Journal of Multivariate Analysis 105 (2012) 222–240

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

An adaptive estimator of the memory parameter and the goodness-of-fit test using a multidimensional increment ratio statistic

Jean-Marc Bardet *, Béchir Dola

SAMM, Université Panthéon-Sorbonne (Paris I), 90 rue de Tolbiac, 75013 Paris, France

ARTICLE INFO

Article history:
Received 1 October 2010
Available online 29 September 2011

AMS subject classifications:
62M07
62M09

Keywords:
Long-memory Gaussian processes
Goodness-of-fit test
Estimation of the memory parameter
Minimax adaptive estimator

ABSTRACT

The increment ratio (IR) statistic was first defined and studied in Surgailis et al. (2007) [19] for estimating the memory parameter either of a stationary or an increment stationary Gaussian process. Here three extensions are proposed in the case of stationary processes. First, a multidimensional central limit theorem is established for a vector composed by several IR statistics. Second, a goodness-of-fit χ^2-type test can be deduced from this theorem. Finally, this theorem allows to construct adaptive versions of the estimator and the test which are studied in a general semiparametric frame. The adaptive estimator of the long-memory parameter is proved to follow an oracle property. Simulations attest to the interesting accuracies and robustness of the estimator and the test, even in the non Gaussian case.

1. Introduction

After almost thirty years of intensive and numerous studies, the long-memory processes now form an important topic of the time series study (see for instance the book edited by Doukhan et al. [7]). The most famous long-memory stationary time series are the fractional Gaussian noises (fGn) with Hurst parameter H and FARIMA(p, d, q) processes. For both these time series, the spectral density f in 0 follows a power law: $f(\lambda) \sim C\,\lambda^{-2d}$ where $H = d + 1/2$ in the case of the fGn. In the case of a long memory process $d \in (0, 1/2)$ but a natural expansion to $d \in (-1/2, 0]$ (short memory) implied that d can be considered more generally as a memory parameter.

There are a lot of statistical results relative to the estimation of this memory parameter d. First and main results in this direction have been obtained for parametric models with the essential articles of [8,6] for Gaussian time series, [11] for linear processes and [12] for non linear functions of Gaussian processes.

However, parametric estimators are not really robust and can induce no consistent estimations. Thus, the research is now rather focused on semiparametric estimators of the memory parameter. Different approaches were considered: the famous R/S statistic (see [13]), the log-periodogram estimator (studied first by Geweke and Porter-Hudack [9], notably improved by Robinson [17], and Moulines and Soulier [16]), the local Whittle estimator (see [18]) or the wavelet based estimator (see [21,15] or [2]). All these estimators require the choice of an auxiliary parameter (frequency bandwidth, scales, etc.) but adaptive versions of these estimators are generally built for avoiding this choice. In a general semiparametric frame, [10] obtained the asymptotic lower bound for the minimax risk in the estimation of d, expressed as a function of the second order parameter of the spectral density expansion around 0. Several adaptive semiparametric estimators are proved to follow an oracle property up to multiplicative logarithm term. But simulations (see for instance [3] or [2]) show that the most accurate estimators are local Whittle, global log-periodogram and wavelet based estimators.

* Corresponding author.
E-mail addresses: bardet@univ-paris1.fr, Jean-Marc.Bardet@univ-paris1.fr (J.-M. Bardet), bechir.dola@malix.univ-paris1.fr (B. Dola).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2011.09.003

In this paper, we consider the IR (Increment Ratio) estimator of a long-memory parameter (see its definition in the next section) for Gaussian time series recently introduced in [19] and we propose three extensions. First, a multivariate central limit theorem is established for a vector of IR statistics with different "windows" (see Section 2) and this induces to consider a pseudo-generalized least squares estimator of the parameter d. Second, this multivariate result allows us to define an adaptive estimator of the memory parameter d based on IR statistics: an "optimal" window is automatically computed (see Section 3). This notably improves the results of [19] in which the choice of m is either theoretical (and cannot be applied to data) or guided by empirical rules without justifications. Third, an adaptive goodness-of-fit test is deduced and its convergence to a chi-square distribution is established (see Section 3).

In Section 4, several Monte Carlo simulations are realized for optimizing the adaptive estimator and exhibiting the theoretical results. Then some numerical comparisons are made with the 3 semiparametric estimators previously mentioned (local Whittle, global log-periodogram and wavelet based estimators) and the results are even better than the theory seems to indicate: as well in terms of convergence rate than in terms of robustness (notably in case of trend or seasonal component), the adaptive IR estimator and goodness-of-fit test provide efficient results. Finally, all the proofs are grouped in Section 5.

2. The multidimensional increment ratio statistic and its statistical applications

Let $X=(X_k)_{k\in\mathbb{N}}$ be a Gaussian time series satisfying the following Assumption $S(d,\beta)$:

Assumption $S(d,\beta)$**.** There exist $\varepsilon>0$, $c_0>0$, $c_0'>0$ and $c_1\in\mathbb{R}$ such that $X=(X_t)_{t\in\mathbb{Z}}$ is a stationary Gaussian time series having a spectral density f satisfying for all $\lambda\in(-\pi,0)\cup(0,\pi)$

$$f(\lambda)=c_0|\lambda|^{-2d}+c_1|\lambda|^{-2d+\beta}+O\big(|\lambda|^{-2d+\beta+\varepsilon}\big)\quad\text{and}\quad|f'(\lambda)|\le c_0'\,\lambda^{-2d-1}. \tag{2.1}$$

Remark 1. Note that here we only consider the case of stationary processes. However, as it was already done in [19], it could be possible, *mutatis mutandis*, to extend our results to the case of processes having stationary increments.

Let $(X_1,\ldots,X_N)$ be a path of X. For $m\in\mathbb{N}^*$, define the random variable $IR_N(m)$ such as

$$IR_N(m):=\frac{1}{N-3m}\sum_{k=0}^{N-3m-1}\frac{\left|\left(\sum\limits_{t=k+1}^{k+m}X_{t+m}-\sum\limits_{t=k+1}^{k+m}X_t\right)+\left(\sum\limits_{t=k+m+1}^{k+2m}X_{t+m}-\sum\limits_{t=k+m+1}^{k+2m}X_t\right)\right|}{\left|\left(\sum\limits_{t=k+1}^{k+m}X_{t+m}-\sum\limits_{t=k+1}^{k+m}X_t\right)\right|+\left|\left(\sum\limits_{t=k+m+1}^{k+2m}X_{t+m}-\sum\limits_{t=k+m+1}^{k+2m}X_t\right)\right|}.$$

From [19], with m such that $N/m\to\infty$ and $m\to\infty$,

$$\sqrt{\frac{N}{m}}\big(IR_N(m)-EIR_N(m)\big)\xrightarrow[N\to\infty]{\mathcal{L}}\mathcal{N}(0,\sigma^2(d)),$$

where

$$\sigma^2(d):=2\int_0^\infty\mathrm{Cov}\left(\frac{|Z_d(0)+Z_d(1)|}{|Z_d(0)|+|Z_d(1)|},\frac{|Z_d(\tau)+Z_d(\tau+1)|}{|Z_d(\tau)|+|Z_d(\tau+1)|}\right)d\tau \tag{2.2}$$

$$\text{and}\quad Z_d(\tau):=\frac{1}{\sqrt{|4^{d+0.5}-4|}}\big(B_{d+0.5}(\tau+2)-2\,B_{d+0.5}(\tau+1)+B_{d+0.5}(\tau)\big) \tag{2.3}$$

with B_H a standardized fractional Brownian motion (FBM) with Hurst parameter $H\in(0,1)$.

Remark 2. This convergence was obtained for Gaussian processes in [19], but there also exist results concerning a modified IR statistic applied to stable processes (see [20]) with a different kind of limit theorem. We may suspect that it is also possible to extend the previous central limit theorem to long memory linear processes (since a Donsker type theorem with FBM as limit was proved for long memory linear processes, see for instance [14]) but such a result requires to prove a non obvious central limit theorem for a functional of a multidimensional linear process. Surgailis et al. [19] also considered the case of i.i.d.r.v. in the domain of attraction of a stable law with index $0<\alpha<2$ and skewness parameter $-1\le\beta\le1$ and concluded that $IR_N(m)$ converges to almost the same limit. Finally, in [4] a "continuous" version of the IR statistic is considered for several kind of continuous time processes (Gaussian processes, diffusions and Lévy processes).

Now, instead of this univariate IR statistic, define a multivariate IR statistic as follows: let $m_j=j\,m$, $j=1,\ldots,p$ with $2\le p[N/m]-4$, and define the random vector $(IR_N(j\,m))_{1\le j\le p}$. Thus, p is the number of considered window lengths of this multivariate statistic. In the sequel, we naturally extend the results obtained for $m\in\mathbb{N}^*$ to $m\in(0,\infty)$ by the convention: $(IR_N(j\,m))_{1\le j\le p}=(IR_N(j\,[m]))_{1\le j\le p}$ (which change nothing to the asymptotic results).

We can establish a multidimensional central limit theorem satisfied by $(IR_N(j\,m))_{1\le j\le p}$.