

How Can We Define The Concept of Long Memory? An Econometric Survey

Guégan D. *†

29th April 2004

Abstract

In this paper we discuss different aspects of long memory behavior and specify what kinds of parametric models follow them. We discuss the confusion which can arise when empirical autocorrelation function of a short memory process decreases in an hyperbolic way.

Keywords: Long-memory - Switching - Estimation theory - Spectral domain - Returns.

JEL classification: C32, C51, G12

1 Introduction

The possibility of confusing long memory behavior with structural changes requires an understanding of the kind of long memory behavior concerned. This paper attempts to be comprehensive in terms of coverage of results that appear of direct relevance for econometricians. While long memory models have only been used by econometricians since around 1980, and by financial researchers since around 1995, they have played a role in physical sciences since at least 1950, with statisticians in fields as diverse as hydrology and climatology detecting the presence of long memory within data over both time and space. One attraction of long memory models is that they imply

*E.N.S. Cachan, Département d'Economie et Gestion et Equipe MORA-IDHE UMR CNRS 8533, 61 Avenue du President Wilson, 94235, Cachan Cedex, France, e-mail : guegan@ecogest.ens-cachan.fr

†QUT Visiting Fellow, April 2004, School of Economics and Finance, QUT, GPO Box 2434, Brisbane QLD 4001, Australia

different long run predictions and effects of shocks to conventional macroeconomic approaches. On the other hand, there is substantial evidence that long memory processes describe rather well financial data such as forward premiums, interest rate differentials, inflation rates and exchange rates. Then, empirical works in macroeconomics and finance have witnessed a renaissance in both the theoretical and empirical econometrics of long memory and fractional integration.

But until now little attention has been given to the possibility of confusing long memory and structural change. This is different from the problem encountered concerning the possible confusion between structural changes and unit roots which is now widely appreciated, see for instance Sowell (1990), Stock (1994) and Granger and Ding (1996). Here we will not consider this point of view and will focus on possible interrelationships between long memory behavior and structural changes. Different classes of structural change models which exhibit some long memory behavior have been proposed. This long memory behavior could be an illusion generated by occasional level shifts then inducing the observed persistence, while most shocks dissipate quickly. In contrast, all shocks are equally persistent in a long memory model. Thus, distinguishing between long memory and level shifts could dramatically improve policy analysis and forecasting performance.

Concerning stock returns, several papers have investigated their long-run properties and suggest that the absolute returns are characterized well by long memory processes. Several attempts try to explain these findings such as existence of temporal higher-order dependence and occasional structural breaks. Thus, different classes of processes have been proposed to explain the long memory property in the stock market, based on one hand on heteroscedastic long memory models and on the other on the assumption that existence of breaks might cause the long memory property of the absolute returns. GARCH models (Engle (1982), Bollerslev (1986)) are well known to be short memory, but long range dependence type behavior of the sample autocorrelation function of these models has been observed by Mikosch and Starica (1999). They consider that statistical tools can "see things that are not there". The sample autocorrelation function reading of long memory behavior can be caused either by stationary long memory time series and equally well by non-stationarity in time series. Then, the asymptotic behavior of sample autocorrelations has to be analysed for all the models in discussion in this paper.

If long memory processes refer to fractional models, the breaks refer to non-

linearity to the data. In certain cases, we can adjust a long memory model to nonlinear data and vice versa. We will see in the following that some nonlinear models may generate time series to which one may want to fit linear long memory models. On the other hand, data generated from long memory models may appear to have nonlinear properties, as occasional structural breaks. It would be interesting also to capture both features of long memory and nonlinearity into a single time series model, in order to be able to access their relative importance. This will be done for instance with the SETAR and STAR models with long memory dynamics. These latter models have shown their capability in the modeling of unemployment for instance. In economy, aggregating variables is important, in particular to build different indexes. Granger (1980) shows that we can provide long memory behavior aggregating dynamic equations. Now, related to the concept of fractional long memory, there exist several notions of long memory behavior and we will specify their interrelationships.

In this paper, we will analyse the different concepts of long memory behavior. We will investigate the different classes of models which present some specific long memory behavior and propose some classification. It is important to note that, throughout we do not consider continuous time series, nor time series with infinite variance. Indeed, the notion of long memory, for continuous time series, often refers to self similarity that it is different of long memory behavior we consider here. Now, infinite variance time series can be very interesting from a theoretical point of view, but very few data sets, and little financial and economic data have this characteristic. Indeed, infinite variance necessitates to work with infinite samples and we always work with finite samples. Then, we are not close to the approaches developed by Cioczek-Georges and Mandelbrot (1995), Taqqu, Willinger and Sherman (1997), Heyde and Yang (1997) or Chen, Hansemand and Carrasco (1999). For a review, see Samorodnitsky and Taqqu (1994).

The paper is organized as follows: in Section two, we specify some notations and results. Section three is devoted to the different notions behind the concept of long memory. In Section four, we present some classes of fractional long memory processes. Section five focuses on models with jumps and switching. In Section six, we discuss specific behavior coming from the aggregation of series and in Section seven we question the asymptotic properties of sample covariance and periodogram in their capacity to create long memory behavior. Section eight concludes.

2 Some recalls on stationary processes

A characterization in terms of the spectral density function is convenient for the concept of long memory. Thus, we need to specify when a stationary process has a spectral density function.

Let $(X_t)_t$ represent a real-valued discrete time stationary process (in the covariance sense). Let us assume that $\forall t, E[X_t] = 0$. We define the autocovariance function γ_X as,

$$\gamma_X(\tau) = E[X_t X_{t+\tau}], \forall t, \forall \tau,$$

and $\gamma_X(\tau) = \gamma_X(-\tau)$, and $|\gamma_X(\tau)| < \infty$, for all τ , by stationarity. Concerning linear time series analysis, it is usually assumed $\gamma_X(\tau) \rightarrow 0$ "fairly rapidly" as $\tau \rightarrow \infty$. This can be represented by the following approximation: $\gamma_X(\tau) \approx Cr^\tau$, $|r| < 1$ and C is a constant. In particular the following condition $\sum_{-\infty}^{\infty} |\gamma_X(\tau)| < \infty$ is evidently satisfied by many classical linear processes, such the class of ARMA processes, see Box and Jenkins (1976) or Brockwell and Davis (1996). These processes are called short memory processes, see Cox (1984) and Guégan (2003a). By extension, it is possible that, for some processes, $\sum_{-\infty}^{\infty} |\gamma_X(\tau)| = \infty$. In that latter case, we say that they present long memory behavior. Now, there exist different characterizations of the concept for long memory behavior that we discuss in the next Section. But, first of all, we need to specify the concept of the spectral density function.

By the Wiener-Khinchin theorem, a necessary and sufficient condition that γ_X be the autocovariance function for some stationary process $(X_t)_t$ is that there exists a function F_X (the spectral distribution function) defined on $[-\pi, \pi]$ such that $F_X(-\pi) = 0$, F_X is non decreasing on $[-\pi, \pi]$ and

$$\gamma_X(\tau) = \int_{-\pi}^{\pi} e^{i\lambda\tau} dF_X(\lambda).$$

If the function F_X is absolutely continuous with respect to Lebesgue measure, then

$$F_X(\nu) = \int_{-\pi}^{\nu} f_X(\lambda) d(\lambda),$$

and f_X is called the spectral density function for the process $(X_t)_t$. The function f_X is necessarily non negative on $[-\pi, \pi]$, is even, and unique up to sets of Lebesgue measure zero since, if f_X^* differs only on a null set from f_X , f_X^* is also a spectral density function for $(X_t)_t$.

Moreover

$$\gamma_X(\tau) = \int_{-\pi}^{\pi} e^{i\lambda\tau} f_X(\lambda) d(\lambda).$$

Thus, the autocovariance function and the spectral density function are equivalent specifications of second order properties for a stationary process $(X_t)_t$.

3 Different aspects of long memory behaviors

Traditionally long memory has been defined in the time domain in terms of decay rates of long-lag autocorrelations, or in frequency domain in terms of rates of explosions of low frequency spectra. The presence of explosions in the spectrum of a process $(X_t)_t$ makes it variable. This variability is difficult to take into account. Thus, different ways have been considered in the literature in order to specify this variability, often linked to the presence of long memory behavior.

3.1 Parzen long memory's concept

We introduce first two definitions of the concept of long memory due to Parzen (1981). The first one concerns the time domain, the second one the frequency domain.

Definition 3.1 *A stationary process $(X_t)_t$ with an autocovariance function γ_X is called a long memory process, in the covariance sense, if*

$$\sum |\gamma_X(\tau)| = \infty.$$

Definition 3.2 *A stationary process $(X_t)_t$ with a spectral density function f_X is called a long memory process in the spectral density sense if the ratio:*

$$\frac{\text{esssup} f_X(\lambda)}{\text{essinf} f_X(\lambda)} = \infty. \quad (1)$$

These two definitions, which introduce concepts of long memory in a general context, are not equivalent. A long memory process in the covariance sense need not be a long memory process in the spectral density sense. The converse is also true. In the other hand, there are processes which are long memory in both senses and also processes that are not long memory in either sense, for instance the ARMA processes for the latter case.

It is also possible to exhibit processes whose spectral density has the property defined by the expression (1) and which do not exhibit long memory behavior by construction. For instance we consider the process $(\eta_t)_t$ defined, for all t , by $\eta_t = \varepsilon_t - \varepsilon_{t-1}$, where $(\varepsilon_t)_t$ is a strong white noise. Its spectral density function is equal to $f_\eta(\lambda) = 4 \sin^2 \frac{\lambda}{2}$. Or, the process $(\eta_t)_t$ cannot be considered as a long memory process.

Thus, definition 3.2 is not sufficient to characterize the long memory behavior for a stationary process. To overcome the limitations of the two previous definitions, other approaches have been considered in a narrow way. A more useful definition, in terms of the spectral approach, is the following:

Definition 3.3 *A stationary process $(X_t)_t$ with a spectral density function f_X is called a long memory process in a restricted spectral density sense if f_X is bounded above on $[\delta, \pi]$, for every $\delta > 0$ and if*

$$f_X(\lambda) = \infty, \quad \text{as } \lambda \rightarrow 0^+. \quad (2)$$

Now a long memory process in the restricted spectral density sense can be a long memory process in the covariance sense. There exist processes that are long memory in both the covariance sense and the spectral density sense, but not in the restricted spectral density sense. In summary, we can distinguish:

- Processes which are long memory in the covariance sense, but not in the spectral density sense (the density spectral does not explode at the origin).
- Processes which are long memory both in the covariance and the spectral density sense.
- Processes which are long memory only in the spectral density sense (the autocovariance function decreases quickly towards zero at the origin).
- Processes which are long memory both in the covariance and the restricted spectral density sense.

Now, in the following, when we work in the spectral domain, we only consider the long memory behavior defined in the restricted spectral density sense.

3.2 The concept of Long memory with rate of convergence

It is also important to characterize the rate at which the spectral density diverges to infinity as $|\lambda| \rightarrow 0^+$, see Cox (1977). We are going to specify this rate using the following definition.

Definition 3.4 A stationary process $(X_t)_t$ with a spectral density function f_X is called a long memory process in the spectral density sense with a power law of order $2d$, with $0 < d < 1/2$, if f_X is bounded above on $[\delta, \pi]$, for every $\delta > 0$ and if

$$\lim_{\lambda \rightarrow 0^+} f_X(\lambda) = h|\lambda|^{-2d}, \quad (3)$$

for some $0 < h < \infty$.

It is clear that a power law long memory process is just a particular type of long memory process in the restricted spectral density sense. We can remark also that when $d \geq 1/2$, the spectral density function is not integrable finitely and hence does not correspond to a stationary process. We do not consider this case here. In the following Section, we will consider analytic models whose autocovariance function and spectral density function will present behaviors as those proposed respectively in the definitions 3.1 and 3.4. Recall that if the process $(X_t)_t$ is an ARMA process then $\lim_{\lambda \rightarrow 0^+} f_X(\lambda) = C$, where C is a constant, taken to be positive.

In the definition 3.4, we only consider the explosion of the spectral density in a frequency close to zero. We can generalize the concept of long memory in terms of spectral density with power law at any frequency, then we get the following definition:

Definition 3.5 A stationary process $(X_t)_t$ with a spectral density function f_X is called a long memory process in the spectral density sense with a power law of order $2d$, with $0 < d < 1/2$, if there exists a frequency $\lambda_0 \in [-\pi, \pi]$ such that

$$\lim_{\lambda \rightarrow \lambda_0} f_X(\lambda) = h|\lambda|^{-2d}, \quad (4)$$

for some $0 < h < \infty$.

It is also possible to give, following the definition 3.4, the equivalence in terms of the speed towards zero of the autocovariance function, see Granger and Joyeux (1980). Then, we get the following definition:

Definition 3.6 A stationary process $(X_t)_t$ with an autocovariance function γ_X is called a long memory process in the covariance sense with a speed of convergence of order $2d$, with $0 < d < 1/2$, if

$$\gamma_X(\tau) = C(d)\tau^{2d-1}, \quad \text{as } \tau \rightarrow \infty, \quad (5)$$

with $C(d)$ a constant which depends on d .

3.3 Long memory behavior in Allan's sense

If there exist explosions in the spectral density, those provoke variability of the spectrum which is very difficult to take into account in practice. To measure this variability, the first idea has been to use $\gamma_X(0)$. But, when there exists long memory behavior in the data, it is well known that sample variance can severely underestimate $\gamma_X(0)$, see Beran (1994). Allan (1966) argues that the main problem with the sample variance is that the expectation depends on the length T of the sample. He proposes to overcome this problem by setting $T = 2$ in this classical expression of the sample variance and by averaging the resulting quantity over all available data. Allan's variance is a peculiar quantity. Allan uses an estimation procedure to define a parameter of interest, see also von Neumann (1941). The partial sums of the process $(X_t)_t$ are given by:

$$S_T = \sum_{t=1}^T X_t. \quad (6)$$

Then, we get the following definition:

Definition 3.7 *The Allan variance is $\text{var}[S_T]$, with S_T defined in (6).*

If the process $(X_t)_t$ is stationary, then

$$\text{var}[S_T] = \gamma_X(0) - \gamma_X(1).$$

It is straightforward that if $(X_t)_t$ is a white noise process, we would have $\text{var}[S_T] = \gamma_X(0)$. Following Allan (1966) and Percival (1983) we get the proposition:

Proposition 3.8 *Let $(X_t)_t$ be a stationary process with a spectral density function f_X which verified the equation (3), for $0 < \lambda < \delta$, with $\delta > 0$, then*

$$\text{var}[S_T] = \frac{C(d)}{T^{1-2d}} + O\left(\frac{1}{T^2}\right), \quad \text{if } 0 < d < 1/2.$$

Here, $C(d)$ is a constant that depends only on d .

Definition 3.9 *A stationary process $(X_t)_t$ is called a long memory process in the Allan variance's sense if the rate of growth of the partial sums' variances is such that*

$$\text{var}[S_T] = O\left(\frac{1}{T^{1-2d}}\right),$$

with $0 < d < 1/2$.

This proposition says that if a process is long memory in the spectral sense with a power law $2d$, then it is long memory in the Allan variance's sense. This means also that $\lim_{T \rightarrow \infty} \text{var}[S_T] = \infty$. Then, we say that a process such that

$$\lim_{T \rightarrow \infty} \text{var}[S_T] = C, \quad (7)$$

where C is a constant, is a short memory process, but it is not always assured that the relationship (7) will imply that $f_X(0)$ is finite.

Since the Allan variance cannot be directly related to the process variance (if it exists finitely) of $(X_t)_t$, the question arises as to what the Allan variance does measure. It seems that the Allan variance is a useful measure of variability for power law long memory process of order $2d$. It can be used to determine the exponent $2d$ for certain power law long memory process. We can also try to use it to discriminate between several long memory parametric processes. Concerning its utility to estimate d , this comes from the following relationship, under the assumptions of the proposition 3.8:

$$\ln \text{var}[S_T] = -(1 - 2d) \ln T + \ln C(d) + o(1).$$

We can determine d by examining, on a log-log plot, the slope of $\text{var}[S_T]$ as a function of T , as T becomes large.

A generalization of the Allan variance has been introduced by Heyde and Yang (1997) to capture models with infinite variance, but we do not investigate it in this paper.

3.4 Long memory behavior in the prediction error variance sense

For some time, prediction error variance has been considered as a reasonable alternative as a measure of variability for long memory processes. This approach has been proposed by Priestley (1981). Consider the Wold representation of the stationary process $(X_t)_t$ which is a power law long memory process of order $2d$:

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

where $(\varepsilon_t)_t$ is an uncorrelated, zero mean stationary sequence and $(\psi_j)_j$ a fixed square summable sequence of constants. We denote σ_ε^2 as the variance of the process $(\varepsilon_t)_t$. We assume that the variance σ_ε^2 and the coefficients ψ_j are known. Now, if we suppose that at time 0 we have available the

present and the past values of X_t and that we want to predict X_{t+1} , then its best predictor is a linear combination of present and all the past values of X_t . Thus, the expected value of the squared prediction error (defined as the difference between the observed value of X_{t+1} and its predicted value) is simply σ_ε^2 . This quantity is known also as the innovations' variance. Here we are going to equate the notion of frequency instability with the notion of unpredictability.

Lemma 3.10 *Let $(X_t)_t$ be a stationary process with a spectral density function f_X which verified the expression (3), for $0 < \lambda < \delta$, with $\delta > 0$, then*

$$\sigma_\varepsilon^2 = \exp\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln(2\pi f_X(\lambda)) d\lambda\right] < \infty.$$

We can remark that the process considered in Lemma 3.10 verifies also the assumptions of the proposition 3.9. Here, the proposal is to let unpredictability be the measure of variability in a long memory process: the more predictable a process, the less variable it is declared to be. The measure of frequency instability is just σ_ε^2 .

Now, if we want to compare the Allan variance and the prediction error variance, we can remark that $\sigma_\varepsilon^2 < \text{Var}[S_T]$. Thus, it is very difficult to establish clear relationship in terms of long memory behavior. Yoshimura (1978) shows that there exists spectral density functions for which the Allan variance is proportional to the prediction error variance when each is regarded as a function of the time t . For most of the spectral density functions this property is not true. This means that there should not be any serious discrepancies between the two measures.

3.5 Long memory behavior in distribution

In order to introduce the Rosenblatt approach, we first recall the mixing conditions, see Ibragimov and Rozanov (1974). Let $(X_t)_{t \in \mathbb{Z}}$ be a second order stationary process on \mathbb{R} . Let F_n^m denote the σ -algebra generated by X_n, X_{n+1}, \dots, X_m and let L_n^m denote the closure in L^2 of the vector space spanned by X_n, X_{n+1}, \dots, X_m . We say that the process $(X_t)_t$ satisfies :

- the strong mixing condition if $\lim_{k \rightarrow +\infty} \alpha(k) = 0$, where :

$$\alpha(k) = \sup_{\substack{A \in F_{-\infty}^n \\ B \in F_{n+k}^\infty}} |P(A \cap B) - P(A)P(B)|; \quad (8)$$

- the completely regular condition if $\lim_{k \rightarrow +\infty} r(k) = 0$, where :

$$r(k) = \sup_{\xi_1, \xi_2} |\text{corr}(\xi_1, \xi_2)|, \quad (9)$$

and the random variables ξ_1 and ξ_2 are respectively measurable with respect to the σ -algebras $F_{-\infty}^n$ and F_{n+k}^∞ ;

- the completely linear regular condition if $\lim_{k \rightarrow +\infty} \rho(k) = 0$, where :

$$\rho(k) = \sup_{\substack{\xi_1 \in L_{-\infty}^n \\ \xi_2 \in L_{n+k}^\infty}} |\text{corr}(\xi_1, \xi_2)|. \quad (10)$$

Note that weakly dependent processes have α -mixing coefficients that decay to zero exponentially, and strongly dependent processes have α -mixing coefficients that are identically one. We have also the following relationships between these coefficients : $\alpha(k) \leq r(k)$ and $\rho(k) \leq r(k)$. Moreover, for Gaussian processes, Kolmogorov and Rozanov (1960) have proved that $\rho(k) = r(k)$ and that $\alpha(k) \leq r(k) \leq 2\pi\alpha(k)$. Rosenblatt (1956) defines short range dependence in terms of a process that satisfies strong mixing condition, so that the maximal dependence between two points of a process becomes trivially small as the distance between these points increases. Thus, it is natural to propose yet another definition for the concept of long memory.

Definition 3.11 *A stationary process is long memory in distribution if it is not strongly mixing.*

We can also say that a process $(X_t)_t$ is "persistent" if, denoting $F(., .)$ the joint distribution of two random variables and $F(.,)$ the margins, we have $\forall \tau$ and $\forall t$:

$$F(X_{t+\tau}, X_t) \neq F(X_{t+\tau})F(X_t).$$

Thus, asymptotically the random variables $X_{t+\tau}$ and X_t are dependent. Otherwise the process $(X_t)_t$ is said to be "not persistent". This characterization is weaker than the one given in Definition 3.11.

Now, there exist several ways to measure the distance $d(., .)$ between the distributions of two random variables, as the χ^2 distance, the Uniform distance, the Kullback distance or the Hellinger distance, for instance. We denote :

$$D(\tau) = d[F(X_{t+\tau}, X_t), F(X_{t+\tau})F(X_t)]. \quad (11)$$

If $D(\tau) \rightarrow 0$, as τ increases, we say that there exists no persistence. If $D(\tau) \rightarrow c\tau^{-d}$, for $d > 0$ and $c > 0$, as τ increases, we say that we have long

memory in distribution and the behavior is similar to the one defined by definition 3.11. If $D(\tau) \rightarrow e^{-c\tau}$, $c > 0$, as τ increases, we get short memory in distribution.

Now it is well known that, there exists a function C , under classical conditions such that:

$$F(X_{t+\tau}, X_t) = C(F^{-1}(X_{t+\tau}), F^{-1}(X_t)). \quad (12)$$

The function C is called the copula associated with $F(., .)$, see Sklar (1959). If we take the derivatives with respect to $X_{t+\tau}$, then we get (with obvious notations):

$$f(X_{t+\tau}, X_t) = f(X_{t+\tau})f(X_t)c(u_{t+\tau}, u_t), \quad (13)$$

where $c(., .)$ is the density copula defined on the unit box $[0, 1] \otimes [0, 1]$ and $u_t = F(X_t)$. If the random variables $X_{t+\tau}$ and X_t are independent, then $c(u_{t+\tau}, u_t) \equiv 1$. Using (11) and (13) we get:

$$D(\tau) = d[f(X_{t+\tau}, X_t), f(X_{t+\tau})f(X_t)c(u_{t+\tau}, u_t)]. \quad (14)$$

Thus, for no persistence, we always need $D(\tau) \rightarrow 0$, as τ increases. Now, if we specify a slow rate of convergence for $D(\tau)$, we can observe a long memory behavior. The use of copulas can be interesting to detect long dependence because we are able to estimate them from real data, see Caillaud and Guégan (2003) for details and some review. See also Granger (2003) for a complementary approach with conditional copulas.

3.6 Comments

In this Section, we have proposed some definitions permitting the specification of the notion of long memory behavior for a process $(X_t)_t$. They are not all equivalent. Some induce others. Now, from an empirical point of view, it is difficult to detect long memory behavior in real data using some of these definitions. Thus, we introduce now parametric models whose properties correspond to some of the previous definitions. This will permit us to specify, in an empirical way, what kind of long memory concept seems more realistic to use in applications. We will see also that different classes of models will produce the same long memory behavior. Then, the question is to be able to distinguish between these different models. Until now, tests are not available and the debate is opened.

4 Fractional long memory processes

In this Section we recall several fractional long memory models and we specify what kind of the previous long memory behavior they exhibit. We indicate some possible confusion when there exists in addition some specific non stationarity. We also discuss the different models proposed in the literature for transformations of returns. Let $(X_t)_t$ a stationary process with covariance function γ_X and spectral density f_X and denote B the backshift operator.

Assume that the centred process $(X_t)_t$ is described by the following recursive scheme:

$$\begin{cases} \phi(B) \prod_{i=1}^k (I - 2u_i B + B^2)^{d_i} X_t = \theta(B) \eta_t \\ L(\eta_t | I_t) \sim N(0, \sigma_t^2) \\ \sigma_{t+1} = \varphi(\eta_t, \sigma_t), \end{cases} \quad (15)$$

where the function $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}_{\geq 0}$ is a measurable function. The polynomials $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ are respectively the autoregressive and moving average polynomials. The shocks $(\eta_t)_t$ are characterized by the conditional variance σ_t^2 and I_t is the σ -algebra generated by the past values $(\eta_{t-1}, \eta_{t-2}, \dots)$. The model (15) includes a lot of models considered in the literature from the eighties.

- In (15), if $k = 1, u = 1$ and if $(\eta_t)_t$ is a strong white noise, we obtain the following FARMA(p, d', q) process (centred or not), with $d' = 2d$:

$$\phi(B)(I - B)^{d'} X_t = \theta(B) \eta_t. \quad (16)$$

This process has been introduced by Granger and Joyeux (1980) and Hosking (1981). Particular models of (16) can be expressed by moving average representations such as:

$$X_t = \begin{cases} \sum_{j=0}^t c_j \varepsilon_{t-j} \\ c_j \sim c j^{2d'-1}. \end{cases} \quad (17)$$

This expression corresponds to a finite moving average representation of process (16).

- If, in (15), $(\eta_t)_t$ is a strong white noise, we get the k -factor GARMA process defined by:

$$\phi(B) \prod_{i=1}^k (I - 2u_i B + B^2)^{d_i} X_t = \theta(B) \eta_t. \quad (18)$$

This process has been introduced by Gray, Zhang and Woodward (1989) for $k = 1$ and generalized by Giraitis and Leipus (1995) and Woodward, Cheng and Gray (1998). If $p = q = 0$, we call it the k -factor Gegenbauer process.

- The process (15), that we call here the GIGARCH process has been introduced by Guégan (2000) and (2003c). When $k = 1$ and $u = 1$, we get the FIGARCH model introduced by Baillie, Bollerslev and Mikkelsen (1996) and Baillie, Chung and Tieslau (1996). A generalization of this last model, called the FIEGARCH model, has also been proposed by Baillie, Bollerslev and Mikkelsen (1996).

We specify now the long memory behaviors of these models when they are second order stationary.

The asymptotic behavior of the spectral density of processes (15) and (18) is :

$$f_X(\lambda) \sim C(\lambda)|\lambda - \lambda_j|^{-2d_j}, \text{ as } \lambda \rightarrow \lambda_j, j = 1, \dots, k, \quad (19)$$

where $C(\lambda)$ depends only on λ . It explodes in the G -frequencies $\lambda_i = \cos^{-1}(u_i)$. For the stationary process (16), this asymptotic behavior reduces to $f(\lambda) \sim C(\lambda)\lambda^{-2d'}$, as $\lambda \rightarrow 0^+$. Now, for the processes (15) and (18), the asymptotic behavior of the autocorrelation function is

$$\rho(\tau) \sim \tau^{2d-1} \cos(\tau\lambda_0), \quad (20)$$

when $\tau \rightarrow \infty$ and λ_0 is the G frequency. For the model (16) $\cos(\tau\lambda_0)$ in (20) reduces to a constant $C(d')$ which depends only on d' and $2d$ is replaced by $2d'$. Thus, all these models, when they are second order stationary, are long memory in the covariance sense and in the spectral density sense with a rate of convergence $2d$ or $2d'$.

The variance $\gamma_X(0)$ of the process $(X_t)_t$ defined by (16) is finite and positive and:

$$\frac{Var[S_T]}{\gamma_X(0)} = 1 - \rho_X(1),$$

with $\rho_X(1) = \frac{-d}{d+1}$, where $\rho_X(1)$ is the lag 1 autocorrelation for $(X_t)_t$. Thus, $\frac{Var[S_T]}{\gamma_X(0)}$ approaches 0 as d' decreases to 1/2 and $Var[S_T] = O(T^{2d'-1})$. This means that the FARMA process, which is long memory in the spectral sense with a convergence rate $2d'$, is also long memory in Allan's sense. Indeed, For the process (16), we observe the dependence on the parameter d' of the growth of the partial sums. The variance of the partial sums of a white noise

process ($d = 0$) grows at a linear rate T because each random shock is uncorrelated with the others and only adds its own variance of the partial sum. In the case of the FARMA process, if $0 < d' < 1/2$, the shocks are positively correlated and the variance of the sum grows faster than the variance of a single shock. It is in that sense that it produces long memory behavior. The processes (15) and (18) are not long memory in the Allan variance sense because $Var[S_T] = O(T)$, but a specification in terms of a generalization of Allan variance has been recently proposed by Collet and Guégan (2004) which provides the same kind of behavior for these processes.

If in equations (16) and (18), the strong noise $(\eta_t)_t$ is a Gaussian noise, and as soon as the variance of the processes is positive and finite, then these models are long memory in distribution, see Guégan and Ladoucette (2001). Actually, we cannot say that the general class of k -factor GIGARCH processes (15) is long memory in distribution. Indeed, if these processes are not regularly linear, the non-mixing property is not proved.

Now, the long memory behavior described by the previous models can be amplified and misinterpreted, in a non-stationary setting. For instance, let us consider the FARMA process (16).

- If $(\eta_t)_t$ is a centered white noise then, $E[X_t] = 0$, and

$$Var[X_t] = \begin{cases} c & t \rightarrow \infty \quad 0 < d' < 1/2 \\ \rightarrow c \ln t & d' = 1/2 \\ \rightarrow ct^{2d'-1} & d' > 1/2. \end{cases} \quad (21)$$

Then this process is non-stationary in variance for $d' \geq 1/2$ and also in covariance sense.

- If $(\eta_t)_t$ is a non-centred white noise with mean m then, $E[X_t] = mt^{d'}$, and this process is non-stationary in mean as soon as $m \neq 0$ but it is also long memory in the covariance and spectral senses with the rate $2d'$. Thus, it appears difficult to distinguish these two characteristics, from real data.
- Let $(X_t)_t$ an ARIMA(0,1,1) process and $(Y_t)_t$ the process defined, $\forall t$, by $Y_t = X_t - X_{t-1}$. Its spectral density function is equal to

$$f_Y(\lambda) = h(1 - \theta^2 - 2\theta \cos \lambda),$$

with $|\theta| < 1$, the moving average parameter and h a constant, $0 < h < \infty$. The process $(Y_t)_t$ is a power law long memory process of

order $d = -1$. It has an improper density function. Hence, f_Y obeys a power law in the limit of order $d = -1$, which is quite different from what spectral analysis indicates that it should be. Nonetheless an ARIMA(0,1,1) model captures the main feature of a long memory process such as (16), namely a spectral density function that diverges to infinity as $|\lambda| \rightarrow 0$. Thus, it is a long memory process in the spectral sense.

- It is also possible to consider transformations of processes such that (16) which keep long memory behavior, which can add some confusion in identification theory. For FARMA processes, Dittman and Granger (2002) use Hermite polynomial developments and consider the transformation $g(\cdot)$ defined by

$$g(x) = g_0 + \sum_{j=J}^k g_j H_j(x), 1 < J < k < \infty, g_j \neq 0. \quad (22)$$

The polynomial $H_j(x)$ are Hermite polynomials. If $(X_t)_t$ is defined by (16), with $0 < d' < 1/2$, then the process $(g(X_t))_t$, is a long memory process such that its autocorrelation function decreases hyperbolically

$$\gamma_{g(X)}(\tau) = \frac{\Gamma(1 - d')}{\Gamma(d')} \tau^{2d'-1}, \text{ for large } \tau. \quad (23)$$

Now, if $-1 < d' < 0$, then the process $(g(X_t))_t$, is a short memory process. We do not observe in that latter case antipersistence like with the FARMA process, when $-1 < d' < 0$ in (16). In theory the property of long memory is theoretically lost by nonlinear transformation, but in practice this long memory effect is observed and can be larger with even functions $g(\cdot)$ than for odd functions.

In a lot of empirical studies, the presence of strong dependence between transformations of returns have been observed although the returns are generally considered short memory with presence of heteroscedasticity. It is well known that the GARCH model (Engle (1982), Bollerslev (1986)) does not take into account some persistence observed via the autocorrelations function of the returns. In a first step the persistence has been modelled using the IGARCH process. Historically two reasons exist to reject this model. The first is because the second moment of the theoretical IGARCH model is infinite or, in practice, the data have no infinite variance, see Mikosch and Starica (1998). The second is because the modelling of IGARCH process induces non-stationarity in the time series which creates switches. Thus, it

appears necessary to consider other models to explain the existence of persistence in variance, i.e., the degree to which past volatility explains current volatility in order to model significant correlation between the present volatility and remote past volatilities. Empirically, it appeared that the effect of shocks on the conditional variance was very persistent though eventually absorbed with time. That is, sample autocorrelations in square returns tend to decline very slowly in contrast to the fast exponential decay implied by standard GARCH-type models.

This is, in turn consistent with the notions of long memory described in the previous Section when theoretical covariances are not summable, or when, alternatively, the power spectrum is unbounded at zero frequency. For these reasons, several processes have appeared with both heteroscedasticity and long memory characteristics, like the model (15). This last model permits one to take into account such behaviors as soon as we use, for the random variables X_t , any transformation of the returns (square, powers of absolute value). With this latter model, we begin to identify or estimate long memory behavior and then we take into account the short memory behavior. This way seems natural in the sense that, as soon as long memory exists in data, it is impossible to separate the characteristics of short memory from those of long memory looking at the autocorrelation function.

In a different way Ding, Granger and Engle (1993) suggest to use the following model for the returns X_t :

$$X_t = \sigma_t \eta_t, \quad (24)$$

where

$$\sigma_t = (I - (I - B)^d) |X_t|, \quad (25)$$

with $0 < d < 1/2$. Then, the spectral density of the process $(|X_t|)_t$ has the property (3) whereas the spectral density of the process $(X_t)_t$ is bounded. The function $\gamma_{|X_t|}(\tau)$ remains strictly positive for many values of τ . It can be shown, that if one defines $\gamma_{|X_t|^\delta}(\tau)$, then, for a wide range of δ values, the series $(|X_t|^\delta)_t$ is long memory too. We refer, for more details, to Granger and Ding (1995) and Avouyi-Dovi, Guégan and Ladoucette (2002), for instance. This means that long memory effects are more important empirically than the variance. In a certain sense, it seems that **if** the process $(X_t)_t$ presents a long memory behavior in the covariance sense with rate of convergence, then the volatility process $(\sigma_t)_t$ will present the same behavior. Now, if we assume that $E[\eta_t] = m$, then $Var[|X_t|] \sim mt^d$ and the variance of the volatility will have a trend. This can be confused with the long memory behavior in covariance.

Following this work, several attempts have been made to construct long memory models taking into account, at the same time, heteroscedasticity and long memory behavior. We will focus on the ARCH (∞) model which has been largely documented and which includes a lot of heteroscedastic models. We assume here that the returns $(X_t)_t$ are explained by the equation (24) with

$$\sigma_t^2 = a + \sum_{j=1}^{\infty} b_j X_{t-j}^2 \quad (26)$$

where $a \geq 0, b_j \geq 0, j = 1, 2, \dots$. We consider the square of σ_t and X_t , but we can consider any transformation as σ_t^δ or $|X_t|^\delta$, $\delta > 0$. The interest of the model defined by (24) and (26) lies on the behavior of its covariance structure which can be quite rich. The behavior of the covariance function will depend on the rate of decay of the sequence $(b_j)_j$ which controls its asymptotic behavior. The exponential decay of $(b_j)_j$ implies the exponential decay of the covariance function of the sequence $(X_t^2)_t$. Now if the weights b_j satisfy

$$b_j \sim c j^{d-1}, 0 < d < 1/2, \quad (27)$$

for some $c > 0$, then the covariance function of $(X_t^2)_t$ also decays hyperbolically (we do not get this result if we use σ_t and X_t in (26)). This means that the covariance structure is close to long memory in structure. Now, there exists some contradiction between the conditions which assure the existence of a stationary solution for the model defined by (24) and (26) and its possible long memory behavior, see Giraitis, Robinson and Surgailis (1999) and Giraitis, Kokoszka, Leipus and Teyssi ere (2000). They show that a sufficient condition for covariance stationarity of the X_t^2 rules out long memory. This result is strengthened by Zaffaroni (2004) who shows that covariance stationarity of the X_t precludes long memory in the X_t^2 . This means that if we want to model both long memory and heteroscedasticity, we need to be careful. Thus, it seems preferable to use the LM(q)-ARCH model proposed by Ding and Granger (1996), as soon as the conditions on the coefficients are modified, or the martingale representation of the ARCH models proposed by Giraitis, Kokoszka and Leipus (2000) or the model (15) under restrictive conditions. On the other hand, the strong dependence in the conditional variance has also been modelled by the long memory stochastic volatility model, see for instance Breidt, Crato and de Lima (1998) and Robinson (2001). All these models are long memory in the covariance sense. Finally, Starica and Granger (2003), studying the absolute value of log returns of SP 500 on a very long period, using a step function to explain the variance dynamics of the data. They say that this approach seem sufficient to explain most of the

dependence structure present in the sample autocorrelation function of long absolute returns.

We can summarize the results presented in this Section saying that most of the models which include fractional dynamics in their expression provide long memory behavior in the covariance and in the spectral senses with rate of convergence. Some exhibit long memory behavior in Allan's sense and few in distribution sense. Very few exhibit all these behaviors. The presence of non stationarity can add confusion and heteroscedatic behavior is very complex to analyse jointly with long memory dynamics.

5 Models with infrequent shocks or switches

There is plenty of evidence throughout economics of structural changes, time varying parameters or regime switches, so it is interesting to ask if such changes occur with long memory models and what are the effects. The results presented in this section suggest that, asymptotically, models with few structural changes can exhibit specific long memory behavior.

A latent question concerning the structural models is to know if the underlying shocks remain transitory or not. Then it is possible to create models in which the long run impact of the shock is time varying and stochastic, and not only transient or permanent. Indeed, the concept of varying the permanent impact of shocks is linked to the familiar topic of structural change. Whatever the impact of a shock, it is possible to interpret it as a specific type of structural break. We can also formalize this question asking if each permanent shock occurs every period with small variance or if it occurs infrequently with random arrival and large variance. For instance, the stationary ARMA processes have no breaks and the random walk can present a break every period. The SETAR processes introduced by Tong and Lim (1980) seems to have no breaks. In these last models, the parameters change values, which causes the innovations to decay at different rate, but nonetheless they remain transitory unless a great percentage of close points stays in a regime. Some authors claim that long memory and structural changes can be easily confused. We will specify, in the following, that it depends on the notion of long memory we consider.

To create breaks and then persistence, a device is to oblige some parameters in the model, such as mixture probabilities, vary with the sample size T . The easiest model that we can build is the Bernouilli $(v_t)_t$ process defined in

the following way:

$$v_t = \begin{cases} 0 & \text{wp } 1-p \\ w_t & \text{wp } p \end{cases} \quad (28)$$

where $(w_t)_t$ is a strong Gaussian white noise $N(0, \sigma_w^2)$. From equation (28), we get:

$$\sum_{t=1}^T v_t = \begin{cases} 0 & \text{wp } 1-p \\ \sum_{t=1}^T w_t & \text{wp } p, \end{cases} \quad (29)$$

and

$$\text{var}\left[\sum_{t=1}^T v_t\right] = pT\sigma_w^2 = O(T). \quad (30)$$

Thus, the Allan variance of this process increases with the sample size T . In that case, this process does not exhibit long memory behavior. Now, assume that p is not constant but changes appropriately with sample size. We define:

$$(H_0) : p = O(T^{2d-1}), \text{ with } 0 < d < 1. \quad (31)$$

Then under the assumption (H_0) , the expression (30) becomes:

$$\text{var}\left[\sum_{t=1}^T v_t\right] = O(T^{(2d-1)+1}), \quad (32)$$

and in that latter case, the partial sum's variance grows consistently with parameter d and the process $(v_t)_t$ has long memory behavior in the Allan's sense. This does not mean that it has long memory in the covariance sense.

Following the above key idea to let the parameter p - which represents the probability to be in a specific state - to decrease with sample size for creating few breaks in the series and then long memory behavior, we can consider the following general class of models. Let $(X_t)_t$ be a process whose recursive scheme is

$$X_t = \mu_{s_t} + \varepsilon_t, \quad (33)$$

where $(\mu_{s_t})_t$ is a process we specify below and $(\varepsilon_t)_t$ a strong white noise, independent to $(\mu_{s_t})_t$. We can distinguish two cases:

1. If $\mu_{s_t} = \mu_t$ then we assume that this process depends on a probability p .
2. If μ_{s_t} depends on an hidden ergodic Markov chain $(s_t)_t$, it is characterized by the transition matrix P of $(s_t)_t$, whose elements are the fixed transition probabilities p_{ij}

$$P[s_t = j | s_{t-1} = i] = p_{ij} \quad 0 \leq p_{ij} \leq 1, \quad (34)$$

with $\sum_{j=0}^1 p_{ij} = 1$, $i = 0, 1$ (here, we assume the existence of two states).

Then, the transition matrix which characterizes the model (33) is,

$$P = \begin{pmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{pmatrix}. \quad (35)$$

The mean stationary switching model defined by (33) - (35) has been introduced, first, by Quandt (1958) then reconsidered by Neftçi (1982, 1984) and popularized later in economics by Hamilton in 1988. This model involves multiple structures that can characterize the time series behaviors in different regimes. By permitting switching between these structures, the model is able to capture complex dynamic patterns, observed in various economic and financial data sets. Different applications have been considered in financial domain to detect existence of low and high volatilities periods.

The first class of models considered above includes "models with breaks" and the second one "models with switches". SETAR models are built in a different way as previous ones but can also create switching and, in certain cases, long memory behavior. We now characterize these three classes of models with respect to their different long memory behavior.

5.1 Models with breaks

In this subsection, first we consider models which create some permanent breaks by their own structure, then we introduce models which need specific assumptions to transform their "natural" short memory behavior in long memory behavior. These models have been introduced first by Balke and Fomby (1991) then investigated by Diebold and Inoue (2001) and Breidt and Hsu (2002).

We assume that the process $(X_t)_t$ is defined by,

$$X_t = \mu_t + \varepsilon_t, \quad (36)$$

and we introduce several dynamics for the process $(\mu_t)_t$ which appears in (36):

- The Binomial model. The dynamics for the process $(\mu_t)_t$ are, $\forall t$:

$$\mu_t = \mu_0 + \sum_{j=1}^t q_j \eta_j, \quad (37)$$

with

$$q_j = \begin{cases} 0 & \text{wp } 1-p \\ 1 & \text{wp } p, \end{cases} \quad (38)$$

where $(\varepsilon_t)_t$, $(\eta_t)_t$ and $(q_t)_t$ are independent sequences. This model implies sudden changes only inside the process $(X_t)_t$. Structural changes may occur gradually. If we denote σ_η^2 the variance of the process $(\eta_t)_t$, then the covariance of the process $(X_t)_t$ is equal to, $\forall t$:

$$\gamma_X(\tau) = tp\sigma_\eta^2. \quad (39)$$

Granger and Hyung (1999) show that - with respect to the value of p - the process defined by (36) - (38) presents asymptotically short memory behavior ($p \rightarrow 0$) or some kind of persistency ($p > 0$). The observed long memory depends on p and can be long memory in covariance sense.

- The mean-plus-noise model. The process $(\mu_t)_t$ is such that, $\forall t$:

$$\mu_t = (1-p)\mu_{t-1} + \sqrt{p}\eta_t, \quad (40)$$

where $(\varepsilon_t)_t$ and $(\eta_t)_t$ are independent noise. The parameter p denotes the persistence of the level component μ_t . If p is small, then the level varies slowly. To prevent the variance of μ_t from blowing up as p goes to zero, the innovation is scaled by \sqrt{p} . This process with small p is highly dependent mean-reverting process and it exhibits persistence, see Chen and Tiao (1990). A generalization of this model is discussed in Breidt and Hsu (2002) and Smith (2003).

- The random walk model with a Bernoulli process. The dynamics of $(\mu_t)_t$ is the following, $\forall t$:

$$\mu_t = \mu_{t-1} + v_t, \quad (41)$$

where the process $(v_t)_t$ is the Bernoulli process defined by the equation (28). If the probability p in (28) verifies the assumption (H_0) , then the partial sums' variance of the process defined by (36) and (41) is equal to (32) and the process $(X_t)_t$ exhibits long memory behavior in the Allan's sense.

- The stop-Break process. The process $(\mu_t)_t$ in (36), is such that, $\forall t$:

$$\mu_t = \mu_{t-1} + \frac{\varepsilon_{t-1}^2}{\gamma_T + \varepsilon_{t-1}^2} \cdot \varepsilon_t. \quad (42)$$

This model has been first introduced by Engle and Smith (1999). In their paper, they are mainly interested by the effect of a permanent shift and not by the long memory behavior. Now as one can show that the covariance γ_X is such that $\gamma_X = O(T^\delta)$, $\delta > 0$, then the variance of the partial sums of the process $(X_t)_t$ defined by the equations (36) and (42) verifies also the equation (32) and this model has a long memory behavior in the Allan's sense, see Diebold and Inoue (2001).

- The stationary random level shift model. Here, the process $(\mu_t)_t$ in (36) depends on the Markov chain $(s_t)_t$ introduced in (34) and (35). Then, $\forall t$ we have:

$$\mu_t = (1 - ks_t)\mu_{t-1} + s_t\eta_t, \quad (43)$$

where $k \in [0, 2)$ and $(s_t)_t$ is independent from (ε_t) and $(\eta_t)_t$. This model provides a framework to assess the performance of a standard time series method on series with level shifts. If the probabilities p_{ij} , $i, j = 1, 2$, of the transition matrix P in (35) verify the assumption (H_0) , then this process has long memory behavior in the spectral sense with a rate equal to $2d$, see Chen and Tiao (1990), Liu (2000), Breidt and Hsu (2002) and Smith (2003).

5.2 Models with switches

Here, we assume that the process $(X_t)_t$ follows the model (33). A process defined by (33) is a short memory process. Its autocovariance function decreases fairly rapidly towards zero, see for instance Andel (1993). Nevertheless it is possible to show empirically that this process has long memory behavior in the spectral sense with a rate of convergence $0 < d < 1/2$, see Guégan and Rioublanc (2003). This means that we can adjust simulated data based from (33) models such that (18) holds. Now, we present two models with switches which exhibit some kind of long memory behavior.

- Diebold and Inoue (2001) consider a process $(X_t)_t$ defined by (33) with a Markov chain whose transition's probabilities satisfy the following assumption:

$$(H_1) : p_{00} = 1 - c_0T^{-\delta_0}, \quad p_{11} = 1 - c_1T^{-\delta_1},$$

where δ_0 and δ_1 are positive. They show that, under this assumption,

$$var\left[\sum_{t=1}^T X_t\right] = O(T^{\max(\min(\delta_0, \delta_1) - |\delta_0, \delta_1| + 1, 1)}), \quad (44)$$

thus the process defined by (33) has a long memory behavior in Allan's sense.

- Another model with switches which does not use a hidden Markov chain - and which creates breaks - is the sign's model introduced in 1999 by Granger and Terasvirta, see also Granger, Spear and Ding (2000). it is defined, $\forall t$, by

$$X_t = \text{sign}(X_{t-1}) + \varepsilon_t, \quad (45)$$

where $(\varepsilon_t)_t$ is a strong white noise $N(0, \sigma^2)$. The sign function is given by $\text{sgn}(x) = 1$ if $x > 0$, $\text{sgn}(x) = 0$ if $x = 0$ and $\text{sgn}(x) = -1$ if $x < 0$. The process $(X_t)_t$ defined by (45) is Markovian with mean zero and autocorrelation function $\rho_X(\tau) = (1 - 2p)^\tau$, where $p = P[\varepsilon_t < -1] = P[\varepsilon_t > 1]$. Thus, this process appears as a short memory process. It produces switches in the mean, as a classical switching model, as soon as, p is small. Now the authors have remarked that the plots of $\frac{\ln \hat{\rho}_X(\tau)}{\ln \tau}$ increase with τ , where $\hat{\rho}_X$ represents the empirical autocorrelation function of $(X_t)_t$. This means that, empirically, this process has a behavior close to long memory behavior in the covariance sense. This phenomenon seems amplified for small values of p . The same behavior occurs even if the process (45) is not generated by a Gaussian noise.

5.3 Models with thresholds

SETAR models can also create jumps from one state to another one even if their structure is completely different from the previous models. Tong and Lim (1980), see also Tong (1990), introduced the famous SETAR model which allows shifts from one state to another one thanks to the existence of a threshold based on the internal structure of the process itself. In that case the shifts in regime are assumed to be directly observable. In the SETAR model the shift from one state to another one is based on the use of a discontinuous function and Terasvirta and Anderson (1992) suggest to use a continuous function. Then, the SETAR model became STAR models, see also van Dijk, Franses and Paap (2002) for a recent review on this last class of models. Several variants of the SETAR model to fit specific dynamics have also be developed. We can cite the SETAR-ARCH model, introduced by Zakoian (1994) and Li and Li (1996) which allows heteroscedastic behavior on each state, the Threshold Stochastic Volatility model (SVM) introduced by So, Li and Lam (2002) which permits SVM on each state and the SETAR model with long memory dynamics on the states introduced by Dufrénot, Guégan and Peguin-Feissolle (2003), for instance. The introduction of a threshold on the conditional heteroscedasticity which characterizes some of the previous

models has been proposed by Pfann, Schotman and Tchernig (1996).

Consider the process $(X_t)_t$ defined by the following general form:

$$\begin{cases} X_t = f(X_{t-1}, \dots)(1 - G(X_{t-d}, \gamma, c)) \\ \quad + g(X_{t-1}, \dots)G(X_{t-d}, \gamma, c) + \varepsilon_t, \end{cases} \quad (46)$$

where the functions f and g can be any linear or non-linear functions of the past values of X_t or ε_t . The process $(\varepsilon_t)_t$ is a strong white noise and G an indicator function or some continuous function. For a given threshold c and the position of the random variable X_{t-d} with respect to this threshold c , the process $(X_t)_t$ follows different models. In its formulation, the model (46) necessitates, for instance:

1. The stationary SETAR(2,2,1) process when G is the indicator function $I_{X_{t-d} > c}$, defined as $I_A = 1$ if A is true and $I_A = 0$ otherwise, $f(X_{t-1}, \dots) = \phi_{0,1} + \phi_{1,1}X_{t-1}$ and $g(X_{t-1}, \dots) = \phi_{0,2} + \phi_{1,2}X_{t-1}$. This model has been introduced by Tong and Lim (1980). On each state, it is possible to propose more complex models like ARMA(p, q) processes (Brockwell and Davis, 1988), bilinear models (Guégan, 1994) or GARCH(p, q) processes (Bollerslev, 1986). Changes on the variance can be also considered, see Pfann, Schotman and Tchernig (1996).
2. The STAR model is obtained using for G a continuous function like the logistic,

$$G(X_{t-d}, \gamma, c) = \frac{1}{1 + \exp(-\gamma(X_{t-d} - c))}. \quad (47)$$

Note that the transition function G is bounded between 0 and 1. For this model, we consider the same functions f and g as before. The parameter c can be interpreted as the threshold between the two regimes in the sense that the logistic function changes monotonically from 0 to 1 with respect to the value of the lagged endogenous variable X_{t-d} . The parameter γ determines the smoothness of the change in the value of the logistic function, and thus, the smoothness of the transition of one regime to the other. As γ becomes very large, the logistic function (47) approaches the indicator function $I_{X_{t-d} > c}$. Consequently, the change of $G(X_{t-d}, \gamma, c)$ from 0 to 1 becomes instantaneous at $X_{t-d} = c$. Then we find the SETAR model as a particular case of this STAR model. When $\gamma \rightarrow 0$, the logistic function approaches a constant (equal to 0.5) and when $\gamma = 0$, the STAR model reduces to a linear AR model. This STAR model has been described by Terasvirta and Anderson (1992), see also van Dijk, Franses and Paap (2002). We can use also time varying coefficients, see Lundbergh, Terasvirta and van Dijk (2003).

3. The SETAR process with long memory dynamics. The equation (46) becomes:

$$(I - B)^d X_t = (\phi_{0,1} + \phi_{1,1} X_{t-1})(1 - G(X_{t-d}, \gamma, c)) \quad (48)$$

$$+ (\phi_{0,2} + \phi_{1,2} X_{t-1})G(Y_{t-d}, \gamma, c) + \varepsilon_t. \quad (49)$$

When G is a continuous function this model has been introduced by van Dijk, Franses and Paap (2002) and when $G = I_{t-d > c}$ and it has been investigated by Dufrénot, Guégan and Peiguin-Feissolle (2003).

The second order stationary model (46) is short memory as soon as the functions f and g correspond to short memory processes. The long memory behavior is present as soon as a long memory process appears in one state like in the process (48). Now to get long memory behavior when short memory processes characterize each state, we need to investigate the conditions established to obtain the stationarity of the model. For instance, for the SETAR(2,2,1) described previously, stationarity is achieved under the following conditions, see Chan (1993).

1. A sufficient condition for stationarity is: $\max |\phi_{1,1}|, |\phi_{1,2}| < 1$.
2. Necessary and sufficient conditions for stationarity are:
 - $\phi_{1,1} < 1, \phi_{1,2} < 1, \phi_{1,1}\phi_{1,2} < 1$,
 - $\phi_{1,1} = 1, \phi_{1,2} < 1, \phi_{0,1} > 0$,
 - $\phi_{1,1} < 1, \phi_{1,2} = 1, \phi_{0,2} > 0$,
 - $\phi_{1,1} = 1, \phi_{1,2} = 1, \phi_{0,2} < 0 < \phi_{0,1}$,
 - $\phi_{1,1}\phi_{1,2} = 1, \phi_{1,1} < 0, \phi_{0,2} + \phi_{1,2}\phi_{0,1} > 0$.

We can remark that these conditions concern mainly the autoregressive parameters. In another hand, we can observe that, even if the process is globally stationary, non-stationary behavior can appear on one regime which can be confused with some long memory behavior.

Now, we consider a particular case of the process (48) defined $\forall t$, by

$$X_t = (1 - B)^{-d} \varepsilon_t^{(1)} I_t(X_{t-d} \leq c) + \varepsilon_t^{(2)} [1 - I_t(X_{t-d} \leq c)]. \quad (50)$$

For this model, the autocovariance function and the spectrum density depend on the regime-shift variable. The "mixture" of a white noise process and of a fractional white noise process produces a memory structure that is function of the distribution function of the variable X_{t-d} across the two regimes at

different dates. If regime 2 is more frequently visited by the observations than regime 1, then this will imply some difficulties to find long-memory dynamics. In that case, the autocovariance and spectrum will exhibit a shape resembling that of a short-memory process. In the opposite case, the autocovariance function and the spectrum exhibit the usual properties of long-memory processes : a slow decay and high values at frequencies near zero. The key parameter here is the threshold c that determines the distribution function of the observations across the two regimes.

Asymptotically, long-memory behavior dominates: the spectrum f_X becomes infinite at the zero frequency and the autocovariance function γ_X is not summable, then:

$$\gamma_X(\tau) \sim \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\tau^{2d-1}, \quad \text{as } \tau \rightarrow +\infty \quad (51)$$

and

$$f_X(\lambda) \sim C\lambda^{-2d}, \quad \text{as } \lambda \rightarrow 0, \quad (52)$$

where C is a positive constant. This model is then long memory in the covariance and the spectral senses with a rate of convergence $2d$.

5.4 Some comments

The previous analysis concerning some specific models with structural breaks suggests that, under certain conditions (more or less plausible), on "small" amounts of structural change, long memory and structural change may be confused. The device proposed in Diebold and Inoue (2001) and Breidt and Hsu (2002) to put certain parameters, such as mixture probabilities, vary with T is a thought experiment to build long memory behavior. The theory suggests that confusion with long memory behavior in Allan's variance will result when only a small amount of breakages occurs, and therefore that the larger is T , the smaller must be the break probability. Thus, the long memory behavior appears asymptotically. In finite samples, short memory break models may be very difficult to distinguish from true long memory models. When the breaks occur each period (this means with a fixed probability), then it seems impossible to confuse this behavior with long memory behavior.

Looking at subsection 5.2, we see that, under specific conditions on the transition matrix, the switching model exhibits long memory in Allan's sense. But long memory is observed empirically, independently of the null assumption (H_0) introduced in (31). In another hand a very simple short memory

model like the sign model exhibits empirically long memory in covariance sense. The models introduced in Subsection 5.3 can exhibit long memory behavior in the covariance sense if we have non-stationarity on one regime or long memory dynamics. In any case the percentage of points in those regimes need to be high to distinguish this behavior, specifically in the non-stationary case.

6 Aggregation of dynamic equations

In 1980 Granger was probably one of the first, in economy, who points out the specific properties obtained by aggregating dynamic equations. Indeed, if these models are found to arise in practice, they have proved useful in improving long run forecasts in economics and also in finding stronger distributed lag relationships between economic variables. We consider here two specific aggregations which produce long memory behavior in the covariance sense.

- The aggregate beta AR(1) process. This model proposed by Granger (1980) rests on the fact that many of the important macroeconomic variables are aggregates of a very large number of micro-variables: total personal income, unemployment, consumption of non-durable goods, inventories and profits, as just a few examples. If we denote $(X_{jt})_t$, these different series, then we consider the aggregate series $(X_t)_t$ defined as:

$$X_t = \sum_{j=1}^n X_{jt}, \quad (53)$$

where, for example, each $(X_{jt})_t$ is generated by an AR(1) model, such that

$$X_{jt} = \phi_j X_{j(t-1)} + \varepsilon_{jt}. \quad (54)$$

Now, if we assume that the parameters ϕ_j follow a beta distribution on the range $(0, 1)$ whose density is defined by:

$$f(y) = \frac{2}{B(p, q)} y^{2p-1} (1 - y^2)^{q-1}, \quad 0 \leq y \leq 1, \quad (55)$$

and $f(y) = 0$ elsewhere, then the asymptotic behavior of the autocovariance function for the process (53) - (55) is such that:

$$\gamma_X(\tau) = C\tau^{1-\frac{q}{2}}, \quad q > 1, \quad \tau \rightarrow \infty. \quad (56)$$

Then, this aggregate process $(X_t)_t$ has a long memory behavior in the sense of Parzen with a rate $1 - \frac{q}{2}$ if $q > 1$. This property can be extended to model (53) when the parameters ϕ_j are drawn from the probability density function $f(x) = cx(1-x^2)$, $x \in [0, 1]$, see Lin (1991). Other examples of distribution functions which permit this long memory behavior can be found in Lukacs (1970). Ding and Granger (1996) generalize the process (53) - (55) using an ARCH or a GARCH on each component (54) instead of a AR(1) model. They exhibit the same long memory behavior for such model $(X_t)_t$.

- The aggregate long memory switching model. Granger and Ding (1996) propose to make switches on the long memory parameter d . They consider two processes (X_{1t}) and (X_{2t}) assuming that they follow the recursive scheme

$$(I - B)^{d_i} X_{it} = \varepsilon_{it}, \quad i = 1, 2. \quad (57)$$

Then, they define the process $(X_t)_t$ such that:

$$X_t = s_t X_{1t} + (1 - s_t) X_{2t}, \quad (58)$$

where $(s_t)_t$ is the Markov chain introduced in (34) and (35). Now if $p_{ij}^\tau = P[s_t = i | s_{t+\tau} = j]$, $i, j = 0, 1$, we get:

$$E[X_{t+\tau} X_t] = \sum_{j=0}^1 p_{jj}^\tau E[X_{it+\tau} X_{it}] + \sum_{i,j=0,1, i \neq j} p_{ij} E[X_{it+\tau} X_{jt}].$$

Then the autocovariance function of the process $(X_t)_t$, is just a weighted average of the autocovariance of the series in the two regimes. Clearly, a variety of correlogram shapes can arise, producing different long memory processes, in the covariance sense.

7 The use of statistical tools to create long memory

The previous study questions the way to identify or estimate all these long memory behaviors. Are they really long memory behavior or "spurious" long memory behavior? When data come from a fractional stationary model such as (15), then methods are available to estimate this long memory behavior, see Ferrara and Guégan (2001), and Diongue and Guégan (2003) and a lot of references therein. But, the field of long memory detection and estimation is

particularly (in)famous for the numerous statistical instruments that behave similarly under the assumptions of large range dependence and stationarity, or under weak dependence affected by some type of non-stationarity. Different statistics used in the detection and estimation of long memory are characterized by their lack of power to discriminate between possible scenarios, in particular presence or not of true long memory behavior. The first one is the R/S statistic, which has the same kind of asymptotic behavior when applied to a stationary long memory time series or a short time series perturbed by a small monotonic trend that even converges to 0 as time goes to infinity, see Bhattacharya, Gupta and Waymire (1983). The second one deals with the sample variance of the time series at various levels of aggregation. Here, we define it as

$$X_t^m = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \quad m \geq 1, \quad k = 1, \dots, [n/m]. \quad (59)$$

Teverovsky and Taqqu (1997), see also Bisaglia and Guégan (1997), showed that this estimator performs similarly when applied to a long memory stationary time series or to a stationary short memory one that was perturbed by shifts in the mean or small trends. This could make cause confusion in detecting existence of long range dependence in the data when in reality a jump or a trend is added to a series with no long range dependence. Some methods to estimate d lead to bias estimators. It is the case for the Geweke-Porter-Hudak method in certain cases. For instance, Smith (2003) shows that if a process is generated by a random level shift model defined by (41) and estimated as a long memory FARMA process, the estimate for d , using the classical Geweke-Porter-Hudak (GPH) method is biased and thus we estimate some kind of "spurious" long memory behavior. This GPH estimate often erroneously indicates the presence of long memory. Finally, we refer to some kinds of tests which detect long memory inside data even if they are simulated with known short memory processes, see de Lima and Crato (1994) for examples.

Now the question is: even if the theoretical model does not present long range dependence and as soon as we observe presence of long range dependence in the reality, what about forecasts?

The detection of long range dependence effects is based on statistics of the underlying time series, such as the sample autocorrelation function, the periodogram and as mentioned before the R/S statistic, the sample variance for aggregated time series, etc. The assumptions that the data are station-

ary and long range dependence imply a certain characteristic behavior of these statistics. The detection of long range dependence is reported when the statistics seem to behave in the way prescribed by the theory. However, as mentioned above, the same behavior can be observed for short memory time series perturbed by some kind of non stationarity. Mikosch and Starica (1999) show for instance that - if the expectations of subsequences $(X_t^j)_t$ of the process $(X_t)_t$ are different - the sample autocorrelation function, for sufficiently large τ , approaches a positive constant. Then, they show that the periodogram becomes arbitrarily large for Fourier frequencies close to zero. Then, this behavior can mislead one in inference on long range dependence when in fact one analyses a non-stationary time series with subsamples that have different unconditional moments.

Then, a general result is the following: Consider a strict stationary process $(X_t)_t$ and for a sample size T , consider r subsamples $(X_t^1)_t, (X_t^2)_t, \dots, (X_t^r)_t$. The i th-subsample comes from an stationary ergodic model with 2nd moment and spectral density f_{X^i} . Clearly, if the subsamples have different marginal distributions, the resulting sample (X_1, \dots, X_T) is non-stationary. Now if we associate with each subsample $(X_t^j)_t$ a positive number p_j , such that $\sum_{j=1}^r p_j = 1$, then if the expectations differ in the subsequences $(X_t^j)_t$ and if the autocovariances $\gamma_{X_t^j}(\tau)$ decay to zero exponentially as $\tau \rightarrow \infty$, the empirical autocovariance $\hat{\gamma}_{X_t^j}(\tau)$ for sufficiently large τ is close to a strictly positive constant given by $p_i p_j (E[X_t^j] - E[X_t^i])^2$, for $1 \leq i < j \leq r$. Then, we can show that the periodogram becomes arbitrarily large for various small values of the frequencies as $T \rightarrow \infty$. This kind of result can apply on GARCH(1,1) process. Then, it seems that the non-stationarity of the unconditional variance is a possible source of both the slow decay of the sample autocorrelation function and the high persistence in the volatility in long log-return time series as measured by ARCH type models. Now, we conjecture that the same phenomenon arises with switching models and some SETAR processes, which empirically exhibit long memory behavior although they are theoretically short memory. This behavior is related to the non stability of the invariant distribution of the underlying process or its heavy tailness behavior.

8 Conclusion

In this paper, we have reviewed different ways to characterize the concept of long memory behavior. We have shown that these concepts are not equivalent. They are all defined in a non parametric context and then can be dif-

difficult to use directly on real data. Then, we have presented different classes of models known to present long memory behavior. We have specified what kind of long memory behavior they exhibit. Some of these models are long memory in covariance or in spectral senses, other in Allan's sense and some in distribution. As these definitions are not equivalent, care is necessary to decide what model we want to use in order to identify long memory behavior.

Some models are used in a lot in applications when one wants to detect long memory. They correspond to the class of GIGARCH processes and derived models from this general class. Other models are theoretically short memory but empirically their autocovariance function decreases in an hyperbolic way. This fact includes the SETAR, the switching and the sign models. For the first one the empirical behavior is close to non-stationary behavior. For the switching and the sign models the empirical covariance function need to be investigated in more detail.

Now, some other models - whose states depend on different probabilities p - with respect to the properties of this probability and of the sample size, are able to create some long run. But this long memory behavior is not the same as the one discussed for the previous models. In that context, Diebold and Inoue (2001) introduce some artificial assumptions on p , which appear drastic, and we question their interest from a economical point of view. It could be interesting to make forecasts using a FARMA(0,d,0) model on data simulated from the model defined in (36) and (41). It seems that this model is not identifiable. If we use Diebold and Inoue's simulations (2001) we have to test $0 < d < 1$, but not $d = 1$, because this case corresponds to a non stationary behavior and not to a long memory behavior, (see P. 148 and examples before in their paper). Now, the paper of Rydén, Terasvirta and Asbrink (1998) present some counter examples at the approach developed by Diebold and Inoue (2001) showing that long memory is not always observed for the Diebold and Inoue's models.

In another context, other structural change models have also been proposed in the literature, but no work has been done concerning their possible long memory behavior in the previous senses proposed in Section 3. We think for instance to the works of Kim and Kon (1999). In their paper, they introduce a model based on a sequential mixture of Gaussian distributions to model the discrete change points in the series.

One of the points which seems the more important is that some models empirically exhibit long memory even if this one does not exist theoretically.

Thus, emphasis on statistical properties of empirical samples need to be more investigated, mainly the assumptions used to get the robustness of any estimates.

For instance, the presence of long memory behavior in the covariance sense can be easily detected, but as soon as we have heteroscedasticity or some non stationarity, the only use of the empirical covariance or the periodogram (which is the estimate of the spectral density) become problematic. In particular, if the variance of the data is finite, but the 4th moment is not, it has no sense to look at the sample autocorrelation function of the square of the data. Now, heavy-tailedness of models would be also a possible explanation for the slow decay of the autocorrelation function. Thus, the long memory type of behavior observed in the sample autocorrelation function of absolute values and squares of the returns would not be in contradiction to the short memory property of the GARCH process. Hence, a careful investigation of the empirical facts given by data sets has to be done.

For most of the models we have discussed, their long memory is in covariance sense or in Allan sense, few are in distribution. For instance, we do not know if breaks in distribution (as those introduced in sign process for instance) will produce long memory behavior in distribution. To get this kind of result, it is necessary to investigate the possibility of mixing or non mixing property for the models with infrequent breaks or which capture switch like the switching models and SETAR models. Nevertheless, sparsely results exist: Francq and Zakoian (2001), prove mixing for specific switching models and Chan (1993) for specific SETAR models. Then, those models are not long memory in distribution.

This work, which is mainly an exploratory work shows the complexity behind the notion of long memory behavior and the carefulness which is necessary for the use of these models on real data.

Acknowledgments This paper has been written during the stay of the author at the School of Economics and Finance at QUT (Brisbane) and she thanks them for financial support.

References

- [1] D.W. Allan (1966) Statistics of Atomic Frequency Standards, Proc. IEEE, 54, 221 - 230.

- [2] J. Andél (1993) A Time Series Model with Suddenly Changing Parameters, *Journal of Time Series Analysis*, 14, 111 - 123.
- [3] S. Avouyi-Dovi, Guégan D., Ladoucette S. (2002) Une mesure de la persistance dans les indices boursiers, *NER 94*, Banque de France, Paris.
- [4] R. T. Baillie, T. Bollerslev, H.O. Mikkelsen (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 74, 3 - 30.
- [5] R.T. Baillie, C.F Chung, M.A. Tieslau (1996) Analysis inflation by the fractionally integrated ARFIMA-GARCH model, *J. of Applied Econometrics*, 711, 23 - 40.
- [6] N.S. Balke, T.F. Fomby (1991) Shifting trends, segmented trends and infrequent permanent shocks, *Journal of Monetary Economics*, 28, 61 - 85.
- [7] J. Beran (1994) *Statistics for long memory processes*, Chapman and Hall, New York.
- [8] R.N. Bhattacharia, V.K. Gupta, E. Waymire (1983) The Hurst effect under trends, *J. Applied Prob.*, 20, 649 - 662.
- [9] L. Bisaglia, D. Guégan (1997) A review of techniques of estimation in long memory processes: application to intra-day data , *Computational Statistics and Data Analysis*, 61 - 81, 1997,
- [10] T. Bollerslev (1986) Generalized autoregressive conditional heteroscedasticity, *J. of Economics*, 31, 302 - 327.
- [11] G.E.P. Box, G.M. Jenkins (1976) *Time series analysis: forecasting and control*, Holden Day, San Francisco.
- [12] F.J. Breidt, N. Crato, P. de Lima (1998) The detection and estimation of long memory in stochastic volatility, *J; of Econometrics*, 83, 325 - 348.
- [13] F.J. Breidt, N.J. Hsu (2002) A class od nearly long memory time series models, 18, 265 - 281.
- [14] P.J. Brockwell, R.A. Davis (1996) *Introduction to time series analysis*, Springer Texts in Statistics, Springer Verlag, N.Y.

- [15] C. Caillault, D. Guégan (2003) Empirical Estimation of Tail Dependence Using Copulas. Application to Asiatic Markets, NR MORA-IDHE, 05 - 2003, ENS - Cachan, France.
- [16] K.S. Chan (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model, *Annals of Statistics*, 21, 520 - 533.
- [17] C. Chen, G.C. Tiao (1990) Random Level-Shift Time Series Models, ARIMA Approximations and Level-Shift Detection, *Journal of Business and Economic Statistics*, 8, 83 - 97.
- [18] X. Chen, L.P. Haussemann, M. Carrasco (1999) Nonlinearity and temporal dependence, Dept. of Economics, University of Chicago.
- [19] R. Cioczek-Georges, B.B. Mandelbrot (1995) A class of micropulses and antipersistent fractional Brownian motion, *Stochastic Processes and their Applications*, 60, 1-18.
- [20] J.J. Collet, D. Guégan (2004) Another characterization of the long memory behavior, NR MORA-IDHE, 02-2004, ENS Cachan, France.
- [21] D.R. Cox (1977) Contribution to discussion of paper by A.J. Lawrance and N.T. Kottegoda, *Journal of the Royal Statistical Society, A* 140, 34.
- [22] F.X. Diebold, A. Inoue (2001) Long memory and regime switching, *J. of Econometrics*, 105, 131 - 159.
- [23] D. van Dijk, P.H. Franses (1999) Modelling multiple regimes in the business cycle, *Macroeconomic dynamics*, 3, 311 - 340.
- [24] Z. Ding, C.W.J. Granger, R. F. Engle (1993) A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, 1, 83 - 106.
- [25] Z. Ding, C.W.J. Granger (1996) Modelling volatility persistence of speculative returns: a new approach, *Journal of Econometrics*, 73, 185 - 215.
- [26] A.K. Diongue, D. Guégan (2004) Processus GIGARCH à k Facteurs: Estimation et Applications aux Prix Spot de l'Électricité, NR IDHE-MORA, 14-2003, ENSC, France.
- [27] I. Dittman, C.W.J. Granger (2002) Properties of non linear transformations of fractionnally integrated processes, *J. of Econometrics*.

- [28] G. Dufrénot, D. Guégan, A. Peguin (2003) A SETAR model with long-memory dynamics, NR IDHE-MORA, 11-2003, ENSC, France.
- [29] R. F. Engle (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom, *Econometrica*, 50, 987 - 1007.
- [30] R. F. Engle, A.D. Smith (1999) Stochastic permanent breaks, *Review of Economics and Statistics*, 81, 553 - 574.
- [31] L. Ferrara, D. Guégan (2001) Forecasting with k -factor Gegenbauer processes: Theory and Applications, *Journal of Forecasting*, 20, 1 - 20.
- [32] Francq C., Zakoian J.-M. (2001), Stationarity of Multivariate Markov-Switching ARMA Models, *Journal of Econometrics* 102, 339-364.
- [33] Giraitis, L. and Leipus, R. (1995), A generalized fractionally differencing approach in long memory modeling, *Lithuanian Mathematical Journal*, 35, 65 - 81.
- [34] Giraitis L., Robinson P., Surgailis D. (1999) A model for long memory conditional heteroscedasticity, *Ann. Appl. Prob.*
- [35] Giraitis L., Kokoszka P., Leipus R. (2000) Stationary ARCH models: dependence structure and central limit theorem, *Economic Theory*, 16, 2 - 22.
- [36] Giraitis L., Kokoszka P., Leipus R., Teyssière G. (2000) Semiparametric estimation of the Intensity of long Memory in Conditional Heteroskedasticity, *Statistical Inference for Stochastic processes*, 3, 113 - 128.
- [37] C.W.J. Granger (1980) Long memory relationships and the aggregation of dynamic models, *J. of Econometrics*, 14, 22è - 238.
- [38] C.W.J. Granger, R. Joyeux (1980) An introduction to long memory time series models and fractional differencing, *Journal of Time Series Analysis*, 1, 15 - 29.
- [39] C.W.J. Granger, Z. Ding (1995) Some properties of absolute returns : an alternative measure of risk, *Annales d'Economie et de Statistique*, 40, 67 - 91.
- [40] C.W.J. Granger, Z. Ding (1996) Varieties of long memory models, *Journal of Econometrics*, 73, 61 - 77.

- [41] C.W.J. Granger, T. Terasvirta (1999) A simple nonlinear time series model with misleading linear properties, *Economic Letters*, 62, 161 - 165.
- [42] C.W.J. Granger, N. Hyung (1999) Occasional structural breaks and long memory, Discussion paper 99-14, University of California, San Diego, USA.
- [43] C.W.J. Granger, S. Spear, Z. Ding (2000) Stylized facts on the temporal and distributional properties of absolute returns: an update, in "Statistics and Finance: an interface", eds. W.S. Chan, W.K. Li, H. Tong, Imperial College Press, 97 - 120.
- [44] C.W.J. Granger (2002) Time series concepts for conditional distributions, *Oxford Bulletin of Economics and Statistics*, 65, 689 - 701.
- [45] Gray, H.L., Zhang, N.-F. and Woodward, W.A. (1989), On generalized fractional processes, *Journal of Time Series Analysis*, 10, 233-257.
- [46] D. Guégan (1994) *Séries chronologiques non linéaires à temps discret*, Economica, Paris.
- [47] D. Guégan (2000) A new model: the k -factor GIGARCH process, *Journal of Signal Processing*, 4, 265 - 271.
- [48] D. Guégan (2003a) *Les chaos en finance: approche statistique*, Economica, Paris.
- [49] D. Guégan (2003c) A prospective Study of the k -factor Gegenbauer processes with the heteroscedastic errors and an application to inflation rates, *Finance India*, XVII, 1-30.
- [50] D. Guégan, S. Ladoucette (2001) Non-mixing properties of long memory processes, *C.R.A.S., Série I*, 333, 373 - 376.
- [51] D. Guégan, S. Rioublanc (2003) Studies of regime switching models. Do they provide long memory behavior? An empirical approach, NR IDHEMORA, 13-2003, ENSC, France.
- [52] J.D. Hamilton (1988) Rational Expectations Econometric Analysis of Change in Regime: An Investigation of the Term Structure of Interest Rates, *Journal of Economic Dynamics and Control*, 12, 385-423.
- [53] C.C. Heyde, Y. Yang (1997) On defining long range dependence, *J. of Applied Prob.*, 34, 939 - 944.

- [54] J.R.M. Hosking, (1981), Fractional differencing, *Biometrika*, 68, 1, 165-176.
- [55] Ibragimov I.A., Rozanov Y.A. (1974) *Processus aléatoires Gaussiens*, Mir Ed., Moscow.
- [56] D. Kim, S.J Kon (1999) Structural change and time dependence in models of stock returns, *Journal of Empirical Finance*, 6, 283 - 308.
- [57] Kolmogorov A.N., Rozanov Y.A. (1960) On strong mixing conditions for stationary Gaussian Processes, *Theory Prob. Applications*, 5, 204-208.
- [58] C.W. Li, W.K. Li (1996) On a Double - Threshold Autoregressive Heteroscedastic Time Series Model, *Journal of Applied Econometrics*, 11, 253 - 274.
- [59] P. de Lima, N. Crato (1994) Long range dependence in the conditional variance of stock returns, *Economic letters*, 45, 281 - 285.
- [60] J.L. Lin (1991) Generalized integrated processes and the aggregation of dynamic time series, *Academia Economic papers*, 19, 207 - 226.
- [61] E. Lukacs (1970) *Characteristic functions*, Griffin, Londres.
- [62] S. Lundbergh, T. Terasvirta, D. van Dijk (2003) Time-Varying Smooth Transition Autoregressive Models, *Journal of Business and Economic Statistics*, 21, 104 - 121.
- [63] T. Mikosch, C. Starica (1999) Change of structure in financial time series, long range dependence and the GARCH model, Manuscript, Dept. of Stat., University of Pennsylvania.
- [64] S. N. Neftçi (1982) Optimal predictions of cyclical downturns, *Journal of Economic Dynamics and Control*, 4, 307-327.
- [65] S. N. Neftçi (1984) Are economic time series asymmetric over the business cycle, *Journal of Political Economy*, 92, 307 - 328.
- [66] J. von Neumann (1941) Distribution of the ratio of the mean square Successive Difference to the Variance, *Annals of Math. Stat.*, 12, 367 - 395.
- [67] E. Parzen (1981) Time Series Model Identification and Prediction Variance Horizon, in "Applied Time Series Analysis II", ed. Findley, Academic Press, 415-447, New York.

- [68] D.B. Percival (1983) The statistics of long memory processes, PhD, University of Washington, U.S.A.
- [69] G.A. Pfann, P.C. Schotman, R. Tchernig (1996) Nonlinear interest rate dynamics and implication for the term structure, *J. of Econometrics*, 74, 149 - 176.
- [70] M.B. Priestley (1981) *Spectral Analysis and Time Series*, Academic Press, New York.
- [71] R.E. Quandt (1958) The Estimation of Parameters of Linear Regression System Obeying Two Separate Regime, *Journal of the American Statistical Association*, 55, 873-880.
- [72] P.M. Robinson (2001) The memory of stochastic volatility models, *Journal of Econometrics*, 101, 195 - 218.
- [73] M. Rosenblatt (1956) A central limit theorem and a strong mixing condition, *Proceeding of the National Academy of Sciences*, 42, 43 - 47.
- [74] T Rydén, T. Terasvirta, S. Asbrink (1998) Stylised facts of daily return series and the hidden Markov model, *J.of Applied Econometrics*, 13, 217 - 244.
- [75] G. Samorodnitsky, M. Taqqu (1994) *Stable non-Gaussian Random Processes: stochastic models with infinite variance*, Chapman and Hall, New York.
- [76] A. Sklar (1959) Fonctions de répartition à n dimensions et leurs marges, *Publications de l'Institut de Statistique de L'Université de Paris*, 8, 229-231.
- [77] A. Smith (2003) Level Shifts and the Illusion of Long Memory in Economic Time series, Preprint Univ. of California.
- [78] M.K.P. So, W.K. Li, K. Lam (2002) A Threshold Volatility Model, *Journal of Forecasting*, 21, 473 - 500.
- [79] C. Starica, C.W.J. Granger (2003) Non stationarities in stock returns, Preprint, Chalmers University.
- [80] J.H. Stock (1994) unit roots and trend breaks, in Engle R.F., McFadden D. eds., *Handbook of Econometrics*, vol IV, North Holland, Amsterdam.

- [81] M.S. Taqqu, W. Willinger, R. Sherman (1997) Proof of a fundamental result in self-similar traffic modeling, *Computer Communication Review*, 27, 5 - 23.
- [82] T. Terasvirta, H.M. Anderson (1992) Characterising Nonlinearities in Business Cycles using Smooth Transition Autoregressive models, *Journal of Applied Econometrics*, 7, S119 - S136.
- [83] V. Teverovsky, M.S. Taqqu (1997) Testing for long range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator, *Journal of Time series analysis*, 18, 279 - 304.
- [84] H. Tong, K.S. Lim (1980) Threshold Autoregression, Limit Cycles and Cyclical Data, *Journal of the Royal Statistical Society, B* 42, 245 - 292.
- [85] H. Tong (1990) *Non-linear Time Series: a Dynamical Approach*, Oxford Scientific Publications, Oxford.
- [86] Woodward, W.A., Cheng Q.C. and Gray H.L. (1998) A k -factor GARMA long-memory model, *Journal of Time Series Analysis*, 19, 485-504.
- [87] K. Yoshimura (1978) Characterization of frequency stability: uncertainty due to the autocorrelation of the frequency fluctuations, *IEEE Trans. Instrum. Meas.*, IM27, 1 - 7.
- [88] P. Zaffaroni (2004) Stationarity and memory of ARCH(∞) models, *Econometric theory*, 20, 147 - 160.
- [89] J.M. Zakoian (1994) Threshold Heteroskedastic Models, *Journal of Economic Dynamics and Control*, 18, 931 - 955.