Nonlinear Persistence and Copersistence

Christian Gourieroux*

Joann Jasiak†

This version: November 23, 1999

*CREST and CEPREMAP , *e-mail*: gouriero@ensae.fr.

†York University, *e-mail*: jasiakj@yorku.ca

The second author gratefully acknowledges financial support of the Natural Sciences and Engineering Research Council of Canada

Acknowledgments: We thank Clive Granger, as well as Marine Carrasco and Benedikt Potscher for helpful comments.

Abstract

Nonlinear Persistence

and Copersistence

In a nonlinear framework, temporal dependence of time series is sensitive to transformations. The aim of this paper is to examine in detail the relationships between various forms of persistence and nonlinear transformations of stationary and nonstationary processes. We introduce the concept of persistence space and use it to define the degrees of persistence of univariate or multivariate processes. For illustration, we examine and compare the persistence structure of a fractionally integrated process and a beta mixture of AR(1) processes. The study of multivariate processes is focused on nonlinear comovements between the components, called the copersistence directions, or cointegration directions in the nonstationary case. We find that, in general, there is a multiplicity of such directions, causing an identification problem in the analysis of nonlinear cointegration.

Keywords: Nonlinear Autocorrelogram, Canonical Analysis, Persistence, Chaos, Unit Root, Cointegration.

JEL : C14

Résumé

Dans un contexte non linéaire, l'importance de la dépendence temporelle peut dépendre de la transformation considérée de la série. Le but de ce papier est de décrire soigneusement les divers degrés de persistence et les transformations qui leur sont associées. Pour des séries multivariées, l'analyse de persistence peut être utilisée pour mettre en évidence des comouvements non linéaires entre les composantes, c'est à dire le phénomène de copersistence (ou de cointégration non linéaire dans le cas non stationnaire). Généralement il y a une multiplicité de directions de copersistence, ce qui induit un problème d'identification.

Mots clés: Autocorrélogramme non linéaire, analyse canonique, persistence, chaos, racine unitaire, cointégration.

JEL : C14

1 Introduction

Although structural econometrics does not ignore nonlinear dynamics in relations between macroeconomic or financial variables, in practice time series are often examined using standard instruments. These instruments, such as the autocorrelation and partial autocorrelation functions, spectral densities, unit root tests, techniques of finding cointegrating vectors, E.C.M. representations [Granger (1986), Engle, Granger (1987), Johansen (1988)], or long memory patterns [Hosking (1981)] have been however designed for linear dynamics. Despite this, researchers commonly apply the traditional methods to various nonlinear transformations of time series. A typical example is the autocorrelogram, which occasionally becomes redefined as the autocorrelation function of a nonlinear transform of the series [Lepnik (1958), Granger, Newbold (1976), Granger, Hallman (1991), Ding, Granger, Engle (1993)]. Widely known empirical evidence shows however that the pattern of autocorrelations may strongly depend on the applied nonlinear transformation. For instance, the autocorrelogram of financial returns often shows an absence of (linear) dependence whereas the autocorrelogram of squares or absolute values of these returns features long memory [Engle (1982), Ding, Granger, Engle (1987)]. Similarly, there exist nonlinear transformations yielding stationary transformed processes whereas some others induce various nonstationary features [1]. This leads to spurious results when, for instance, a standard Dickey-Fuller test is applied to the transformed series [Granger, Hallman (1989), Corradi (1995)].

Rather than investigating the properties of linear techniques applied to specific nonlinear dynamics, we prefer in this paper to come back on the notion of persistence in a nonlinear framework. We propose persistence measures which show how temporal dependence evolves with the prediction horizon and depends on the nonlinear transformations of interest. Our approach is developped for both stationary and nonstationary processes, and concerns both short and long memory phenomena. The nonstationarity case covers nonlinear unit roots processes [see, Park, Phillips (1998a, 1998b)] and nonlinear cointegration analysis. Recently, the nonlinear comovements of nonstationary processes have received a considerable attention [see, Bierens (1999), Chang, Park, Phillips (1999), Karlsen, Myklebust, Tjosheim (1999), De Jong (1999)]. Researchers often draw parallels between the linear and nonlinear cointegration, and expect to find a finite number of cointegrating relations. We caution against it, since as our approach indicates, there exist in general a multiplicity of cointegrating relations which requires a judicious choice of identifying constraints.

The first five sections cover strongly stationary processes. We introduce in section 2 the transformed autocorrelogram, which measures the dependence between a transformed future value

[1] (i)If (ϵ_t) is a gaussian white noise and $Z_t = \sum_{\tau=1}^{t} \epsilon_\tau$ the associated random walk, the process defined by $X_t = sgn\,(\epsilon_t) Z_t^2$ is such that: $sgn\,(X_t) = sgn\,(\epsilon_t)$ is stationary whereas $|X_t| = Z_t^2$ is nonstationary.

(ii) A convex transformation of a zero mean random walk (Z_t) will include a deterministic trend [Corradi(1995)].

$g(X_{t+h})$ and the information contained in the current observation X_t. We compare this measure with the autocorrelogram of $g(X_t)$ usually considered in the literature. The transformed autocorrelogram is used in section 3 to define the persistence decomposition into degrees of persistence and the associated persistence spaces. In section 4 we illustrate the approach by showing persistence decompositions of various processes including long memory processes, processes generated from gaussian time series or discretized diffusions. The definition of copersistence for multivariate time series is introduced in section 5. We emphasize there the importance of the concept of underlying universe. The analysis is extended in section 6 to processes featuring nonstationarities. In particular, we focus on the persistence by trajectory of nonlinear unit root processes, and discuss nonlinear cointegration. Section 7 covers statistical inference and presents a simulation study. Section 8 concludes the paper.

2 Transformed Autocorrelograms

2.1 Definition of the Autocorrelogram

A common approach to investigating the effect of a nonlinear scalar transformation g on a strongly stationary time series is to examine the autocorrelation function:

$$r_h(g) = \text{Corr}[g(X_{t+h}), g(X_t)], \quad h \geq 0, \tag{2.1}$$

[see, e.g Ding, Engle, Granger (1993), Ding, Granger (1996), Granger, Terasvirta (1997), He, Terasvirta (1997), Gourieroux, Jasiak (1998)]. This method may not be precise enough to define appropriately the degree of persistence. The dependence between $g(X_{t+h})$ and the information contained in X_t appears to be better measured by:

$$\rho_h(g) = \max_{g_2} \text{Corr}[g(X_{t+h}), g_2(X_t)], \quad h \geq 0. \tag{2.2}$$

Note that this autocorrelation function is similar to the traditional one in that it represents the dependence between a pair of variables, i.e. X_t and X_{t+h} [2]. We can easily verify that it admits the optimum for $g_2(X_t) = E[g(X_{t+h})|X_t]$ [see Appendix 1], and that:

$$\rho_h(g) = \text{Corr}[g(X_{t+h}), E(g(X_{t+h})|X_t)] = \sqrt{\frac{\text{Var}\, E(g(X_{t+h})|X_t)}{\text{Var}\, g(X_t)}}. \tag{2.3}$$

From this formula we deduce that:

[2] An alternative is to measure instead the dependence between $g(X_{t+h})$ and the past of the process, i.e. $\underline{X_t} = [X_t, X_{t-1}, X_{t-2}, ...]$. Then the transformed autocorrelogram would be: $\rho^*_h(g) = \max_{g_2} \text{Corr}[g(X_{t+h}), g_2(\underline{X_t})]$. Even if the autocorrelogram $\rho^*_h(g)$ is more informative than the autocorrrelogram $\rho_h(g)$, it is also more difficult to implement in practice. Of course the two types of autocorrelograms coincide for a Markov process of order one.

$$1 \Leftrightarrow \rho_h^2(g) = \frac{E \operatorname{Var}[g(X_{t+h})|X_t]}{\operatorname{Var} g(X_t)},$$

measures the accuracy of the nonlinear prediction of the transformed series at horizon h [3].

By construction, the transformed autocorrelogram ($\rho_h(g)$, h varying) [T-autocorrelogram henceforth] takes nonnegative values, and we have $\rho_h(g) \geq |r_h(g)|$, $\forall\ h, g$ [4].

In practice we can estimate the T-autocorrelogram in the following way. Let us denote by K a kernel function, and by $\hat{g}_{2,h}$ the associated regressogram of $g(X_{t+h})$ on X_t:

$$\hat{g}_{2,h}(x) = \frac{\sum_{\tau=1}^{T-h} K\left[(X_\tau \Leftrightarrow x)/\eta\right] g(X_{\tau+h})}{\sum_{\tau=1}^{T-h} K\left[(X_\tau \Leftrightarrow x)/\eta\right]}, \tag{2.4}$$

where η denotes the bandwidth. The T-autocorrelogram is approximated by the empirical correlation between $g(X_{t+h})$ and $\hat{g}_{2,h}(X_t)$:

$$\hat{\rho}_h(g) = \operatorname{Corr}_e[g(X_{t+h}), \hat{g}_{2,h}(X_t)].$$

2.2 Nonlinear Canonical Analysis

Let us consider a stationary process with a continuous distribution, and denote by f_h the joint density of (X_t, X_{t+h}), and by f the marginal density of X_t. Under weak conditions [5], the joint density can be decomposed as [Dunford, Schwartz (1968), Lancaster(1968)]:

$$f_h(x_t, x_{t+h}) = f(x_t) f(x_{t+h}) \{1 + \sum_{j=1}^{\infty} \lambda_{j,h} a_{j,h}(x_{t+h}) b_{j,h}(x_t)\}, \tag{2.5}$$

where the canonical correlations $\lambda_{j,h}$, j varying, are decreasing $\lambda_{1,h} \geq \lambda_{2,h} \geq \ldots \geq 0$, $\forall h$, the canonical directions satisfy the orthogonality conditions:

$E[a_{j,h}(X_t) a_{k,h}(X_t)] = 0, \ \forall k \neq j, \ \forall h,$

$E[b_{j,h}(X_t) b_{k,h}(X_t)] = 0, \ \forall k \neq j, \ \forall h,$

$E a_{j,h}(X_t) = E b_{j,h}(X_t) = 0, \ \forall j, \ h,$

and the normalization conditions:

$$\operatorname{Var} a_{j,h}(X_t) = \operatorname{Var} b_{j,h}(X_t) = 1, \ \forall j, \ h.$$

Therefore in nonlinear framework, the canonical analysis involves the joint density function instead of linear regression coefficients, which are used in the standard linear setup [see e.g. Tiao, Tsay (1989), Johansen (1988)].

[3] When $E[g(X_{t+h})|X_t]$ is constant, the correlation $\rho_h(g)$ is conventionally set equal to zero.

[4] If we defined $\rho_h(g)$ for all h, the resulting autocorrelogram would not be an even function, i.e. we would have $\rho_h(g) \neq \rho_{-h}(g)$, violating the condition of the traditional autocorrelation function.

[5] For instance, if $\int \int [f_h^2(x_t, x_{t+h})] \, / \, [f(x_t) f(x_{t+h})] dx_t dx_{t+h} < +\infty, \ \forall h.$

The transformed autocorrelogram can be easily written in terms of the components of the canonical decomposition.

Property 2.1: We have:

$$\rho_h^2(g) = \frac{\sum_{j=1}^{\infty} \lambda_{j,h}^2 < g, a_{j,h} >^2}{\sum_{j=1}^{\infty} < g, a_{j,h} >^2},$$

where $< g, a_{j,h} >= E[g(X_t)a_{j,h}(X_t)] = \text{Cov}[g(X_t), a_{j,h}(X_t)]$.

Proof: For any nonlinear function g, we get:

$$E[g(X_{t+h})|X_t] = Eg(X_{t+h}) + \sum_{j=1}^{\infty} \lambda_{j,h} E[g(X_{t+h})a_{j,h}(X_{t+h})]b_{j,h}(X_t).$$

Since $b_{j,h},\ h$ varying, and $a_{j,h},\ j$ varying, form an orthonormal basis, we directly deduce that:

$$\rho_h(g) = \sqrt{\frac{\text{Var}E[g(X_{t+h})|X_t]}{\text{Var}[g(X_{t+h})]}} = \left[\frac{\sum_{j=1}^{\infty} \lambda_{j,h}^2 < g, a_{j,h} >^2}{\sum_{j=1}^{\infty} < g, a_{j,h} >^2}\right]^{1/2}.$$

Q.E.D.

3 Persistence Decomposition

In this section we define the degrees of persistence and the associated persistence spaces. The results are derived for a strictly stationary univariate or multivariate process (X_t).

3.1 Persistence Spaces

Let $\alpha = (\alpha_h,\ h \geq 0)$ denote a positively valued sequence converging to zero at infinity. Henceforth the sequence α will be a priori called the degree of persistence and will be used to measure how the temporal dependence of the transformed series evolves with the prediction horizon [6]. For example, we may consider: $\alpha_h = r^h$ or $\alpha_h = h^{2d-1}$ where $r,\ r \geq 0$, and $d,\ d \leq 1/2$, are predetermined to describe geometric or hyperbolic decay patterns.

Definition 3.1: The persistence space of order α is defined by [7]:

$$E_\alpha = \{g:\ \rho_h(g) = 0(\alpha_h)\}, \tag{3.1}$$

where the symbol 0 means that there exists a constant c such that $|\rho_h(g)| \leq c\,\alpha_h$.

Property 3.1: E_α is a vector space.

Proof: Let us consider a linear combination of two elements g and g^* in E_α. We have:

[6] More precisely a degree of persistence is a class of equivalence of scalar sequences with the same asymptotic behaviour for large h; α and β provide the same degree of persistence, if $\alpha_h \sim \beta_h$ for large h.

[7] Since $\text{Corr}(a, g_2(X_{t-h})) = 0,\ \forall a \in R$, this space contains at least the constants.

$$\begin{aligned}
\text{Corr}\,[(ag+a^*g^*)(X_{t+h}),g_2(X_t)] &= \frac{\text{Cov}\,[(ag+a^*g^*)(X_{t+h}),g_2(X_t)]}{\sqrt{\text{Var}\,(ag+a^*g^*)(X_t)}\sqrt{\text{Var}\,g_2(X_t)}} \\
&= \frac{a\text{Cov}\,[g(X_{t+h}),g_2(X_t)]}{\sqrt{\text{Var}\,(ag+a^*g^*)(X_t)}\sqrt{\text{Var}\,g_2(X_t)}} \\
&\quad + \frac{a^*\text{Cov}\,[g^*(X_{t+h}),g_2(X_t)]}{\sqrt{\text{Var}\,(ag+a^*g^*)(X_t)}\sqrt{\text{Var}\,g_2(X_t)}} \\
&= a\frac{\sqrt{\text{Var}\,g}}{\sqrt{\text{Var}\,(ag+a^*g^*)}}\ \text{Corr}[g(X_{t+h}),g_2(X_t)] \\
&\quad + a^*\frac{\sqrt{\text{Var}\,g^*}}{\sqrt{\text{Var}\,(ag+a^*g^*)}}\ \text{Corr}\,[g^*(X_{t+h}),g_2(X_t)].
\end{aligned}$$

This implies that:

$$\rho_h(ag+a^*g^*) \le a\frac{\sqrt{\text{Var}\,g}}{\sqrt{\text{Var}\,(ag+a^*g^*)}}\rho_h(g) + a^*\frac{\sqrt{\text{Var}\,g^*}}{\sqrt{\text{Var}\,(ag+a^*g^*)}}\rho_h(g^*) = 0(\alpha_h).$$

Q.E.D.

Next, we set $\alpha_h = 0$.

Corollary: $E_0 = \{g: \ \text{Corr}\,[g(X_{t+h}),g_2(X_t)] = 0, \ \forall g_2, h\}$ is a vector space.

It forms the space of nonlinear independence directions.

From (2.3), $g \in E_0$ if and only if $E[g(X_{t+h})|X_t] = Eg(X_{t+h}), \ \forall h \geq 1$. In particular for a Markov process, $g \in E_0$ if and only if the process $g(X_t) \Leftrightarrow Eg(X_t)$ is a martingale difference sequence with respect to the filtration $\underline{X_t}$ [8].

At this point, it is interesting to note that the set of functions $\mathcal{E}_0 = \{g : r_h(g) = 0, \quad \forall h\}$ does not define a vector space. Indeed, even if $[g(X_t)]$ and $[g^*(X_t)]$ are two weak white noises, $[g(X_t) + g^*(X_t^*)]$ may feature temporal dependence due to cross-correlations between the two processes. This difficulty does not arise with our definition of the nonlinear autocorrelogram, since the past information $(g_2(X_t), g_2$ varying) is independent of the transformation g [9].

3.2 Properties of the Persistence Spaces

Various properties of the persistence space can easily be derived.

monotonicity

If (α_h) and (β_h) are two sequences such that $\beta_h = 0(\alpha_h)$, then $E_\beta \subset E_\alpha$.

[8] If we modify the definition of 0 by considering that $\rho_h(g) = 0(\alpha_h)$ iff there exists a constant c such that $\lim_{h\to\infty}\sup|\rho_h(g)| \leq c\alpha_h$, the space E_0 becomes: $E_0 = \{g : \ \exists q$ with $E[g(X_{t+h})|X_t] = Eg(X_{t+h}) \ \forall h \geq q + 1\}$. Therefore E_0 defines the (nonlinear) moving average directions called codependence directions in the linear framework [Gourieroux, Peaucelle (1992), Vahid, Engle (1997)].

[9] The same type of remark applies to the analysis by Corradi (1995), where the markovian properties of the transformed process $g(X_t)$ are considered with respect to the filtration generated by the transformation, not by the initial process X_t itself.

invariance by one-to-one transformation

Let us consider a one-to-one transformation $Y_t = a(X_t)$, (say) of the initial process. The persistence spaces of X and Y can now be compared. Indeed we have:

$$\begin{aligned} E_\alpha^Y &= \{g : \max_{g_2} \text{ Corr}\,[g(Y_{t+h}), g_2(Y_t)] = 0(\alpha_h)\} \\ &= \{g : \max_{g_2} \text{ Corr}\,[g \cdot a(X_{t+h}), g_2(X_t)] = 0(\alpha_h)\}. \end{aligned}$$

We deduce that:

$$g \in E_\alpha^Y \Leftrightarrow g \cdot a \in E_\alpha^X.$$

3.3 Degree of Persistence

In this section we proceed to define the set of degrees of persistence. This task is not easy due to the absence of a complete ordering of the set of sequences (α_h). Indeed, there can exist sequences (α_h) and (β_h) which are not comparable, i.e. such that neither the condition $\alpha_h = 0(\beta_h)$, nor $\beta_h = 0(\alpha_h)$ are satisfied. There are two possible ways of solving this problem.

i) We can select from the set of sequences a restricted subset which can be ordered without ambiguity. For example, we could consider only either the geometric sequences:

$$\alpha_h = r^h, \text{ with } r \geq 0,$$

or the hyperbolic ones.

Let us suppose, for example, such a family $\alpha_h(\theta)$ depending continuously on a parameter $\theta \in R^+$, and such that if $\theta_1 > \theta_2$, $\alpha_h(\theta_2) = o[\alpha_h(\theta_1)]$. In this case, the persistence spaces can be introduced for various values of θ and, according to the monotonicity property, we have $E_{\alpha(\theta_1)} \subset E_{\alpha(\theta_2)}$. This allows us to define the limiting parameter values:

$$\Theta^* = \{\theta : E_{\alpha(\theta)} \neq \cup_{\theta^* < \theta} E_{\alpha(\theta^*)}\}. \tag{3.2}$$

The set of sequences $[\alpha_h(\theta), \theta \in \Theta^*]$ is the set of constrained degrees of persistence of the process. This set can be quite complex and contain an infinite number of basic sequences. It can however be defined for any process.

ii) We can alternatively impose some constraints on the dynamics in order to define the implied degrees of persistence without ambiguity.

Definition 3.2: The process (X_t) admits a decomposition of persistence iff there exists a countable set of sequences $(\alpha_{n,h})$, $n \in N$, such that:

(i) $\forall n \quad \alpha_{n,h} = o(\alpha_{n-1,h})$,

and

(ii) $\forall g$ in the supplement of E_{α_n} in $E_{\alpha_{n-1}}$, we have:

$$\rho_h(g) \sim \alpha_{n-1,h}, \quad \text{for} \quad h \quad \text{large}.$$

For such a process, the sequences $(\alpha_{n,h})$, n varying, define the succession of persistence degrees, whereas E_{α_n} are the corresponding persistence spaces. In general all processes do not admit a persistence decomposition. In section 4 we describe classes of processes for which a persistence decomposition exists.

3.4 Change of Universe

The definitions of the transformed autocorrelogram and the associated persistence decomposition can be extended to a universe Z_t (say), possibly different from X_t. More precisely, if the process $(X_t,\, Z_t)$ is jointly stationary, we can introduce the transformed autocorrelogram of X with respect to the universe :

$$\rho_h(g;\, Z) = \max_{g_2}\ \text{Corr}\,[g(X_{t+h}), g_2(Z_t)]. \tag{3.3}$$

The concept of universe is of particular importance in the multivariate framework. Let us consider a bivariate stationary process $X_t = (X_{1,t}, X_{2,t})'$, and analyze the persistence properties of the first component. We can naturally consider two types of transformed autocorrelograms.

· The *marginal* transformed autocorrelogram corresponds to the universe $Z = X_1$ and is defined by:

$$\rho_h^{1,m}(g) = \max_{g_2}\ \text{Corr}\,[g(X_{1,t+h}), g_2(X_{1,t})]. \tag{3.4}$$

· The *global* transformed autocorrelogram corresponds to the universe $Z = X$ and is defined by:

$$\rho_h^{1}(g) = \max_{g_2}\ \text{Corr}\,[g(X_{1,t+h}), g_2(X_t)]. \tag{3.5}$$

We denote by $E_\alpha^{1,m}$ and E_α^1 the associated persistence spaces. The property below relates the marginal and global persistence decompositions.

Property 3.3: We have:

(i) $\rho_h^{1,m}(g) \leq \rho_h^1(g), \ \forall\, h, g.$

(ii) $E_\alpha^{1,m} \supset E_\alpha^1, \ \forall \alpha.$

4 Examples of Persistence Decomposition

4.1 Processes with Stable Canonical Decompositions

Let us consider a stationary process, with canonical variates independent of the horizon h:

$$f_h(x_t, x_{t+h}) = f(x_t)f(x_{t+h})\{1 + \sum_{j=1}^{\infty} \mu_{j,h} a_j(x_{t+h}) b_j(x_t)\}, \tag{4.1}$$

where $|\mu_{1,h}| \geq |\mu_{2,h}| .... \geq 0$ and the functions $a_j, b_j,\ j$ varying, satisfy the orthogonality and normalization conditions described in subsection 2.2.

Property 4.1: Any process with a stable canonical decomposition (4.1), where $\mu_{j,h} = o(\mu_{j-1,h})$, $\forall j$, admits a persistence decomposition. The degrees of persistence are the sequences $(|\mu_{j,h}|)$, j varying, whereas the associated persistence spaces are:

$$E_{|\mu_j|} = \{g : < g, a_k >= 0,\ \forall\, k \leq j \Leftrightarrow 1\}.$$

Proof: This is a direct consequence of the property 2.1.

Q.E.D.

Therefore the degrees of persistence coincide with the sequences of canonical correlations of various orders. These correlations, jointly displayed, form the nonlinear autocorrelogram [Gourieroux, Jasiak (1998)].

Example 1: Gaussian processes

Let us consider a gaussian process with zero mean, unitary variance and autocorrelation function (ρ_h). The canonical decomposition is given by [Cramer (1963), Barrett, Lampard (1955)]:

$$f_h(x_t, x_{t+h}) = \phi(x_t)\phi(x_{t+h})\{1 + \sum_{j=1}^{\infty} (\rho_h)^j H_j(x_{t+h}) H_j(x_t)\},$$

where ϕ denotes the p.d.f. of the standard normal distribution and H_j is the Hermite polynomial of order j. The persistence degrees are: $(\alpha_{j,h}) = (|\rho_h|^j)$, j varying, and the persistence spaces E_{α_j} are generated by the Hermite polynomials of degrees larger or equal to j. In particular, we can consider an AR(1) gaussian process with $\rho_h = \rho^h$, and a fractional gaussian process $(1 \Leftrightarrow L)^d X_t = \epsilon_t,\ d < 1/2$, with $\rho_h \sim A h^{2d-1}$. Since $(\rho^h)^j = (\rho^j)^h$ in the autoregressive case and

$$(\rho^h)^j \sim A h^{(2d-1)j} = A h^{2(dj + \frac{1-j}{2}) - 1},$$

we directly deduce the patterns of the persistence degrees.

Table 4.1: Persistence degrees for gaussian processes

Process	pattern	parametrization
AR(1) gaussian	geometric: $\alpha_{j,h} = r_j^h$	$r_j = \lvert\rho\rvert^j$
I(d) gaussian	hyperbolic: $\alpha_{j,h} = A_j h^{2d_j - 1}$	$A_j = A,\ d_j = dj + \frac{1-j}{2}$

In the hyperbolic case the fractional degrees d_j decrease arithmetically at rate $d = 0.5$. They become equal in the limiting case $d = 0.5$ of nonstationarity.

Example 2: Mixture of gaussian processes

This example illustrates non gaussian long memory processes derived from a mixture of AR(1) gaussian processes. More precisely, let us consider an AR(1) gaussian process, with a stochastic autoregressive parameter following a distribution π. The joint distribution of (X_t, X_{t+h}) becomes:

$$\begin{aligned} f_h(x_t, x_{t+h}) &= \int \phi(x_t)\phi(x_{t+h})\{1 + \sum_{j=1}^{\infty} \rho^{jh} H_j(x_{t+h}) H_j(x_t)\} d\pi(\rho) \\ &= \phi(x_t)\phi(x_{t+h})\{1 + \sum_{j=1}^{\infty} E_\pi(\rho^{jh}) H_j(x_{t+h}) H_j(x_t)\}. \end{aligned}$$

In general $E_\pi(\rho^{jh}) \neq (E_\pi \rho^h)^j$. This implies that this process is not gaussian, even if it is marginally gaussian and admits the same canonical directions as a gaussian process.

We can easily find the long memory persistence degrees when the heterogeneity distribution is a beta distribution $B(\delta, 1 \Leftrightarrow \delta),\ 0 < \delta < 1$ [Granger, Joyeux (1980)]:

$$\pi(\rho) = \frac{\rho^{\delta-1}(1 \Leftrightarrow \rho)^{-\delta}}{,\,(\delta),\ (1 \Leftrightarrow \delta)} 1_{[0,1]}(\rho).$$

In this case $E_\pi \rho^k = \frac{\Gamma(\delta+k)}{\Gamma(\delta)} \frac{1}{\Gamma(1+k)}$. We deduce:

$$\alpha_{j,h} = E_\pi \rho^{jh} = \frac{,\,(\delta + jh)}{,\,(\delta)} \frac{1}{,\,(1 + jh)} \sim \frac{1}{,\,(\delta)} \frac{1}{j^{\delta-1}} \frac{1}{h^{\delta-1}},$$

and the patterns of the persistence degrees.

Table 4.2: Persistence degrees of a mixture of gaussian processes

Process	pattern	parametrization
beta mixture	hyperbolic $\alpha_{j,h} = A_j h^{2d_j - 1}$	$A_j = \frac{1}{\Gamma(\delta) j^{\delta-1}},\ d_j = \delta\,/\,2$

It is interesting to note the striking difference between the nonlinear dynamics of this beta mixture and the $I(d)$ gaussian process, despite their common feature of long memory in all directions. We observe that the fractional order $d_j = \delta\,/\,2$ is independent of j in the mixture case.

4.2 Deterministic Autoregression

It is difficult to discuss nonlinear dynamics without mentioning the chaos. Let us consider a deterministic stationary autoregression:

$$X_{t+1} = c(X_t), \tag{4.2}$$

where the function c is not constant.

Property 4.2: The stationary process (X_t) is a deterministic autoregression if and only if $\rho_h(g) = 1$, $\forall h$, $\forall g$ not constant.

Proof:

Necessary condition:

Without loss of generality we can choose $h = 1$. We get:

$$\begin{aligned} \rho_1(g) &= \max_{g_2} \operatorname{Corr}[g(X_{t+1}), g_2(X_t)] \\ &= \max_{g_2} \operatorname{Corr}[g(c(X_t)), g_2(X_t)] \\ &\geq \operatorname{Corr}[g(c(X_t)), g(c(X_t))] = 1. \end{aligned}$$

Sufficient condition:

If we consider the identity function g, there exists g_2^* such that: $\operatorname{Corr}[X_{t+1}, g_2^*(X_t)] = 1$. We deduce that $\exists\, a, b: \; X_{t+1} = a g_2^*(X_t) + b$, a.s.

Q.E.D.

The above property gives a characterization of the deterministic autoregressions in terms of persistence decomposition. It is interesting to note that this characterization follows from the appropriate definition of the transformed autocorrelogram. Indeed, if we study the well-known quadratic map: $X_{t+1} = 4X_t(1 \Leftrightarrow X_t)$, [10] [Tong (1990), section 3.3.2] and the identity transformation $g = Id$, we get:

$$\rho_h(g) = 1, \;\; \forall h,$$

whereas $r_h(g) = \operatorname{Corr}(X_{t+h}, X_t) = 0, \; \forall h$. The quadratic map represents a weak white noise process featuring strong nonlinear dependence.

Finally note that stationary deterministic autoregressions feature a "unit root" property (since $\lim_{h\to\infty} \rho_h(g) = 1$). Therefore the relationship between unit roots and nonstationarity needs to be precisely characterized in a nonlinear framework.

[10] The marginal distribution is uniform on the interval $[0, 1]$.

4.3 Discretized Unidimensional Diffusion Process

Let us now consider a stationary unidimensional diffusion process defined by:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \tag{4.3}$$

where (W_t) is a brownian motion and μ and σ denote the drift and volatility respectively. Under compactness conditions [11], this process admits the canonical decomposition:

$$f_h(x_t, x_{t+h}) = f(x_t)f(x_{t+h})\{1 + \sum_{j=1}^{\infty}(\mu_j)^h a_j(x_t)a_j(x_{t+h})\}, \tag{4.4}$$

where the canonical elements are derived from the spectral analysis of the infinitesimal generator: $A = \mu\frac{d}{dx} + \frac{1}{2}\sigma^2\frac{d^2}{dx^2}$. More precisely, they satisfy:

$$\mu(x)\frac{da_j(x)}{dx} + \frac{1}{2}\sigma^2(x)\frac{d^2a_j(x)}{dx^2} = \log\mu_j\, a_j(x). \tag{4.5}$$

We deduce geometric patterns of the persistence degrees:

$$\alpha_{j,h} = |\mu_j|^h,\ j \text{ varying.}$$

It is interesting to note that Sturm-Liouville theorem provides information on the canonical variates. The canonical variate a_j admits exactly j zeros, and its derivative is equal to zero at a single point between two successive zeros of a_j. In particular the first canonical variate is a monotone function.

4.4 Discretized Reflected Brownian Motion

The brownian motion reflected on the interval $[0, l]$ is another example of a stationary Markov continuous time process, which is not a diffusion process and admits a persistence decomposition. The canonical decomposition is:

$$f_h(x_t, x_{t+h}) = f(x_t)f(x_{t+h})\left\{1 + \sum_{j=1}^{\infty}\exp\left[\Leftrightarrow\frac{h}{2}\left(\frac{j\pi}{l}\right)^2\right]\cos\left(\frac{j\pi}{l}x_{t+h}\right)\cos\left(\frac{j\pi}{l}x_t\right)\right\}. \tag{4.6}$$

5 Copersistence

5.1 Definition

The concept of linear comovements between time series and their analysis has been developped for both stationary series [the so-called codependence, see e.g. Kugler, Neusser (1990), Engle, Kozicki

[11] see Hansen, Scheinkman, Touzi (1998) for discussion

(1993), Gourieroux, Peaucelle (1994)] and nonstationary series [the so-called cointegration, see e.g. Granger (1986), Engle, Granger (1987)]. Broadly speaking there exists a linear comovement between the (stationary) series $(X_{1t}), (X_{2t})$ with equivalent patterns of time dependence, if and only if there exists a non degenerate combination $Y_t = a_1 X_{1t} + a_2 X_{2t}$, say, whose autocorrelation pattern is negligible with respect to the autocorrelation patterns of the initial series $(X_{1t}), (X_{2t})$.

Such linear combinations are unique, up to some multiplicative factor, and both coefficients a_1, a_2 are different from zero. Then it is common to write the relation:

$$a_1 X_{1t} + a_2 X_{2t} = \eta_t, \tag{5.1}$$

where (η_t) features less persistence than the initial series. This relation is often interpreted in structural terms either as a stable relationship (e.g. less sensitive to shocks), or a long run relationship.

A similar approach can be followed in the nonlinear framework to derive nonlinear relationships between the initial series. Let us consider a bivariate stationary series $X_t = (X_{1t}, X_{2t})$, and introduce the (maximal) degree of persistence of the univariate series X_1 say, as:

$$\rho_h^*(X_1) = \max_g \rho_h[g(X_{1t})] = \max_{g_1, g_2} \text{ Corr}[g(X_{1,t+h}), g_2(X_t)]. \tag{5.2}$$

Note that this (maximal) degree of persistence is computed with respect to the whole universe.

Definition 5.1: (X_{1t}) and (X_{2t}) are copersistent if and only if there exists a transformation $Y_t = a(X_{1t}, X_{2t})$, which depends both on X_1 and X_2 (non degeneracy condition), is one to one with respect to one of the arguments, and such that:

$$\rho_h^*(Y) = o[\rho_h^*(X_1)] = o[\rho_h^*(X_2)].$$

It is important to impose the non degeneracy condition in the nonlinear framework. Indeed, we know that some nonlinear transformations of (X_{1t}) , for example, may feature less persistence that (X_{1t}) itself, but they clearly do not generate a relationship between X_1 and X_2.

Moreover, if the transformation a is one to one with respect to X_1 (say), we can write:

$$a(X_{1t}, X_{2t}) = \eta_t,$$

or

$$X_{1t} = a_1^{-1}(\eta_t, X_{2t}), \tag{5.3}$$

where a_1^{-1} denotes the inverse with respect to the first argument. Therefore it is possible to express X_{1t} as a nonlinear function of X_{2t} and η_t with different persistence.

Finally note that the definition of copersistence is invariant with respect to one-to-one nonlinear transformations of either (X_{1t}) or (X_{2t}).

By allowing for nonlinear transformations, we have a better chance to find nonlinear comovements along with the linear ones. There may exist however series without copersistence.

5.2 The Multiplicity of Copersistence Directions

The aim of this subsection is to show that there may exist a very large number of copersistence directions, which renders difficult the identification of a nonlinear relationship with structural interpretation.

We consider a bivariate Markov model with finite dimensional dependence [Gourieroux, Jasiak (1999)]. The joint distribution of $(X_t, X_{t+h}) = [(X_{1t}, X_{2t}), (X_{1,t+h}, X_{2,t+h})]$ is:

$$f_h(x_t, x_{t+h}) = 1 + \lambda^h [\sqrt{12}(x_{1,t+h} + x_{2,t+h} \Leftrightarrow 1)][\sqrt{12}(x_{1t} \Leftrightarrow 1/2)],$$

wher $\lambda \in [0, 1/12]$. The marginal distribution of X_t is such that X_{1t} and X_{2t} are independent, with a marginal uniform distribution on $[0,1]$. The canonical decomposition involves a single term [one dimensional time dependence], with the canonical directions $a_1(x) = \sqrt{12}(x_1 + x_2 \Leftrightarrow 1)$, $b_1(x) = \sqrt{12}(x_1 \Leftrightarrow 1/2)$. The parameter λ is bounded to ensure the positivity of the joint p.d.f..

For any nonlinear transformation $a(x)$, we have:

$$\begin{aligned} \rho_h(a(x)) &= \sqrt{\frac{\text{Var}\, E(a(X_{t+h})|X_t)}{\text{Var}\, a(X_t)}} \\ &= \frac{\lambda^h |E[\sqrt{12}(X_{1t} + X_{2t} \Leftrightarrow 1) a(X_t)]|}{\text{Var}\, a(X_t)^{1/2}}. \end{aligned}$$

Therefore we have two different patterns for the transformed autocorrelogram:

· If $E[\sqrt{12}(X_{1t} + X_{2t} \Leftrightarrow 1) a(X_t)] \neq 0$, it features a geometric decay λ^h;

· If $E[\sqrt{12}(X_{1t} + X_{2t} \Leftrightarrow 1) a(X_t)] = 0$, it is equal to zero.

We easily verify that the initial series X_{1t} and X_{2t} admit transformed autocorrelograms (and also maximal degree of persistence) with the same geometric decay λ^h.

Let us now consider a nonlinear transformation of the type:

$$a(X) = X_1 + \alpha c(X_2),$$

where c is a given function such that $\text{Cov}\,[X_2, c(X_2)] \neq 0$. We get:

$$\begin{aligned} E[\sqrt{12}(X_{1t} + X_{2t} \Leftrightarrow 1)a(X_t)] &= \text{Cov}\left[\sqrt{12}(X_{1t} + X_{2t} \Leftrightarrow 1), X_{1t} + \alpha c(X_{2t})\right] \\ &= \sqrt{12}\left(\text{Var}\, X_{1t} + \alpha \text{Cov}\,[X_{2t}, c(X_{2t})]\right). \end{aligned}$$

We deduce that, for any function c such that $\text{Cov}\,[X_2, c(X_2)] \neq 0$, the transformation:

$$a(X) = X_1 \Leftrightarrow \frac{\text{Var}\, X_{1t}}{\text{Cov}\,[X_{2t}, c(X_{2t})]} c(X_2) \tag{5.4}$$

is a copersistence direction. Therefore we get an infinite number of relations of the type:

$$X_{1t} = d(X_{2t}) + \eta_t,$$

with an interpretation in terms of copersistence.

6 Persistence by Trajectories and Nonlinear Cointegration

The persistence degrees have been defined in section 3.5 for stationary processes as measures of the effect of declining memory in terms of both correlation and prediction [see equation (2.3)]. The aim of this section is to extend this notion to homogenous processes, which may feature non stationarities. We first discuss the case of the gaussian random walk to show that persistence assessment in terms of correlations and predictions may differ significantly in the presence of nonstationarities. This leads to the notion of persistence by trajectory, which is defined in the second subsection. Finally we discuss nonlinear cointegration.

6.1 Gaussian Random Walk

Let us consider a gaussian random walk:

$$X_t = \sum_{\tau=1}^{t} \epsilon_\tau, \tag{6.1}$$

where the components of the noise are i.i.d., with N(0,1) distribution. This is a homogenous Markov process. It has been proven in Ermini, Granger (1993) that the exponential transformations of the random walk are such that:

$$E[\exp \mu X_{t+h} | X_t] = \exp \frac{h\mu^2}{2} \exp \mu X_t, \tag{6.2}$$

and

$$\begin{aligned}\rho_{t,h} &= \max_{g_2} \operatorname{Corr}[\exp \mu X_{t+h},\, g_2(X_t)] \\ &= \operatorname{Corr}(\exp \mu X_{t+h},\, \exp \mu X_t) \\ &= \sqrt{\frac{\exp \mu^2 t \Leftrightarrow 1}{\exp \mu^2 (t+h) \Leftrightarrow 1}}. \end{aligned} \tag{6.3}$$

The exponential random walk clearly features an explosive behaviour, implied by the autoregressive representation (6.2), where the autoregressive coefficient $\exp \frac{h\mu^2}{2}$ is larger than one. However when we consider the behaviour of the autocorrelation function for large t, we get:

$$\bar{\rho}_h = \lim_{t\to\infty} \rho_{t,h} = \exp \Leftrightarrow \frac{\mu^2 h}{2} = \left(\exp \Leftrightarrow \frac{\mu^2}{2}\right)^h,$$

which resembles the autocorrelation of a stationary AR(1) process. This implies that, in general, an assessment of persistence based on the asymptotic properties of the prediction is preferable.

6.2 Persistence by Trajectory (b.t.)

We consider a process (X_t) and a possibly extended universe (Z_t). We assume that the process is homogenous with respect to the universe, i.e. that the conditional distributions of X_t given $\underline{Z_{t-1}}$ do not depend on the date t. Then, the conditional expectations $E[g(X_{t+h})|Z_t]$ are also time independent.

Let us now introduce a positively valued sequence $\alpha = (\alpha_h,\ h \geq 0)$ converging to zero at infinity.

Definition 6.1: The by trajectory (b.t.) persistence space of order α is defined by:
$E_\alpha^{b.t.} = \{g$ such that there exists a scalar $c(g)$ with: $E(g(X_{t+h})|Z_t) \Leftrightarrow c(g) = 0(\alpha_h)$ a.s. $\}$.

Therefore we have $|E(g(X_{t+h})|Z_t) \Leftrightarrow c(g)| \leq \alpha(Z_t)\, \alpha_h$ a.s.. It is easily verified that $E_\alpha^{b.t.}$ is a vector space, has the monotonicity property, and b.t. persistence degrees can be defined along the lines of subsection 3.3. However we also need to separate the stationary and nonstationary components of the process (X_t). For this purpose, we can consider the vector space:

$$E^{b.t.}(0) = \cup_\alpha E_\alpha^{b,t}, \tag{6.4}$$

where the union is taken over all possible α sequences. A transformation g belongs to the space $E^{b.t.}(0)$ if and only if the prediction $E[g(X_{t+h})|Z_t]$ becomes independent of Z_t when the horizon h tends to infinity. This condition is satisfied for stationary regular processes, usually called I(0) in the literature. It is not satisfied when $g(X_t)$ features nonstationarities or nonregularities.

Definition 6.2: If g belongs to $E^{b.t.}(0)$, the transformed process is NLI(0) (nonlinearly integrated of order 0).

If g does not belong to $E^{b.t.}(0)$, the transformed process is NLI (nonlinearly integrated).

6.3 Nonlinear Cointegration

Let us consider a bivariate process $X_t = (X_{1t}, X_{2t})'$ homogenous with respect to the information $(Z_t) = (X_t)$.

Definition 6.3: The components X_1 and X_2 are nonlinearly cointegrated iff:
(i) (X_{1t}) and (X_{2t}) are NLI with respect to the universe $(Z_t) = (X_t)$;
(ii) There exists a transformation $Y_t = a(X_{1t}, X_{2t})$, which depends both on X_1 and X_2, is one to one with respect to one of the arguments and such that (Y_t) is NLI(0) with respect to the universe (X_t).

We see that as comovements, there may exist a large multiplicity of cointegration directions, causing an identification problem.

Example 6.1:

Let us consider three independent gaussian white noises $(\epsilon_{1,t}), (\epsilon_{2,t}), (\epsilon_{3,t})$, and define:

$Z_{1,t} = \rho_1 Z_{1,t-1} + \epsilon_{1,t}$ (AR(1) process),

$Z_{2,t} = \rho_2 Z_{2,t-1} + \epsilon_{2,t}$ (AR(1) process),

$Z_{3,t} = \sum_{\tau=1}^{t} \epsilon_{3,\tau}$ (random walk),

$X_{1,t} = sgn(\epsilon_{1,t})|Z_{1,t}|\,|Z_{3,t}|$,

$X_{2,t} = sgn(\epsilon_{2,t})|Z_{2,t}|\,|Z_{3,t}|$.

The components $(X_{1,t})$ and $(X_{2,t})$ are NLI due to the presence of the random walk $|Z_{3,t}|$. Moreover any transformation of $X_{1,t}/X_{2,t}$, $sgn(X_{1,t})$, $sgn(X_{2,t})$ is NLI(0). For instance the directions $X_{1,t}/X_{2,t} + a\, sgn(X_{2,t})$ are cointegration directions for any scalar a.

There exist various ways to handle the multiplicity problem.
(i) We can try to find the set of all nonlinear cointegration directions in a nonparametric approach. Such an approach, however is likely unfeasible except for cases where either the associated space, or its supplement are included in a vector space of a small dimension.
(ii) Alternatively we can consider the problem under constraints. We can restrict the set of admissible distributions of the process (X_t) and/or of the admissible forms of cointegration directions. This approach is generally followed in the literature as illustrated by the examples below.

Example 6.2

For a bivariate process we know that linear cointegration directions are included in a space of dimension equal or less than one. Therefore the linearity of the cointegration direction is a kind of identifying constraint. However there may exist strictly nonlinear cointegration directions in the absence of linear cointegration directions, and some nonlinear directions may coexist with the linear ones.

Example 6.3

In a series of papers Park, Phillips (1998;a,b), Karlsen, Myklebust, Tjostheim (1999) considered the estimation of nonlinear relations of the type $X_{1,t} = f(X_{2,t}) + u_{1,t}$, where $X_{2,t}$ features nonstationarities, and they introduced various assumptions on the error term $u_{1,t}$. In Park, Phillips (1998;a), Karlsen, Myklebust, Tjostheim (1999) the assumption imposes the independence between the processes X_2 and u_1. It is easily seen that this is an identification condition of the regression function f. Indeed, if we consider two admissible decompositions $X_{1,t} = f(X_{2,t}) + u_{1,t} = \tilde{f}(X_{2,t}) + \tilde{u}_{1,t}$, we deduce $u_{1,t} = \tilde{f}(X_{2,t}) \Leftrightarrow f(X_{2,t}) + \tilde{u}_{1,t}$, and $(u_{1,t})$ is independent of $(X_{2,t})$ if and only if $f = \tilde{f}$. In Park, Phillips (1998;b), the error term is assumed to be a martingale difference sequence with respect to the information (X_t). If we consider two admissible decompositions $X_{1,t} = f(X_{2,t}) + u_{1,t} = \tilde{f}(X_{2,t}) + \tilde{u}_{1,t}$, we get $u_{1,t} \Leftrightarrow \tilde{u}_{1,t} = \tilde{f}(X_{2,t}) \Leftrightarrow f(X_{2,t})$. We find that f is identifiable if and only if the only martingale difference sequence deterministic function of $(X_{2,t})$ is zero.

6.4 Markov Process with Finite Dimensional Dependence

Let us consider a homogenous Markov process with the transition function:

$$\begin{aligned} p(x_{t+1}|x_t) &= \sum_{j=1}^{J} b_j(x_{t+1}) a_j(x_t) \\ &= b'(x_{t+1}) a(x_t), \quad \text{say.} \end{aligned} \tag{6.5}$$

It is equivalent to assume the previous decomposition or a finite dimensional predictor space [Gourieroux, Jasiak (1999)]. The predictors, such as $E[g(X_{t+h})|X_t]$, g, h varying, belong to the finite dimensional space generated by $a_j(X_t)$, $j = 1, ..., J$. The elements a and b of the decomposition (6.5) are defined up to an inversible linear transformation.

The transition function h-step ahead is given by:

$$p^{(h)}(x_{t+h}|x_t) = b'(x_{t+h}) C^{h-1} a(x_t), \tag{6.6}$$

where the elements of the C matrix are $c_{i,j} = \int a_i(x) b_j(x) dx$. The predictors are easily derived from:

$$\begin{aligned} E[g(X_{t+h})|X_t] &= \int g(x_{t+h}) p^{(h)}(x_{t+h}|X_t) dx_{t+h} \\ &= \int g(x) b'(x) dx\, C^{h-1} a(X_t). \end{aligned} \quad (6.7)$$

To simplify the discussion we assume that the matrix C can be diagonalized with real eigenvalues $\lambda_j,\ j = 1, ..., J$. We denote by $C = \sum_{j=1}^{J} \lambda_j u_j v_j'$ a spectral decomposition of C. We get:

$$E[g(X_{t+h})|X_t] = \sum_{j=1}^{J} \lambda_j^{h-1} \int g(x)[b'(x)u_j] dx\, v_j' a(X_t). \quad (6.8)$$

We can search now for the NLI(0) directions. Two cases have to be distinguished, depending on the eigenspace associated to the unitary eigenvalues [12].

Case 1: The eigenspace associated to the unitary eigenvalue does not contain the constant function.

The space $E^{b.t.}(0)$ is:

$$E^{b.t.}(0) = \{g : \int g(x)[b'(x)u_j] dx = 0, \quad \text{for any } j \text{ with } |\lambda_j| \geq 1\}.$$

Case 2: The eigenspace associated to the unitary eigenvalue contains the constant function.

Up to a change of basis we can always assume that $\lambda_1 = 1$ is associated with $v_1' a(x_t) = 1$. We get:

$$E^{b.t.}(0) = \{g : \int g(x)[b'(x)u_j] dx = 0, \quad \text{for any } j \geq 2, \text{ with } |\lambda_j| \geq 1\}.$$

The spaces $E^{b.t.}(0)$ have an infinite dimension and for a bivariate process $X_t = (X_{1t}, X_{2t})$ there is in general a multiplicity of cointegration directions.

We note that there exist NLI(0) transformations which are linear combinations of the b_j functions. Therefore, after an appropriate change of the factor a and b, we can write (in case 1):

$$p(x_{t+1}|x_t) = \sum_{j=1}^{J_1} b_j(x_{t+1}) a_j(x_t) + \sum_{j=J_1+1}^{J} b_j(x_{t+1}) a_j(x_t), \quad (6.9)$$

where $b_j,\ j = J_1 + 1, ..., J$ [resp. $j = 1, ..., J_1$] are NLI(0) transformations [resp. NLI transformations]. Then, it is possible to separate the stationary and "nonstationary" components of the transition function.

[12] Markov processes with finite dimensional dependence are direct extensions of Markov chains with a finite state space. Let us denote by $j = 1, ..., J$ the admissible states and by $P = (p_{i,j})$ the transition matrix. The transition function can be written:

$$p(x_{t+1}|x_t) = a'(x_{t+1}) P a(x_t),$$

where $a_j(x_t) = 1$, if $x_t = j$, 0 otherwise. In this particular case $b = P'a$, $C = P$, and $p^{(h)}(x_{t+1}|x_t) = a(x_{t+1})' P^h a(x_t)$. The persistence analysis is based on the analysis of the eigenvalues of P.

7 Statistical Inference

7.1 Empirical Nonlinear Canonical Analysis

In practice the analysis of persistence and copersistence directions can be based on the empirical nonlinear canonical decomposition of univariate or bivariate time series.

More precisely, let us consider the covariance operator at lag h:

$$< \phi(X_t), \psi(X_{t+h}) >_h = \int\int \phi(x_t)\psi(x_{t+h}) f_h(x_t, x_{t+h}) dx_t dx_{t+h}. \tag{7.1}$$

It can be approximated by replacing the unknown joint p.d.f. f_h by a kernel estimator. The approximated operator is:

$$< \phi(X_t), \psi(X_{t+h}) >_{h,T} = \int\int \phi(x_t)\psi(x_{t+h}) \hat{f}_{h,T}(x_t, x_{t+h}) dx_t dx_{t+h}, \tag{7.2}$$

where $\hat{f}_{h,T}(x,y) = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{h^{2d}} K\left(\frac{X_t - x}{h}\right) K\left(\frac{X_{t+h} - y}{h}\right)$, K is a kernel, and h is the bandwidth.

Next, we use the approximated kernel, to obtain the estimated canonical correlations and canonical variates, i.e.:

$$\hat{f}_{h,T}(x_t, x_{t+h}) = \hat{f}_{h,T}(x_t)\hat{f}_{h,T}(x_{t+h})\{1 + \sum_{j=1}^{\infty} \hat{\mu}_{j,h} \hat{a}_{j,h}(x_{t+h}) \hat{b}_{j,h}(x_t)\}. \tag{7.3}$$

The consistency and asymptotic distributional properties of the estimated correlations and of the functional approximations of canonical directions have been derived in Darolles, Florens, Gourieroux (1998), for stationary, geometrically mixing processes. In particular these results can be used to check if the canonical directions become stable for large lag h [13]. When this condition is satisfied it is possible to proceed with the persistence decomposition as shown in subsection 4.1.

7.2 Simulations of Fractional Gaussian Processes

The theoretical properties of nonlinear kernel-based canonical analysis have not been derived yet in the long memory framework. To provide some insights on the performance of this method, we examine simulated realizations of a gaussian, fractionally integrated process with parameter d. From Example 1 of section 4 we know that the expressions of canonical directions are Hermite polynomials, and that the limiting behaviour of the canonical correlations is a hyperbolic decay. We wish to investigate the performance of correlations obtained from the canonical analysis by comparing the following correlation estimators:

(i) $\hat{\mu}^{(1)}_{j,h} = \hat{\mathrm{Corr}}(H_j(x_t), H_j(x_{t+h}))$, i.e. the autocorrelogram computed for the Hermite polynomial of degree j;

[13] It is interesting to note that this stability has been observed in intertrade duration data, from financial markets [Gourieroux, Jasiak (1998)].

(ii) $\hat{\mu}^{(2)}_{j,h} = \hat{\rho}_h(H_j(x_{t+h}))$, i.e. the transformed autocorrelogram for the Hermite polynomial of degree j;

(iii) $\hat{\mu}^{(3)}_{j,h}$ the kernel based canonical correlation of order j.

These estimators require a diminishing amount of information on the canonical directions. Indeed $\hat{\mu}^{(1)}$ requires the knowledge of the current and lagged canonical directions, whereas $\hat{\mu}^{(2)}$ only requires the current directions, and $\hat{\mu}^{(3)}$ none of them, i.e. no information about the canonical directions at all. For real series $\hat{\mu}^{(3)}$ is the only implementable method. In the second step our analysis consists in using the estimated correlations to build crude estimators of the fractional order, and comparing their performance. This study is based on the following simulation experiment.

We simulate a trajectory of a fractionally integrated process $(1 \Leftrightarrow L)^d X_t = \epsilon_t, \;\; \epsilon_t \sim N(0,1)$ of length T = 4000. Figures 7.1, 7.2, 7.3 display the three types of autocorrelograms, for J=1,...,4, h =1,...,500, and d=0.45.

[Insert Figure 7.1: Autocorrelogram of $H_j(x)$]

[Insert Figure 7.2: Transformed Autocorrelogram of $H_j(x)$]

[Insert Figure 7.3: Nonlinear Autocorrelogram]

The formulas of estimators imply that in general $\hat{\mu}^{(1)}_{j,h} \leq \hat{\mu}^{(2)}_{j,h} \leq \hat{\mu}^{(3)}_{j,h}$, as it is observed from the Figures. In the second step we use the estimated autocorrelograms to derive various estimators of the fractional degree d. For a given autocorrelogram $\hat{\mu}^{(k)}_{j,h}, \;\; k = 1, ..., 3, \;\; j = 1, ..., 4$ we regress $\log |\hat{\mu}^{(k)}_{j,h}|$ on 1 and $\log h$ for large h. The regression coefficient of $\log h$ provides an estimator of $2d_j \Leftrightarrow 1$, and of d using the formulas of Table 4.1. We have run such regressions for h=1,...500 and two simulated series of length 4000 corresonding to d=0.45, and d=0.3 respectively.

Table 7.1: Estimation of d (true value $d = 0.45$)

j	acf	T-acf	nonlinear acf
1	0.465 *(2.72)*	0.475 *(3.66)*	0.485 *(3.57)*
2	0.45 *(6.62)*	0.48 *(8.53)*	0.485 *(6.85)*
3	0.44 *(3.00)*	0.485 *(3.94)*	0.49 *(3.76)*
4	0.45 *(2.65)*	0.48 *(3.88)*	0.49 *(3.71)*

Table 7.2: Estimation of d (true value $d = 0.3$)

j	acf	T-acf	nonlinear acf
1	0.34 *(3.07)*	0.42 *(1.44)*	0.41 *(2.08)*
2	0.54 *(1.41)*	0.50 *(1.79)*	0.45 *(1.90)*
3	0.50 *(0.72)*	0.52 *(2.42)*	0.47 *(2.05)*
4	0.50 *(0.01)*	0.49 *(0.71)*	0.49 *(1.35)*

The crude estimator of the fractional order is known for its lack of precision. Indeed, we find estimates differing significantly from true d's in the columns showing results based on standard a.c.f. of Hermite polynomials in both Tables. Especially this difference is more pronounced in Table 7.2 for $d = 0.3$ which lies further from the region of nonstationarity, and for polynomials of higher degrees. In fact, for large values of the (X_t) process, the polynomials display an explosive behavior, resulting in more extreme values, with a stronger serial correlation in polynomials of higher orders. The large number of extremes reduces the precision of autocorrelation estimators whereas their serial correlation induces a finite sample positive bias in the estimated persistence coefficient [see, Deo, Hurvich (1999) for a similar analysis].

In Figure 7.4 we display the extreme values of Hermite polynomials for a simulated long memory process with $d = 0.3$. The extremes are defined as observations differing from the mean by more than three standard deviations.

[Insert Figure 7.4: Extremes of Hermite Polynomials, j = 1,2,3,4,]

The behaviour of polynomials may even result in estimated values of d beyond the stationarity region (see Table 7.2). This effect may be weaker for the nonlinear a.c.f., where the (unknown) canonical directions are kernel smoothed, and hence extreme values of the transformed series are less frequent.

8 Conclusions

In this paper we investigated various aspects of persistence in nonlinear time series. We introduced the concept of persistence space and defined the degrees of persistence of nonlinear processes. Several examples of persistence decomposition were discussed, including the long memory processes where we highlighted the difference between a fractionally integrated process and a beta mixture of AR(1) processes, continuous time processes and chaos.

In the multivariate framework we pointed out the problem of multiplicity of copersistence or cointegration directions. We emphasized the role of identifying constraints, and commented on some results from recent literature on nonlinear cointegration, where this problem may arise.

Appendix 1

The Analytical Expression of the T-autocorrelogram

The definition of correlation implies that:

$$
\begin{aligned}
\mathrm{Corr}[g(X_{t+h}), g_2(X_t)] &= \frac{\mathrm{Cov}\,[E(g(X_{t+h})|X_t), g_2(X_t)]}{\sqrt{\mathrm{Var}\, g(X_{t+h})}\sqrt{\mathrm{Var}\, g_2(X_t)}} \\
&= \sqrt{\frac{\mathrm{Var}\, E(g(X_{t+h})|X_t)}{\mathrm{Var}\, g(X_{t+h})}}\ \mathrm{Corr}[E(g(X_{t+h})|X_t), g_2(X_t)].
\end{aligned}
$$

It admits the maximum for $g_2(X_t) = E(g(X_{t+h})|X_t)$. At this point $\mathrm{Corr}[g(X_{t+h}), g_2(X_t)] = 1$, and the expression of $\rho_h(g)$ is easily found.

Figure 7.1: Autocorrelogram of H_j(x)

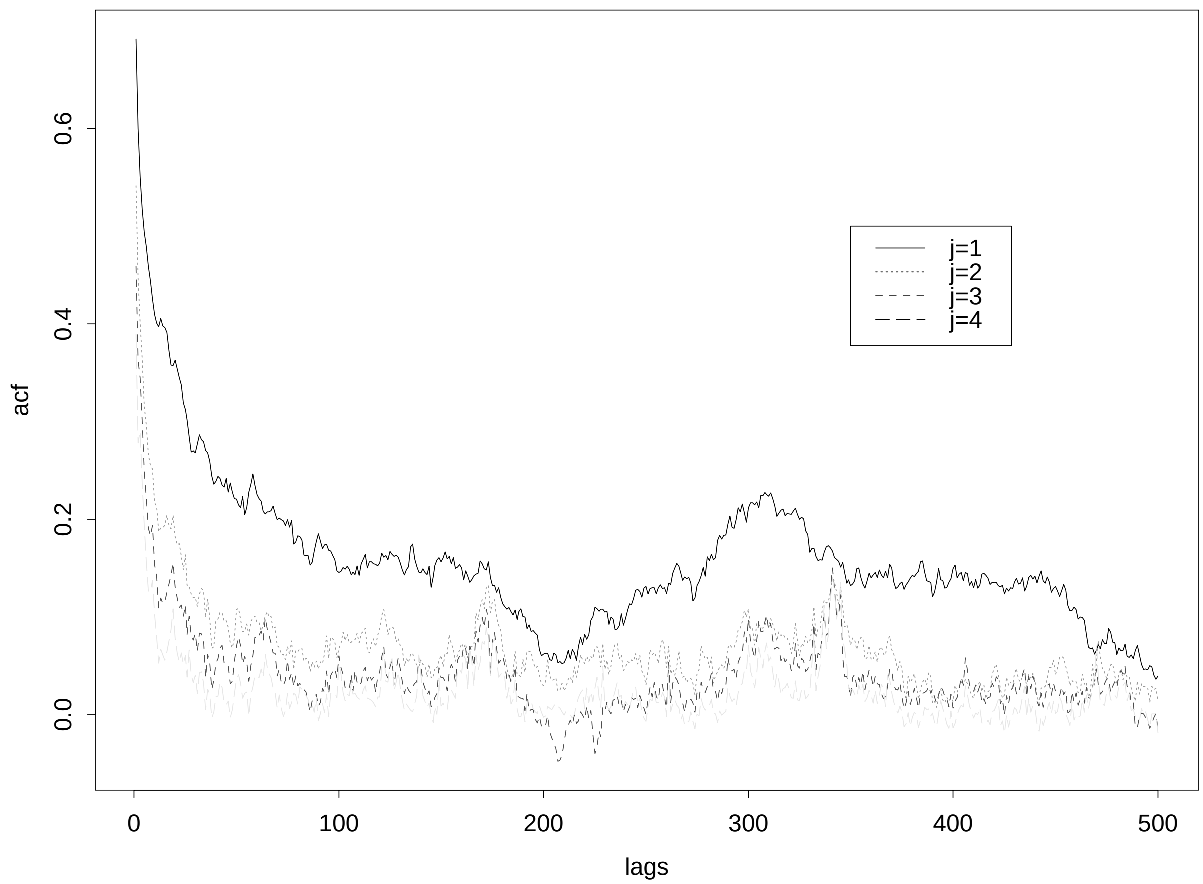


Figure 7.2: Transformed Autocorrelogram of H_j(x)

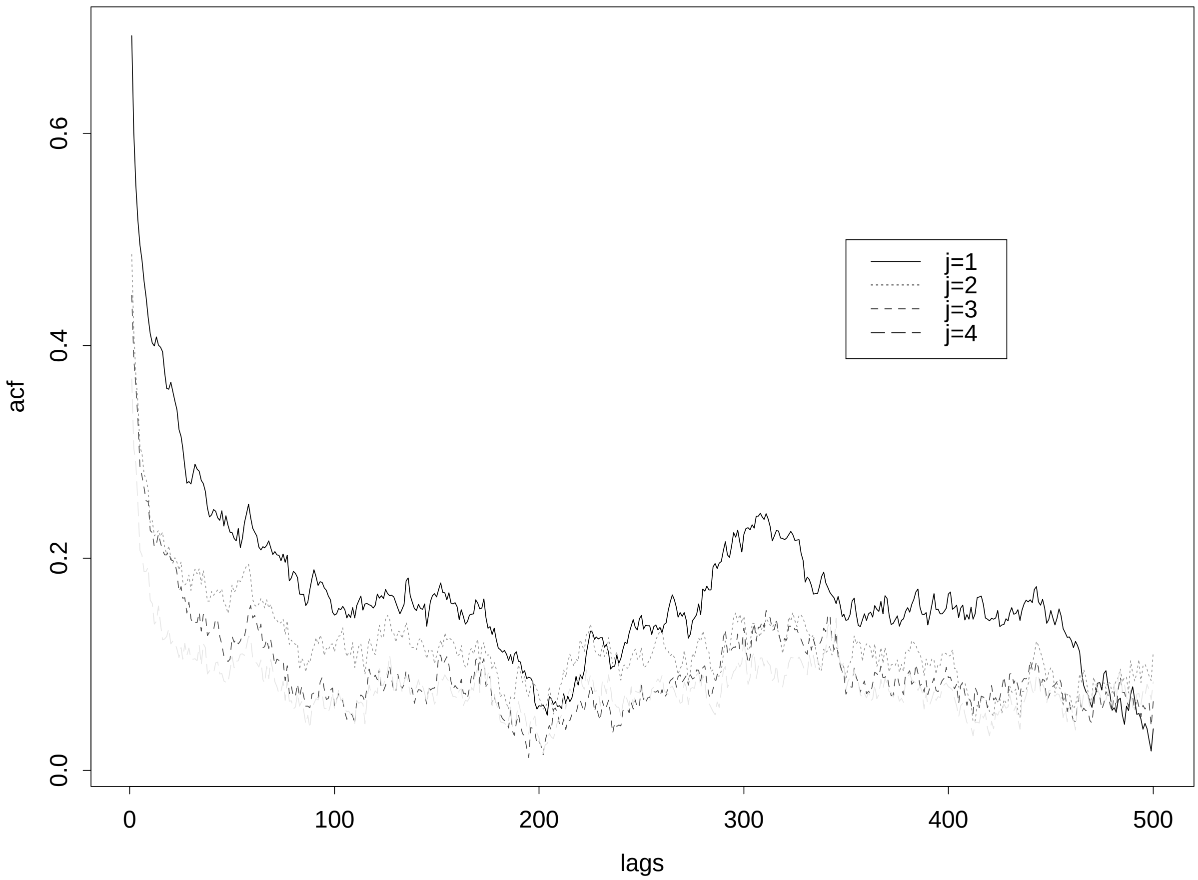


Figure 7.3: Nonlinear Autocorrelogram of H_j(x)

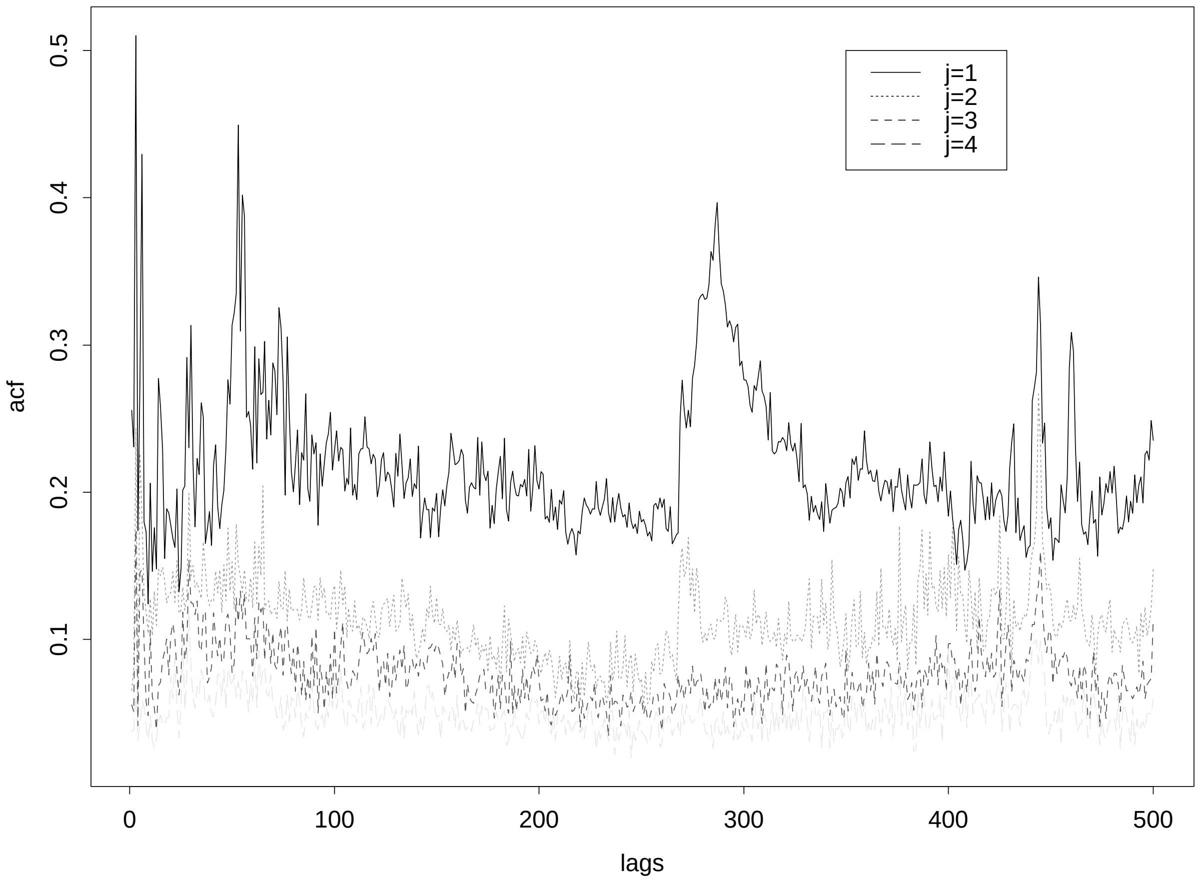
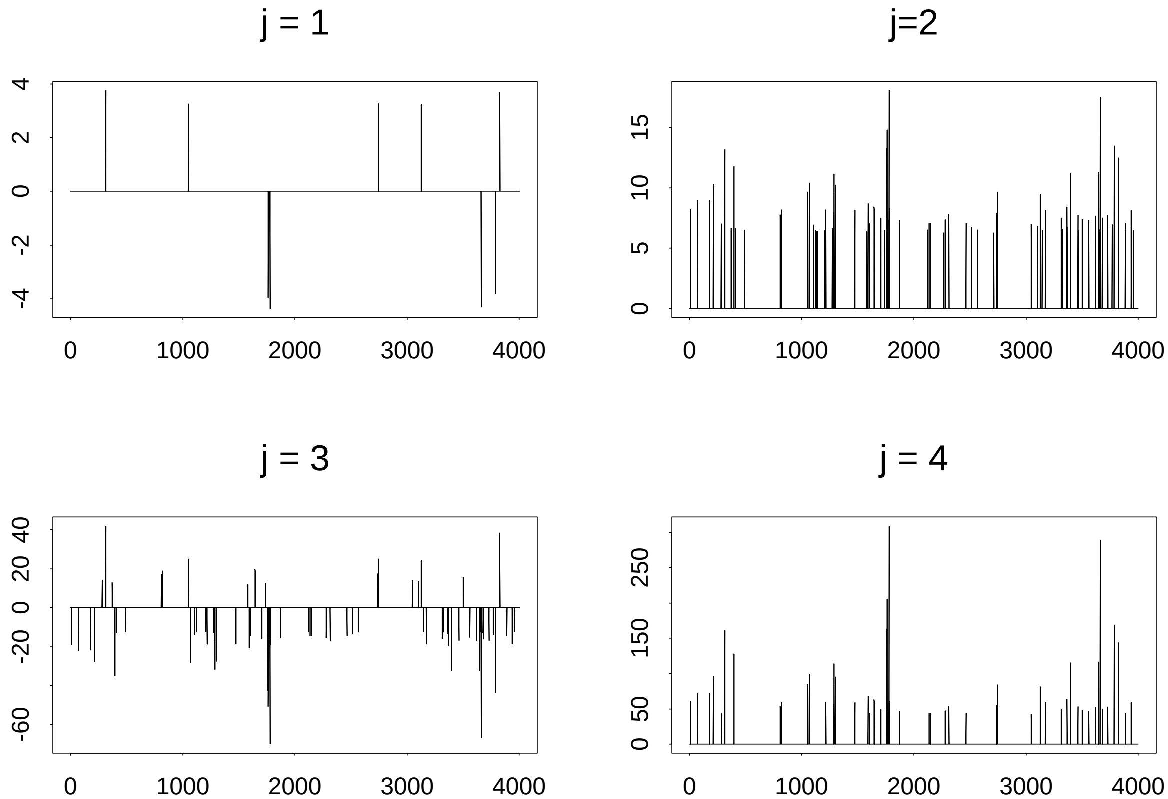


Figure 7.4: Extremes of Hermite Polynomials, j=1,2,3,4



References

[1] Anderson, H., and F. Vahid (1998): "Testing Multiple Equation Systems for Common Nonlinear Components", Journal of Econometrics, 84, 1-36.

[2] Abadir, K. and G. Talmain (1999): " Autocovariance Functions of Series and their Transforms", Manuscript, University of York.

[3] Barrett, J. and D. Lampard (1955): "An Expansion for Some Second Order Probability Distributions and its Application to Noise Problems", I.R.E. Trans, PGIT, IT-1, 10-15.

[4] Bierens,H.J. (1999): "Nonparametric Nonlinear Co-Trending Analysis, With an Application to Interest and Inflation in the U.S.", paper presented at the Cowles Foundation Conference on "New Developments in Time Series Econometrics", Yale University, October 1999.

[5] Chang, Y., Park, J., and P.C.B. Phillips (1999): "Nonlinear Econometric Models with Cointegrated and Deterministically Trending Regressors", paper presented at the Cowles Foundation Conference on "New Developments in Time Series Econometrics", Yale University, October 1999.

[6] Corradi, V. (1995): "Nonlinear Transformations of Integrated Time Series: A Reconsideration", Journal of Time Series Analysis, 16, 6, 537-549.

[7] Corradi, V. and H. White (1994): "Testing for Stationarity, Ergodicity and for Comovements Between Nonlinear Discrete Time Markov Processes", Manuscript, University of Pennsylvania.

[8] Cramer, H. (1963): "Mathematical Methods of Statistics", Princeton University Press.

[9] Darolles, S., Florens, J.P., and C. Gourieroux, (1998): "Kernel Based Canonical Analysis", CREST, D.P. 9855.

[10] Davydoff, J. (1973): "Mixing Conditions for Markov Chains", Theory of Probability and its Applications, 18, 2, 312-328.

[11] Deo, R., and C. Hurvich (1999): "On the Log Periodogram Regression Estimator of the Memory Parameter in Long Memory Stochastic Volatility Models", paper presented at the Cowles Foundation Conference on "New Developments in Time Series Econometrics", Yale University, October 1999.

[12] Ding, Z. and C. Granger (1996): "Modeling Volatility Persistence of Speculative Returns: A New Approach", Journal of Econometrics 73, 185-216.

[13] Ding, Z., Granger, C. and R. Engle (1993): "A Long Memory Property of Stock Market Returns and a New Model", Journal of Empirical Finance, 1, 83-106.

[14] Dunford, N. and J. Schwartz (1968): "Linear Operators: Part I", Wiley, New York.

[15] Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the U.K. Inflation", Econometrica, 40, 987-1008.

[16] Engle, R. and S. Kozicki (1993): "Testing for Common Features", Journal of Business and Economic Statistics, 11, 369-380.

[17] Engle, R. and C. Granger (1987): "Cointegration and Error Correction Representation, Estimation and Testing", Econometrica, 55, 251-276.

[18] Ermini, L. and C. Granger (1993): "Some Generalizations on the Algebra of I(1) Processes, Journal of Econometrics, 58, 369-384.

[19] Geweke, J. and S. Porter-Hudak (1983): "The Estimation and Application of Long Memory Time Series Models", Journal of Time Series Analysis, 4, 221-238.

[20] Gourieroux, C. and J. Jasiak (1998): "Nonlinear Autocorrelograms: An Application to Inter Trade Durations", CREST D.P. 9841.

[21] Gourieroux, C. and J. Jasiak (1999): "Finite Factor Markov Models", CREST D.P.

[22] Gourieroux, C. and I. Peaucelle (1992): "Series Codependantes: Application à l'hypothèse du Pouvoir d'Achat", Revue d'Analyse Economique, 68, 283-304.

[23] Granger, C. (1986): "Developments in the Study of Cointegrated Economic Variables", Oxford Bulletin of Economics and Statistics", 48, 213-228.

[24] Granger, C. (1993): "Strategies for Modelling Nonlinear Time Series Relationships", Economic Record, 69, 233-238.

[25] Granger, C. (1995): "Modelling Nonlinear Relationships Between Extended Memory Variables", Econometrica, 63, 265-279.

[26] Granger, C. (1999): "Aspects of Research Strategies for Time Series Analysis", paper presented at the Cowles Foundation Conference on "New Developments in Time Series Econometrics", Yale University, October 1999.

[27] Granger, C. and Hallman (1991): "Nonlinear Transformations of Integrated Time Series", Journal of Time Series Analysis", 12, 207-234.

[28] Granger, C. and R. Joyeux (1980): "An Introduction to Long Memory Time Series Models and Fractional Differencing", Journal of Time Series Analysis, 1, 15-29.

[29] Granger, C. and P. Newbold (1976): "Forecasting Transformed Series", Journal of the Royal Statistical Society, B, 38, 189-203.

[30] Granger, C. and T. Terasvirta (1993): "Modelling Nonlinear Economic Relationships", Oxford University Press.

[31] Granger, C. and T. Terasvirta (1999): " A Simple Nonlinear Time Series Models with Misleading Linear Properties", Economic Letters, 62, 161-165.

[32] Hamilton, J. (1994): "Time Series Analysis", Princeton University Press.

[33] Hansen, L., Scheinkman, J, and N. Touzi (1998): "Spectral Methods for Identifying Scalar Diffusions", Journal of Econometrics, 86, 1-32.

[34] He, C. and T. Terasvirta (1997): "Statistical Properties of the Asymmetric Power ARCH Process", D.P. Stockholm School of Economics.

[35] Hong, Y., and T. Lee (1999): "Diagnostic Checking for Adequacy of Linear and Nonlinear Time Series Models", paper presented at the Cowles Foundation Conference on "New Developments in Time Series Econometrics", Yale University, October 1999.

[36] Hosking, J. (1981): "Fractional Differencing", Biometrika, 68, 165-176.

[37] Johansen, S (1988): "Statistical Analysis of Cointegration Vectors", Journal of Economic Dynamics and Control, 12, 231-254.

[38] de Jong, R.M. (1999): "Nonlinear Minimization Estimators in the Presence of Cointegrating Relations", paper presented at the Cowles Foundation Conference on "New Developments in Time Series Econometrics", Yale University, October 1999.

[39] Karlsen, H., Myklebust C., and D. Tjostheim (1999): "Nonparametric Estimates in a Nonlinear Cointegration Type Model", paper presented at the Cowles Foundation Conference on "New Developments in Time Series Econometrics", Yale University, October 1999.

[40] Kugler, P. and K. Neusser (1990): "International Real Interest Rate Equalization", Wien University, D.P.

[41] Lancaster, H. (1968): "The Structure of Bivariate Distributions", Annals of Mathematical Statistics, 29, 719-736.

[42] Lee, T., White H., and C. Granger (1993): "Testing for Neglected Nonlinearity in Time Series Models", Journal of Econometrics, 56, 268-290.

[43] Lepnik, R. (1958): "The Effect of Instantaneous Nonlinear Devices on Cross-Correlation", I.R.E. Trans. Information Theory, IT-4, 73-76.

[44] Park, J., and P.C.B. Phillips (1998a): "Asymptotics for Nonlinear Transformations of Integrated Time Series", Cowles Foundation D.P.

[45] Park, J., and P.C.B. Phillips (1998b): "Nonlinear Regression with Integrated Time Series", paper presented at the Cowles Foundation Conference on "New Developments in Time Series Econometrics", Yale University, October 1999.

[46] Tiao, G., and R. Tsay (1989): "Model Specification in Multivariate Time Series", Journal of the Royal Statistical Society, B51, 157-213.

[47] Tong, H. (1990): "Nonlinear Time Series: A Dynamical System Approach", Oxford University Press.

[48] Vahid,F., and R. Engle (1993): "Common Trends and Common Cycles", Journal of Applied Econometrics, 8, 341-360.

[49] Vahid,F., and R. Engle (1997): "Codependent Cycles", Journal of Econometrics, 80, 199-221.