

GREQAM

Groupement de Recherche en Economie
Quantitative d'Aix-Marseille - UMR-CNRS 6579
Ecole des Hautes Etudes en Sciences Sociales
Universités d'Aix-Marseille II et III

Document de Travail

n°2006-01

**Limiting distribution of the least squares
estimates in polynomial regression with long
memory noises**

Mohamed BOUTAHAR

January 2006

DT-GREQAM

Limiting distribution of the least squares estimates in polynomial regression with long memory noises ^{*}

Mohamed BOUTAHAR [†]

January 5, 2006

Abstract

We give the limiting distribution of the least squares estimator in the polynomial regression model driven by some long memory processes. We prove that with an appropriate normalization, the estimation error converges, in distribution, to a random vector which components are a mixture of stochastic integrals. These integrals are with respect to a Lebesgue measure, and can be computed recursively where the seed is a random variable which depends on the assumptions made on the noise process. The limiting distribution can be Gaussian or non Gaussian.

Keywords: Fractional Brownian motion; Long memory; Multiple Wiener-Itô integral; Polynomial regression; Stochastic integral

1 Introduction

Consider the univariate polynomial regression

$$y_t = \theta_0 + \theta_1 t + \dots + \theta_p t^p + \varepsilon_t, \quad (1)$$

^{*}L'auteur remercie le Conseil Régional de la région PACA pour son soutien financier.

[†]Département de mathématiques case 901. Faculté des Sciences de Luminy. 163 Av. de Luminy 13288 MARSEILLE Cedex 9, and GREQAM. boutahar@lumimath.univ-mrs.fr

where y_t is the t th observation on the dependent variable, and ε_t is a zero mean process with finite variance and satisfying the following functional non-central limit theorem (FNCLT)

$$n^{-H}L(n)^{-1/2}\sum_{k=1}^{[nt]}\varepsilon_k\Longrightarrow Z^{(0)}(t), 0 < H < 1, \quad (2)$$

where L is a slowly varying function (i.e., $L(an)/L(n) \rightarrow 1$ as $n \rightarrow \infty$ for any $a > 0$), bounded in all finite intervals, \Longrightarrow denotes the convergence in the CADLAG space $D[0,1]$ endowed with the Skorokhod topology, $Z^{(0)}(t)$ is a random element of $D[0,1]$.

In section 3, we will give some processes who satisfy the FNCLT (2).

The parameter $\theta = (\theta_0, \dots, \theta_p)'$ is unknown and estimated by the least-squares estimator (L.S.E.)

$$\hat{\theta}_n = \left(\sum_{k=1}^n \Phi_k \Phi_k' \right)^{-1} \sum_{k=1}^n \Phi_k y_k, \quad (3)$$

where $\Phi_t = (1, t, \dots, t^p)'$. Recall that the least-squares estimation error satisfies:

$$\hat{\theta}_n - \theta = \left(\sum_{k=1}^n \Phi_k \Phi_k' \right)^{-1} \sum_{k=1}^n \Phi_k \varepsilon_k. \quad (4)$$

The process (1) is very useful in many area. For example many nonstationary (in trend) macroeconomic time series exhibit long memory behavior after extracting a linear trend from them, see Ajmi and Boutahar (2004).

Yajima ((1988),(1991)) considered the L.S.E. in the regression model with deterministic design, he proved its strong consistency and under a condition on the higher-order cumulants of the white-noise process of the errors (ε_t) he proved also the asymptotic normality of the L.S.E. He also proposed to estimate the parameter H by applying Whittle estimator to the least-squares residuals.

If we assume that the noise (ε_t) is a Gaussian process with spectral density satisfying (7) below, then θ can be estimated by using the weighted least squares estimator proposed by Dahlhaus (1995) who proved that the WLSE is asymptotically normal.

Giraitis et al.(1996) proved the asymptotic normality of certain classes of M- and R-estimators of the slope parameter vector in linear regression models with non Gaussian linear errors i.e. (ε_t) is a long memory moving average.

In this paper we will consider a large class of processes (ε_t) . We don't impose Yajima(1991)'s condition on the higher-order cumulants of the white-noise process of (ε_t) . Moreover, we don't assume the Gaussianity of (ε_t) imposed by Dahlhaus (1995), nor the linearity of (ε_t) assumed by Giraitis et al. (1996). We establish a new asymptotic distribution for the L.S.E. $\hat{\theta}_n$. It contains a mixture of recursive stochastic integrals based on the random limit process $Z^{(0)}(t)$ of the FNCLT (2), and can be Gaussian or non Gaussian.

To understand the effects of spuriously detrending a nonstationary fractionally integrated (NFI) process (i.e. the process (14) with $d \geq 1/2$), we refer the reader to the work of Marmol and Velasco (2002). Our process under study is assumed to be a trend-stationary long memory process which is not structurally the same as the NFI, but the two processes can have trajectories which can be visually confused. Moreover, for a given data, the empirical debate of whether such series is best described as being trend-stationary or difference-stationary continues unresolved (see Perron (1989), Andrews and Zivot (1992) for analysis of some macroeconomic time series). Note also that the NFI processes, considered by Marmol and Velasco (2002), don't possess the self-similarity property even supposing, in their assumption A, the normality of (u_t) and that $|d| < 1/2$ (i.e. the process is stationary and invertible). For instance, the fractional Gaussian noise defined by (8) doesn't belong to the class of processes considered by them. In our paper we obtain a limiting distribution which is a functional of the random variable $Z^{(0)}(t)$ of the FNCLT (2), which can be Gaussian (fractional Brownian motion), non Gaussian (a multiple Wiener-Itô integral) or any other process, whereas the one obtained by Marmol and Velasco (2002) is a Gaussian distribution which is a functional of Holmgren-Riemann-Liouville integral (see Lévy (1953) for more details about this process)

defined by $B_H^0(t) = \int_0^t (t-s)^{H-1/2} dB(s) / \Gamma(H + \frac{1}{2})$.

2 Main result

Lemma 1. Assume that the process (ε_t) satisfies the FNCLT (2) and let

$$Z_n(\tau_0, \dots, \tau_p) = n^{-H} L(n)^{-1/2} \left(\sum_{k=1}^{[n\tau_0]} \varepsilon_k, n^{-1} \sum_{k=1}^{[n\tau_1]} k \varepsilon_k, \dots, n^{-p} \sum_{k=1}^{[n\tau_p]} k^p \varepsilon_k \right). \quad (5)$$

Then

$$Z_n \Rightarrow Z,$$

where

$$Z(\tau_0, \dots, \tau_p) = \left(Z^{(0)}(\tau_0), \dots, Z^{(p)}(\tau_p) \right),$$

$$\text{and for all } 1 \leq j \leq p, \quad Z^{(j)}(\tau) = \tau Z^{(j-1)}(\tau) - \int_0^\tau Z^{(j-1)}(s) ds.$$

Proof. Let $Z_n^{(j)}(\tau) = \sum_{k=1}^{[n\tau]} k^j \varepsilon_k$. It's not difficult to show that for all $1 \leq j \leq p$

$$Z_n^{(j)}(\tau) = [n\tau] Z_n^{(j-1)}(\tau) - \sum_{k=1}^{[n\tau]-1} Z_k^{(j-1)}(\tau),$$

since $Z_n^{(0)}(\tau) = \sum_{k=1}^{[n\tau]} \varepsilon_k$, the conclusion of the lemma holds from (2) by applying the continuous mapping theorem, by induction on j , and by using the fact that for all $1 \leq j \leq p$, the following functional

$$(f_1(\bullet), \dots, f_j(\bullet)) \mapsto \left(f_1(\bullet), \dots, f_j(\bullet), \bullet f_j(\bullet) - \int_0^\bullet f_j(s) ds \right)$$

is continuous from $D[0, 1]^j$ to $D[0, 1]^{j+1}$. \square

To state the main theorem let us consider the normalization matrix

$$D_n = \text{diag}(n^{-\frac{1}{2}}, n^{-\frac{3}{2}}, \dots, n^{-\frac{(2p+1)}{2}}).$$

Theorem 2. *The limiting distribution of $\hat{\theta}_n$ is given by*

$$n^{-H+\frac{1}{2}}L(n)^{-1/2}D_n^{-1}(\hat{\theta}_n-\theta) \xrightarrow{d} \mathcal{H}^{-1}\xi,$$

where $\xi = Z(1, \dots, 1)' = (Z^{(0)}(1), \dots, Z^{(p)}(1))'$, the processes $Z^{(j)}(\tau)$, $0 \leq j \leq p$, are defined in Lemma 1, \mathcal{H} is the Hilbert matrix with the (i, j) -th entry $h(i, j) = \frac{1}{i+j-1}$, $1 \leq i, j \leq p+1$.

Proof. We have

$$\begin{aligned} n^{-H+\frac{1}{2}}L(n)^{-1/2}D_n^{-1}(\hat{\theta}_n-\theta) &= \left(D_n \sum_{k=1}^n \Phi_k \Phi_k' D_n \right)^{-1} n^{-H+\frac{1}{2}}L(n)^{-1/2}D_n \sum_{k=1}^n \Phi_k \varepsilon_k, \\ &= \mathcal{H}_n^{-1} Z_n(1, \dots, 1)' \end{aligned}$$

where

$$\begin{aligned} \mathcal{H}_n &= D_n \sum_{k=1}^n \Phi_k \Phi_k' D_n \quad \text{and} \quad Z_n(\tau_0, \dots, \tau_p) \text{ is given by (5).} \\ \mathcal{H}_n(i, j) &= \frac{1}{n^{i+j-1}} \sum_{k=1}^n k^{i+j-2} \longrightarrow \frac{1}{i+j-1}, \end{aligned}$$

hence

$$\mathcal{H}_n \longrightarrow \mathcal{H}. \tag{6}$$

\mathcal{H} is an Hilbert matrix, therefore it is positive definite (cf. Choi (1983)). Consequently the conclusion follows from (6) and the Lemma 1. \square

3 Some processes satisfying the FNCLT

There are many long memory processes who satisfy the FNCLT (2). The following four classes are usually considered in the literature.

3.1 Gaussian processes

Let (ε_t) be a stationary Gaussian process with regularly varying spectral density $f(\lambda)$ of the form

$$f(\lambda) = |\lambda|^{1-2H} L(|\lambda|^{-1}), \quad 0 < H < 1, \tag{7}$$

where L is a slowly varying function (i.e., $L(an)/L(n) \rightarrow 1$ as $n \rightarrow \infty$ for any $a > 0$), bounded in all finite intervals and f is integrable on $[-\pi, \pi]$.

There are two popular processes satisfying (7). The first one is the fractional Gaussian noise, introduced by Mandelbrot and Van Ness (1968), defined by

$$X_H(t) = B_H(t) - B_H(t-1), \quad (8)$$

where $B_H(t)$ is the reduced fractional Brownian motion

$$B_H(t) = \frac{1}{\Gamma(H + \frac{1}{2})} \left\{ \int_{-\infty}^0 [(t-s)^{H-1/2} - (s)^{H-1/2}] dB(s) + \int_0^t (t-s)^{H-1/2} dB(s) \right\}, \quad (9)$$

$\Gamma(\cdot)$ standing for the gamma function, and $B(t)$ is the standard Brownian motion.

An important property of fractional Brownian motion is the self-similarity of its increments i.e.

$$B_H(t_0 + \tau) - B_H(t_0) \stackrel{d}{=} n^{-H} (B_H(t_0 + n\tau) - B_H(t_0)), \quad \forall(t_0, n, \tau), \quad (10)$$

where $Y \stackrel{d}{=} Z$ means that Y and Z have the same finite joint distribution functions.

The equality (10) implies the following property for the fractional Gaussian noise $X_H(t)$

$$A_n = \sum_{k=1}^n X_H(t-1+k) \stackrel{d}{=} n^H X_H(t), \quad (11)$$

i.e. the aggregated and normalized process (A_n/n^H) has the same finite joint distribution function as the starting value $X_H(t)$.

The spectral density of the fractional Gaussian noise (see Sinai (1976)) is given by

$$f(\lambda) = V(H)F(H)(1 - \cos \lambda) \sum_{k=-\infty}^{+\infty} |\lambda + 2k\pi|^{-1-2H}, \quad (12)$$

where

$$V(H) = \frac{1}{[\Gamma(H + \frac{1}{2})]^2} \left\{ \int_{-\infty}^0 [(1-s)^{H-1/2} - (-s)^{H-1/2}]^2 ds + \frac{1}{2H} \right\}, \quad (13)$$

$$F(H) = \left(\int_{-\infty}^{+\infty} (1 - \cos x) |x|^{-1-2H} dx \right)^{-1},$$

Further properties of fractional Gaussian noise are discussed in Mandelbrot and Taqqu (1979). The fractional Gaussian noise (or fractional Brownian motion) has been used as a model or noise in various domains: geophysical data (Beran and Terrien (1996), Graf (1983)), communication (Mandelbrot (1965), Leland et al. (1994)), see also the references therein.

The second process who satisfies (7) is the fractional autoregressive-moving average ARFIMA(p, d, q)

$$\phi(L)(1-L)^d(Y_t - \mu) = \theta(L)u_t, \quad (14)$$

where

$$\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p, \quad \theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q,$$

$d \in \mathbb{R}$, is the memory parameter such that $|d| < 1/2$, L is the backshift operator $LY_t = Y_{t-1}$, $\mu = E(Y_t)$, $u_t \sim i.i.d.(0, \sigma^2)$ and

$$(1-L)^d = \sum_{k \geq 0} b_k(d) L^k, \quad (15)$$

where

$$b_k(d) = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} \quad \text{for every } k \geq 0. \quad (16)$$

The spectral density of the ARFIMA process is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} |1 - e^{i\lambda}|^{-2d} |\theta(e^{i\lambda})|^2 |\phi(e^{i\lambda})|^{-2}. \quad (17)$$

Let $\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda$ denote the autocovariance function of (Y_n) , then we have (see Zygmund (1959), theorem (2.22), page 190)

$$\gamma(k) \sim 2k^{2H-2} L(k) \Gamma(2-2H) \sin(\pi(H - \frac{1}{2})) \quad \text{as } k \rightarrow \infty.$$

This class of processes was introduced by Granger and Joyeux (1980) and Hosking (1981) and is frequently used in modelling economic time series. The motivation behind is that the aggregation of stationary short memory processes (for example ARMA) can generate a stationary process with long-range dependence, i.e. a process with persistent memory, which has the following two properties: 1) the autocorrelation function $\gamma(k)$ decays hyperbolically and is not absolutely summable, $\sum |\gamma(k)| = +\infty$, 2) the spectral density $f(\lambda)$ is unbounded at the frequency 0 and is integrable on $[-\pi, \pi]$ (see Granger (1980), Hassler and Wolters (1995)). However, this class of processes doesn't possess the self-similarity property (11).

To obtain the FNCLT (2) for Gaussian processes satisfying (7), let $\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda$ denote the autocovariance function of (ε_n) , then

$$\begin{aligned} \sum_{j=1}^n \sum_{l=1}^n \gamma(j-l) &= \int_{-\pi}^{\pi} \left| \sum_{k=1}^n e^{ik\lambda} \right|^2 |\lambda|^{1-2H} L(|\lambda|^{-1}) d\lambda \\ &= \int_{-\pi}^{\pi} \left| \frac{e^{i(n+1)\lambda} - e^{i\lambda}}{e^{i\lambda} - 1} \right|^2 |\lambda|^{1-2H} L(|\lambda|^{-1}) d\lambda \\ &= n^{2H} L(n) \frac{1}{L(n)} \int_{-\pi}^{\pi} \left| \frac{e^{\frac{i(n+1)\lambda}{n}} - e^{\frac{i\lambda}{n}}}{e^{\frac{i\lambda}{n}} - 1} \right|^2 |\lambda|^{1-2H} L\left(\left|\frac{\lambda}{n}\right|^{-1}\right) d\lambda \\ &\sim n^{2H} L(n) \int_{-\infty}^{\infty} \left| \frac{e^{i\lambda} - 1}{i\lambda} \right|^2 |\lambda|^{1-2H} d\lambda \\ &= n^{2H} L(n) \frac{\pi}{H \sin(H\pi) \Gamma(2H)}. \end{aligned}$$

By applying the Lemma 5.1 of Taqqu (1975) we have

$$n^{-H} L(n)^{-1/2} \sum_{k=1}^{[nt]} \varepsilon_k \Rightarrow Z^{(0)}(t) = \frac{C(H)}{V(H)^{1/2}} B_H(t), \text{ where } C^2(H) = \frac{\pi}{H \sin(H\pi) \Gamma(2H)}. \quad (18)$$

3.2 Non Gaussian linear processes

For non Gaussian linear processes, many versions of the FNCLT exist in the literature, for example Davydov (1970) proved (2) with $Z^{(0)}(t) = \frac{B_H(t)}{V(H)^{1/2}}$, by assuming that (ε_t) is linear, i.e. $\varepsilon_t = \sum_{j \in \mathbb{Z}} c_j u_{t-j}$, and satisfies the following conditions:

i) the process (ε_t) has a finite variance $\sum_{j \in \mathbb{Z}} c_j^2 < \infty$; ii) (u_t) is an i.i.d. sequence with zero mean; iii) $E(|u_1|^{2k}) < \infty$ for some $k \geq 2$; iv) $\text{var}(\sum_{t=1}^n \varepsilon_t) = n^{2H} L(n)$, $\frac{1}{2+k} \leq H \leq 1$.

An other FNCLT for fractionally integrated processes was also established by Davidson and De Jong (2000) who assumed that

$$(1 - L)^d \varepsilon_t = u_t, \quad (19)$$

and the process (u_t) is such that

i) $E(u_t) = 0$; ii) $\sup_t E(|u_t|^r) < \infty$ for some $r > 2$; iii) u_t is $L_2 - NED$ of size $-1/2$ on V_t , where V_t is either α -mixing sequence of size $-r/(r-2)$ or a ϕ -mixing of size $-r/(2(r-1))$; iv) u_t is covariance stationary and $0 < \sigma_u^2 < \infty$, where

$$\sigma_u^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n E(u_t u_s).$$

For the process (19), the FNCLT (2) is obtained by choosing $H = d+1/2$, $\text{var}(\sum_{t=1}^n \varepsilon_t) = n^{2H} L(n)$ and $Z^{(0)}(t) = \frac{B_H(t)}{V(H)^{1/2}}$.

3.3 Nonlinear functionals of stationary Gaussian process

Let $\varepsilon_t = G(X_t)$ where $X_t = \frac{1}{\sigma} \int_{-\pi}^{\pi} e^{it\lambda} f^{1/2}(\lambda) W(d\lambda)$, $\sigma^2 = \int_{-\pi}^{\pi} f(\lambda) d\lambda$, $f(\lambda)$ is given by (7)

with $1 - \frac{1}{2m} < H < 1$, $W(\cdot)$ is the Gaussian random measure corresponding to a white noise, G is a nonlinear function and m denotes the Hermite rank of G .

Then by using the results of (Taqqu (1979), Dobrushin and Major (1979)) we can deduce the FNCLT

$$n^{-1+m(1-H)} L(n)^{-m/2} \sum_{k=1}^{[nt]} \varepsilon_k \Longrightarrow Z^{(0)}(t) \quad (20)$$

where

$$\begin{aligned}
Z^{(0)}(t) &= \sigma^{2m} K(m, H) K_2(m, H) J(m) \bar{Z}_m(t), \\
\bar{Z}_m(t) &= K_1(m, H) \int_{\mathbb{R}^m} \frac{e^{i(\lambda_1 + \dots + \lambda_m)t} - 1}{i(\lambda_1 + \dots + \lambda_m)} |\lambda_1|^{1/2-H} \dots |\lambda_m|^{1/2-H} W(d\lambda_1) \dots W(d\lambda_m), \\
J(m) &= E(G(X_1) H_m(X_1)), \quad K_2(m, H) = \left\{ 2\Gamma(2-2H) \sin(\pi(H - \frac{1}{2})) \right\}^{m/2}, \\
K(m, H) &= \left\{ \frac{m!(m(H-1)+1)(2m(H-1)+1) \Gamma^m(3/2-H)}{(\Gamma(H-1/2) \Gamma(2-2H))^m} \right\}^{1/2}, \\
K_1(m, H) &= \left\{ \frac{(m(H-1)+1)(2m(H-1)+1)}{m! K_2^2(m, H)} \right\}^{1/2},
\end{aligned}$$

$H_m(x)$ are the Hermite polynomials, and $\int_{\mathbb{R}^m}''$ is the multiple Wiener-Itô integral defined in Major (1981).

3.4 Polynomial functionals of linear processes

Suppose that $X_t = \sum_{j \leq t} c_{t-j} u_j$ satisfies the following conditions:

- i) (u_t) is an i.i.d. sequence with zero mean; ii) $E(|u_1|^{2m}) < \infty$, $1 - \frac{1}{2m} < H < 1$; iii) $c_n \sim n^{H-3/2} L^{1/2}(n)$.

Let $\varepsilon_t = A_m(X_t)$ where A_m is the m th Appell polynomial associated with the distribution of X_0 . The FNCLT (20) was established by Surgailis (1982) and Avram and Taqqu (1987) where

$$Z^{(0)}(t) = K_3(m, H) \bar{Z}_m(t), \quad K_3(m, H) = \left\{ \lim_{n \rightarrow \infty} \text{var} \left(\sum_{t=1}^n \varepsilon_t \right) / (L(n)^m n^{2(1-m(1-H))}) \right\}^{1/2}.$$

Remark. If we consider the process (1) and assume that (ε_t) belongs to one of the four classes of processes cited above then the L.S.E. $\hat{\theta}_n$ is consistent. Moreover, for the first two classes, $\hat{\theta}_n$ is such that

$$|\hat{\theta}_{i,n} - \theta_i| = O_P \left(L(n)^{\frac{1}{2}} n^{H-i-1} \right), \quad 0 \leq i \leq p; \quad (21)$$

whereas for the last two classes it satisfies

$$\left| \hat{\theta}_{i,n} - \theta_i \right| = O_P \left(L(n)^{\frac{m}{2}} n^{m(H-1)-i} \right), 0 \leq i \leq p. \quad (22)$$

4 A Monte-Carlo study

Usually, as in a macroeconomic modelling, we extract a linear function on time to remove the trend of the nonstationary time series, hence we will consider the process

$$y_t = \theta_0 + \theta_1 t + \varepsilon_t, \quad (23)$$

We consider two processes for the noise (ε_t) . The first one is an I(d) process and the second one is a quadratic function of the I(d) process. The limiting distribution of $\hat{\theta}_n$ is Gaussian for the former, while is not for the later.

We have generated time series by choosing $\theta_0 = 3, \theta_1 = 0.02$ in the process (23), and taking six values for $d = 0.1, 0.2, 0.3, 0.4, 0.45, 0.49$ i.e. the noise ε_t moves from a region which is nearly a white noise to a region which is nearly nonstationary. The number of time series generated from each process was 1000. We consider two sample sizes, $n = 100$ and $n = 500$.

4.1 Gaussian limiting distribution

The noise (ε_t) is an I(d) i.e.

$$(1 - L)^d \varepsilon_t = u_t, \quad u_t \sim i.i.d. N(0, \sigma_u^2). \quad (24)$$

For the process (23) we have the following convergence in distribution

$$\begin{pmatrix} n^{\frac{1}{2}-d}(\hat{\theta}_{0,n} - \theta_0) \\ n^{\frac{3}{2}-d}(\hat{\theta}_{1,n} - \theta_1) \end{pmatrix} \xrightarrow{d} \zeta = \frac{\sigma_u}{(2\pi)^{1/2}} \frac{C(H)}{V(H)^{1/2}} \begin{pmatrix} B_H(1) \\ \int_0^1 (B_H(1) - B_H(s)) ds \end{pmatrix}, H = d + \frac{1}{2} \quad (25)$$

and

$$\text{cov}(\zeta) = \frac{\sigma_u^2}{2\pi(2H+1)} C^2(H) \begin{pmatrix} 28-16H & 36H-54 \\ 36H-54 & 108-72H \end{pmatrix}.$$

Table 1. The L.S.E. of $\theta = (\theta_0, \theta_1)'$, $n = 100$, (Gaussian case).

d $n = 100$	0.1	0.2	0.3	0.4	0.45	0.49
$\hat{\theta}_{0,n}$	2.995 (0.279)	2.978 (0.389)	3.027 (0.604)	2.969 (1.057)	2.952 (1.687)	3.179 (3.778)
$\hat{\theta}_{1,n}$	0.02 (0.005)	0.020 (0.006)	0.020 (0.009)	0.020 (0.012)	0.021 (0.014)	0.019 (0.016)

Figure 1 represents the histograms of the distribution of $\hat{\theta}_n$ based on the 1000 realizations corresponding to the Table 1.

Table 2. The L.S.E. of $\theta = (\theta_0, \theta_1)'$, $n = 500$, (Gaussian case).

d $n = 500$	0.1	0.2	0.3	0.4	0.45	0.49
$\hat{\theta}_{0,n}$	2.996 (0.146)	2.998 (0.244)	3.003 (0.441)	2.962 (0.926)	2.982 (1.601)	2.940 (3.744)
$\hat{\theta}_{1,n}$	0.020 (0.0005)	0.020 (0.0008)	0.020 (0.001)	0.020 (0.002)	0.020 (0.003)	0.020 (0.003)

Figure 2 represents the histograms of the distribution of $\hat{\theta}_n$ based on the 1000 realizations corresponding to the Table 2.

For n fixed, the standard errors of the L.S.E. increase as the long range dependent parameter d increases. For d fixed, the standard errors decrease as the sample size increases. Note that the L.S.E. is unbiased and that the standard errors of the slope estimator $\hat{\theta}_{1,n}$ are lesser than those of the intercept estimator $\hat{\theta}_{0,n}$. These conclusions are naturally in accordance with the result (25) which indicates that the speed of convergence of $\hat{\theta}_{1,n}$ is greater than the one of $\hat{\theta}_{0,n}$. Finally for $d = 0.49$, the standard

errors do not converge to 0 for the estimator $\hat{\theta}_{0,n}$, this indicates that $\hat{\theta}_{0,n}$ will not be consistent if the noise (ε_t) is nonstationary. However, it seems that the consistency of $\hat{\theta}_{1,n}$ still holds.

4.2 Non Gaussian limiting distribution

The noise (ε_t) is defined by $\varepsilon_t = X_t^2/E(X_t^2) - 1$, where (X_t) is an $I(d)$ process

$$(1-L)^d X_t = u_t, \quad u_t \sim i.i.d. N(0, \sigma_u^2), \quad \frac{1}{4} < d < \frac{1}{2}. \quad (26)$$

The limiting distribution of the L.S.E. is given by

$$\begin{pmatrix} n^{1-2d}(\hat{\theta}_{0,n} - \theta_0) \\ n^{3-2d}(\hat{\theta}_{1,n} - \theta_1) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} -Z^{(0)}(1) + 3 \int_0^1 Z^{(0)}(s) ds, \\ 3 Z^{(0)}(1) - 6 \int_0^1 Z^{(0)}(s) ds \end{pmatrix}, \quad Z^{(0)}(t) = K_2(d)Z(t), \quad (27)$$

where

$$K_2(d) = \frac{4\sigma_u^6}{\pi} \frac{d(4d-1)\Gamma(1-2d)}{\Gamma(d)\Gamma^3(1-d)},$$

$$Z(t) = \int_{\mathbb{R}^2} \frac{e^{i(\lambda_1+\lambda_2)t} - 1}{i(\lambda_1 + \lambda_2)} |\lambda_1|^{-d} |\lambda_2|^{-d} W(d\lambda_1) W(d\lambda_2).$$

Table 3. The L.S.E. of $\theta = (\theta_0, \theta_1)'$, $n = 100$, (non Gaussian case).

d $n = 100$	0.1	0.2	0.3	0.4	0.45	0.49
$\hat{\theta}_{0,n}$	3.010 (0.285)	3.003 (0.326)	2.960 (0.458)	3.026 (0.812)	3.003 (0.976)	2.993 (1.245)
$\hat{\theta}_{1,n}$	0.020 (0.005)	0.020 (0.005)	0.020 (0.007)	0.019 (0.011)	0.020 (0.011)	0.020 (0.008)

Figure 3 represents the histograms of the distribution of $\hat{\theta}_n$ based on the 1000 realizations corresponding to the Table 3.

Table 4. The L.S.E. of $\theta = (\theta_0, \theta_1)'$, $n = 500$, (non Gaussian case).

d $n = 500$	0.1	0.2	0.3	0.4	0.45	0.49
$\hat{\theta}_{0,n}$	3.000 (0.128)	3.000 (0.151)	2.994 (0.240)	2.982 (0.532)	2.992 (0.868)	3.003 (1.311)
$\hat{\theta}_{1,n}$	0.020 (0.0004)	0.020 (0.0005)	0.020 (0.0008)	0.020 (0.0015)	0.020 (0.0021)	0.020 (0.0015)

Figure 4 represents the histograms of the distribution of $\hat{\theta}_n$ based on the 1000 realizations corresponding to the Table 4.

The behavior of the standard errors of the L.S.E. is the same as in the Gaussian case. The distribution of $\hat{\theta}_n$ (see Figures 3-4) is clearly non Gaussian and asymmetric, except the case $d = 0.1$ for which (ε_t) is nearly a white noise. Moreover as d approaches 0.5, the overestimation of the intercept θ_0 becomes severe.

References

- [1] Ajmi, A.N., Boutahar, M. (2004), Estimateurs des Moindres Carrés dans le Cas de Régressions Linéaires Avec Bruit à Mémoire Longue : Théorie et Simulation, *Working paper GREQAM* No 2004-03.
- [2] Andrews, D.W.K., Zivot, E. (1992), Further evidence of the great crash, the oil-price shock, and the unit root hypothesis, *Journal of Business and Economic Statistics*, 10, 251-270.
- [3] Avram, F., Taqqu, M.S. (1987), Noncentral limit theorem and Appell polynomials, *The Annals of Probability*. Vol 15, No. 2, 767-775.
- [4] Beran, J. (1994), *Statistics for Long-Memory Processes* (Chapman and Hall, New York).
- [5] Beran, J., Terrin, N. (1996), Testing for a change of the long-memory parameter, *Biometrika* 83, N°3, 627 – 638.

- [6] Billingsley, P.(1968), *Convergence of probability measures* (New York, Wiley).
- [7] Choi, M. D. (1983), Tricks or treats with the Hilbert matrix, *Ann. Math. Month* 90:301-312.
- [8] Dahlhaus, R. (1995), Efficient location and regression estimators for long range dependent regression models. *The Annals of Statistics*. Vol 23, No. 3, 1029-1047.
- [9] Davidson, J., De Jong R. M. (2000), The Functional Central Theorem and Weak Convergence to Stochastic Integrals II: Fractionally Integrated Processes, *Econometric Theory*, Vol 16,5, 643-666.
- [10] Davydov, Yu. A. (1970), The invariance principle for stationary processes, *Theory Probab. Appl.* 15, 487-498.
- [11] Dehling, H., Mikosch, T., Sørensen, M. (2002), *Empirical process techniques for dependent data* (Birkhäuser Boston, Inc., Boston, MA).
- [12] Dobrushin, R.L., Major, P. (1979), Non-central limits theorem for non-linear functionals of Gaussian fields, *Z. Wahrsch. verw. Gebiete*.50, 27-52.
- [13] Doukhan, P., Oppenheim, G., Taqqu, M.S.(2003), *Theory and applications of long-range dependence* (Birkhäuser Boston, Inc., Boston, MA)
- [14] Giraitis, L., Koul, H.L., Surgailis, D. (1996), Asymptotic normality of regression estimators with long memory errors, *Statist. Probab. Lett.* 29, 317-335
- [15] Graf, H. (1983), Long range correlations and estimation of self-similarity parameter, *Ph.D thesis*, ETH Zürich.
- [16] Granger, C.W.J. (1980), Long memory relationships and aggregation of dynamic models, *Journal of Econometrics* 227 – 238.
- [17] Granger, C.W.J., Joyeux, R. (1980), An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis*, N°1, 15 – 29.

- [18] Hassler, U., Wolters, J. (1995), Long memory in inflation rates: International evidence, *Journal of Business and Economic Statistics*, 13, 37 – 45.
- [19] Hosking, J.R.M. (1981), Fractional Differencing, *Biometrika*, 68, 165-176.
- [20] Hosking, J.R.M. (1984), Asymptotic distributions of the sample mean, autocovariances and autocorrelations of long memory time series, *Mathematics Research technical summary report # 2752*, University of Wisconsin, Madison, WI. and Fractional Differencing.
- [21] Leland, W.E., Taqqu, M.S., Willinger, W., Wilson D.V. (1994), On the self-similar nature of Ethernet Traffic, *IEEE/ACM Trans. Networking* 2, No 1, 1-15.
- [22] Lévy, P. (1953), Random functions: General theory with special reference to Laplacian random functions. Univ. California Publ. Statist. 1, 331-390.
- [23] Major, P.(1981), Multiple Wiener-Itô integrals, *lecture Notes in Math. 849*, Springer, New-York.
- [24] Mandelbrot, B.B. (1965), Self-similar error clusters in communication and the concept of conditional stationarity, *IEEE, Trans. Commun. Technol.* COM-13, 71-90.
- [25] Mandelbrot, B.B., Van Ness, J.W. (1968), Fractional Brownian motion, fractional noises and applications, *SIAM Rev.* 10, 422-437.
- [26] Mandelbrot, B.B., Taqqu, M.S.(1979), Robust R/S analysis of long run serial correlation. *Proceeding of the 42nd session of the International Statistical Institute, Manila, Bulletin of the I.S.I* 48 (Book 2) 69-104.
- [27] Marmol, F., Velasco, C. (2002), Trend stationarity versus long-range dependence in time series analysis, *Journal of Econometrics* 108, 25 – 42.
- [28] Perron, P. (1989), The great crash, the oil price shock, and the unit root hypothesis, *Econometrica*, 55, 277-532.

- [29] Sinai, Ya. G. (1976), Self-similar probability distributions, *Theory Probab. Appl.* 21, 64-80.
- [30] Sowell, F. (1990), The Fractional unit root distribution, *Econometrica* 58, 495-505.
- [31] Surgailis, D. (1982), Zones of attraction of self-similar multiple integrals, *Lithuanian.Math. J.* 22, 327-340.
- [32] Taqqu, M.S. (1975), Weak Convergence to Fractional Brownian Motion and to the Rosenblatt Process, *Z. Wahrsch. verw. Gebiete.* 31, 287-302.
- [33] Taqqu, M.S. (1979), Convergence of Integrated Processes of Arbitrary Hermite Rank, *Z. Wahrsch. verw. Gebiete.* 50, 83.
- [34] Yajima, Y. (1988), On estimation of a regression model with long-memory stationary errors, *The Annals of Statistics.* Vol 16, No. 2, 791-807.
- [35] Yajima, Y. (1991), Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *The Annals of Statistics.* Vol 19, No. 1, 158-177.
- [36] Zygmund, A.(1959), *Trigonometric series*, Cambridge University Press, London.

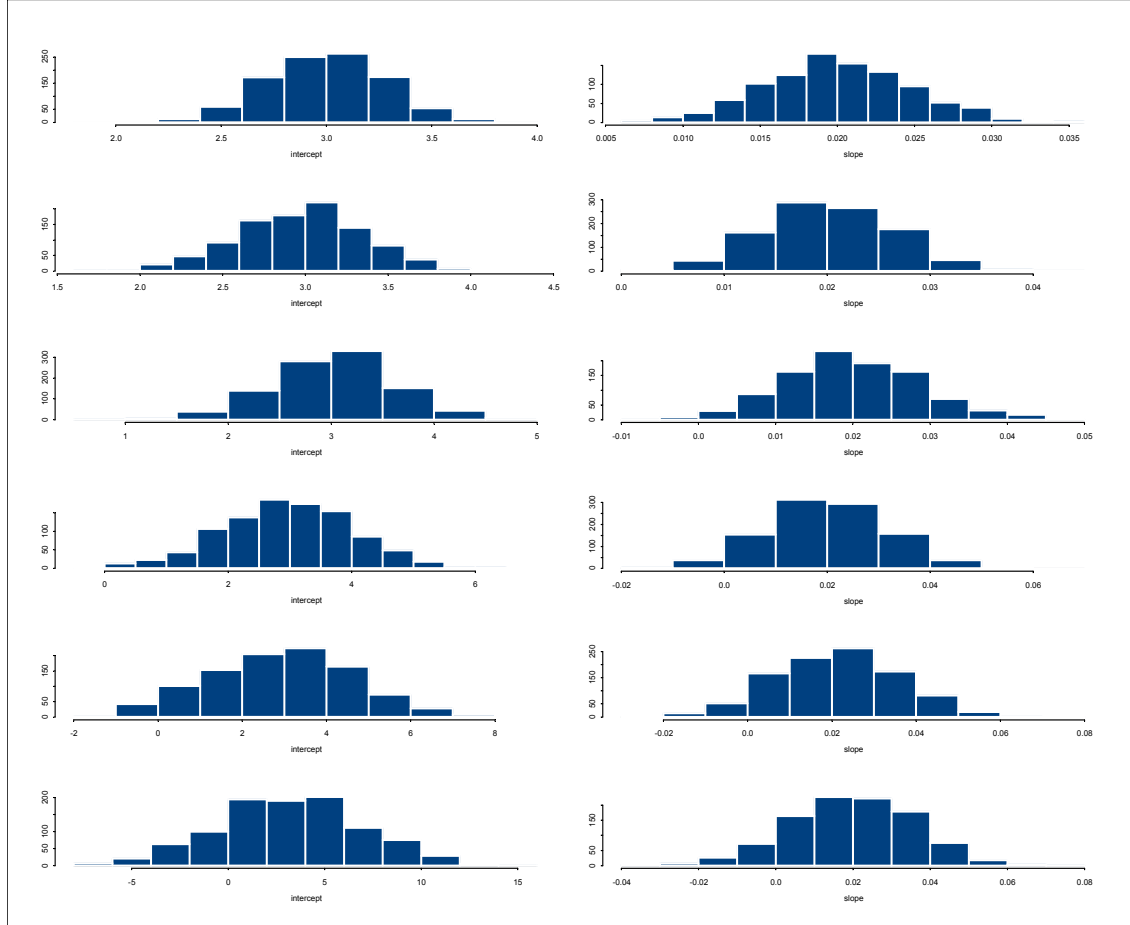


Figure 1: The histograms of the distributions of the L.S.E. On the left the intercept $\hat{\theta}_{0,n}$ and on the right the slope $\hat{\theta}_{1,n}$. The sample size is $n = 100$. On the top $d = 0.1$, next $d = 0.2$ and so on until on the bottom $d = 0.49$. The means and the standard deviations are given in Table 1.

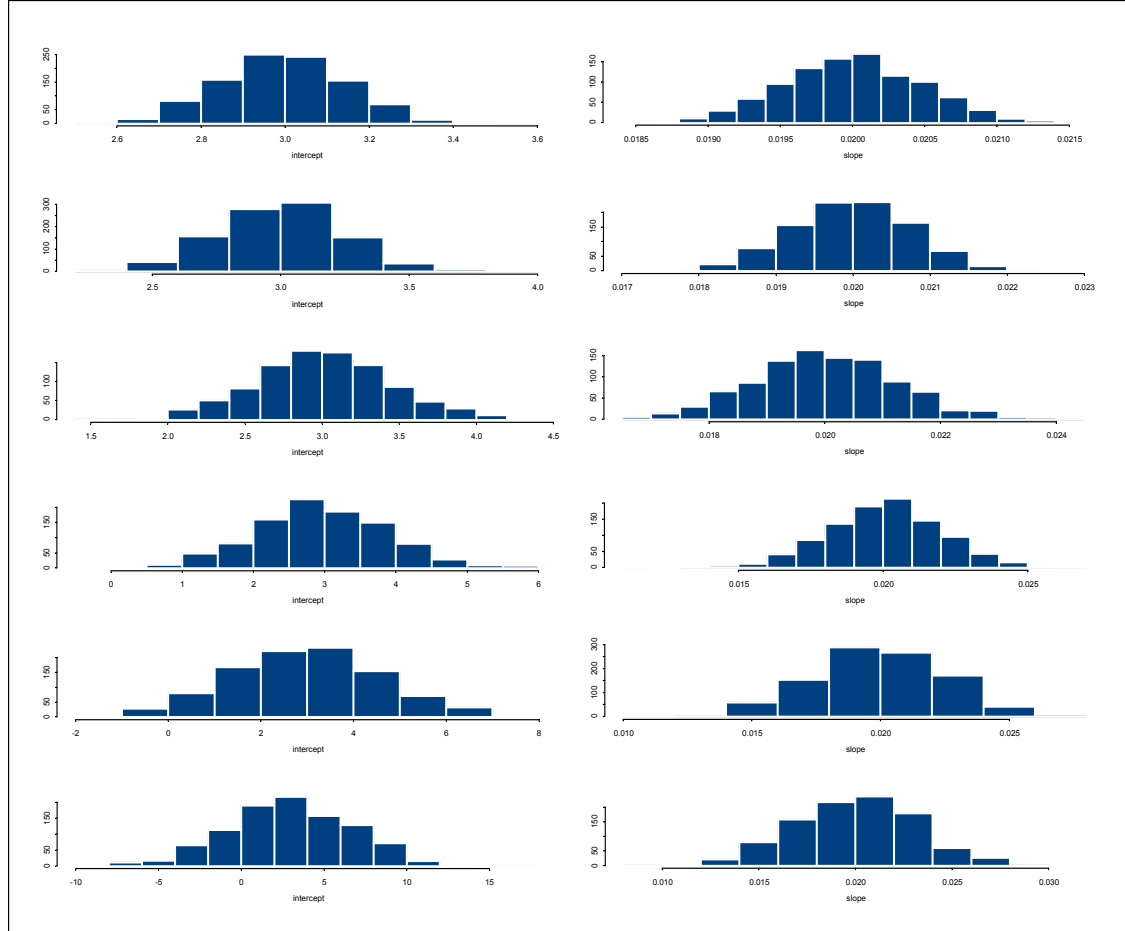


Figure 2: The histograms of the distributions of the L.S.E. On the left the intercept $\hat{\theta}_{0,n}$ and on the right the slope $\hat{\theta}_{1,n}$. The sample size is $n = 500$. On the top $d = 0.1$, next $d = 0.2$ and so on until on the bottom $d = 0.49$. The means and the standard deviations are given in Table 2.

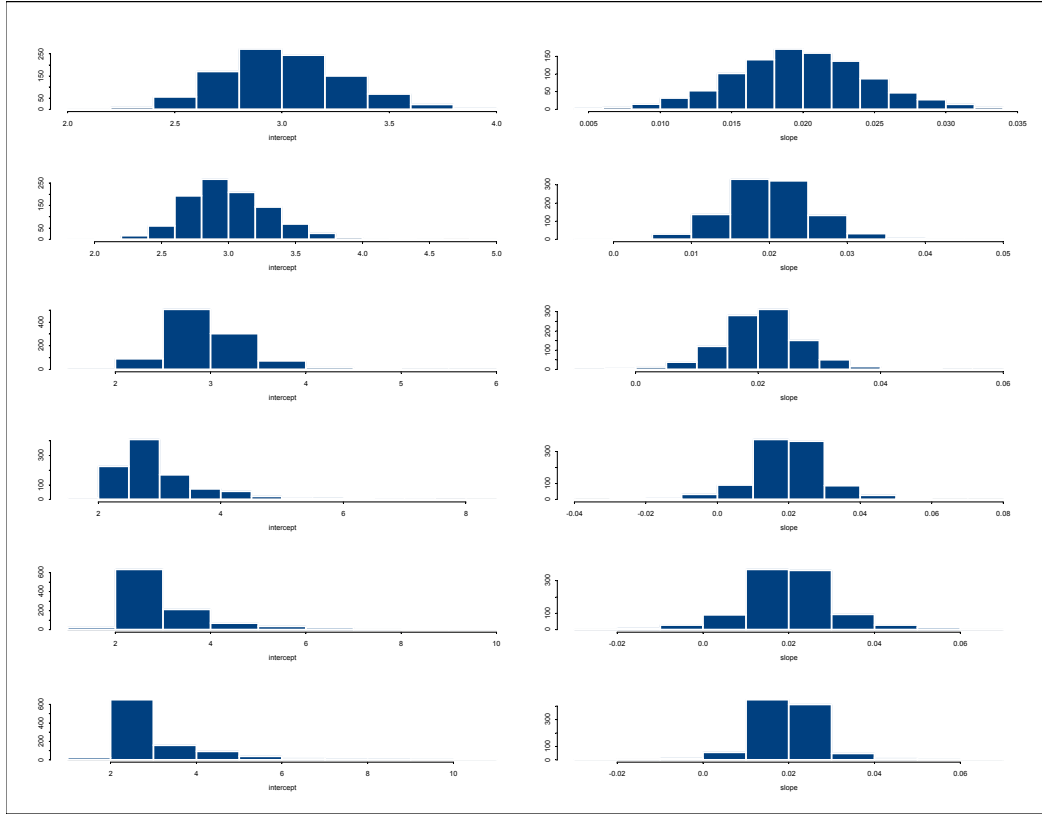


Figure 3: The histograms of the distributions of the L.S.E. On the left the intercept $\hat{\theta}_{0,n}$ and on the right the slope $\hat{\theta}_{1,n}$. The sample size is $n = 100$. On the top $d = 0.1$, next $d = 0.2$ and so on until on the bottom $d = 0.49$. The means and the standard deviations are given in Table 3. (Non Gaussian limit).

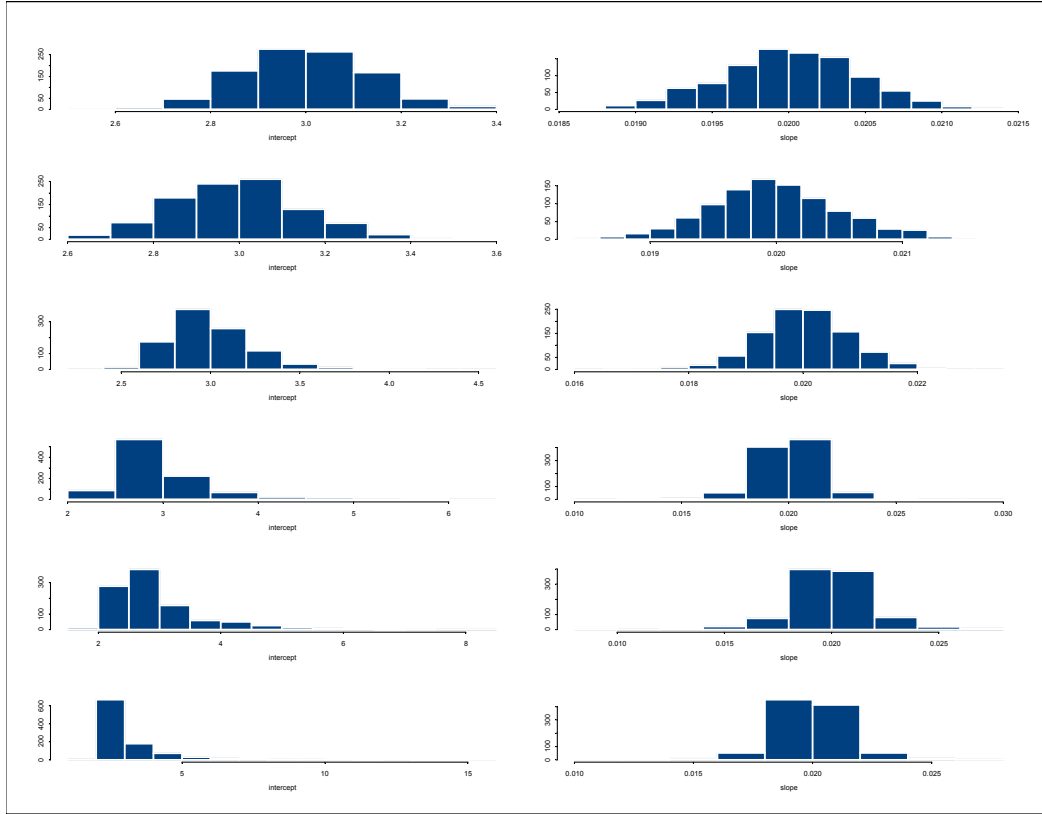


Figure 4: The histograms of the distributions of the L.S.E. On the left the intercept $\hat{\theta}_{0,n}$ and on the right the slope $\hat{\theta}_{1,n}$. The sample size is $n = 500$. On the top $d = 0.1$, next $d = 0.2$ and so on until on the bottom $d = 0.49$. The means and the standard deviations are given in Table 4. (Non Gaussian case)