A Study on "Spurious Long Memory in Nonlinear Time Series Models"

Heri Kuswanto and Philipp Sibbertsen
Leibniz University Hanover, Department of Economics, Germany

Discussion Paper No. 410
November 2008
ISSN 0949-9962

Abstract

This paper discusses the existence of spurious long memory in common nonlinear time series models, namely Markov switching and threshold models. We describe the asymptotic behavior of the process in terms of autocovariance and autocorrelation function and support the theoretical evidences by providing Monte Carlo simulation. The existence of long memory in these nonlinear processes is induced by the nature of the process in certain conditions. In addition, GPH estimator itself introduces bias.

JEL - classifications: C12, C22
Keywords: long memory, nonlinear time series, regime switching

1 Introduction

In this paper we discuss the asymptotic behavior of nonlinear processes which are able to create spurious long memory. In the recent years econometric research addressed the problem of finding spurious long memory when the data contains structural breaks. A growing literature proposed models which able to capture both phenomena, as well as developed tests to distinguish between long memory and structural changes. Granger and Hyung (2004) notice that a linear process with breaks can mimic long memory. For an overview about structural breaks and long memory, see Sibbertsen (2004) or Banarjee and Urga(2005).

However, long memory can appear in various processes. Granger and Ding (1996) demonstrate that some processes can generate long memory as for instance processes containing an aggregation scheme, time changing coefficient models and possibly nonlinear time series. For the existence of long memory in aggregated processes see also Robinson (1978). Leipus and Surgailis (2003) show that random coefficient autoregressive models may exhibit long memory, in the sense that the covariance function decays hyperbolically.

Breidt and Hsu (2002) consider extensively a class of nearly long memory time series. They consider regime switching with a dynamic mean structure and show that for special cases such as random level shift, AR(1), random walk and regime switching the processes have similar properties than a long memory process. Morever, Leipus et al. (2005) discuss the long memory properties and large sample behavior of partial sums in a general regime switching scheme. Parke (1999) introduces an error duration representation for fractional integration. Gourieoux and Jasiax (2001) study how processes with infrequent regime switching may generate a long memory effect in the autocorrelation function. Another related discussions about the relation of long memory and nonlinearity can be found in Deo et al. (2007) and Davidson and Sibbertsen (2005).

In this paper, we consider whether Markov switching and threshold models can exhibit long-range dependencies. These models are very popular in empirical applications and have been identified to create similar empirical characteristics as a long memory process. We study the asymptotic behavior of these nonlinear processes and perform a simulation study to support the theory. We describe in which sense nonlinear time series can create a spurious long memory behavior.

This paper is organized as follows: section 2 discusses some basic characteristics of long memory processes, section 3 discusses the estimation of the long memory parameter and the possible sources for a bias of the GPH estimator. The existence of spurious long memory in nonlinear processes is discussed in section 4 and section

5 concludes.

2 Characteristic of Long Memory Processes

Long memory or long range dependence means that observations far away from each other are still strongly correlated. The correlations of long memory processes decay slowly that is with a hyperbolic rate.

Long memory can be defined in different ways. The definition is always related to the asymptotic behavior of the process. In this paper we use those definitions of long memory which are used later for our considerations.

Definition 1 *Let (X_t) be a stationary process for which the following holds. There exists a real number $d \in (0, 1/2)$ and a constant $C_\rho > 0$ such that*

$$\lim_{\tau \to \infty} \frac{\rho(\tau)}{\tau^{2d-1}} = C_\rho$$

then (X_t) is called a stationary process with long memory.

From the definition above, it is known that the correlations of a long memory process decay with a hyperbolic rate. They are not summable. If definition 1 gives a definition for long memory in terms of the asymptotic decay of the autocovariance function, the equivalent definition below uses another characteristic of long memory in terms of the shape of the spectral density.

Definition 2 *Let (X_t) be a stationary process for which the following holds. There exists a real number $d \in (0, 1/2)$ and a constant $C_f > 0$ and a frequency $\lambda_0 \in [0, \pi]$ such that*

$$\lim_{\lambda \to \lambda_0} \frac{f(\lambda)}{|\lambda - \lambda_0|^{-2d}} = C_f$$

then (X_t) is called a stationary process with long memory.

Both definitions are equivalent as the spectral density links to the autocovariance function via a Fourier transformation. Another related definition is the asymptotic behavior of the variances of partial sums:

Definition 3 *Let (X_t) be a stationary process and denote by $\sigma_X(T)$ the variance of the partial sums $S_T = \sum_{t=1}^{T} X_t$. If the variance $\sigma_X(T)$ has the following asymptotic behavior*

$$\sigma_X(T) \sim O(T^{2d-1}), \quad when \;\; T \to \infty$$

with $d \in (0, 1/2)$, then (X_t) is called a stationary process with long memory.

In order to give a more severe understanding of the definitions above, Figure 1 shows an example of a typical path of a long memory time series and the autocorrelation function of this long memory process with parameter d equal to 0.4 . It can be seen that the autocorrelations are significant even after 50 lags and that they decay slowly.

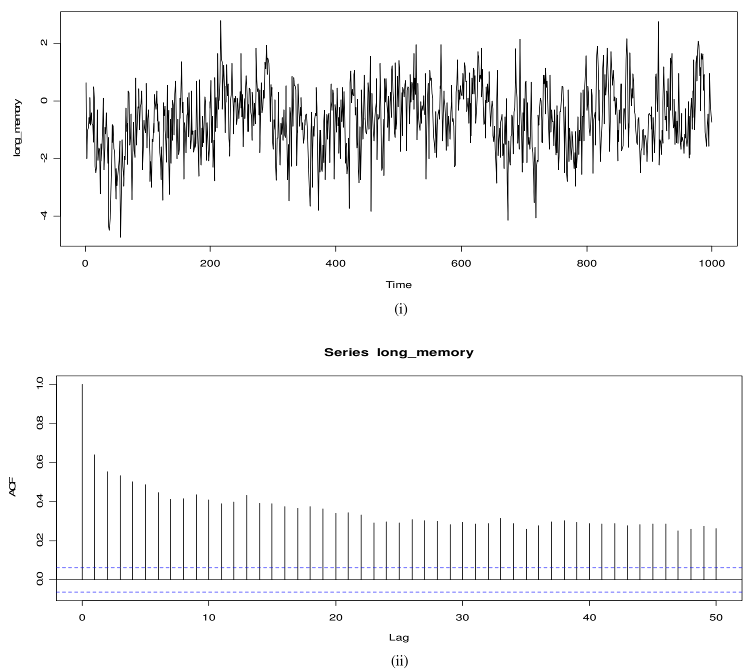

Figure 1: Long memory process with d=0.4. (i) time series plot (ii) autocorrelation function

3 Modeling Long Memory and Bias of the GPH estimator

ARFIMA models introduced by Granger and Joyeux (1980) and independently by Hosking (1981) are a popular class of long memory processes. They allow for a fractional degree of integration in order to generalize the class of ARIMA models. ARFIMA Model are defined as follows:

$$\phi(B)(1-B)^d X_t = \psi(B)\epsilon_t$$

where B is the backshift operator, $\phi(B)$ and $\psi(B)$ are the AR and MA polynomials respectively and ϵ_t is a white noise process.
The operator $(1-B)^d$ can be written as:

$$(1-B)^d = \sum_{j=0}^{\infty} \frac{d\Gamma(j+d)}{\Gamma(1+d)\Gamma(j+1)}, \tag{1}$$

The spectral density of an ARFIMA process behaves like a constant C_f times $|\lambda|^{-2d}$ near the origin. Thus the process exhibits long range dependence for $0 < d < 1/2$, where d characterizes the memory parameter (see Beran (1994) for details).

A popular semiparametric method to estimate d is the log-periodogram or GPH estimator proposed by Geweke-Porter Hudak (1983). It is based on the first J periodogram ordinates

$$I_j = \frac{1}{2\pi T}|\sum_{t-1}^{T} X_t \exp(i\lambda_j t)|^2$$

where $\lambda_j = 2\pi j/T$ with $j = 1, ..., J$ and J is a positive integer smaller than T. The idea is to estimate the spectral density by the periodogram and to take the logarithm on both sides of the equation. This gives a linear regression model in the memory parameter which can be estimated by least squares.

The estimator is given by $-1/2$ times the least squares estimator of the slope parameter in the regression of $\{\log I_j\}$ on a constant and the regressor variable

$$Y_j = \log|1-\exp(-i\lambda_j)| = \frac{1}{2}\log(2-2\cos\lambda_j).$$

By definition the GPH estimator is

$$\hat{d}_p = \frac{-0.5 \sum_{j=1}^{J} (Y_j - \bar{Y}) \log I_j}{\sum_{j=1}^{J} (Y_j - \bar{Y})^2} \tag{2}$$

where $\bar{Y} = \frac{1}{J} \sum_{j=1}^{J} Y_j$. This estimator can be motivated using the model:

$$\log I_j = \log C_f - 2dY_j + \log \xi_j \tag{3}$$

where Y_j denotes the j-th Fourier frequency and the ξ_j are identically distributed error variables with $-E[\log \xi_j] = 0.577$, known as Euler constant.

A short memory process is characterized by the value of $d = 0$. Thus, whenever the data generating process is short memory but creates a positive estimate of the memory parameter means that the GPH-estimator has to be biased. Next we are interested in the possible sources of the bias. Let us consider the popular version of GPH regression (3). The term in (2) can be arranged as follows:

$$d_p = \hat{d} + \frac{\sum_{j=1}^{J} (Y_j - \bar{Y}) \log \hat{I}_j / I_j}{\sum_{j=1}^{J} (Y_j - \bar{Y})^2} \tag{4}$$

where $\hat{d}$ is the GPH estimator and $\hat{I}_j$ is the estimated periodogram. Due to the fact that short memory process is characterized by the memory parameter equal to zero, thus the bias of GPH estimator is:

$$\begin{aligned} bias(\hat{d}) &= \mathbf{E}(\hat{d}) \\ &= d_p - \frac{\sum_{j=1}^{J} (Y_j - \bar{Y}) \mathbf{E}(\log \hat{I}_j / I_j)}{\sum_{j=1}^{J} (Y_j - \bar{Y})^2} \end{aligned}$$

From the last expression above, it is clear that there are two sources of bias. The first term d_p represents the bias induced by the short memory components and the second arises from the fact that the log periodogram is a biased estimator of the log spectrum (see Smith (2002) for details). To get a clear illustration about the bias of the GPH estimator, see Agiakloglou et al. (1993) and Choi and Wohar (1992). They provide an illustration for biases of the GPH estimator for simple AR(1) and MA(1) process.

4 Spurious Long Memory in Nonlinear Processes

We restrict the consideration in this paper to Markov switching and threshold models. There are several ways to show that the properties of such short memory processes can resemble long memory by means of the autocovariance function, the conditional mean, the variance of partial sums and the autocorrelation function as well as the spectral density.

The behavior of the periodogram as an estimator of the spectral density is one characteristic which might look similar for different processes in finite samples. Figure 2 and 3 present the spectral density and periodogram of a long memory, SETAR and Markov-Switching process respectively. Note that the long term behavior of a process is specified by the small frequencies of the periodogram. For long memory processes the spectral density has a pole at the origin.

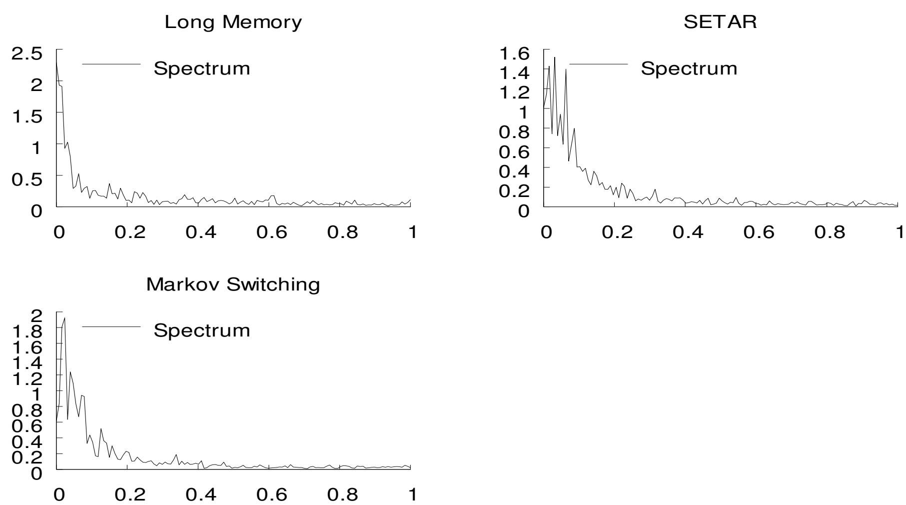

Figure 2: Plot spectrum of long memory, threshold and Markov switching process

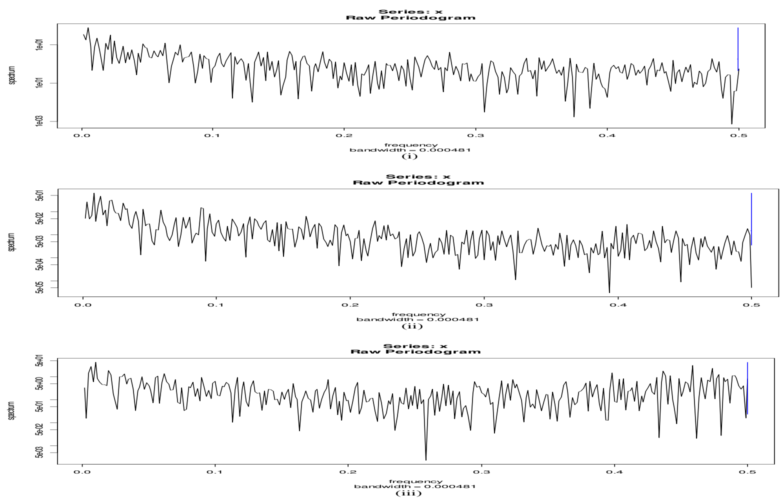

Figure 3: Plot periodogram of (i) long memory process (ii) Markov switching process (iii) Threshold process

From the figures it is clear that the periodogram as well as the spectrum of the processes are hardly to distinguish. They are identical and flat near the origin.

The following subsections discuss the asymptotic behavior of the processes as well as the simulation results giving evidence of long memory in the considered non-linear processes. Firstly, a simulation study applies the GPH estimator with the original bandwidth frequency proposed by GPH, which is $J \sim o(T^{0.5})$.

4.1 Markov Switching Models

In this paper we consider a simple two-state Markov switching model. The parameters of the process are time varying and are governed by an unobservable random variable s_t. Lets define the following first order Markov switching model with an AR(1) process in each regime (Hamilton(1989)):

$$X_t = \begin{cases} \mu_1 + \phi_1 X_{t-1} + \sigma_1 \epsilon_t & \text{if } s_t = 1 \\ \mu_2 + \phi_2 X_{t-1} + \sigma_2 \epsilon_t & \text{if } s_t = 2 \end{cases} \tag{5}$$

The model above can be written as:

$$X_t = \mu_{s_t} + \phi_{s_t} X_{t-1} + \sigma_{s_t} \epsilon_t \tag{6}$$

where μ_{s_t}, ϕ_{s_t} and σ_{s_t} are parameters under corresponding states s_t for $s_t = 1, 2$. The states represent different situation in a time series for instance expansion and recession, congestion and non-congestion, and so forth. The process s_t is a Markov chain, characterized by a transition probability $\mathbf{P}$ given by the following matrix:

$$\mathbf{P} = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix} \tag{7}$$

The properties of Markov switching models have been widely considered in recent papers. Yao and Attali (2000) give a sufficient condition for geometric ergodicity of Markov switching autoregressive models. Geometric ergodicity ensures the existence of stationary distribution, meaning that if X_0 is drawn from any stationary distribution, then X_t is also stationary and geometrically β-mixing. Higher moments of Markov switching process can be found in Timmermann (2000).

By assuming that the chain is irreducible and recurrent, and that there exists a stationary probability for the chain as matrix $\mathbf{P}$, Liu (2000) demonstrated the inability of the Markov switching model to generate long memory behavior. The following theorem formalizes the result:

Theorem 4 . *If the Markov chain is stationary, then the Markov chain regime switching model is in the class of short memory models.*

Proof: see Liu (2000)

The result is based on the behavior of the covariance which indicates that the process is short memory. The asymptotic behavior of the process is always used identify long memory. Guégan and Rioublanc (2003) derived the autocovariance function for model (5). They employ the following assumptions:
(1): The process $(\epsilon_t)_t$ is a strong white noise and all its moments exist
(2): $(s_t)_t$ is an irreducible, aperiodic and stationary Markov chain

(3): The process $(\epsilon_t)_t$ is independent of $(s_t)_t$

(4): $\parallel \Phi \mathbf{P} \parallel < 1$, with $\Phi = diag(\phi_1, \phi_2)$

(5): There are an integer $h \geq 1$ and a nonempty subset $K_1 = \{k_1, ..., k_{t_1}\}$ of the state space $K = \{1, 2\}$ such that

$$\min_{i \in K, j \in K_1} q_{ij}^{(h)} = \theta > 0$$

where $q_{ij}^{(h)}$ is the (i, j)th element of the matrix $(\mathbf{P}^h)'$, where $\mathbf{P}$ is defined in (7).

Assumption $(1)-(3)$ are needed to develop the unique strict stationarity condition and assumption (4) and (5) imply that the stable unconditional probabilities $\pi_i = \mathbf{P}[s_t = i], i = 1, 2$ exist and can be expressed as $\pi_i = \lim_{h \to \infty} q_{ij}^{(h)}, i = 1, 2$. Then, it can be shown that the convergence speed of the autocovariance function for the process X_t follows the theorem below:

Theorem 5 *Let X_t be the process defined in (5), by assuming that the assumption $(1) - (5)$ hold, then the autocovariance function $\gamma(\tau)$ of the process X_t converges to 0 with the rate $O(\tau v^{\tau})$, when $\tau \to \infty$, with $0 < v < 1$.*

Proof: see Guégan and Rioublanc (2003).

The theorem gives the rate of decay of the autocovariance function and confirms that the process defined in (5) asymptotically behaves as a short memory process in terms of the autocovariance function.

Below we present simulation results to confirm whether Markov switching processes as defined in (5) can be detected as long memory process or not. For all of our simulation settings, we use 1000 replications with sample size equal to $T = 200$ and $T = 600$ and parameter $\sigma_1 = \sigma_2$ are set to be one. For the first simulation in Table 1 we generate a data set following model (5) with the parameters $\mu_1 = 0.5$ and $\mu_2 = -0.5$ and $p_{11} = p_{22} = 0.1$. Different sample sizes are considered to assess the consistency of the estimator.

Table 1: GPH estimator for Markov switching process(5) with $p_{11} = p_{22} = 0.1$

	$T = 200$		$T = 600$	
$\phi_1 = -\phi_2$	d	$t - stat$	d	$t - stat$
0.1	-0.064	-8.419	-0.0347	-6.755
0.2	-0.066	-8.569	-0.0457	-8.324
0.3	-0.069	-8.815	-0.0416	-7.910
0.4	-0.067	-8.793	-0.0319	-6.102
0.5	-0.079	-10.089	-0.0383	-6.865
0.6	-0.082	-10.559	-0.0416	-7.770
0.7	-0.070	-10.039	-0.0394	-7.439
0.8	-0.084	-10.940	-0.0346	-6.459
0.9	-0.083	-11.003	-0.0496	-9.512

From the table above, it can be seen that for all cases the GPH estimator is indicating that the considered Markov process is a short memory process. This is also supported by the value of the t-statistic indicating that the estimator is not significantly different from zero. However, note that the results in Table 1 are obtained by setting the value of the transition probability $p_{11} = p_{22} = 0.1$.

Since the transition probabilities are a key element for Markov processes which are considered as "persistence" parameter, it is necessary to do further investigations by using other values. The higher the value of the transition probability p_{ii} the longer the process is expected to remain in state i and the process becomes more persistent.

Let us consider the following table, which contains the results when the same parameters of data generating process are used as above but now with $p_{11} = p_{22} = 0.9$. We expect that the GPH estimator will be biased towards long memory since these parameters leads to a higher persistence of the Markov switching process and long memory itself is also a persistent process.

Table 2: GPH estimator for the Markov switching process(5) with $p_{11} = p_{22} = 0.9$

	$T = 200$		$T = 600$	
$\phi_1 = -\phi_2$	d	$t - stat$	d	$t - stat$
0.1	0.1194	15.691*	0.0527	9.722*
0.2	0.1158	15.968*	0.0553	10.902*
0.3	0.1178	16.119*	0.0531	10.233*
0.4	0.1219	16.727*	0.0598	11.311*
0.5	0.1292	16.821*	0.0541	9.933*
0.6	0.1468	19.313*	0.0639	11.772*
0.7	0.1765	22.311*	0.0763	14.830*
0.8	0.2272	28.447*	0.0928	16.970*
0.9	0.3358	39.529*	0.1367	23.436*

Note: The asterisk indicates significance at 5% level

Now, a value of the transition probability leads to a positively biased GPH estimator. Table 2 shows that all values are in the range of stationary long memory, for $T = 200$ and $T = 600$. The t-statistic indicates that the estimator d is significantly greater than zero. This means that in certain cases, Markov switching process can exhibit long memory depending on the value of the transition probability. This result shows that instead of the autocovariances there should be other asymptotic properties of Markov switching processes (5) which resemble long memory and depend on the transition probability parameter.

To assess the behavior of the GPH estimator against sample size, we see that the value of d decreases with increasing sample size. This permits easy assessment of the extent to which the problem of bias diminishes with increasing sample size. This finding is consistent with the result of Agiakloglou et al. (1993).

Model (5) is a general Markov switching model and it contains several special cases. Therefore, we now discuss whether some of these representations do behave asymptotically like a long memory process. One of the processes which attract many considerations in the literatures is regime switching in the mean defined as follows, $\forall_t$:

$$X_t = \begin{cases} \mu_1 + \epsilon_t & \text{if } s_t = 1 \\ \mu_2 + \epsilon_t & \text{if } s_t = 2 \end{cases} \tag{8}$$

Thus, the the ij-th element of $\mathbf{P}$ gives the probability of moving from state i (at time $t-1$) to state j at time t. The process (8) is called as mean switching model where X_t switches from μ_1 to μ_2 and ϵ_t is Gaussian white noise with variance one, independent of the Markov chain s_t.

Model (8) is a candidate for a Markov switching process which is able to create a spurious long memory. Andel (1993) showed that the autocovariance function of a two state model such as (8) is similar to the autocovariance function of an ARMA(1,1) process. It is well known that ARMA processes are short memory with geometrically decaying autocorrelation functions. However, certain ARMA processes have autocorrelation functions which decay slowly enough to resemble long memory. The following lemma provides the autocorrelation function of the process (Guegan and Rioublanc (2005)):

Lemma 6 *The autocorrelation function $\rho(\tau)$ of the process X_t defined by (8) is equal to*

$$\rho(\tau) = \frac{(\mu_1 - \mu_2)^2(1-p_{11})(1-p_{22})r^\tau}{(2-p_{11}-p_{22})^2[\pi_1\mu_1^2 + \pi_2\mu_2^2 + 1 - (\pi_1\mu_1 + \pi_2\mu_2)^2]} \tag{9}$$

where $r = -1+p_{11}+p_{22}$, $\pi_1 = \frac{1-p_{22}}{2-p_{11}-p_{22}}$ and $\pi_2 = \frac{1-p_{11}}{2-p_{11}-p_{22}}$ are the non conditional probabilities.

From the lemma above, the autocorrelation function $\rho(\tau)$ can be written as $\rho(\tau) = A_{\mu_i,p_{ii}} r^h$, with $A_{\mu_i,p_{ii}}$ is defined as the following:

$$A_{\mu_i,p_{ii}} = \frac{(\mu_1 - \mu_2)^2(1-p_{11})(1-p_{22})}{(2-p_{11}-p_{22})^2[\pi_1\mu_1^2 + \pi_2\mu_2^2 + 1 - (\pi_1\mu_1 + \pi_2\mu_2)^2]}, i = 1, 2.$$

The levels μ_i and the transition probabilities p_{ii} determine the decay of the autocorrelation function with the rate of convergence is $r^\tau = (-1+p_{11}+p_{22})^\tau$.

Having r as defined above implies that for any value of transition probabilities p_{ii} will yield on r in the range of -1 and 1. r will close to 1 if the transition probabilities are high and therefore the autocorrelation function decreases slowly. In other words, if jumps are rare relative to sample size, then the process has a behavior similar to that of a long memory process. Otherwise, when r is close to 0 (the case of $p_{11} + p_{22}$ close to 1), the autocorrelation function will decay faster and shows the characteristic of a short memory process.

Consistent to the Lemma above another behavior of such Markov switching process is examined in Diebold and Inoue (2001). They point out that the variance of partial sums of the Markov switching process (8) matches those of long memory processes under certain conditions. The following proposition holds:

Proposition 4.1 *Assume that (a)* $\mu_1 \neq \mu 2$ *and that (b)* $p_{11} = 1 - C_1 T^{-\delta_1}$ *and* $p_{22} = 1 - C_2 T^{-\delta_2}$*, with* $\delta_1, \delta_2 > 0$ *and* $0 < C_1, C_2 < 1$*, then the variances of the partial sums of* X_t *grow at a rate corresponding to* $I((1/2)\max(\min(\delta_1, \delta_2) - |\delta_1 - \delta_2|, 0))$.

Proof: see Diebold and Inoue (2001).

By introducing those assumptions, they use the sample size to normalize the distance between the parameters p_{11} and p_{22} and the non-ergodic values. From this, Diebold and Ineue (2001) determine that the variance of the partial sums of X_t has the same order as the variance of the partial sums of fractionally integrated process for any value of $\delta_1, \delta_2 > 0$.

The tables below provide simulation results for the presence of long memory in the regime switching in mean process. The data generating process is based on different values for the mean and different settings of the transition probabilities following the lemma above. We consider mean value of $\mu_1 = 0.5$ and $\mu_2 = -0.5$ for the first, and $\mu_1 = 5$ and $\mu_2 = -5$ for the second simulation.

Table 3: GPH estimator for Markov switching process(8) with $\mu_1 = 0.5$ and $\mu_2 = -0.5$

	$T = 200$		$T = 600$	
$p_{11} = p_{22}$	d	$t - stat$	d	$t - stat$
0.1	-0.0465	-6.321	-0.0389	-7.131
0.2	-0.0521	-6.992	-0.0346	-6.597
0.3	-0.0570	-7.850	-0.0283	-5.511
0.4	-0.0618	-8.499	-0.0325	-6.148
0.5	-0.0551	-7.674	-0.0295	-5.620
0.6	-0.0418	-5.591	-0.0303	-5.687
0.7	-0.0461	-6.134	-0.0291	-5.267
0.8	-0.0015	-0.199	-0.0172	-3.249
0.9	0.1093	14.701*	0.0522	10.211*

In line with the result of the previous simulations long memory appears in the case of high transition probabilities. The Table below presents the simulation result by setting $\mu_1 = 5$ and $\mu_2 = -5$ to asses the behavior against μ.

Table 4: GPH estimator for Markov switching process(8) with $\mu_1 = 5$ and $\mu_2 = -5$

	$T = 200$		$T = 600$	
$p_{11} = p_{22}$	d	$t - stat$	d	$t - stat$
0.1	-0.0509	-7.212	-0.0373	-7.030
0.2	-0.0639	-8.507	-0.0404	-7.584
0.3	-0.0515	-6.596	-0.0299	-5.588
0.4	-0.0592	-7.780	-0.0353	-6.544
0.5	-0.0678	-8.348	-0.0256	-4.804
0.6	-0.0396	-5.293	-0.0291	-5.512
0.7	-0.0212	-2.880	-0.0315	-5.839
0.8	0.0537	6.976*	-0.0133	-2.481
0.9	0.2442	31.549*	0.1137	20.892*

The results suggest that a higher distance of the means leads to a higher possibility that long memory appears. For instance, if the value for $p_{11} = p_{22} = 0.8$ than for a higher μ the GPH estimator is biased towards long memory. Changing the transition probabilities yields to a consistent result with the previous experiment where a higher p_{ii} results in a higher probability that the GPH estimator is biased towards long memory.

The discussion about the bias of the GPH estimator leads to the question whether it is possible to reduce it and how the bandwidth frequency J has to be chosen. For the mean switching process Smith (2002) extends the results above to derive the limiting value of the GPH estimator d_p for a particular value of δ and shows that the choice of J will influence the GPH estimator.

Theorem 7 *Consider the Markov switching process in (6), let* $p_{11} = 1 - C_1 T^{-\delta}$ *and* $p_{22} = 1 - C_2 T^{-\delta}$*, and* $J = \theta T^{\gamma}$*, where* $\delta = 1 - \gamma$*, then*

$$\lim_{T\to\infty} d_p = 1 - 0.25 \sum_{m=0}^{\infty} (-1)^m (\frac{2\pi\theta}{(C_1 + C_2)})^{2m} (0.5 + m)^{-2}.$$

Proof: see Smith (2002)

The theorem implies that d has the limiting value which lies in $(0, 1)$ and therefore $\sup_{p_{11}+p_{22}\in(0,1)} d_p$ does not converge to zero as $T \to \infty$. Note that the function

$$\sum_{m=0}^{\infty} (-1)^m (\frac{2\pi\theta}{(C_1 + C_2)})^{2m} (0.5 + m)^{-2}$$

is special function called as the Lerch transcendent function evaluated at $(-((2\pi\theta)/(C_1+C_2))^2, 2, 0.5)$. This function generalizes the zeta function.

The fraction $\frac{(C_1+C_2)}{2\theta}$ can be written in terms of J as

$$\frac{(C_1 + C_2)}{2\theta} = (T(1 - p_{11}) + T(1 - p_{22}))/2J.$$

Thus, by setting different values of J will yield on the values of d in the range between zero and one, which characterize long memory. To see the behavior of the bias depending on the bandwidth selection, the Table below presents the GPH estimator by allowing for several choices of J dependent on γ.

Table 5: GPH estimator for Markov switching processes with different γ

	$T = 200$		$T = 600$	
γ	d	$t - stat$	d	$t - stat$
0.2	-0.2248	-5.832	-0.1702	-6.709
0.3	-0.0942	-4.870	-0.0798	-5.565
0.4	0.0273	2.447*	-0.0282	-3.501
0.5	0.1146	15.349*	0.0517	9.761*
0.6	0.1612	31.313*	0.1241	34.135*
0.7	0.1591	41.960*	0.1652	67.397*
0.8	0.1443	54.140*	0.1564	96.904*
0.9	0.1277	57.340*	0.1294	109.160*

The estimation cannot be carried out for $\gamma = 0.1$ as the bandwidth is too short. The results in Table 5 clearly show that the estimated value of d changes with a changing value of γ. In this case $\gamma = 0.5$ and $\gamma = 0.8$ correspond to the value suggested by Geweke and Porter Hudak (1983) and Hurvich et al. (1998), respectively. Hurvich et al.(1998) show that $J = T^{0.8}$ results in a minimal mean squared error(MSE). The Table below presents the value of the GPH estimator with the same parameter setting as in Table 3 and 4, but using $\gamma = 0.8$.

Table 6: GPH estimator for Markov switching processes (8) with $\mu_1 = 0.5$,$\mu_2 = -0.5$ and $\gamma = 0.8$

	$T = 200$		$T = 600$	
$p_{11} = p_{22}$	d	$t - stat$	d	$t - stat$
0.1	-0.0231	-8.213	-0.0095	-5.580
0.2	-0.0289	-10.468	-0.0133	-7.626
0.3	-0.0310	-11.340	-0.0170	-9.853
0.4	-0.0264	-9.390	-0.0162	-9.331
0.5	-0.0130	-4.841	-0.0049	-2.808
0.6	0.0150	5.470*	0.0119	6.742*
0.7	0.0464	16.272*	0.0408	23.532*
0.8	0.0905	32.745*	0.0880	52.817*
0.9	0.14601	56.887*	0.1577	98.305*

Table 7: GPH estimator for Markov switching processes (8) with $\mu_1 = 5$,$\mu_2 = -5$ and $\gamma = 0.8$

	$T = 200$		$T = 600$	
$p_{11} = p_{22}$	d	$t - stat$	d	$t - stat$
0.1	-0.1609	-58.384	-0.0864	-51.630
0.2	-0.1636	-58.137	-0.0903	-53.559
0.3	-0.1425	-51.357	-0.0806	-45.587
0.4	-0.0923	-31.657	-0.0588	-35.777
0.5	-0.0110	-3.841	-0.0072	-4.043
0.6	0.0900	32.191*	0.0685	39.630*
0.7	0.2217	78.366*	0.1761	108.502*
0.8	0.3932	142.849*	0.3297	187.361*
0.9	0.6002	217.947*	0.5523	314.112*

Comparing Table (3) with (6) and Table (4) with (7) leads to the conclusion that the choice of the bandwidth frequency is important in order to determine the bias. Using $\gamma = 0.8$, the GPH estimator will frequently be biased towards long memory.

Using $\gamma = 0.5$ results in a lower bias but it is considered as inefficient as it is not MSE optimal. In addition, if we see the nature of the process the closer the process is to ergodicity the higher is the persistence and the process will resemble long memory. We can say that the choice of $\gamma = 0.8$ gives a better explanation of the nature of the process in terms of persistency.

4.2 Threshold Models

Threshold models differ from Markov switching models on the way to create jumps from one state to another. Threshold models assume that the shifts between the regimes are observable and not exogenous. There are two different types of threshold models, namely SETAR and STAR models. The difference between them is that the regime switching in a SETAR model is based on a discontinuous function, whereas in STAR models it is based on continuous function. Threshold models especially the TAR model have a close relationship to the Markov switching process in a certain case (see Carrasco (2002) and Gourieroux (1997)). However, In case of the delay parameter equal to one, threshold models are not Markov switching because the Markov chain (indicator) function is not exogenous.

In this part we describe the existence of spurious long memory generated by threshold models. Point of departure is the following SETAR representation:

$$X_t = F_1(X_{t-1}, \Phi)(1 - I(X_{t-l} > c) + F_2(X_{t-1}, \Phi)(I(X_{t-l} > c)) + \epsilon_t, \qquad (10)$$

where the functions F_1 and F_2 are autoregressive processes depending on the past values of X_t and ϵ_t. The process ϵ_t is white noise and I an indicator function. The model (10) becomes a STAR model and the regime changes smoothly by setting the indicator function to a continuous function, $G(X_{t-l}, \gamma, c)$. If the function F_1 and F_2 are short memory, then the process in (10) is short memory.

In the case that one state has long memory, the process is long memory. Investigations on the existence of long memory in the processes is done by examining the stationarity conditions of the processes. Let us consider the SETAR (2,1) process, a simple SETAR with two regimes and autoregressive order one in each regime as described below:

$$X_t = \begin{cases} \phi_{0,1} + \phi_{1,1} X_{t-1} + \epsilon_t & \text{if } X_{t-1} \leq c \\ \phi_{0,2} + \phi_{1,2} X_{t-1} + \epsilon_t & \text{if } X_{t-1} > c \end{cases} \tag{11}$$

The delay parameter in the model above is set to be one. Chan (1993) and Dijk, et al. (2002) define the stationary conditions of (11) as follows:

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

From the conditions above, stationarity depends on the setting of the autoregressive parameters. A non-stationary behavior can appear in one regime whereas the process is still globally stationary, which can lead to a confusion with long memory.

The following tables present simulation results on spurious long memory in threshold models. Let us consider the case where the necessary and sufficient conditions for the stationarity above are fulfilled. Under the first condition the results are given in Kuswanto and Sibbertsen (2007) showing that the GPH estimator is biased towards long memory. To investigate the second condition, we set the parameters for the data generating process in Table 8 as $\phi_{1,1} = 1$, $\phi_{1,2} = 0.1$ and c is set to be zero.

Table 8: GPH estimator for TAR processes with $\phi_{1,1} = 1$ and $\phi_{1,2} = 0.1$

	$T = 200$		$T = 600$	
$\phi_{0,1}$	d	$t - stat$	d	$t - stat$
0.1	0.8249	207.604*	0.8441	307.957*
0.2	0.7678	162.380*	0.7121	231.900*
0.3	0.5928	125.017*	0.5913	183.095*
0.4	0.4956	106.163*	0.4812	147.967*
0.5	0.4069	94.433*	0.3885	129.739*
0.6	0.3306	76.876*	0.3094	105.290*
0.7	0.2723	65.739*	0.2421	88.749*
0.8	0.2157	53.311*	0.1889	74.089*
0.9	0.1723	45.949*	0.1459	60.820*

From the table, we can see that the mixing parameter in case of global stationarity can generate a long memory behavior. To know how this behavior depends on the choice of the autoregressive parameter $\phi_{1,2}$, we do simulation by setting $\phi_{1,2} = 0.9$. This shows that the more persistent the autoregressive part of the process is (a higher value of ϕ close to unity), the higher is the possibility that long memory will appear. This can be seen from the table below.

Table 9: GPH estimator for TAR processes with $\phi_{1,1} = 1$ and $\phi_{1,2} = 0.9$

	$T = 200$		$T = 600$	
$\phi_{0,1}$	d	$t - stat$	d	$t - stat$
0.1	0.8165	237.264*	0.7954	309.919*
0.2	0.8040	258.839*	0.7696	347.590*
0.3	0.7979	260.175*	0.7633	366.930*
0.4	0.7911	262.086*	0.7518	396.365*
0.5	0.7869	266.074*	0.7433	419.715*
0.6	0.7848	283.256*	0.7436	402.063*
0.7	0.7842	266.810*	0.7383	426.123*
0.8	0.7799	280.985*	0.7360	427.559*
0.9	0.7811	275.666*	0.7340	428.562*

All GPH estimators are biased towards long memory. This result is also consistent under condition (3). Below you find the result under condition (4), where $\phi_{1,1} = 1, \phi_{1,2} = 1$ with various values of $\phi_{0,1}$ and $\phi_{0,2}$.

Table 10: GPH estimator for TAR processes with $\phi_{1,1} = 1, \phi_{1,2} = 1$

	$T = 200$		$T = 600$	
$\phi_{0,1} = -\phi_{0,1}$	d	$t - stat$	d	$t - stat$
0.1	0.9208	299.877*	0.9364	487.486*
0.2	0.8590	251.864*	0.8479	376.124*
0.3	0.7759	205.877*	0.7429	300.329*
0.4	0.6733	172.456*	0.6406	244.489*
0.5	0.5805	148.509*	0.5377	199.488*
0.6	0.4871	129.712*	0.4448	169.450*
0.7	0.4034	106.050*	0.3629	144.674*
0.8	0.3330	91.155*	0.2915	121.350*
0.9	0.2728	73.922*	0.2317	98.254*

Again, all GPH estimators are biased towards long memory, either stationary or non-stationary. The results above are obtained under the bandwidth $J \sim o(T^{0.8})$. Using $J \sim o(T^{0.5})$ might give different result. However, we examine only $J \sim o(T^{0.8})$ due to reasons mentioned in the previous subsection.

Now consider a special case of SETAR models given in Dufrenot et al. (2005) as follow:

$$X_t = \begin{cases} (1-B)^{-d}\epsilon_t^{(1)} & \text{if } X_{t-1} \leq c \\ \epsilon_t^{(2)} & \text{if } X_{t-1} > c \end{cases} \tag{12}$$

The similar model was considered by Guégan (2004). This model has the specific characteristic that one regime has long memory dynamics and the other has weak dependencies. The switching in the regimes determines the autocovariance function and the spectral density of the process. The autocovariance function of (12) can be expressed as

$$\gamma(\tau) \sim \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\tau^{2d-1}, \ as \ \tau \to +\infty \tag{13}$$

which is not summable and the spectrum has the following representation

$$f(\lambda) \sim C\lambda^{-2d}+, \ as \ \lambda \rightarrow 0 \tag{14}$$

where C is a positive constant. We see that at zero frequency the spectrum f goes to infinite. This indicates that long memory dominates asymptotically. The existence of long memory is induced by the switching behavior across the two regimes. If regime 1 is more frequently visited by the observations than regime 2, then the autocorrelations will decay slowly and the spectral density at frequencies near zero will have high values. The opposite condition results to the short memory process.

5 Conclusion

This paper has been written to give the reader a clear description in a structural way about the existence of spurious long memory in some nonlinear processes which are most interesting in practice. The paper makes the following contributions. First, general Markov switching model as well as mean shift process can mimic long memory. This mimicking phenomena emerges under certain settings of the parameters. Long memory processes more likely emerge in case of transition probabilities close to unity, indicating that the process is becoming more persistent. Second, threshold models are clearly able to generate spurious long memory under locally or globally stationarity conditions especially if the process is highly persistent. Third, the GPH estimator itself introduces a bias and the choice of the bandwidth frequency plays an important role in generating spurious long memory. The bias decreases with an increasing sample size.

References

[1] Agiakloglou, C., Newbold, P., and Wohar, M., Bias in an estimator of the fractional difference parameter, *Journal of Time Series Analysis,* **14** (1993),

235-46.

[2] Andel, J., A time series model with suddenly changing parameters. *Journal of Time Series Analysis,* **14** (1993), 111-123.

[3] Banarjee, A. and Urga, G., Modelling structural breaks, long memory and stock market volatility: an overview, *Journal of Econometrics,* **129** (2005), 1-34.

[4] Beran, J., *Statistics for long memory processes,* Chapman & Hall, New York, 1994.

[5] Breidt, F. J. and Hsu, N. J., A class of nearly long memory time series models, *International Journal of Forecasting,* **18** (2002), 265-281

[6] Carrasco, M., Misspecifed structural change, threshold, and markov-switching models, *Journal of Econometrics,* **109** (2002), 239-273.

[7] Chan, K.S., Consistency and limiting distribution of the least square estimator of a threshold autoregressive model, *Annals of Statistics,* **21** (1993), 520-533.

[8] Choi, S., and Wohar, M.E., The performance of the GPH estimator of the fractional difference parameter : simulation results, *Review of Quantitative Finance and Accounting,* **2** (1992), 409-417.

[9] Davidson, J. and Sibbertsen, P., Generating scheme for long memory process : Regimes, Aggregation and linearity, *Journal of Econometrics,* **127(2)** (2005), 253-282.

[10] Deo, R., Hseih, M., Hurvich, C.M., and Soulier, P., Long memory in nonlinear processes, *Lecturer notes in Statistics,* **187** (2006), Springer.

[11] Dick van Dijk, D., Teräsvirta, T., and Franses, P.H., Smooth transition autoregressive models - A survey of recent developments, *Econometrics Reviews,* **21** (2002), 1-47.

[12] Dieblod, F.X., and Inoue, A., Long memory and regime switching, *Journal of Econometrics,* **105** (2001), 131-159.

[13] Dufrenot, G. Guégan, D. and Peguin, A., Modeling squarec returns using a SETAR model with long memory dynamics, *Economics Letters,* **86** (2005), 237-243

[14] Geweke, J. and Porter-Hudak, S., The estimation and application of long memory time series models, *Journal of Time Series Analysis,* **4** (1983), 221-238.

[15] Gourieroux, C., *ARCH models and financial applications,* Springer, New York, 1997.

[16] Gourieroux, C. and Jasiak, J., Memory and infrequent breaks, *Economics Letters,* **70** (2001), 29-41.

[17] Granger, C.W.J. and Ding, Z., Varieties of long memory models, *Journal of Econometrics,* **73** (1996), 61-77.

[18] Granger, C.W.J. and Hyung, N., Occasional structural breaks and long-memory, *Journal of Empirical Finance,* **11** (2004), 399-421.

[19] Granger, C. J. W. and Joyeux, R., An introduction to long memory time series and fractional differencing, *Journal of Time Series Analysis,* **1** (1980), 15-29.

[20] Guégan, D., How can we define the concept of long memory? an econometric survey, *Working Paper of Queensland University of Technology,* **No. 13**, (2004).

[21] Guégan, D. and Rioublanc, S., Study of regime switcing models. Do they provide long memory behavior? an empirical approach, *Note de Recherche IDHE-MORA,* **No. 13**, (2003).

[22] Guégan, D. and Rioublanc, S., Regime switching models: real or spurious long memory?, *Note de Recherche IDHE-MORA,* **No. 02**, (2005).

[23] Hamilton, J.D., A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica,* **57** (1989), 357-384.

[24] Hosking, J. R. M., Fractional differencing, *Biometrika,* **68** (1981), 165-176.

[25] Hurvich, C.M., Deo, R. and Brodsky, J., The mean squared error of Geweke and Porter Hudak's estimator of a long memory time series, *Journal of Time Series Analysis,* **19** (1998), 19-46.

[26] Kuswanto, H. and Sibbertsen, P., Can we distinguish between common nonlinear time series models and long memory?, *Working paper of Department of Economic and Management, Hannover University,* **380** (2007).

[27] Leipus, R., Paulauska, V. and Surgailis, D., Renewal regime switching and stable limit laws, *Journal of Econometrics,* **129** (2005), 299-327.

[28] Leipus, R. and Surgailis, D., Random coefficient autoregression, regime switching and long memory, *Advances in Applied Probability,* **35** (2003), 737-754.

[29] Liu, M., Modeling long memory in stock market volatility, *Journal of Econometrics,* **99** (2000), 139-171.

[30] Parke, W. R., What is fractional integration, *Review of Economics and Statistics,* , **81** (1999), 632-638.

[31] Robinson, P.M., Statistical inference for a random coefficient autoregressive model, *Scandinavian Journal of Statistics,* **5** (1978), 163-168.

[32] Sibberten, P., Long-memory versus structural change: An overview, *Statistical Papers* **45** (2004), 465 - 515.

[33] Smith, A., Why regime switching creates the illusion of long memory, *Department of Agriculture and Resource Economics, University of California, Davis, mimeo,* (2002).

[34] Timmermann, A., Moments of Markov swiching models, *Journal of Econometrics,* **96** (2000), 75-111.

[35] Yao, J. and Attali, J. G., On stability of nonlinear AR process Markov switching, *Advance in Applied Probability,* **32** (2000), 394-407.