Can we distinguish between common nonlinear time series models and long memory?

Heri Kuswanto and Philipp Sibbertsen[1]
Leibniz Universität Hannover, Department of Economics, Institute of Statistics

Abstract

We show that specific nonlinear time series models such as SETAR, LSTAR, ESTAR and Markov switching which are common in econometric practice can hardly be distinguished from long memory by standard methods such as the GPH estimator for the memory parameter or linearity tests either general or against a specific nonlinear model. We show by Monte Carlo that under certain conditions, the nonlinear data generating process can have misleading either stationary or non-stationary long memory properties.

KEYWORDS: Nonlinear models, long - range dependencies

JEL - classifications: C12, C22

1 Introduction

Long memory attracts attention among practical and theoretical econometricians in the recent years. In econometrics it is mainly applied to model financial time series such as volatilities of stock returns and exchange rate dynamics.

[1]Corresponding author, Email: sibbertsen@statistik.uni-hannover.de

However, so far it is not clear whether the evidence of long - range dependencies in economic time series is due to a real long memory or whether it is because of other phenomena such as structural breaks. Recent works show that structural instability may produce spurious evidence of long memory. Dieblod and Inoue(2001) show that stochastic regime switching can easily be confused with long memory. Davidson and Sibbersten(2005) prove that the aggregation of processes with structural breaks converges to a long - memory process. For an overview about the problem of misspecifying structural breaks and long - range dependence see Sibbertsen(2004). Whereas these papers consider regime switching in the sense of a structural break in the mean of the process there can be many other ways of regime switching leading to the various nonlinear models such as TAR, STAR or Markov - Switching which are considered in this paper. Carasso (2002) shows that simply testing for structural breaks might lead to a wrong usage of linear models although the true data generating process is a nonlinear Markov Switching model.

Granger and Ding (1996) pointed out that there are a number of processes which can also exhibit long memory, including generalized fractionally integrated models arising from aggregation, time changing coefficient models and nonlinear models as well. Granger and Teräsvirta (1999) demonstrate that by using the fractional difference test of Gewecke and Porter - Hudak (1983), a simple nonlinear time series model, which is basically a sign model, generates an autocorrelation structure which could easily be mistaken to be long memory. In this paper, we examine specific nonlinear time series models which are short memory and show by Monte Carlo that they can hardly be distinguished from long memory by standard methodology. In order to do this we estimate the long memory parameter by applying the Gewecke and Porter - Hudak(1983) (further on denoted by GPH) estimator to the nonlinear SETAR, LSTAR, ESTAR and Markov switching model. It turns out that not accounting for the nonlinear structure will bias the GPH estimator and give evidence of long memory. On the other hand we generate linear long memory time series and apply linearity tests to them. We apply the general Teräsvirta Neural Network test as well as linearity tests constructed specially for the considered nonlinear models. It turns out that none of these tests can correctly specify the linear structure of the long memory process. All of these tests are biased towards a rejection of linearity. As a result nonlinearity and long - range dependence are two phenomena which can easily be misspecified and standard methodology is not able to distinguish between these phenomena.

This paper is organised as follows. Section 2 presents briefly the concept of long memory, an overview of the nonlinear time series models used in this paper is given in section 3. The results of our Monte Carlo study are presented in section 4 and 5 and section 6 concludes.

2 Long memory, GPH estimator and rescaled variance test

Long memory or long-range dependence means that observations far away from each other are still strongly correlated. A stationary time series $Y_t, t = 1, \ldots, T$ exhibits long memory or long-range dependence when the correlation function $\rho(k)$ behaves for $k \to \infty$ as

$$\lim_{k \to \infty} \frac{\rho(k)}{c_\rho k^{2d-1}} = 1 \tag{1}$$

Here c_ρ is a constant and $d \in (0, 0.5)$ denotes the long memory parameter. The correlation of a long memory process decays slowly that is with a hyperbolic rate. For $d \in (-0.5, 0)$ the process has short memory. In this situation the spectral density is zero at the origin and the process is said to be antipersistent. For $d \in (0.5, 1)$ the process is non-stationary but still mean reverting. Further discussion about long memory can be found for example in Beran (1994).

A popular semiparametric procedure of estimating the memory parameter d is the GPH estimator introduced by Geweke / Porter-Hudak (1983). It is based on the first m periodogram ordinates

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

where $\lambda_j = 2\pi j/N$ and m is a positive integer smaller than N. 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 $\{\log I_j : j = 1, \ldots, m\}$ on a constant and the regressor variable

$$X_j = \log|1 - \exp(-i\lambda_j)| = \frac{1}{2}\log(2 - 2\cos\lambda_j). \tag{3}$$

By definition the GPH estimator is

$$\hat{d}_{GPH} = \frac{-0.5 \sum_{j=1}^{m} (X_j - \bar{X}) \log I_j}{\sum_{j=1}^{m} (X_j - \bar{X})^2} \tag{4}$$

where $\bar{X} = \frac{1}{m} \sum_{j=1}^{m} X_j$.

This estimator can be motivated using the model:

$$\log I_j = \log c_f - 2dX_j + \log \xi_j \tag{5}$$

where X_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. Besides simplicity another advantage of the GPH-estimator is that it does not require a knowledge about further short-range dependencies in the underlying process. Referring to Hurvich et al. (1998) to get the optimal MSE, we include $o(N^{0.8})$ frequencies in the regression equation.

As an alternative to the GPH estimator we also apply a nonparametric V/S test proposed by Giraitis et al. (2003) to the series. The V/S statistic has better power properties than either the R/S statistic by Mandelbrot/Wallis(1969) or the modified R/S of Lo (1991). Defining $S_k^* = \sum_{j=1}^{k}(X_j - \bar{X})$ as the partial sums of the observations with the sample variance $\widehat{Var}(S_1^*, ..., S_N^*) = N^{-1}\sum_{j=1}^{N}(S_j^* - \bar{S}_N^*)^2$, the V/S statistic is given by

$$Q_N = N^{-1}\frac{\widehat{Var}(S_1^*, ..., S_N^*)}{\hat{s}_{N,q}^2} \tag{6}$$

with

$$\hat{s}_{N,q}^2 = \frac{1}{N}\sum_{j=1}^{N}(X_j - \bar{X}_N)^2 + 2\sum_{j=1}^{q}\omega_j(q)\hat{\gamma}_j. \tag{7}$$

In (7), $\omega_j(q) = 1 - \frac{j}{q+1}$ are the Bartlett weights. The clasical R/S statistic of Mandelbrot/Wallis (1969) corresponds to $q = 0$. We consider the statistic for several different values of q including the optimal q proposed by Andrews(1991).

3 Nonlinear time series models

Nonlinear time series models have become popular in recent years and are widely used in applied macroeconometrics. This paper analyzes three types of models that are most commonly used in nonlinear modelling particularly in modelling economic and financial time series. These include self exciting threshold autoregressive (SETAR), smooth transition autoregressive (STAR) and Markov switching models. These are regime switching models. They share the property of being mean reverting with a long - memory process and they also mimic the persistence of long - range dependent models by exhibiting only short - range dependencies. Therefore, these models are natural candidates to be misspecified with long memory. In the following they are briefly introduced.

The SETAR model by Tong(1983) has been widely considered in the econometric literature as it is a very simple though extremely flexible nonlinear time series model. A time series Y_t is said to be a self-exciting threshold autoregressive (SETAR) process if it follows the model

$$Y_t = \Phi_0^{(j)} Y_{t-1} + a_t^{(j)}, \quad c_{j-1} \leq Y_{t-l} < c_j \tag{8}$$

where $j = 1, \ldots, k$ and l is the lag parameter. The thresholds are $-\infty = c_0 < c_1 < \ldots < c_k = \infty$. SETAR models drive the regime switching by themselves by looking in which interval the observations with lag l are. However, the switches are rapid. Since for $p \geq 2$ these models can mimic a cyclical behaviour, they are expected to be particularly applicable to series with a strong cyclical component. Tong(1990) gives a thorough discussion of these models.

The smooth transition autoregressive (STAR) model is a regime switching model similar to the SETAR model but allowing for a smooth transition between the regimes. It has been considered in detail for example by Teräsvirta(1994). Generally, a STAR process of order p is defined by

$$y_t = \phi' x_t[1 - G(s_t; \gamma, c)] + \theta' x_t G(s_t; \gamma, c) + a_t, \tag{9}$$

where $x_t' = (1, y_{t-1}, \ldots, y_{t-p})$ is an $((p+1) \times 1)$ vector containing lagged values of y_t and $\phi' = (\phi_0, \phi_1, \ldots, \phi_p)$ and $\theta' = (\theta_0, \theta_1, \ldots, \theta_p)$ are parameter vectors of the same dimension. a_t is a Gaussian white noise, $G(s_t, \gamma, c)$ is the transition function governing the movement from one regime to another and s_t is a transition variable so that $s_t = y_{t-l}$.

According to Taylor, Peel and Sarno (2001), the transition variable is commonly chosen to be lagged by one period that is $l = 1$. This is what we use in this paper as well. The variable γ determines the degree of curvature of the transition function and c is a threshold parameter.

The exponential transition function can be written as:

$$G(s_t; \gamma, c) = 1 - \exp\{-\gamma(s_t - c)^2\} \tag{10}$$

with $\gamma > 0$. The limiting cases are $\lim_{\gamma \to \infty} G(s_t; \gamma, c) = 0$ and $\lim_{\gamma \to \infty} G(s_t; \gamma, c) = 1$. Generally speaking, the transition function could be either a logistic function (resulting in LSTAR), or an exponential function (resulting in ESTAR). And the logistic transition function can be written as :

$$G(s_t; \gamma, c) = \frac{1}{1 + \exp(-\gamma(s_t - c))}. \tag{11}$$

This function is asymmetric as it does not depend on the fact that the transition variable moves above or below the threshold. The parameter γ controls the degree of nonlinearity. When $\gamma \to 0$, the transition function tends towards 0, and the model will be a simple autoregressive process. When $\gamma \to \infty$, the transition function converges towards unity, which implies that the model is a different autoregressive model with coefficients equal to the mean of the autoregressive parameters of the two regimes. A survey about recent developments related to STAR models can be found in van Dijk et al. (2002).

The last class of regime switching models we consider in this paper are Markov switching models developed by Hamilton (1989). In this model class, nonlinearities arise as discrete shifts between the regimes. Most importantly these shifts are breaks in the mean of the process. By permitting switching between N regimes, in which the dynamic behaviour of series is markedly different, more complex dynamic patterns can be described.

In the following we will focus on a two-regime Markov switching AR model of order one. The general form of the model is given by

$$y_t = \alpha^{s(t)} + \phi^{s(t)}(L)y_{t-1} + \sigma^{s(t)}a_t \tag{12}$$

where s_t is an m-state Markov chain taking values $1, \dots, m$, with transition matrix P. $\phi_i(L), i = 1, \dots, m$ is a lag polynomial of order $p_1, \dots, p_m$ respectively.
The switching mechanism is controlled by an unobservable state variable that follows a first order Markov chain. Thus, the probability that the state variable s_t equals some particular value j depends on the past only through the most recent value s_{t-1}:

$$P\{s_t = j | s_{t-1} = i, s_{t-2} = k, \dots\} = P\{s_t = j | s_{t-1} = i\} = p_{ij} \tag{13}$$

The transition probability p_{ij} gives the probability that state i will be followed by step j.

Investigating whether nonlinear models can be misspecified as long memory contains two steps. First, we show that nonlinearity leads to a bias in estimators for the memory parameter. Second, we show that standard linearity tests reject the null of a linear process when the data exhibits long - range dependence.

4 Testing for long memory

In this section, we simulate various data generating processes from the above nonlinear time series models, apply the V/S test and estimate the long memory parameter. All nonlinear models considered in our Monte Carlo study are stationary and short - range dependent. The autoregressive order is chosen to be one. Each model is simulated with 1000 replications and different sample sizes of $N = 250$ and $N = 600$ after discarding the first 200 observations to minimize the effect of the initial value of the simulated series. The error terms are modeled to be nid $(0, \sigma_\varepsilon^2)$.

This procedure is used to investigate whether the considered short memory nonlinear models could be detected as to exhibit long memory. We apply the V/S test to the models and compute the rejection probabilities. When applying the V/S test we do this for several values of $q = 0, 5, 10, 25$ and the q following Andrews(1991). They are denoted by q_1, q_2, q_3, q_4 and q_5 respectively. It should be kept in mind when interpreting the simulation results below that by construction of the V/S - test the rejection probability decreases for an increasing value of q.

For our simulation experiments we at first consider the simple 2 regimes SETAR process as follows:

$$y_t = \begin{cases} \phi_1 y_{t-1} + a_t & \text{if } y_{t-1} \leq 0 \\ \phi_2 y_{t-1} + a_t & \text{if } y_{t-1} > 0 \end{cases} \tag{14}$$

and we restrict our consideration on stationary nonlinear processes. We use $\phi_1 = -\phi_2$. The table below presents the rejection probabilities of the V/S test. All rejection probabilities are given to the 5% level.

Table 1: Rejection probabilities of V/S test for the SETAR process

	$N = 250$					$N = 600$				
$\phi_1 = -\phi_2$	q_1	q_2	q_3	q_4	q_5	q_1	q_2	q_3	q_4	q_5
0.1	0.05	0.045	0.034	0.012	0.04	0.054	0.05	0.047	0.031	0.037
0.2	0.061	0.047	0.031	0.01	0.049	0.069	0.055	0.049	0.038	0.04
0.3	0.084	0.04	0.034	0.005	0.041	0.082	0.053	0.05	0.043	0.044
0.4	0.112	0.05	0.036	0.008	0.041	0.107	0.055	0.049	0.034	0.052
0.5	0.148	0.049	0.04	0.011	0.055	0.172	0.066	0.06	0.032	0.063
0.6	0.23	0.065	0.048	0.006	0.047	0.26	0.072	0.049	0.036	0.063
0.7	0.408	0.071	0.046	0.013	0.061	0.424	0.103	0.074	0.043	0.076
0.8	0.687	0.138	0.071	0.012	0.037	0.645	0.142	0.075	0.039	0.085
0.9	0.944	0.309	0.13	0.019	0.019	0.969	0.321	0.177	0.053	0.084

For q_1, which is the classical R/S test, the test tends to reject the null hypothesis too often under both sample sizes. The rejection probability increases with an increasing autoregressive parameter. Using small lags (q_1, q_2 and q_3) the test has a strong bias towards rejecting the nonlinear short memory null hypothesis. The longer the lag

q, the lower is the probability to reject the null in general as we expected. We see that q_4 has the lowest probability compared to the others. Interestingly, q_5 which is considered to be the optimal q rejects the null hypothesis with a probability of around 5% and therefore gives reasonable values.

The following table presents the GPH estimator in order to see whether the GPH estimator is biased towards long memory[2].

Table 2: GPH estimator for the SETAR process

	$N = 250$		$N = 600$	
$\phi_1 = -\phi_2$	d	$t - stat$	d	$t - stat$
0.1	-0.004	-2.5468	-0.0024	-1.5664
0.2	0.0109	5.3870	0.0097	5.9102
0.3	0.0342	15.257	0.0215	14.0443
0.4	0.0662	32.3718	0.0481	31.6628
0.5	0.1137	51.3184	0.0865	53.6984
0.6	0.1676	73.9037	0.1428	87.9102
0.7	0.2566	108.9098	0.2199	126.8156
0.8	0.3750	148.9203	0.3419	183.3246
0.9	0.5360	203.8089	0.5291	287.3747

It can be seen that the GPH estimator indicates either stationary or non-stationary long memory for the SETAR process. In most cases the GPH estimator is in the stationary long memory region. Only for $\phi_1 = -\phi_2 = 0.1$, the GPH estimator is not significantly different from zero according to the t - statistic. The memory parameter increases with the autoregressive parameter. Increasing the sample size does not reduce the bias significantly.

[2]We use $m = o(N^{0.8})$ as number of frequencies employed for the estimation as Hurvich et al (1998) proved that this rate results in an optimal MSE. However, we did also the simulation with $m = N^{0.5}$ as originally proposed by Geweke/Porter-Hudak (1983) for a comparison. The results indicate that the GPH estimator might be biased towards long memory for a higher amount of frequencies used. This is in line with the findings of Davidson and Sibbertsen (2006)

As the GPH estimator is computed by means of the periodogram it seems useful to compare the periodograms of the nonlinear process and the long - memory process. The upper panel of Figure 4 in the appendix presents a sample ACF plot and the periodogram of the SETAR model with $\phi_1 = -\phi_2 = 0.8$. The lower panel shows the ACF and the periodogram of a long - memory process with the same memory parameter $d = 0.3419$ as estimated above. The periodograms of these two DGPs do not show much significant difference. The periodogram of the nonlinear process seems to be more flat near the origin. However, the ACF of the SETAR model shows even more pronounced correlations than the ACF of the long memory process indicating also long term correlations in the nonlinear time series model.

After considering SETAR models we examine both types of STAR models, LSTAR as well as ESTAR. We use the transition variable $s_t = y_{t-1}$ and $c = 0$. The degree of non-linearity in the LSTAR / ESTAR model is determined by the parameter γ in the transition function. Thus, we use two values of γ to examine the behavior of the GPH estimator depending on the transition function. The parameters under consideration are $\gamma = 5$ and $\gamma = 25$.

The model equation for the LSTAR or ESTAR series is given by:

$$y_t = \phi_1 y_{t-1} - (\phi_1 - \phi_2) y_{t-1} F(y_{t-1}, \gamma) + a_t. \tag{15}$$

The tables below give the results for the V/S test for ESTAR processes with both γ[3].

[3] We do not present the results for the LSTAR processes since they are similar to the ESTAR results.

Table 3: Rejection probabilities of V/S test for ESTAR ($\gamma = 5$)

	N=250					N=600				
$\phi_1 = -\phi_2$	q_1	q_2	q_3	q_4	q_5	q_1	q_2	q_3	q_4	q_5
0.1	0.102	0.044	0.03	0.008	0.05	0.087	0.05	0.049	0.038	0.058
0.2	0.154	0.065	0.052	0.013	0.051	0.154	0.072	0.054	0.043	0.057
0.3	0.221	0.051	0.033	0.006	0.052	0.277	0.068	0.048	0.036	0.065
0.4	0.396	0.082	0.061	0.01	0.06	0.362	0.07	0.051	0.042	0.069
0.5	0.511	0.094	0.062	0.015	0.069	0.536	0.09	0.057	0.031	0.06
0.6	0.679	0.124	0.063	0.017	0.048	0.722	0.139	0.078	0.05	0.058
0.7	0.849	0.179	0.082	0.019	0.044	0.858	0.189	0.107	0.052	0.063
0.8	0.969	0.29	0.139	0.027	0.042	0.978	0.339	0.168	0.064	0.064
0.9	0.998	0.587	0.335	0.056	0.025	1	0.665	0.387	0.144	0.06

Table 4: Rejection probabilities of V/S test for ESTAR ($\gamma = 25$)

	N=250					N=600				
$\phi_1 = -\phi_2$	q_1	q_2	q_3	q_4	q_5	q_1	q_2	q_3	q_4	q_5
0.1	0.098	0.05	0.041	0.014	0.05	0.099	0.054	0.048	0.038	0.058
0.2	0.163	0.053	0.042	0.012	0.054	0.167	0.059	0.05	0.034	0.062
0.3	0.274	0.057	0.031	0.003	0.051	0.283	0.072	0.051	0.039	0.061
0.4	0.392	0.098	0.057	0.012	0.056	0.384	0.084	0.06	0.043	0.06
0.5	0.548	0.103	0.049	0.01	0.063	0.565	0.108	0.068	0.046	0.049
0.6	0.676	0.116	0.058	0.013	0.07	0.738	0.165	0.099	0.062	0.075
0.7	0.869	0.185	0.08	0.015	0.062	0.894	0.236	0.128	0.065	0.073
0.8	0.972	0.305	0.146	0.023	0.058	0.986	0.386	0.205	0.079	0.072
0.9	0.999	0.61	0.332	0.06	0.034	1	0.684	0.389	0.137	0.053

Again, the classical R/S test fails to detect the short memory property for all considered nonlinear processes. Similar to the results of the V/S test for SETAR processes, the rejection probability increases with the autoregressive parameter. For the lag length q_5 the null hypothesis is rejected with a probability around 5% though usually a bit higher in almost all cases. For the lag length q_4 the test shows a better performance but the probability still reaches values above 5% for high autoregressive parameters and a large sample size. This seems also to be rather an artefact of the V/S statistic. Interestingly, changing the transition functions does not change the rejection probability.

Table 5: GPH estimator for the ESTAR process ($\gamma = 5$)

	$N = 250$		$N = 600$	
$\phi_1 = -\phi_2$	d	$t - stat$	d	$t - stat$
0.1	0.0309	11.9327	0.0244	14.5522
0.2	0.0777	30.8505	0.0647	36.5351
0.3	0.1385	54.9238	0.1095	62.9958
0.4	0.2043	78.5431	0.1717	98.3033
0.5	0.2856	111.5377	0.2399	140.2350
0.6	0.3713	181.1988	0.3244	182.7908
0.7	0.4780	227.128	0.4294	242.483
0.8	0.6057	141.1256	0.5611	309.304
0.9	0.7679	290.485	0.7344	404.768

Table 6: GPH estimator for the ESTAR process ($\gamma = 25$)

	$N = 250$		$N = 600$	
$\phi_1 = -\phi_2$	d	$t - stat$	d	$t - stat$
0.1	0.02965	11.4185	0.0263	15.1419
0.2	0.0888	34.9036	0.0689	40.4176
0.3	0.1536	60.7541	0.1219	73.4514
0.4	0.2178	83.0316	0.1812	104.261
0.5	0.3043	118.4546	0.2545	154.1513
0.6	0.3929	152.4049	0.3427	191.467
0.7	0.4999	182.6267	0.4484	249.69
0.8	0.6261	240.7276	0.5782	322.25
0.9	0.7729	292.6791	0.7426	419.5613

Table 5 and 6 show that the GPH estimator is biased towards long memory either stationary or non-stationary depending on the parameter settings for the ESTAR model. Furthermore, even doubling the sample size (increasing the sample from 250 to 600) does not decrease the bias significantly. These results are also robust against changing the γ parameter in the transition function. This confirms the simulation results of Choi and Wohar (1992) which investigate the performance of the GPH estimator if the DGP is a stationary AR(1) process. The GPH estimator is seriously biased with an increasing bias for an increasing value of the autoregressive parameter, even for a relatively large sample size.[4]

Figure 5 in the appendix shows the ACF and periodograms of an ESTAR and LSTAR process with $\gamma = 5$, $\phi_1 = -\phi_2 = 0.6$ and a true long memory process generated by using the according memory parameter as estimated above ($d = 0.3427$). All periodograms show a clear long memory behaviour which is shown by the negative slope of the fitted line. However,the ESTAR process shows the most pronounced peak in the periodogram near the origin indicating some long - memory behaviour. The sample ACFs can hardly be distinguished. However the ACF of the true long memory process seems to decay hyperbolically for the first few lags.

Finally, we investigate the behaviour of the GPH estimator when the true DGP is a Markov switching model. The DGP in this section is simulated based on the general Markov switching process:

$$y_t = \begin{cases} \phi_1 y_{t-1} + a_t & \text{if } S_t = 1 \\ \phi_2 y_{t-1} + a_t & \text{if } S_t = 2 \end{cases} \tag{16}$$

with $u_t \sim NID(0,1)$

In line with other considered nonlinear models, we set $\phi_1 = -\phi_2$ in all of our simulations in order to generate a stationary nonlinear process. The transition probabilities

[4]Choi and Wohar (1992) consider a stationary AR(1) process and use $N^{0.5}$ frequencies for their simulation.

are taken from Hamilton (1989), which are $P = (0.1, 0.25, 0.75, 0.9)$.

Table 7: Rejection probabilities of the V/S test for Markov switching processes

	N=250					N=600				
$\phi_1 = -\phi_2$	q_1	q_2	q_3	q_4	q_5	q_1	q_2	q_3	q_4	q_5
0.1	0.065	0.041	0.03	0.008	0.045	0.064	0.052	0.045	0.037	0.047
0.2	0.109	0.047	0.034	0.007	0.047	0.12	0.066	0.057	0.041	0.049
0.3	0.147	0.05	0.034	0.009	0.038	0.128	0.062	0.05	0.038	0.059
0.4	0.227	0.062	0.039	0.009	0.048	0.229	0.088	0.067	0.038	0.058
0.5	0.306	0.077	0.046	0.006	0.037	0.321	0.099	0.068	0.049	0.083
0.6	0.431	0.116	0.067	0.015	0.041	0.462	0.121	0.079	0.056	0.095
0.7	0.551	0.138	0.055	0.008	0.05	0.607	0.177	0.108	0.069	0.099
0.8	0.723	0.175	0.087	0.013	0.052	0.759	0.196	0.101	0.05	0.157
0.9	0.873	0.3	0.137	0.019	0.03	0.893	0.35	0.172	0.069	0.174

From table 7 we see that the result of the V/S test has a similar tendency as the previous results. However, for Markov switching models the rejection probabilities for q_5 are relatively higher and reach 0.174 for a sample size of $N = 600$.

Table 8: GPH estimator for Markov switching processes

	$N = 250$		$N = 600$	
$\phi_1 = -\phi_2$	d	$t - stat$	d	$t - stat$
0.1	0.0245	11.8715	0.0159	10.2812
0.2	0.0585	27.7900	0.0449	28.6479
0.3	0.0903	40.4370	0.0821	52.2730
0.4	0.1231	51.2194	0.1272	79.0536
0.5	0.1586	61.0404	0.1839	112.041
0.6	0.1952	70.9674	0.2407	144.956
0.7	0.2325	80.5741	0.3136	179.8552
0.8	0.2691	91.5120	0.3916	214.0388
0.9	0.3040	97.4647	0.4774	249.6712

The GPH estimator does not show any surprising result. It is biased towards stationary long memory and increases with the autoregressive parameter but with a relatively slow rate. However, the bias increases with the sample size for a very small amount in contrast to the other processes. These results therefore confirm Smith (2002) who shows that the GPH estimator is substantially biased for a stationary Markov switching process which does not contain long memory.

To investigate the impact of the transition probabilities to the GPH estimator, we consider another Markov process by considering the various P values given above and the parameter setting $\phi_1 = -\phi_2 = 0.9$. We use this autoregressive parameter, since it leads to a higher bias of the GPH estimator and therefore shows the relevant effect more clearly. Table 9 presents the results for the considered process.

Table 9: GPH estimator for the Markov switching process

	$N = 250$		$N = 600$	
$P_{11} = P_{22}$	d	$t - stat$	d	$t - stat$
0.1	-0.1363	-61.264	-0.4441	-223.5914
0.2	-0.1099	-45.2729	-0.3314	-156.9377
0.3	-0.070	-27.990	-0.2216	-102.034
0.4	-0.0437	-16.5335	-0.1174	-52.1006
0.5	-0.0064	-2.4088	-0.0081	-3.7331
0.6	0.034	11.9585	0.1006	43.9457
0.7	0.0769	26.1633	0.2230	100.850
0.8	0.1203	39.9354	0.3453	150.018
0.9	0.1734	58.8143	0.4800	198.999

Note that when $P_{11} = P_{22} = 0.5$ it implies that $P_{11} + P_{22} = 1$ and thus there is no persistence in the Markov process because the probability that s_t switches from state 1 to state 2 is independent of the previous state. This is a rather simple switching model. From the table we see that for some values of the transition probabilities above 0.5 (close to one), they are biased towards stationary long memory and the process is detected as to be short memory when the transition probability is less than 0.5. It is natural since as the parameters approach the non-ergodicity point (when P_{11} and P_{22} are equal one), the AR component gets more persistent and causes the dominant component of the GPH bias (see Smith (2000) for details)

Similar to the other nonlinear models periodograms which are generated from the Markov Switching model do not show much difference than those of the true long memory process (see figure 6 in the appendix). On the other hand we see that the ACF of the Markov Switching model does not decay as slow as the true long memory process.

From the above results, we come to the conclusion that although the process under the null is nonlinear but still a short memory process, the above results for the V/S test are consistent with Lo (1997) and Giraitis et al. (2003), where the probability to reject the short memory null hypothesis is lower for large q, since the imprecision with which the higher order autocovariances can introduce considerable noise into the statistic is reduced. The classical R/S test fails to identify the short memory properties. The Andrews procedure also reject the null relatively often and might reach a probability of more than 5%.

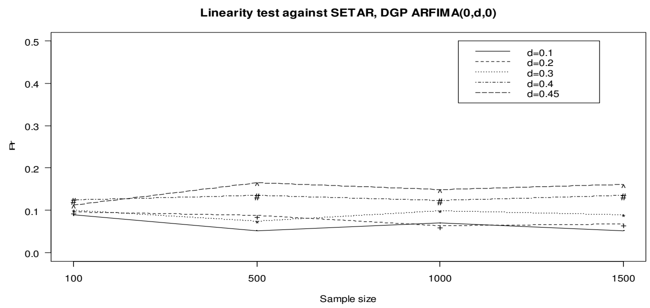
The GPH estimator, which is also quiet popular as a semiparametric procedure to detect long memory fails to distinguish the considered nonlinear processes from long range dependencies. Most of the processes are biased towards long memory. The periodogram of nonlinear and long memory processes behave quite similar near the origin. Thus, we can say that by these quite common tests, it is difficult to distinguish between nonlinearity and long memory. Long memory tests as well as point estimates can lead to a misleading inference. However, using a higher lag order in the V/S test gives more reliable results.

5 Testing linearity

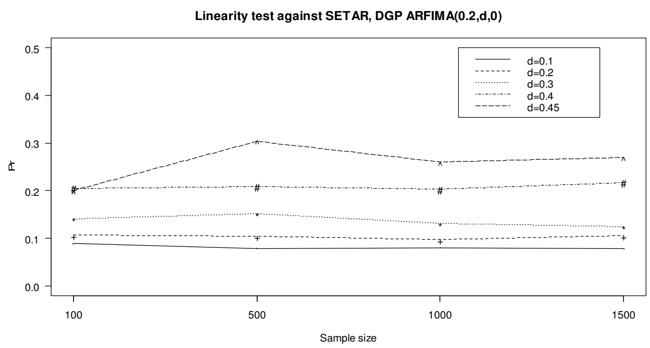
In this section we apply a general linearity test, namely the Neural Network test of Teräsvirta et al. (1991), as well as specific linearity tests constructed to test the null hypothesis of linearity against the alternative of a specific nonlinear structure, namely SETAR or STAR.

We compute the rejection probabilities of the 5% significance level with 10000 replications and various sample sizes $N = 100, 500, 1000$ and 1500.

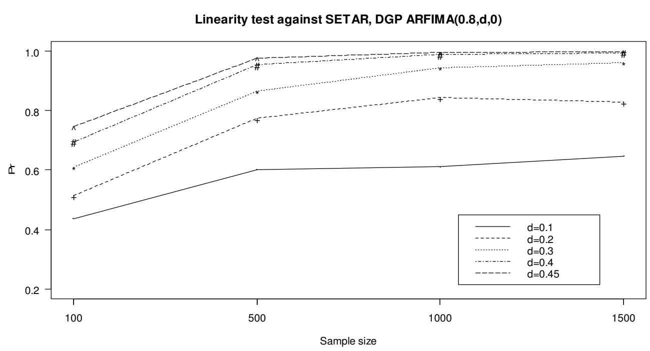
First, we use a portmanteau test in order to test for a SETAR type nonlinearity. For a detailed discussion of this test, see Petrucelli and Davies (1986). This test was also considered by Chan and Ng (2004) who show that the test is not robust against misspecifications of the model. It is also not robust against outliers. Figure 1 shows the rejection probabilities of this test when the true DGP is long memory.
If the DGP is a pure long memory processes (Figure 1(i)) the probability to reject the null hypothesis of linearity reaches a maximum of 0.165. The probability increases with higher values of the memory parameter and larger sample sizes. The same tendency appears when the DGP follows an ARFIMA $(\phi, d, 0)$ process, this is a long memory process with an additional autoregressive root. The rejection probability increases with an increase of the autoregressive parameter. For a value of 0.8 the rejections probabilities are already close to 1 even for moderate sample sizes. This is due to an increase of the persistence of the process induced by the positive autoregressive parameter. However, we clearly see that the portmanteau test is not able to capture the linearity of the long memory DGP.



(i)



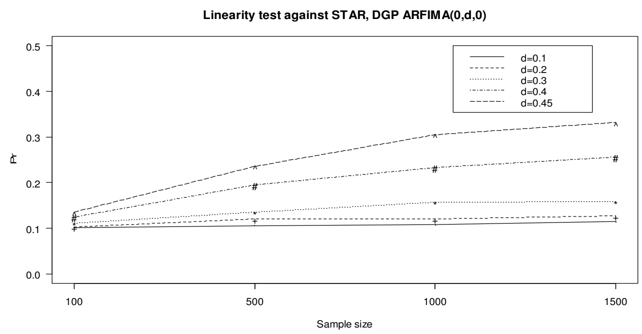
(ii)



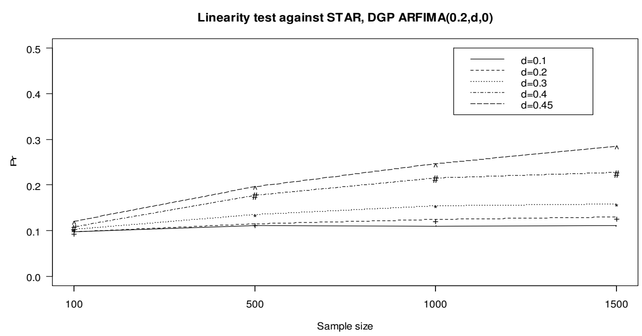
(iii)

Figure 1: Rejection probabilities of linearity test against SETAR model (i) DGP is ARFIMA(0,d,0),(ii) DGP is ARFIMA(0.2,d,0) and (iii) DGP is ARFIMA(0.8,d,0)

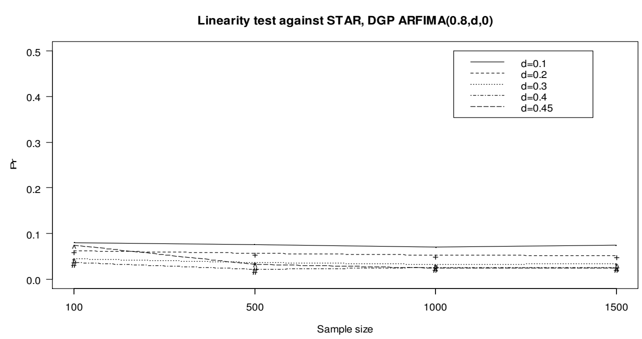
As a second test we consider a linearity test against the STAR alternative. The test is a Lagrange Multiplier type test proposed by Luukkonen et al. (1988). It is based on a third-order Taylor approximation of the transition function. By this procedure, testing against ESTAR is not distinguishable from testing against LSTAR, when a second-order logistic transition function is employed (see also Saikkonen and Luukkonen,1988). Figure 2 below presents the results of the test.



(i)



(ii)



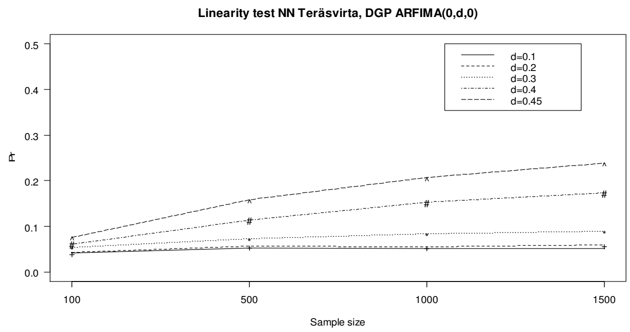
(iii)

Figure 2: Rejection probabilities of linearity test against STAR model (i) DGP is ARFIMA(0,d,0),(ii) DGP is ARFIMA(0.2,d,0) and (iii) DGP is ARFIMA(0.8,d,0)

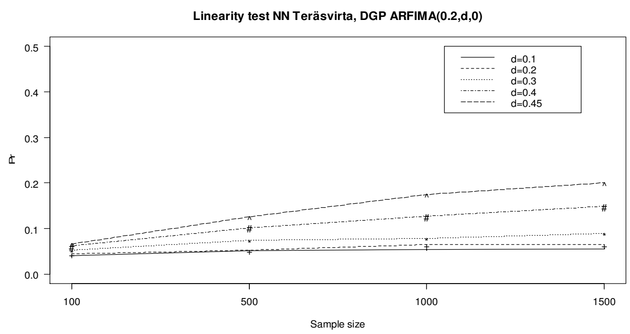
If the DGP is a pure long memory process, the results are similar to those of Anderson et al. (1999). The rejection probability increases with the value of the memory parameter and with the sample sizes. The same results are obtained for an ARFIMA $(\phi, d, 0)$ - process with a small autoregressive parameter ($\phi = 0.2$). The rejection probability reaches a value of up to 0.25 in our study. Interestingly, for the same process but with a higher autoregressive parameter ($\phi = 0.8$) the rejection probability decreases with sample size. It actually collapses even under the nominal size of the test.

Finally, we apply the neural network based linearity test proposed by Teräsvirta et al. (1991). This test is a special neural network model with a single hidden layer. This test is a Lagrange Multiplier (LM) type test derived from a neural network model based on the "dual" of the Volterra expansion representation for nonlinear series.

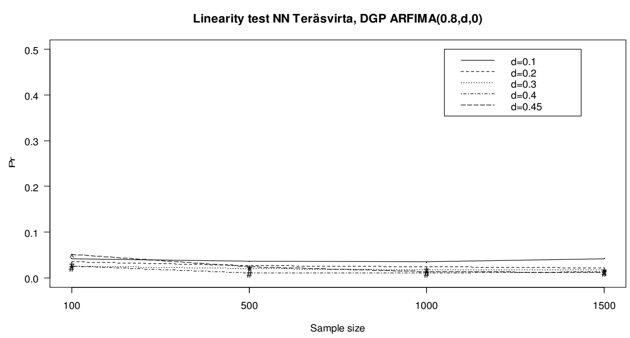
Let consider Figure 3 for the results of this test. The results are similar to those of the STAR test considered before. For a pure long memory DGP as well as for an ARFIMA $(\phi, d, 0)$ - process with a small autoregressive parameter ($\phi = 0.2$), the values of the rejection probability increase with d and with the sample size. Again, for an increasing autoregressive parameter the rejection probability collapses under the nominal size of the test and converges to zero. Since the two tests are Lagrange multiplier test, which involves the estimation of the autoregressive parameter to compute the statistic, the higher AR and d parameter are confounded as a simple AR(1) parameter. This leads to a higher sum of squared errors (SSE_0) in the denominator and the statistic tends to not reject the null hypothesis.



(i)



(ii)



(iii)

Figure 3: Rejection probabilities of linearity test against Markov switching model (i) DGP is ARFIMA(0,d,0),(ii) DGP is ARFIMA(0.2,d,0) and (iii) DGP is ARFIMA(0.8,d,0)

6 Conclusion

In this paper we show by Monte Carlo that popular nonlinear models such as TAR, STAR and Markov - Switching models can easily be misspecified as long memory. We estimate the memory parameter for various specifications of the above models and find that the GPH estimator is positively biased indicating long - range dependence. However, applying the V/S test with an optimal lag - length as suggested by Andrews (1991) seems to give reasonable results. On the other hand do linearity tests reject the null hypothesis of linearity when the true data generating process exhibits long memory with a rejection probability tending to one. The rejection probabilities increase with the memory parameter. This effect is more pronounced for tests against a specific alternative such as TAR or STAR. The more general neural network test shows a favorable behaviour. However, a strong autoregressive root can collapse the rejection probabilities.

Therefore, nonlinear models can easily be misspecified as long - range dependence and vice versa by using standard methodology. Methods for distinguishing between these two phenomena are subject to future research.

References

Anderson, M. K., Eklund, B., and Lyhagen, J., 1999: A simple linear time series model with misleading nonlinear properties. *Economics Letters 65, 281-284*

Andrews, D., 1991: Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59, 817-858*

Beran, J., 1994: Statistics for long memory processes. Chapman & Hall, New York

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

Choi, W.S., and Ng, M.W., 2004: Robustness of alternative non-linearity tests for SETAR models. *Journal of Forecasting 23, 215-231*

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

Davidson, J. and Sibbertsen, P., 2006: Test of Bias in log - periodogram regression. *Working Paper, University of Hannover.*

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

Dieblod, F.X., and Inoue, A., 2001: Long memory and regime switching. *Journal of Econometrics 105, 131-159*

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

Granger, C.W.J. and Ding, Z., 1996: Varieties of long memory models. *Journal of Econometrics 73, 61-77*

Granger, C.W.J, and Teräsvirta T., 1999: A simple nonlinear time series model with misleading linear properties. *Economics Letters 62, 161-165*

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

Hurst, H., 1951: Long-term storage capacity of reservoirs.*Transactions of the American Society of Civil Engineers, 116*, 770-808

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

Lo, A., 1991: Long term memory in stock market prices. *Econometrica 59, 1279-1313*

Giraitis, L., Kokoszka, P., Leipus, R. and Teyssiere, G., 2003: Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics 112, 265-294*

Luukkonen, R., Saikkonen, P. and Teräsvirta, T., 1988: Testing linearity against smooth transition autoregressive models. *Biometrika 75, 491-499*

Mandelbrot, B.B. and Wallis, J.M., 1969: Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research 5, 967-988*

Petrucelli, J.D. and Davies, N., 1986: A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. *Biometrika 73, No 3, 687-694*

Saikkonen, P. and Luukkonen, R., 1988: Lagrange multiplier tests for testing nonlinearities in time series models. *Scand. J. Statist. 15, 55-68*

Sibbertsen, P., 2004: Long-memory versus structural change: An overview. *Statistical Papers 45, 465 - 515*

Smith, A., 2002: Why regime switching creates the illusion of long memory. Department of Agriculture and Resource Economics, University of California, Davis, mimeo.

Taylor, M.P., Peel, D.A. and Sarno, L., 2001: Non-linearity in real exchange toward solution of the purchasing power parity puzzles. *International Economic Review 42, 1015-1042*

Teräsvirta, T., 1994: Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association, 208-218*

Teräsvirta, T., Lin, C.F. and Granger, C.W., 1991: Power of the neural network linearity test. *Journal of Time Series Analysis 14, No. 2, 209-220*

Tong, H., 1983: Threshold model in nonlinear time series analysis. Lecture Notes in Statistics, Springer-Verlag, New York

Tong, H., 1990: Non-linear time series: A dynamical system approach, Oxford University Press, Oxford

Appendix

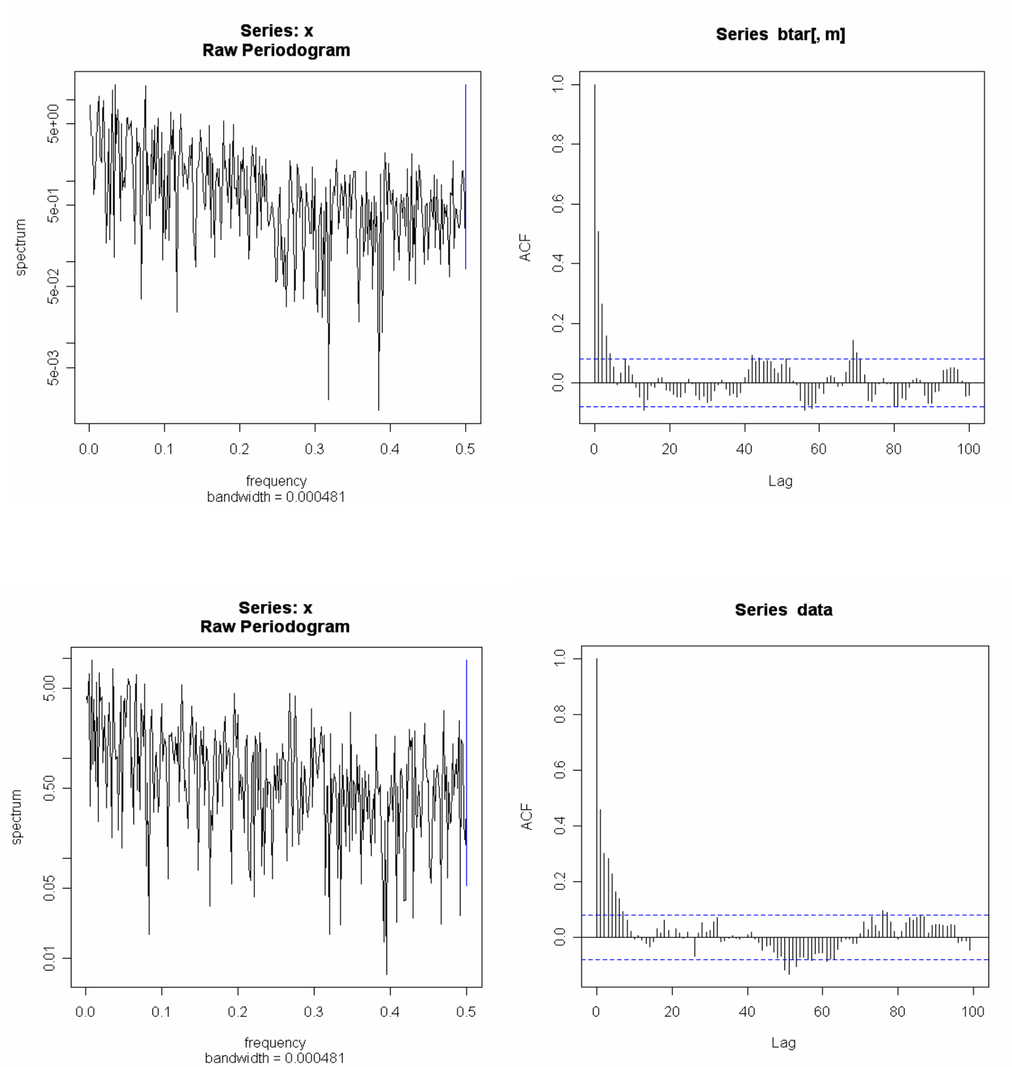
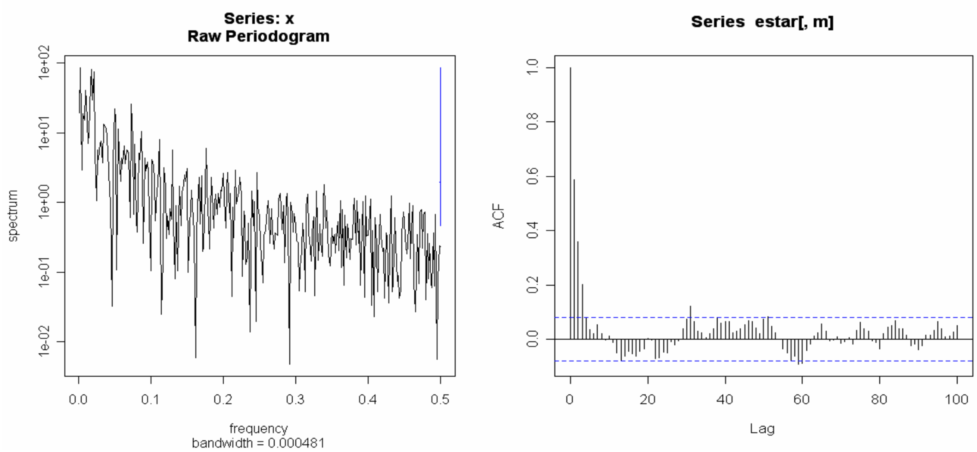
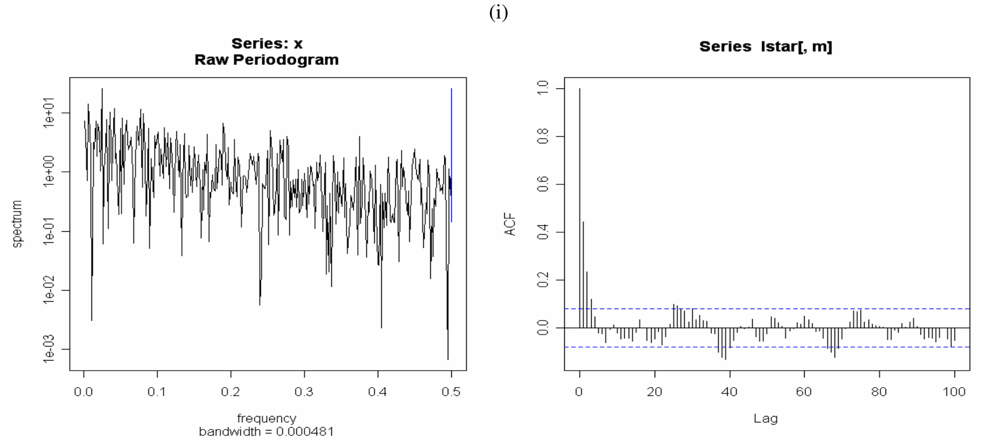


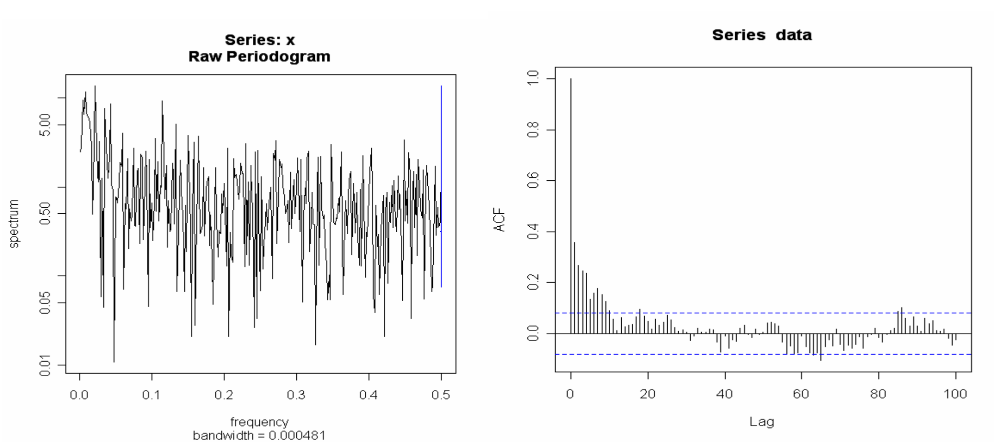
Figure 4: Sample periodograms and ACF plots (i) SETAR process (ii) Long memory with $d = 0.3419$



(i)

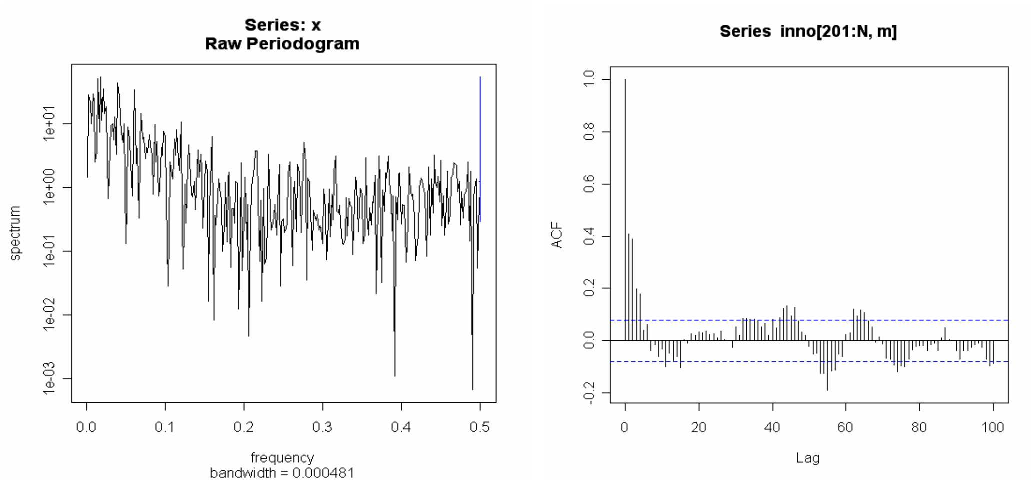


(ii)

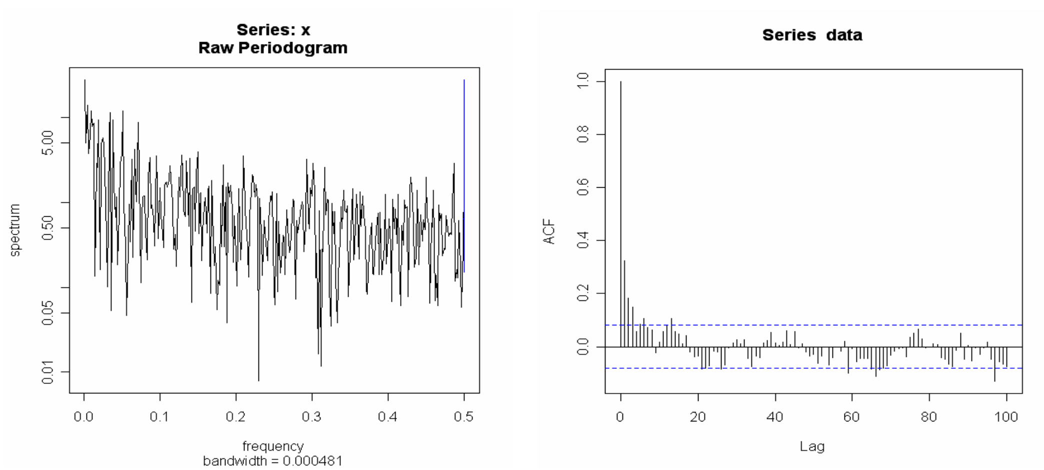


(iii)

Figure 5: Sample periodograms and ACF plots (i) ESTAR process (ii) LSTAR process (iii) Long memory with $d = 0.3427$



(i)



(ii)

Figure 6: Sample periodograms and ACF plots (i) Markov switching process (ii) Long memory with $d = 0.3916$