\Delta H_{t-1}, \dots, \Delta H_{t-q}$ . Save the residuals  $R_{qt}$ .

$$S_{jk} = T^{-1} \sum_{t=1}^T R_{jt} R'_{kt}, \quad j, k = 0, 1.$$

Solve the eigenproblem  $(S_{q0} S_{00}^{-1} S_{0q}) \hat{\alpha}_i = v_i S_{qq} \hat{\alpha}_i$ ,  $i = 1, 2$ .This is equivalent to solve the symmetric eigenproblem  $[(F')^{-1} S_{q0} S_{00}^{-1} S_{0q} (F)^{-1} - v_i I] F \hat{\alpha}_i = 0$ , where  $FF' = S_{qq}$ .

Rank the eigenvalues in descending order.

$$\hat{\beta}_{fml} = -\hat{\alpha}_{12}/\hat{\alpha}_{11}.$$

*Principal components (PC)*

$$M = \sum_{t=1}^T H_t H'_t.$$

Solve the eigenproblem  $M \hat{p}_i = \mu_i \hat{p}_i$ ,  $i = 1, 2$ .

Rank the eigenvalues in descending order.

$$\hat{\beta}_{pc} = -\hat{p}_{22}/\hat{p}_{21}.$$

*Canonical correlations (CC)*

$$M_{jk} = \sum_{t=1}^T H_{t-j} H'_{t-k}, \quad j, k = 0, 1.$$

Solve the eigenproblem  $(M_{01} M_{11}^{-1} M_{10}) \hat{c}_i = \delta_i M_{00} \hat{c}_i$ ,  $i = 1, 2$ , or the symmetric equivalent.

Rank the eigenvalues in descending order.

$$\hat{\beta}_{cc} = -\hat{c}_{22}/\hat{c}_{21}.$$

Table 3  
Asymptotic distributions of estimators of the cointegrating vector.<sup>a, b</sup>

$$\text{DGP: } \begin{aligned} y_t &= \beta x_t + z_t, & z_t &= \rho z_{t-1} + e_{z_t}, & \begin{pmatrix} e_{z_t} \\ e_{x_t} \end{pmatrix} &\equiv \text{iid N} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \theta \sigma_1 \sigma_2 \\ \theta \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \right] \\ \Delta x_t &= e_{x_t}, & |\rho| &< 1, \end{aligned}$$

## OLS

$$T(\hat{\beta}_{\text{ols}} - \beta) \Rightarrow (\int B_2^2)^{-1} \left\{ \left( \frac{\sigma_1}{1-\rho} \right) (1-\theta^2)^{1/2} \int B_2 dW + \left( \frac{1}{1-\rho} \right) (\alpha_1 - \beta) \int B_2 dB_2 + \left( \frac{1}{1-\rho} \right) \theta \sigma_1 \sigma_2 \right\}$$

## NLS

$$T(\hat{\beta}_{\text{nls}} - \beta) \Rightarrow (\int B_2^2)^{-1} \left\{ \left( \frac{\sigma_1}{1-\rho} \right) (1-\theta^2)^{1/2} \int B_2 dW + \left( \frac{1}{1-\rho} \right) \alpha_1 \int B_2 dB_2 \right\}$$

## MLECM

$$T(\hat{\beta}_{\text{mlecm}} - \beta) \Rightarrow (\int B_2^2)^{-1} \left\{ \left( \frac{\sigma_1}{1-\rho} \right) (1-\theta^2)^{1/2} \int B_2 dW \right\}$$

## PC

$$\begin{aligned} T(\hat{\beta}_{\text{pc}} - \beta) &\Rightarrow (\int B_2^2)^{-1} \left\{ \left( \frac{\sigma_1}{1-\rho} \right) (1-\theta^2)^{1/2} \int B_2 dW + \left( \frac{1}{1-\rho} \right) (\alpha_1 - \beta) \int B_2 dB_2 \right. \\ &\quad \left. + \left( \frac{1}{1-\rho} \right) \theta \sigma_1 \sigma_2 + \left( \frac{\beta}{1+\beta^2} \right) \left( \frac{\sigma_1^2}{1-\rho^2} \right) \right\} \end{aligned}$$

## CC

$$\begin{aligned} T(\hat{\beta}_{\text{cc}} - \beta) &\Rightarrow (\int B_2^2)^{-1} \left\{ \left( \frac{\sigma_1}{1-\rho} \right) (1-\theta^2)^{1/2} \int B_2 dW + \left( \frac{1}{1-\rho} \right) (\alpha_1 - \beta) \int B_2 dB_2 \right. \\ &\quad \left. + \left( \frac{\rho}{1+\rho} \right) \left[ \left( \frac{1}{1-\rho} \right) \theta \sigma_1 \sigma_2 \right] \right\} \end{aligned}$$

<sup>a</sup>  $B_2 \equiv BM(\sigma_2^2)$  and  $W \equiv BM(1)$  are independent.

<sup>b</sup>  $\alpha_1$  is defined by  $E(\Delta y_t | \Delta x_t, z_{t-1}) = \alpha_1 \Delta x_t + (\rho - 1) z_{t-1}$ .

### 2.1. Ordinary least squares (OLS)

The asymptotic distribution of OLS involves three different parts. The first has a distribution that is a mixture of normals ( $W$  and  $B_2$  are independent Brownian motions). The second is the usual unit root term,  $(\int B_2^2)^{-1} \int B_2 dB_2$ , that among other things, makes the distribution nonsymmetric. The third one is a kind of simultaneous equation bias caused by the long-run covariance between

$x_t$  and  $z_t$ . The last two terms produce a finite sample bias (asymptotically the bias will vanish) in median and mean, respectively, and invalidate the use of standard distributions for testing hypothesis about the cointegrating vector [see Phillips (1991)].

Notice that if the long-run ( $\beta$ ) and the short-run ( $\alpha_1$ ) multipliers are equal, the last two terms in the a.d. of  $\hat{\beta}_{ols}$  vanish, and OLS is asymptotically equivalent to MLECM. This is because  $\alpha_1 = \beta$  implies  $\theta = 0$ , and therefore  $x_t$  is strictly exogenous. There is no information in eq. (2.2) to estimate  $\beta$  in eq. (2.1).

## 2.2. Nonlinear least squares (NLS)

The idea behind this method is that of two-stage least squares (2SLS). Eqs. (2.1) and (2.2) form the structural model, and the unrestricted VAR (UVAR),

$$\Delta y_t = \pi_{11} y_{t-1} + \pi_{12} x_{t-1} + \text{lags}(\Delta y_t, \Delta x_t) + u_{1t}, \quad (2.6)$$

$$\Delta x_t = \pi_{21} y_{t-1} + \pi_{22} x_{t-1} + \text{lags}(\Delta y_t, \Delta x_t) + u_{2t},$$

is the reduced form. Let us assume for simplicity that  $z_t$  is white noise ( $\rho = 0$ ). Then, the 2SLS estimator of  $\beta$  is obtained from the regression of  $y_t$  on  $x_{t-1}$  (an instrument for  $x_t$ ). Because the long-run covariance between  $x_{t-1}$  and  $z_t$  is zero,  $\beta$  is estimated without the simultaneous equation bias that characterizes the OLS estimator (the third element of its a.d.). Yet, while this bias is actually eliminated, the unit root part (the second term) does create additional problems. If the UVAR (2.6) is estimated by OLS equation by equation (as NLS does), we implicitly assume that the total multiplier matrix  $\Pi$  has full rank, and therefore we do not incorporate the information about the presence of unit roots. This omission will induce a median bias and will complicate inference, first, through the presence of nuisance parameters, and second, because the asymptotic distribution will not be a mixture of normals.

Comparing the a.d. of OLS and NLS, it is seen that although for the NLS estimator the 'simultaneous equation bias' term has disappeared, the fact that the unit root term,  $\int B_2 dB_2$ , does not vanish can make OLS to perform 'better' than NLS in finite samples. This is the case if the short-run multiplier  $\alpha_1$  is big enough and close to the long-run multiplier  $\beta$ . This result agrees with our Monte Carlo study, and explains why in other Monte Carlo experiments, NLS has always beaten OLS. For instance, in Stock (1987) the DGP used is

$$(1 - \lambda L) \begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = - \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} (1 - 1) \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, \quad (2.7)$$

where  $u_t \equiv \text{iid } N(0, I)$  and  $\gamma_2 = 0$  (or  $\gamma_2 = -\gamma_1$ ). In this experiment  $\alpha_1$ , the coefficient of  $\Delta x_t$  in the conditional expectation of  $\Delta y_t$  with respect to  $\Delta x_t, z_{t-1}$ , and lags  $(\Delta y_t, \Delta x_t)$ , equals zero. Notice that with the DGP(1), NLS and MLECM are asymptotically equivalent when  $\alpha_1 = 0$ .

Notice that NLS does not necessarily perform better if  $\Delta x_t$  and  $z_t$  are uncorrelated. In fact, in the DGP(1) this can make things worse ( $\theta = 0$  implies  $\alpha_1 = \beta$ ), depending on the size of  $\beta$ . The relevant parameter in the a.d. of  $\hat{\beta}_{\text{nls}}$  is  $\alpha_1$ , which in some way measures the importance of the mistake made in estimating  $\beta$  from the marginal distribution of  $\Delta y_t$  [the first equation of the ECM (2.3)] rather than from the conditional distribution (2.4). Note that the value of  $\alpha_1$  has nothing to do with the fact that  $x_t$  is weakly (or strongly) exogenous with respect to  $\beta$  in the DGP(1).

### 2.3. Maximum likelihood in an error correction model (MLECM)

The method of MLECM estimates  $\alpha$  by maximum likelihood in the fully specified ECM,

$$\Delta H_t = \gamma \alpha' H_{t-1} + \text{lags}(\Delta H_t) + u_t \quad (2.8)$$

[see Johansen (1988)]. This is a nonlinear estimation problem that can be simplified to a simple eigenproblem by applying reduced rank regression techniques<sup>4</sup> [see appendix B in Gonzalo (1991) for a general review of these techniques]. Table 2 shows how to implement this method. An empirical application can be found in Johansen and Juselius (1988).

The asymptotic distribution of MLECM has the smallest variance among all these five estimators. As an example compare the a.d. of OLS and MLECM. The covariance between

$$\left\{ \left( \int B_2^2 \right)^{-1} \left\{ \left( \frac{\sigma_1}{1-\rho} \right) (1-\theta^2)^{1/2} \int B_2 dW \right\} \right\}$$

and

$$\left\{ \left( \int B_2^2 \right)^{-1} \left\{ \frac{(\alpha_1 - \beta)}{1-\rho} \int B_2 dB_2 + \frac{\theta \sigma_1 \sigma_2}{1-\rho} \right\} \right\}$$

<sup>4</sup>The main results concerning the estimation of a reduced rank regression coefficient were obtained by Anderson (1951), Rao (1965), and Brillinger (1981) and restated in Izenman (1980), Velu et al. (1986), and Robinson (1973). Most of these results are based on an application of the Eckart-Young Theorem [see Eckart and Young (1936)].



is zero because  $B_2$  and  $W$  are two independent Brownian motions. Therefore the variance of the a.d. of  $\hat{\beta}_{mlecm}$  has to be smaller than that of  $\hat{\beta}_{ols}$ . Nevertheless, this does not imply that MLECM is more efficient since they do not have the same asymptotic distribution. To prove the asymptotic efficiency of MLECM, see Saikkonen (1991).

It is important to notice that in cointegrated systems the inclusion of the correct information will not only increase the efficiency, as in the stationary case, but will also decrease the bias in mean and median (simultaneous equation bias and unit root bias, respectively) and make the asymptotic distribution symmetric. This can be seen when comparing the a.d. of  $\hat{\beta}_{mlecm}$  with the a.d. of the other estimators.

As important as the above remarks is the fact that the asymptotic distribution of  $\hat{\beta}_{mlecm}$  is a mixture of normals ( $W$  and  $B_2$  are two independent Brownian motions). Therefore, hypothesis tests may be conducted using standard asymptotic chi-squared tests. This solves the problem of inference in cointegrated systems [see Johansen (1988) and Phillips (1991)].

Two objections have been raised against Johansen's method:

- (a) The number of lags in the ECM (2.8) is unknown.
- (b)  $\{u_t\}$  may be non-Gaussian.

Both of them are discussed in sections 3.1 and 3.2.

#### 2.4. Principal components (PC)

The method of PC is designed, as OLS, to find the linear combination of  $(y_t, x_t)$  with minimum variance. This is the linear combination described by the cointegrating vector  $(1, -\beta)$ .

Let  $\Sigma$  be the covariance matrix of  $H_t = (y_t, x_t)'$ , then PC solves the following minimization (or maximization) problem:

$$\min p' \Sigma p \quad \text{subject to} \quad p' p = 1. \quad (2.9)$$

From the first-order conditions we have

$$\Sigma p_i = \mu_i p_i, \quad i = 1, 2. \quad (2.10)$$

Therefore  $\text{var}(p_i' H_t) = \mu_i$ , and hence the cointegrating vector is the eigenvector corresponding to the smallest eigenvalue of  $\Sigma$ . An easy way of understanding this technique is thinking in terms of orthogonal regression [see Malinvaud (1980, pp. 9–13)]. The principal component with minimum variance is orthogonal to the line that minimizes the sum of the squared perpendicular distances

from the points  $(x_1, y_1), \dots, (x_T, y_T)$ ,

$$\min_{\{\beta\}} (1 + \beta^2)^{-1} \sum_{t=1}^T (y_t - \beta x_t)^2. \quad (2.11)$$

The asymptotic distribution of OLS and PC are very similar, since both are regression methods in the same single equation  $y_t = \beta x_t + z_t$ . The only difference between the two lies in the mean bias. The mean of the asymptotic distribution of PC contains the variance of  $z_t$ , and this makes the method very sensitive to the units of measurement (something confirmed by the simulation results). This is so, because the normalization  $p'/p = 1$  is not the correct one. Notice that the mean bias of PC can be bigger or smaller than that of OLS.

### 2.5. Canonical correlations (CC)

CC analysis searches for linear combinations of elements of  $H_t = (y_t, x_t)'$  (which define canonical variables  $c^0 H_t$ ) and linear combinations of  $H_{t-1}$  (generating corresponding canonical variables  $c^1 H_{t-1}$ ) that are maximally correlated.

CC analysis solves the following maximization problem:

$$\max \{c^0 \Sigma_{01} c^1 / [c^0 \Sigma_{00} c^0 \cdot c^1 \Sigma_{11} c^1]^{1/2}\} \quad (2.12)$$

$$\text{subject to } c^0 \Sigma_{00} c^0 = 1, \quad c^1 \Sigma_{11} c^1 = 1,$$

where

$$\Sigma_{00} = \text{var}(H_t), \quad \Sigma_{11} = \text{var}(H_{t-1}), \quad \Sigma_{01} = \text{cov}(H_t, H_{t-1}).$$

The solution from the first-order conditions is

$$\Sigma_{01} \Sigma_{11}^{-1} \Sigma_{10} c_i^0 = \delta_i \Sigma_{00} c_i^0, \quad \Sigma_{10} \Sigma_{00}^{-1} \Sigma_{01} c_i^1 = \delta_i \Sigma_{11} c_i^1, \quad i = 1, 2. \quad (2.13)$$

It can be shown that  $\delta_i = [\text{correlation}(c_i^0 H_t, c_i^1 H_{t-1})]^2$ , then if  $\delta_1 > \delta_2$ , it is clear that the first canonical variable will be nonstationary. In fact,  $c_1^0 = c_1^1$  since that will give maximum correlation (namely one). The crucial point is that the remaining canonical variables cannot be defined by vectors  $c_2^0$  and  $c_2^1$  in the complement of the space spanned by  $\{\alpha\}$ , since in that case they would be nonstationary, and cointegrated with the first canonical variable (otherwise cannot be one cointegrated vector). Therefore  $(c_1^0 H_t)$  and  $(c_1^1 H_{t-1})$  does not satisfy the requirements of being mutually uncorrelated. Consequently  $c_2^0$  (and  $c_2^1$ ) must be defined by elements in the space spanned  $\{\alpha\}$ , i.e., it will be the cointegrated

vector. This method proposed by Bossaerts (1988) is related to the one used by Aoki (1988) to find permanent and transitory components.

The difference between MLECM and CC in the DGP(1) is that, while MLECM performs a canonical correlation analysis between  $\Delta H_t$  and  $H_{t-1}$ , CC does so between  $H_t$  and  $H_{t-1}$ . Such a slight difference has a big repercussion in the asymptotic distribution. The main reason for this is that CC does not incorporate any information on the presence of unit roots into the estimation process.

### 3. Simulation results

The data generating process used for the Monte Carlo study is based on the one used by Banerjee et al. (1986), Engle and Granger (1987), and Hansen and Phillips (1990):

DGP(2)

$$y_t - \beta x_t = z_t, \quad z_t = \rho z_{t-1} + e_{z_t}, \quad (3.1)$$

$$\begin{pmatrix} e_{z_t} \\ e_{w_t} \end{pmatrix} \equiv \text{iid } N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \theta\sigma \\ \theta\sigma & \sigma^2 \end{pmatrix} \right]. \quad (3.2)$$

The difference between DGP(2) and DGP(1) is that in the former  $y_t$  can appear (if  $a_1 \neq 0$ ) in the second equation, and therefore the error correction term ( $z_{t-1}$ ) can be present in both equations of the ECM. In this case  $x_t$  is no longer weakly exogenous [see Gonzalo (1991)].

In order for the reader to better understand our results, it is important to notice that the above DGP can be rewritten as DGP(1) with MA errors [model used in Phillips and Loretan (1991)],

$$\begin{aligned} y_t &= \beta x_t + z_t, \\ \Delta x_t &= e_{x_t}, \end{aligned} \quad \begin{bmatrix} z_t \\ e_{x_t} \end{bmatrix} = \begin{bmatrix} \frac{1}{1-\rho L} & 0 \\ -a_1 & \frac{(1-L)}{(1-\rho L)} & \frac{1}{a_1\beta - a_2} \end{bmatrix} \begin{bmatrix} e_{z_t} \\ e_{w_t} \end{bmatrix}, \quad (3.3)$$

or as an ECM [DGP used in Stock (1987)],

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = - \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} [1 - \beta] \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, \quad (3.4)$$

with

$$\gamma_1 = (\rho - 1) \frac{a_2}{d}, \quad \gamma_2 = (\rho - 1) \frac{a_1}{d},$$

$$u_{1t} = \frac{1}{d} (-a_2 e_{zt} + \beta e_{wt}), \quad u_{2t} = \frac{1}{d} (-a_1 e_{zt} + e_{wt}),$$

and

$$d = (a_1 \beta - a_2).$$

An equivalent model to (3.4), but with uncorrelated errors, is obtained diagonalizing the covariance matrix of  $u_t$ ,

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ 0 \end{pmatrix} \Delta x_t + \begin{pmatrix} \alpha_1 \gamma_2 - \gamma_1 \\ \gamma_2 \end{pmatrix} (1 - \beta) \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \quad (3.5)$$

where

$$\alpha_1 = \frac{\text{cov}(\varepsilon_{1t}, \varepsilon_{2t})}{\text{var}(\varepsilon_{1t})}, \quad e_{1t} = u_{1t} - \alpha_1 u_{2t}, \quad e_{2t} = u_{2t}.$$

We have considered the parameter space  $(a_1 \times a_2 \times \beta \times \rho \times \sigma \times \theta \times T)$ , where  $a_1 = (0, 1)$ ,  $a_2 = (-1, 1)$ ,  $\beta = (0.5, 1)$ ,  $\rho = (0.5, 0.8, 0.9)$ ,  $\sigma = (0.25, 0.33, 0.5)$ ,  $\theta = (-0.5, 0.5)$ , and  $T = (100, 300)$ , giving rise to 180 experiments.

In all the simulations we generated 500 series of length  $T + 20$ , starting with  $e_{z0} = 0$  and  $e_{w0} = 0$ , and then discarding the initial 20 observations. The GAUSS matrix programming language and its RNDN functions were used to generate the pseudo-normal innovations. The latter function uses the fast acceptance-rejection algorithm proposed by Kinderman and Ramage (1976). For consistency, given a point in the parameter space, all the methods have been computed using the same random numbers.

A brief review of the methods studied is in table 2. The methods based on dynamic regressions (MLECM and NLS) have been used in two ways: one with the correct number of lags, the other with more lags than the true model. The simulation results are summarized in tables 4, 5, 6, and 7. For simplicity, we only present the case where  $\theta = -0.5$  and  $\rho = 0.8$  (a copy of the entire experiment is available on request). The results are presented in terms of the two parameters that appear to be the relevant ones from the asymptotic distributions: (1) the short-run multiplier  $\alpha_1$  and (2) the ratio signal-noise  $\sigma$  that measures how big the random walk component of the variables is. The five estimators are compared in terms of the mean, median, standard deviation (sd), and interquartile

Table 4  
Characteristics of the empirical distribution of estimators of the cointegrating vector

$$\text{DGP: } \begin{aligned} y_t - \beta x_t &= z_t, & z_t &= \rho z_{t-1} + e_{z,t}, & \begin{pmatrix} e_{z,t} \\ e_{x,t} \end{pmatrix} &\equiv \text{iid } N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \theta\sigma \\ \theta\sigma & \sigma^2 \end{pmatrix} \right) \\ a_1 y_t - a_2 x_t &= w_t, & w_t &= w_{t-1} + e_{w,t} \end{aligned}$$

Parameters:  $\beta = 1$ ,  $a_1 = 0$ ,  $a_2 = -1$ ,  $\rho = 0.8$ ,  $\theta = -0.5$ ,  $T = 100$

$\alpha_1$	$\sigma$	OLS	NLS (0)	NLS (4)	MLECM (0)	MLECM (4)	PC	CC
Bias in mean								
-1	0.25	-0.2219	0.0798	0.0970	0.0101	0.1183	19.6298	-0.0079
0	0.5	-0.1110	-0.0014	0.0053	0.0051	0.0591	0.1113	-0.0039
0.5	1	-0.0555	-0.0400	-0.0378	0.0025	0.0296	-0.0007	-0.0020
0.75	2	-0.0277	-0.0891	-0.0067	0.0013	0.0148	-0.0147	-0.0010
Bias in median								
-1	0.25	-0.1764	0.0422	0.0490	0.0044	0.0099	0.4649	-0.0317
0	0.5	-0.0882	0.0028	0.0039	0.0022	0.0050	0.0628	-0.0159
0.5	1	-0.0441	-0.0150	-0.0131	0.0011	0.0025	-0.0080	-0.0079
0.75	2	-0.0221	-0.0241	-0.0264	0.0005	0.0012	-0.0119	-0.0040
IQR <sub>(50)</sub>								
-1	0.25	0.3927	0.3930	0.4078	0.3749	0.4096	1.3063	0.3793
0	0.5	0.1963	0.1877	0.1861	0.1875	0.2048	0.2595	0.1897
0.5	1	0.0982	0.1087	0.1064	0.0937	0.1024	0.0914	0.0948
0.75	2	0.491	0.1161	0.1181	0.0469	0.0512	0.0452	0.0474
SD								
-1	0.25	0.3785	0.4683	0.5220	0.4889	1.666	411.1135	0.4854
0	0.5	0.1893	0.2125	0.2334	0.2444	0.8333	1.0734	0.2427
0.5	1	0.0946	0.1256	0.1415	0.1222	0.4166	0.0983	0.1214
0.75	2	0.0473	0.6257	0.9838	0.0611	0.2083	0.0450	0.0607
Prob ( $ \hat{\beta} - \beta  < 0.05$ )								
-1	0.25	14.4	15.2	12.2	15.0	15.2	7.6	17.0
0	0.5	26.6	28.8	27.8	29.6	26.0	22.8	30.2
0.5	1	46.4	51.6	46.8	54.8	49.0	50.2	48.6
0.75	2	73.6	53.2	50.2	78.4	74.8	77.2	77.0

\* 500 replications were used to calculate the Monte Carlo numerical summaries.

<sup>b</sup>  $\alpha_1$  is defined by  $E(\Delta y_t | \Delta x_t, z_{t-1}) = \alpha_1 \Delta x_t + (\rho - 1)z_{t-1}$ .

Table 5  
Characteristics of the empirical distribution of estimators of the cointegrating vector.<sup>a,b</sup>

$$\text{DGP: } y_t = \beta x_t = z_t, \quad z_t = \rho z_{t-1} + e_{z,t}, \quad \begin{pmatrix} e_{z,t} \\ e_{w,t} \end{pmatrix} \equiv \text{iid N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \theta\sigma \\ \theta\sigma & \sigma^2 \end{pmatrix} \right)$$

$$a_1 y_t = a_2 x_t = w_t, \quad w_t = w_{t-1} + e_{w,t},$$

Parameters:  $\beta = 1$ ,  $a_1 = 0$ ,  $a_2 = -1$ ,  $\rho = 0.8$ ,  $\theta = -0.5$ ,  $T = 300$

$\alpha_1$	$\sigma$	OLS	NLS (0)	NLS (4)	MLECM (0)	MLECM (4)	PC	CC
Bias in mean								
-1	0.25	-0.1132	0.0187	0.0196	0.0018	0.0013	0.7759	-0.0217
0	0.5	-0.0566	-0.0050	-0.0046	0.0009	0.0006	0.0454	-0.0109
0.5	1	-0.0283	-0.0173	-0.0177	0.0005	0.0003	-0.0053	-0.0054
0.75	2	-0.0142	-0.0284	-0.0257	0.0002	0.0002	-0.0085	-0.0027
Bias in median								
-1	0.25	-0.0816	0.0086	0.0048	0.0025	-0.0012	0.1610	-0.0290
0	0.5	-0.0408	0.0031	-0.0030	0.0012	-0.0006	0.0151	-0.0145
0.5	1	-0.0204	-0.0076	-0.0071	0.0006	-0.0003	-0.0070	-0.0072
0.75	2	-0.0102	-0.0108	-0.0094	0.0003	-0.0001	-0.0070	-0.0036
IQR <sub>(50)</sub>								
-1	0.25	0.1578	0.1438	0.1417	0.1378	0.1406	0.4584	0.1428
0	0.5	0.0789	0.0682	0.0694	0.0689	0.0703	0.0902	0.0714
0.5	1	0.0395	0.0438	0.0453	0.0344	0.0352	-0.0350	0.0357
0.75	2	0.0197	0.0422	0.0453	0.0172	0.0176	0.0171	0.0178
SD								
-1	0.25	0.1542	0.1390	0.1466	0.1379	0.1489	5.2901	0.1419
0	0.5	0.0771	0.0660	0.0694	0.0690	0.0744	0.1200	0.0710
0.5	1	0.0386	0.0452	0.0488	0.0345	0.0372	0.0337	0.0355
0.75	2	0.0193	0.0811	0.2030	0.0172	0.0186	0.0168	0.0177
Prob ( $ \hat{\beta} - \beta  < 0.05$ )								
-1	0.25	28.8	40.2	37.8	37.4	37.2	23.2	36.8
0	0.5	52.2	65.4	65.2	64.2	64.8	59.2	61.6
0.5	1	78.0	81.8	82.0	88.8	87.2	89.4	88.4
0.75	2	95.6	80.0	77.6	98.2	97.8	97.4	98.2

<sup>a</sup> 500 replications were used to calculate the Monte Carlo numerical summaries.

<sup>b</sup>  $\alpha_1$  is defined by  $E(\Delta y_t | \Delta x_t, z_{t-1}) = \alpha_1 \Delta x_t + (\rho - 1) z_{t-1}$ .

Table 6  
Characteristics of the empirical distribution of estimators of the cointegrating vector  $\beta$ <sup>a</sup>

$$\text{DGP: } \begin{matrix} y_t & \beta_1 x_t & z_t & w_t & v_t & u_t \\ a_1 y_t & a_2 x_t & w_t & w_t & w_t & w_t \end{matrix} \left( \begin{matrix} y_t \\ x_t \\ z_t \\ w_t \\ v_t \\ u_t \end{matrix} \right) \sim \text{iid N} \left[ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right]$$

Parameters:  $\beta = 1, a_1 = 1, a_2 = 1, \rho = 0.8, \theta = 0.5, T = 100$

$\alpha_1$	$\sigma$	OLS	NLS (0)	NLS (4)	MLECM (0)	MLECM (4)	PC	CC
-0.7	0.25	-1.0117	0.7443	-0.2242	0.9187	1.7060	-0.4505	3.7943
-0.4	0.5	-0.5422	-0.0023	0.4562	0.0630	0.1759	-0.1946	0.1463
0	1	-0.2359	-0.0534	-0.0421	0.0435	0.0973	-0.1011	0.0329
0.4	2	-0.0940	-0.1205	-0.0472	0.0102	0.0039	-0.0518	0.0057
0.7	4	-0.0383	-0.2559	-0.0944	0.0031	-0.0246	-0.0266	0.0099
-0.7	0.25	-1.0379	0.0413	0.0383	-0.0183	-0.0309	-1.0892	-0.0939
-0.4	0.5	-0.5146	0.0036	0.0078	-0.0001	-0.0014	-0.2246	-0.0326
0	1	-0.1982	-0.0296	-0.0258	0.0022	0.0029	-0.0905	-0.0157
0.4	2	-0.0751	-0.0493	-0.0531	0.0011	0.0014	-0.0441	-0.0079
0.7	4	-0.0295	-0.0500	-0.0304	0.0005	0.0009	-0.0218	-0.0040
-0.7	0.25	0.6516	0.8142	0.8848	0.7281	0.8269	1.3796	0.6677
-0.4	0.5	0.4937	0.3735	0.3827	0.3712	0.4139	0.5177	0.3617
0	1	0.2594	0.2081	0.2056	0.1879	0.2065	0.2040	0.1870
0.4	2	0.1138	0.2168	0.2243	0.0938	0.1026	0.0965	0.0941
0.7	4	0.0543	0.2000	0.1811	0.0469	0.0514	0.0481	0.0472
-0.7	0.25	0.4248	10.8594	17.6214	39.7996	35.6712	11.1484	79.0213
-0.4	0.5	0.3480	1.6007	6.4467	1.2540	1.9705	2.3015	1.4918
0	1	0.2048	0.2008	0.2395	0.3732	0.9739	0.2217	0.3449
0.4	2	0.1017	0.7915	0.8681	0.1301	0.2976	0.0917	0.1297
0.7	4	0.0493	2.6322	6.9943	0.0615	0.6236	0.0457	0.0615
-0.7	0.25	0.2	7.8	6.4	8.2	8.0	2.4	7.4
-0.4	0.5	4.4	12.8	14.2	15.2	14.8	10.4	17.0
0	1	13.6	27.2	26.2	29.6	26.2	25.2	30.2
0.4	2	37.0	32.2	29.0	54.6	49.6	45.6	49.2
0.7	4	67.2	31.2	34.0	78.8	75.0	74.0	77.6

<sup>a</sup> 500 replications were used to calculate the Monte Carlo numerical summaries.

<sup>b</sup>  $\alpha_1$  is defined by  $E(\Delta y_t | \Delta x_t, z_{t-1}) = \alpha_1 \Delta x_t + f(z_{t-1})$ .

Table 7  
 Characteristics of the empirical distribution of estimators of the cointegrating vector.<sup>a,b</sup>

$$\text{DGP: } y_t = \beta x_t = z_t, \quad z_t = \rho z_{t-1} + e_{z_t}, \quad \begin{pmatrix} e_{z_t} \\ e_{w_t} \end{pmatrix} \equiv \text{iid } N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \theta\sigma \\ \theta\sigma & \sigma^2 \end{pmatrix} \right]$$

$$a_1 y_t - a_2 x_t = w_t, \quad w_t = w_{t-1} + e_{w_t},$$

Parameters:  $\beta = 1$ ,  $a_1 = 1$ ,  $a_2 = -1$ ,  $\rho = 0.8$ ,  $\theta = -0.5$ ,  $T = 300$ .

$\alpha_1$	$\sigma$	OLS	NLS (0)	NLS (4)	MLECM (0)	MLECM (4)	PC	CC
					Bias in mean			
-0.7	0.25	-0.7011	0.1149	0.1737	0.2960	-0.1740	-0.5485	-0.1395
-0.4	0.5	-0.3200	-0.0014	0.0003	0.0089	0.0138	-0.1401	-0.0100
0	1	-0.1249	-0.0305	-0.0306	0.0016	0.0035	-0.0566	-0.0082
0.4	2	-0.0476	-0.0474	-0.0830	0.0001	0.0010	-0.0277	-0.0048
0.7	4	-0.0192	-0.3787	-0.1089	-0.0001	0.0003	-0.0139	-0.0026
					Bias in median			
-0.7	0.25	-0.6315	0.0173	0.0097	-0.0050	-0.0028	-0.2816	-0.0575
-0.4	0.5	-0.2519	-0.0063	-0.0059	-0.0025	-0.0012	-0.0915	-0.0286
0	1	-0.0942	-0.0151	-0.0141	-0.0012	-0.0006	-0.0411	-0.0144
0.4	2	-0.0338	-0.0213	-0.0188	-0.0006	-0.0003	-0.0204	-0.0072
0.7	4	-0.0133	-0.0200	-0.0144	-0.0003	-0.0001	-0.0102	-0.0036
					IQR <sub>(.50)</sub>			
-0.7	0.25	0.5998	0.2968	0.2883	0.2748	0.2811	0.8903	0.2680
-0.4	0.5	0.3336	0.1355	0.1377	0.1374	0.1403	0.1851	0.1390
0	1	0.1393	0.0854	0.0884	0.0688	0.0702	0.0782	0.0704
0.4	2	0.0550	0.0816	0.0904	0.0344	0.0351	0.0389	0.0354
0.7	4	0.0239	0.0922	0.0900	0.0172	0.0176	0.0195	0.0178
					SD			
-0.7	0.25	0.3782	0.8197	1.9326	5.9084	5.0064	0.7565	3.0758
-0.4	0.5	0.2389	0.1321	0.1381	0.1622	0.1762	0.2406	0.1713
0	1	0.1141	0.0803	0.0855	0.0724	0.0779	0.0795	0.0750
0.4	2	0.0497	0.1049	0.4319	0.0351	0.0378	0.0377	0.0362
0.7	4	0.0222	8.7760	1.9424	0.0174	0.0187	0.0189	0.0179
					Prob ( $ \hat{\beta} - \beta  < 0.05$ )			
-0.7	0.25	0.8	20.4	20.2	19.2	19.6	10.4	18.4
-0.4	0.5	5.6	37.8	38.0	37.0	37.4	26.0	37.8
0	1	28.6	60.4	60.8	64.2	65.2	51.4	62.0
0.4	2	64.4	61.0	60.2	88.8	87.0	78.4	88.4
0.7	4	91.8	56.8	54.4	98.2	97.8	96.0	98.4

<sup>a</sup> 500 replications were used to calculate the Monte Carlo numerical summaries.

<sup>b</sup>  $\alpha_1$  is defined by  $E(\Delta y_t | \Delta x_t, z_{t-1}) = \alpha_1 \Delta x_t + f(z_{t-1})$ .



range (IQR) of their empirical distributions. Our conclusion will be based on the median and IQR because all the estimators, except OLS which imposes a normalization before the estimation, are ratios of random variables; therefore their finite sample moments may not exist, making the comparison in terms of the second moments very unfair [see Anderson (1976)]. If one uses the criterion of mean square error (MSE), then one must prefer (for  $T = 100$ )  $\hat{\beta}_{ols}$  to  $\hat{\beta}_{mlecm}$ ; however the probability of  $\hat{\beta}_{mlecm}$  falling in an interval about  $\beta$  may be greater than the probability of  $\hat{\beta}_{ols}$  falling in that interval for many intervals of interest as it is shown in the bottom part of tables 4 through 7.

Our objective is to decide which estimator performs 'better' in finite samples. We interpret 'better' in terms of the empirical distribution being more concentrated around the true value of the parameter. That is why we chose as the best estimator the one that has smaller bias in median and smaller IQR.

The following comments summarize the Monte Carlo study. For simplicity we focus most of our remarks on the case of  $a_1 = 0$ . The only estimator that can take advantage of it is MLECM. To make a fair comparison our MLECM estimator has not incorporated this constraint. In doing that, the conclusion is not affected by the value of  $a_1$ .

All the results of the Monte Carlo are consistent with the asymptotic distribution shown in section 2. The bias and dispersion of all the empirical distributions, except that of NLS, decrease as the ratio signal-noise  $\sigma$  increases and as the difference between the short-run and long-run multiplier ( $\alpha_1 - \beta$ ) decreases. In the case of NLS, its performance becomes worse when  $\alpha_1$  increases. This is because NLS is estimating  $\beta$  from the first equation of the ECM (3.4) and  $\alpha_1$  is measuring the cost of omitting a relevant variable  $\Delta x_t$  in this equation [see (3.5)].

ML in a fully specified ECM is less biased in mean and median and has smaller sample dispersion (measured by IQR) than the other estimators. This holds for both sample sizes  $T = 100$  and  $T = 300$ . In the latter case the SD is also smaller than that of the other methods, as the asymptotic distributions suggest. The true value  $\beta = 1$  is contained in a 95% confidence interval around the sample mean of  $\hat{\beta}_{mlecm}$  in all the cases. Notice that, even in the situations where MLECM does not perform so well (i.e., for small values of  $\sigma$ ) in the sense of having a bigger standard deviation and bias in mean than NLS (see tables 6 and 7), it still has less bias in median and smaller IQR than its competitors.

OLS is downward biased in mean and median; the sample distribution is skewed, but the sample dispersion (measured by the SD) is smaller than that of the other estimators when  $T = 100$  and similar to the one of MLECM when  $T = 300$ . OLS has the smaller MSE when  $T = 100$  because an arbitrary normalization has been imposed before the estimation. If we regress  $x$  and  $y$  instead of  $y$  on  $x$ , and we invert the estimates to get  $\hat{\beta}_{ols}$ , the MSE is higher than that of the other estimators in most of the cases ( $\sigma < 2$ ). Notice that even when OLS has

smaller MSE than that of MLECM, the latter estimator has a greater probability of obtaining an accurate estimate,

$$\text{Prob}(|\hat{\beta}_{\text{MLECM}} - \beta| < c) > \text{Prob}(|\hat{\beta}_{\text{OLS}} - \beta| < c),$$

as it is shown at the bottom of tables 4 to 7, for  $c = 0.05$ . Different values of  $c$  have been tried with the same conclusion. The true value  $\beta = 1$  is not contained in a 95% confidence interval around the sample mean of  $\hat{\beta}_{\text{OLS}}$  in any case. The rate of convergence is slower than the theoretical one, in the sense that when the sample size increases from 100 to 300, the bias decreases in a lower proportion. OLS performs very well when the size of the random walk component of the variables is big ( $\sigma > 1$ ).

NLS performs better than OLS when  $\alpha_1 < 0.5$ , but not in other cases. As we have mentioned earlier, the reason why in other experiments NLS always beats OLS is because a DGP with a very small  $\alpha_1$  has been used. In tables 4 through 7 we can see that this is a situation clearly favorable to NLS. In fact, when  $x_t$  is weakly exogenous with respect to  $\beta$  and  $\alpha_1 = 0$ , NLS and MLECM are asymptotically equivalent and their finite sample performances are very similar (see tables 4 and 5).

PC (orthogonal regression) performs a little better than OLS when  $\sigma$  (or  $\alpha_1$ ) is not very small; otherwise it is the worst method because of its high sensitivity to units of measurement.

CC performs very similarly to MLECM when  $H_t = (y_t, x_t)'$  follows an AR(1), although with a higher bias in median. In other situations (when the ECM has more lags), CC could be worse than OLS, especially in the MA case (see table 8).

From the above remarks, it is clear why we propose MLECM (the Johansen method) to estimate the cointegrating vector. It has the smallest bias in median and sample dispersion (measured by IQR) and presents the greater probability of obtaining a very accurate estimate (see the bottom part of tables 4–7). The next two sections relate to the two major objections that have been raised against this method.

### 3.1. Unknown number of lags in the ECM

The example in table 1 shows that Johansen's method gives different estimates of  $\beta$  depending on the number of lags we choose in the ECM. In tables 4 to 7, where  $H_t$  follows an AR(1), it seems that the cost of overparametrizing by including more lags in the ECM is small in terms of efficiency lost – MLECM(4) still performs better than the other methods. This is not the case if the ECM is underparametrized, as is shown in the first part of table 8, where the true model for  $H_t$  is an AR(2) and an AR(1) is fitted – OLS is superior to MLECM(0).

Table 8

Characteristics of the empirical distribution of estimators of the cointegrating vector.<sup>a,b</sup>

$$\text{DGP} \quad \begin{aligned} y_t - \beta x_t &= z_t, & z_t &= \begin{cases} 1.6z_{t-1} - 0.8z_{t-2} + e_{z_t}, \\ e_{z_t} - 0.9e_{z_{t-1}}, \end{cases} & \begin{pmatrix} e_{z_t} \\ e_{w_t} \end{pmatrix} &\equiv \text{iid } N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \theta\sigma \\ \theta\sigma & \sigma^2 \end{pmatrix} \right] \\ u_t y_t - a_2 x_t &= w_t, & w_t &= w_{t-1} + e_{w_t}, \end{aligned}$$

Parameters:  $\beta = 1$ ,  $a_1 = 0$ ,  $a_2 = -1$ ,  $\theta = -0.5$ ,  $\sigma = 1$ ,  $\alpha_1 = 0.5$ ,  $T = 100$ 

	OLS	NLS (0)	NLS (4)	MLECM (0)	MLECM (4)	PC	CC
Bias in mean							
AR (2)	-0.0652	-0.0726	-0.0302	0.1584	0.0042	0.2990	0.0699
MA (2)	-0.0015	-0.0024	-0.0013	0.0076	0.0009	0.0377	0.0279
Bias in median							
AR (2)	-0.0507	0.0040	-0.0089	0.0360	0.0049	0.1276	0.0163
MA (2)	-0.0008	0.0028	-0.0004	0.0053	0.0004	0.0204	0.0141
IQR <sub>(50)</sub>							
AR (2)	0.1041	0.2103	0.1053	0.1568	0.0985	0.2732	0.1224
MA (2)	0.0048	0.0077	0.0043	0.0069	0.0029	0.0395	0.0279
SD							
AR (2)	0.1169	0.3667	0.1216	3.3394	0.1051	0.5428	0.4104
MA (2)	0.0066	0.0076	0.0049	0.0077	0.0031	0.0498	0.0400

<sup>a</sup> 500 replications were used to calculate the Monte Carlo numerical summaries.<sup>b</sup>  $z_t$  is defined by  $E(\Delta y_t | \Delta x_t, z_{t-1}) = \alpha_1 \Delta x_t + (\rho - 1)z_{t-1}$ .

Fitting an ECM with four lags solves the problem. The same remains true in the second part of table 8, where the true model for  $H_t$  is an  $AR(\infty)$ .

Further research will be oriented towards cases where there are not enough degrees of freedom to choose the right number  $q$  of lags. According to the Monte Carlo study, this seems to be the most serious drawback of Johansen's method.

### 3.2. Nonnormal errors

When the errors are non-Gaussian, there is no reason for Johansen's method to perform worse than the other four methods, if one realizes that this procedure is a particular case of reduced rank simultaneous least squares (RRSLS) where no assumptions about any particular distribution of the error term is made. In fact, it can be proved that the asymptotic distribution of  $\hat{\beta}_{rrsls}$  is equivalent to that of  $\hat{\beta}_{mlecm}$  [see appendix B in Gonzalo (1991)].

The experiment has been performed with different distributions of the errors  $e_1$  and  $e_2$ ,

$$e_{w_t} = \sigma e_1, e_{z_t} = (1/\sigma)\theta e_{w_t} + (1 - \theta^2)^{1/2} e_2, \quad t = 1, \dots, T.$$

The distributions considered are two nonsymmetric (exponential and chi-squared with one degree of freedom) ones and two with heavy tails (logistic and student's  $t$  with three degrees of freedom). Simulations with the uniform  $[0, 1]$ , extreme value (type I), and ARCH distributions were also carried out. In all these cases the ranking of the five methods did not change at all. Tables 9, 10, and 11 present the results for the standardized chi-squared, student's  $t$ , and ARCH distributions (other results are available upon request).

Table 9

Characteristics of the empirical distribution of estimators of the cointegrating vector.<sup>a</sup>

$$\text{DGP: } y_t - \beta x_t = z_t, \quad z_t = \rho z_{t-1} + e_{z_t}, \quad e_{z_t} = (1/\sigma)\theta e_{w_t} + (1 - \theta^2)^{1/2} e_{2t} \\ a_1 y_t - a_2 x_t = w_t, \quad w_t = w_{t-1} + e_{w_t}, \quad e_{w_t} = \sigma e_{1t}, \quad e_{it} \equiv \chi^2(1\text{df}), \quad i = 1, 2$$

Parameters:  $\beta = 1$ ,  $a_1 = 0$ ,  $a_2 = -1$ ,  $\rho = 0.8$ ,  $\theta = 0.5$ , and  $T = 100$

$\sigma$	OLS	NLS (0)	NLS (4)	MLECM (0)	MLECM (4)	PC	CC
Bias in mean							
0.25	-0.3557	0.2141	0.2311	0.0526	0.2894	1.0395	0.1121
0.5	-0.1778	-0.0152	-0.0141	0.0263	0.1447	0.5142	0.0561
1	-0.0889	-0.1300	-0.1305	0.0132	0.0724	0.0231	0.0280
2	-0.0445	-0.0929	-0.2283	0.0066	0.0362	-0.0184	0.0140
Bias in median							
0.25	-0.2817	0.1395	0.1375	0.0124	0.0094	1.8123	0.0659
0.5	-0.1408	0.0227	-0.0158	0.0062	0.0047	0.2525	0.0329
1	-0.0704	-0.0969	-0.0896	0.0031	0.0023	0.0156	0.0165
2	-0.0352	-0.1368	-0.1221	0.0015	0.0012	-0.0124	0.0082
IQR <sub>(50)</sub>							
0.25	0.5570	0.5400	0.5433	0.5124	0.5537	3.7472	0.4855
0.5	0.2785	0.2477	0.2379	0.2562	0.2769	0.4552	0.2427
1	0.1392	0.1706	0.1762	0.1281	0.1384	0.1144	0.1214
2	0.0696	0.1883	0.2011	0.0640	0.0692	0.0555	0.0607
SD							
0.25	0.4555	0.5363	0.6364	0.6777	3.3583	33.1260	0.6765
0.5	0.2277	0.2404	0.2829	0.3388	1.6792	1.0452	0.3382
1	0.1139	0.1621	0.3124	0.1694	0.8396	0.1192	0.1691
2	0.0569	0.3113	0.2092	0.0847	0.4198	0.0521	0.0846

<sup>a</sup> 500 replications were used to calculate the Monte Carlo numerical summaries.

Table 10

Characteristics of the empirical distribution of estimators of the cointegrating vector.<sup>a</sup>

DGP:  $y_t - \beta x_t = z_t$ ,  $z_t = \rho z_{t-1} + e_{z_t}$ ,  $e_{z_t} = (1/\sigma)\theta e_{w_t} + (1 - \theta^2)^{1/2} e_{2t}$   
 $a_1 y_t - a_2 x_t = w_t$ ,  $w_t = w_{t-1} + e_{w_t}$ ,  $e_{w_t} = \sigma e_{1t}$ ,  $e_{it} \equiv t$  student (3df),  $i = 1, 2$

Parameters:  $\beta = 1$ ,  $a_1 = 0$ ,  $a_2 = -1$ ,  $\rho = 0.8$ ,  $\theta = -0.5$ ,  $T = 100$ 

$\sigma$	OLS	NLS (0)	NLS (4)	MLECM (0)	MLECM (4)	PC	CC
Bias in mean							
0.25	-0.3190	0.2074	0.2215	0.0693	0.0303	2.2794	0.1176
0.5	-0.1595	-0.0108	-0.0001	0.0347	0.0152	0.4851	0.0588
1	-0.0798	-0.1196	-0.0905	0.0173	0.0076	0.0256	0.0294
2	-0.0399	-0.2455	-0.1354	0.0087	0.0038	-0.0151	0.0147
Bias in median							
0.25	-0.2378	0.1410	0.1281	0.0235	0.0138	1.7760	0.0743
0.5	-0.1189	0.0013	-0.0016	0.0117	0.0069	0.2376	0.0371
1	-0.0594	-0.0870	-0.0718	0.0059	0.0034	0.0167	0.0186
2	-0.0297	-0.1385	-0.1117	0.0029	0.0017	-0.0114	0.0093
IQR <sub>(50)</sub>							
0.25	0.4868	0.4693	0.4961	0.4147	0.5337	3.5026	0.4522
0.5	0.2434	0.2089	0.2358	0.2074	0.2669	0.4607	0.2261
1	0.1217	0.1500	0.1570	0.1037	0.1334	0.0942	0.1130
2	0.0608	0.1858	0.1730	0.0518	0.0667	0.0494	0.0565
SD							
0.25	0.4360	0.4962	0.5683	0.6281	1.3282	16.8231	0.5766
0.5	0.2180	0.2282	0.2555	0.3141	0.6641	1.4542	0.2883
1	0.1090	0.1601	0.3728	0.1570	0.3321	0.1158	0.1441
2	0.0545	1.0003	1.1177	0.0785	0.1660	0.0514	0.0721

<sup>a</sup>500 replications were used to calculate the Monte Carlo numerical summaries.

#### 4. Conclusions

The objective of the paper was to address the question of how best to proceed in the estimation of a cointegrated system in empirical research. To answer this question we have to recognize three elements in any cointegrated system. First the existence of unit roots, second the multivariate aspect, and third the dynamics. Not taking these elements into account may create problems in estimation. In general the coefficient estimates will be biased in mean and median as well as inefficient. The distribution will be nonsymmetric and nonstandard, and there will be nuisance parameter dependencies.

Table 11

Characteristics of the empirical distribution of estimators of the cointegrating vector.<sup>a, b</sup>

$$\text{DGP: } y_t - \beta x_t = z_t, \quad z_t = \rho z_{t-1} + e_{z_t}, \quad e_{z_t} = e_{1t}((1 - \lambda_1) + \lambda_1 e_{z_{t-1}}^2)^{1/2}, \quad \text{ARCH} \\ a_1 y_t - a_2 x_t = w_t, \quad w_t = w_{t-1} + e_{w_t}, \quad e_{w_t} = \theta e_{z_t} + (1 - \theta^2)^{1/2} e_{2t}, \quad e_{it} \equiv N(0, 1)$$

Parameters:  $\beta = 1$ ,  $a_1 = 0$ ,  $a_2 = -1$ ,  $\rho = 0.8$ ,  $\theta = -0.5$ , and  $T = 100$ 

$\lambda_1$	OLS	NLS (0)	NLS (4)	MLECM (0)	MLECM (4)	PC	CC
Bias in mean							
0.3	-0.0657	-0.0895	-0.0437	-0.0198	0.0304	-0.0052	-0.0374
0.5	-0.0637	-0.0356	-0.0819	-0.0377	-0.0087	-0.0065	0.9080
0.7	-0.0587	-0.0925	-0.0862	-0.0235	0.0101	-0.0077	-0.0018
Bias in median							
0.3	-0.0482	-0.0229	-0.0180	-0.0045	-0.0053	-0.0155	-0.0181
0.5	-0.0446	-0.0200	-0.0168	-0.0046	-0.0055	-0.0162	-0.0159
0.7	-0.0390	-0.0213	-0.0161	-0.0029	-0.0055	-0.0134	-0.0140
IQR <sub>(50)</sub>							
0.3	0.0987	0.1198	0.1267	0.0977	0.1090	0.0935	0.0938
0.5	0.0991	0.1167	0.1255	0.0987	0.1026	0.0945	0.0969
0.7	0.0934	0.1199	0.1316	0.0919	0.1014	0.0887	0.0948
SD							
0.3	0.1007	0.8515	0.2735	0.2973	0.9565	0.1299	0.4342
0.5	0.0991	0.3963	0.6054	0.5530	0.3441	0.1279	22.4588
0.7	0.0936	0.9217	0.8750	0.5163	0.3659	0.1058	0.1975

<sup>a</sup> 500 replications were used to calculate the Monte Carlo numerical summaries.

$$^b \{E[e_{z_t}^2]\}^{-2} E[e_{z_t}^4] = \begin{cases} 3(1 - 3\lambda_1^2)^{-1}(1 - \lambda_1^2) & \text{if } 3\lambda_1^2 < 1 \\ \infty & \text{otherwise} \end{cases}$$

The method we are looking for should have the following characteristics:

- Incorporate all prior knowledge about the presence of unit roots; this eliminates the median bias, the nonsymmetry, part of the nuisance parameter dependencies, and increases efficiency.
- Full system estimation; this eliminates the simultaneous equation bias and increases efficiency.
- Flexible enough to capture the dynamics of the system.

Of the five methods we have compared (OLS, NLS, ML in an ECM, PC, and CC) only maximum likelihood in an error correction model satisfies these requirements. As it is shown in section 2 and in Phillips (1991), this approach ensures that coefficient estimates are symmetrically distributed and median

unbiased, and that hypothesis tests may be conducted using standard asymptotic chi-squared tests. None of the other methods analyzed offer these properties (see table 3).

Although the above properties are based on asymptotic theory, the paper shows, via a Monte Carlo study, that this conclusion is still valid for finite samples. ML in an ECM (Johansen's procedure) performs better than single-equation methods (OLS and NLS) and multivariate methods (PC and CC), even when the errors are nonnormal distributed or when the dynamics are unknown and we overparametrize by including additional lags in the ECM.

This paper suggests that further research on the estimation of a cointegrated system should proceed in the direction of maximum likelihood in a fully specified error correction model. In particular, first it should be investigated whether the same conclusion is obtained when we analyze a large set of variables, and second, which kind of solution can be offered when not enough degrees of freedom to choose the right number of lags are available.

#### Appendix: Proof of the asymptotic results

Define

$$\varepsilon_t = \begin{pmatrix} z_t \\ e_{x_t} \end{pmatrix}.$$

The partial sum process constructed from  $\{\varepsilon_t\}$  satisfies the multivariate invariance principle [see Theorem 23.1 of Billingsley (1968)],

$$X_T(r) = T^{-1/2} \sum_{t=1}^{[Tr]} \varepsilon_t \Rightarrow B(r) \quad \text{for } r \in [0, 1] \quad \text{and } T \rightarrow \infty, \quad (\text{A.1})$$

where

$$B(r) = \begin{pmatrix} B_1(r) \\ B_2(r) \end{pmatrix} \equiv BM(\Omega).$$

The covariance matrix of the Brownian motion is

$$\Omega = 2\pi f_{\varepsilon\varepsilon}(0) = \lim_{T \rightarrow \infty} T^{-1} \mathbf{E} \left\{ \left( \sum_{t=1}^T \varepsilon_t \right) \left( \sum_{t=1}^T \varepsilon_t \right)' \right\},$$

which reduces to

$$\begin{aligned}\Omega &= \lim_{T \rightarrow \infty} T^{-1} \mathbb{E} \left\{ \sum_{t=1}^T \varepsilon_t \varepsilon_t' + \sum_{t=2}^{T-1} \left( \sum_{s=1}^{t-1} \varepsilon_s \right) \varepsilon_t' + \sum_{t=2}^{T-1} \varepsilon_t \left( \sum_{s=1}^{t-1} \varepsilon_s \right)' \right\} \quad (\text{A.2}) \\ &= \mathbb{E}(\varepsilon_0 \varepsilon_0') + \sum_{k=1}^{\infty} \mathbb{E}(\varepsilon_0 \varepsilon_k') + \sum_{k=1}^{\infty} \mathbb{E}(\varepsilon_k \varepsilon_0') = \Omega_0 + \Omega_1 + \Omega_1' .\end{aligned}$$

For the DGP(1), we have

$$\Omega_0 = \begin{bmatrix} \frac{\sigma_1^2}{1-\rho^2} & \theta \sigma_1 \sigma_2 \\ \theta \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}, \quad \Omega_1' = \begin{bmatrix} \frac{\rho \sigma_1^2}{(1-\rho)(1-\rho^2)} & \frac{\rho}{(1-\rho)} \theta \sigma_1 \sigma_2 \\ 0 & 0 \end{bmatrix}. \quad (\text{A.3})$$

Therefore

$$\Omega = \begin{bmatrix} \frac{\sigma_1^2}{(1-\rho)^2} & \theta \sigma_1 \sigma_2 \left( \frac{1}{1-\rho} \right) \\ \theta \sigma_1 \sigma_2 \left( \frac{1}{1-\rho} \right) & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}. \quad (\text{A.4})$$

*Proof of the a.d. of  $\hat{\beta}_{\text{ols}}$ :*

$$T(\hat{\beta}_{\text{ols}} - \beta) = \left( T^{-2} \sum_{t=1}^T x_t^2 \right)^{-1} \left( T^{-1} \sum_{t=1}^T x_t z_t \right). \quad (\text{A.5})$$

By applying the invariance principle (A.1) it can be shown [see Phillips and Durlauf (1986)] that

$$T(\hat{\beta}_{\text{ols}} - \beta) \Rightarrow (\int B_2^2)^{-1} (\int B_2 dB_1 + \lambda), \quad (\text{A.6})$$

where

$$\lambda = \sum_{k=0}^{\infty} \mathbb{E}(\varepsilon_{20} \varepsilon_{1k}') = (\Omega_0 + \Omega_1)_{21} = \left( \frac{1}{1-\rho} \right) \theta \sigma_1 \sigma_2. \quad (\text{A.7})$$



$B_1$  can be decomposed into the sum of two independent Brownian motions,

$$B_1 \equiv \omega_{12}\omega_{22}^{-1}B_2 + d_{11}W, \quad (\text{A.8})$$

where

$$d_{11} = (\omega_{11} - \omega_{12}^2\omega_{22}^{-1})^{1/2} = \frac{\sigma_1}{1-\rho}(1-\theta^2)^{1/2}.$$

The a.d. of OLS follows from substituting (A.7), (A.8), and  $\alpha_1 = \beta + \theta(\sigma_1/\sigma_2)$  into (A.6). ■

*Proof of the a.d. of  $\hat{\beta}_{\text{nls}}$ :* The NLS estimator solves

$$\min_{\{\hat{\beta}\}} \sum_{t=2}^T [\Delta y_t - \gamma_1(y_{t-1} - \beta x_{t-1})]^2. \quad (\text{A.9})$$

From the first-order conditions,

$$T(\hat{\beta}_{\text{nls}} - \beta) = \left( T^{-2} \sum_{t=2}^T x_t^2 \right)^{-1} \left( T^{-1} \sum_{t=2}^T (\Delta y_t - \hat{\gamma}_1 z_{t-1}) x_{t-1} / \hat{\gamma}_1 \right). \quad (\text{A.10})$$

Since  $\Delta y_t = \gamma_1 z_{t-1} + u_{1t}$ , with  $\gamma_1 = (\rho - 1)$  and  $u_{1t} = \beta e_{x_t} + e_{z_t}$ , we may write part of the numerator in (A.10) as

$$(\Delta y_t - \hat{\gamma}_1 z_{t-1}) x_{t-1} = [(\gamma_1 - \hat{\gamma}_1) z_{t-1} x_{t-1} + u_{1t} x_{t-1}].$$

Noting that  $\hat{\gamma}_1$  is a consistent estimator of  $\gamma_1$  and that

$$\text{var} \left( T^{-1} \sum_{t=2}^T z_{t-1} x_{t-1} \right) = O(1),$$

then

$$T(\hat{\beta}_{\text{nls}} - \beta) \equiv \left( T^{-2} \sum_{t=2}^T x_t^2 \right)^{-1} \left( T^{-1} \sum_{t=2}^T (\beta e_{x_t} + e_{z_t}) x_{t-1} / \gamma_1 \right). \quad (\text{A.11})$$

From the invariance principle (A.1),

$$\begin{aligned} T(\hat{\beta}_{\text{nls}} - \beta) \Rightarrow & -(\int B_2^2)^{-1} \{ \beta \int B_2 dB_2 - \theta \sigma_1 \sigma_2 + (1 - \rho) \lambda \\ & + (1 - \rho) \int B_2 dB_1 \} / \gamma_1. \end{aligned} \quad (\text{A.12})$$

The a.d. of NLS follows from substituting (A.7), (A.8), and  $\gamma_1 = (\rho - 1)$  into (A.12). ■

*Proof of the a.d. of  $\hat{\beta}_{\text{mlec}}$ :* From the DGP(1) we derive the expression (ECM),

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} (1 - \beta) \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}, \quad (\text{A.13})$$

where

$$v_{1t} = z_t + \beta e_{x_t} \quad \text{and} \quad v_{2t} = e_{x_t}.$$

By (A.1), the following multivariate invariance principle holds:

$$T^{-1/2} \sum_{t=1}^{[Tr]} v_t \Rightarrow S(r) \equiv BM(\Psi), \quad (\text{A.14})$$

where

$$\Psi = F \Omega F' \quad \text{with} \quad F = \begin{bmatrix} 1 & \beta \\ 0 & 1 \end{bmatrix}.$$

Then

$$\begin{aligned} \Psi &= \begin{bmatrix} \frac{\sigma_1^2}{1-\rho^2} + 2\beta\theta\sigma_1\sigma_2\left(\frac{1}{1-\rho}\right) + \beta^2\sigma_2^2 & \theta\sigma_1\sigma_2\left(\frac{1}{1-\rho}\right) + \beta\sigma_2^2 \\ \theta\sigma_1\sigma_2\left(\frac{1}{1-\rho}\right) + \beta\sigma_2^2 & \sigma_2^2 \end{bmatrix} \\ &= \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}. \end{aligned} \quad (\text{A.15})$$

Since the ML estimator of  $\beta$  maximizes

$$L(\tilde{\gamma}, \beta, \Psi) = -\frac{T}{2} \ln |\Psi| - \frac{1}{2} \sum_{t=1}^T (\Delta H_t - \tilde{\gamma} \alpha' H_{t-1})' \Psi^{-1} (\Delta H_t - \tilde{\gamma} \alpha' H_{t-1}), \quad (\text{A.16})$$

with  $\tilde{\gamma} = (-1, 0)$  and  $\alpha = (1, -\beta)'$ , it can be proved [see Phillips (1991)] that

$$T(\hat{\beta}_{\text{MLECM}} - \beta) \Rightarrow (\int S_2^2)^{-1} (\int S_2 dS_{1,2}), \quad (\text{A.17})$$

where

$$S_{1,2} = S_1 - \Psi_{12} \Psi_{22}^{-1} S_2.$$

From (A.14) it is clear that  $S_2 \equiv B_2$  and  $S_1 \equiv B_1 + \beta B_2$ . By (A.8) it is obtained that

$$S_{1,2} \equiv \left( \frac{\sigma_1}{1 - \rho} \right) (1 - \theta^2)^{1/2} W. \quad (\text{A.18})$$

The a.d. of MLECM follows from inserting (A.18) in (A.17). ■

*Proof of the a.d. of  $\hat{\beta}_{\text{pc}}$ :* The method of PC solves the eigenproblem

$$M \hat{p}_i = \hat{\mu}_i \hat{p}_i, \quad i = 1, 2, \quad (\text{A.19})$$

with

$$M = \sum_{t=1}^T H_t H_t'.$$

Ranking the eigenvalues in descending order ( $\hat{\mu}_1 > \hat{\mu}_2$ ), the PC estimator of  $\alpha$  is  $\hat{p}_2$ ;  $\hat{p}_2$  can be decomposed as follows

$$\hat{p}_2 = \alpha b + \alpha_{\perp} c, \quad (\text{A.20})$$

where  $\alpha = (1, -\beta)'$ ,  $\alpha_{\perp} = (\beta, 1)'$ ,  $b = (\alpha' \alpha)^{-1} \alpha' \hat{p}_2$ , and  $c = (\alpha'_{\perp} \alpha_{\perp})^{-1} \alpha'_{\perp} \hat{p}_2$ . From (A.20) we have

$$T(\hat{p}_2 b^{-1} - \alpha) = T \alpha_{\perp} c b^{-1}. \quad (\text{A.21})$$

Since the eigenvectors  $\hat{p}_i$  and eigenvalues  $\hat{\mu}_i$  satisfy

$$S(\hat{\mu}_i) \hat{p}_i = 0, \quad (\text{A.22})$$

where  $S(\hat{\mu}_i) = \hat{\mu}_i I - M$ , it follows from substituting (A.20) into (A.22) and multiplying (A.22) by  $\alpha'_{\perp}$  from the left that

$$\alpha'_{\perp} S(\hat{\mu}_2) \alpha b + \alpha'_{\perp} S(\hat{\mu}_2) \alpha_{\perp} c = 0. \quad (\text{A.23})$$

Thus,

$$cb^{-1} = -(\alpha'_{\perp} S(\hat{\mu}_2) \alpha_{\perp})^{-1} \alpha'_{\perp} S(\hat{\mu}_2) \alpha, \quad (\text{A.24})$$

and inserting (A.24) in (A.21) we obtain

$$T(\hat{p}_2 b^{-1} - \alpha) = -\alpha_{\perp} (T^{-2} \alpha'_{\perp} S(\hat{\mu}_2) \alpha_{\perp})^{-1} (T^{-1} \alpha'_{\perp} S(\hat{\mu}_2) \alpha). \quad (\text{A.25})$$

It can be shown that  $\hat{\mu}_1$  and  $\hat{\mu}_2$  are  $O_p(T^2)$  and  $O_p(T)$ , respectively. Therefore

$$\begin{aligned} T^{-2}(\alpha'_{\perp} S(\hat{\mu}_2) \alpha_{\perp}) &= T^{-2}(\hat{\mu}_2 \alpha'_{\perp} \alpha_{\perp} - \alpha'_{\perp} M \alpha_{\perp}) \\ &= T^{-2} \left\{ \hat{\mu}_2 (\beta^2 + 1) - \left[ \beta^2 \sum_{t=1}^T y_t^2 + 2\beta \sum_{t=1}^T (x_t y_t) + \sum_{t=1}^T x_t^2 \right] \right\} \\ &\Rightarrow -(\beta^2 + 1)^2 \int B_2^2 \end{aligned} \quad (\text{A.26})$$

and

$$\begin{aligned} T^{-1}(\alpha'_{\perp} S(\hat{\mu}_2) \alpha) &= T^{-1}(\hat{\mu}_2 \alpha'_{\perp} \alpha_{\perp} - \alpha'_{\perp} M \alpha) \\ &= -T^{-1} \sum_{t=1}^T (\beta y_t + x_t) z_t \\ &\Rightarrow -\{(1 + \beta^2)(\int B_2 dB_1 + \lambda) + \beta \text{var}(z_t)\}. \end{aligned} \quad (\text{A.27})$$

Substituting (A.26) and (A.27) in (A.25), we have

$$\begin{aligned} T(\hat{p}_2 b^{-1} - \alpha) &\Rightarrow -\alpha_{\perp} \{[(\beta^2 + 1)^2 \int B_2^2]^{-1} [(1 + \beta^2)(\int B_2 dB_1 + \lambda) \\ &\quad + (\beta \text{var}(z_t))]\} \\ &= -\alpha_{\perp} D_1. \end{aligned} \quad (\text{A.28})$$

From (A.20) we may write  $b^{-1} = (1 + \beta^2)/(p_{21} - \beta p_{22})$  with  $\hat{p}_2 = (p_{21}, p_{22})'$ . Then multiplying (A.28) by  $\alpha'_{\perp}$  from the left, it follows that

$$\begin{aligned} \alpha'_{\perp} T(\hat{p}_2 b^{-1} - \alpha) &= T(p_{21} - \beta p_{22})^{-1} (\beta p_{21} + p_{22})(1 + \beta^2) \\ &= T(1 - \beta \hat{\beta}_{\text{pc}})^{-1} (\beta - \hat{\beta}_{\text{pc}})(1 + \beta^2) \\ &\Rightarrow -(1 + \beta^2) D_1, \end{aligned} \quad (\text{A.29})$$

where  $\hat{\beta}_{pc} = -(p_{22}/p_{21})$ . Noting that  $\hat{\beta}_{pc}$  is consistent (PC is equivalent to orthogonal regression), then

$$T(\hat{\beta}_{pc} - \beta) \Rightarrow (1 + \beta^2)D_1 \equiv (\int B_2^2)^{-1}(\int B_2 dB_1 + \lambda) + \frac{\beta}{1 + \beta^2} \text{var}(z_t) \quad (\text{A.30})$$

The a.d. of PC follows from substituting (A.7), (A.8), and the expression of  $\text{var}(z_t)$  in (A.30). ■

*Proof of the a.d. of  $\hat{\beta}_{cc}$ :* The method of CC solves the eigenproblem

$$M_{01}M_{11}^{-1}M_{10}\hat{c}_i = \delta_i M_{00}\hat{c}_i, \quad i = 1, 2, \quad (\text{A.31})$$

where

$$M_{jk} = \sum_{t=1}^T H_{t-j}H'_{t-k}, \quad j, k = 0, 1.$$

Ranking the eigenvalues in descending order, the CC estimator of  $\alpha$  is  $\hat{c}_2$ .

Following (A.20)–(A.25) with  $S(\hat{\delta}_i) = \hat{\delta}_i M_{00} - M_{01}M_{11}^{-1}M_{10}$ , we obtain

$$T(\hat{c}_2 b^{-1} - \alpha) = -\alpha_{\perp} [T^{-2} \alpha'_{\perp} S(\hat{\delta}_2) \alpha_{\perp}]^{-1} [T^{-1} \alpha'_{\perp} S(\hat{\delta}_2) \alpha]. \quad (\text{A.32})$$

The expressions (A.33) and (A.34) are the asymptotic distributions of the elements of the denominator in (A.32),

$$\begin{aligned} T^{-2}(\hat{\delta}_2 \alpha'_{\perp} M_{00} \alpha_{\perp}) &= T^{-2} \hat{\delta}_2 \sum_{t=1}^T (\beta y_t + x_t)^2 \\ &\Rightarrow \hat{\delta}_2 (\beta^2 + 1)^2 \int B_2^2, \end{aligned} \quad (\text{A.33})$$

$$\begin{aligned} T^{-2}(\alpha'_{\perp} M_{01} M_{11}^{-1} M_{10} \alpha_{\perp}) &= T^{-2} [\alpha'_{\perp} M_{01} (T^{1/2} \alpha, \alpha_{\perp})] \\ &\quad \times [(T^{1/2} \alpha', \alpha'_{\perp})' M_{11} (T^{1/2} \alpha, \alpha_{\perp})]^{-1} \\ &\quad \times [(T^{1/2} \alpha', \alpha'_{\perp})' M_{10} \alpha_{\perp}] \\ &\Rightarrow (\beta^2 + 1)^2 \int B_2^2, \end{aligned} \quad (\text{A.34})$$

and (A.35) and (A.36) are the asymptotic distributions of the numerator,

$$\begin{aligned} T^{-1}(\hat{\delta}_2' \alpha_1' M_{00} \alpha) &= T^{-1} \hat{\delta}_2' \sum_{t=1}^T (\beta y_t + x_t) z_t \\ &\Rightarrow \hat{\delta}_2' \{(\beta^2 + 1)(\int B_2 dB_1 + \lambda) + \beta \text{var}(z_t)\} , \end{aligned} \quad (\text{A.35})$$

$$\begin{aligned} T^{-1}(\alpha_1' M_{01} M_{11}^{-1} M_{10} \alpha) &\Rightarrow (1 + \beta^2 \{(\int B_2 dB_1 + \lambda) - \theta \sigma_1 \sigma_2\} \\ &\quad + \{\beta \rho \text{cov}(z_t, z_{t-1})\} . \end{aligned} \quad (\text{A.36})$$

The last line is obtained using the same intermediate step as (A.34). Noting that

$$\hat{\delta}_2 \xrightarrow{P} \delta_2 = (\text{corr}(z_t, z_{t-1}))^2 = \rho^2 ,$$

we have

$$T(\hat{\beta}_{cc} - \beta) \Rightarrow (B_2^2)^{-1} \left\{ \left( \int B_2 dB_1 + \frac{\rho}{1 - \rho} \lambda \right) \right\} , \quad (\text{A.37})$$

from inserting (A.33)–(A.36) in (A.32) and following (A.29).

The a.d. of CC is obtained (A.7) and (A.8) in (A.37). ■

## References

- Ahn, S.K. and G.C. Reinsel, 1990, Estimation for partially nonstationary multivariate autoregressive models, *Journal of the American Statistical Association* 85, 813–823.
- Anderson, T.W., 1951, Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Annals of Mathematical Statistics* 22, 327–351.
- Anderson, T.W., 1976, Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics (with discussion), *Journal of the Royal Statistical Society B* 38, 1–36.
- Aoki, M., 1988, On alternative state space representations of time series models, *Journal of Economic Dynamics and Control* 12, 595–607.
- Banerjee, A., J.J. Dolado, D.F. Hendry, and G.W. Smith, 1986, Exploring equilibrium relationships in econometrics through static models: Some Monte Carlo evidence, *Oxford Bulletin of Economics and Statistics* 48, 253–277.
- Billingsley, P., 1968, *Convergence of probability measures* (Wiley, New York, NY).
- Bossaerts, P., 1988, Common nonstationary components of asset prices, *Journal of Economic Dynamics and Control* 12, 347–364.
- Brillinger, D.R., 1981, *Time series: Data analysis and theory* (Holden-Day, San Francisco, CA).
- Eckart, C. and G. Young, 1936, The approximation of one matrix by another of lower rank, *Psychometrika* 1, 211–218.
- Engle, R.F. and C.W.J. Granger, 1987, Cointegration and error correction: Representation, estimation and testing, *Econometrica* 55, 251–276.

- Engel, R.F. and B.S. Yoo, 1989, Cointegrated economic time series: A survey with new results, Discussion paper 87-26R (University of California, San Diego, CA).
- Gonzalo, J., 1991, Five alternative methods of estimating long run equilibrium relationships, Discussion paper 89-55R (University of California, San Diego, CA).
- Granger, C.W.J., 1983, Co-integrated variables and error-correcting models, Discussion paper 83-13 (University of California, San Diego, CA).
- Hansen, B. and P.C.B. Phillips, 1990, Estimation and inference in models of cointegration: A simulation study, *Advances in Econometrics* 8, 225-248.
- Izenman, A.J., 1980, Assessing dimensionality in multivariate regression, in: P.R. Krishnaiah, ed., *Handbook of statistics 1* (North Holland, New York, NY) 571-592.
- Johansen, S., 1988, Statistical analysis of cointegrating vectors, *Journal of Economic Dynamics and Control* 12, 231-254.
- Johansen, S. and K. Juselius, 1988, Hypothesis testing for cointegration vectors with an application to the demand for money in Denmark and Finland, Discussion paper 8805 (University of Copenhagen, Copenhagen).
- Kinderman, A.J. and J.G. Ramage, 1976, Computer generation of normal random numbers, *Journal of the American Statistical Association* 71, 893-896.
- Malinvaud, E., 1980, *Statistical methods of econometrics* (North-Holland, Amsterdam).
- Phillips, P.C.B., 1988a, Optimal inference in cointegrated systems, *Econometrica* 59, 283-306.
- Phillips, P.C.B., 1988b, Spectral regression for cointegrated time series, in: W. Barnett, ed., *Non-parametric and semiparametric methods in economics and statistics* (Cambridge University Press, Cambridge) 413-436.
- Phillips, P.C.B. and S.N. Durlauf, 1986, Multiple time series regression with integrated process, *Review of Economic Studies* 53, 473-496.
- Phillips, P.C.B. and M. Loretan, 1991, *Estimating long run economic equilibria*, *Review of Economic Studies* 58, 407-436.
- Rao, C.R., 1965, *Linear statistical inference and its applications* (Wiley, New York, NY).
- Robinson, P.M., 1973, Generalized canonical analysis for time series, *Journal of Multivariate Analysis* 3, 141-160.
- Saikkonen, P., 1991, Asymptotically efficient estimation of cointegration regression, *Econometric Theory* 7, 1-21.
- Stock, J.H., 1987, Asymptotic properties of least squares estimators of cointegrating vectors, *Econometrica* 55, 1035-1056.
- Stock, J.H. and M.W. Watson, 1988, Testing for common trends, *Journal of the American Statistical Association* 83, 1097-1107.
- Stock, J.H. and M.W. Watson, 1991, A simple estimator of cointegrating vectors in higher order integrated systems, Manuscript.
- Velu, R.P., G.C. Reinsel, and D.W. Wichern, 1986, Reduced rank models for multiple time series, *Biometrika* 73, 105-118.