Working Paper Series, N. 1, February 2005

Department of Statistical Sciences

University of Padua

Italy

Switching regime and ARFIMA processes

Luisa Bisaglia
Department of Statistical Sciences
University of Padua
Italy

Margherita Gerolimetto
Department of Statistical Sciences
University of Padua
Italy

Abstract: The phenomena of long range dependence and regime switching are intimely related and it is very difficult to separate the two effects. In this paper we consider the problem of inference on the order of integration in presence of infrequent structural breaks. For this purpose we consider the most important estimators of long memory parameters used in literature and compare their performances in finite samples via Monte Carlo experiments.

Keywords: Long memory, Occasional structural breaks, Regime switching, Markov switching.

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE STATISTICHE

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Luisa Bisaglia
tel: +39 049 827 4168
luisa.bisaglia@unipd.it

Switching regime and ARFIMA processes

Luisa Bisaglia
Department of Statistical Sciences
University of Padua
Italy

Margherita Gerolimetto
Department of Statistical Sciences
University of Padua
Italy

Abstract: The phenomena of long range dependence and regime switching are intimely related and it is very difficult to separate the two effects. In this paper we consider the problem of inference on the order of integration in presence of infrequent structural breaks. For this purpose we consider the most important estimators of long memory parameters used in literature and compare their performances in finite samples via Monte Carlo experiments.

Keywords: Long memory, Occasional structural breaks, Regime switching, Markov switching.

1 Introduction

Since the seminal papers of Granger and Joyeux (1980) and Hosking (1981) there has been a considerable interest in modelling strong persistence in time series. This interest is motivated by the analysis of many empirical time series, such as, for instance, the hydrological time series of Nile River minima (Hurst, 1951), the time series of Ethernet traffic (Leland *et al.*, 1993) and financial time series (Granger and Ding, 1995, 1996). For all these series the autocorrelation function decreases to zero like a power function rather than exponentially and the spectral density diverges as the frequencies tend to zero.

Many authors have studied the problem of estimating the long memory parameter with both parametric and non/semi-parametric methods (see Beran, 1994 for a good review on this argument). However, recently, it has been shown that inference on the long memory parameter and persistence tests are severely compromised in series which display occasional breaks, since these processes give the impression of persistence. In other words, neglecting structural breaks causes an over estimation of the long memory parameter, leading the reseacher to believe in a long memory data generating process. Granger and Terasvirta (1999), Granger and Hyung (2004), Diebold and Inoue (2001), Mikosch and Starica (1999) and Gourieroux and Jasiak (2001) provide both theoretical justification and Monte Carlo evidence that models with structural breaks may exhibit spurious long memory properties and it is troublesome in practice to distinguish between the occasional break process and the long memory process. Indeed, if an occasional break process (for instance a STOBREAK process) is generate and the long memory parameter, d, is estimated, it is likely that the estimation method does not recognize the process as a short memory one with breaks in mean, but it produces instead a value of $\hat{d}$ bigger than zero.

Morana and Beltratti (2004) analyse the realized variance process for the DM/US$ and Yen/US$ exchange rates. They find that while long memory is evident in the actual process, a structural break analysis reveals that this feature is partially explained by anaccounted changes in regime. In all these papers only the Geweke-Porter Hudak (1983) method (hereaftere GPH) or Robinson's LM test (1994) are used to distinguish between long memory and switching regime.

On the other hand, as formally demonstrated by Bai (1998), when the disturbances of a regression model follow an $I(1)$ process there is a tendency to estimate a break point in the middle of the sample, even though a break point does not actually exist. Moreover, Granger and Hyung (2004) demonstrated that, in case of a DGP with long memory characteristics, the estimation techniques, as for instance quasi-maximum likelihood or least squares, usually used to estimate break points, encounter some difficulties in distinguishing between structural change and long memory. In particular, as the value of the long memory parameter d increases, it grows also the estimated number of breaks even when the DGP have no break points. So the long memory properties of the DGP might cause many breaks to be detected spuriously by standard estimation methods and unlike $I(0)$ processes, caution should be exercised in estimating break points in presence of $d > 0$.

The choice between long memory and structural break model is very important especially if we devoted much attention to forecasting. In fact Diebold and Inoue (2001) suggest that long memory models may be a useful description of the real process in terms of providing good forecasts, even if the DGP possesses structural breaks. However, the authors do not offer empirical evidence supporting their view and, moreover, Gabriel and Martins (2004) show, by Monte Carlo study and an empirical comparison using a real time series, that although long memory models may capture some in-sample features of the data, their forecasting performance is relatively poor when shifts occur in the series, compared to simple linear and Markov switching models.

In this paper, we consider the problem of distinguishing between long memory and swithing regime processes and in particular we are interested in evaluating and comparing the performance of some estimation methods often used in the literature. We orient our interest to the semiparametric GPH method, the parametric pseudo-likelihood Whittle method and some non parametric methods like the Higuchi method, the aggregate variance method and the riscaled range method. We want to find out whether the bad performance exhibited by the long memory tests considered so far in the literature (in particular the GPH) is typical also of the other estimation methods or some of them manage to be somehow more robust in case of structural breaks.

The plan of the paper is as follows. In Section 2, we briefly introduce long memory ARFIMA models and some simple linear models with occasional or infrequent breaks in mean. Section 3 reviews the link between long memory and occasional switching regime. Section 4 is devoted to Monte Carlo simulations to compare the performance of the various estimation methods used to distinguish between long memory and structural breaks and Section 5 concludes.

2 Long memory ARFIMA and Occasional Structural Break processes

In this Section we want to recall some basic notions about long memory ARFIMA processes, some estimation metodologies for the long memory parameter and three simple models of stochastic regime switching that we will use in our experiments.

2.1 Long memory ARFIMA processes

There exists different definitions of long memory processes. In particular, long memory could be expressed either in the time domain or in the frequency domain. In the time domain, a stationary discrete time series is said to be long memory if its autocorrelation function decays to zero like a power function. This definition implies that the dependence between successive observations decays slowly as the number of lags tends to infinity. On the other hand, in the frequency domain, a stationary discrete time series is said to be long memory if its spectral density is unbounded at low frequencies.

Alternatively, the memory of a process can be expressed in terms of the rate of growth of variances of partial sums, $\mathrm{var}(S_T) = O(T^{2d+1})$, where $(S_t) = \Sigma_{t=1}^{t} x_t$ and d is the long memory parameter. These definitions are not completly equivalent but there is a tight connection between them (Beran, 1994).

In this paper we consider one of the most popular long memory processes that is the Autoregressive Fractionally Integrated Moving Average process, ARFIMA(p, d, q) in the following, independently introduced by Granger and Joyeux (1980) and Hosking (1981). This process simply generalizes the usual ARIMA(p, d, q) process by assuming d to be fractional.

Let ϵ_t be a white noise process having $E[\epsilon_t^2] = \sigma^2$. The process $\{X_t,\ t \in \mathbf{Z}\}$ is said to be an ARFIMA(p, d, q) process with $d \in (-1/2, 1/2)$, if it is stationary and satisfies the difference equation

$$\Phi(B)\,\Delta(B)\,(X_t - \mu) = \Theta(B)\,\epsilon_t,$$

where $\Phi(\cdot)$ and $\Theta(\cdot)$ are polynomials in the backward shift operator B of degree p and q, respectively, $\Delta(B) = (1 - B)^d = \sum_{j=0}^{\infty} \pi_j B^j$ with $\pi_j = \Gamma(j - d)/[\Gamma(j + 1)\Gamma(-d)]$, and $\Gamma(\cdot)$ is the gamma function.

If $p = q = 0$ the process $\{X_t,\ t \in \mathbf{Z}\}$ is called Fractionally Integrated Noise and denoted by $I(d)$. When $d \in (0, 1/2)$ the ARFIMA(p, d, q) process is stationary and the autocorrelation function decays to zero hyperbolically at a rate $O(k^{2d-1})$, where k denotes the lag. In this case we say that the process has a long-memory behavior. When $d \in (-1/2, 0)$ the ARFIMA(p, d, q) process is a stationary process with intermediate memory. In the following we will concentrate on $I(d)$ processes with $d \in (0, 1/2)$: for this range of values the process is stationary, invertible and possesses long-range dependence. Moreover, we will assume for convenience and without loss of generality that $\sigma^2 = 1$ and $\mu = 0$.

2.2 Estimation techniques for ARFIMA processes

Now we briefly describe the methods we consider to estimate the long memory parameter d and so to detect long memory behaviour on real data.

1. **The R/S method.** The R/S or *rescaled adjusted range* statistic was first introduced by Hurst (1951), who investigated the question of how to regularize the flow of the Nile river. This problem has been refined among others by Mandelbrot (1972, 1975) and Mandelbrot and Taqqu (1979).
   The statistic R/S can be used both to estimate the long-memory parameter and to test the presence of long-range dependence in a process.

   Let $(X_t,\ t = 1, 2, \ldots)$ be the observed data and let $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$ denote the sample mean and $s_n^2 = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ the usual sample variance.
   The R/S statistic is the range of partial sums of deviations of the time series from its mean rescaled by its standard deviation. For every $n > 0$ it is given by:

   $$\tilde{Q}(n) = \frac{1}{s_n} \left[\max_{1 \le k \le n} \sum_{i=1}^{k} (X_i - \bar{X}_n) - \min_{1 \le k \le n} \sum_{i=1}^{k} (X_i - \bar{X}_n) \right]$$

   where n represents a lag. Mandelbrot (1975) proves that if the process $(X_t,\ t \in \boldsymbol{T})$ is such that $(X_t^2,\ t \in \boldsymbol{T})$ is ergodic and $N^{-H} \sum_{t=1}^{N} X_t$ converges weakly to a Fractional Brownian Motion as $N \to \infty$, where N is the sample size of the series, then as $n \to \infty$

   $$n^{-H} \tilde{Q}(n) \xrightarrow{d} \xi \tag{1}$$

   where ξ is a nondegenerate random variable and $\xrightarrow{d}$ denotes the convergence in distribution. If we consider $H = 1/2$, these assumptions are generally satisfied by a short-memory process. H is called Hurst exponent and its relation with the long memory parameter d is $H = d + 1/2$. Moreover if the random variables X_t are i.i.d. with $E(X_t^2) < \infty$ or with $E(X_t^2) = \infty$ and X_t in the domain of attraction of a stable process with $0 < \alpha < 2$, then $H = 1/2$ (Mandelbrot, 1972). In this case:

   $$\frac{\tilde{Q}(n)}{\sqrt{n}} \xrightarrow{d} V$$

   where V is the range of a Brownian Bridge on the unit interval.

   This means that if we plot the $\log(\tilde{Q}(n))$ against $\log(n)$, the points should be scattered randomly around a straight line with slope H. Thus if the process $(X_t,\ t \in \boldsymbol{T})$ is a short-memory process, then $H = 1/2$ (or $d = 0$), and if the process $(X_t,\ t \in \boldsymbol{T})$ has a long-memory dependence then $H > 1/2$ (or $d > 0$).

2. **The aggregate variance method.** If the process $(X_t,\ t \in \boldsymbol{T})$ is a stationary, long-memory process then

$$Var(\bar{X}_n) \sim Cn^{2H-2}$$

when $n \to \infty$, where C is a positive constant. This means that the variance of the sample mean converges to zero slower than at the usual rate of convergence n^{-1} (nevertheless the loss of efficiency with respect to the best linear unbiased estimator is very small: about two percent if the model is Gaussian). This suggests that we can estimate H or d proceeding as follows:

Let $(X_t,\ t = 1, 2, \ldots, N)$ be the time series.

(a) For $m = 2, 3, \ldots, N/2$ we divide the series into N/m subseries of size m. For each subseries we calculate the sample mean

$$\bar{X}_k(m) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \quad k = 1, 2, \ldots, \frac{N}{m}$$

and then the overall mean

$$\bar{X}(m) = \frac{1}{N/m} \sum_{k=1}^{N/m} \bar{X}_k(m).$$

(b) For each m we calculate the sample variance of the sample means $\bar{X}_k(m)$, $k = 1, 2, \ldots, N/m$, that is:

$$Var(\bar{X}(m)) = \frac{1}{N/m - 1} \sum_{k=1}^{N/m} (\bar{X}_k(m) - \bar{X}(m))^2.$$

(c) We plot $\log Var(\bar{X}(m))$ against $\log m$.

For large values of m the data should produce a straight line with slope $(2H-2)$ or $(2d-1)$. The straight line is fitted according to the least squares procedure. If the data have not long-range dependence $(H = 1/2,\ d = 0)$, then the slope of the straight line should equal -1. In practice we have to consider m and N large enough and $m \ll N$ so that both the length of each subseries and the number of subseries are large.

3. **The Higuchi method.** This method was introduced by Higuchi (1988) for measuring the fractal dimension of a non periodic and irregular time series, like for instance $H-$ self-similar processes.

Let $(X_t,\ t = 1, 2, \ldots, N)$ be a Fractional Gaussian Noise series. Then the partial sums of X_t

$$Y_t = \sum_{i=1}^{t} X_i$$

reproduce the Fractional Brownian Motion series. We assume here that $Y_0 = 0$ with probability 1. For the time series $Y_t,\ t = 1, 2, \dots, N$ we can build the k time series

$$Y_k^m = X_m,\ X_{m+k},\ X_{m+2k}, \dots, X_{m+[\frac{N-m}{k}]}$$

where $m = 1, 2, \dots, k$ is the initial time and k indicates the interval time. For each series it is possible to define the normalized length of the curve:

$$L_m(k) = \frac{N-1}{k^2[\frac{N-m}{k}]} \sum_{i=1}^{[\frac{N-m}{k}]} \mid Y_{m+ik} - Y_{m+(i-1)k} \mid .$$

Finally we define the length of the curve for the time interval k as:

$$L(k) = \frac{1}{k} \sum_{m=1}^{k} L_m(k).$$

In the case of a self-similar series we have the following result:

$$E[L(k)] \sim C_H k^{-D}$$

where $D = 3/2 - d$. Then if we plot $\log L(k)$ with respect to $\log k$ the data should produce a straight line with slope $-D$. The straight line is fitted according to the least squares procedure. If the time series has not long-dependence, then $d = 0$ and thus we should find $D = 3/2$.

4. **The log-periodogram method** This is one of the best known methods to estimate in a semiparametric way the fractional parameters d of long-range dependence behaviour. The advantage of this method is that the specification of the model is not really necessary because the only information we need is the behaviour of the spectral density near the origin. Furthermore, the long-memory parameter can be estimated alone.

   This method was first introduced by Geweke and Porter-Hudak (1983) for the Gaussian case when d belongs to $(-1/2, 0)$ and then it was developed by Robinson (1995).

   Assume that the process $(X_t,\ t = 1, 2, \dots, N)$ is a ARFIMA(p,d,q) model:

$$\Phi(B)\Delta^d X_t = \Theta(B)\epsilon_t\ , \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2)$$

   then we can observe that the spectral density of this model is proportional to $(4\sin^2(\lambda/2))^{-d}$ near the origin, i.e.

$$f(\lambda) \sim c_f(4\sin^2(\frac{\lambda}{2}))^{-d}, \tag{2}$$

   when λ tends to 0. Since the periodogram $I(\lambda)$ is an asymptotically unbiased estimate of the spectral density, that is:

$$\lim_{\lambda \to 0} E[I(\lambda)] = f(\lambda)$$

it is possible to estimate d applying the least squares method to the following equation:

$$\ln(I(\lambda_{j,N})) = \ln c_f + \beta \ln(4\sin^2(\frac{\lambda}{2})) + \epsilon_j \quad (3)$$

where $\alpha = \ln c_f$, $\beta = -d$ and ϵ_j, $j = 1, 2, \ldots, n^*$ are the i.i.d. error terms, evaluating the equation (2) at the Fourier frequencies $\lambda_{j,N} = (2\pi j/N)$, $j = 1, 2, \ldots, n^*$, where n^* is the integer part of $(N-1)/2$.
If we define $y_j = \ln(I(\lambda_{j,N}))$, $x_j = \ln(4\sin^2(\lambda/2))$, $e_j = \epsilon_j + C$, and $\alpha = \ln c_f + C$, then the equation (3) can be written as:

$$y_j = \alpha + \beta x_j + e_j, \qquad j = 1, 2, \ldots, N$$

where now the e_t are i.i.d. zero mean variables.

The equation (2) is an asymptotic relation that holds only in the neighborhood of the origin, then if we use this relation from all periodogram ordinates ($-\pi < \lambda < \pi$) the estimator of d can be highly biased. GPH proposed to estimate the least squares line (3) from the periodogram ordinates at low frequencies only. Nevertheless the results are not always optimal.

Even if this method presents some problems, it is the only non parametric method to estimate d for which it is possible to derive an asympototic distribution of the estimator. Thus it is possible to build tests and confidence intervals for d.

5. **Whittle estimator** Several theoretical and practical advantages are possessed by the frequency domain approximate maximum likelihood method proposed by Fox and Taqqu (1986), also called Whittle estimator. This estimator extends the results of Hannan (1973), who applied Whittle's method to the estimation of the parameters of ARMA models. Fox and Taqqu's result, later generalized by Dahlhaus (1989) to the exact maximum likelihood estimator, is the basis of one of the most used methods for estimating the long (and short, if both are present) memory parameters in Gaussian time series. Giraitis and Surgailis (1990) generalized the result of Fox and Taqqu in order to prove the asymptotic normality of Whittle's estimator without the Gaussianity assumption.

   The exact maximum likelihood estimator has the drawback of implying a large computational burden and it might also cause computational problems when calculating the autocovariances needed to evaluate the likelihood function (Sowell, 1992). These difficulties do not occur when using the Whittle estimator, which has the further advantage of not requiring the estimation of the mean of the series (generally unknown in practice). Besides, under some regularity assumptions (Fox and Taqqu, 1986, Dahlhaus, 1989) fulfilled by ARFIMA(p, d, q) processes, it is possible to prove that the Whittle estimator has the same asymptotic distribution as the exact maximum likelihood estimator and it converges to the true values of the parameters at the usual rate of $n^{-1/2}$, where n is the length of the series. Finally, for Gaussian processes the Whittle estimator is asymptotically efficient in the sense of Fisher.

If the Whittle approximation to the log-likelihood function is used, the parameter vector $\boldsymbol{\theta} = (d, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ is estimated by minimizing with respect to $\boldsymbol{\theta}$ the estimated variance of the underlying white noise process: $\hat{\sigma}^2(\boldsymbol{\theta}) = \frac{1}{2\pi}\sum_{j=1}^{n'} \frac{I_n(\lambda_j)}{f(\lambda_j,\boldsymbol{\theta})}$, where n' is the integer part of $(n-1)/2$, $I_n(\lambda_j)$ denotes the periodogram of the series, defined at the Fourier frequencies $\lambda_j = 2\pi j/n$ $(j = 1, \ldots, n')$, and $f(\lambda_j, \boldsymbol{\theta})$ indicates the spectral density of the ARFIMA process at the Fourier frequency λ_j.

The disadvantage of this estimator is that it assumes the parametric form of the spectral density to be known *a priori*. If the specified spectral density function is not the correct one (as it is often the case) the estimated parameters may be biased and consequently the forecast will not be optimal in the sense of minimizing the mean square prediction error.

2.3 Occasional break processes

In this section we will describe some well known structural break processes that we will use in our finite sample experiments.

In all the models we are describing there are only occasional breaks in mean, which means that the number of breaks that can occur in a specific period of time is somehow bounded.

More formally, we assume that the probabiliy of breaks, p, converges to zero slowly as the sample size increases, i.e. $p \to 0$ as $T \to \infty$, yet $\lim_{T\to\infty} Tp$ is a non-zero finite constant. The meaning of this assumption is that the number of breaks, Tp, is bounded from above, even in the extreme case that T increases to infinity. To maintain memory property induced by breaks the number of breaks has to remain finite as the size of sample increases. Yet this sample size dependent probability is not truly time varying parameter, it is only a condition under which regarless the sample size, realization tends to have finite breaks.

The first model we consider and describe is the so called mean plus noise or occasional break model (Chen and Tiao, 1990; Engle and Smith, 1999)

$$y_t = m_t + \epsilon_t, \qquad t = 1, ..., T, \tag{4}$$

where ϵ_t is a noise variable and occasional level shifts, m_t, are controlled by two variables q_t (date of breaks) and η_t (size of jump), as

$$m_t = m_{t-1} + q_t\eta_t \tag{5}$$

where η_t is i.i.d. $(0, \sigma_\eta^2)$. In the following sections the distribution of η_t has been taken to be normal although this distribution has no particular relevance. We assume that q_t follows an i.i.d. binomial distribution, that is,

$$q_t = \begin{cases} 0, & \text{with probability } 1-p \\ 1, & \text{with probability } p \end{cases} \tag{6}$$

Combine Eqs. 4 and 5, then y_t is represented by

$$y_t = (m_0 + q_1\eta_1 + ... + q_t\eta_t\} + \epsilon_t \tag{7}$$

The time-varying mean of y_t is $(m_0 + q_1\eta_1 + ... + q_t\eta_t\}$, which shows infrequent level shifts.

The binomial model we described is so characterized by sudden changes only. It might be also the case that structural changes occur gradually. In this case a Markov switching model (Hamilton, 1989), a simple generalization of Eq.7, is more adapt. Suppose s_t is a latent random variable that can assume only the two discrete values 0 or 1. Each value of s_t represents a different state in the length of memory of shock. s_t is assumed to be governed by the following Markov probability law: $p_{ij} = Pr\,(s_t = j/s_{t-1} = i)$. Then, it is possible to use a switching model of q_t such that $q_t = 0$ when $s_t = 0$ and $q_t = 1$ when $s_t = 1$. In this specification, called Markov switching model a regime with $s_t = 1$ represents a period of structural change. Therefore, there is structural change everytime $s_t = 1$, both if $s_{t-1} = 1$ and if $s_{t-1} = 0$. We have to observe that graphically only in the second case the swithing regime will be so visible, whereas in the first case, there will be a switching regime followed by another switching regime of approximately the same magnitude order. This causes the presence of local trends that look like those of a long memory process and because of this we expect the tests for long memory perform particularly bad in case of a Markov switching model.

Finally we present the so called Stocastic Permanent Break model (STOPBREAK model) formulated by Engle and Smith (1999) to bridge the gap between transience and permanence of the shocks. Typically time series analysis tends to draw a sharp line between processes where shocks have a permanent effect and those where they do not. The most notable example of this is the distinction between stationary $AR(1)$ where all the shocks are transitory and random walk where all the shocks have infinite memory. The STOPBREAK is a stocastic process in which the long run impact of each obervation is time varying and stochastic. At one extreme all innovations are transitory, at the other all shocks are permanent.

The formulation is as follows:

$$y_t = m_t + \epsilon_t \tag{8}$$

$$m_t = m_{t-1} + q_{t-1}\epsilon_{t-1} \tag{9}$$

where $q_t = q\,(|\epsilon_t|)$ is non decreasing in $|\epsilon_t|$ and bounded by zero and one, so that bigger innovations have more permanent effects, and ϵ_t are i.i.d $N\left(0, \sigma_\epsilon^2\right)$, moreover $q_t = \frac{\epsilon_t^2}{(\gamma+\epsilon_t^2)}$ for $\gamma > 0$. Therefore in the STOPBREAK process permanent shocks can be indentified by their larger magnitude. In this approach the effects of shocks can fluctuate between transitory and permanent and typically such data exhibit periods of apparent stationarity punctuated by occasional mean shifts.

3 Long memory and infrequent structural breaks

The appareance of long memory may be a genuine feature of time series or it could occur when several processes are aggregated (Granger, 1980) or because of the presence of structural breaks and regime switching (Granger and Hyung, 2004; Diebold and Inoue, 2001). In this Section we will discuss how recent literature has dealt with the presence of occasional structural breaks and infrequent regime switching in studying long memory in the data.

Granger and Terasvirta (1999) give an example of a very simple nonlinear autoregressive short memory model (that is essentially a regime switching model) that produces empirical autocorrelations declining slower to zero like that of an $I(d)$ process.

Gourieroux and Jasiak (2001) study how processes with infrequent stochastic breaks can generate a long memory effect in the estimated autocorrelation function. According with the idea of Granger and Terasvirta (1999) they assess that the hyperbolic decay rate of the estimated autocorrelograms may results from nonlinear dynamics with infrequent switching regimes and not from genuine fractional dynamics.

Granger and Hyung (2004) study the sample autocorrelation function of an occasional break model, that is a model with only a few breaks in the mean. The authors prove that the sample autocorrelation function of the model (4) can be approximated by the following expression:

$$\hat{\rho}_T\left(k\right) \approx \frac{\frac{Tp\sigma_\eta^2}{6}\left(1-\frac{k}{T}\right)\left(1-2\frac{k}{T}+4\left(\frac{k}{T}\right)^2\right)}{\frac{Tp\sigma_\eta^2}{6}+\sigma_\epsilon^2} \tag{10}$$

where $\approx$ denotes approximate equality for any k such that $k/T \to 0$ as T increases. The value of Tp implies an expected number of structural breaks with the sample size T, and σ_η^2 is related to the size of breaks. Moreover, the authors show that for this process, if the probability of breaks p converges to zero slowly as the sample size increases, i.e., $p \to 0$ as $T \to \infty$, then

$$\hat{\rho}_T\left(k\right) \to \left(1+\frac{6\sigma_\epsilon^2}{Tp\sigma_\eta^2}\right)^{-1}$$

where $0 < \left(1+\frac{6\sigma_\epsilon^2}{Tp\sigma_\eta^2}\right)^{-1} < 1$. That is the $k-th$ sample autocorrelation converges to nonzero value for any k such that $k/T \to 0$ as T increases. Obviously, when $0 < p < 1$ and p is fixed the break process is $I(1)$ and $\hat{\rho}_T\left(k\right) \to 1$ as $T \to \infty$, while when $p = 0$ then there are not break, the process is $I(0)$ and $\hat{\rho}_T\left(k\right) \to 0$ as $T \to \infty$.

Note that, although the autocorrelations in eq. (10) decay very slowly as k increases, there is a fondamental difference with the theoretical autocorrelation function of an ARFIMA process. For this process, in fact, the asymptotic decay of the correlations is hyperbolic, while for the occasional break process considered by Granger and Hyung (2004) is polynomial. Moreover, the autocorrelation function for the occasional break process depends on the sample size T because the process is non stationary, while the ARFIMA process is stationary for the interest value of the parameter, $0 < d < 1/2$. Thus we make a mistake if we calculate the autocorrelation function for an occasional break process because we are considering just linear properties of the data.

Engle and Smith (1999) show that the STOPBREAK model they propose is an approximation to the mean-plus-noise model considered by Granger and Hyung (2004). So, we suppose, the autocorrelation function of this model behaves like eq. (10).

Moreover, Diebold and Inoue (2001) introduce a slight variation in the STOPBREAK model allowing γ changing with the sample size, i.e. $q_t = \frac{\epsilon_t^2}{(\gamma(T)+\epsilon_t^2)}$., The autors showed that if $E(\epsilon_t^\delta) < \infty$ and $\gamma \to \infty$ as $T \to \infty$ with $\gamma(T) = O(T^\delta), \quad \delta > 0$, then the variances of the partial sums match those of an $I(d)$ process with $d = 1-\delta$. When γ is fixed then $\delta = 0$ and the STOPBREAK process is $I(1)$.

4 Monte Carlo study

In this Section we describe the Monte Carlo study we carried out to show and compare the performance of the different long memory estimation methods when the real DGP is a structural break process.

The functions we use are written in R language (Ihaka and Gentlemen, 1996) and are available upon request by the authors.

Simulations were conducted for the following models:

1. DGP1: mean plus noise model (4), with $p = 0.0025$, 0.005, 0.01, 0.05, 0.1 and $\sigma_\eta^2 = 0.005$, 0.01, 0.05, 0.1;

2. DGP2: Markov switching model, with $(p, q) = (0.95, 0.95; 0.95, 0.99; 0.99, 0.95; 0.99, 0.99; 0.999, 0.999)$ and $\sigma_\eta^2 = 0.005$, 0.01, 0.05, 0.1. In this case the initial state s_1 is generated by a Bernoulli random variable with $p = 0.5$;

3. DGP3: STOPBREAK model (8), with $\gamma = (10^{-5}, 10^{-3}, 10^{-1}, 1, 10, 10^2, 10^3)$, following Diebold and Inuoe (2001), in order to make comparisons.

For each model we have considered $\sigma_\epsilon^2 = 1$ and $s = 1000$ independent realizations. Thus for a given estimation method we obtain $s = 1000$ estimated values for d. Moreover, to evaluate the effects of different sample size, we have considered $T =$ 500, 1000, 2000. All series are generated with 200 additional values in order to obtain random starting values.

In the following we present the tables with the results we obtained.

By observing the tables, in general we can notice that the performance of all methods becomes poor as the sample size increases. This could be justified by thinking that as the series become longer, the non stationary behaviours are more emphasized, therefore a structural break process is more similar to a long memory one.

From a very general point of view, it is also worth noticing that the Whittle estimator exhibits always the best performance. Indeed, if compared to the other techniques, the estimates obtained with the Whittle method are often much closer to zero over all the DGP's we considered. On the other hand, it is interesting to point out that the GPH technique has often a bad perfomance. This is because the GPH procedure focuses on the elasticity of the spectral density close to zero frequency: when $p = 0$ (no break), the elasticity is 0 which implies no long memory in the frequency domain, but when $0 < p < 1$ and p is fixed, then the elasticity tends to one as the frequency tends to zero (for details, see Granger and Hyung, 2004).

Going into further details, if we read tables 1-3 about DGP1 we can notice that for all methods the bias tend to grow with the increase both of the number of expected breaks Tp, and the size of jump σ_η. In those conditions the level of similarity between long memory and structural break processes is much higher and because of this it is more difficult for the estimation methods to distinguish the two different typologies.

As we already pointed out, the Whittle method is the one which performs best. Indeed, for low values of σ_η the Whittle estimation manages to approach zero quite well even when p reaches the higher value. At the same time when the expected number of breaks is not too high the Whittle method produces an estimate close to zero also when σ_η is big. This good performance of the Whittle method is very interesting and we are still working on a justification for it.

The Higuchi and GPH methods are those which perform worst. Indeed, only for the lowest value of p and σ_η the two estimates manage to be reasonably close to zero.

If we read table 4-6 about DGP2, first of all we notice that the general performance of all methods is worse than in case of DGP1. Even the Whittle method does not manage to produce an estimate close to zero, although it still performs better than the other procedures.

When $p = q$ the unconditional probability of having a break is always 1/2. If p and q increase, the bias reduces, but the variabiliy of the estimates grows because the probability of remaining in the same initial state is bigger and so the estimate can be evaluated from very low values of d to high ones. For instance, if the initial state is 1 and $p = 0.999$ it is very likely that the process will remain in state 1, so we would end up having a $I(1)$ process, on the contrary, but following the same logic, if the initial state is 0, and $q = 0.999$, we would end up with an $I(0)$ process.

When $p \neq q$ the performances are definitely not symmetric and that is because the unconditional probabiliy are different. The latter can be computed following Hamilton (1989) $\tilde{p} = \frac{1-q}{2-p-q}$. So, as expected, when $p = 0.95$ and $q = 0.99$ the performance is generally better since there is a higher probability of remaining in the state 0 and so the process has a more stationary appearance.

As far as tables 7-9 about DGP3 are concerned, first of all we can notice that with the increase of γ the performance of the estimator improves since the size of the jump becomes smaller. Also we observe that the GPH and the Whittle methods seem to be more sensitive to the variations of γ, than the other methods. However the standard error of the Whittle method is always the smallest.
For all the DGP's we considered the results we obtained are consistent with previous results presented in literature (Granger and Hyung, 2004; Diebold and Inoue, 2001), where, in fact, only the GPH (in test version) has been used.

Table 1: Estimation results for d: DGP1, T=500

σ_η^2		$p = 0.0025$	$p = 0.005$	$p = 0.01$	$p = 0.05$	$p = 0.1$
	rs	0.046 (0.077)	0.054 (0.076)	0.065 (0.087)	0.148 (0.106)	0.199 (0.119)
	av	-0.013 (0.077)	-0.008 (0.077)	0.006 (0.081)	0.086 (0.095)	0.133 (0.104)
0.005	hi	0.034 (0.128)	0.073 (0.139)	0.125 (0.149)	0.274 (0.152)	0.335 (0.134)
	gph	0.014 (0.171)	0.022 (0.166)	0.041 (0.177)	0.162 (0.175)	0.220 (0.187)
	wh	0.012 (0.019)	0.012 (0.018)	0.015 (0.022)	0.035 (0.033)	0.053 (0.042)
	rs	0.050 (0.081)	0.065 (0.086)	0.091 (0.094)	0.204 (0.125)	0.285 (0.139)
	av	-0.008 (0.078)	0.008 (0.082)	0.030 (0.088)	0.136 (0.106)	0.208 (0.111)
0.01	hi	0.067 (0.142)	0.130 (0.155)	0.190 (0.157)	0.337 (0.134)	0.380 (0.118)
	gph	0.018 (0.178)	0.042 (0.185)	0.080 (0.180)	0.225 (0.188)	0.326 (0.193)
	wh	0.012 (0.020)	0.015 (0.022)	0.019 (0.024)	0.055 (0.045)	0.089 (0.056)
	rs	0.095 (0.111)	0.136 (0.131)	0.194 (0.138)	0.376 (0.157)	0.474 (0.154)
	av	0.032 (0.104)	0.071 (0.117)	0.129 (0.121)	0.284 (0.109)	0.357 (0.083)
0.05	hi	0.157 (0.184)	0.253 (0.169)	0.327 (0.144)	0.424 (0.095)	0.446 (0.080)
	gph	0.077 (0.196)	0.134 (0.208)	0.219 (0.203)	0.458 (0.206)	0.584 (0.198)
	wh	0.023 (0.031)	0.034 (0.038)	0.054 (0.048)	0.141 (0.069)	0.203 (0.072)
	rs	0.125 (0.139)	0.176 (0.150)	0.256 (0.161)	0.474 (0.160)	0.546 (0.156)
	av	0.062 (0.124)	0.109 (0.130)	0.181 (0.133)	0.350 (0.093)	0.398 (0.069)
0.1	hi	0.205 (0.190)	0.292 (0.174)	0.372 (0.133)	0.444 (0.077)	0.458 (0.069)
	gph	0.122 (0.211)	0.196 (0.213)	0.297 (0.218)	0.582 (0.198)	0.688 (0.189)
	wh	0.033 (0.043)	0.050 (0.051)	0.080 (0.062)	0.199 (0.077)	0.262 (0.076)

Table 2: Estimation results for d: DGP1, T=1000

σ_η^2		$p = 0.0025$	$p = 0.005$	$p = 0.01$	$p = 0.05$	$p = 0.1$
	rs	0.055 (0.081)	0.082 (0.093)	0.117 (0.100)	0.260 (0.132)	0.346 (0.142)
	av	0.006 (0.075)	0.028 (0.085)	0.065 (0.089)	0.193 (0.104)	0.264 (0.098)
0.005	hi	0.093 (0.139)	0.170 (0.152)	0.231 (0.146)	0.355 (0.121)	0.404 (0.101)
	gph	0.038 (0.143)	0.058 (0.148)	0.106 (0.151)	0.261 (0.158)	0.359 (0.166)
	wh	0.010 (0.015)	0.014 (0.018)	0.020 (0.023)	0.059 (0.039)	0.093 (0.048)
	rs	0.080 (0.092)	0.111 (0.109)	0.168 (0.117)	0.343 (0.148)	0.438 (0.147)
	av	0.030 (0.083)	0.058 (0.097)	0.112 (0.099)	0.263 (0.103)	0.331 (0.087)
0.01	hi	0.151 (0.153)	0.217 (0.162)	0.284 (0.143)	0.399 (0.104)	0.431 (0.085)
	gph	0.067 (0.141)	0.099 (0.154)	0.163 (0.152)	0.359 (0.171)	0.472 (0.167)
	wh	0.014 (0.019)	0.020 (0.024)	0.033 (0.030)	0.091 (0.050)	0.135 (0.055)
	rs	0.160 (0.144)	0.240 (0.157)	0.325 (0.155)	0.538 (0.152)	0.604 (0.153)
	av	0.106 (0.122)	0.172 (0.125)	0.244 (0.113)	0.392 (0.073)	0.426 (0.057)
0.05	hi	0.267 (0.171)	0.338 (0.141)	0.391 (0.112)	0.453 (0.072)	0.462 (0.067)
	gph	0.154 (0.178)	0.239 (0.183)	0.335 (0.177)	0.616 (0.171)	0.717 (0.164)
	wh	0.034 (0.036)	0.055 (0.045)	0.085 (0.051)	0.199 (0.064)	0.254 (0.068)
	rs	0.208 (0.176)	0.311 (0.172)	0.416 (0.164)	0.602 (0.146)	0.648 (0.156)
	av	0.146 (0.144)	0.233 (.129)	0.311 (0.105)	0.425 (0.056)	0.446 (0.048)
0.1	hi	0.302 (0.176)	0.373 (0.129)	0.425 (0.094)	0.465 (0.059)	0.468 (0.062)
	gph	0.212 (0.206)	0.328 (0.202)	0.437 (0.184)	0.714 (0.160)	0.813 (0.151)
	wh	0.051 (0.049)	0.083 (0.057)	0.125 (0.063)	0.250 (0.066)	0.312 (0.068)

Table 3: Estimation results for d: DGP1, T=2000

σ^2_η		$p = 0.0025$	$p = 0.005$	$p = 0.01$	$p = 0.05$	$p = 0.1$
	rs	0.096 (0.101)	0.143 (0.111)	0.206 (0.124)	0.408 (0.149)	0.493 (0.149)
	av	0.051 (0.088)	0.101 (0.094)	0.153 (0.104)	0.310 (0.091)	0.368 (0.073)
0.005	hi	0.189 (0.155)	0.265 (0.152)	0.323 (0.131)	0.416 (0.084)	0.440 (0.074)
	gph	0.071 (0.124)	0.124 (0.122)	0.174 (0.132)	0.373 (0.144)	0.486 (0.147)
	wh	0.011 (0.014)	0.019 (0.019)	0.030 (0.026)	0.091 (0.042)	0.130 (0.047)
	rs	0.143 (0.119)	0.226 (0.127)	0.303 (0.135)	0.515 (0.154)	0.580 (0.151)
	av	0.097 (0.102)	0.166 (0.104)	0.235 (0.096)	0.371 (0.072)	0.407 (0.060)
0.01	hi	0.250 (0.152)	0.324 (0.131)	0.375 (0.111)	0.439 (0.077)	0.461 (0.062)
	gph	0.118 (0.130)	0.182 (0.132)	0.268 (0.124)	0.487 (0.140)	0.596 (0.142)
	wh	0.019 (0.020)	0.034 (0.028)	0.054 (0.035)	0.130 (0.048)	0.170 (0.052)
	rs	0.304 (0.180)	0.408 (0.167)	0.501 (0.157)	0.681 (0.155)	0.706 (0.157)
	av	0.223 (0.126)	0.302 (0.103)	0.362 (0.080)	0.445 (0.045)	0.458 (0.039)
0.05	hi	0.361 (0.133)	0.412 (0.102)	0.438 (0.084)	0.466 (0.056)	0.471 (0.056)
	gph	0.263 (0.168)	0.378 (0.166)	0.482 (0.145)	0.736 (0.142)	0.816 (0.134)
	wh	0.059 (0.041)	0.089 (0.048)	0.127 (0.051)	0.242 (0.059)	0.292 (0.063)
	rs	0.385 (0.192)	0.493 (0.170)	0.579 (0.164)	0.710 (0.151)	0.729 (0.156)
	av	0.281 (0.122)	0.350 (0.088)	0.402 (0.067)	0.461 (0.038)	0.470 (0.031)
0.1	hi	0.393 (0.121)	0.423 (0.091)	0.454 (0.071)	0.475 (0.052)	0.474 (0.052)
	gph	0.347 (0.183)	0.470 (0.163)	0.584 (0.155)	0.826 (0.126)	0.895 (0.121)
	wh	0.083 (0.051)	0.122 (0.055)	0.169 (0.059)	0.352 (0.059)	0.352 (0.059)

Table 4: Estimation results for d: DGP2, T=500

σ^2_η		$p = 0.95$	$p = 0.95$	$p = 0.99$	$p = 0.99$	$p = 0.999$
		$q = 0.95$	$q = 0.99$	$q = 0.95$	$q = 0.99$	$q = 0.999$
	rs	0.386 (0.159)	0.231 (0.163)	0.461 (0.155)	0.365 (0.187)	0.286 (0.238)
	av	0.291 (0.104)	0.161 (0.133)	0.343 (0.076)	0.273 (0.131)	0.201 (0.188)
0.005	hi	0.423 (0.099)	0.355 (0.134)	0.459 (0.068)	0.414 (0.113)	0.296 (0.220)
	gph	0.469 (0.200)	0.272 (0.224)	0.665 (0.194)	0.446 (0.235)	0.344 (0.320)
	wh	0.145 (0.067)	0.072 (0.062)	0.246 (0.074)	0.139 (0.082)	0.116 (0.104)
	rs	0.472 (0.165)	0.299 (0.183)	0.529 (0.154)	0.432 (0.190)	0.318 (0.265)
	av	0.350 (0.090)	0.216 (0.142)	0.389 (0.076)	0.319 (0.125)	0.218 (0.208)
0.01	hi	0.444 (0.083)	0.396 (0.120)	0.459 (0.068)	0.435 (0.099)	0.293 (0.230)
	gph	0.584 (0.198)	0.355 (0.244)	0.664 (0.194)	0.542 (0.230)	0.392 (0.364)
	wh	0.200 (0.074)	0.103 (0.076)	0.246 (0.074)	0.182 (0.089)	0.145 (0.128)
	rs	0.634 (0.158)	0.483 (0.219)	0.661 (0.173)	0.599 (0.210)	0.407 (0.330)
	av	0.439 (0.053)	0.341 (0.135)	0.455 (0.044)	0.412 (0.101)	0.266 (0.233)
0.05	hi	0.470 (0.059)	0.454 (0.080)	0.473 (0.056)	0.467 (0.069)	0.315 (0.234)
	gph	0.822 (0.184)	0.598 (0.272)	0.890 (0.173)	0.781 (0.235)	0.538 (0.455)
	wh	0.357 (0.083)	0.221 (0.111)	0.424 (0.085)	0.341 (0.119)	0.254 (0.205)
	rs	0.679 (0.178)	0.558 (0.222)	0.694 (0.189)	0.651 (0.208)	0.439 (0.341)
	av	0.455 (0.046)	0.382 (0.119)	0.464 (0.040)	0.434 (0.087)	0.285 (0.235)
0.1	hi	0.474 (0.057)	0.463 (0.074)	0.477 (0.052)	0.468 (0.065)	0.328 (0.234)
	gph	0.891 (0.174)	0.693 (0.259)	0.931 (0.167)	0.854 (0.226)	0.588 (0.464)
	wh	0.436 (0.089)	0.282 (0.123)	0.499 (0.082)	0.419 (0.121)	0.308 (0.234)

Table 5: Estimation results for d: DGP2, T=1000

σ_η^2		$p = 0.95$ $q = 0.95$	$p = 0.95$ $q = 0.99$	$p = 0.99$ $q = 0.95$	$p = 0.99$ $q = 0.99$	$p = 0.999$ $q = 0.999$
	rs	0.527 (0.155)	0.383 (0.173)	0.587 (0.151)	0.529 (0.167)	0.405 (0.275)
	av	0.386 (0.075)	0.285 (0.116)	0.419 (0.059)	0.380 (0.085)	0.286 (0.194)
0.005	hi	0.455 (0.068)	0.414 (0.104)	0.461 (0.067)	0.450 (0.076)	0.348 (0.201)
	gph	0.609 (0.169)	0.407 (0.198)	0.693 (0.164)	0.599 (0.187)	0.485 (0.329)
	wh	0.192 (0.064)	0.112 (0.064)	0.239 (0.064)	0.193 (0.072)	0.159 (0.113)
	rs	0.601 (0.151)	0.473 (0.173)	0.640 (0.149)	0.579 (0.160)	0.426 (0.299)
	av	0.425 (0.058)	0.345 (0.102)	0.443 (0.051)	0.409 (0.072)	0.297 (0.205)
0.01	hi	0.465 (0.061)	0.435 (0.084)	0.475 (0.053)	0.459 (0.074)	0.353 (0.207)
	gph	0.723 (0.161)	0.515 (0.198)	0.789 (0.156)	0.695 (0.177)	0.532 (0.369)
	wh	0.251 (0.066)	0.156 (0.069)	0.299 (0.067)	0.239 (0.074)	0.194 (0.137)
	rs	0.709 (0.172)	0.640 (0.175)	0.725 (0.189)	0.708 (0.175)	0.503 (0.337)
	av	0.465 (0.040)	0.428 (0.072)	0.471 (0.034)	0.459 (0.048)	0.328 (0.214)
0.05	hi	0.475 (0.054)	0.466 (0.060)	0.475 (0.053)	0.475 (0.053)	0.358 (0.214)
	gph	0.898 (0.143)	0.767 (0.192)	0.934 (0.135)	0.879 (0.163)	0.661 (0.413)
	wh	0.401 (0.072)	0.285 (0.089)	0.457 (0.071)	0.391 (0.084)	0.304 (0.194)
	rs	0.713 (0.180)	0.691 (0.176)	0.723 (0.207)	0.749 (0.186)	0.534 (0.341)
	av	0.469 (0.035)	0.447 (0.060)	0.474 (0.034)	0.467 (0.040)	0.342 (0.212)
0.1	hi	0.475 (0.054)	0.473 (0.059)	0.479 (0.052)	0.480 (0.046)	0.371 (0.204)
	gph	0.936 (0.134)	0.836 (0.176)	0.962 (0.132)	0.938 (0.149)	0.715 (0.419)
	wh	0.472 (0.075)	0.348 (0.095)	0.532 (0.071)	0.469 (0.086)	0.367 (0.219)

Table 6: Estimation results for d: DGP2, T=2000

σ_η^2		$p = 0.95$ $q = 0.95$	$p = 0.95$ $q = 0.99$	$p = 0.99$ $q = 0.95$	$p = 0.99$ $q = 0.99$	$p = 0.999$ $q = 0.999$
	rs	0.669 (0.148)	0.552 (0.165)	0.714 (0.151)	0.655 (0.168)	0.559 (0.265)
	av	0.446 (0.045)	0.384 (0.079)	0.461 (0.038)	0.434 (0.059)	0.376 (0.151)
0.005	hi	0.471 (0.055)	0.441 (0.084)	0.474 (0.049)	0.462 (0.062)	0.409 (0.152)
	gph	0.736 (0.133)	0.533 (0.159)	0.826 (0.123)	0.715 (0.148)	0.637 (0.264)
	wh	0.240 (0.055)	0.151 (0.055)	0.293 (0.058)	0.231 (0.064)	0.204 (0.103)
	rs	0.713 (0.148)	0.636 (0.158)	0.728 (0.165)	0.724 (0.157)	0.601 (0.304)
	av	0.462 (0.037)	0.424 (0.059)	0.469 (0.032)	0.458 (0.041)	0.381 (0.169)
0.01	hi	0.476 (0.051)	0.455 (0.069)	0.471 (0.058)	0.473 (0.054)	0.409 (0.165)
	gph	0.824 (0.120)	0.670 (0.169)	0.881 (0.114)	0.816 (0.142)	0.702 (0.314)
	wh	0.295 (0.057)	0.205 (0.064)	0.353 (0.062)	0.291 (0.065)	0.252 (0.128)
	rs	0.741 (0.213)	0.742 (0.147)	0.755 (0.208)	0.768 (0.184)	0.683 (0.315)
	av	0.476 (0.028)	0.464 (0.038)	0.478 (0.028)	0.473 (0.032)	0.410 (0.159)
0.05	hi	0.475 (0.055)	0.475 (0.052)	0.476 (0.055)	0.476 (0.050)	0.419 (0.163)
	gph	0.965 (0.104)	0.870 (0.135)	0.964 (0.097)	0.947 (0.109)	0.837 (0.303)
	wh	0.496 (0.061)	0.330 (0.073)	0.495 (0.059)	0.431 (0.067)	0.384 (0.157)
	rs	0.744 (0.224)	0.779 (0.169)	0.739 (0.243)	0.788 (0.214)	0.695 (0.298)
	av	0.478 (0.026)	0.467 (0.039)	0.481 (0.025)	0.477 (0.030)	0.420 (0.141)
0.1	hi	0.481 (0.047)	0.472 (0.059)	0.479 (0.048)	0.480 (0.049)	0.426 (0.142)
	gph	0.968 (0.096)	0.870 (0.135)	0.987 (0.096)	0.975 (0.109)	0.837 (0.272)
	wh	0.502 (0.061)	0.330 (0.073)	0.573 (0.062)	0.501 (0.069)	0.448 (0.162)

Table 7: Estimation results for d: DGP3, T=500

γ	10^{-5}	10^{-3}	10^{-1}	1	10	10^2	10^3
rs	0.722 (0.308)	0.719 (0.308)	0.731 (0.330)	0.714 (0.285)	0.709 (0.191)	0.337 (0.152)	0.044 (0.085)
av	0.476 (0.030)	0.476 (0.033)	0.478 (0.033)	0.475 (0.033)	0.463 (0.041)	0.234 (0.107)	-0.012 (0.075)
hi	0.476 (0.051)	0.479 (0.051)	0.479 (0.049)	0.479 (0.047)	0.475 (0.052)	0.394 (0.117)	0.076 (0.132)
gph	1.003 (0.158)	1.005 (0.160)	0.997 (0.161)	0.996 (0.161)	0.923 (0.166)	0.368 (0.189)	0.009 (0.165)
wh	0.982 (0.023)	0.982 (0.023)	0.958 (0.033)	0.825 (0.059)	0.485 (0.082)	0.104 (0.058)	0.012 (0.018)

Table 8: Estimation results for d: DGP3, T=1000

γ	10^{-5}	10^{-3}	10^{-1}	1	10	10^2	10^3
rs	0.729 (0.348)	0.725 (0.354)	0.707 (0.339)	0.731 (0.302)	0.728 (0.199)	0.481 (0.158)	0.061 (0.084)
av	0.478 (0.029)	0.478 (0.028)	0.478 (0.029)	0.478 (0.046)	0.473 (0.031)	0.349 (0.083)	0.011 (0.073)
hi	0.477 (0.052)	0.478 (0.049)	0.476 (0.052)	0.479 (0.046)	0.477 (0.054)	0.445 (0.075)	0.147 (0.144)
gph	0.995 (0.129)	1.001 (0.126)	0.994 (0.127)	1.000 (0.118)	0.955 (0.125)	0.513 (0.168)	0.045 (0.141)
wh	0.987 (0.015)	0.988 (0.016)	0.965 (0.025)	0.834 (0.049)	0.514 (0.069)	0.149 (0.060)	0.010 (0.015)

5 Conclusion

In this article we have presented further Monte Carlo evidence about the issue of distinguishing between a time series process exhibiting long-range dependence and one with short memory but suffering from structural shifts. We considered several different estimation methods of the long memory parameter d, in particular we focused our attention on the GPH method, Whittle, Higuchi, aggregate variance and riscaled range.

We considered three differnet DGP's among the most used in recent literature on this topic, the mean plus noise, the Markov switching and the STOPBREAK model.

We found out that almost all the test tend to be biased and this is because the process with a occasional breaks looks very similar to a long memory one, especially when the size of the jumps is relatively big and also it is high the expected number of breaks.

The Markov switching model is the one where it is too difficult to detect the structural breaks, probably because in this case the shifts take place gradually, giving the impression of a local trend more than a break point.

Among our future research lines we can mention:

1. the analysis of the other side of the problem, that is to understand the performance of the tests to detect structural breaks in hypothesis of long memory data generating processes;

2. to develop an appropriate test procedure to distinguish real long memory from

Table 9: Estimation results for d: DGP3, T=2000

γ	10^{-5}	10^{-3}	10^{-1}	1	10	10^2	10^3
rs	0.725 (0.384)	0.702 (0.369)	0.692 (0.362)	0.708 (0.327)	0.741 (0.225)	0.611 (0.156)	0.119 (0.098)
av	0.482 (0.026)	0.479 (0.026)	0.480 (0.026)	0.481 (0.026)	0.478 (0.028)	0.421 (0.056)	0.070 (0.083)
hi	0.474 (0.049)	0.478 (0.051)	0.479 (0.049)	0.479 (0.052)	0.476 (0.053)	0.458 (0.067)	0.244 (0.139)
gph	1.002 (0.104)	1.003 (0.101)	1.004 (0.108)	1.002 (0.103)	0.976 (0.097)	0.634 (0.139)	0.089 (0.117)
wh	0.992 (0.010)	0.992 (0.010)	0.968 (0.019)	0.841 (0.049)	0.542 (0.062)	0.191 (0.052)	0.014 (0.015)

spurious one generate by infrequent breaks;

3. the problem of forecasting with long memory processes when the real DGP is a switching regime process and viceversa.

References

Bai, J. (1998). A note on spurious break *Econometric Theory*, **14**, 663–669.

Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, London.

Chen, C. and Tiao, G.C. (1990). Random level-shift time series models, ARIMA approximations, and level-shift detection *Journal of Business and Economics Statistics*, **8**, 83–97.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics*, **17**, 1749–1766.

Diebold, F.X. and Inoue, A. (2001). Long memory and regime switching *Journal of Econometrics*, **105**, 131–159.

Engle, R.F. and Smith, A.D. (1999). Stochastic permanent breaks *The Review of Economics and Statistics*, **81**, 553–574.

Fox, R. and Taqqu, M.S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, **14**, 517–532.

Gabriel, V.J. and Martins, L.F. (2004). On the forecasting ability of ARFIMA models when infrequent breaks occur *Econometrics Journal*, **7**, 455–475.

Geweke J. and Porter-Hudack S. (1983) The Estimation and Application of Long-Memory Time Series Models *J.T.S.A.*, **4**, (1983) 221-237.

Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields*, **86**, 87–104.

Gourieux, C. and Jasiak, J. (2001). Memory and infrequent breaks. *Economics Letters*, **70**, 29–41.

Granger, C.W.J. (1980). Long-memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, **14**, 227–238.

Granger, C.W.J. and Ding, Z. (1995). Some properties of absolute returns. An alternative measure of risk. *Annales d'Economie et de Statistique*, **40**, 67–91.

Granger, C.W.J. and Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, **73**, 61–78.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing. *Journal of Time Series Analysis*, **1**, 15–30.

Granger, C.W.J. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S$P500 absolute stock returns. *Journal of Empirical Finance*, **11**, 399–421.

Granger, C.W.J. and Terasvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. *Economics Letters*, **62**, 161–165.

Hannan, E.J. (1973). The asymptotic theory of linear time series models. *Journal of Applied Probability*, **10**, 130–145.

Higuchi, T. (1988). pproach to an irregular time series on the basis of the fractal theory. *Physica D*, **31**, 277-283.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, **68**, 165–176.

Hurst, H.E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **116**, 770–799.

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.

Leland, W.E., Taqqu, M.S., Willinger, W. and Wilson, D.V. (1993). High time-resolution measurement and analysis of LAN traffic: implications for LAN interconnection. *Proceedings of the IEEE INFOCOM '91, Bal Harbor, FL*, 1360–1366.

Mandelbrot, B.B. (1972). Statistical Methodology for nonperiodic cycles: from the Covariance to R/S Analysis. *Ann. Economic. Soc. Measurement*, **1**, (1972) 259-290.

Mandelbrot, B.B. (1975). Limit theorems of the self-normalized range for weakly and strongly dependent processes. *Z. Wahr. verw. Geb.*, **31**, (1975) 271-285.

Mandelbrot, B.B. and Taqqu, M.S. (1979). Robust analysis of long-run serial correlation *Proceedings of the 42nd Session of the International Statistical Institute, Manila, 1979.* Bulletin of the International Statistical Institute, Vol. 48, Book 2, 69-104.

Mikosh, T. and Starica, C. (1999). Change of structure in financial time series, long range dependence and the GARCH model. Working Paper. The Wharton School, Chalmers University of Technology

Morana, C. and Beltratti, A. (2004). Structural change and long-range dependence in volatility of exchange rates: either, neither or both? *Journal of Empirical Finance*, **11**, 629–658.

Robinson P. M. (1995) Log-periodogram regression of time series with long range dependence *The Annals of Statistics*, **23**, (1995) 1048-1072.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it
Most of the working papers can also be found at the following url: http://wp.stat.unipd.it