CEIS Tor Vergata
RESEARCH PAPER SERIES

Working Paper No. 100 March 2007

Polynomial Cointegration between Stationary Processes with Long Memory

Marco Avarucci Domenico Marinucci

CEIS Tor Vergata - Research Paper Series, Vol. 34, No.100 , March 2007

Polynomial Cointegration between Stationary Processes with Long Memory*

Marco Avarucci
SEFeMEQ, University of Rome "Tor Vergata"

Domenico Marinucci
Department of Mathematics, University of Rome "Tor Vergata"

November 20, 2006

Abstract

In this paper we consider polynomial cointegrating relationships between stationary processes with long range dependence. We express the regression functions in terms of Hermite polynomials and we consider a form of spectral regression around frequency zero. For these estimates, we establish consistency by means of a more general result on continuously averaged estimates of the spectral density matrix at frequency zero.

Keywords and phrases: Nonlinear cointegration, Long memory, Hermite polynomials, Spectral regression, Diagram formula.

AMS classification: Primary 62M15, Secondary 62M10, 60G10

1 Introduction

The extension of the standard cointegration paradigm to more general, fractional circumstances has drawn growing attention in the time series literature over the last decade, prompting the development of many novel estimation approaches. Robinson (1994) introduced the idea of using degenerating, narrow band regression in a long memory context, establishing also consistency for fractional cointegrating relationships in the stationary case. The

*This paper is part of the first author PhD dissertation. We are grateful to Franco Peracchi and an Associated Editor for useful comments.

properties of this estimator (which has become known as NBLS) were then investigated under nonstationary circumstances by Marinucci and Robinson (2001), Robinson and Marinucci (2001, 2003). Chen and Hurvich (2003a,b) considered principal components methods in the frequency domain, whereas Velasco (2003), Robinson and Hualde (2003) advocate pseudo-maximum likelihood methods which improve the efficiency of the estimates and yield standard asymptotic properties. Cointegration among stationary processes has also been considered, for instance by Marinucci (2000), Christensen and Nielsen (2006).

All these papers have focused on the case of linear cointegration. Nevertheless, the possibility of polynomial cointegrating relationships seems of practical interest, for instance (but not exclusively) for applications to financial data. Nonlinear cointegration has been considered in the literature (most recently by Karlsen, Myklebust, and Tjostheim (2006)), but only in non-fractional circumstances, to the best of our knowledge. In this paper, we shall focus on nonlinear cointegrating relationships between stationary long memory processes; the restriction to a stationarity framework is made necessary by the need to exploit expansions into Hermite polynomials, a powerful tool to investigate nonlinear transformations (see for instance Giraitis and Surgailis (1985), Arcones (1994), Surgailis (2003)). Our general setting can be explained as follows. Let $\{A_t\} = \{x_t, e_t\}$, $t \in \mathbb{Z}$ be a stationary bivariate time series with mean zero and covariance such that

$$\mathbb{E}A_t A'_{t+\tau} := \Gamma(\tau) = \int_0^{2\pi} f(\lambda) e^{i\tau\lambda} d\lambda \ ,$$

where

$$f(\lambda) = \begin{bmatrix} f_{xx}(\lambda) & f_{xe}(\lambda) \\ f_{ex}(\lambda) & f_{ee}(\lambda) \end{bmatrix} ,$$

is the spectral density matrix of $\{A_t\}$. We shall take $\{x_t, e_t\}$ to be long memory, in the sense that

$$\gamma_{ab}(\tau) \simeq G_{ab} \tau^{d_a + d_b - 1} \tag{1}$$

for $a, b = x, e$, $0 < d_a, d_b < \frac{1}{2}$, $G_{xx}, G_{ee} > 0$, $|G_{xe}| \geq 0$. We write $z_t \sim I(d_z)$ for long memory processes with memory parameter d_z, and $\simeq$ to denote that the ratio of the left- and right-hand sides tends to 1.

Now assume there is a polynomial function $g(\cdot)$ such that $\mathbb{E}[g(x_t)] = 0$ and

$$y_t = g(x_t) + e_t, \qquad 0 < d_e < d_y \leq d_x < 1/2 \ ; \tag{2}$$

in this case, we say that y_t, x_t are nonlinearly cointegrated. Clearly, the standard (stationary) fractional cointegrating relationship is obtained in the special case where $d_x = d_y$ and $g(\cdot)$ is a linear function. It is very important to stress that x_t, e_t are allowed to be correlated, which entails Ordinary Least Squares are typically inconsistent in these stationary circumstances (Robinson (1994)).

Our main idea in this paper is to write $g(\cdot)$ as a sum of Hermite polynomials; the coefficients of these polynomials will be estimated by means of a spectral regression method, as in Robinson (1994), Marinucci (2000) and Marinucci and Robinson (2001). We shall show that, by using a degenerating band of frequencies around the origin, then the estimator of these coefficients is consistent, despite the lack of orthogonality between x_t and e_t.

In the sequel, C denotes a generic, positive, finite constant, which need not to be the same all the time it is used; for two generic matrices A and B, of equal dimension, we say that $A \simeq B$ if, for each (i, j), the ratio of the (i, j)-th elements of A and B tends to unity.

2 Nonlinear cointegration

There is now a well-established literature on the analysis of nonlinear transformation of stationary Gaussian time series by means of Hermite polynomials; see Taqqu (1975, 1979) and Dobrushin and Major (1979) and more recently Dittmann and Granger (2002) and Dalla, Giratis, and Hidalgo (2006).

These polynomials are defined through the formula:

$$H_j(z; \sigma^2) = (-1)^j \sigma^{2j} \exp\left(\frac{z^2}{2\sigma^2}\right) \frac{d^j}{dz^j} \exp\left(-\frac{z^2}{2\sigma^2}\right), \quad j = 1, 2, \ldots \quad .$$

It is well-known that, for any mean zero Gaussian random variables v and u, we have:

$$\mathbb{E}\left[H_p(u)H_q(v)\right] = \begin{cases} p!\left[\mathbb{E}(uv)\right]^p & \text{for } p = q \\ 0 & \text{for } p \neq q \end{cases} . \tag{3}$$

The index of the first non null coefficient b_k is termed *Hermite rank* of $g(.)$. Of course, y_t is non-Gaussian unless $g(\cdot)$ is linear. For our aims, the most important property of Hermite polynomials is their orthogonality. This property allows us to characterize in a simple way the dependence structure of a nonlinear transformation of a stationary Gaussian process that exhibit long range dependence. Let $z_t \sim I(d_z)$; in view of (1) and (3) it is easy to see that

$$\mathbb{E}[H_k(z_0)H_k(z_\tau)] = k!\gamma^k(\tau) \simeq C\tau^{k(2d_z-1)}, \quad \text{as } \tau \to \infty$$

so the sequence $H_{k+1}(z_t)$ is "less" dependent then $H_k(z_t)$. More precisely, if $z_t \sim I(d_z)$, then $H_k(z_t)$ can be viewed as a long memory series that is fractional integrated of order d_k ,

$$d_k := \left\{ k \left(d_z - \frac{1}{2} \right) + \frac{1}{2} \right\} \vee 0 \leq d_z \ . \tag{4}$$

The above equation follows straightforwardly from the equality $2d_k - 1 = k(2d_z - 1)$.

We state here more precisely our full set of assumptions.
ASSUMPTION A

1) $(x_t, \varepsilon_t)'$ are jointly Gaussian and long memory, that is, as $\tau \to \infty$

$$\begin{array}{ll} \gamma_{xx}(\tau) \simeq G_{xx}\tau^{2d_x-1}, & 0 < G_{xx} < \infty \\ \gamma_{\varepsilon\varepsilon}(\tau) \simeq G_{\varepsilon\varepsilon}\tau^{2d_\varepsilon-1}, & 0 < G_{\varepsilon\varepsilon} < \infty \\ \gamma_{x\varepsilon}(\tau) \simeq G_{x\varepsilon}\tau^{d_x+d_\varepsilon-1}, & |G_{x\varepsilon}| < \infty \end{array}$$

for $0 \leq d_\varepsilon, d_x < \frac{1}{2}$.

2) The following equation holds:

$$y_t = g(x_t) + e_t \ , \tag{5}$$

where for $t = 1, 2, ...$

$$\begin{array}{rl} g(x_t) = \sum_{k=k_0}^{K} a_k x_t^k = \sum_{k=k_0}^{K} b_k H_k(x_t) \ , & b_{k_0} \neq 0 \ , \\ e_t = \sum_{\widetilde{k}=\widetilde{k}_0}^{\widetilde{K}} \theta_{\widetilde{k}} \varepsilon_t^{\widetilde{k}} = \sum_{\widetilde{k}=\widetilde{k}_0}^{\widetilde{K}} \xi_{\widetilde{k}} H_{\widetilde{k}}(\varepsilon_t) \ , & \xi_{\widetilde{k}_0} \neq 0 \ , \end{array}$$

3) The parameters $K, \widetilde{k}_0$ are such that

$$K(2d_x - 1) > \left\{ -1 \vee \widetilde{k}_0(2d_\varepsilon - 1) \right\}.$$

Assumptions A1-A2 identify a polynomial cointegration model where the residual is a Gaussian subordinated process. By assumption A2, the cointegrating relation (2) can be rewritten as

$$y_t = \beta' H(x_t) + e_t, \quad \text{where} \quad H(x_t) = [H_1(x_t), ..., H_K(x_t)]' \ .$$

The possible correlation between x_t and e_t leads to the inconsistency of OLS and justifies the use of the spectral regression techniques. As it shall be apparent from the proofs, in our arguments a great simplification occurs for $G_{xe} = 0$. However, in our view, in general it seems difficult to assess a priori what sort of behavior will characterize the covariance between regressors and residual at long lags ($G_{xe} = 0$ would entail these covariances to be eventually zero). Therefore in this paper we focus on the general situation where the value of this parameter is left unconstrained, abiding in some sense to the traditional cointegration framework.

Assumption A3 ensures that $H_K(x_t)$ is still a long memory process, with stronger memory than e_t. This is also a necessary identification condition: there are no means to distinguish $H_k(x_t)$ and e_t if they are not orthogonal unless the former has stronger long range dependence. In this paper, we take k_0 and K to be known, whereas their estimation will be addressed in a different work. Note that to implement our estimates we need no a priori information on $\widetilde{k}_0, \widetilde{K}$, although the value of $\widetilde{k}_0(2d_\varepsilon - 1)$ does affect the rate of consistency of our estimators.

Let us now define:

$$f_{HH}(\lambda) = \begin{bmatrix} f_{11}(\lambda) & 0 & \cdots & \cdots \\ 0 & f_{22}(\lambda) & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \vdots & f_{KK}(\lambda) \end{bmatrix}, \; f_{He}(\lambda) = \begin{bmatrix} f_{1e}(\lambda) \\ f_{2e}(\lambda) \\ \vdots \\ f_{Ke}(\lambda) \end{bmatrix}$$

and let also, for $a, b = 1, 2, \ldots K$.

$$\begin{aligned} \gamma_{ab}(\tau) &= \mathbb{E}\left[H_a(x_t)H_b(x_{t+\tau})\right] = a!\delta_a^b \left\{\mathbb{E}\left(x_t x_{t+\tau}\right)\right\}^a, \\ \gamma_{ae}(\tau) &= \mathbb{E}\left[H_a(x_t)e_{t+\tau}\right] = \mathbb{E}\left[H_a(x_t)\sum_{\widetilde{k}=\widetilde{k}_0}^{\widetilde{K}} \xi_{\widetilde{k}} H_{\widetilde{k}}(\varepsilon_t)\right] \\ &= \begin{cases} a!\xi_a \left\{\mathbb{E}\left(x_t \varepsilon_{t+\tau}\right)\right\}^a \text{ for } a \leq \widetilde{K} \\ 0 \text{ , otherwise} \end{cases}. \end{aligned}$$

where δ_a^b represents the Kronecker delta function. Likewise

$$f_{az}(\lambda) = (2\pi)^{-1} \sum_{\tau=-\infty}^{\infty} \gamma_{az}(\tau) \mathrm{e}^{-i\lambda\tau}$$

where $z = a, y, e$.

The Weighted Covariance Estimator (WCE) (see Marinucci (2000)) of $\beta' = (\beta_1, \ldots, \beta_K)$ is defined as

$$\hat{\beta}_M = \hat{f}_{HH}(0)^{-1}\hat{f}_{Hy}(0) \ ,$$

whence

$$\hat{\beta}_M - \beta = \hat{f}_{HH}(0)^{-1}\hat{f}_{He}(0) \ ;$$

as usual, we assume $\hat{f}_{HH}(0)$ is non-singular, where

$$\hat{f}_{HH}(0) = \frac{1}{2\pi} \begin{bmatrix} \sum_{\tau=-M}^{M} k(\tau/M)c_{11}(\tau) & \cdots & \sum_{\tau=-M}^{M} k(\tau/M)c_{1K}(\tau) \\ \vdots & \ddots & \vdots \\ \sum_{\tau=-M}^{M} k(\tau/M)c_{K1}(\tau) & \cdots & \sum_{\tau=-M}^{M} k(\tau/M)c_{KK}(\tau) \end{bmatrix},$$

$$\hat{f}_{Hz}(0) = \frac{1}{2\pi} \begin{bmatrix} \sum_{\tau=-M}^{M} k(\tau/M)c_{1z}(\tau) \\ \vdots \\ \sum_{\tau=-M}^{M} k(\tau/M)c_{Kz}(\tau) \end{bmatrix}$$

and $k(\cdot)$ is a kernel function to be discussed later, M is a positive integer representing a bandwidth parameter,

$$\begin{aligned} c_{ab}(\tau) &= n^{-1} \sum_{t=1}^{n-\tau} H_a(x_t) H_b(x_{t+\tau}) \\ c_{az}(\tau) &= n^{-1} \sum_{t=1}^{n-\tau} H_a(x_t) z_{t+\tau} \end{aligned}$$

for $a, b = 1, 2, \ldots K$, $z = y, e.$ and $\tau \leq 0$. For $\tau < 0$, we have $c_{aw}(\tau) = c_{wa}(-|\tau|)$, $w = b, z$.

$\hat{\beta}_M$ can be interpreted as resulting from a continuously averaged lest square regression of the discrete Fourier transform (DFT) of y_t on the DFT of $H(x_t)$ around a band of frequencies degenerating to zero as n goes to infinity. It is thus a continuously averaged analogous of the NBLS estimator considered for the linear case by Robinson (1994), Robinson and Marinucci (2001); in both cases, the numerator and the denominator can be viewed as spectral density estimates at zero frequency.

While there is certainly scope to consider discretely averaged estimates in this framework, we stick to the continuous case because we believe a time-domain expression can be appealing for practitioners and in view of the greater transparency of the proofs. Also, an explicit allowance for a general kernel grants more flexibility in the analysis of real data.

The last two assumptions concern the kernel and the bandwidth condition.

Assumption B: The kernel $k(\cdot)$ is a real-valued, symmetric Lebesgue measurable function that, for $\upsilon \in \mathbb{R}$, satisfies

$$\int_{-1}^{1} k(\upsilon) d\upsilon = 1 \qquad 0 \leq k(\upsilon) \leq \infty, \quad k(\upsilon) = 0 \quad \text{for } |\upsilon| > 1.$$

Assumption C: Let $\eta = K \vee \widetilde{k}_0$; as $n \to \infty$,

$$\frac{1}{M} + \frac{M^{3 \vee (\eta - 2)}}{n} \to 0 \ .$$

Assumption B is common for spectral estimates, and it is satisfied by (normalized version of) truncated lag windows such as Bartlett, modified Bartlett, Parzen, and many others, see Brillinger (1981) for a review.

Assumption C imposes a minimal lower bound and a significant upper bound on the behaviour of the user-chosen bandwidth parameter M. The need for this bandwidth condition is made clear by inspection of the proof in the appendix; heuristically, as K grows the signal in $H_K(x_t)$ decreases, which makes the estimation harder; on the other hand an increase in $\widetilde{k}_0$ makes the convergence rates in Lemma 1 and Theorem 1 faster, whence the need for tighter bandwidth conditions. We are not claiming Assumption C is sharp, however an inspection of the Proof of Lemma 1 reveals that any improvement is likely to require at least almost unmanageable computations.

The following lemma is the main tool for our consistency result, compare Lemma 1 in Marinucci (2000). By (4) we write

$$d_a := a\left(d_x - \frac{1}{2}\right) + \frac{1}{2} \ , \ d_e = \left\{\widetilde{k}_0 \left(d_\varepsilon - \frac{1}{2}\right) + \frac{1}{2}\right\} \vee 0 \ ;$$

by Assumption A3 we have $d_a > 0$, $a = k_0, ..., K$.

LEMMA 1 Under Assumptions A-C, as $n \to \infty$ we have:

$$\sum_{\tau=-M}^{M} k\left(\frac{\tau}{M}\right)\{c_{ab}(\tau) - \gamma_{ab}(\tau)\} = o_p(M^{d_a + d_b}) \tag{6}$$

$$\sum_{\tau=-M}^{M} k\left(\frac{\tau}{M}\right)\{c_{ae}(\tau) - \gamma_{ae}(\tau)\} = o_p(M^{d_a + d_e}) \tag{7}$$

for $a, b = 1, 2, \ldots K$

Proof See Appendix

We are now ready to state the main result of this paper. Let

$$B_{ab} \ : \ = a! G_{xx}^a \delta_a^b \int_{-1}^{1} k(\upsilon)|\upsilon|^{a(2d_x - 1)} d\upsilon < \infty \ ,$$

$$B_{ae} \ : \ = a! \xi_a \left\{G_{x\varepsilon}\right\}^a \int_{-1}^{1} k(\upsilon)|\upsilon|^{a(d_x + d_\varepsilon - 1)} d\upsilon < \infty \text{ , for } a \leq \widetilde{K} \ ,$$

see also Assumption B, $a, b = k_0, ..., K$. Let

$$\mathcal{B}_{HH} = \operatorname{diag}\{B_{11}, \ldots B_{KK}\} \ , \mathcal{B}_{He} = \{B_{1e}, ..., B_{Ke}\} \ , \mathcal{M} = \operatorname{diag}\left\{M^{-d_1}, \ldots M^{-d_K}\right\} \ .$$

Note that $B_{ae} = 0$ unless $a \leq \widetilde{K}$, due to the orthogonality of Hermite polynomials.

Theorem 1 Under the Assumptions A-C, as $n \to \infty$

$$\begin{bmatrix} M^{d_1-d_e} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & M^{d_K-d_e} \end{bmatrix} \left(\hat{\beta}_M - \beta\right) = \mathcal{B}_{HH}^{-1}\mathcal{B}_{He} + o_p(1) \ .$$

Proof By the dominated convergence theorem, as $M \to \infty$

$$M^{-(d_a+d_b)} \sum_{\tau=-M}^{M} k\left(\frac{\tau}{M}\right)\gamma_{ab}(\tau) = \sum_{\tau=-M}^{M} k\left(\frac{\tau}{M}\right)\frac{\gamma_{ab}(\tau)}{M^{d_a+d_b-1}}\frac{1}{M} \to B_{ab}$$

$$M^{-(d_a+d_e)} \sum_{\tau=-M}^{M} k\left(\frac{\tau}{M}\right)\gamma_{ae}(\tau) = \sum_{\tau=-M}^{M} k\left(\frac{\tau}{M}\right)\frac{\gamma_{1e}(\tau)}{M^{d_a+d_e-1}}\frac{1}{M} \to B_{ae}$$

From Lemma 1, it follows easily that

$$\hat{f}_{HH}(0) = \begin{bmatrix} \zeta_1 + o_p(M^{2d_1}) & o_p(M^{d_1+d_2}) & \cdots & o_p(M^{d_1+d_p}) \\ o_p(M^{d_2+d_1}) & \zeta_2 + o_p(M^{2d_2}) & \cdots & o_p(M^{d_2+d_p}) \\ \vdots & \vdots & \ddots & \vdots \\ o_p(M^{d_K+d_1}) & \cdots & \cdots & \zeta_K + o_p(M^{2d_K}) \end{bmatrix}$$

where

$$\zeta_a := \frac{1}{2\pi}\sum_{\tau=-M}^{M} k(\frac{\tau}{M})\gamma_{aa}(\tau) \ .$$

Moreover

$$\mathcal{M}\hat{f}_{HH}(0)\mathcal{M} = \begin{bmatrix} B_{11} + o_p(1) & \cdots & o_p(1) \\ \vdots & \ddots & \vdots \\ o_p(1) & \cdots & B_{KK} + o_p(1) \end{bmatrix} \to \mathcal{B}_{HH} \ .$$

Therefore, for $M \to \infty$

$$\hat{f}_{HH}(0) = \mathcal{M}^{-1}\mathcal{B}_{HH}\mathcal{M}^{-1} + o_p(1) = \mathcal{B}_{HH}\mathcal{M}^{-2} + o_p(1) \ ,$$

since $\mathcal{B}_{HH}$ is diagonal and hence commutes with $\mathcal{M}^{-1}$. Using the same arguments, it follows easily that:

$$M^{-d_e}\mathcal{M}\hat{f}_{he}(0) \to \mathcal{B}_{He} \ , \text{as } n \to \infty \ .$$

Finally, as $n \to \infty$,

$$M^{-d_e}\mathcal{M}^{-1}\left\{\hat{\beta}_M - \beta\right\} = \left\{\mathcal{M}\hat{f}_{hh}(0)\mathcal{M}\right\}^{-1}\mathcal{M}M^{-d_e}\hat{f}_{he}(0) \to \mathcal{B}_{HH}^{-1}\mathcal{B}_{He} \ ,$$

which completes the proof of Theorem 1.

$\square$

Remark In Theorem 1 we have proved the consistency of the WCE estimator of the cointegrating vector, $\hat{\beta}_M \overset{p}{\to} \beta$. In a very loose sense, this result follows from consistency of a continuously averaged estimate of the spectral density at frequency zero, see Lemma 1. It is also possible to use Lemma 1 to derive a robust estimate for the memory parameter of an observed, Gaussian subordinated series $w_t := g(x_t)$, $(k_0(d_x - \frac{1}{2}) + \frac{1}{2} =: d_w$, say). We use a very similar idea to the averaged periodogram estimate advocated by Robinson (1994). More precisely, with an obvious notation we can consider

$$\begin{aligned} \widetilde{d}_w \quad : \quad &= \frac{\log\left|\sum_{\tau=-M}^{M} k(\frac{\tau}{M})c_{ww}(\tau)\right|}{2\log M} = d_w + \frac{\log B_{ww}}{2\log M} + o_p(1) \ , \\ &= \quad d_w + o_p(1) \ , \end{aligned}$$

where we have used Lemma 1. This estimate converges at a mere logarithmic rate and it is not asymptotically centered around zero; it is however consistent under broader circumstances than usually allowed for in the literature (see Velasco (2006) for a recent survey).

3 Comments and conclusions

We view this paper as a first step in a new research direction, and as such we are well aware that it leaves several questions unresolved and open for future research. A first issue relates to the choice of the Hermite rank k_0 and of K. As far as the former is concerned, we remark that for the great majority of practical applications, k_0 can be taken a priori as 1 or 2. Under the assumption that $k_0 = 1$, the equality $d_x = d_y$ holds; this trivial observation immediately suggests a naive test for $k_0 = 1$, which can be simply implemented by testing for equality of the two memory parameters. It should be noted, however, that when x_t and y_t are cointegrated the standard asymptotic results on multivariate long memory estimation (for instance Robinson (1995)) do not hold. Incidentally, we note that the nonlinear framework allows to cover the possibility of cointegration between time series with different integration orders, a significant extension over the standard paradigm.

For K, we can take as an identifying assumption

$$K := \text{argmax}(k : k(2d_x - 1) > (2d_e - 1)) \ ; \tag{8}$$

higher order terms can be thought of as included by definition in the residuals, to make identification possible. Indeed, it is natural to suggest to view $g(.)$ as a general nonlinear function and envisage K as growing with n; we expect, however, that only the projection coefficients b_k with k satisfying (8) could be consistently estimated in this broader framework. On the other hand, we note that the it is also possible to estimate consistently $K^* < K$ regression coefficients, by simply dropping the higher order regressors: it is immediate to see that their inclusion in the residual would not alter any of our asymptotic result (there may be an effect in finite samples, however). We stress that a lower number of regressors allows in general a weaker bandwidth condition, see Assumption C.

The extension to multivariate regressors does not seem to pose any new theoretical problem: multivariate generalizations of Hermite expansions are well known to the literature. The non-Gaussian case is more complicated to consider, even if some results using Appell polynomials are provided by Surgailis (2000).

Of course, much more challenging seems to be the possibility to allow for multiple cointegrating relationships. An important point to remark is the following. In standard cointegration theory, the role of the variables on the left and on on the right-hand sides is, by all means, symmetric: this is no longer the case when nonlinear relationships are allowed. In particular, it should be noted that the memory parameter of the dependent variable y_t is always smaller or equal than d_x; this information can be exploited in an obvious way to decide the form of the regression, provided that first step estimates of the long memory parameters are available. We also remark that our procedure requires a preliminary knowledge on the variance of the regressor x_t; such knowledge can clearly be derived from first step estimates, and we leave for future research the analysis of its consequences in finite samples.

In this paper, we restricted ourselves to consistency results, and gave no hint on asymptotic distributions. The latter are likely to be non-Gaussian, at least if the Hermite rank is larger than one and/or the memory of the raw series is such to make their autocovariances not square summable (see for instance Fox and Taqqu (1985, 1986). A much wider issue relates to the possible extension to nonstationary circumstances. Here, a major technical difficulty arises: the higher order terms in Hermite expansions need no longer be of smaller order in the presence of nonstationarity. We believe, however, that the stationary framework considered in this paper is of sufficient interest by itself for applications to real data, see again Christensen and Nielsen (2006) for examples on how fractional cointegration between stationary variables may be implied by some models of volatility, based on the Black-Scholes formula for option pricing.

Appendix

Proof of Lemma 1 Recall we have

$$\gamma_{ab}(\tau) \simeq G|\tau|^{d_a+d_b-1} \quad \text{as} \quad \tau \to \infty \ ,$$

where for $a, b = 1, ..., K$, d_a is such that

$$d_a := \begin{cases} \frac{a}{2}(2d_x - 1) + \frac{1}{2} & \text{for} \quad a(2d_x-1) > -1 \\ 0 & \text{for} \quad a(2d_x-1) < -1 \end{cases} . \tag{9}$$

The first part of the proof follows closely Marinucci (2000). For (6), it is sufficient to show that

$$\begin{aligned} Var\left\{\sum_{\tau=-M+1}^{M-1} k\left(\frac{\tau}{M}\right) c_{ab}(\tau)\right\} &= \mathbb{E}\left\{\sum_{\tau=-M+1}^{M-1} k\left(\frac{\tau}{M}\right)\left[c_{ab}(\tau) - \left(1-\frac{\tau}{n}\right)\gamma_{ab}(\tau)\right]\right\}^2 \\ &\leq C\sum_{p=-M}^{M}\sum_{q=-M}^{M}|Cov\{c_{ab}(p), c_{ab}(q)\}| = o(M^{2d_a+2d_b}) \end{aligned}$$

From Hannan (1970), p.210 we have:

$$\begin{aligned} & Cov\{c_{ab}(p), c_{ab}(q)\} \\ = \ & \frac{1}{n}\sum_{r=-n+1}^{n-1}\left(1-\frac{|r|}{n}\right)\{\gamma_{aa}(r)\gamma_{bb}(r+q-p) + \gamma_{ab}(r+q)\gamma_{ba}(r-p)\} \quad (10) \\ & + \frac{1}{n^2}\sum_{r=-n+1}^{n-1}\sum_{s=1-r}^{n-r} \mathrm{cum}_{abab}(s, s+p, s+r, s+r+q) \ , \quad (11) \end{aligned}$$

where

$$\mathrm{cum}_{abab}(s, s+p, s+r, s+r+q) = \mathrm{cum}\{H_a(x_s), H_b(x_{s+p}), H_a(x_{s+r}), H_b(x_{s+r+q})\} \ .$$

Likewise, for (7) we shall show that

$$\begin{aligned} & Var\left\{\sum_{\tau=-M+1}^{M-1} k\left(\frac{\tau}{M}\right) c_{ae}(p)\right\} \\ \leq \ & C\sum_{p=-M}^{M}\sum_{q=-M}^{M}|Cov\{c_{ae}(p), c_{ae}(q)\}| = o(M^{2d_a+2d_e}) \end{aligned} \tag{12}$$

For (10) we have

$$
\begin{aligned}
& \sum_{p=-M}^{M} \sum_{q=-M}^{M} \frac{1}{n} \left| \sum_{r=-n+1}^{n-1} \left(1 - \frac{|r|}{n}\right) \{\gamma_{aa}(r)\gamma_{bb}(r+q-p)\} \right| \\
\leq \;& C\frac{M}{n} \sum_{\tau=-2M}^{2M} \left(\sum_{|r|\leq 2M} (|r|+1)^{2d_a-1}(|r+\tau|+1)^{2d_b-1} + \right. \\
& \left. + \sum_{|r|>2M} (|r|+1)^{2d_a-1}(|r+\tau|+1)^{2d_b-1} \right) \\
= \;& C\frac{M}{n} \left[\sum_{|r|\leq 2M} \left((|r|+1)^{2d_a-1} \sum_{\tau=-2M}^{2M} (|r+\tau|+1)^{2d_b-1} \right) \right. \\
& \left. + \sum_{\tau=-2M}^{2M} \left(\sum_{2M<|r|<n} (|r|+1)^{2d_a-1}(|r+\tau|+1)^{2d_b-1} \right) \right] \\
= \;& O(Mn^{-1}M^{2d_a}M^{2d_b}) + O(M^2n^{-1}n^{2d_a+2d_b-1}) = o(M^{2d_a+2d_b}) \; .
\end{aligned}
$$

As usual, summations over empty sets are taken to be equal to zero. For the second term we have:

$$
\begin{aligned}
& \sum_{p=-M}^{M} \sum_{q=-M}^{M} \frac{1}{n} \left| \sum_{r=-n+1}^{n-1} \left(1 - \frac{|r|}{n}\right) \gamma_{ab}(r+p)\gamma_{ba}(r-q) \right| \\
\leq \;& C \sum_{p=-M}^{M} \sum_{q=-M}^{M} \frac{1}{n} \sum_{r=-n+1}^{n-1} \left(1 - \frac{|r|}{n}\right) \frac{1}{2} \left| \gamma_{ab}^2(r+p) + \gamma_{ba}^2(r-q) \right| \\
\leq \;& \frac{C}{n} \sum_{|r|\leq 2M} \left[\sum_{p=-M}^{M} (|r+p|+1)^{2d_a+2d_b-2} + \sum_{q=-M}^{M} (|r-q|+1)^{2d_a+2d_b-2} \right] \\
& + C\frac{M^2}{n} \sum_{2M<|r|<n} \left[(|r+p|+1)^{2d_a+2d_b-2} + (|r-q|+1)^{2d_a+2d_b-2} \right] \\
= \;& O(Mn^{-1}M^{2d_a+2d_b}) + O(M^2n^{-1}n^{2d_a+2d_b-1}) = o(M^{2d_a+2d_b}) \; .
\end{aligned}
$$

The argument for (11) is entirely analogous, but simpler, to that we are giving below for (12). More precisely, to bound (12) we need to focus on

$$
\begin{aligned}
& \mathrm{cum}_{aeae}\,(s, s+p, s+r, s+r+q) \\
= \;& \mathrm{cum}\,\{H_a(x_s), e_{s+p}, H_a(x_{s+r}), e_{s+r+q}\} \\
= \;& \sum_{k=\widetilde{k}_0}^{\widetilde{K}} \sum_{k'=\widetilde{k}_0}^{\widetilde{K}} \mathrm{cum}\,\{H_a(x_s), H_k(\varepsilon_{s+p})H_a(x_{s+r}), H_{k'}, (\varepsilon_{s+r+q})\} \, .
\end{aligned}
$$

The orders of magnitude of the cumulants are investigated by means of the diagram formula (see Arcones (1994), Surgailis (2003); some diagrams emerging from our arguments are represented in Figures 1 to 7). The proof is quite tedious. From the diagram formula it follows easily that, for any finite $k, k' \geq \widetilde{k}_0$

$$\sum_{k=\widetilde{k}_0}^{\widetilde{K}} \sum_{k'=\widetilde{k}_0}^{\widetilde{K}} \left|\mathrm{cum}\left\{H_a(x_s), H_k(\varepsilon_{s+p}), H_a(x_{s+r}), H_{k'}(\varepsilon_{s+r+q})\right\}\right|$$
$$\leq \; C\left|\mathrm{cum}\left\{H_a(x_s), H_{\widetilde{k}_0}(\varepsilon_{s+p}), H_a(x_{s+r}), H_{\widetilde{k}_0}(\varepsilon_{s+r+q})\right\}\right|. \tag{13}$$

Indeed, increasing the value of $\widetilde{k}_0$ to k, k' entails including more products of covariances in the cumulant, and these covariances are bounded. In order to simplify the presentation, we divide it in three parts, that is

1) $a = 1$, $\widetilde{k}_0 \geq 2$ or $a \geq 2$, $\widetilde{k}_0 = 1$
2) $a = 2$, $\widetilde{k}_0 \geq 2$ or $a \geq 2$, $\widetilde{k}_0 = 2$
3) $a, \widetilde{k}_0 \geq 3$.

Throughout the proof, we shall assume for brevity's sake $\widetilde{k}_0(2d_\varepsilon - 1) > -1$; it is simple to check that for $\widetilde{k}_0(2d_\varepsilon - 1) \leq -1$ the proof is analogous, indeed slightly simpler.

Part I: $a = 1$, $\widetilde{k}_0 \geq 2$ **or** $a \geq 2$, $\widetilde{k}_0 = 1$

For $a = 1$, $\widetilde{k}_0 = 2$ we have

$$\sum_{p=-M}^{M} \sum_{q=-M}^{M} \frac{1}{n^2} \left|\mathrm{cum}\left\{x_s, H_2(\varepsilon_{s+p}), x_{s+r}, H_2(\varepsilon_{s+r+q})\right\}\right|$$
$$\leq \sum_{p=-M}^{M} \sum_{q=-M}^{M} \frac{C}{n^2} \left| \sum_{r=-n+1}^{n-1} \sum_{s=1-r}^{n-r} \gamma_{x\varepsilon}(p)\gamma_{x\varepsilon}(q)\gamma_{\varepsilon\varepsilon}(r+q-p) \right.$$
$$\left. + \gamma_{\varepsilon\varepsilon}(r+q-p)\gamma_{\varepsilon x}(r-p)\gamma_{x\varepsilon}(r+q) \right|$$
$$\leq \frac{C}{n} \sum_{p=-M}^{M} (|p|+1)^{d_x+d_\varepsilon-1} \sum_{q=-M}^{M} (|q|+1)^{d_x+d_\varepsilon-1} \left(\sum_{|r|\leq 3M} (|r+q-p|+1)^{2d_\varepsilon-1} \right.$$
$$\left. + \sum_{3M<|r|\leq n} (|r+q-p|+1)^{2d_\varepsilon-1} \right)$$
$$+ \frac{C}{n} \sum_{p=-M}^{M} \sum_{q=-M}^{M} \left(\sum_{|r|\leq 3M} (|r+q-p|+1)^{2d_\varepsilon-1} (|r-p|+1)^{d_x+d_\varepsilon-1} (|r+q|+1)^{d_x+d_\varepsilon-1} \right.$$

$$
\begin{aligned}
& + \sum_{3M<|r|\le n} (|r+q-p|+1)^{2d_\varepsilon-1}(|r-p|+1)^{d_x+d_\varepsilon-1}(|r+q|+1)^{d_x+d_\varepsilon-1}\Bigg) \\
= \;& O(n^{-1}M^{d_x+d_\varepsilon}M^{d_x+d_\varepsilon}M^{2d_\varepsilon}) + O(n^{-1}M^{2d_x+2d_\varepsilon}n^{2d_\varepsilon}) \\
& +O(n^{-1}MM^{d_x+d_\varepsilon}M^{2d_\varepsilon}) + O(n^{-1}M^2 n^{2d_x+4d_\varepsilon-3}) \\
= \;& O\left(\frac{M}{n}M^{2d_x+4d_\varepsilon-1}\right) + o(M^{4d_\varepsilon+2d_x}) + O\left(\frac{M^2}{n}M^{d_x+3d_\varepsilon-1}\right) + o(M^{4d_\varepsilon+2d_x}\Big) \\
= \;& o(M^{2d_x+4d_\varepsilon-1}) = o(M^{2d_x+2d_e}) \quad \text{ because } 2d_e = 4d_\varepsilon - 1 \ .
\end{aligned}
$$

The extension to $\widetilde{k}_0 > 2$ is trivial:

$$
\begin{aligned}
& \operatorname{cum}\left\{x_s, H_{\widetilde{k}_0}(\varepsilon_{s+p}), x_{s+r}, H_{\widetilde{k}_0}(\varepsilon_{s+r+q})\right\} \\
= \;& \sum_{p=-M}^{M}\sum_{q=-M}^{M}\frac{C}{n^2}\left|\sum_{r=-n+1}^{n-1}\sum_{s=1-r}^{n-r}\gamma_{x\varepsilon}(p)\gamma_{x\varepsilon}(q)\gamma_{\varepsilon\varepsilon}^{\widetilde{k}_0-1}(r+q-p)\right. \\
& \left. +\gamma_{\varepsilon\varepsilon}^{\widetilde{k}_0-1}(r+q-p)\gamma_{\varepsilon x}(r-p)\gamma_{x\varepsilon}(r+q)\right| \\
= \;& O(n^{-1}M^{2d_x+2d_\varepsilon}M^{(\widetilde{k}_0-1)(2d_\varepsilon-1)+1}) + O(n^{-1}M^{2d_x+2d_\varepsilon}n^{(\widetilde{k}_0-1)(2d_\varepsilon-1)+1}) \\
& +O(n^{-1}M^{d_x+d_\varepsilon+1}M^{(\widetilde{k}_0-1)(2d_\varepsilon-1)+1}) + O(n^{-1}M^2 n^{2d_x+2d_\varepsilon-2+(\widetilde{k}_0-1)(2d_\varepsilon-1)+1}) \\
= \;& O\left(\frac{M}{n}M^{2d_x+2d_\varepsilon-1}M^{(\widetilde{k}_0-1)(2d_\varepsilon-1)+1}\right) + o(M^{2d_x+\widetilde{k}_0(2d_\varepsilon-1)+1}) \\
& +O\left(\frac{M^2}{n}M^{d_x+d_\varepsilon-1}M^{(\widetilde{k}_0-1)(2d_\varepsilon-1)+1}\right) + O(n^{-1}M^2 n^{2d_x+\widetilde{k}_0(2d_\varepsilon-1)}) \\
= \;& o(M^{2d_x+2d_e}),
\end{aligned}
$$

by the same argument as before. The proof for $a \ge 2$, $\widetilde{k}_0 = 1$ is entirely analogous and hence omitted.

Part II: $a = 2$, $\widetilde{k}_0 \ge 2$ **or** $a \ge 2$, $\widetilde{k}_0 = 2$

For $a = 2$, $\widetilde{k}_0 = 2$ we have

$$
\begin{aligned}
& \sum_{p=-M}^{M}\sum_{q=-M}^{M}\frac{1}{n^2}\left|\sum_{r=-n+1}^{n-1}\sum_{s=1-r}^{n-r}\operatorname{cum}\{H_2(x_s)H_2(\varepsilon_{s+p})H_2(x_{s+r})H_2(\varepsilon_{s+r+q})\}\right| \\
\le \;& \sum_{p=-M}^{M}\sum_{q=-M}^{M}\frac{C}{n^2}\left|\sum_{r=-n+1}^{n-1}\sum_{s=1-r}^{n-r}\gamma_{x\varepsilon}(p)\gamma_{x\varepsilon}(q)\gamma_{xx}(r)\gamma_{\varepsilon\varepsilon}(r+q-p)\right. \\
& +\gamma_{x\varepsilon}(p)\gamma_{x\varepsilon}(r+q)\gamma_{\varepsilon x}(r-p)\gamma_{x\varepsilon}(q) \\
& \left. +\gamma_{xx}(r)\gamma_{\varepsilon\varepsilon}(r+q-p)\gamma_{\varepsilon x}(r-p)\gamma_{x\varepsilon}(r+q)\right|
\end{aligned}
$$

$$\begin{aligned}
&\le \frac{C}{n}\Bigg\{\sum_{p=-M}^{M}\sum_{q=-M}^{M}\sum_{|r|\le 3M}\Big[(|p|+1)^{d_x+d_\varepsilon-1}(|q|+1)^{d_x+d_\varepsilon-1}(|r|+1)^{2d_x-1}(|r+q-p|+1)^{2d_\varepsilon-1}\\
&\quad+(|p|+1)^{d_x+d_\varepsilon-1}(|q|+1)^{d_x+d_\varepsilon-1}(|r+q|+1)^{d_x+d_\varepsilon-1}(|r-p|+1)^{d_x+d_\varepsilon-1}\\
&\quad+(|r|+1)^{2d_x-1}(|r+q-p|+1)^{2d_\varepsilon-1}(|r-p|+1)^{d_x+d_\varepsilon-1}(|r+q|+1)^{d_x+d_\varepsilon-1}\Big]\Bigg\}\\
&\quad+\frac{C}{n}\Bigg\{\sum_{p=-M}^{M}\sum_{q=-M}^{M}\sum_{3M<r\le n}\Big[(|p|+1)^{d_x+d_\varepsilon-1}(|q|+1)^{d_x+d_\varepsilon-1}(|r|+1)^{2d_x-1}(|r+q-p|+1)^{2d_\varepsilon-1}\\
&\quad+(|p|+1)^{d_x+d_\varepsilon-1}(|q|+1)^{d_x+d_\varepsilon-1}(|r+q|+1)^{d_x+d_\varepsilon-1}(|r-p|+1)^{d_x+d_\varepsilon-1}\\
&\quad+(|r|+1)^{2d_x-1}(|r+q-p|+1)^{2d_\varepsilon-1}(|r-p|+1)^{d_x+d_\varepsilon-1}(|r+q|+1)^{d_x+d_\varepsilon-1}\Big]\Bigg\}\\
&= O(n^{-1}M^{2d_x+2d_\varepsilon}M^{2d_\varepsilon})+O(n^{-1}M^{d_x+d_\varepsilon}M^{d_x+d_\varepsilon}M^{d_x+d_\varepsilon})+O(n^{-1}M^{3d_\varepsilon+3d_x})\\
&\quad+O(n^{-1}M^{2d_x+2d_\varepsilon}n^{2d_x+2d_\varepsilon-1})+O(n^{-1}M^{2d_x+2d_\varepsilon}n^{2d_x+2d_\varepsilon-1})+O(M^2n^{-1}n^{4d_x+4d\varepsilon-3})
\end{aligned}$$

$$\begin{aligned}
&= O\left(\frac{M^2}{n}M^{2d_x+4d_\varepsilon-2}\right)+O\left(\frac{M^2}{n}M^{2d_x+4d_\varepsilon-2}\right)+o(M^{4d_x+4d_\varepsilon-2})\\
&= o(M^{2d_2+2d_e}) \text{ because } 2d_2=4d_x-1 \text{ and } 2d_e=4d_\varepsilon-1\ .
\end{aligned}$$

For $\widetilde{k}_0>2$ the argument is very much the same:

$$\begin{aligned}
&\sum_{p=-M}^{M}\sum_{q=-M}^{M}\frac{1}{n^2}\left|\sum_{r=-n+1}^{n-1}\sum_{s=1-r}^{n-r}\mathrm{cum}\{H_2(x_s)H_{\widetilde{k}_0}(\varepsilon_{s+p})H_2(x_{s+r})H_{\widetilde{k}_0}(\varepsilon_{s+r+q})\}\right|\\
&\le \sum_{p=-M}^{M}\sum_{q=-M}^{M}\frac{C}{n^2}\Bigg|\sum_{r=-n+1}^{n-1}\sum_{s=1-r}^{n-r}\gamma_{x\varepsilon}(p)\gamma_{x\varepsilon}(q)\gamma_{xx}(r)\gamma_{\varepsilon\varepsilon}^{\widetilde{k}_0-1}(r+q-p)\\
&\quad+\gamma_{x\varepsilon}(p)\gamma_{x\varepsilon}(q)\gamma_{x\varepsilon}(r+q)\gamma_{\varepsilon x}(r-p)\gamma_{\varepsilon\varepsilon}^{\widetilde{k}_0-2}(r+q-p)\\
&\quad+\gamma_{xx}(r)\gamma_{\varepsilon x}(r-p)\gamma_{x\varepsilon}(r+q)\gamma_{\varepsilon\varepsilon}^{\widetilde{k}_0-1}(r+q-p)\Bigg|\\
&= O(n^{-1}M^{2d_x+2d_\varepsilon}M^{(\widetilde{k}_0-1)(2d_\varepsilon-1)+1})+O(n^{-1}M^{2d_x+2d_\varepsilon}M^{(\widetilde{k}_0-2)(2d_\varepsilon-1)+1})\\
&\quad+O(n^{-1}M^{2d_x}M^{d_x+d_\varepsilon}M^{(\widetilde{k}_0-1)(2d_\varepsilon-1)+1})+O(n^{-1}M^{2d_x+2d_\varepsilon}n^{2d_x-1+(\widetilde{k}_0-1)(2d_\varepsilon-1)+1})\\
&\quad+O(n^{-1}M^{2d_x+2d_\varepsilon}n^{2d_x+2d_\varepsilon-1+(\widetilde{k}_0-2)(2d_\varepsilon-1)})+O(M^2n^{-1}n^{4d_x+2d_\varepsilon-2+(\widetilde{k}_0-1)(2d_\varepsilon-1)})\\
&= O\left(\frac{M^2}{n}M^{2d_x-1}M^{\widetilde{k}_0(2d_\varepsilon-1)+1}\right)+o\left(\frac{M^2}{n}M^{2d_x+2d_\varepsilon-1+\widetilde{k}_0(2d_\varepsilon-1)+1}\right)\\
&\quad+O\left(\frac{M^2}{n}M^{3d_x-1-d_\varepsilon}M^{\widetilde{k}_0(2d_\varepsilon-1)+1}\right)+O\left(\frac{M^2}{n}M^{4d_x-1}M^{\widetilde{k}_0(2d_\varepsilon-1)+1}\right)\\
&= o(M^{2d_2+2d_e})\ \text{, because } 2d_e=\widetilde{k}_0(2d_\varepsilon-1)+1\ .
\end{aligned}$$

Part III: $a \geq 3,\ \widetilde{k}_0 \geq 3$

We note that, by the diagram formula (as in (13))

$$\begin{aligned} & \left|\mathrm{cum}\left[H_a(x_s)H_{\widetilde{k}_0}(\varepsilon_{s+p})H_a(x_{s+r})H_{\widetilde{k}_0}(\varepsilon_{s+r+q})\right]\right| \\ \leq\ & C\left|\mathrm{cum}\left[H_3(x_s)H_3(\varepsilon_{s+p})H_3(x_{s+r})H_3(\varepsilon_{s+r+q})\right]\right| \ . \end{aligned}$$

It suffices then to focus on $a = \widetilde{k}_0 = 3$. There are seven different kinds of connected diagrams, which are represented in Figures 1 to 7. We have

$$\begin{aligned} & \sum_{p=-M}^{M}\sum_{q=-M}^{M}\frac{1}{n^2}\left|\sum_{r=-n+1}^{n-1}\sum_{s=1-r}^{n-r}\mathrm{cum}\{H_3(x_s)H_3(\varepsilon_{s+p})H_3(x_{s+r})H_3(\varepsilon_{s+r+q})\}\right| \\ = & \sum_{p=-M}^{M}\sum_{q=-M}^{M}\frac{C}{n^2}\Big|\sum_{r=-n+1}^{n-1}\sum_{s=1-r}^{n-r}\gamma_{x\varepsilon}^2(p)\gamma_{x\varepsilon}^2(q)\gamma_{\varepsilon x}(r-p)\gamma_{x\varepsilon}(r+q)+ \\ & \quad + \ \gamma_{x\varepsilon}(p)\gamma_{xx}(r)\gamma_{x\varepsilon}(r+q)\gamma_{\varepsilon\varepsilon}(r+q-p)\gamma_{\varepsilon x}(r-p)\gamma_{x\varepsilon}(q) \\ & \quad + \ \gamma_{xx}^2(r)\gamma_{\varepsilon\varepsilon}^2(r+q-p)\gamma_{\varepsilon x}(r-p)\gamma_{x\varepsilon}(r+q) \\ & \quad + \ \gamma_{x\varepsilon}^2(p)\gamma_{x\varepsilon}^2(q)\gamma_{xx}(r)\gamma_{\varepsilon\varepsilon}(r+q-p) \\ & \quad + \ \gamma_{xx}^2(r)\gamma_{\varepsilon\varepsilon}^2(r+q-p)\gamma_{x\varepsilon}(p)\gamma_{x\varepsilon}(q) \\ & \quad + \ \gamma_{x\varepsilon}^2(r-p)\gamma_{x\varepsilon}^2(r+q)\gamma_{x\varepsilon}(p)\gamma_{x\varepsilon}(q) \\ & \quad + \ \gamma_{x\varepsilon}^2(r-p)\gamma_{x\varepsilon}^2(r+q)\gamma_{xx}(r)\gamma_{\varepsilon\varepsilon}(r+q-p)\Big| \end{aligned}$$

$$\begin{aligned} \leq\ & \frac{C}{n}\sum_{p=-M}^{M}(|p|+1)^{2(d_x+d_\varepsilon-1)}\sum_{q=-M}^{M}(|q|+1)^{2(d_x+d_\varepsilon-1)} \\ \times\ & \Bigg[\sum_{|r|\leq 2M}(|r-p|+1)^{d_x+d_\varepsilon-1}(|r+q|+1)^{d_x+d_\varepsilon-1} \\ +\ & \sum_{2M<|r|\leq n}(|r-p|+1)^{d_x+d_\varepsilon-1}(|r+q|+1)^{d_x+d_\varepsilon-1}\Bigg] \end{aligned} \tag{14}$$

$$\begin{aligned} +\ & \frac{C}{n}\sum_{p=-M}^{M}(|p|+1)^{d_x+d_\varepsilon-1}\sum_{q=-M}^{M}(|q|+1)^{d_x+d_\varepsilon-1} \\ & \times\Bigg[\sum_{|r|\leq 3M}(|r|+1)^{2d_x-1}(|r+q-p|+1)^{2d_\varepsilon-1}(|r-p|+1)^{d_x+d_\varepsilon-1}(|r+q|+1)^{d_x+d_\varepsilon-1} \\ +\ & \sum_{3M<|r|\leq n}(|r|+1)^{2d_x-1}(|r+q-p|+1)^{2d_\varepsilon-1}(|r-p|+1)^{d_x+d_\varepsilon-1}(|r+q|+1)^{d_x+d_\varepsilon-1}\Bigg] \end{aligned} \tag{15}$$

$$
\begin{aligned}
+\quad & \frac{C}{n}\left[\left(\sum_{|r|\leq 3M}(|r|+1)^{2(2d_x-1)}\sum_{p=-M}^{M}\sum_{q=-M}^{M}(|r+q-p|+1)^{2(2d_\varepsilon-1)}\right.\right.\\
& \times(|r-p|+1)^{d_x+d_\varepsilon-1}(|r+q|+1)^{d_x+d_\varepsilon-1}\Big)+\sum_{p=-M}^{M}\sum_{q=-M}^{M}\left(\sum_{3M<|r|\leq n}(|r|+1)^{2(2d_x-1)}\right.\\
& \left.\left.\times\ |r+q-p|+1)^{2(2d_\varepsilon-1)}(|r-p|+1)^{d_x+d_\varepsilon-1}(|r+q|+1)^{d_x+d_\varepsilon-1}\right)\right] \qquad (16)
\end{aligned}
$$

$$
\begin{aligned}
+\quad & \frac{C}{n}\sum_{p=-M}^{M}(|p|+1)^{2(d_x+d_\varepsilon-1)}\sum_{q=-M}^{M}(|q|+1)^{2(d_x+d_\varepsilon-1)}\left[\left(\sum_{|r|\leq 3M}(|r|+1)^{2d_x-1}\right.\right.\\
& \left.\left.\times(|r+q-p|+1)^{2d_x-1}\right)+\sum_{3M<|r|\leq n}(|r|+1)^{2d_x-1}(|r+q-p|+1)^{2d_x-1}\right] \qquad (17)
\end{aligned}
$$

$$
\begin{aligned}
+\quad & \frac{C}{n}\sum_{p=-M}^{M}(|p|+1)^{d_x+d_\varepsilon-1}\sum_{q=-M}^{M}(|q|+1)^{d_x+d_\varepsilon-1}\left[\left(\sum_{|r|\leq 3M}(|r|+1)^{2(2d_x-1)}\right.\right.\\
& \left.\left.\times(|r+q-p|+1)^{2(2d_\varepsilon-1)}\right)+\sum_{3M<|r|\leq n}(|r|+1)^{2(2d_x-1)}(|r+q-p|+1)^{2(2d_\varepsilon-1)}\right] \quad (18)
\end{aligned}
$$

$$
\begin{aligned}
+\quad & \frac{C}{n}\sum_{p=-M}^{M}(|p|+1)^{d_x+d_\varepsilon-1}\sum_{q=-M}^{M}(|q|+1)^{d_x+d_\varepsilon-1}\left[\left(\sum_{|r|\leq 3M}(|r+q|+1)^{2(d_x+d_\varepsilon-1)}\right.\right.\\
& \left.\left.\times(|r-p|+1)^{2(d_x+d_\varepsilon-1)}\right)+\sum_{3M<|r|\leq n}(|r+q|+1)^{2(d_x+d_\varepsilon-1)}(|r-p|+1)^{2(d_x+d_\varepsilon-1)}\right] (19)
\end{aligned}
$$

$$
\begin{aligned}
+\quad & \frac{C}{n}\sum_{|r|\leq 3M}(|r|+1)^{2d_x-1}\sum_{p=-M}^{M}\sum_{q=-M}^{M}(|r+p-q|+1)^{2d_\varepsilon-1}\\
& \times(|r-p|+1)^{2(d_x+d_\varepsilon-1)}(|r+q|+1)^{2(d_x+d_\varepsilon-1)}+\sum_{3M<|r|\leq n}(|r|+1)^{2d_x-1}\\
& \times\sum_{p=-M}^{M}\sum_{q=-M}^{M}(|r+p-q|+1)^{2d_\varepsilon-1}(|r-p|+1)^{2(d_x+d_\varepsilon-1)}(|r+q|+1)^{2(d_x+d_\varepsilon-1)} (20)
\end{aligned}
$$

After lengthy but straightforward computations, it is not difficult to see that

$$\begin{aligned}
(14) &= O(n^{-1}M^{2d_x+2d_\varepsilon-1}M^{2d_x+2d_\varepsilon-1}M^{d_x+d_\varepsilon}) + O(n^{-1}M^{4d_x-1}M^{4d_\varepsilon-1}n^{2d_x+2d_\varepsilon-1}) \\
&= o\left(\frac{M^{4d_x+4d_\varepsilon-1}}{n}\right) + O(M^{4d_x-1}M^{4d_\varepsilon-1}n^{2d_x+2d_\varepsilon-2}) \\
(15) &= O(n^{-1}M^{d_x+d_\varepsilon}M^{d_x+d_\varepsilon}M^{2d_x+2d_\varepsilon-1}) + O(n^{-1}M^{d_x+d_\varepsilon}M^{d_x+d_\varepsilon}n^{4d_x+4d_\varepsilon-3}) \\
&= O\left(\frac{M^{4d_x+4d_\varepsilon-1}}{n}\right) + o(M^{4d_x-1}M^{4d_\varepsilon-1}n^{2d_x+2d_\varepsilon-2}) \\
(16) &= O(n^{-1}M^{4d_x-1}M^{4d_\varepsilon}) + O(n^{-1}M^{2}n^{6d_x+6d_\varepsilon-5}) \\
&= O\left(\frac{M^{4d_x+4d_\varepsilon-1}}{n}\right) + o(M^{4d_x-1}M^{4d_\varepsilon-1}n^{2d_x+2d_\varepsilon-2}) \\
(17) &= O(n^{-1}M^{2d_x+2d_\varepsilon-1}M^{2d_x+2d_\varepsilon-1}M^{2d_x}) + O(n^{-1}M^{4d_x-1}M^{4d_\varepsilon-1}n^{2d_x+2d_\varepsilon-1}) \\
&= o\left(\frac{M^{4d_x+4d_\varepsilon-1}}{n}\right) + O(M^{4d_x-1}M^{4d_\varepsilon-1}n^{2d_x+2d_\varepsilon-2}) \\
(18) &= O(n^{-1}M^{d_x+d_\varepsilon}M^{d_x+d_\varepsilon}M^{4d_\varepsilon-1}) + O(n^{-1}M^{2d_x}M^{2d_x}n^{4d_x+4d_\varepsilon-3}) \\
&= O\left(\frac{M^{4d_x+4d_\varepsilon-1}}{n}\right) + o(M^{4d_x-1}M^{4d_\varepsilon-1}n^{2d_x+2d_\varepsilon-2})
\end{aligned}$$

$$\begin{aligned}
(19) &= O(n^{-1}M^{d_x+d_\varepsilon}M^{d_x+d_\varepsilon}M^{2d_x+2d_\varepsilon-1}) + O(n^{-1}M^{2d_x+2d_\varepsilon}n^{4d_x+4d_\varepsilon-3}) \\
&= O\left(\frac{M^{4d_x+4d_\varepsilon-1}}{n}\right) + o(M^{4d_x-1}M^{4d_\varepsilon-1}n^{2d_x+2d_\varepsilon-2}) \\
(20) &= O(n^{-1}M^{2d_x}M^{2d_x+2d_\varepsilon}) + O(n^{-1}M^{2}n^{6d_x+6d_\varepsilon-5}) \\
&= o(\frac{M^{4d_x+4d_\varepsilon-1}}{n}) + o(M^{4d_x-1}M^{4d_\varepsilon-1}n^{2d_x+2d_\varepsilon-2}) \\
&= O(\frac{M^{4d_x+4d_\varepsilon-1}}{n}) + o(M^{4d_x-1}M^{4d_\varepsilon-1}n^{2d_x+2d_\varepsilon-2}) \ .
\end{aligned}$$

In view of the previous results, our proof will be completed if we show that

$$\frac{M^{4d_x+4d_\varepsilon-1}}{n} + M^{4d_x-1}M^{4d_\varepsilon-1}n^{2d_x+2d_\varepsilon-2} = o(M^{2d_a+2d_e}) \ ,$$

where

$$2d_a + 2d_e = 2ad_x + 2\widetilde{k}_0 d_\varepsilon - (a + \widetilde{k}_0) + 2 \ .$$

We note first that

$$\frac{M^{4d_x+4d_\varepsilon-1}}{nM^{2d_a+2d_e}} = \frac{M^{4d_x+4d_\varepsilon-1}}{n(M^{2ad_x+2\widetilde{k}_0 d_\varepsilon-(a+\widetilde{k}_0)+2})} = \frac{M^{2(2-a)d_x+2(2-\widetilde{k}_0)d_\varepsilon-3+a+\widetilde{k}_0}}{n} \ .$$

From (9) and $d_e > 0$ it follows that

$$d_x > \frac{1}{2} - \frac{1}{2a} \quad \text{and} \quad d_\varepsilon > \frac{1}{2} - \frac{1}{2\widetilde{k}_0} \ ,$$

whence, because $a, \widetilde{k}_0 \geq 2$ we have

$$
\begin{aligned}
\frac{M^{2(2-a)d_x+2(2-\widetilde{k}_0)d_\varepsilon-3+a+\widetilde{k}_0}}{n} &\leq \frac{M^{2(2-a)(\frac{1}{2}-\frac{1}{2a})+2(2-\widetilde{k}_0)(\frac{1}{2}-\frac{1}{2\widetilde{k}_0})-3+a+\widetilde{k}_0}}{n} \\
&= o\left(\frac{M^3}{n}\right) = o(1) \ ,
\end{aligned}
$$

in view of Assumption C. To complete the proof, note that, again from Assumption C, for some $\alpha > a-2, \widetilde{k}_0 - 2$ we have $M^\alpha = O(n)$, whence

$$
\frac{M^{4d_x+4d_\varepsilon-2}n^{2d_x+2d_\varepsilon-2}}{M^{2ad_x+2\widetilde{k}_0 d_\varepsilon-(a+\widetilde{k}_0)+2}} = o(M^{(2-a+\alpha)(2d_x-1)+(2-\widetilde{k}_0+\alpha)(2d_\varepsilon-1)}) = o(1) \qquad \text{as} \quad n \to \infty \ .
$$

Thus the proof is completed. $\square$

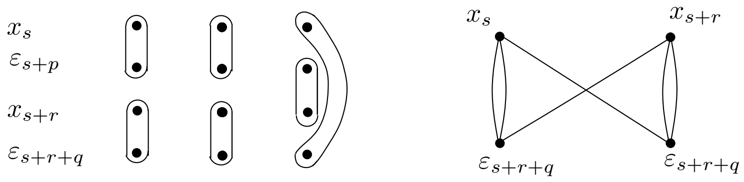

Figure 1: $\gamma_{x\varepsilon}^2(p)\gamma_{x\varepsilon}^2(q)\gamma_{x\varepsilon}(r+q)\gamma_{\varepsilon x}(r-p)$

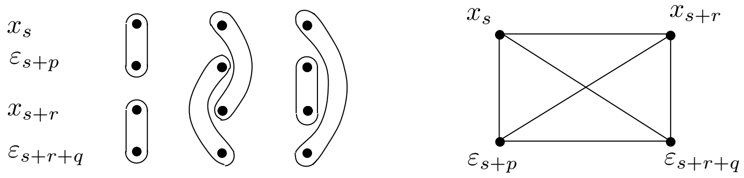

Figure 2: $\gamma_{x\varepsilon}(r+q)\gamma_{\varepsilon x}(r-p)\gamma_{xx}^2(r)\gamma_{\varepsilon\varepsilon}^2(r+q-p)$

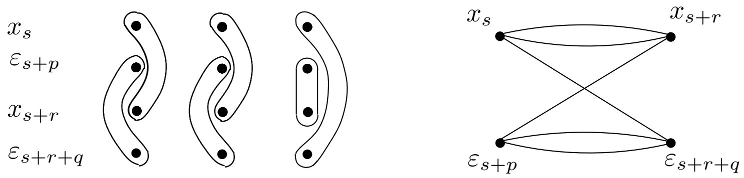

Figure 3: $\gamma_{x\varepsilon}(p)\gamma_{x\varepsilon}(q)\gamma_{x\varepsilon}^2(r+q)\gamma_{\varepsilon x}^2(r-p)\gamma_{xx}(r)\gamma_{\varepsilon\varepsilon}(r+q-p)$

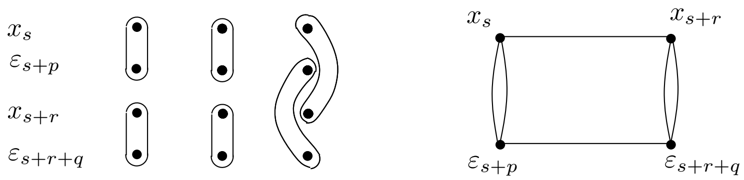

Figure 4: $\gamma_{x\varepsilon}^2(p)\gamma_{x\varepsilon}^2(q)\gamma_{xx}(r)\gamma_{\varepsilon\varepsilon}(r+q-p)$

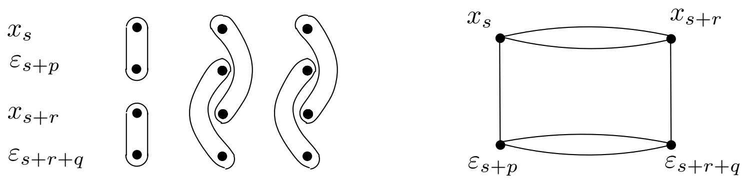

Figure 5: $\gamma_{x\varepsilon}^2(p)\gamma_{x\varepsilon}^2(q)\gamma_{xx}^2(r)\gamma_{\varepsilon\varepsilon}^2(r+q-p)$

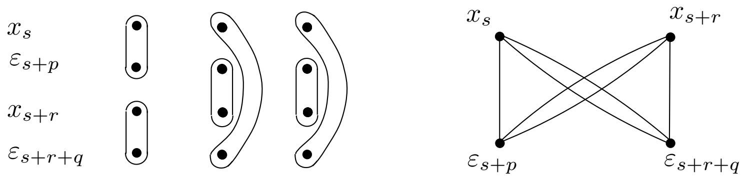

Figure 6: $\gamma_{x\varepsilon}(p)\gamma_{x\varepsilon}(q)\gamma^2_{x\varepsilon}(r+q)\gamma^2_{\varepsilon x}(r-p)$

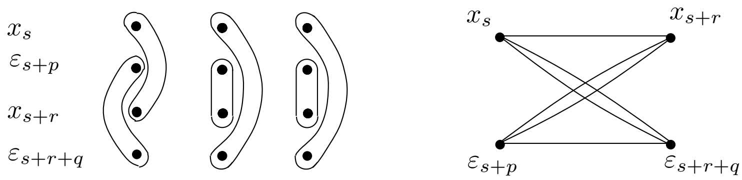

Figure 7: $\gamma_{xx}(r)\gamma_{\varepsilon\varepsilon}(r+q-p)\gamma^2_{x\varepsilon}(r+q)\gamma^2_{\varepsilon x}(r-p)$

References

ARCONES, M. A. (1994): "Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors," *Annals of Probability*, 22(4), 2242–2274.

BRILLINGER, D. R. (1981): *Time series: data analysis and theory.* Springer, New York.

CHEN, W. W., AND C. M. HURVICH (2003a): "Estimating fractional cointegration in the presence of polynomial trends," *Journal of Econometrics*, 117(1), 95–121.

——— (2003b): "Semiparametric estimation of multivariate fractional cointegration," *Journal of American Statistical Association*, 98(463), 629–642.

CHRISTENSEN, B. J., AND M. NIELSEN (2006): "Asymptotic normality of narrow-band least square in the stationary fractional cointegration model and volatility forecasting," *Journal of Econometrics*, 133, 343–371.

DALLA, V., L. GIRATIS, AND J. HIDALGO (2006): "Consistent estimation of the memory parameter for nonlinear time series," *Journal of Time Series Analysis*, 27(2), 211–251.

DITTMANN, I., AND C. W. J. GRANGER (2002): "Properties of nonlinear transformations of fractionally integrated processes," *Journal of Econometrics*, 110(2), 113–13.

Dobrushin, R. L., and P. Major (1979): "Non central limit theorems for non-linear functionals of Gaussian fields," *Zeitschrift fur Wahrscheinlichkeitstheorie verwandte Gebiete*, 50, 27–52.

Fox, R., and M. S. Taqqu (1985): "Non-central limit theorems for random variable in quadratic forms having long-range dependence," *Annals of Probability*, 13(2), 428–436.

——— (1986): "Large sample properties of parameter estimates for strongly dependent stationary gaussian time series," *Annals of Statistics*, 14(2), 517–532.

Giraitis, L., and D. Surgailis (1985): "CLT and other limit theorems for functionals of Gaussian processes," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 70(2), 191–212.

Hannan, E. J. (1970): *Multiple time series*. Wiley, New York.

Karlsen, K. A., T. Myklebust, and D. Tjostheim (2006): "Nonparametric estimation in a nonlinear cointegration type model," *The Annals of Statistics*, forthcoming.

Marinucci, D. (2000): "Spectral regression for cointegrated time series with long-memory innovations," *Journal of Time Series Analysis*, 21(6), 685–705.

Marinucci, D., and P. M. Robinson (2001): "Semiparametric fractional cointegration analysis," *Journal of Econometrics*, 105(1), 225–247.

Robinson, P. M. (1994): "Semiparametric analysis of long-memory time series," *The Annals of Statistics*, 22(1), 515 – 539.

——— (1995): "Log-periodogram regression of time series with time series with long range dependence," *The Annals of Statistics*, 23(3), 1048–1072.

Robinson, P. M., and J. Hualde (2003): "Cointegration in fractional systems with unknown integration orders," *Econometrica*, 71(6), 1727–1766.

Robinson, P. M., and D. Marinucci (2001): "Narrow-band analysis of nonstationary process," *The Annals of Statistics*, 29(4), 947 – 986.

——— (2003): "Semiparametric frequency domain analysis of fractional cointegration," in *Time series with long memory*, ed. by P. Robinson, pp. 334–373. Oxford University Press.

Surgailis, D. (2000): "Long range dependence and Appell rank," *Annals of Probability*, 28(1), 478–479.

——— (2003): "CLTs for polynomials of linear sequences: diagram formula with illustration," in *Theory and applications of long-range dependence*, ed. by P. Doukan, G. Oppenheim, and M. S. Taqqu, pp. 111–128. Birkäuser.

Taqqu, M. S. (1975): "Weak convergence to fractional Brownian motion and to the Rosenblatt process," *Zeitschrift fur Wahrscheinlichkeitstheorie verwandte Gebiete*, 31, 287–302.

——— (1979): "Convergence of integrated processes of arbitrary Hermite rank," *Zeitschrift fur Wahrscheinlichkeitstheorie verwandte Gebiete*, 50, 53–83.

Velasco, C. (2003): "Gaussian semi-parametric estimation of fractional cointegration," *Journal of Time Series Analysis*, 24(3), 345–378.

——— (2006): "Semiparametric estimation of long-memory models," in *Palgrave Book Hanbook of Econometrics, Vol. 1 Econometric Theory*, ed. by K. Patterson, and T. C. Mills. Palgrave, MacMillian.

Corresponding author:
Domenico Marinucci
Dipartimento di Matematica
Universita' di Roma Tor Vergata
via della Ricerca Scientifica, 1
00133 Roma, Italy
email: marinucc@mat.uniroma2.it