Studies in Nonlinear Dynamics & Econometrics

Volume 3, *Issue* 3 1998 *Article* 1

Information-Theoretic Analysis of Serial Dependence and Cointegration

F. M. Aparicio* A. Escribano†

*Universidad Carlos III de Madrid

†Universidad Carlos III de Madrid

http://www.bepress.com/snde

ISSN: 1558-3708

This volume was previously published by MIT Press.

Information-Theoretic Analysis of Serial Dependence and Cointegration

F. M. Aparicio*
A. Escribano†
***Department of Statistics and Econometrics**
†**Department of Economics**
Universidad Carlos III de Madrid
28903 Getafe (Madrid), Spain
aparicio@est-econ.uc3m.es

Abstract. *This paper is devoted to presenting wider characterizations of memory and cointegration in time series, in terms of information-theoretic statistics such as the entropy and the mutual information between pairs of variables. We suggest a nonparametric and nonlinear methodology for data analysis and for testing the hypotheses of long memory and the existence of a cointegrating relationship in a nonlinear context. This new framework represents a natural extension of the linear-memory concepts based on correlations. Finally, we show that our testing devices seem promising for exploratory analysis with nonlinearly cointegrated time series.*

Keywords. information-theoretic statistics, long memory, cointegration, nonlinearity

Acknowledgments. This research, initiated at the Department of Economics (Aparicio and Granger 1995a) and the Institute for Nonlinear Science of the University of California at San Diego, was partially supported by the Spanish DGICYT under grant PB95–0298. We are grateful to Ignacio Pena and to Z. Ding for kindly providing the foreign exchange rate and the stock-return series, respectively, and to Matt Kennel for helping with the software. Any remaining errors or inconsistencies are entirely of our responsibility.

1 Introduction

Many economic time series exhibit important random changes in their mean behavior. These series are sometimes said to be *integrated*, in the sense that it is possible to simulate the most important features in their patterns with sums of an increasing number of weakly dependent random variables. Integrated series can be expressed in terms of *unobserved components*, where one of the components is a *stochastic trend*. The fact that remote shocks have a persistent influence on the levels of these series is known as *long memory* or *extended memory*, depending on whether this influence is linear or nonlinear (Granger 1995).

In some cases, the accumulated changes in mean behavior may be correlated across series. In the context of macroeconomics and finance, certain models suggest the presence of economic or social forces preventing two or more series from drifting too far apart from each other. Pairs of series which exhibit a common long-memory component or stochastic trend are said to be *cointegrated*. The concept of *cointegration* was coined by Granger (1981), and later on developed by Engle and Granger (1987). Well-known examples of cointegrating relationships can be found between income and expenditure, prices of a particular good in different markets, interest rates in different parts of a country, etc.

Underlying the idea of cointegration is the existence of a long-run *equilibrium* (i.e., a deterministic relationship that holds on the average for the levels) between two integrated variables, x_t, y_t. A strict (linear) equilibrium exists when for some $a \neq 0$, one has $y_t = ax_t$. This unrealistic situation is replaced, in practice, by that of a (linear) cointegrating relationship, in which the equilibrium error $z_t = y_t - ax_t$ is different from zero but fluctuates around this value much more frequently than each of the individual series (i.e., z_t is mean-reverting), and the size of these fluctuations is much smaller.

During the last years, nonlinear time-series models have been a pole of attraction of econometric research (Granger and Teräsvirta [1993] offer an overview of recent contributions). Indeed, economic theory suggests the presence of nonlinearities in many economic variables, as well as in their relationships. Such nonlinear behavior can, for example, take the form of asymmetries in adjustment costs and convexities in production (for example, see the work of Escribano and Pfann [1998]). A growing number of authors are considering nonlinear versions of the classical cointegration model. For example, Granger and Hallman (1991) proposed the definition of a *nonlinear attractor*, whereby the strength of attraction toward a linear long-run equilibrium depends on the level of the series. Escribano and Mira (1997) presented a different definition of nonlinear cointegration, based on the *near-epoch dependence* concept, and Balke and Fomby (1997) proposed a *threshold-cointegration* model, that is, one that switches between different regimes, and which seems to mimic well the nonlinear adjustment process describing many economic phenomena.

A most difficult aspect of nonlinear cointegration is that of testing. A common belief of many authors is that conventional cointegration tests may have low power for most forms of nonlinearity in a long-run relationship between the variables (Aparicio 1995; Aparicio and Granger 1995b; DeJong 1992; Granger and Hallman 1991; Schotman and Van Dijk 1991; Sims 1988). A consequence of this is that many pairs of series that are considered noncointegrated by standard tests could, in fact, have a *nonlinear equilibrium* relationship. It is therefore important to investigate new methods that are capable of detecting long-run equilibriums other than linear.

There have been few attempts to address this problem. One was due to Hallman (1990), who proposed applying standard cointegration tests (*unit-root* tests) to the ranks rather than to the levels of the series in order to make these tests more robust against monotonic nonlinear transformations of cointegrated variables. However, this strategy is unable to cope with more complex types of nonlinearity in the relationship of the variables. Moreover, Hallman's approach relies on an invariance assumption regarding the distributional properties of the conventional tests when applied to the ranks.

Granger and Hallman (1991) proposed estimating the nonlinear transformations using a nonparametric technique known as the *alternate conditional expectation* (ACE) algorithm (Breiman and Friedman 1985). This was followed by a standard cointegration test applied on the transformed variables, obtained as the ACE estimates. These estimates also allowed the possibility of checking the hypothesis of linearity in cointegration. But as remarked by these authors, it is unclear how nonparametric estimators of the transformations affect the asymptotic distribution of the standard cointegration test statistics.

More recently, Balke and Fomby (1997) proposed a two-step threshold-cointegration testing device by adapting the Engle-Granger approach. This was motivated by their findings of a certain robustness of standard cointegration tests in the presence of threshold nonlinearities in the long-run relation. Thus the analysis could be focused first on the global behavior (test for cointegration) and later on the local one (test for threshold nonlinearity). However, this test is tailored to the specific nonlinear cointegration model considered by these authors (here a threshold cointegration model), who also leave open the question of testing more general forms of nonlinear cointegration. Their results also suggest a higher effectiveness of nonparametric testing devices for detecting cointegration in this nonlinear context.

All of these findings call for a nonparametric characterization of cointegration which could be used to check this hypothesis in the general context (i.e., where any form of nonlinearity is allowed), and more importantly, where prior estimation of the nonlinear relation is not required.

In this paper, we first review the concepts of mean reversion, short and long memory, and cointegration, and introduce a new characterization of these properties using information-theoretic ideas. This will lead us to propose some new schemes for exploratory data analysis and for testing the hypotheses of long memory and cointegration between two long-memory time series. Although we only address here the bivariate case, these ideas could be extended to a multivariate context. The derivation of an asymptotic theory for the test statistics that we propose is beyond the scope of this paper, but work on this subject is currently in progress.

The rest of the paper is structured as follows. Section 2 introduces a general framework for the analysis of mean reversion, short (long) memory, and cointegration, in a nonlinear context. Section 3 presents the information-theoretic tools to be used later. In particular, we review the definitions of *entropy* and *mutual information* for random variables and stochastic processes. In Section 4, we propose an interpretation of nonlinear dependence in and among time series using the previous tools, which leads us to a more general definition of long memory and cointegration. In Section 5, we turn the previous characterization into exploratory tests of long memory and cointegration. We also provide some simulation results for the cointegration test, and apply it to pairs of financial series from a stock and a foreign exchange-rate market. Finally, Section 6 gives a concise summary of the paper.

2 Toward a General Characterization of Memory and Cointegration

Standard definitions for long memory and cointegration are inadequate when dealing with non-Gaussian and nonlinear time series, and with pairs of series which are nonlinearly related. In the first case, the trouble is that the autocorrelation function (ACF) fails to capture the higher-order dependencies in the data. In the second case, the problem lies with series which do not appear to be "aligned" in their mean behavior, but which could be cointegrated after being nonlinearly transformed. In fact, what we need is a nonlinear measure of serial dependence, and to reformulate the cointegration concept in terms of the latter.

2.1 A general characterization of memory in time series

The standard characterization of memory in a time series x_t is given in terms of its ACF, say $\rho_x(\tau, t) = \frac{\mathrm{cov}(x_t, x_{t-\tau})}{\sigma_{x_t}\sigma_{x_{t-\tau}}}$, which we consider to be generally dependent on a time index, so as to allow for some heterogeneity.

Definition 1. *A process x_t is said to be mean reverting if $\forall t$, $\lim_{\tau\to\infty} \rho_x(\tau, t) = 0$.*

Intuitively, the process x_t is mean reverting if $x_t - E(x_t)$ changes sign frequently enough. When the process is not mean reverting, its memory span is larger and $\lim_{\tau\to\infty} |\rho_x(\tau, t)| > 0$. Thus any two infinitely distant variables from such a process are still correlated (persistent behavior). However, even for a mean-reverting process, the memory span can be very large in the sense that its ACF decays very slowly as τ grows. This motivates the distinction between short and long memory.

Definition 2. *A process x_t is said to have short memory (in short, $I(0)$) if $\forall t$, $\exists b_t < \infty$ such that $\lim_{T\to\infty} \sum_{\tau=1}^{T} |\rho_x(\tau, t)| = b_t$.*

Definition 3. *A process x_t is said to have long memory if $\forall t$, $\lim_{T\to\infty} \sum_{\tau=1}^{T} |\rho_x(\tau, t)| = \infty$.*

Definition 4. *A time series of x_t is said to be integrated of order d (in short, $I(d)$), if $\lim_{T\to\infty} \sum_{\tau=1}^{T} |\rho_x(\tau, t|) = \infty$, $\forall t$, and d is the smallest positive real number such that $\lim_{T\to\infty} \sum_{\tau=1}^{T} |\rho_z(\tau, t)| < \infty$, $\forall t$, with $z_t = (1-B)^d x_t$.*

The parameter d that appears in this latter definition serves to quantify the memory span in the series. The previous characterization of memory in terms of the ACF is adequate for Gaussian series, since all of the dependence structure is captured by its second-order moments. With non-Gaussian time series, in particular, nonlinear time series, the ACF cannot provide a full account of the serial dependence structure. A first attempt to establish a general characterization of memory in a non-Gaussian context was due to Granger and Teräsvirta (1993). They proposed a general definition of mean reversion in terms of the conditional distribution function of the process. Let X_t denote the random variable at time t from a time series of a stochastic process x_t, and let $F_h(x) = P(X_{t+h} \leq x \mid I_t)$ represent the conditional distribution function of the random variable X_{t+h} given its h-horizon past, $I_t = \mathcal{F}_x^{-\infty,t}$, where $\mathcal{F}_x^{-\infty,t}$ denotes the σ-field generated by the random variables $X_t, X_{t-1}, \ldots$.

Definition 5. *A process x_t has no extended memory if $\lim_{h\to\infty} F_h(x)$ does not depend on the conditioning past, I_t.*

As a consequence, for any Borel sets C_1, C_2 and for any integer k such that $P(X_{t-k} \in C_2) > 0$, we would have

$$\lim_{h\to\infty} \mid P(x_{t+h} \in C_1 | x_{t-k} \in C_2) - P(x_{t+h} \in C_1)| = 0 \quad (1)$$

This property reminds us of the concept of *ϕ-mixing*, since it means that the dependence among temporarily nonoverlapping blocks of random variables from the process vanishes in the limit, when the temporal distance between the blocks becomes infinite.

In the sequel, we propose an alternative generalization of the memory concept for time series based on a measure of serial dependence which generalizes the ACF, such as the *mutual information measure* proposed in Section 4. Suppose $i_x(\tau, t)$ is such a non-negative measure of serial dependence that captures the higher-order dependency structure in the series. A most general characterization of mean reversion, short and long memory, and of integration can then be proposed. A process x_t could be said to be:

- mean reverting in information, if $\forall t \lim_{\tau\to\infty} i_x(\tau, t) = 0\ \forall t$;
- short memory in information (in short, $II(0)$), if $\forall t \lim_{T\to\infty} \sum_{\tau=1}^{T} i_x(\tau, t) < \infty$;
- long memory in information, if $\forall t \lim_{T\to\infty} \sum_{\tau=1}^{T} i_x(\tau, t) = \infty$;[1] or
- integrated of order d in information, say $x_t \sim II(d)$, if $\lim_{T\to\infty} \sum_{\tau=1}^{T} i_x(\tau, t) = \infty,\ \ \forall t$, and d is the smallest positive real number such that $\lim_{T\to\infty} \sum_{\tau=1}^{T} i_z(\tau, t) < \infty,\ \ \forall t$, with $z_t = (1-B)^d x_t$.

Remark 1. In principle, the function $i_x(\tau, t)$ could be any serial dependence measure capable of capturing nonlinear dependencies between the variables in the series. Remark that $\sum_{\tau=1}^{\infty} i_x(\tau, t)$ rather than $\sum_{\tau=1}^{\infty} \rho_x(\tau)$, with $\rho_x(\tau)$ representing the ACF of x_t, should be used as a persistence measure for non-Gaussian time series.

Remark 2. Note that the rates of convergence of $i_x(\tau, t)$ toward 0 as $\tau \to \infty$ are different for long- and for short-memory processes. A short-memory process is also mean reverting, according to these definitions.

2.2 A general characterization of cointegration

A minor modification of the standard definition of cointegration, due to Granger (1981), goes as follows:

Definition 6. *Two long-memory time series x_t, y_t, with long-memory parameter d, are said to be (linearly[2]) cointegrated if $\exists\, a \in \Re - \{0\}$ such that the series $z_t = y_t - a x_t$ is $I(d_z)$ with $d_z < d$.*

Figure 1 illustrates a simulation example of linear cointegration with a pair of correlated random walks ($d = 1$) and for $a = 0.72$. The scatter plot clearly shows the linearity of the relationship between x_t and y_t.

An important shortcoming in this definition of cointegration is that it requires the cointegrating relationship between the series to be linear. As as consequence, classical cointegration testing techniques relying on these definitions yield misleading results when nonlinearity enters the true equilibrium relationship. Evidence of this problem with Definition 6 was first reported by Hallman (1990), who proposed applying standard cointegration tests to the ranks rather than the levels of the series. However, even though this trick succeeds in making the test more robust against monotonic nonlinearities, it fails when confronted with general forms of nonlinearity.

In general, it should be possible to find time series that are cointegrated only after applying certain nonlinear transformations to them. Indeed, an extension of the (linear) cointegration concept follows by noticing that the common low-frequency component may "live" in a moment of higher-order than the mean; that is, in nonlinear transformations of the series. For example, x_t and y_t could be cointegrated when squared, while being more or less uncorrelated in their levels. To explain, suppose $y_t = x_t \epsilon_t$, with x_t an $I(1)$ series, and ϵ_t a zero mean iid sequence, and thus $y_t \sim I(0)$. It follows that $y_t^2 = \sigma_\epsilon^2 x_t^2 + (\epsilon_t^2 - \sigma_\epsilon^2) x_t^2$, where the rightmost term must be short memory, since it is the product of an $I(0)$ process $(\epsilon_t^2 - \sigma_\epsilon^2)$ and an $I(1)$ process (x_t^2). Thus y_t^2 is linearly cointegrated with x_t^2, although y_t is not cointegrated with x_t.

Example 1. Consider the following nonlinear factor model:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} a \\ 1 \end{pmatrix} w_t + \begin{pmatrix} -b \\ 0 \end{pmatrix} w_t^2 + \begin{pmatrix} v_t \\ \xi_t \end{pmatrix}, \tag{2}$$

where $a \neq 0$, $w_t = w_{t-1} + \epsilon_t$ with $w_0 = 0$, and (v_t, ξ_t, ϵ_t) are independent sequences of independent and identically normally distributed random variables with zero mean and joint covariance matrix equal to the

[1] As shown by example later in the paper, $i_x(\tau, t)$ need not converge to zero as τ tends to infinity for a long-memory process.

[2] In the work of Granger (1983), there is no explicit mention of the term "linear," although it is implied.

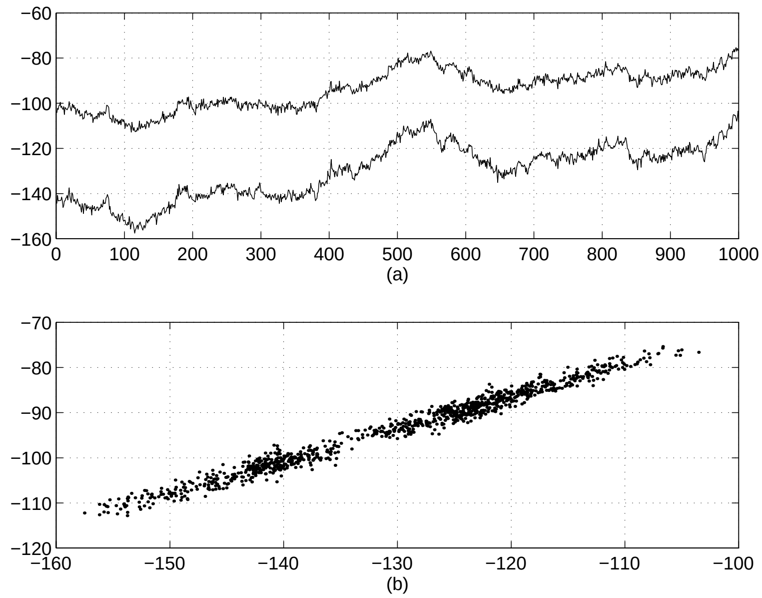

Figure 1
Two simulated linearly cointegrated random walks (a), and their scatter plot (b). The series x_t, y_t were generated with the model $y_t = aw_t + v_t$, $x_t = w_t + \xi_t$, and $w_t = w_{t-1} + \epsilon_t$, with $w_0 = 0$, and where ϵ_t, v_t, ξ_t were independent sequences of iid Gaussian random variables.

identity matrix. Let $\beta'_{l,\perp} = (a, 1)$, and let $\beta'_{n,\perp} = (-b, 0)$. Thus the orthogonal complements of $\beta'_{l,\perp}$ and $\beta'_{l,\perp}$ are, respectively, $\beta'_l = (1, -a)$ and $\beta'_n = (0, b)$. The nonlinear cointegrating relationship can be obtained as

$$\begin{aligned} z_t &= \beta'_l \begin{pmatrix} y_t \\ x_t \end{pmatrix} + \beta'_n \begin{pmatrix} y_t \\ x_t^2 \end{pmatrix} \\ &= y_t - ax_t + bx_t^2. \end{aligned} \tag{3}$$

Thus the cointegration errors are given by $z_t = 2bw_t\xi_t + b\xi_t^2 + v_t - a\xi_t$, and can be easily shown to be short memory according to our definition.

Figure 2 illustrates a simulation experiment of nonlinear cointegration with series having a common factor, and obtained with the model shown by Equation (2), with $a = 2.0$ and $b = 0.05$. Figure 3 shows a real example of an apparently nonlinearly cointegrating relationship. In both cases, the scatter plots clearly show that the dependence between the variables is not linear.

Some previous concepts of nonlinear cointegration are the following:

Definition 7 (Granger and Hallman 1991). *A pair of series x_t, y_t, are said to have a cointegrating nonlinear attractor if there are nonlinear measurable functions $f(.)$, $g(.)$ such that $f(x_t)$ and $g(y_t)$ are both $I(d)$, $d > 0$, and $z_t = g(y_t) - f(x_t)$ is $\sim I(d_z)$, with $d_z < d$.*

Remark 3. Assuming that f and g can be expanded in a Taylor series up to some order $p \geq 2$ around the origin, we may write $z_t = c_0 + c_1 u_t + \mathrm{HOT}(x_t, y_t)$, where $u_t = y_t - ax_t$, and HOT denotes higher-order terms. It follows that the linear approximation, u_t, to the true cointegration residuals differs from the latter by some higher-order terms. These terms express that the strengh of attraction onto the cointegration line $y_t = ax_t$ may vary with the levels of the series, x_t, y_t, when nonlinearities exist in their relationship.

As stated in the introduction, a difficulty with the application of this definition is the need to find proper estimates of the cointegrating functions $f(.)$ and $g(.)$ in order to test for cointegration.

Escribano and Mira (1997) propose the following definition of nonlinear cointegration, based on the concepts of *α-mixing* (Rosenblatt 1974) and *near-epoch dependence* (NED) (Wooldridge 1986).

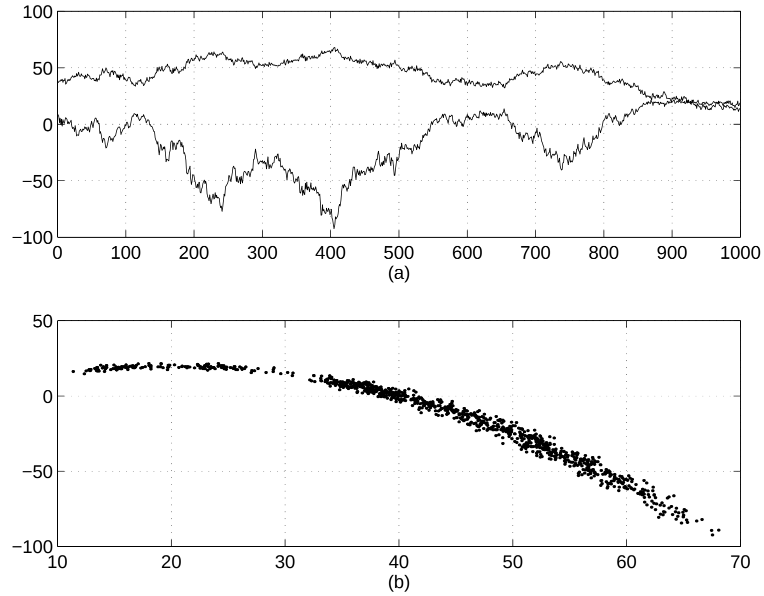

Figure 2
Two simulated nonlinearly cointegrated series (a), and their scatter plot (b). The upper series was obtained as $x_t = w_t + \xi_t$, where $w_t = w_{t-1} + \epsilon_t$ with $w_0 = 0$, and the lower series corresponds to $y_t = 2w_t - 0.05w_t^2 + v_t$. The errors ϵ_t, v_t, and ξ_t are independent sequences of iid Gaussian random variables.

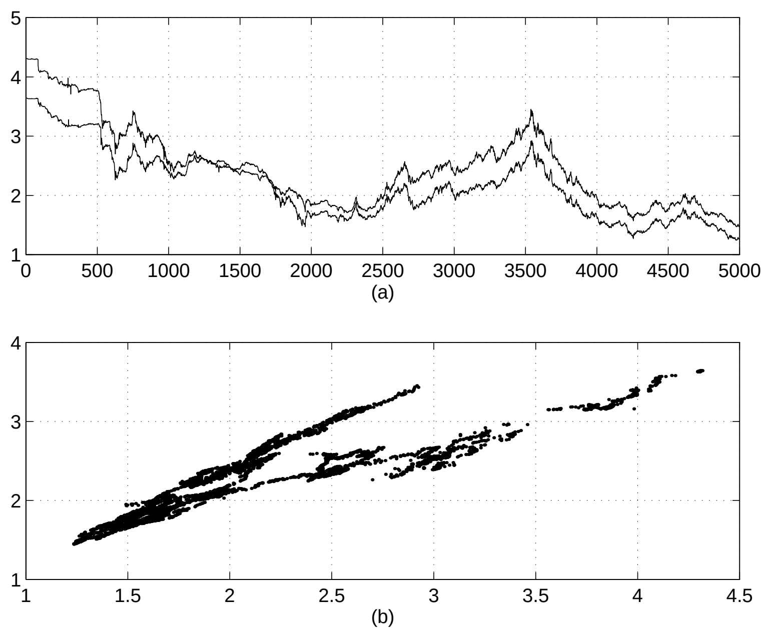

Figure 3
Two apparently nonlinearly cointegrated time series of stock prices from a Japanese food company, Ajinomoto (a). Clearly, the strength of attraction varies across time, as shown in the scatter plot (b).

Definition 8 (Escribano and Mira 1997). *A pair of series x_t, y_t are nonlinear cointegrated with cointegration function $g(\cdot, \cdot, \gamma)$ (where γ is a parameter), if $g(y_t, x_t, \gamma^*)$ is NED (α-mixing) on some α-mixing series, but $g(y_t, x_t, \gamma)$ is not NED (α-mixing) for any $\gamma \neq \gamma^*$.*

The main drawbacks of this definition are that it relies on concepts of dependence that are usually difficult to test in practice, and that it requires the consistent estimation of nonlinear functions of nonstationary variables.

We propose now an alternative definition by looking at the relative asymptotic behavior of our general measure of serial dependence, $i_{x,y}(\tau, t)$, and that of a general measure of serial cross-dependence, say $i_{x,y}(\tau, t)$, between a pair of series x_t, y_t. This latter measure reduces to $i_x(\tau, t)$ when $y_t = x_t$.

Definition 9. *A pair of time series x_t, y_t that are long memory in information are said to be cointegrated in information (in short, CII) if*

$$\lim_{\tau\to\infty} \frac{i_{x,y}(\tau, t)}{i_x(\tau, t)} = 1, \quad \forall > \tau. \tag{4}$$

Intuitively, the definition states that under cointegration, the remote past of y_t should be as useful as the remote past of x_t in the long-term forecasting of x_t.

Remark 4. This more general characterization of cointegration relies on the different limit behavior of $i_x(\tau, t)$ and $i_{x,y}(\tau, t)$, under noncointegration. Notice that when cointegration holds, we cannot have different convergence rates for $i_x(\tau, t)$ and for $i_{x,y}(\tau, t)$. The possibly different rates of convergence could be used to construct a measure of the degree of noncointegration. Suppose that $i_x(\tau, t) \sim \tau^{-\alpha}$, and that $i_{x,y}(\tau, t) \sim \tau^{-\beta}$ for τ large enough. In numerical applications we may find that neither $i_{x,y}(\tau, t)$ nor $i_x(\tau, t)$ is either infinite or zero for any finite τ. So we may safely take the logarithm of the ratio $i_x(\tau, t)/i_{x,y}(\tau, t)$ and plot it as a function of $\log \tau$. This function will tend toward an asymptote as τ grows to infinity. The slope of this asymptote is given by $\alpha - \beta$, and it is always non-negative, since we expect that $\alpha \leq \beta$. Thus the larger its value and the more unlikely the hypothesis of information cointegratedness between the series.

Remark 5. If we replace $i_x(\tau, t)$ by the ACF of x_t, and $i_{x,y}(\tau, t)$ by the cross-correlation function between x_t and y_t, say $\rho_{x,y}(\tau, t)$, then our definition amounts to comparing the behavior at the origin of the spectral densities of the series (Aparicio and Escribano 1998).[3]

Let $S_n^{(x,y)} = \sum_{\tau=1}^{n} i_{x,y}(\tau, t)$. A necessary condition for cointegration in information is:

Proposition 1. *If the series y_t, x_t are cointegrated in information, then the sequence of partial sums $S_n^{(x,y)}$ diverges as $n \to \infty$.*

Proof. Suppose the series are cointegrated in information. Then from our definition, it follows that there exists a finite real number C such that $\lim_{n\to\infty} S_n^{(x,y)} = \lim_{n\to\infty} S_n^{(x,x)} + C$. Also, the divergence of $S_n^{(x,y)}$ follows from the divergence of $S_n^{(x,x)}$, since x_t has long memory in information. ∎

3 Some Information-Theoretic Measures of Data Variability and Dependence

In this section, we present the information-theoretic concepts that form the basis of our new characterization of the relationship between integrated time series.

3.1 Information-theoretic measures for partitions

A most basic problem in information theory is that of assigning a measure of uncertainty to the occurrence or nonoccurrence of any event in a partition $\mathcal{P}$ of the set of outcomes of an underlying experiment. We call this measure of uncertainty the *entropy* of the partition, and denote it by $H(\mathcal{P})$. The construction of this functional stems from some postulates which must be satisfied in order to provide such a measure of uncertainty.

[3]The limit number 1 in Equation 4 could be changed into a constant value $b \neq 0$, as discussed by Aparicio and Escribano (1999). The appropriate value of b depends on the particular concept of dependence used in each case.

Suppose now that we have a partition of a sample space $\mathcal{S}$ with M events $\mathcal{A}_i$, $i = 1, \ldots, M$, and that the event $\mathcal{A}_i$ occurs with probability p_i. It can be shown that the convex functional

$$H(\mathcal{P}) = -\sum_{i=1}^{M} p_i \log(p_i) \tag{5}$$

yields a proper measure of average uncertainty in the partition $\mathcal{P}$.

Similarly, when we know about the occurrence of a subset $\mathcal{M}$ of events from a different partition, $\mathcal{Q}$ of $\mathcal{S}$, the remaining uncertainty in the partition $\mathcal{P}$ can be measured by the non-negative functional

$$H(\mathcal{P} \mid \mathcal{M}) = -\sum_{i=1}^{M} P(\mathcal{A}_i \mid \mathcal{M}) \log P(\mathcal{A}_i \mid \mathcal{M}), \tag{6}$$

which is called the *conditional entropy* of $\mathcal{P}$ given $\mathcal{M}$. Notice that if the events in $\mathcal{P}$ are independent of those in $\mathcal{M}$, then $H(\mathcal{P} \mid \mathcal{M}) = H(\mathcal{P})$. In general, $\mathcal{M}$ may convey information about the events in $\mathcal{P}$, and this mutual information can be quantified by the functional

$$I(\mathcal{P}, \mathcal{M}) = H(\mathcal{P}) - H(\mathcal{P} \mid \mathcal{M}). \tag{7}$$

That is, the observation of $\mathcal{M}$ reduces the uncertainty about $\mathcal{P}$ from $H(\mathcal{P})$ to $H(\mathcal{P} \mid \mathcal{M})$, so the information that $\mathcal{M}$ conveys about $\mathcal{P}$ is just $I(\mathcal{P}, \mathcal{M})$. Notice that $\mathcal{M}$ can convey at most $H(\mathcal{P})$ bits of information about the events in $\mathcal{P}$, and since $H(\mathcal{P} \mid \mathcal{M}) < H(\mathcal{P})$, $I(\mathcal{P}, \mathcal{M})$ must also be non-negative.

Now let us denote by $H(\mathcal{P}, \mathcal{Q})$ the *joint entropy* functional for the partition whose events are the intersections of the events in $\mathcal{P}$ and $\mathcal{Q}$. The resulting partition is called a *refinement* of both $\mathcal{P}$ and $\mathcal{Q}$. Notice that to observe the joint partition we must observe both $\mathcal{P}$ and $\mathcal{Q}$. It follows that the uncertainty in the joint partition must be at least equal to that of the elementary partitions. Rigorously speaking, by convexity of the entropy functional it is easy to show that $H(\mathcal{P}, \mathcal{Q}) \geq H(\mathcal{P})$ and that $H(\mathcal{P}, \mathcal{Q}) \geq H(\mathcal{Q})$ (i.e., Papoulis [1991]). In fact, we have

$$\begin{aligned} H(\mathcal{P}, \mathcal{Q}) &= H(\mathcal{Q}) + H(\mathcal{P} \mid \mathcal{Q}) \\ &= H(\mathcal{P}) + H(\mathcal{Q} \mid \mathcal{P}) \qquad (8) \\ &\leq H(\mathcal{P}) + H(\mathcal{Q}). \qquad (9) \end{aligned}$$

Clearly, the maximum value of $H(\mathcal{P}, \mathcal{Q})$ is attained when $\mathcal{P}$ and $\mathcal{Q}$ are independent. Also, by manipulating Equations 7 and 8, we obtain

$$I(\mathcal{P}, \mathcal{M}) = H(\mathcal{P}) + H(\mathcal{Q}) - H(\mathcal{P}, \mathcal{Q}). \tag{10}$$

3.2 Information-theoretic measures for random variables

So far we have introduced the concept of entropy of a given partition of the sample space of an experiment. It is possible to define the entropy of a random variable by forming a suitable partition. This is straightforward for discrete-valued random variables. For example, if a random variable X takes a countable set of values $\{x_i\}$, $i = 1, 2, \ldots$, with probabilities p_i, we can form the partition in which each event corresponds to a different value of X. Thus the definition of entropy as given in the previous paragraph also applies here, and we can define the entropy of the random variable X as

$$H(X) = -\sum_{i} p_i \log(p_i). \tag{11}$$

The definitions for the rest of the uncertainty measures discussed in the preceeding paragraph, such as conditional and joint entropies and the mutual information, also remain valid in this case.

When dealing with continuous-valued random variables, the extension of these concepts is not immediate. The difficulty here is that the events $\{X = x_i\}$ no longer form a partition, since they are not countable. Therefore, to define the entropy we must first convert X into a discrete-valued random variable. That is, we can define the entropy of a quantized version of X given by $X_\delta = m\delta$ if $X \in (m\delta - \delta, m\delta]$. If we assume that X has a probability density function (pdf), $f_x()$ is then easy to show as

$$\lim_{\delta \to 0} \left[H(X_\delta) + \log \delta \right] = -\int_{-\infty}^{\infty} f_x(X) \log f_x(X) \, dX. \tag{12}$$

We remark that $\lim_{\delta \to 0} H(X_\delta) = \infty$. However, in practice, we can only observe X with finite accuracy because of noise and quantification errors from the measurement instrument. Since the term $-\log \delta$ only reflects this lack of observation accuracy (which is instrument-dependent), we may define an uncertainty measure intrinsic to the variable, by leaving this term out:

$$h(X) = -\int_{-\infty}^{\infty} f_x(X) \log f_x(X) \, dX. \tag{13}$$

However, contrary to the entropy of a partition, the latter measure can take negative values, and thus it does only have sense when used to measure changes in uncertainty. This is why it is often referred to as *differential entropy*. In the same way, we may define joint and conditional differential entropies for any two continuous random variables, X, Y:

$$h(X, Y) = -E(\log f_{x,y}(X, Y)), \tag{14}$$

$$h(X \mid Y) = -E(\log f_{x|y}(X)), \tag{15}$$

where $f_{x,y}(\,,\,)$ and $f_{x|y}()$ denote the joint and conditional pdfs of the variables (respectively), and $E(.)$ is the expectation operator. Clearly, when X is independent of Y, we have $h(X, Y) = h(X) + h(Y)$, and $h(X \mid Y) = h(X)$. The previous expressions generalize straightforwardly to more than two variables.

In general, the different information-theoretic concepts discussed for partitions also apply to continuous-valued random variables as long as they only refer to differences of entropies. Thus the mutual information for continuous random variables, defined as

$$I(X, Y) = h(X) + h(Y) - h(X, Y), \tag{16}$$

$$= E\left[\log \frac{f_{x,y}(X, Y)}{f_x(X) f_y(Y)}\right], \tag{17}$$

conveys the same idea of dependence among the variables as for partitions.

For the purpose of illustration, we give the values of these information-theoretic quantities for Gaussian random variables.

Let X, Y be two jointly Gaussian random variables, such that $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$, and suppose that their joint pdf is given by

$$f_{x,y}(X, Y) = \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_x\sigma_y} e^{-\left((X-\mu_x)^2/\sigma_x^2 + (Y-\mu_y)^2/\sigma_y^2 - 2\rho(X-\mu_x)(Y-\mu_y)/(\sigma_x\sigma_y)\right)}, \tag{18}$$

where ρ is the correlation coefficient between the X and Y variables. Then the following can be shown (i.e., as by Papoulis [1991]):

$$h(X) = \log(\sigma_x \sqrt{2\pi e}), \tag{19}$$

$$h(Y) = \log(\sigma_y \sqrt{2\pi e}), \tag{20}$$

$$h(X, Y) = \log(2\pi e) + \log(\sqrt{\Delta}), \tag{21}$$

$$h(X \mid Y) = \log(\sigma_x \sqrt{2\pi e}) + \frac{1}{2}\log(1-\rho^2), \tag{22}$$

$$I(X, Y) = -\frac{1}{2}\log(1-\rho^2), \tag{23}$$

where Δ is the determinant of the variance-covariance matrix of the variables, that is, $\Delta = \sigma_x^2\sigma_y^2(1-\rho^2)$. In general, given n jointly Gaussian random variables, $X_1, \ldots, X_n$, with variance-covariance matrix Σ, the joint differential entropy is given by

$$h(X_1, \ldots, X_n) = \frac{n}{2}\log(2\pi e) + \log(\sqrt{\Delta}), \tag{24}$$

where Δ is the determinant of Σ.

3.3 Information-theoretic measures for stochastic processes

Stochastic processes are defined in terms of the joint distributions for all subsets of their random variables. In particular, the information gained when the m random variables $X_{t_1}, \ldots, X_{t_m}$ of a continuous-valued stochastic process x_t are observed, is given by their m^{th}-order joint-differential entropy, defined as

$$h(X_{t_1}, \ldots, X_{t_m}) = -E\left(\log f_{t_1,\ldots,t_m}(X_{t_1}, \ldots, X_{t_m})\right). \tag{25}$$

Obviously, the uncertainty about the values of x_t on any finite interval of t is infinite. However, if x_t can be expressed in terms of its samples on a countable set of sampling instants $\{t_i\}_i$ (i.e., to the extent that x_t can be approximated by a narrow-band process), it may be possible to define entropy measures. Henceforth we will assume that this is the case. Now, if there exists a conditional stationary pdf for x_t, we can define a measure of the uncertainty about any variable of the process, when its most recent values are known. For example, the m^{th}-order (differential) conditional entropy of x_t, $h(X_n \mid X_{n-1}, \ldots, X_{n-m})$, captures the remaining uncertainty about any random variable from x_t, when information about its m^{th} history has been collected. This functional is, obviously, decreasing in m, and its rate of decay contains important information about the type of serial dependence in the process. For $m \to \infty$ we obtain a measure of the unknown information about any variable X_n once we know its entire past. Clearly, for a deterministic process, this measure, call it $h_r(x) = \lim_{m\to\infty} h(X_n \mid X_{n-1}, \ldots, X_{n-m})$, equals zero. It is customary to call $h_r(x)$ the *entropy rate* of the process x_t. This name acknowledges the fact that when x_t is stationary, we can write

$$h_r(x) = \lim_{m\to\infty} \frac{1}{m} h(X_1, \ldots, X_m). \tag{26}$$

Clearly, the limit on the right of the previous equality measures the speed at which the uncertainty grows as we try to guess at the values of an ever-increasing number of random variables from the process.

As a way of illustration, for a wide-sense stationary Gaussian process, x_t, we have

$$h_r(x) = \log(\sqrt{2\pi e}) + \frac{1}{2}\lim_{m\to\infty} \log\left(\frac{\Delta_{m+1}}{\Delta_m}\right), \tag{27}$$

where Δ_m is the determinant of the m^{th}-order variance-covariance matrix of the process.

4 An Information-Theoretic Characterization of Memory

In the previous section, we saw that the mutual information in a pair of random variables could be interpreted as a measure of general dependence between them, in contrast with their correlation, which only measures the adequacy of any variable for linearly predicting the other. Similarly, we can establish the serial dependence and cross-dependence properties of wide-sense stationary stochastic processes, in terms of a *mutual information function* (MIF), which generalizes the standard autocorrelation function (ACF). However, in order to extend the new characterization to processes having stochastic trends, we must again allow some scope for heterogeneity, and thus our measures will in general depend on time. Let the MIF of x_t be $i_x(\tau, t) = I(X_t, X_{t-\tau})$. Our information-theoretic characterization of mean reversion and of short and long memory follows from the definitions in Section 2.1. Correspondingly, we will say that a series is either mean reverting, short memory, long memory, or integrated in information.

Remark 6. In the Gaussian case, $i_x(\tau, t)$ is related to the ACF, and thus for a Gaussian short-memory process, $i_x(\tau, t)$ must converge exponentially fast to zero. For a Gaussian long-memory process, this convergence must be slower (typically, hyperbolically fast).

Remark 7. The information quantities can be rewritten as (differential) entropy changes. That is,

$$i_x(\tau, t) = h(X_t) - h(X_t \mid X_{t-\tau}). \tag{28}$$

This supports our intuition that entropy differences are most useful for characterizing the dependence properties of a process.

Remark 8. There are some connections between Granger's most general definition of mean reversion (introduced in a previous paragraph) and the MIF. This can be seen by reinterpreting the latter as some sort of mixing of coefficients. Given a stochastic process x_t, the standard α-mixing coefficients are given by (Rosenblatt 1974)

$$\alpha(\tau, t) = \sup_t \sup_{X \in \mathcal{F}_x^{-\infty,t}; X^* \in \mathcal{F}_x^{t+\tau,\infty}} |P(X^*, X) - P(X^*)P(X)|, \tag{29}$$

where $P(.)$ is a probability measure defined on the Borel σ-field of x_t. In contrast, the information-mixing coefficients $i_x(\tau, t)$ can be expressed as

$$i_x(\tau, t) = E\left(\log f_{x,x}(X_t, X_{t-\tau}) - \log f_x(X_t) f_x(X_{t-\tau})\right), \tag{30}$$

where $f_{x,x}(,)$ and $f_x(.)$ denote the bivariate and univariate pdf for x_t. We remark that both types of mixing coefficients allow for heterogeneity in the process.

Remark 9. An alternative characterization could be made in terms of the conditional densities. Let $\mathcal{F}_{-\infty,t-\tau+1}^{t-\tau-1,t-1}$ denote the σ-field generated by the random variables $X_{t-1}, \ldots, X_{t-\tau+1}; X_{t-\tau-1}, \ldots$. A generally nonstationary time series of x_t can be said to be conditionally short memory in information, if the sequence of partial sums $R_n^{(x)} = \sum_{\tau>0}^{n} I(X_t, X_{t-\tau} \mid \mathcal{F}_{-\infty,t-\tau+1}^{t-\tau-1,t-1})$ converges as n grows to infinity. If, on the contrary, $R_n^{(x)}$ diverges, then x_t can be said to be conditionally long memory in information. These alternative definitions rephrase the former ones in terms of a partial serial-dependence measure, which can be regarded as a generalization of the concept of the partial autocorrelation function (PACF) in the linear context. However, when working with conditional densities, we may encounter severe computational difficulties (i.e., the need for very large data sets, the curse of dimensionality, etc.), which make us prefer the former approach, despite that the marginal densities are not well defined in the nonstationary context.

A few examples may help to illustrate the behavior of the new unconditional dependence measures. Consider the following cases:

- Let $x_t = ax_{t-1} + \epsilon_t$, where ϵ_t is an iid sequence of Gaussian random variables with zero mean and variance σ^2; in short, $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$, and $|a| < 1$. This model generates a stationary Gaussian Markov process, for which $\text{cov}(x_t, x_{t-\tau}) = \sigma^2 a^\tau$, which converges to zero exponentially fast as $\tau \to \infty$. The information-mixing coefficients, defined for $\tau > 0$, are given in this case by

$$i_x(\tau, t) = i_x(\tau) = -\frac{1}{2}\log(1 - a^{2\tau}), \tag{31}$$

  which clearly converges exponentially fast to zero as τ grows to ∞, thus implying that $\sum_{\tau>0} i_x(\tau, t) < \infty$. We may therefore conclude that x_t is both $I(0)$ and $II(0)$. On the contrary, if $a = 1$, we have a nonmixing process with a unit root, for which $\text{corr}(x_t, x_{t-\tau}) = 1$, and $i_x(\tau, t) = \infty$ for any τ and any t. Therefore, we may classify this $I(1)$ process as $II(1)$.

- Let x_t be a Gaussian stationary long-memory process with long-memory parameter d $(0 < d < 0.5)$; that is, $(1 - B)^d x_t = \epsilon_t$, with ϵ_t representing a stationary zero-mean short-memory Gaussian process. This mean-reverting process is characterized by an ACF which decays hyperbolically fast, that is, $\text{cov}(x_t, x_{t-\tau}) \sim \tau^{2d-1}$ for large τ (e.g., Hosking 1981), and thus we write $x_t \sim I(d)$. On the other hand, we obtain the following approximation for large τ:

$$i_x(\tau, t) = i_x(\tau) \sim -\frac{1}{2}\log(1 - c_d \tau^{4d-2}), \tag{32}$$

  where c_d is a constant depending only on d. Clearly, $i_x(\tau)$ also converges to zero, but this time the convergence is only hyperbolically fast. Noting that $\log(1 - c_d \tau^{4d-2}) \approx c_d \tau^{4d-2}$ for sufficiently large τ, the divergence of $\sum_{\tau>0} i_x(\tau, t)$ follows inmediately. Therefore, x_t is long memory in information.

Now let us look at these measures from the viewpoint of the conditional (differential) entropies. Let $h_{c,\tau}(X_t) = h(X_t \mid X_{t-1}, \ldots, X_{t-\tau})$, or equivalently, $h^*_{c,\tau}(X_t) = h(X_t \mid X_{t-\tau}, \ldots, X_{t-\infty})$. Notice that $h^*_{c\tau}(X_t) \leq h(X_t)$.

Proposition 2. If $h^*_{c,\tau}(X_t) < h(X_t)\ \forall \tau$ and $\forall t$, then the process is neither mean reverting nor short memory in information.

Proof. Let $I(X_t; X_{t-\tau}, X_{t-\tau-1}, \ldots, X_{t-\infty})$ denote the information on X_t conveyed by the variables $X_{t-\tau}, X_{t-\tau-1}, \ldots$. We can write

$$I(X_t, X_{t-\tau}) \geq I(X_t; X_{t-\tau}, X_{t-\tau-1}, \ldots, X_{t-\infty}) = h(X_t) - h(X_t \mid X_{t-\tau}, X_{t-\tau-1}, \ldots, X_{t-\infty}) \quad (33)$$

$$\geq 0, \quad (34)$$

If $h^*_{c,\tau}(X_t) < h(X_t)$, the inequality holds strictly and thus $\lim_{\tau\to\infty} I(X_t, X_{t-\tau}) > 0$, implying that x_t is neither mean reverting nor short memory in information. ∎

We shall assume in the following examples that our processes are Gaussian. Therefore, recalling Equation 27, the τ^{th}-order conditional (differential) entropy for a Gaussian process x_t is

$$h_{c,\tau}(X_t) = \log(\sqrt{2\pi e}) + \frac{1}{2}\log\left(\frac{\Delta_{\tau+1,t}}{\Delta_{\tau,t}}\right), \quad (35)$$

where $\Delta_{\tau,t}$ is the determinant of the τ^{th}-order variance-covariance matrix of x_t.

In the following, we will determine the conditional entropies and some implications for the classes of processes previously characterized in terms of the MIF.

- Let $x_t = ax_{t-1} + \epsilon_t$ where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. If $|a| < 1$, then we can write $h_{c,\tau}(X_t) = h(X_t \mid X_{t-1}) = \log\left(\sigma\sqrt{2\pi e}\right)$ for any $\tau > 0$. It follows that $I(X_t; X_{t-1}, \ldots, X_{t-\tau}) = I(X_t, X_{t-1}) = h(X_t) - h_{c,\tau}(X_t) = -\frac{1}{2}\log(1-|a|^2) < \infty$, for any $\tau > 0$. On the contrary, if $a = 1$, then $I(X_t; X_{t-1}, \ldots, X_{t-\tau})$ is infinity for any $\tau > 0$.

- Let x_t be a stationary autoregressive process of order p; in short, $x_t \sim \text{AR}(p)$. If x_t is Gaussian, then we have the following result from Kay (1988, pp. 169–178):

$$\frac{\Delta_{\tau+1,t}}{\Delta_{\tau,t}} = \frac{\Delta_{\tau+1}}{\Delta_\tau} = \sigma^2\prod_{k=1}^{\tau}(1-|r_k|^2), \quad (36)$$

  where r_k is the partial autocorrelation at lag k. Thus, at long lags,

$$\frac{\Delta_{\tau+1}}{\Delta_\tau} = \sigma^2\prod_{k=1}^{p}(1-|r_k|^2), \quad (37)$$

  since $r_k = 0$ for $k > p$. Now, since $|r_k| < 1$, $\forall k$, it follows from Equation 27 that $h_r(x)$ is bounded, and that $I(X_t; X_{t-1}, \ldots, X_{t-\infty}) < \infty$.

- Suppose x_t is a Gaussian stationary long-memory process with long-memory parameter d $(0 < d < 0.5)$. Then, since the partial autocorrelations of this process r_k satisfy $0 < r_k < 1$ for any finite k (Hosking 1981), then

$$\lim_{\tau\to\infty}\frac{\Delta_{\tau+1}}{\Delta_\tau} = \sigma^2\lim_{\tau\to\infty}\prod_{k=1}^{\tau}(1-|r_k|^2)$$

$$= 0. \quad (38)$$

  The latter implies that $h_r(x) = -\infty$, which in turns leads to an infinite value for the mutual information between X_t and its infinite history; that is, $I(X_t; X_{t-1}, \ldots, X_{t-\infty}) = \infty$.

These examples show that the persistence of the shocks in a process results in its entire past containing an infinite amount of information about its present. On the contrary, this amount of information is bounded for mixing processes.

The connection of the latter discussion with our characterization of dependence in terms of the information-mixing numbers $i_x(\tau)$ appears by noticing that each variable from the past contributes a small portion of information about the present variable, X_t. In other words, we must have

$$I(X_t; X_{t-1}, \ldots, X_{t-\infty}) \leq \sum_{\tau=1}^{\infty} i_x(\tau, t). \tag{39}$$

Now, the fact that $I(X_t; X_{t-1}, \ldots, X_{t-\infty}) = \infty$ for persistent Gaussian processes implies that $i_x(\tau, t)$ cannot decrease with τ faster than $O(\tau^{-2+\delta})$ for some $\delta > 0$. Alternatively, for stationary Gaussian processes, we obtained $I(X_t; X_{t-1}, \ldots, X_{t-\infty}) < \infty$, which is consistent with an exponentially fast decay of $i_x(\tau, t)$ for growing τ.

4.1 Some implementation issues

We briefly explain how the mutual-information quantities were estimated in the experiments that follow. The MIF, $i_x(\tau)$, was evaluated using the following estimator, where N is the sample size,

$$\begin{aligned} \hat{\imath}_x^{(N)}(\tau) &= N^{-1} \sum_{t=1}^{N} \hat{\imath}_x(\tau, t) \\ &\approx N_\gamma^{-1} \sum_{t \in \mathcal{S}} c_t(\gamma) \log \left(\frac{\hat{f}_{x,x}(X_t, X_{t-\tau})}{\hat{f}_x(X_t) \hat{f}_x(X_{t-\tau})} \right), \end{aligned} \tag{40}$$

with

$$c_t(\gamma) = \begin{cases} 1+\gamma, & \text{for } t \text{ odd,} \\ 1-\gamma, & \text{for } t \text{ even,} \end{cases}$$

where $\gamma \geq 0$, $N_\gamma = N$ for N even, and $N_\gamma = N + \gamma$ for N odd. Here X_t represents a generic vector variable, $\hat{f}_{x,x}(\ ,\ )$, and $\hat{f}_x(\ )$ are estimators of the bivariate and univariate pdfs (which may be time varying), respectively and the set $\mathcal{S}$ is introduced to make explicit the exclusion of certain inocuous summands, which can occur, for example, when $\hat{f}_{x,x}(\ ,\ ) \leq 0$ or $\hat{f}_x(\ ) \leq 0$, or when logarithms cannot be taken. The densities can be estimated using *kernel smoothers* (Breiman, Meisel, and Purcell 1977). In general, given a set of $N - n$ n-dimensional vectors $\{X_t\}_{t=1,N-n}$, a kernel-density estimator with kernel K and bandwidth α of their unconditional pdf, say f, has the form

$$\hat{f}(X) = (N-n)^{-1} \alpha^{-1} \sum_{t=1}^{N-n} K[\alpha^{-1}(X - X_t)], \tag{41}$$

where the kernel K is a function verifying $\int_{\mathfrak{R}^n} K(Y)dY = 1$. Robinson (1991) proved the consistency of a similar estimator under the assumption of stationarity in the series and for $n = 1$. For the experiments, we choose Gaussian kernels, that is

$$K(X) = (2\pi)^{-n/2} \exp(-X'X/2). \tag{42}$$

Even though the form of the kernel is not critical to the results, the bandwidth is. We can deal with this problem by means of adaptive bandwidths. This technique consists of allowing the kernels to shrink in rather densely populated regions of the n-dimensional embedding space, and to widen in regions with few data points. The likelihood of introducing important biases is greatly reduced in this way, since the smoothing becomes only important at those regions of the embedding space containing a large number of points. Initially, we took a fixed bandwidth for the kernels, say α, and the initial density estimates were subsequently plugged in to obtain locally adapted bandwidths, say $\beta(X)$, according to

$$\beta(X) \propto 1/\hat{f}_\alpha(X), \tag{43}$$

where $\hat{f}_\alpha(X)$ denotes a rough estimate of the pdf at X using a kernel estimator with the fixed bandwidth, α.

4.2 Information-theoretic characterization of cointegration

Let x_t, y_t be long memory in information. The concept of cointegration in information arises when letting $i_{x,y}(\tau, t) = I(X_t, Y_{t-\tau})$ in the characterization of cointegration proposed in Section 2.2 (see Definition 9). The information-cointegratedness concept states that for any long-run predictor of X_t based on $X_{t-\tau}$, we can find a predictor based on $Y_{t-\tau}$ which conveys exactly the same information about X_t.

Remark 10. Our characterization applies to both integer and fractionally integrated processes. Also, the processes involved are not required to have the same integration order. For instance, consider the case in which $x_t \sim II(d_x)$, $y_t \sim II(d_y)$, with $d_x \neq d_y$, and $\phi(\ ,\ )$ is a nonlinear one-to-one transformation such that $z_t = \phi(y_t) \sim II(d_z)$ with $d_z = d_x$. This situation can be understood noting that both the entropy and the mutual information of the variables in a process are invariant to one-to-one transformations of the latter (see, for instance, the work by Papoulis [1991, p. 565]).

An alternative condition for the information-cointegratedness of (x_t, y_t) can be given using conditional entropies,

$$\lim_{\tau\to\infty} \frac{h(Y_t \mid \mathcal{F}_x^{-\infty,t-\tau})}{h(Y_t \mid \mathcal{F}_y^{-\infty,t-\tau})} \neq 0, \quad \forall t, \tag{44}$$

$$\lim_{\tau\to\infty} \frac{h(X_t \mid \mathcal{F}_y^{-\infty,t-\tau})}{h(X_t \mid \mathcal{F}_x^{-\infty,t-\tau})} \neq 0, \quad \forall t. \tag{45}$$

Remark 11. The information-cointegration definition can equally handle multivariate processes, which enter naturally as arguments of the information measures.

Example 2. Consider the following linear common factor model:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} a \\ 1 \end{pmatrix} w_t + \begin{pmatrix} v_t \\ \xi_t \end{pmatrix}, \tag{46}$$

where $a \neq 0$, $w_t = w_{t-1} + \epsilon_t$ with $w_0 = 0$, and (v_t, ξ_t, ϵ_t) are independent sequences of independent and identically normally distributed random variables with zero mean and joint-covariance matrix equal to the identity matrix. If we now define $z_t = y_t - ax_t$, and

$$\rho_x(\tau, t) = \frac{\operatorname{cov}(x_t x_{t-\tau})}{\sigma_{x_t}\sigma_{x_{t-\tau}}}, \tag{47}$$

$$\rho_{x,y}(\tau, t) = \frac{\operatorname{cov}(y_t x_{t-\tau})}{\sigma_{x_t}\sigma_{x_{t-\tau}}}, \tag{48}$$

after some algebra, we obtain

$$\rho_x(\tau, t) = \frac{(t-\tau)\sigma_\epsilon^2}{\sqrt{(t\sigma_\epsilon^2 + \sigma_\xi^2)}\sqrt{((t-\tau)\sigma_\epsilon^2 + \sigma_\xi^2)}}, \tag{49}$$

$$\rho_{x,y}(\tau, t) = \frac{a(t-\tau)\sigma_\epsilon^2}{\sqrt{(a^2(t\sigma_\epsilon^2 + \sigma_\xi^2) + \sigma_z^2)}\sqrt{((t-\tau)\sigma_\epsilon^2 + \sigma_\xi^2)}}. \tag{50}$$

It follows that for sufficiently large t, $\rho_{x,y}(\tau, t) \approx \rho_x(\tau, t)$.

Now, since $i_{x,y}(\tau, t) = -\frac{1}{2}\log(1 - \rho_{x,y}^2(\tau, t))$, and $i_x(\tau, t) = -\frac{1}{2}\log(1 - \rho_x^2(\tau, t))$, it follows that $i_{x,y}(\tau, t)/i_x(\tau, t) \approx 1$ for any τ.

In Figure 4, we compare the behavior of a normalized version of the generalized sample correlations, $\hat{\imath}_{x,y}^{(N)}(\tau)/\hat{\imath}_x^{(N)}(1)$ and $\hat{\imath}_x^{(N)}(\tau)/\hat{\imath}_x^{(N)}(1)$ as functions of τ, by means of Monte Carlo simulations. Here $\hat{\imath}_{x,y}^{(N)}(\tau)$ is given by

$$\begin{aligned} \hat{\imath}_{x,y}^{(N)}(\tau) &= N^{-1}\sum_{t=1}^{N} \hat{\imath}_{x,y}(\tau, t) \\ &\approx N_\gamma^{-1} \sum_{t\in\mathcal{S}} c_t(\gamma) \log\left(\frac{\hat{f}_{x,y}(X_t, Y_{t-\tau})}{\hat{f}_x(X_t)\hat{f}_x(Y_{t-\tau})}\right), \end{aligned} \tag{51}$$

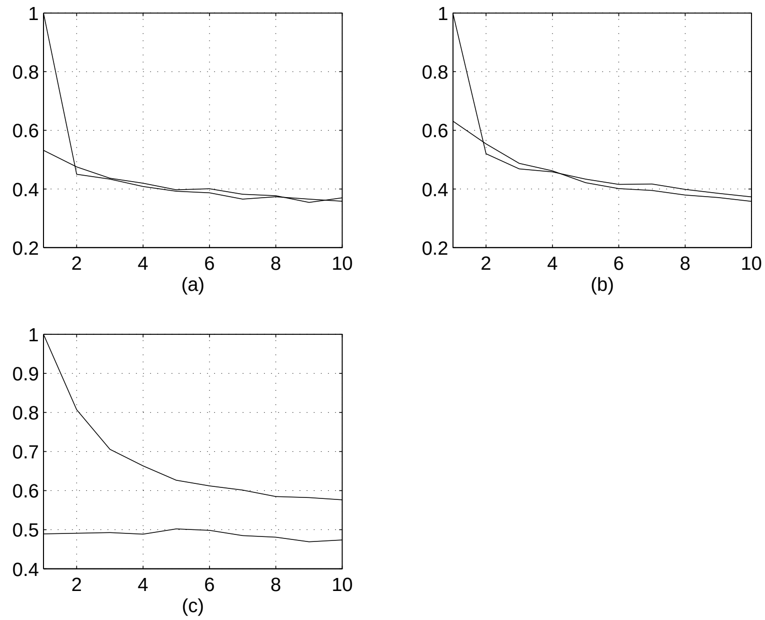

Figure 4
Plots of the generalized correlations $\hat{\imath}_{x,y}^{(N)}(\tau)/\hat{\imath}_x^{(N)}(1)$ and $\hat{\imath}_x^{(N)}(\tau)/\hat{\imath}_x^{(N)}(1)$ (vertical axes) versus $\tau+1$ (horizontal axes) for linearly (a), nonlinearly (b), and noncointegrated (c) series. The plots show the average curves obtained from 20 Monte Carlo simulations. The nonlinearly cointegrated series were generated by applying third-order polynomial transformations to a common random-walk component. The noncointegrated series were independent random walks.

and the coefficients $c_t(\gamma)$ and γ are as in the previous section. The curves shown in the figure represent statistical averages computed from 20 simulated pairs of series. Plots (a), (b), and (c) correspond to linearly cointegrated, nonlinearly cointegrated, and noncointegrated series, respectively. The horizontal scale shows $\tau + 1$. The linear cointegrated series were generated as those in Figure 1, while the nonlinearly cointegrated ones were obtained by applying third-order polynomial transformations to a common random-walk component.

5 Testing for Cointegration in Information

Testing for a cointegrating relationship usually involves two steps: (1) a test for long memory in the series; and (2) a test for a relation between the long-memory series. Long-memory and cointegration tests are commonly used to study the dynamics of financial time series. Most of the empirical work in this area has been based on classical methods. However, there is a growing consensus among econometricians and time-series analysts on the poor power properties of these methods in nonlinear settings.

In this section, we first present a long-memory testing device involving the MIF, which is potentially able to deal succesfully with nonlinearly generated time series. Then, we propose a statistic, that also involves the MIF, for testing the hypothesis of a relation in the series. Some simulation experiments suggest the possibility of doing general cointegration testing using only this statistic.

5.1 Long-memory testing

Evidence of the presence of long memory in stock price and in exchange-rate series has been reported many times (see, for instance, the works of Baillie and Bollerslev [1989]; Berg and Lyhagen [1998]; Cerchi and Havenner [1988]; DeJong and Whiteman [1991]; Greene and Fielitz [1977]; Koop [1994]; Lo [1991]) and has important implications in financial economics. For example, portfolio decisions become, under long memory, extremely sensitive to the investment horizon.

Most existing approaches to long-memory testing use linear analysis tools such as the ACF, and therefore, their results could be misleading when the series-generating mechanism is nonlinear. The older *range-scale analysis* (see Mandelbrot's [1972] work for a discussion), based on the R/S statistic proposed by Hurst (1951), and its subsequent modifications, led to some long-memory tests not involving the ACF (Davies and Harte 1987; Lo 1991). An alternative test for long memory in information for a time-series variable can be obtained by working out our characterization of short and long memory in information, as in the previous sections.

Recall that for x_t to be short memory in information we must have $\sum_{\tau>0} i_x(\tau, t) < \infty$, which implies that for any $\delta > 0$ and any t, $i_x(\tau, t) = o(\tau^{-2+\delta})$; that is, there exist positive real numbers τ_0 and b such that $i_x(\tau, t) < b\tau^{-2}$ $\forall\tau > \tau_0$ and $\forall t$. On the contrary, if x_t is long memory in information, then there exist positive real numbers τ_1, c_t, and $2 > r > 0$ such that $i_x(\tau, t) \approx c_t\tau^{-r}$ $\forall\tau > \tau_1$. Or taking logs,

$$\log i_x(\tau, t) \approx \log c_t - r\log\tau + \xi_{\tau,t}, \quad \forall\tau \gg \tau_1, \tag{52}$$

where $\xi_{\tau,t}$ is an error sequence. Therefore, we could check the property of short memory in information by testing the null hypothesis, H_0: $r \geq 2$, against the alternative H_1: $r < 2$.

A frequency-domain version of this testing device, similar in spirit to the one proposed by Geweke and Porter-Hudak (1983), allows us to do the analysis at low frequencies ($\lambda \to 0$) instead of at very long lags ($\tau \to \infty$). For this, let us first define a *generalized periodogram* as

$$G_x^{(N)}(\lambda, t) = \sum_{\tau=1}^{N} w_\tau\, i_x(\tau, t)\exp(-j2\pi\lambda\tau), \tag{53}$$

where $j^2 = -1$, w_τ is a spectral window, and N is the sample size. Now, if x_t is long memory in information, we should have

$$G_x^{(N)}(\lambda, t) \approx u_x(\lambda, t)\lambda^{-2d}, \tag{54}$$

for small λs where $d > 0$ and $u_x(\tau, t)$ is a slowly varying function of τ; that is, $\lim_{\lambda\to a} u_x(c\lambda, t)/u_x(\lambda, t) = 1$ $\forall c$, for $a = 0$ and $a = \infty$.

Again, taking logs we obtain

$$\log G_x^{(N)}(\lambda, t) = \log u_x(\lambda, t) - 2d\log\lambda + \upsilon_{\lambda,t}, \tag{55}$$

for small λs and with $\upsilon_{\lambda,t}$ representing an error sequence. Now we can test the null hypothesis of short memory in information, H_0: $d = 0$, once we have an estimate of the slope of the previous regression line.

Different procedures for estimating the parameter d in the information-integrated time-series model could be borrowed, for example, from the works of Geweke and Porter-Hudak (1983), and Robinson (1995).

5.2 Cointegration testing

Common stochastic trends in financial time series, such as interest rates, stock prices, commodity prices from goods that are close substitutes, purchasing-power parity (PPP), and cross-country exchange rates, point to common underlying factors driving them. If exchange rates and/or stock prices from different countries are found to move together in the long run, that suggests an increasing degree of international integration of capital markets. Moreover, the finding that stock prices are cointegrated allows us the possibility of rejecting the hypothesis of *strong-market efficiency* (see the work of Escribano and Granger [1998] for a recent application of nonlinear models).

The following sets of financial time series (among many others) have been found to be linearly cointegrated: exchange rates (Baillie and Bollerslev 1989); interest rates of different maturities (Engle and Granger 1987); interest rates in different countries (Akella and Patel 1991); foreign currency spot and forward rates (Barnhart and Szakmary 1991); dividends and prices (Campbell and Shiller 1987); equity markets in different countries (Taylor and Tonks 1989); stock prices within a given industry (Cerchi and Havenner 1988); and size-ranked portfolios (Bossaerts 1988). Interestingly, other sets have been found not to be linearly cointegrated, such as commodity spot and futures prices (Baillie and Myers 1991) and purchasing-power parity relationships (Corbae and Ouliaris 1988; Enders 1988).

The standard definitions and tests of cointegration assume a uniform and smooth tendency for the series to move toward a unique long-run equilibrium. This is in contrast to the behavior of many pairs of economic variables. In fact, economic agents will adjust continuously only as far as their adjustment benefits exceed their costs (Balke and Fomby 1997). For example, transaction costs in financial markets 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 nonuniformities in the adjustment of agents. These nonuniformities could translate into departures from the linear-cointegration hypothesis, or into nonlinear error-correction models (Escribano and Mira 1997; Escribano and Pfann 1998), and could eventually mask the existence of a long-run equilibrium in standard cointegration tests. Meese and Rose (1991) found that linear cointegration between exchange rates and the fundamentals of four different structural models was rejected, and suggested that this could be traced to the existence of unsuspected nonlinearities. Therefore, a cointegration-testing device contemplating the possibility of a nonlinear attractor could be an interesting alternative.

Definition 9 suggests the following test statistic, which measures the strength of the relation between a pair of series x_t, y_t:

$$c_{m,q}(x, y) = N^{-1} \sum_{t=1}^{N} \sum_{\tau=m}^{m+q} \left(1 - \hat{\imath}_{x,y}(\tau, t)/\hat{\imath}_x(\tau, t)\right), \tag{56}$$

where m is supposed to be sufficiently large so as to capture as little as possible of the short-run dependencies,[4] and q is supposed to be such that $m + q < N$, where N is the sample size.

As we pointed out in the preceding section, under cointegration $i_{x,y}(\tau, t)$ will be of the same order of magnitude as $i_x(\tau, t)$ for sufficiently large τ. On the contrary, under noncointegration, $i_{x,y}(\tau, t) \ll i_x(\tau, t) > 0$ for sufficiently large τ. This implies a tendency for the values of $c_{m,q}(x, y)$ to cluster around 1 under noncointegration.

The limiting distribution of our statistic is difficult to obtain since we are dealing with nonmixing processes, and its formal derivation is beyond the scope of this paper. However, we suggest to test the null hypothesis of cointegration by constructing an empirical confidence interval for the test statistic. That is, for fixed values of m and q, we estimate the empirical critical value b_α such that $P\left(c_{m,q}(x, y) > b_\alpha\right) = \alpha$ under the assumption of information cointegratedness, for the given significance level, α. Therefore, this hypothesis will be rejected at this level when $c_{m,q}(x, y) > b_\alpha$.

To assess the potentialities of the statistic $c_{m,q}(x, y)$ in Equation 56 as a general cointegration-testing device, we did a small Monte Carlo experiment involving 100 pairs of linearly cointegrated, nonlinearly cointegrated, and noncointegrated series. The data-generating mechanism for the linearly cointegrated series was the same as for the series in Figure 1. The nonlinearly cointegrated series were computed by applying third-order polynomial transformations to a common random-walk component. The coefficients of these polynomials were chosen at random (the program randomly selected a set of coefficients each time it generated a pair of time series). Finally, the noncointegrated series were either pairs of independent random walks ($H_{2,1}$) or mutually dependent short-memory series ($H_{2,2}$). In the latter case, the series were generated according to the model $y_t = x_t + \epsilon_t$, where $x_t = a_4 e_{t-2} e_{t-1} + e_t$, ϵ_t, e_t are mutually independent iid sequences, and the a_i were chosen at random. For the experiment, we selected $q = 0$, $m = 10$, and a sample size of $N = 1{,}000$. In the 1,000 replications done, the value of $c_{10,0}(x, y)$ was comparatively large and positive under noncointegration, but small and with varying sign under cointegration, both in the linear and the nonlinear cases. Table 1 shows the mean, standard deviation, and mean absolute value of $c_{10,0}(x, y)$ obtained in the experiment, and the histogram plots of $c_{10,0}(x, y)$ for the different cases are given in Figure 5. Using the 5% empirical critical values of this statistic under $H_{2,1}$, estimated from 1,000 Monte Carlo replicas, the power of the test (percentage rejection) approached 85% of the simulated cointegrated pairs. This is in contrast with the low-power results of standard cointegration tests when applied to nonlinearly related integrated time series, as reported in many empirical studies (DeJong 1992; Granger and Hallman 1991; Schotman and Van Dijk 1991; Sims 1988).

The comparatively large values taken by our test statistic on pairs of mutually dependent short-memory

[4] A criterion for selecting m in the linear case is suggested by the theory of filter design, and consists of choosing for this parameter the value m^* satisfying $\int_{m^*}^{\infty} |\rho_x(\tau, t)| d\tau < \delta$, where δ is a real number arbitrarily close to zero. In fact, m^* is related to the bandwidth of an optimally designed low-pass filter. This criterion can be extended straightforwardly to the more general nonlinear case by replacing $\rho_x(\tau, t)$ with $i_x(\tau, t)$.

Table 1
Mean, standard deviation, and absolute mean values of $c_{10,0}(x, y)$ for linearly cointegrated, nonlinearly cointegrated, and noncointegrated series.

Test Statistic	Linear Cointegration	Nonlinear Cointegration	Noncointegration ($H_{2,1}$)	Noncointegration ($H_{2,2}$)
$E(c_{10,0}(x, y))$	0.0619	0.0189	0.2953	0.8307
$\text{std}(c_{10,0}(x, y))$	0.117	0.061	0.12	0.07
$E(\|c_{10,0}(x, y)\|)$	0.0718	0.0434	0.2953	0.8307

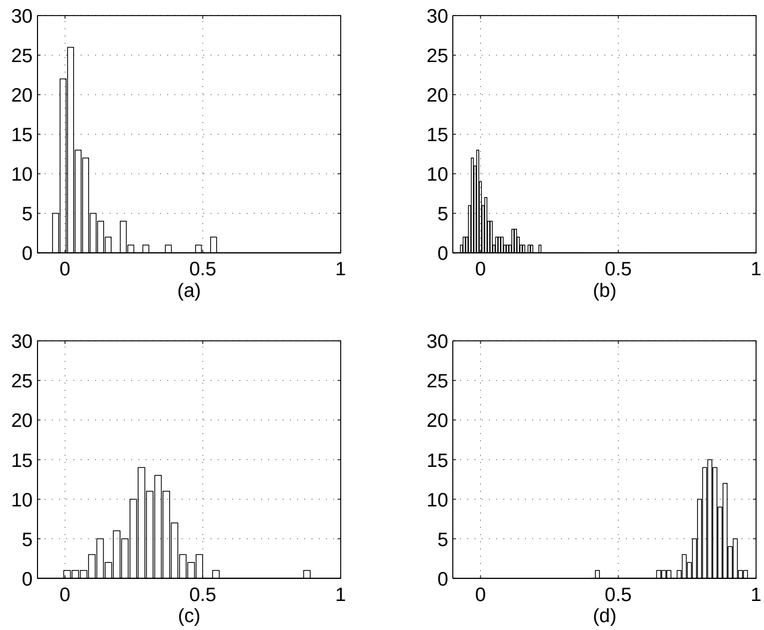

Figure 5
Histogram plots of $c_{10,0}(x, y)$, where (x, y) represents linearly (a), nonlinearly (b), and noncointegrated (c), (d) pairs of series. Plots (c) and (d) correspond to noncointegrated series from the alternative hypotheses $H_{2,1}$ and $H_{2,2}$, respectively. The nonlinearly cointegrated series were obtained as in Figure 2.

time series could be explained by the smaller values of their mutual information, as compared with those for pairs of independent random walks (this may be a consequence of the spurious regressions problem[5]).

As a simple empirical illustration, we used our cointegration-testing device based on the mutual information to investigate the joint properties of certain stock prices and cross-country exchange-rate series. The interest in these types of series is clear from an economic point of view, following our previous discussion.

First, we considered two pairs of cross-country exchange-rate series (Figure 7), namely, those of the U.S. dollar (EXRPD), the Deutsch mark (EXRPM), and the Japanese yen (EXRPY), against the Spanish peseta. A sample size of $N = 1{,}000$ daily observations was taken for these series, starting from the date of January 1, 1987. We also considered a pair of series of stock prices (STR1, STR2) from the Japanese food company Ajinomoto (Figure 6). The data records were chosen on the basis of the sample period for which these data were available, and on the absence of gross disparities appearing in plots of the series.

Under the absolute version of the PPP, the exchange rate is cointegrated (with cointegration coefficient 1) with the ratio of the relative prices of the same goods in two different countries. If this also happens pairwise with another country, those exchange rates should as well be cointegrated with their relative prices. However, following Meese and Rose (1991), most cointegration relationships could be nonlinear. In our analysis, we found that for all series and for the given sample sizes, with the critical values given by MacKinnon (1990), the hypothesis of a unit root could not be rejected by an augmented Dickey-Fuller (ADF) test at the 10%

[5]We agree with an anonymous referee that this problem is hard to discard, and thus this topic deserves further attention in a nonlinear context.

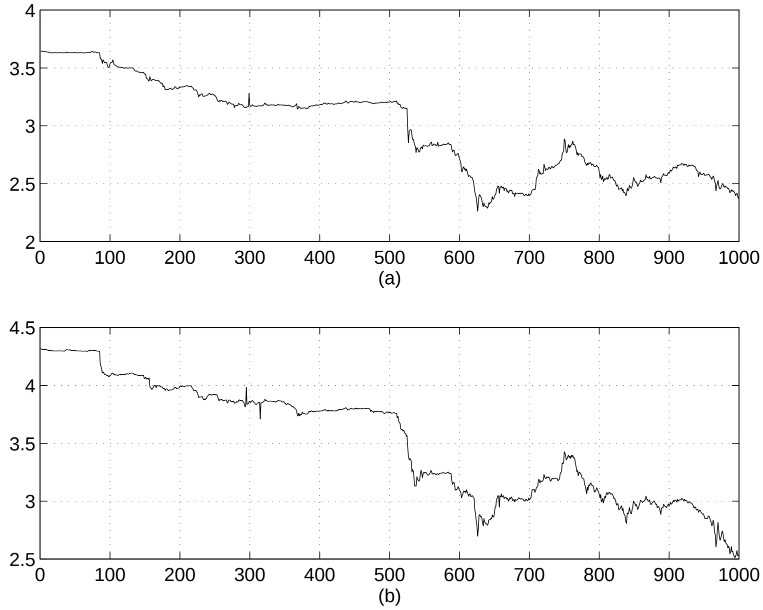

Figure 6
Two stock-price series from the Japanese food company, Ajinomoto.

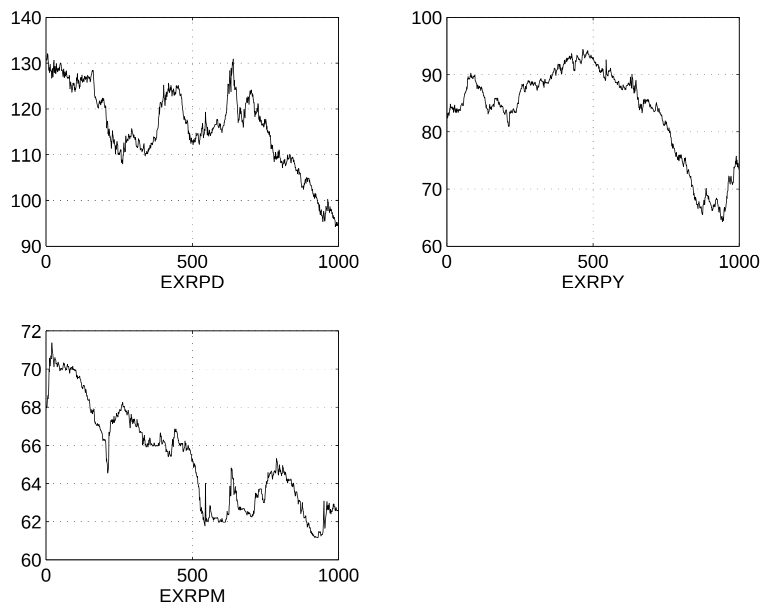

Figure 7
Daily foreign exchange-rate series from January 1987: EXRPD (peseta/U.S. dollar), EXRPY (peseta/100 yen), and EXRPM (peseta/Deutsch mark).

significance level. Performing this same test on the cointegration residuals led to rejection of the null hypothesis of noncointegration in all cases except for the pair of stock prices (STR1, STR2). On the other hand, the values of the test statistic $c_{10,0}(x, y)$ of Equation 56, shown in Table 2, suggest evidence of cointegration in both (EXRPD, EXRPY) and (STR1, STR2), when using a one-standard-deviation empirical confidence interval. These findings point to a certain robustness of our test to departures from linearity of the cointegration relationship.

Table 2
Values taken by the cointegration test statistic $c_{10,0}(x, y)$ on two pairs of foreign exchange-rate series and a pair of stock-return series.

Series	EXRPY/EXRPD	EXRPM/EXRPD	STR1/STR2
$c_{10,0}(x, y)$	−0.0113	0.257	0.1169

6 Conclusions

Long memory and cointegration are two important features of many economic variables. Usual testing techniques for those features are not robust to the presence of eventual nonlinearities, either in the univariate data-generating process or in the cointegrating relationship. In this paper, we have proposed an alternative methodology for measuring dependence and memory, based on the concepts of entropy and mutual information. We showed by some algebraic examples that this approach works well in a linear context, when the variables are $I(1)$ and $I(0)$, and also with certain fractional differencing models. We suggest a testing device for the null hypothesis of cointegration (either linear or not), and show, by a small Monte Carlo experiment and by an empirical application, that this test statistic could be promising in detecting nonlinear cointegrating relationships. In particular, we found that even though the peseta/dollar and the peseta/yen exchange rates are not (linearly) cointegrated, they seem to be nonlinearly cointegrated. This means that if the absolute PPP hypothesis holds, then the relative prices of those countries move together in a nonlinear way, and they might have eventual departures from the linear equilibrium along the business cycles.

References

Akella, S., and A. Patel (1991). "Are international real rates of interest cointegrated? Evidence from the recent floating exchange rate period." Working Paper, University of Missouri.

Aparicio, F. M. (1995). "Nonlinear modeling 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. M., and A. Escribano (1999). "A model-free characterization of cointegration and a linearity test." Working Paper, Dept. of Statistics, Universidad Carlos III de Madrid.

Aparicio, F. M., and C. W. J. Granger (March 1995a). "Information-theoretic schemes for linearity testing under long-range dependence and under cointegration." Working Paper, Department of Economics, University of California at San Diego.

Aparicio, F. M., and C. W. J. Granger (March 1995b). "Nonlinear cointegration and some new tests for comovements." Working Paper, Department of Economics, University of California at San Diego.

Baillie, R., and T. Bollerslev (1989). "Common stochastic trends in a system of exchange rates." *Journal of Finance*, 44: 167–182.

Baillie, R. T., and J. Myers (1991). "Bivariate GARCH estimation of the optimal commodity futures hedge." *Journal of Applied Econometrics*, 6: 109–124.

Balke, N. S., and T. B. Fomby (1997). "Threshold cointegration." *International Economic Review*, 38(3): 627–645.

Barnhart, S. W., and A. C. Szakmary (1991). "Testing the unbiased forward-rate hypothesis: Evidence on unit roots, cointegration, and stochastic coefficients." *Journal of Financial and Quantitative Analysis*, 26: 245–267.

Berg, L., and J. Lyhagen (1998). "Short- and long-run dependence in Swedish stock returns." *Applied Financial Economics*, 8: 435–443.

Bossaerts, P. (1988). "Common nonstationary components of asset returns." *Journal of Economic Dynamics and Control*, 12: 347–364.

Breiman, L., and J. H. Friedman (1985). "Estimating optimal transformations for multiple regressions and correlation (with discussion)." *Journal of the American Statistical Association*, 80(391): 580–619.

Breiman, L., W. S. Meisel, and E. Purcell (1977). "Variable kernel estimates of multivariate densities and their calibration." *Technometrics*, 19: 135–144.

Campbell, J. Y., and R. J. Shiller (1987). "Cointegration and tests of present value models." *Journal of Political Economy*, 95: 1062–1088.

Cerchi, M., and A. Havenner (1988). "Cointegration and stock prices: The random walk on Wall Street revisited." *Journal of Economic Dynamics and Control*, 12: 333–346.

Corbae, D., and S. Ouliaris (1988). "Cointegration and tests of purchasing power parity." *Review of Economics and Statistics*, 70: 508–511.

Davies, R. B., and D. S. Harte (1987). "Tests for Hurst effect." *Biometrika*, 74(1): 95–101.

DeJong, D. (1992). "Cointegration and trend-stationarity in macroeconomic time series: Evidence from the likelihood function." *Journal of Econometrics*, 52: 347–370.

DeJong, D., and C. Whiteman (1991). "The temporal stability of dividends and stock prices: Evidence from the likelihood function." *American Economic Review*, 81: 600–617.

Enders, W. (1988). "ARIMA and cointegration tests of purchasing-power parity." *Review of Econometrics and Statistics*, 70: 504–508.

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

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

Escribano, A., and S. Mira (1997). "Nonlinear error-correction models." Working Paper 97-26, Department of Statistics, Universidad Carlos III de Madrid.

Escribano, A., and G. A. Pfann (1998). "Nonlinear error correction, asymmetric adjustment, and cointegration." *Economic Modeling*, 15: 197–216.

Geweke, J., and S. Porter-Hudak (1983). "The estimation and application of long-memory time-series models." *Journal of Time-Series Analysis*, 4(4): 221–238.

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. (1983). "Cointegrated variables and error-correcting models." Discussion Paper 83-13, Department of Economics, University of California at San Diego.

Granger, C. W. J. (1995). "Modeling nonlinear relationships between extended-memory variables." *Econometrica*, 63(2): 265–279.

Granger, C. W. J., and J. J. Hallman (1991). "Long-memory series with attractors." *Oxford Bulletin of Economics and Statistics*, 53(1): 11–26.

Granger, C. W. J., and T. Teräsvirta (1993). *Modeling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Greene, M. T., and B. D. Fielitz (1977). "Long-term dependence in common stock returns." *Journal of Financial Economics*, 4: 339–349.

Hallman, J. J. (1990). "Nonlinear integrated series, cointegration, and an application." PhD Thesis, Department of Economics, University of California at San Diego, La Jolla, California, USA.

Hosking, J. R. M. (1981). "Fractional differencing." *Biometrika*, 68: 165–176.

Hurst, H. (1951). "Long-term storage capacity of reservoirs." *Transactions of the American Society of Civil Engineers*, 116: 778–808.

Kay, S. M. (1988). *Modern Spectral Estimation*. New York: Prentice-Hall.

Koop, G. (1994). "An objective Bayesian analysis of common stochastic trends in international stock prices and exchange rates." *Journal of Empirical Finance*, 1(3–4): 343–364.

Lo, A. W. (1991). "Long-term memory in stock-market prices." *Econometrica*, 59(4): 1279–1313.

MacKinnon, J. G. (January 1990). "Critical values for cointegration." Working Paper, Department of Economics, University of California at San Diego.

Mandelbrot, B. (1972). "Statistical methodology for nonperiodic cycles: From the covariance to R/S analysis." *Annals of Economic and Social Measurements*, 1(Dec.): 257–288.

Meese, R. A., and A. K. Rose (1991). “An empirical assessment of nonlinearities in models of exchange-rate determination.” *Review of Economic Studies*, 58: 603–619.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill.

Robinson, P. M. (1991). “Consistent nonparametric entropy-based testing.” *Review of Economic Studies*, 58: 437–453.

Robinson, P. M. (1995). “Log-periodogram regression of time series with long-range dependence.” *Annals of Statistics*, 23(3): 1048–1072.

Rosenblatt, M. (1974). *Random Processes*. New York: Springer-Verlag.

Schotman, P., and H. K. Van Dijk (1991). “A Bayesian analysis of the unit root in real exchange rates.” *Journal of Econometrics*, 49: 195–238.

Sims, C. A. (1988). “Bayesian skepticism on unit-root econometrics.” *Journal of Economics, Dynamics, and Control*, 12: 436–474.

Taylor, M. P., and I. Tonks (1989). “The internationalisation of stock markets and the abolition of U.K. exchange control.” *Review of Economics and Statistics*, 71: 332–336.

Wooldridge, J. M. (1986). “Asymptotic properties of econometric estimators.” PhD Thesis, Department of Economics, University of Califnoria at San Diego, La Jolla, California, USA.

Editorial Policy

The *SNDE* is formed in recognition that advances in statistics and dynamical systems theory may increase our understanding of economic and financial markets. The journal will seek both theoretical and applied papers that characterize and motivate nonlinear phenomena. Researchers will be encouraged to assist replication of empirical results by providing copies of data and programs online. Algorithms and rapid communications will also be published.

ISSN 1081-1826