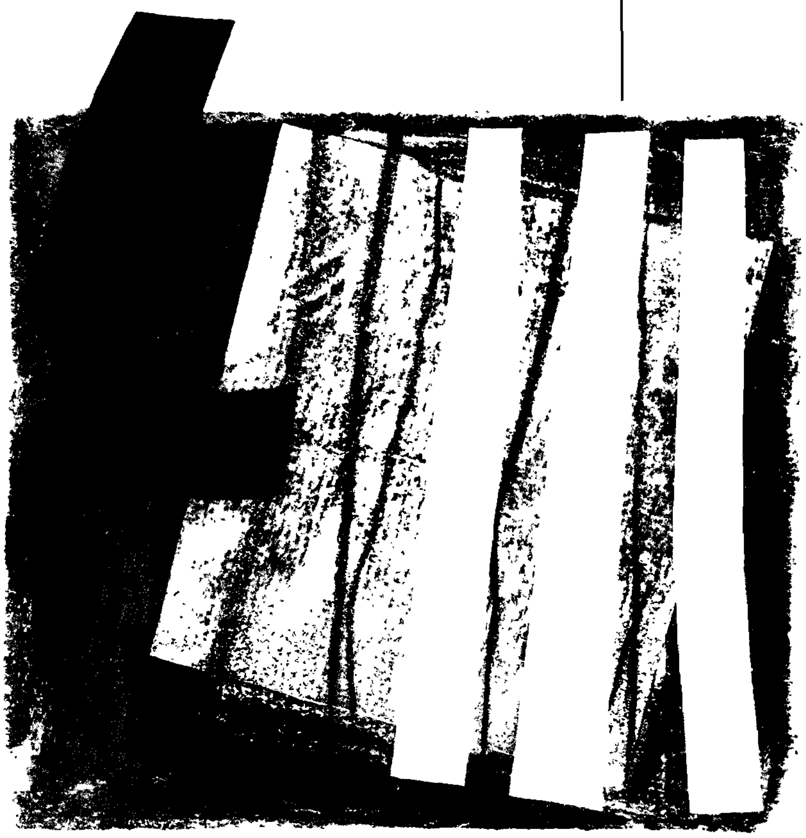
A MODEL-FREE COINTEGRATION APPROACH FOR PAIRS OF I(d) VARIABLES

Felipe M. Aparicio,
Miguel A. Arranz
and Alvaro Escribano

00-59



WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 00-59
Statistics and Econometrics Series 24
July 2000

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

A MODEL-FREE COINTEGRATION APPROACH FOR PAIRS OF I(d) VARIABLES

Felipe M. Aparicio, Miguel A. Arranz, and Alvaro Escribano *

Abstract

In this paper we propose several model free (non parametric) statistics to measure serial dependence that are useful to characterize the short and the long memory properties of series in the time and the frequency domain. Conditions on the joint memory properties of the series such as cointegration are introduced by means of these statistics. We show that the relationship between the nonparametric concept of cointegration and the cross-covariance functions of the series, has a natural interpretation as an instrumental variable estimator. We show that its small sample behavior is better than the usual least squares estimator. Finally, from our characterization it is posibble to discriminate between fractional and integer cointegration.

Keywords: Cointegration, fractional cointegration, instrumental variable (IV) estimation.

* Aparicio, Dept. of Statistics and Econometrics, Univ. Carlos III de Madrid, Spain, E-mail: aparicio@est-econ.uc3m.es; Arranz Dept. of Statistics and Econometrics, Univ. Carlos III de Madrid, Spain, E-mail: arranz@pvi.uc3m.es; Escribano, Dept. of Economics, Univ. Carlos III de Madrid, Spain. E-mail: alvaroe@ eco.uc3m.es. This paper is a revised and extended version of Aparicio & Escribano (1997). Research partially supported by the Spanish DGICYT under grant PB95-0298, and by the European Union under grant TMR ERB-4061 PL97-0994.

1. Introduction

Many time series exhibit permanent shocks that affect the mean and the variance. These series are usually said to be *integrated*, since the most important features in their patterns can be generated by sums of an increasing number of weakly-dependent random variables. The fact that remote shocks have a persistent influence on the levels of these series is known as the *long-memory* property of the series. Integrated series can be expressed in terms of *unobserved components*, where one of the components is a *stochastic trend*. In some cases, these permanent movements in the levels of the series may be strongly correlated across series. When the series exhibit a common long-memory component (common stochastic trend) are said to be *cointegrated*.

The concept of *cointegration* was coined by Granger (1981), and later developed by Engle & Granger (1987) and Johansen (1988) in an error correction context that is and widely applied in macroeconomics and finance. Examples of these series are stock and commodity prices, income and expenditure, and cross-country exchange-rates. It is also possible to find cointegrating relationships among variables in engineering applications such as meteorology and telecommunications. One example could be the output signals from different sensing or processing devices having a nonlinear characteristic response and driven by a common persistent input flow, see Aparicio (1995*b*).

These common stochastic trends that affect cointegrated series allow us to predict their relative long term evolution and to acquire information about the nature of their data generating processes. For example, if exchange-rates and/or stock prices from different countries are found to move together in the long run, this may suggest an integration of international capital markets. Moreover, the finding that commodity prices from different countries are cointegrated means that the Purchasing Power Parity Hypothesis should hold in the long run.

On the engineering side, suppose we are interested in recognizing patterns or features for classification or identification purposes, and suppose that our discretised signals have an overwhelming low-frequency energy content. A problem arises because significant correlations could be detected among this sort of signals even when they are completely unrelated. This problem is known in econometrics as the *spurious regression problem*, see Granger & Newbold (1974). Obviously, the consequences of such false alarms cannot always be neglected, especially when the recovery of an individual under medical treatment, or the security of a population against natural disasters depend on it. Cointegration tests are aimed at solving the spurious regression problem, and thus can tell us when a pair of signals are really related in the long run.

Underlying the cointegration concept is the idea of a long run *equilibrium* (i.e. a deterministic relationship that holds on the average) between two integrated (or long-memory) variables, x_t, y_t.

A strict (linear) equilibrium exists when for some $\alpha \neq 0$, one has $y_t = \alpha x_t$. This unrealistic deterministic situation is replaced in practice by that of cointegration, where the stochastic equilibrium error $z_t = y_t - \alpha x_t$ can be different from zero but must fluctuate around the mean value much more frequently than the individual series (this behavior the equilibrium errors is called *short-memory*).

The standard definitions and tests of cointegration assume a certain type of data generating mechanism for both the individual series and their relationship and the mechanism is always linear. As a consequence, cointegration is usually interpreted as a uniform tendency for the series to move towards a unique long-run equilibrium. This is in contrast with the multiple equilibria observed in many pairs of economic variables (see for example Escribano (1986) and Escribano & Granger (1998)).

Dynamic general equilibrium models are usually nonlinear by construction. Other nonlinear relationships are for example derived from consumer and production theory, and can take the form of asymmetries in adjustment costs or convexities in production as in Escribano & Pfann (1998). This theory suggests that economic agents will adjust continuously only as far as their adjustment benefits exceed their costs (Balke & Fomby, 1997). For example, in financial markets, transaction costs allow for a band to appear in which returns can diverge, thereby introducing inefficiencies and the possibility of arbitrage. Policy interventions such as exchange-rate management via the central banks, and commodity price stabilization through government intervention by buying or selling stocks, may also induce non-uniformities in the adjustment of agents. These non-uniformities translate into departures from the linear cointegration hypothesis, see Escribano (1986) for an example based on the money demand of the U.K.

Nonlinear long-memory time series and nonlinear cointegrating relationships are present in many engineering applications as well. This sort of time series appears, for example, when we analyze and compare the sampled output signals from inherently nonlinear sensing or processing devices in response to persistent inputs. Aparicio (1993, 1995*a*) discusses a few applications of the analysis of time series with these features obtained from the output of wind speed sensors. Here the nonlinear characteristic response of each sensor modulates a nonstationary wind speed flow. The problem of estimating cointegrating relationships arises naturally when comparing the performances of these sensors driven by the same input wind flow. Similar problems occur in communications engineering when dealing with time series from signals transmitted through a distorting communication channel, many of which has long-range correlations, as reported for example in Mandelbrot (1965, 1967), Leland et al. (1994), Beran et al. (1991), and Willinger et al. (1995).

It is important to introduce model free (nonparametric) measures of cointegration since that opens the opportunity to extend the concept of cointegration to a nonlinear context, see for example the nonlinear information-theoretic measures proposed in Aparicio & Escribano (1999). To reach this goal. we first discuss general concepts of *short–memory*, *long–memory*, and *cointegration*, that lead us to introduce a new characterization of cointegration.

The structure of the paper is as follows. Section 2 introduces several measures of linear dependence (correlation) in the context of time series that are integrated of order d, $I(d)$, where d can be any real or integer number. In section 3, a new model–free measure of linear cointegration in the time domain is introduced. We find the relationship with the usual concepts for $I(1)$ and long memory series (fractional cointegration). We show that it has a natural interpretation as an instrumental variable estimator where the instrument is a lagged explanatory variable. Section 4 includes a small sample analysis of the instrumental variable estimator based on Monte Carlo simulation experiments with $I(1)$ and fractionally integrated variables. In section 5, the concept of cointegration is analyzed in the frequency domain. The relationship between the cointegration concepts defined in the frequency domain and in the time domain is explained in section 6, based on analytical results and some Monte Carlo simulation experiments. Finally, section 7 includes some concluding remarks and topics for further research.

2. Definitions of memory in time series

The usual characterization of integrated time series is in terms of *ARIMA* models.

Definition 1. A time series x_t is said to be $ARIMA(p,d,q)$. where $d \in \Re$, if after being differenced d times. it has a stationary and invertible $ARMA(p,q)$ representation, where p, q are nonnegative integers.

Thus if $x_t \sim ARIMA(p,d,q)$ there exists polynomials $\Phi(B)$ and $\Theta(B)$ in the backward shift operator B, of order $p \geq 0$ and $q \geq 0$ respectively, with all roots outside the unit circle and no factors in common, such that we can write

$$\Phi(B)(1-B)^d x_t = \Theta(B)\epsilon_t, \tag{2.1}$$

where ϵ_t is generally assumed to be a sequence of zero-mean, independent and identically Normally-distributed errors. Let Δ be the difference operator, and thus let $\Delta^d x_t = (1-B)^d x_t$. Following

Hosking (1981), when d is not an integer we can write :

$$\begin{aligned}\Delta^d x_t &= \sum_{k=0}^{\infty} \frac{d!}{k!(d-k)!}(-B)^k x_t \\ &= x_t - dx_{t-1} - \frac{1}{2}d(1-d)x_{t-2} - \frac{1}{6}d(1-d)(2-d)x_{t-3} - \ldots \end{aligned} \tag{2.2}$$

When the parameter d is positive, it is sometimes referred to as the *long-memory parameter*, and it determines the rate of decay of the serial dependence of x_t with increasing lag. That is, if $d > 0$ the process x_t has long memory, while it has short-memory when $d = 0$. For $d < 0$ x_t is often called *anti-persistent*. Moreover, if $d < \frac{1}{2}$ then x_t is stationary, while it is nonstationary for $d \geq \frac{1}{2}$, see Granger & Joyeux (1980). Finally, only when $d < 1$ is x_t mean-reverting.

It is also known that if x_t is Gaussian and short-memory ($d = 0$) then its autocorrelation function (ACF) converges to zero at an *exponential* rate as τ grows to infinity, see Box & Jenkins (1970), while this rate is *hyperbolical* for $0 < d < 1$, see Granger & Joyeux (1980).

In the linear stationary case the standard measures of serial dependence for a given time series x_t with mean $E(x_t) = \mu_x$ are based on its autocovariance function, say $cov(x_t, x_{t-\tau})$, defined as

$$cov(x_t, x_{t-\tau}) = E\big[(x_t - \mu_x)(x_{t-\tau} - \mu_x)\big]. \tag{2.3}$$

In dealing with integrated time series, we allow both the autocovariance function and the variance of the process to depend on when the stochastic processes are initiated at a finite value at time $t = 0$, say $x_0 = 0$. Notice that in this case the variance of a non stationary process is finite only for finite t. Therefore, from now on all the expectations considered in this paper, $E(x_t) = E(x_t|x_0)$, $cov(x_t, x_{t-\tau}) = E\big[(x_t - \mu_x)(x_{t-\tau} - \mu_x)|x_0\big]$, and $cov(x_t, y_{t-\tau}) = E[(y_t - \mu_{y_t})(x_{t-\tau} - \mu_{x_{t-\tau}})|x_0, y_0]$, are all conditional on some starting values at $t = 0$, say $x_0 = 0$, and $y_0 = 0$.

Definition 2. A stochastic process x_t is said to be **short-memory** [1] if there exists a finite and positive real number b such that $\lim_{T\to\infty} \sum_{\tau=1}^{T} |cov(x_t, x_{t-\tau})| = b$.

Definition 3. A stochastic process x_t is said to be **long-memory** if $\lim_{T\to\infty} \sum_{\tau=1}^{T} |cov(x_t, x_{t-\tau})| = \infty$.

Based on these definitions, it is natural to define the following model–free concept of integration:

[1] This concept of short-memory is directly related to that of asymptotic uncorrelation. See, for example, White (1984), and Escribano (1987).

Definition 4. A stochastic process x_t is said to be **integrated of order d**, (in short $x_t \sim I(d)$) if it is long-memory and d is the smallest real number such that $(1-B)^d x_t$ is short-memory.

Notice that the time series generated by $ARIMA(p,d,q)$ models, defined in equation (2.1) satisfy this definition of $I(d)$ process. In the following section we extend this approach to the concept of cointegration.

3. Linear cointegration in the time domain

Let x_t, y_t be two zero-mean integrated time series of orders $d_x, d_y \in \Re^+$, respectively. In short, $x_t \sim I(d_x)$, $y_t \sim I(d_y)$, which means that $(1-B)^{d_x} x_t = \epsilon_t$, $(1-B)^{d_y} y_t = \xi_t$, where ϵ_t, ξ_t are *short-memory* series. Let $z_t = y_t - \alpha x_t$, for some generic nonzero real number α.

Definition 5. (Granger, 1981) Two $I(d)$ time series x_t, y_t, with $d > 0$, are said to be (linearly) [2] **cointegrated** if $\exists\, \alpha \in \Re - \{0\}$ such that the series $z_t = y_t - \alpha\, x_t$ is $I(d_z)$ with $d_z < d$.

Remarks:

- When x_t, y_t are cointegrated z_t is zero mean and $I(0)$, and y_t, and x_t tend to move jointly in the long-run, even though their short-run movements may not be "aligned".
- From the economic point of view, a most important case is when $d = 1$, $d_z = 0$, since this situation can be interpreted as the existence of a (linear) long-run equilibrium for the series. However, the previous definition of cointegration does not imply the existence of an equilibrium between the two $I(d)$ series ($d > 0$), since for the former we need that their cointegration residuals z_t be $I(d_z)$ with $d_z < \min(1,d)$. However, an observable equilibrium requires that z_t be mean-reverting ($d_z < 1$).

In the sequel, we propose an alternative characterization of linear cointegration. We restrict our discussion here to the non-trivial case where all series are mutually dependent. Let x_t, y_t denote two $I(d)$ time series with $d > 0$, initialized at $t = 0$ at the values $x_0 = 0$ and $y_0 = 0$, respectively. Let $cov(x_t, y_{t-\tau})$ represent the *cross-covariance function* of x_t, y_t conditional on x_0, y_0 for $t \geq \tau$, defined as

$$cov(y_t, x_{t-\tau}) = E\big[(y_t - \mu_{y_t})(x_{t-\tau} - \mu_{x_{t-\tau}})\big] \qquad (3.1)$$

where $\mu_{x_t} = E(x_t)$ and $\mu_{y_t} = E(y_t)$ are the expectations of x_t and y_t conditional on $x_0 = 0$, and $y_0 = 0$ respectively.

We propose the following model–free characterization of cointegration.

[2] In Granger (1981) there is no explicit mention to the term *linear*.

Theorem 1. *A pair of $I(d)$ time series x_t, y_t are* ***(linearly) cointegrated*** *with cointegrating vector $\beta' = (1, -\alpha)$ if, for some fixed $\tau < t$ or for $\tau = o(t)$*

$$\lim_{\tau \to \infty} \frac{cov(y_t, x_{t-\tau})}{cov(x_t, x_{t-\tau})} = \alpha. \tag{3.2}$$

Proof. In Appendix A. □

Corollary 1. *Let y_t, x_t be $I(d)$, and linearly cointegrated with cointegrating vector $\beta' = (1, -\alpha)$, with d restricted to the interval $(0, 1]$. If $\lim_{\tau\to\infty} \frac{cov(y_t, x_{t-\tau})}{var(x_{t-\tau})} = \alpha$ with $\alpha \neq 0$, then $x_t, y_t \sim I(1)$*

Proof. In Appendix A. □

Remarks:

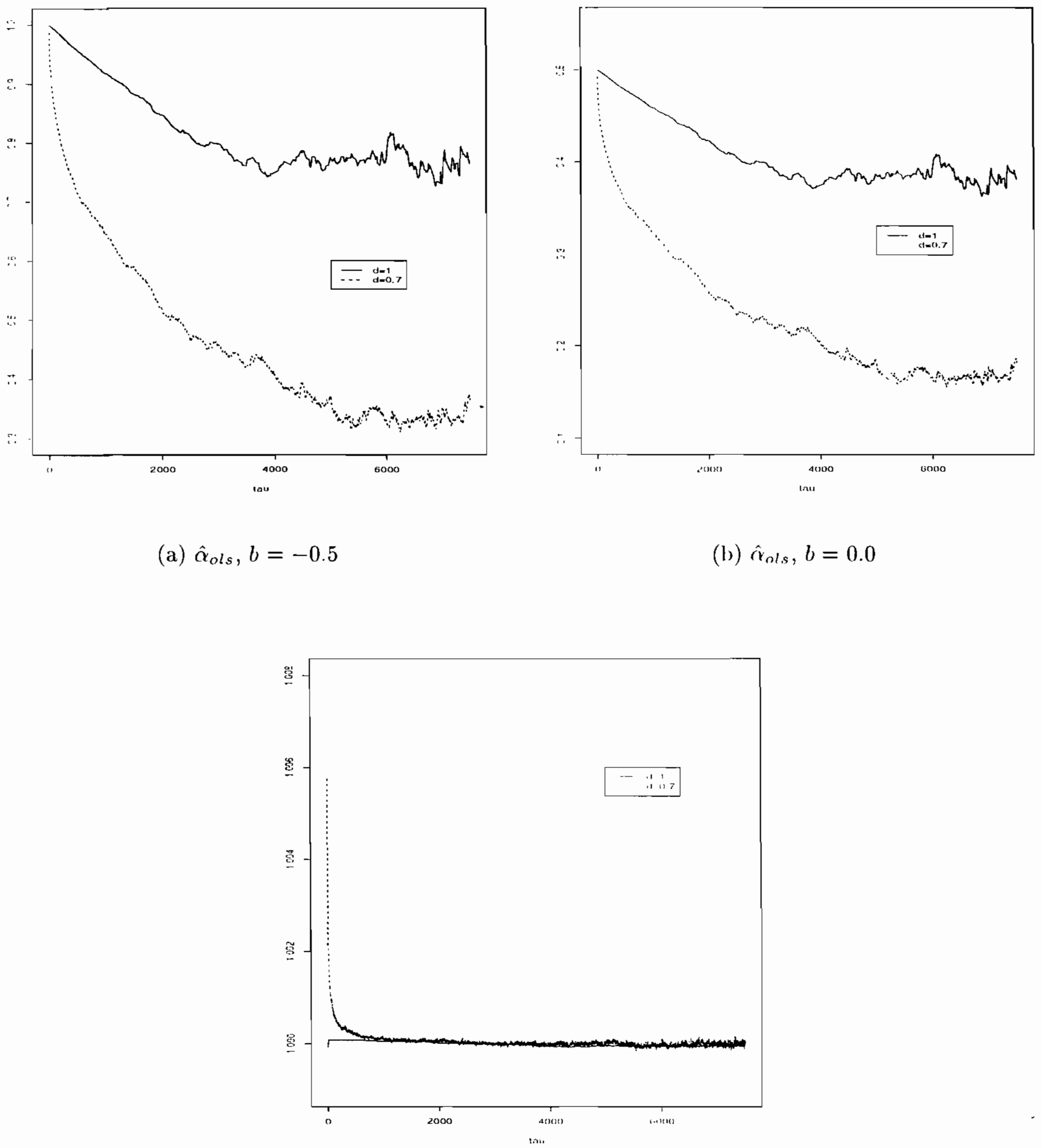
- Intuitively, Theorem 1 states that, under cointegration, the remote past of x_t should be as useful as the remote past of y_t in long-term forecasting y_t.
- Theorem 1 implies that the rates of convergence of $cov(y_t, x_{t-\tau})$ and of $cov(x_t, x_{t-\tau})$ as τ increases without bound, should be the same. For example, suppose $cov(y_t, x_{t-\tau}) \sim \alpha\, \tau^{-b}\, var(x_{t-\tau})$ and that $cov(x_t, x_{t-\tau}) \sim \tau^{-a}\, var(x_{t-\tau})$ for large τ. In general, we expect $a \leq b$, but equality should hold under linear cointegration. In a preliminary analysis, a plot of $cov(y_t, x_{t-\tau})/cov(x_t, x_{t-\tau}) \sim \alpha\, \tau^{a-b}$ versus τ should help in identifying the existence of a linear cointegrating relationship ($a = b$). Figure 1(c) shows that this is true even for small values of τ.
- If y_t, x_t have zero means, the ratio $\frac{cov(y_t, x_{t-\tau})}{cov(x_t, x_{t-\tau})} = \frac{E(y_t, x_{t-\tau})}{E(x_t, x_{t-\tau})}$, and provides an instrumental variable (IV) estimator of the cointegration parameter, where the instrument is $x_{t-\tau}$. That is, consider the linear cointegration relationship

$$y_t = \alpha\, x_t + z_t, \tag{3.3}$$

where x_t, y_t are $I(d)$ with $d > 0$ and z_t is $I(0)$, then the IV estimator of α is given by

$$\hat{\alpha}_{iv} = \frac{\sum_{t=\tau+1}^{N} y_t x_{t-\tau}}{\sum_{t=\tau+1}^{N} x_t x_{t-\tau}}, \tag{3.4}$$

with N denoting the available sample size. Figure 1(c) reports the quick convergence of $\hat{\alpha}_{iv}$ even for small values of τ. Furthermore, Engle & Granger (1987) and Stock (1987) for $I(1)$, and Dolado & Marmol (1998) for $I(d)$ where d is fractional, show that α can be consistently



(a) $\hat{\alpha}_{ols}$, $b = -0.5$

(b) $\hat{\alpha}_{ols}$, $b = 0.0$

(c) $\hat{\alpha}_{ivc}(\tau)$, $b = -0.5$

FIGURE 1. Simulation examples of $\hat{\alpha}_{ols}(\tau)$, equation (3.5) and $\hat{\alpha}_{iv}(\tau)$, equation (3.4) $a = 0.5$, $s = 1$, sample size 10000, averaging over 500 replications.

estimated by OLS from the regression (3.3), where

$$\hat{\alpha}_{ols} = \frac{\sum_{t=1}^{N} y_t x_t}{\sum_{t=1}^{N} x_t^2}. \tag{3.5}$$

- The limit condition in Theorem 1 cannot be checked in practice unless τ is finite. However, when looking for linear cointegration in empirical applications, this condition is generally satisfied for small values of τ as will be seen in Section 4. An important example occurs when $y_t = \alpha\, x_t + z_t$, with $\alpha \neq 0$, and where x_t, y_t are $I(1)$ and u_t is a sequence of *i.i.d. r.v.*'s and independent of x_t. In this case, $cov(y_t, x_{t-\tau})/cov(x_t, x_{t-\tau}) = \alpha$ for any $\tau < t$.
- If the series are short-memory but mutually correlated then the ratio of the cross-covariance function to the autocovariance function could also eventually converge to a nonzero value, and therefore this is the reason for imposing in Theorem 1 that the individual series are long-memory.
- The expression of Corollary 1 can be estimated by OLS running the regression of y_t on $x_{t-\tau}$.

The following examples show that Theorem 1 is not satisfied by pairs of non-cointegrated time series (Example 1), while it holds for cointegrated variables generated by a linear *common factor model* (Example 2).

Example 1: Consider the following pair of non-cointegrated series:

$$x_t = w_t + \xi_t \tag{3.6}$$

$$y_t = q_t + v_t \tag{3.7}$$

where w_t, q_t are mutually independent $I(d)$ series, and where ξ_t, v_t are series from $ARMA(p,q)$ processes with possibly different AR and MA orders, and independent of w_t and q_t, respectively. Then it is straightforward to see that

$$cov(y_t, x_{t-\tau}) = cov(v_t, \xi_{t-\tau}). \tag{3.8}$$

Now, the covariance $cov(v_t, \xi_{t-\tau})$ will tail off exponentially as τ grows to infinity since the series v_t, ξ_t are both $I(0)$. On the contrary, $cov(x_t, x_{t-\tau})$ will grow without bound if $d = 1$, that is $cov(x_t, x_{t-\tau}) = (t-\tau)\sigma_\epsilon^2$, or decay hyperbolically with growing τ for $\tau = o(t)$. In both cases, we will have

$$\lim_{t\to\infty} \frac{cov(y_t, x_{t-\tau})}{cov(x_t, x_{t-\tau})} = 0. \tag{3.9}$$

Example 2: Consider the following linear common factor model:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \alpha \\ 1 \end{pmatrix} w_t + \begin{pmatrix} v_t \\ \xi_t \end{pmatrix} \tag{3.10}$$

with $\alpha \neq 0$ and where $w_t = w_{t-1} + \epsilon_t$ and (v_t, ξ_t, ϵ_t) are independent sequences of independent and identically Normally distributed *r.v.*'s with zero mean and joint covariance matrix equal to the identity matrix. Let $\beta'_{\perp} = (\alpha, 1)$, where $\beta'_{\perp}$ is the transpose of $\beta_{\perp}$. Thus the orthogonal complement of $\beta'_{\perp}$ is $\beta' = (1, -\alpha)$. The cointegrating relationship is therefore obtained as

$$z_t = \beta' \begin{pmatrix} y_t \\ x_t \end{pmatrix} = y_t - \alpha\, x_t = v_t - \alpha\, \xi_t, \tag{3.11}$$

which shows that z_t is $I(0)$.

From equation 2, the autocovariance of x_t is

$$cov(x_t, x_{t-\tau}) = (t - \tau)\sigma_\epsilon^2 \tag{3.12}$$

while the cross-covariance of y_t and x_t is given by

$$cov(y_t, x_{t-\tau}) = \alpha(t - \tau)\sigma_\epsilon^2. \tag{3.13}$$

Clearly. their ratio is α, and since $\alpha \neq 0$ by assumption, we conclude that the series y_t, x_t are linearly cointegrated.

In the following section we provide small sample evidence of the improvements obtained with the sample analog of the model-free cointegration concept.

4. Monte Carlo simulation experiment

The data generating process (DGP) follows the experimental design of Arranz & Escribano (2000). This DGP is an extension of the one introduced by Engle & Granger (1987), Banerjee et al. (1986). and Gonzalo (1994) to study the small sample properties of the superconsistent OLS cointegrating estimator.

The DGP is a bivariate cointegrating system of fractionally integrated series of order $I(d)$ and has two equivalent representations. The first one is based on the *cointegrating regression*, with stationary $I(0)$ equilibrium errors (z_t)

$$y_t = \alpha x_t + z_t \tag{4.1a}$$

$$z_t = (a - \alpha)\Delta^d x_t + (1 + b) z_{t-1} + u_{1,t} \tag{4.1b}$$

$$\Delta^d x_t = u_{2,t} \tag{4.1c}$$

where the error terms $u_{1,t}$, $u_{2,t}$ are $iidN(0, \sigma_1^2)$, and $iidN(0, \sigma_2^2)$ with $cov(u_{1,t}, u_{2,t}) = 0$.

The second representation is based on the following *error correction representation* (ECM) of fractionally integrated series

$$\Delta y_t = \alpha \Delta x_t + (a-\alpha)\Delta^d x_t + b(y_{t-1} - \alpha x_{t-1}) + u_{1,t} \tag{4.2a}$$

$$\Delta^d x_t = u_{2,t} \tag{4.2b}$$

Notice that the ECM does not have the usual specification is the sense of having only variables in first differences and levels. However, when the common factor restriction, $(a-\alpha=0)$, is satisfied, the ECM becomes

$$\Delta y_t = \alpha \Delta x_t + b(y_{t-1} - \alpha x_{t-1}) + u_{1,t} \tag{4.3}$$

even when the variables are fractionally integrated, $I(d)$, say $d = 0.7$.

Theorem 2 shows that the values of the parameters a and b are inportant when doing inference on least squares estimators of the cointegrating parameter α.

Theorem 2. *Given the DGP process* (4.1a)–(4.1c) *with* $d = 1$, *under standard regularity conditions, as* $T \to \infty$,

$$T(\hat{\alpha}_{ols} - \alpha) \xrightarrow{d} \left\{ \int_0^1 B_2^2(r)dr \right\}^{-1} \left\{ \frac{\sigma_1}{-b}\left[1 - (a-\alpha)^2 \frac{\sigma_2^2}{\sigma_1^2}\right] \int_0^1 B_2(r)dW_1(r) + \right. \\ \left. + \left(\frac{a-\alpha}{-b}\right) \int_0^1 B_2(r)dB_2(r) + \left(\frac{(a-\alpha)\sigma_2}{-b}\right) \right\} \tag{4.4}$$

where $W_1(r)$ *is a standard Wiener process,* $B_2(r)$ *is a Brownian motion with long-run variance* Σ_{22} *and* $W_1(r)$ *and* $B_2(r)$ *are independent.*

Proof. In Appendix A □

In the Monte Carlo simulations we take the following set of parameter values: $\alpha = 1$, $a = -1, -0.5, 0$, and 1 (common factor restriction), $b = -0.2, -0.5, -1$, $\sigma_1^2 = 1$, $\sigma_2^2 = s^2$, where $s = 1, 6, 16$. The small sample size is $N = 100$, the degree of integration of the variables is $I(d)$ for $d = 1, 0.7, 1.3$ and the number of replications is $T = 20000$.

We analyze the small sample biases of the IV estimator (3.4) for different values of the lag τ of the instrumental variable $(x_{t-\tau})$. Results are displayed in Tables 1 to 5. Notice that when $\tau = 0$, the IV estinator is just the OLS estimator (3.5), with the asymptotic distribution given in equation (4.4).

We consider five different statistics in order to compare the small sample properties of the estimators:

1. Mean bias: $(1/T)\sum_{i=1}^{T}\hat{\alpha}_i - \alpha$
2. Median bias: $med(\hat{\alpha}_1, \ldots, \hat{\alpha}_T) - \alpha$
3. MSE: $(\text{mean bias})^2 + \text{variance of } (\hat{\alpha}_1, \ldots, \hat{\alpha}_T)$
4. Interquartile range (IQR): $Q_3 - Q_1$
5. Concentration probabilities: $\Pr(|\hat{\alpha} - \alpha| \leq 0.05)$

Tables 1 to 3 of Appendix B report the simulation results obtained for $I(1)$ variables, whereas Tables 4 and 5 display the results for fractionally integrated variables, $I(d)$, for $d = 0.7, 1.3$. We show the results obtained for $\tau = 1, 2, \ldots, 5$, and the different values of the parameter a. Although we analyzed different values of s, in particular $s = 1, 6, 16$, we only display results for $s = 6$ since the conclusions hold for all cases. As it was previously stated, the parameter b, corresponding to the error adjustment term in equation (4.2a), is critical in terms of the bias of the estimator. The closer is b to 0, the higher the autocorrelation of the equilibrium errors, z_t. For example, for $b = -0.2$ corresponds to an $AR(1)$ coefficient of 0.8 in equation (4.1b).

When $a = 1$ there is a *common factor restriction*, $(a - \alpha = 0)$, and therefore OLS should perform well in equation (4.1a) (small biases and concentration probability close to 1). The reason is clear since the contemporaneous regressor, Δx_t, drops out of the conditional expectation of equation (4.1b), and Δy_t and Δz_t are uncorrelated in the short-run. On the other hand, the OLS estimator bias should increase with the absolute value of $(a - 1)$. In particular, see Table 1, for $b = -0.2$, and $(a - 1) = -0.5$, the mean bias is -4.7%, the median bias -3.5%, the MSE 0.4%, the IQR 4.5%, and the concentration probability 66%. For $(a - 1)$=-1 the mean bias is -9.5%, the median bias -6.9%, the MSE is 17%, the IQR 8.8% and the concentration probability is 35%. The worst results are for $(a - 1) = -2$: mean bias is -19%, median bias is -13.9%, MSE is 6.6%, IQR is 17.4%, and the concentration probability is 14%.

The improvements of the new IV estimator for $\tau = 3$ are quite impressive. Once again, when the common factor restriction holds $(a-1) = 0$, both the OLS and IV estimator perform well. For other parameter values like $(a - 1) = -0.5$, the IV estimator outperforms by far OLS in terms of biases and dispersion: the mean bias is 0.87%, the median bias is -1.1%, the MSE is 0.1%, the IQR is 3%, and the concentration probability is 90%. For $(a-1) = -1$ the previous difference increases in favor of the IV estimator: the mean bias is -1.7%, the median bias is -2.1%, the MSE is 0.3%, the IQR is 5.4% and the concentration probability is 69%. Finally, for the largest value, $(a - 1) = -2$, the results are better when compared with the corresponding OLS estimator: the mean bias is -3.4%, the median bias is −4%, the MSE 1.3%, the IQR is 10.3% and the concentration probability is 41%.

When $b = -0.5$, the AR(1) coefficient of the equilibrium errors (z_t) of equation (4.1b) is 0.5. therefore. the memory or temporal dependence is reduced generating lower biases and lower required values of τ, see Table 2. Now the best results are obtained for $\tau = 1$ and $\tau = 2$. OLS perform worse than the IV estimator but better than the previous OLS estimator when b was $b = -0.2$.

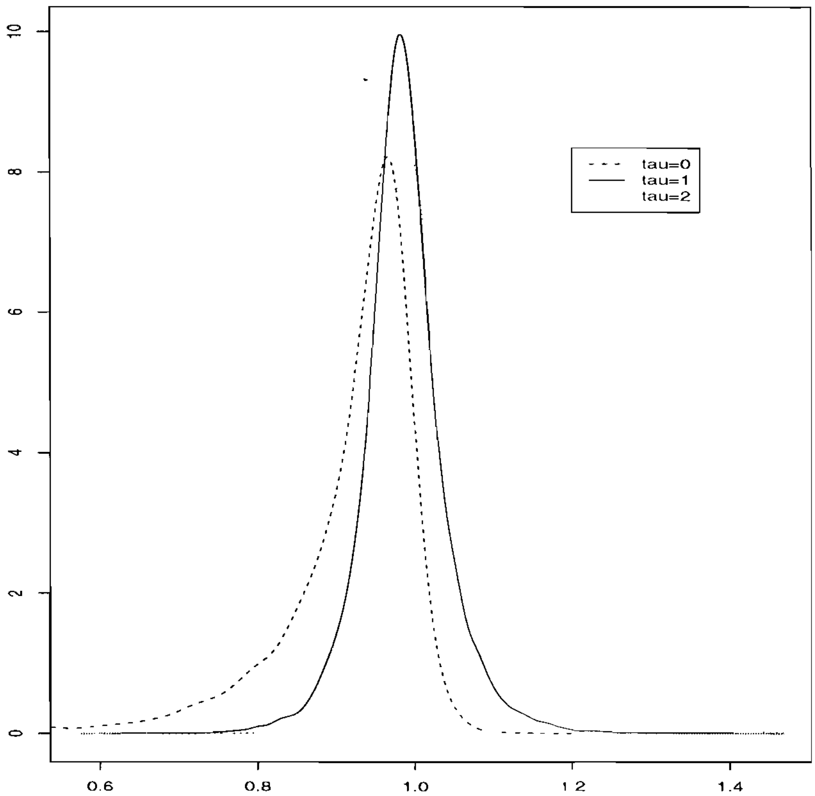


FIGURE 2. Density of IV estimator for $b = -0.5$, $a - \alpha = -2$

Figure 2 reveals about the effects of τ on the distribution of the IV estimator. Notice that when $\tau = 0$ (OLS case) the distribution of the estimator of α is skewed to the left. When we set $\tau = 1$, the mean and the median of $\hat{\alpha}_{iv}$ (the central measures) are closer to 1.0, the true value of α, the dispersion decreases and the distribution is symmetrical. However, when we take $\tau = 2$ the mean bias increases, although the median is the closest one to 1.0, and the dispersion is between the values corresponding to $\tau = 0$ and $\tau = 1$.

As expected with 'weakly exogenous variables' for the cointegrating parameter, the improvement of the IV estimator over OLS is marginal when the equilibrium errors are white noise, $b = -1$, see Table 3. The best IV estimator in terms of mean bias, MSE, and IQR is now $\tau = 1$, although $\tau = 2$ provides almost the same IQR and a slightly smaller median bias.

When the variables are fractionally integrated, $I(d)$, with $d = 0.7$, the improvement of the IV estimator over OLS is again remarkable, see Table 4. OLS only performs well when the common factor restriction, $(a-1) = 0$, is satisfied. In all the other cases the OLS concentration probabilities are never larger than 11%, and in most cases are well below 1%, and the mean biases are very large, reaching 53% when $(a-1) = -2$. On the other hand, by considering the new IV estimator with $\tau = 3$ the results improve dramatically. Even when there is no common factor, in the worst case $(a-1) = -2$, the concentration probability is 28%, the mean bias is 5.6%. the median bias 1.54% (compared to 46.3% obtained with the OLS estimator), and the IQR is 19.2% (compared to 36.5% of the OLS estimator). Therefore, the new IV cointegration estimator produces important reductions in bias relative to OLS.

In the next section we interpret the new concept of cointegration in the frequency domain.

5. Linear cointegration in the frequency domain

Consider again the series $x_t \sim I(d_x)$, $y_t \sim I(d_y)$, and the series z_t formed as $z_t = y_t - \alpha x_t$. To illustrate the meaning of linear cointegration in the frequency domain, we consider the spectrum and the cross-spectrum for the different series [3]. Following Granger (1981), it is easy to see from the definition of z_t that for $0 < \lambda < \pi$,

$$S_z(\lambda) = S_y(\lambda) + \alpha^2 S_x(\lambda) - \alpha\big(S_{y,x}(\lambda) + S^*_{y,x}(\lambda)\big), \tag{5.1}$$

where $S_u(\lambda)$ and $S_{y,x}(\lambda)$ represent the spectrum of u_t $(u = x, y, z)$ and the cross-spectrum of the pair x_t, y_t, respectively, and $S^*_{y,x}(\lambda)$ denotes the complex conjugate of $S_{y,x}(\lambda)$. Since $|S_{y,x}(\lambda)|^2 \leq S_x(\lambda)S_y(\lambda)$, for any λ, and $S_x(\lambda) \sim A_x\lambda^{-2d_x}$, $S_y(\lambda) \sim A_y\lambda^{-2d_y}$ as $\lambda \to 0$, it is clear that the term $\lambda^{-2\max(d_x,d_y)}$ will dominate at low frequencies, and thus $z_t \sim I(\max[d_x, d_y])$ unless a cointegrating restriction applies. Indeed, under (linear) cointegration, the previous algebraic rule breaks down, and there exists a nonzero real number α such that $z_t = y_t - \alpha x_t \sim I(d_z), d_z < d$. This amounts to saying that there exists a positive and finite real number c such that $\lim_{\lambda\to 0}\big(S_y(\lambda)/S_x(\lambda)\big) = c$, and on the other hand, that $\lim_{\lambda\to 0}\big(S_z(\lambda)/S_x(\lambda)\big) = 0$. Therefore, $\lim_{\lambda\to 0}\big[(\alpha^2 + c)S_x(\lambda) - \alpha S_{y,x}(\lambda)\big] = 0$, and since both α and c are nonzero, the limit $\lim_{\lambda\to 0} S_{y,x}(\lambda)/S_x(\lambda)$ must be nonzero and finite.

[3] For nonstationary integrated series we can always consider the *pseudo-spectrum* or *pseudo-cross-spectrum* as in Harvey (1989, p. 64).

Intuitively, this means that x_t, y_t have the long-wave component in common, which amounts to the statement of Theorem 1. Indeed, define

$$\sigma_x^2(t) = var(x_t) \tag{5.2}$$

$$S_x(\lambda, t) = \sigma_x^2(t)\left(1 + 2\sum_{\tau=1}^{t} cov(x_t, x_{t-\tau})exp(-j\lambda\tau)\right) \tag{5.3}$$

$$S_{y,x}(\lambda, t) = \sigma_x^2(t)\left(1 + 2\sum_{\tau=1}^{t} cov(y_t, x_{t-\tau})exp(-j\lambda\tau)\right), \tag{5.4}$$

where $j^2 = -1$. Notice that by using this notation, the (pseudo) spectra $S_x(\lambda)$ and $S_{y,x}(\lambda)$ can be obtained as the limit of t, that is,

$$S_x(\lambda) = \lim_{t\to\infty} S_x(\lambda, t) \tag{5.5}$$

$$S_{y,x}(\lambda) = \lim_{t\to\infty} S_{y,x}(\lambda, t). \tag{5.6}$$

The advantage of introducing this notation is that both $S_x(\lambda, t)$ and $S_{y,x}(\lambda, t)$ exist for finite t. Therefore for finite t we obtain

$$\lim_{\lambda\to 0}\frac{S_{y,x}(\lambda, t)}{S_x(\lambda, t)} = \frac{1}{1+2s_t^{(x,x)}} + \frac{2s_t^{(y,x)}}{1+2s_t^{(x,x)}}. \tag{5.7}$$

Now, since $s_t^{(x,x)}$ diverges for $t \to \infty$, then for all T satisfying $\tau < T < t$ we have that

$$\lim_{t\to\infty,\lambda\to 0}\frac{S_{y,x}(\lambda, t)}{S_x(\lambda, t)} = \lim_{t\to\infty}\frac{\sum_{\tau=1}^{T} cov(y_t, x_{t-\tau}) + \sum_{\tau>T}^{t} cov(y_t, x_{t-\tau})}{\sum_{\tau=1}^{T} cov(x_t, x_{t-\tau}) + \sum_{\tau>T}^{t} cov(x_t, x_{t-\tau})}, \tag{5.8}$$

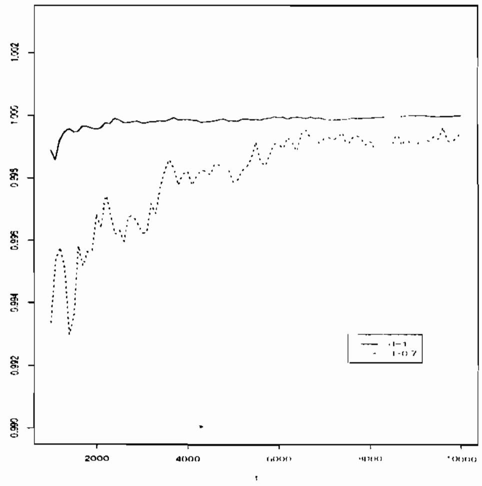
and since the first term in both the numerator and the denominator of the right-hand side of the previous formula is bounded for finite T, they can be neglected for large t, and thereby

$$\lim_{t\to\infty,\lambda\to 0}\frac{S_{y,x}(\lambda, t)}{S_x(\lambda, t)} = \lim_{t\to\infty}\frac{\sum_{\tau>T}^{t} cov(y_t, x_{t-\tau})}{\sum_{\tau>T}^{t} cov(x_t, x_{t-\tau})}. \tag{5.9}$$

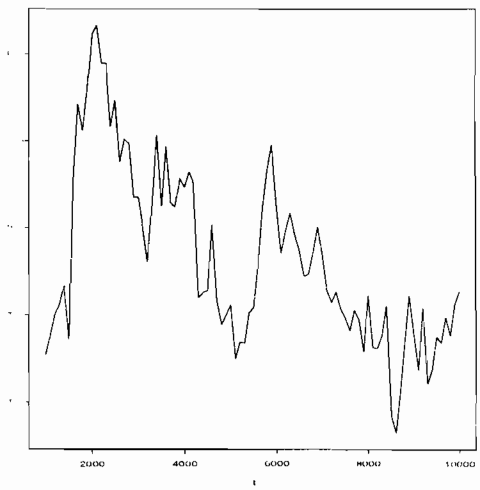
The previous expression is valid for all T finite. Letting T increase without bound as t grows to infinity, we obtain that for $\tau = o(t)$

$$\lim_{t\to\infty,\lambda\to 0}\frac{S_{y,x}(\lambda, t)}{S_x(\lambda, t)} = \lim_{t\to\infty}\frac{cov(y_t, x_{t-\tau})}{cov(x_t, x_{t-\tau})} = \alpha. \tag{5.10}$$

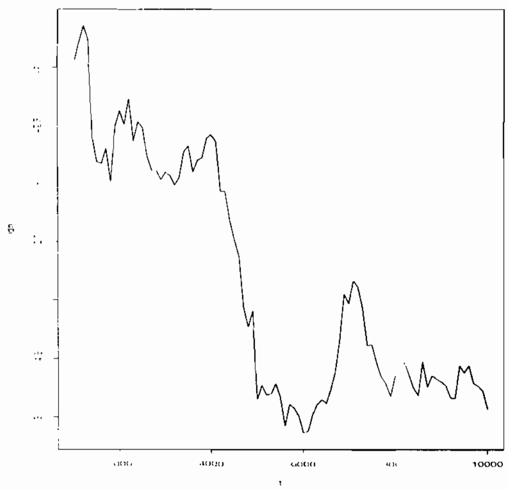
In order to check the condition (5.9) we decided to simulate 500 pairs of series x_t, y_t of sample size 10000 following the DGP of equations (4.2a)–(4.2b) for $d = 1, 0.7$, $a = 0.5$, with $\alpha = 1$. As we can see in Figure 3(a), the mean of the ratios $\frac{\sum_{\tau>T}^{t} cov(y_t, x_{t-\tau})}{\sum_{\tau>T}^{t} cov(x_t, x_{t-\tau})}$ is around the true value of α if



(a) $b = -0.5$.



(b) $b = 0.0, I(1)$



(c) $b = 0.0, I(0.7)$

FIGURE 3. $\left[\sum_{\tau>T}^{t} \widehat{cov}(y_t, x_{t-\tau})\right] / \left[\sum_{\tau>T}^{t} \widehat{cov}(x_t, x_{t-\tau})\right]$

the variables are cointegrated, while it is not convergent when the series are not cointegrated, see Figures 3(b) and 3(c).

6. Relationship between cointegration in the time domain and in the frequency domain.

Notice that both Theorem 1 and equation (5.10) point to the same spectral result, namely that under linear cointegration $\lim_{\lambda\to 0}(S_{y,x}(\lambda)/S_x(\lambda)) = \alpha$, where α is nonzero and finite constant value. If we assume that the series x_t, y_t have spectra bounded away from zero at $\lambda = 0$, the latter result suggests the possibility of using the time domain estimator of $\frac{cov(y_t,x_{t-\tau})}{cov(x_t,x_{t-\tau})}$ (IV estimator), as t approaches infinity, or alternatively, as λ approaches 0, the spectral estimator of $\frac{S_{y,x}(\lambda)}{S_x(\lambda)}$ for the linear cointegration parameter.

Lemma 1. *Let $\sum p_n$ and $\sum q_n$ be two divergent series of positive terms. If $q_n/p_n \to 0(\infty)$ then $\sum q_n$ diverges less (more) rapidly than $\sum p_n$.*

Proof. See Knopp (1990, p. 280) □

Theorem 3. *Let $s_n^{(y,x)} = \sum_{\tau=1}^{n} cov(y_t, x_{t-\tau})/var(x_{t-\tau})$, where $n = o(t)$. The series y_t, x_t are* **long-memory** *and* **linearly cointegrated** *if as $n \to \infty$, we have that:*

1. *the sequence of partial sums $s_n^{(y,x)}$ diverges and,*
2. *the ratio of the sequences $s_n^{(y,x)}$ and $s_n^{(x,x)}$ converges to a nonzero and finite real number, α (the cointegration parameter).*

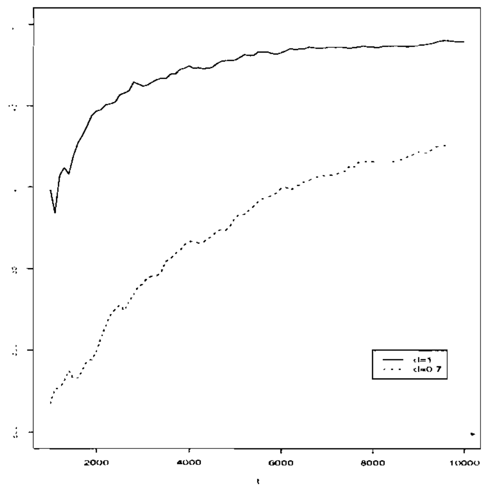
Proof. In Appendix A. □

Corollary 2. *If y_t, x_t are $I(d_y), I(d_x)$, respectively, with d_x, d_y restricted to the interval $[0,1]$, then a necessary and sufficient condition for the series to be* **fractionally cointegrated**, *or that they are linearly cointegrated with $0 < d_x = d_y = d < 1$ is that, as $n \to \infty$, one has:*

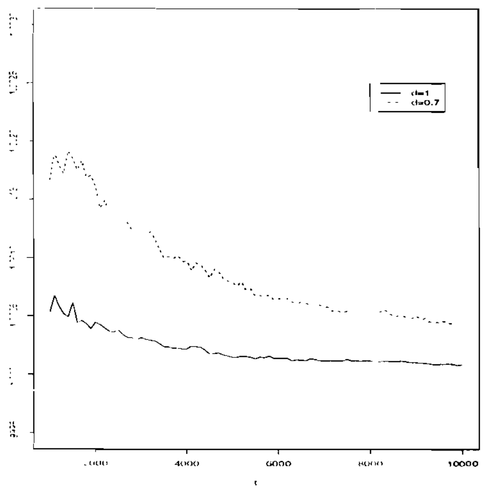
1. *The sequence $s_n^{(y,x)}$ diverges.*
2. *The ratio of sequences $s_n^{(y,x)}$ and $s_n^{(x,x)}$ converges to the nonzero and finite real number, α.*
3. *For some fixed $\tau < t$ or for $\tau = o(t)$, $\lim_{t\to\infty} cov(y_t, x_{t-\tau})/var(x_{t-\tau}) = 0$.*

Proof. In Appendix A. □

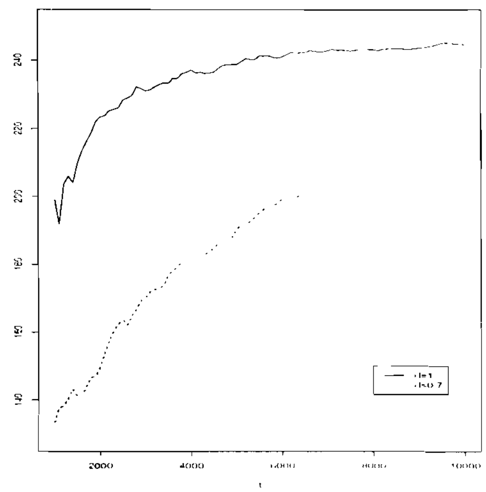
Again, Figure 4 shows the intuition behind Theorem 3 and Corollary 2. We have simulated 500 pairs of x_t and y_t series of sample size t following the DGP of equations (4.2a) (4.2b) for $d = 1$, and 0.7, $a = 0.5$, with $\alpha = 1$, and take the average of the sample analog of the desired statistics. The results when the series are cointegrated are as expected, the partial sums diverge and the ratio converges even for small values of τ. In the case that the series are not cointegrated, the partial sums diverge and the ratio is exactly the value of the parameter a.



(a) $s_n^{(y,x)}$, $b = -0.5$



(b) $s_n^{(y,x)}/s_n^{(x,x)}$, $b = -0.5$



(c) $s_n^{(y,x)}$, $b = 0.0$

FIGURE 4. Simulation experiment of Theorem 3 and Corollary 2, when $d = 1$, and $d = 0.7$, $\alpha = 1$, $a = 0.5$, $n = 500$

Remark:

- Notice that there is no loss of generality by restricting d_x, d_y to lie within the unit interval, since by proper differencing of the series we can determine the integers closest to d_x and d_y.

- Once we know that $0 < d < 1$, we can inquire as to whether $d < 1/2$ (in which case the series are stationary) or not. To answer the question we can check whether the variance of x_t diverges or not.

7. Conclusions and extensions

In this paper we propose a model-free (non parametric) characterization of long-memory and linear cointegration, for integer and fractional $I(d)$, based on simple statistics constructed as the ratio of the cross-covariance and the covariance functions of the series. We show the relationship of this definition with the usual concept of cointegration.

Our nonparametric concept of cointegration has a natural interpretation as an instrumental variable estimator where the instrument is a lagged value of the explanatory variable $(x_{t-\tau})$. The analysis is performed in the time and the frequency domains. From the comparison of the results we suggest an alternative estimator of the cointegration parameter, α, based on the ratio of the partial sums of the cross–covariances and the partial sums of the autocovariances.

Finally, we show by Monte Carlo simulation experiments the good small sample behavior of our cointegration estimators relative to the usual estimators based on least squares, The formal derivation of the asymptotic distribution properties of the instrumental variable estimator is beyond the scope of this paper, and we are currently working in that direction.

Appendix A. Proofs

Proof of Theorem 1

For simplicity, we will assume that both x_t and y_t have zero mean. Also let us refer to the covariance function of a pair of processes y_t, x_t, conditional on some initial conditions, say $y_0 = x_0 = 0$, as $cov(x_t, y_{t-\tau}) = \gamma_{x,y}(\tau, t)$. Notice that if one of the processes is nonstationary, the covariance will be a function of t. Similarly, we will refer to the autocovariance function of a process, say x_t, as $cov(x_t, x_{t-\tau}) = \gamma_x(\tau, t)$, conditional on $x_0 = 0$.

Using Definition 5 and letting $\tau = o(t)$, under linear cointegration, say $y_t = \alpha x_t + z_t$ there exists a nonzero finite real number α such that

$$\lim_{\tau\to\infty} cov(y_t, x_{t-\tau}) = \alpha \lim_{\tau\to\infty} cov(x_t, x_{t-\tau}) + \lim_{\tau\to\infty} cov(z_t, x_{t-\tau}) \tag{A.1}$$

Thus

$$\lim_{\tau\to\infty} \frac{cov(y_t, x_{t-\tau})}{cov(x_t, x_{t-\tau})} = \alpha + \lim_{\tau\to\infty} \frac{\gamma_{z,x}(\tau, t)}{\gamma_x(\tau, t)}. \tag{A.2}$$

Now under cointegration z_t must be an $I(d_z)$ series, with $d_z < d$, while x_t is $I(d)$ that is,

$$\Delta^d x_t = u_t \tag{A.3}$$

$$\Delta^{d_z} z_t = v_t, \tag{A.4}$$

with u_t, v_t representing zero-mean $I(0)$ series. Inversion of the differencing operator leads to an infinitely long moving-average expansion. By truncating this expansion to take into account that the series is available only from $t = 0$ given the initial conditions $x_0 = y_0 = z_0 = 0$, we obtain:

$$z_t = \sum_{k=0}^{t} \theta_k v_{t-k} \tag{A.5}$$

$$x_{t-\tau} = \sum_{l=0}^{t} \phi_l u_{t-\tau-l}. \tag{A.6}$$

Thus:

$$\gamma_{z,x}(\tau, t) = \sum_{k=0}^{t} \sum_{l=0}^{t} \phi_l \theta_k \gamma_{u,v}(\tau + l - k). \tag{A.7}$$

Now recalling that the covariance function of two stationary series u_t and v_t, $\gamma_{u,v}(\tau)$, is linked to their *cross-spectrum*, $S_{u,v}(\lambda)$, through

$$\gamma_{u,v}(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{u,v}(\lambda) \exp(j\lambda\tau)\, d\lambda. \tag{A.8}$$

with $j^2 = -1$. (see Granger & Hatanaka (1964) and Priestley (1981)), we obtain after a little algebra :

$$\gamma_{z,x}(\tau, t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_t^*(\lambda)\Theta_t(\lambda) S_{u,v}(\lambda) \exp(j\lambda\tau)\, d\lambda, \tag{A.9}$$

where $\Phi_t(\lambda)$ and $\Theta_t(\lambda)$ are defined as

$$\Phi_t(\lambda) = \sum_{k=0}^{t} \phi_k \exp(-j\lambda k) \tag{A.10}$$

$$\Theta_t(\lambda) = \sum_{l=0}^{t} \theta_k \exp(-j\lambda l), \tag{A.11}$$

and $\Phi_t^*(\lambda)$ denotes the complex conjugate of $\Phi_t(\lambda)$. Notice that as t grows to infinity these operators become

$$\Phi(\lambda) = \lim_{t\to\infty} \Phi_t(\lambda) = [1 - \exp(-j\lambda)]^{-d} \quad (A.12)$$

$$\Theta(\lambda) = \lim_{t\to\infty} \Theta_t(\lambda) = [1 - \exp(-j\lambda)]^{-d_z}. \quad (A.13)$$

Moreover, it is easy to see that as λ approaches zero they become :

$$\Phi(\lambda) \sim \lambda^{-d} \quad (A.14)$$

$$\Theta(\lambda) \sim \lambda^{-d_z}. \quad (A.15)$$

As a consequence, when multiplying $S_{u,v}(\lambda)$ in equation (A.9), both operators $\Phi(\lambda)$ and $\Theta(\lambda)$ act as very sharp *low-pass filters*, that is, they emphasize the values of $S_{u,v}(\lambda)$ for λ very close to zero, while damping those for the higher frequencies. Since u_t, v_t are $I(0)$ processes, their cross-spectrum $S_{u,v}(\lambda)$ will be bounded away from zero and infinity at $\lambda = 0$, we will have for very large t that $\gamma_{z,x}(\tau, t)$ will be dominated by the term $\int_{-\pi}^{\pi} \lambda^{-d-d_z} \exp(j\lambda\tau)\, d\lambda$. Finally, noting that the inverse Fourier transform of the function $f(\lambda) = \lambda^{-d-d_z}$ is τ^{d+d_z-1}, we obtain :

$$\frac{\gamma_{z,x}(\tau, t)}{\gamma_x(\tau, t)} \sim \tau^{d_z - d}. \quad (A.16)$$

The statement of the Theorem follows immediately since $d_z < d$.

Proof of Corollary 1

Since the series y_t and x_t are cointegrated, we know from Theorem 1 that

$$\lim_{\tau\to\infty} \frac{cov(y_t, x_{t-\tau})}{cov(x_t, x_{t-\tau})} = \alpha \quad \text{for } \tau < t \text{ or } \tau = o(t)$$

The same is true if we multiply and divide by $var(x_{t-\tau})$

$$\lim_{\tau\to\infty} \frac{\dfrac{cov(y_t, x_{t-\tau})}{var(x_{t-\tau})}}{\dfrac{cov(x_t, x_{t-\tau})}{var(x_{t-\tau})}} = \alpha$$

If $\lim_{\tau\to\infty} \dfrac{cov(y_t, x_{t-\tau})}{var(x_{t-\tau})} = \alpha$ it must occur that $\lim_{\tau\to\infty} \dfrac{cov(x_t, x_{t-\tau})}{var(x_{t-\tau})} = 1$. Given that $d \in (0, 1]$, this is only true for $d = 1$, since for any other value of d, x_t is mean reverting.

Proof of Theorem 2

Gonzalo (1994) proved that when the DGP is $y_t = \alpha x_t + z_t$, where $z_t = \varphi z_{t-1} + \eta_{1,t}$, and $\Delta x_t = u_{2,t}$, with

$$\begin{pmatrix} \eta_{1,t} \\ u_{2,t} \end{pmatrix} \sim NID\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \theta\sigma_1\sigma_2 \\ \theta\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right]$$

then

$$T(\hat{\alpha}_{ols} - \alpha) \xrightarrow{d} \left\{\int_0^1 B_2^2(r)dr\right\}^{-1} \left\{\left(\frac{\sigma_1}{1-\varphi}\right)\left[1-\theta^2\right]^{1/2} \int_0^1 B_2(r)dW_1(r) + \right.$$
$$\left. + \left(\frac{1}{1-\varphi}\right)\int_0^1 B_2(r)dB_2(r) + \left(\frac{1}{1-\varphi}\right)\theta\sigma_1\sigma_2\right\}$$

The DGP used by Gonzalo (1994) can be written as (4.1a)–(4.1c) with $d = 1$.

$$y_t = \alpha x_t + z_t$$
$$z_t = (1+b)z_{t-1} + (a-\alpha)\Delta x_t + u_{1,t}$$
$$\Delta x_t = u_{2,t}$$

where

$$\begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} \sim NID\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right]$$

The result follows by noticing that $\varphi = (1+b)$ and that $\theta = (a-\alpha)\sigma_2/\sigma_1$.

Proof of Theorem 3

First, let us prove that our conditions are necessary if the series are long-memory and linearly cointegrated. Under the latter assumption, from Theorem 1, there exists a nonzero and finite real number b such that

$$cov(y_t, x_{t-\tau}) = \alpha\, cov(x_t, x_{t-\tau}) + o\left(cov(x_t, x_{t-\tau})\right), \tag{A.17}$$

On the other hand, since x_t is long-memory, $s_n^{(x,x)}$ diverges. Now, taking partial sums in the previous equation, it is immediate that $s_n^{(y,x)}$ must also diverge, whereas the ratio $(s_n^{(y,x)}/s_n^{(x,x)})$ will converge to α as n grows to infinity.

Proof of Corollary 2

Fractionally integrated series with long-memory parameter $d < 1$ series satisfy condition 1, while conditions 2 and 3 are implied by Theorems 1 and 2 when the series are linearly cointegrated.

APPENDIX B. TABLES

TABLE 1. OLS and IV estimator of α in equation $y_t = \alpha x_t + z_t$, where the DGP is $\Delta y_t = a\Delta x_t - 0.2(y_{t-1} - \alpha x_{t-1}) + u_{1t}$, $\Delta x_t = u_{2t}$, $\alpha = 1$, $u_{1t} \sim N(0,1)$, $u_{2t} \sim N(0, s^2)$. $N = 100$, $T = 20000$ replications.

$a - \alpha$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\alpha} - \alpha\rvert \leq 0.05)$
			$\tau = 0$ (OLS)		
-2.0	-0.190	-0.139	0.664E-01	0.174	0.139
-1.0	-0.948E-01	-0.691E-01	0.168E-01	0.882E-01	0.351
-0.5	-0.474E-01	-0.346E-01	0.433E-02	0.455E-01	0.659
0.0	-0.920E-04	-0.184E-03	0.190E-03	0.134E-01	0.994
			$\tau = 1$		
-2.0	-0.129	-0.102	0.322E-01	0.128	0.216
-1.0	-0.647E-01	-0.512E-01	0.820E-02	0.656E-01	0.479
-0.5	-0.324E-01	-0.257E-01	0.221E-02	0.350E-01	0.774
0.0	-0.106E-03	-0.185E-03	0.209E-03	0.136E-01	0.991
			$\tau = 2$		
-2.0	-0.782E-01	-0.706E-01	0.158E-01	0.102	0.320
-1.0	-0.392E-01	-0.353E-01	0.413E-02	0.528E-01	0.615
-0.5	-0.196E-01	-0.177E-01	0.120E-02	0.298E-01	0.872
0.0	-0.109E-03	-0.159E-03	0.232E-03	0.138E-01	0.988
			$\tau = 3$		
-2.0	-0.346E-01	-0.398E-01	0.129E-01	0.103	0.407
-1.0	-0.173E-01	-0.208E-01	0.340E-02	0.538E-01	0.692
-0.5	-0.872E-02	-0.109E-01	0.104E-02	0.301E-01	0.904
0.0	-0.107E-03	-0.198E-03	0.261E-03	0.140E-01	0.984
			$\tau = 4$		
-2.0	0.408E-02	-0.184E-01	0.243E-01	0.124	0.393
-1.0	0.200E-02	-0.993E-02	0.632E-02	0.630E-01	0.675
-0.5	0.963E-03	-0.581E-02	0.182E-02	0.342E-01	0.885
0.0	-0.777E-04	-0.185E-03	0.308E-03	0.143E-01	0.981
			$\tau = 5$		
-2.0	0.436E-01	-0.437E-02	0.374	0.147	0.369
-1.0	0.218E-01	-0.228E-02	0.992E-01	0.745E-01	0.643
-0.5	0.110E-01	-0.163E-02	0.279E-01	0.389E-01	0.854
0.0	0.110E-03	-0.195E-03	0.746E-03	0.146E-01	0.976

TABLE 2. OLS and IV estimator of α in equation $y_t = \alpha x_t + z_t$, where the DGP is $\Delta y_t = a\Delta x_t - 0.5(y_{t-1} - \alpha x_{t-1}) + u_{1t}$, $\Delta x_t = u_{2t}$, $\alpha = 1$, $u_{1t} \sim N(0,1)$, $u_{2t} \sim N(0, s^2)$. $N = 100$, $T = 20000$ replications.

$a - \alpha$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\alpha} - \alpha\rvert \leq 0.05)$
			$\tau = 0$ (OLS)		
-2.0	-0.819E-01	-0.561E-01	0.139E-01	0.754E-01	0.449
-1.0	-0.410E-01	-0.279E-01	0.350E-02	0.381E-01	0.729
-0.5	-0.205E-01	-0.141E-01	0.902E-03	0.197E-01	0.911
0.0	-0.397E-04	-0.679E-04	0.352E-04	0.559E-02	1.00
			$\tau = 1$		
-2.0	-0.149E-01	-0.149E-01	0.193E-02	0.418E-01	0.799
-1.0	-0.746E-02	-0.767E-02	0.510E-03	0.217E-01	0.961
-0.5	-0.376E-02	-0.408E-02	0.156E-03	0.122E-01	0.997
0.0	-0.521E-04	-0.566E-04	0.390E-04	0.569E-02	1.000
			$\tau = 2$		
-2.0	0.211E-01	0.461E-02	0.575E-02	0.645E-01	0.711
-1.0	0.105E-01	0.214E-02	0.147E-02	0.324E-01	0.889
-0.5	0.523E-02	0.900E-03	0.398E-03	0.169E-01	0.969
0.0	-0.504E-04	-0.681E-04	0.439E-04	0.575E-02	1.000
			$\tau = 3$		
-2.0	0.408E-01	0.130E-01	0.122E-01	0.794E-01	0.654
-1.0	0.204E-01	0.638E-02	0.309E-02	0.398E-01	0.830
-0.5	0.102E-01	0.311E-02	0.809E-03	0.205E-01	0.936
0.0	-0.418E-04	-0.695E-04	0.507E-04	0.585E-02	0.999
			$\tau = 4$		
-2.0	0.530E-01	0.172E-01	0.220E-01	0.882E-01	0.628
-1.0	0.265E-01	0.853E-02	0.556E-02	0.442E-01	0.803
-0.5	0.132E-01	0.418E-02	0.144E-02	0.228E-01	0.915
0.0	-0.140E-04	-0.973E-04	0.615E-04	0.596E-02	0.998
			$\tau = 5$		
-2.0	0.647E-01	0.193E-01	0.234	0.932E-01	0.616
-1.0	0.324E-01	0.957E-02	0.609E-01	0.468E-01	0.788
-0.5	0.163E-01	0.464E-02	0.165E-01	0.240E-01	0.904
0.0	0.105E-03	-0.789E-04	0.183E-03	0.603E-02	0.997

TABLE 3. OLS and IV estimator of α in equation $y_t = \alpha x_t + z_t$, where the DGP is $\Delta y_t = a\Delta x_t - (y_{t-1} - \alpha x_{t-1}) + u_{1t}$, $\Delta x_t = u_{2t}$, $\alpha = 1$. $u_{1t} \sim N(0,1)$, $u_{2t} \sim N(0, s^2)$. $N = 100$, $T = 20000$ replications.

$a - \alpha$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\alpha} - \alpha\rvert \leq 0.05)$
			$\tau = 0$ (OLS)		
-2.0	-0.422E-01	-0.282E-01	0.391E-02	0.390E-01	0.727
-1.0	-0.211E-01	-0.141E-01	0.984E-03	0.196E-01	0.906
-0.5	-0.106E-01	-0.705E-02	0.253E-03	0.100E-01	0.984
0.0	-0.195E-04	-0.294E-04	0.941E-05	0.284E-02	1.00
			$\tau = 1$		
-2.0	0.277E-01	0.112E-01	0.387E-02	0.461E-01	0.789
-1.0	0.138E-01	0.551E-02	0.973E-03	0.230E-01	0.914
-0.5	0.691E-02	0.270E-02	0.250E-03	0.117E-01	0.981
0.0	-0.322E-04	-0.394E-04	0.106E-04	0.289E-02	1.00
			$\tau = 2$		
-2.0	0.283E-01	0.109E-01	0.430E-02	0.463E-01	0.786
-1.0	0.142E-01	0.540E-02	0.108E-02	0.233E-01	0.911
-0.5	0.707E-02	0.264E-02	0.279E-03	0.119E-01	0.979
0.0	-0.226E-04	-0.347E-04	0.120E-04	0.292E-02	1.00
			$\tau = 3$		
-2.0	0.290E-01	0.107E-01	0.483E-02	0.470E-01	0.785
-1.0	0.145E-01	0.535E-02	0.122E-02	0.234E-01	0.908
-0.5	0.723E-02	0.261E-02	0.315E-03	0.120E-01	0.976
0.0	-0.146E-04	-0.431E-04	0.141E-04	0.300E-02	1.000
			$\tau = 4$		
-2.0	0.303E-01	0.105E-01	0.656E-02	0.473E-01	0.784
-1.0	0.151E-01	0.524E-02	0.166E-02	0.238E-01	0.905
-0.5	0.757E-02	0.255E-02	0.434E-03	0.122E-01	0.973
0.0	0.649E-05	-0.482E-04	0.171E-04	0.302E-02	1.000
			$\tau = 5$		
-2.0	0.330E-01	0.103E-01	0.513E-01	0.480E-01	0.781
-1.0	0.166E-01	0.504E-02	0.133E-01	0.241E-01	0.902
-0.5	0.831E-02	0.248E-02	0.362E-02	0.124E-01	0.969
0.0	0.726E-04	-0.323E-04	0.473E-04	0.306E-02	1.000

TABLE 4. OLS and IV estimator of α in equation $y_t = \alpha x_t + z_t$, where the DGP is $\Delta y_t = \alpha \Delta x_t + (a-1)\Delta^d x_t - 0.5(y_{t-1} - \alpha x_{t-1}) + u_{1t}$, $\Delta x_t = u_{2t}$, $\alpha = 1$,$d = 0.7$, $u_{1t} \sim N(0,1)$, $u_{2t} \sim N(0,s^2)$. $N = 100$, $T = 20000$ replications.

$a-\alpha$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\alpha}-\alpha\rvert \leq 0.05)$
			$\tau = 0$ (OLS)		
-2.0	-0.526	-0.463	0.365	0.430	0.001
-1.0	-0.263	-0.232	0.914E-01	0.215	0.009
-0.5	-0.131	-0.116	0.230E-01	0.108	0.116
0.0	0.480E-05	-0.134E-04	0.182E-03	0.157E-01	0.998
			$\tau = 1$		
-2.0	-0.250	-0.227	0.829E-01	0.195	0.038
-1.0	-0.125	-0.113	0.209E-01	0.980E-01	0.128
-0.5	-0.625E-01	-0.560E-01	0.544E-02	0.504E-01	0.434
0.0	-0.609E-04	-0.907E-04	0.272E-03	0.174E-01	0.989
			$\tau = 2$		
-2.0	-0.648E-01	-0.852E-01	0.263E-01	0.132	0.220
-1.0	-0.324E-01	-0.427E-01	0.678E-02	0.678E-01	0.468
-0.5	-0.163E-01	-0.214E-01	0.194E-02	0.375E-01	0.809
0.0	-0.118E-03	0.306E-04	0.373E-03	0.186E-01	0.975
			$\tau = 3$		
-2.0	0.558E-01	-0.154E-01	0.103	0.192	0.282
-1.0	0.278E-01	-0.764E-02	0.259E-01	0.971E-01	0.528
-0.5	0.138E-01	-0.413E-02	0.669E-02	0.518E-01	0.791
0.0	-0.177E-03	-0.406E-05	0.541E-03	0.195E-01	0.962
			$\tau = 4$		
-2.0	0.184	0.202E-01	45.2	0.247	0.268
-1.0	0.921E-01	0.947E-02	11.7	0.125	0.493
-0.5	0.464E-01	0.475E-02	3.12	0.650E-01	0.728
0.0	0.726E-03	0.880E-04	0.141E-01	0.206E-01	0.948
			$\tau = 5$		
-2.0	0.253	0.387E-01	38.2	0.285	0.253
-1.0	0.125	0.192E-01	8.48	0.144	0.466
-0.5	0.613E-01	0.962E-02	1.64	0.744E-01	0.691
0.0	-0.255E-02	0.107E-04	0.141	0.216E-01	0.931

TABLE 5. OLS and IV estimator of α in equation $y_t = \alpha x_t + z_t$, where the DGP is $\Delta y_t = \alpha \Delta x_t + (a-1)\Delta^d x_t - 0.5(y_{t-1} - \alpha x_{t-1}) + u_{1t}$, $\Delta x_t = u_{2t}$, $\alpha = 1, d = 1.3$, $u_{1t} \sim N(0,1)$, $u_{2t} \sim N(0, s^2)$. $N = 100$, $T = 20000$ replications.

$a - \alpha$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\alpha} - \alpha\rvert \leq 0.05)$
		$\tau = 0$ (OLS)			
-2.0	-0.120E-02	-0.146E-02	0.283E-03	0.156E-01	0.983
-1.0	-0.605E-03	-0.797E-03	0.753E-04	0.788E-02	0.997
-0.5	-0.308E-03	-0.421E-03	0.227E-04	0.414E-02	1.000
0.0	-0.119E-04	0.394E-05	0.434E-05	0.151E-02	1.000
		$\tau = 1$			
-2.0	0.791E-02	0.216E-02	0.620E-03	0.207E-01	0.947
-1.0	0.395E-02	0.111E-02	0.157E-03	0.104E-01	0.991
-0.5	0.197E-02	0.532E-03	0.421E-04	0.525E-02	0.999
0.0	-0.158E-04	-0.268E-05	0.446E-05	0.151E-02	1.000
		$\tau = 2$			
-2.0	0.124E-01	0.365E-02	0.111E-02	0.237E-01	0.916
-1.0	0.619E-02	0.179E-02	0.278E-03	0.118E-01	0.977
-0.5	0.309E-02	0.887E-03	0.719E-04	0.604E-02	0.997
0.0	-0.147E-04	-0.459E-05	0.461E-05	0.152E-02	1.000
		$\tau = 3$			
-2.0	0.146E-01	0.422E-02	0.144E-02	0.250E-01	0.904
-1.0	0.727E-02	0.208E-02	0.361E-03	0.124E-01	0.970
-0.5	0.363E-02	0.103E-02	0.925E-04	0.642E-02	0.994
0.0	-0.170E-04	-0.492E-05	0.480E-05	0.150E-02	1.000
		$\tau = 4$			
-2.0	0.157E-01	0.434E-02	0.166E-02	0.258E-01	0.896
-1.0	0.782E-02	0.217E-02	0.415E-03	0.127E-01	0.968
-0.5	0.390E-02	0.108E-02	0.106E-03	0.656E-02	0.993
0.0	-0.180E-04	0.381E-05	0.521E-05	0.151E-02	1.000
		$\tau = 5$			
-2.0	0.161E-01	0.437E-02	0.177E-02	0.262E-01	0.893
-1.0	0.805E-02	0.219E-02	0.445E-03	0.130E-01	0.965
-0.5	0.402E-02	0.107E-02	0.115E-03	0.662E-02	0.991
0.0	-0.118E-04	0.745E-05	0.559E-05	0.153E-02	1.000

References

Aparicio, F. (1993), Performances and limitations of wind speed sensor equalization, *in* 'Proceedings of the Colloque GRETSI', Vol. 2, Juan-Les-Pins, France, pp. 1407–1410.

Aparicio, F. (1995*a*), A linearity test for a class of processes with long-range dependence, *in* 'Proceedings of the Colloque GRETSI', Juan-Les-Pins, France.

Aparicio, F. (1995*b*), Nonlinear modelling and analysis under long-range dependence with an application to positive time series, PhD thesis, Swiss Federal Institute of Technology (EPFL) -Signal Processing Laboratory-, Lausanne, Switzerland.

Aparicio, F. & Escribano, A. (1997), Searching for linear and nonlinear cointegration: A new approach, Working Paper 97–65, Dpt. of Statistics and Econometrics, Universidad Carlos III de Madrid.

Aparicio, F. & Escribano, A. (1999), 'Information-theoretic analysis of serial dependence and cointegration', *Studies in Nonlinear Dynamics and Econometrics*.

Arranz, M. & Escribano, A. (2000), 'Cointegration testing under structural breaks: A robust extended error correction model', *Oxford Bulletin of Economics and Statistics* **62**, 23–52.

Balke, N. S. & Fomby, T. B. (1997), 'Threshold cointegration', *International economic review* **38**(3), 627–645.

Banerjee, A., Dolado, J., Hendry, D. & Smith, G. (1986), 'Exploring equilibrium relationships in econometrics through static models: Some Monte Carlo evidence', *Oxford Bulletin of Economics and Statistics* **48**, 853–877.

Beran, J., Sherman, R., Taqqu, M. & Willinger, W. (1991), Variable-bit-rate video traffic and long range dependence, Working paper (Dept. of Mathematics of Boston University).

Box, G. & Jenkins, G. (1970), *Time series analysis: forecasting and control*, Holden-Day, San Francisco.

Dolado, J. & Marmol, F. (1998), FM-OLS estimation of cointegrating relationships among nonstationary fractionally integrated processes, Working Paper 98–41, Dpt. of Statistics and Econometrics, Universidad Carlos III de Madrid.

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

Escribano, A. (1986), Nonlinear error-correction. The case of money demand in the U.K. (1878-1970). Chapter IV, PhD thesis, Dept. of Economics of the University of California at San Diego, La Jolla, USA.

Escribano, A. (1987), Error-correcting systems: nonlinear adjustment to linear long-run relationships, CORE Discussion paper 8730, C.O.R.E.

Escribano, A. & Aparicio, F. M. (1999), Cointegration: linearity, nonlinearity, outliers, and structural breaks, *in* 'The Current State of Economic Science', Spellbound Publications, pp. 383–407. Forthcoming.

Escribano, A. & Granger, C. W. J. (1998), 'Investigating the relationships between gold and silver prices', *Journal of forecasting* **17**, 81–107.

Escribano, A. & Pfann, G. A. (1998), 'Nonlinear error correction, asymmetric adjustment and cointegration', *Economic modelling* **15**, 197–216.

Gonzalo, J. (1994), 'Five alternative methods of estimating long-run relationships', *Journal of Econometrics* **60**, 233–233.

Granger, C. W. J. (1981), 'Some properties of time series data and their use in econometric model specification', *Journal of econometrics* **16**, 121–130.

Granger, C. W. J. & Hatanaka, M. (1964), *Spectral analysis of economic time series*, Princeton University Press, New Jersey.

Granger, C. W. J. & Joyeux, R. (1980), 'An introduction to long-memory time series models and fractional differencing', *Journal of Time Series Analysis* 1(1), 15–29.

Granger, C. W. J. & Newbold, P. (1974), 'Spurious regressions in econometrics', *Journal of Econometrics* **2**, 111–120.

Harvey, A. (1989), *Forecasting structural time series models and the Kalman filter*, Cambridge University Press, Cambridge.

Hosking, J. (1981), 'Fractional differencing', *Biometrika* **68**, 165–176.

Johansen, S. (1988), 'Statistical analysis of cointegration vectors', *Journal of economic dynamics and control* **12**, 231–254.

Knopp, K. (1990), *Theory and application of infinite series*, Dover, N.Y.

Leland, W., Taqqu, M., Willinger, W. & Wilson, D. (1994), 'On the self-similar nature of ethernet traffic', *IEEE/ACM Transactions on Networking* **2**, 1–15.

Mandelbrot, B. (1965), 'Self-similar error clusters in communication systems and the concept of conditional stationarity', *IEEE Transactions on communication technology* **13**, 71–72.

Mandelbrot, B. (1967), 'Some noises with 1/f spectrum: a bridge between direct current and white noise', *IEEE Transactions on information theory* **13**, 289–298.

Priestley, M. (1981), *Spectral Analysis and Time Series*, Academic Press.

Stock, J. (1987), 'Asymptotic properties of LS estimators of cointegrating vectors', *Econometrica* **55**(5), 1035–1056.

White, H. (1984), *Asymptotic theory for econometricians*, Academic Press, San Diego.

Willinger, W., Taqqu, M., Leland, W. & Wilson, D. (1995), 'Self-similarity in high-speed packet traffic: analysis and modelling of ethernet traffic measurements', *Statistical Science* **10**(1), 67–85.

Felipe M. Aparicio, Department of Statistics and Econometrics, Universidad Carlos III de Madrid

E-mail address: aparicio@est-econ.uc3m.es

Miguel A. Arranz, Department of Statistics and Econometrics,Universidad Carlos III de Madrid.

E-mail address: arranz@pvi.uc3m.es

Alvaro Escribano, Department of Economics,Universidad Carlos III de Madrid. C/Madrid 126-128, Getafe. E-28903 Madrid (SPAIN).

E-mail address: alvaroe@eco.uc3m.es