Essays on the problem of distinguishing between long memory and nonlinear time series

Von der Wirtschaftswissenschaftlichen Fakultät
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenchaften
- Doctor rerum politicarum -

genehmigte Dissertation

von

Heri Kuswanto (M.Sc.)
geboren am 26. 03. 1982 in Gresik (Indonesien)

2009

Referent : Prof. Dr. Philipp Sibbertsen

Korreferent : Prof. Dr. Hendrik Hakenes

Tag der Promotion: 25.06.2009

ABSTRACT

Long memory and nonlinear time series have both been extensively applied in empirical studies on the business cycle and other macroeconomic time series leading to different economic implications. This dissertation considers the problem of confusion between long memory and several nonlinear processes. Chapter 1 demonstrates inability of standard methods to distinguish between these two phenomena. The analysis is done through intensive simulation study. Chapter 2 discusses theoretical explanations about the sources of the confusion.

The following two chapters introduce new tests to distinguish between long memory and several processes. Chapter 3 considers a new test to distinguish between long memory and ESTAR nonlinearities. We develop the test based on two approaches, and use directed-Wald to overcome the problem of restricted parameter under alternative hypothesis. New critical values are provided and intensive simulation study is done to assess the test performance in finite sample size. Moreover, we apply the test to real exchange rate data.

Furthermore, we propose a new test against spurious long memory in chapter 4. The test is developed based on the invariance principle of estimated long memory parameters to aggregation. The test performance in finite sample size is evaluated through simulation study. It indicates that the test has good power, and is able to detect spurious long memory in German stock returns.

Keywords: spurious long memory, nonlinear, directed-Wald, aggregation.

KURZFASSUNG

Eine langfristige Abhängigkeitsstruktur (sog. "lange Gedächtnis"), und nicht lineare Zeitreihen wurden ausführlich in empirischen Studien über Konjunkturzyklen sowie anderen makroökonomischen Zeitreihen angewendet und führen zu unterschiedlichen ökonomischen Implikationen. Die vorliegende Dissertation befasst sich mit der Problematik derVerwechslung zwischen einem "langen Gedächtnis" und verschiedenen nichtlinearen Prozessen. In Kapitel 1 wird gezeigt, dass es mit Standardmethoden nicht möglich ist, zwischen diesen beiden Phänomenen zu unterscheiden. Die Analyse wird mit Hilfe einer intensiven Simulation durchgeführt. Im zweiten Kapitel werden theoretische Erklärungen für die Ursache der Verwechselung diskutiert.Im zwei letzte Kapitel werden neue Tests vorgestellt, mit denen zwischen einem "langen Gedächtnis" und verschiedenen Prozessen unterschieden werden kann.

In Kapitel 3 wird ein neuer Test eingeführt, mit dem zwischen einem "langen Gedächtnis" und ESTAR Nichtlinearitäten unterschieden werden kann. Der von uns entwickelte Test basiert auf zwei Ansätzen und verwendet einen directed-Wald Test, um das Problem beschränkter Parameter bei der Alternativhypothese zu beseitigen. Es werden neue kritische Werte geliefert und eine intensive Simulation durchgeführt, um die Test-Performance für einen begrenzten Stichprobenumfang abzuschätzen. Außerdem wenden wir diesen Test für Daten von realenWechselkursen an.

Darüber hinaus entwickeln wir in Kapitel 4 einen neuen Test um ein unechtes "langes Gedächtnis" zu verifizieren. Der Test wurde auf Basis des Prinzips der Invarianz von geschätzten Parametern des "langen Gedächtnisses" zu aggregation entwickelt.

Test-Performance der begrenzten Stichprobengröße wird anhand einer Simulation ausgewertet. Dabei wird deutlich, dass der Test ein hohes Potential hat und in der Lage ist, unechte "lange Gedächtnisse" in den Renditen deutscher Aktien aufzudecken.

Schlagwörter: lange Gedächtnis, nicht linear, directed-Wald, aggregation.

TABLE OF CONTENTS

ACKNOWLEDGMENT

Firstly, I would like to thank my supervisor Prof. Dr. Philipp Sibbertsen for his excellent support, encouragement and generosity of time over the last three years. I owe to him some very valuable suggestions which are now part of this thesis. I would also like thank Prof. Dr. Hendrik Hakenes and Prof. Dr. Lukas Menkhoff for their kindness and support during my PhD study here.

Thanks also to all the members of the Institute of Statistics, for providing a stimulating and cooperative environment of research and study. I gratefully acknowledge the financial support of the German Academic Exchange Service (DAAD) by way of a PhD scholarship.

Finally, my greatest debt is to my beloved wife "Nora", my beloved son "Ryan" and my family, for their continuing support and encouragement, without which this thesis could not have been written.

Chapter 1

SUMMARY

This dissertation deals with the problem of confusing between long memory and some common classes of nonlinear time series process. It essentially contains of four contributions to the problem. The dissertation pursuits also three major goals. One is to assess the ability of existing methods to distinguish long memory from the nonlinear processes. The second goal is to identify the sources of the confusion, and the last is to develop tests which are able to distinguish long memory and nonlinear processes.

The outline of each chapter is described as follows. The second chapter deals with the first goal. Through an intensive simulation study, 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. This problem is thus known as confusion between long memory and nonlinear processes.

Chapter 3 identifies the sources of the confusion by exploring some theoretical explanations. We describe the asymptotic behavior of the process in terms of the autocovariance and the autocorrelation function, and support the theoretical evidences by providing some Monte Carlo simulations. The existence of long memory in these nonlinear processes is induced by the nature of the process in certain conditions. In

addition, the GPH estimator itself is biased.

Most of the contributions in the dissertation transpire in the forth and fifth chapter. In chapter 4, we develop a Wald type test to distinguish between long memory and ESTAR nonlinearity by using a directed-Wald statistic to overcome the problem of restricted parameters under the alternative. The test is derived from two basic model specifications, where the first is the standard model based on an auxiliary regression and the second allows the parameter γ to appear as a nuisance parameter in the transition function. A simulation study indicates that both approaches lead to tests with good size and power properties to distinguish between stationary long memory and ESTAR. Moreover, the second approach is shown to have more power.

As an empirical application, we apply the test to the problem addressed by Cheung and Lai (2001). They faced difficulties in distinguishing long memory and nonlinear mean reversion in the case of bilateral real exchange rate against Japan YEN. In this thesis, we consider several cases including real exchange rates of developed and developing countries. The proposed tests are able to provide a solution to the problem. Given the fact that nonlinear adjustment towards PPP more likely holds in developing countries, the tests are able to capture this phenomena. We apply also a test proposed by Baillie and Kapetanios (2008) to detect any neglected nonlinearity in long memory processes and the result is consistent with ours.

Finally, the last chapter discusses a new simple test against spurious long memory. The test statistic is developed based on the invariance of long memory parameters to aggregation, combined with the idea of testing for a change in the long memory parameter. By using the local Whittle estimator, the statistic takes a maximum value among combinations of paired aggregated series. Simulations show that the test performs well in finite samples, and is able to distinguish long memory from spurious processes with excellent power. Moreover, the empirical application gives further evidence that the observed long memory in German stock returns is spurious.

Chapter 2

CAN STANDARD TESTS DISCRIMINATE BETWEEN COMMON NONLINEAR TIME SERIES MODELS AND LONG MEMORY?[1]

2.1 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. 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. Diebold and Inoue (2001) show that stochastic regime switching can easily be confused with long memory. Davidson and Sibbertsen (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 (2004b). 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. Here the regimes are different short memory processes, usually of an autoregressive type. Therefore, regime switching in the sense

[1] Co-author: Prof. Dr. Philipp Sibbertsen, Leibniz Universität Hannover Germany. This paper is available as Discussion Paper no. 380 of the discussion paper series of the Faculty of Economics at Leibniz Universität Hannover.

of nonlinear model building is substantially different from the long memory versus structural break case. It is not clear whether nonlinear regime switching processes can produce spurious long memory in the sense that standard tests cannot distinguish between these two models. Carrasco (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 Geweke 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. However, these examples are hardly comparable with nonlinear models used in economic practice. There is by now a huge literature on nonlinear modeling in economics. This literature mainly contains of TAR, STAR and Markov switching models which prove useful especially for modeling exchange rate behaviour. In this paper, we concentrate on these model classes 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 Geweke 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's Neural Network test of Teräsvirta et al. (1993) 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 misspecifed, and standard methodology is not able to distinguish between these phenomena. In this respect we extend the earlier study of Andersson et al. (1999).

This paper is organized as follows. Section 2.2 presents briefly the concept of long memory, an overview of the nonlinear time series models used in this paper is given in section 2.3. The results of our Monte Carlo study are presented in section 2.4 and 2.5 and section 2.6 concludes.

2.2 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

$$\lim_{k\to\infty} \frac{\rho(k)}{C_\rho k^{2d-1}} = 1 \tag{2.1}$$

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 anti-persistent. 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 semi-parametric procedure of estimating the memory parameter d is the GPH estimator introduced by Geweke and Porter-Hudak (1983). It is based on the first J periodogram ordinates

$$I_j = \frac{1}{2\pi T} | \sum_{t-1}^{T} y_t \exp(i\lambda_j t)|^2 \tag{2.2}$$

where $\lambda_j = 2\pi j/T$, $j = 1, \ldots, J$ 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 $\{\log I_j : j = 1, \ldots, J\}$ on a constant and the regressor variable

$$x_j = \log |1 - \exp(-i\lambda_j)| = \frac{1}{2} \log(2 - 2\cos\lambda_j). \tag{2.3}$$

By definition the GPH estimator is

$$\hat{d}_{GPH} = \frac{-0.5 \sum_{j=1}^{J} (x_j - \bar{x}) \log I_j}{\sum_{j=1}^{J} (x_j - \bar{x})^2} \tag{2.4}$$

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

$$\log I_j = \log C_f - 2dx_j + \log \xi_j \tag{2.5}$$

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 $T^{0.8}$ 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. It tests the short memory process under null hypothesis against alternative of long memory process. The V/S statistic has better power properties than either the R/S statistic by Mandelbrot and Wallis (1969) or the modified R/S of Lo (1991). Defining $S_k^* = \sum_{j=1}^{k}(y_j - \bar{y})$ as the partial sums of the observations with the sample variance $\widehat{Var}(S_1^*, ..., S_T^*) = T^{-1}\sum_{j=1}^{T}(S_j^* - \bar{S}_T^*)^2$, the V/S statistic is given by

$$Q_T = T^{-1}\frac{\widehat{Var}(S_1^*, ..., S_T^*)}{\hat{s}_{T,q}^2} \tag{2.6}$$

with

$$\hat{s}_{T,q}^2 = \frac{1}{T}\sum_{j=1}^{T}(y_j - \bar{y}_T)^2 + 2\sum_{j=1}^{q}\omega_j(q)\hat{\gamma}_j. \tag{2.7}$$

In (2.7), $\omega_j(q) = 1 - \frac{j}{q+1}$ are the Bartlett weights. The classical R/S statistic of Mandelbrot and Wallis (1969) corresponds to $q = 0$. We consider the statistic for several different values of q including the optimal q proposed by Andrews (1991).

2.3 Nonlinear time series models

Nonlinear time series models have become popular in recent years and are widely used in applied macro-econometrics. This paper analyzes three types of models that are most commonly used in nonlinear modeling particularly in modeling 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.

Unlike the simple autoregressive processes, SETAR model allows the model parameters to change according to the value of threshold variable y_{t-l}:

$$y_t = \mathbf{X_t}\phi^{(j)} + \epsilon_t^{(j)}, \quad c_{j-1} \leq y_{t-l} < c_j \tag{2.8}$$

where $\mathbf{X_t} = (1, y_{t-1}, y_{t-1}, ..., y_{t-p})$, $\phi = (\mu, \phi_1, \phi_2, ..., \phi_p)'$, $j = 1, 2, ..., k$ and $-\infty = c_0 < c_1 < \ldots < c_k = \infty$. In essence, the $k-1$ nontrivial thresholds $(c_1, c_2, ..., c_{k-1})$ divide the domain of the threshold variable into k different regimes. In each different regime, the time series y_t follows a different autoregressive model. In the threshold variable, the delay parameter l being a positif integer and the lagged value y_{t-l} determine the dynamic or regime of y_t. 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

$$y_t = \mathbf{X_t}\phi[1 - G(s_t; \gamma, c)] + \mathbf{X_t}\theta G(s_t; \gamma, c) + \epsilon_t, \tag{2.9}$$

where $\mathbf{X_t} = (1, y_{t-1}, \ldots, y_{t-p})$ 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. ϵ_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:

$$G(s_t; \gamma, c) = 1 - \exp\{-\gamma(s_t - c)^2\} \tag{2.10}$$

with $\gamma > 0$. 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:

$$G(s_t; \gamma, c) = \frac{1}{1 + \exp(-\gamma(s_t - c))}. \tag{2.11}$$

The parameter γ controls the degree of nonlinearity. If γ is small, both transition functions switch between 0 and 1 very smoothly and slowly. If γ is large, both transition functions switch between 0 and 1 more quickly. As $\gamma \to \infty$, both transition functions become binary. However, for logistic function, the model reduces to SETAR model, while for exponential function, the model does not nest the SETAR model as a special case. The logistic function is monotonic and the LSTAR model switches between two regimes smoothly depending on how much the transition variable s_t is smaller than or greater than the threshold c. The exponential function is symmetrical and the ESTAR model switches between two regimes smoothly depending on how far the transition variable s_t is from the threshold c. 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 regimes, in which the dynamic behavior of series is markedly different, more complex dynamic patterns can be described.

The general form of the model is given by

$$y_t = \mu_{s_t} + \mathbf{X_t}\phi_{s_t} + \epsilon_t \tag{2.12}$$

where $\mathbf{X_t} = (y_{t-1}, y_{t-2}, ..., y_{t-p})$, ϕ_{s_t} is the p x1 vector of AR coefficients , ϵ_t follows $N(0, \sigma^2_{s_t})$ and s_t is an m-state Markov chain taking values $1, \ldots, m$, with transition matrix $\mathbf{P}$. 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}:

$$P\{s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots\} = P\{s_t = j | s_{t-1} = i\} = p_{ij} \tag{2.13}$$

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.

2.4 Testing for long memory

A popular research strategy to see if a time series exhibits long-range dependencies is to estimate the memory parameter by means of the GPH estimator and perform a t-test based on this estimator to prove the significance of the results. Therefore, we apply the GPH estimator to various nonlinear models of the before mentioned model classes in order to see if this research strategy might lead to misleading results. In addition to the GPH test we apply another popular long memory test, the V/S test, to the series in order to check the robustness of the results. All nonlinear models considered in our Monte Carlo study are stationary and short-range dependent in the sense that the central limit theorem still holds (for a more extensive discussion of this point see Davidson (2002)). The autoregressive order is chosen to be one. Each model is simulated with 1000 replications and different sample sizes of $T = 250$ and $T = 600$ 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(0,1).

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

$$y_t = \begin{cases} \phi_1 y_{t-1} + \epsilon_t & \text{if } y_{t-1} \leq 0 \\ \phi_2 y_{t-1} + \epsilon_t & \text{if } y_{t-1} > 0 \end{cases} \tag{2.14}$$

and we restrict our consideration on stationary nonlinear processes. We use $\phi_1 = -\phi_2$.

The following table presents the GPH estimator in order to see whether the GPH estimator is biased towards long memory[2]. In all tables, asterisk in $t - stat$ indicates that the value of d is significantly greater than zero with 5% significant level.

Table 2.1: GPH estimator for the SETAR process

	$T = 250$		$T = 600$	
$\phi_1 = -\phi_2$	d	$t - stat$	d	$t - stat$
0.1	-0.004	-2.546	-0.002	-1.566
0.2	0.010	5.387*	0.009	5.910*
0.3	0.034	15.257*	0.021	14.044*
0.4	0.066	32.371*	0.048	31.662*
0.5	0.113	51.318*	0.086	53.698*
0.6	0.167	73.903*	0.142	87.910*
0.7	0.256	108.909*	0.219	126.815*
0.8	0.375	148.920*	0.341	183.324*
0.9	0.536	203.808*	0.529	287.374*

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

[2] We use $J = T^{0.8}$ 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 $J = T^{0.5}$ as originally proposed by Geweke and 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 (2009).

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.

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 2.4 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 $d = 0.341$ 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.

In order to check the robustness of these results we also apply the V/S test for these DGPs. 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. Table 2.2 presents the rejection probabilities of the V/S test. All rejection probabilities are given to the 5% level.

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

Table 2.2: Rejection probabilities of V/S test for the SETAR process

	$T = 250$					$T = 600$				
$\phi_1 = -\phi_2$	q_1	q_2	q_3	q_4	q_5	q_1	q_2	q_3	q_4	q_5
0.1	0.050	0.045	0.034	0.012	0.040	0.054	0.050	0.047	0.031	0.037
0.2	0.061	0.047	0.031	0.010	0.049	0.069	0.055	0.049	0.038	0.040
0.3	0.084	0.040	0.034	0.005	0.041	0.082	0.053	0.050	0.043	0.044
0.4	0.112	0.050	0.036	0.008	0.041	0.107	0.055	0.049	0.034	0.052
0.5	0.148	0.049	0.040	0.011	0.055	0.172	0.066	0.060	0.032	0.063
0.6	0.230	0.065	0.048	0.006	0.047	0.260	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	0.130	0.019	0.019	0.969	0.321	0.177	0.053	0.084

Note: the critical value of V/S test are 0.2685, 0.1869 and 0.1518 for the significant level of 1%, 5% and 10% respectively.

considered to be the optimal q rejects the null hypothesis with a probability of around 5% and therefore gives reasonable values.

After considering SETAR models, we examine STAR models. As the results for LSTAR models are in line with our findings for ESTAR models, we only present the results for the latter. We use the transition variable $s_t = y_{t-1}$ and $c = 0$. The degree of non-linearity in the 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 $\gamma = 0.5$ and $\gamma = 5$. The model equation for the ESTAR series is given by:

$$y_t = \phi_1 y_{t-1} - (\phi_1 - \phi_2) y_{t-1} F(y_{t-1}, \gamma) + \epsilon_t. \tag{2.15}$$

Table 2.3 and 2.4 show that the GPH estimator is biased towards long memory either stationary or non-stationary depending on the parameter settings for the ESTAR

Table 2.3: GPH estimator for the ESTAR process ($\gamma = 0.5$)

	$T = 250$		$T = 600$	
$\phi_1 = -\phi_2$	d	$t - stat$	d	$t - stat$
0.1	0.000	1.334	0.021	0.113
0.2	0.012	7.737*	0.014	4.876*
0.3	0.034	15.871*	0.028	13.412*
0.4	0.049	23.445*	0.042	17.774*
0.5	0.080	37.723*	0.067	28.802*
0.6	0.116	49.399*	0.097	38.926*
0.7	0.179	70.451*	0.154	55.025*
0.8	0.261	91.278*	0.245	69.956*
0.9	0.415	123.204*	0.420	87.820*

Table 2.4: GPH estimator for the ESTAR process ($\gamma = 5$)

	$T = 250$		$T = 600$	
$\phi_1 = -\phi_2$	d	$t - stat$	d	$t - stat$
0.1	0.031	11.932*	0.024	14.552*
0.2	0.077	30.850*	0.064	36.535*
0.3	0.138	54.923*	0.109	62.995*
0.4	0.204	78.543*	0.171	98.303*
0.5	0.285	111.537*	0.239	140.235*
0.6	0.371	181.198*	0.324	182.790*
0.7	0.478	227.128*	0.429	242.483*
0.8	0.605	141.125*	0.561	309.304*
0.9	0.767	290.485*	0.734	404.768*

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 value of the autoregressive parameter, even for a relatively large sample size.[3]

Figure 2.5 in the appendix shows the ACF and periodogram of an ESTAR 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 ($d = 0.342$). All periodogram show a clear long memory behavior 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 behavior. 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.

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

Table 2.5: Rejection probabilities of V/S test for ESTAR ($\gamma = 0.5$)

	$T = 250$					$T = 600$				
$\phi_1 = -\phi_2$	q_1	q_2	q_3	q_4	q_5	q_1	q_2	q_3	q_4	q_5
0.1	0.055	0.044	0.039	0.004	0.044	0.056	0.054	0.051	0.043	0.050
0.2	0.071	0.040	0.042	0.009	0.051	0.080	0.058	0.042	0.040	0.058
0.3	0.086	0.042	0.038	0.009	0.061	0.089	0.057	0.042	0.035	0.056
0.4	0.115	0.053	0.036	0.006	0.051	0.114	0.060	0.041	0.035	0.053
0.5	0.150	0.051	0.037	0.009	0.044	0.151	0.070	0.047	0.038	0.064
0.6	0.198	0.059	0.037	0.009	0.044	0.225	0.076	0.055	0.042	0.055
0.7	0.321	0.069	0.042	0.011	0.072	0.331	0.081	0.066	0.049	0.068
0.8	0.496	0.122	0.054	0.005	0.056	0.572	0.138	0.059	0.042	0.063
0.9	0.788	0.289	0.114	0.007	0.066	0.895	0.344	0.163	0.058	0.092

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

Table 2.6: Rejection probabilities of V/S test for ESTAR ($\gamma = 5$)

	$T = 250$					$T = 600$				
$\phi_1 = -\phi_2$	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

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.

Finally, we investigate the behavior 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:

$$y_t = \begin{cases} \phi_1 y_{t-1} + \epsilon_t & \text{if } s_t = 1 \\ \phi_2 y_{t-1} + \epsilon_t & \text{if } s_t = 2 \end{cases} \tag{2.16}$$

with ϵ_t ~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 are taken from Hamilton (1989), which are $\mathbf{P} = (0.9, 0.1; 0.25, 0.75)$.

Table 2.7: GPH estimator for Markov switching processes

	$T = 250$		$T = 600$	
$\phi_1 = -\phi_2$	d	$t - stat$	d	$t - stat$
0.1	0.024	11.871*	0.015	10.281*
0.2	0.058	27.790*	0.044	28.647*
0.3	0.090	40.437*	0.082	52.273*
0.4	0.123	51.219*	0.127	79.053*
0.5	0.158	61.040*	0.183	112.041*
0.6	0.195	70.967*	0.240	144.956*
0.7	0.232	80.574*	0.313	179.855*
0.8	0.269	91.512*	0.391	214.038*
0.9	0.304	97.464*	0.477	249.671*

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 (2005) 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 transition probabilities 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 2.8 presents the results for the considered process.

Table 2.8: GPH estimator for the Markov switching process

	$T = 250$		$T = 600$	
$p_{11} = p_{22}$	d	$t - stat$	d	$t - stat$
0.1	-0.136	-61.264	-0.444	-223.591
0.2	-0.109	-45.272	-0.331	-156.937
0.3	-0.070	-27.990	-0.221	-102.034
0.4	-0.043	-16.533	-0.117	-52.100
0.5	-0.006	-2.408	-0.008	-3.733
0.6	0.034	11.958*	0.100	43.945*
0.7	0.076	26.163*	0.223	100.850*
0.8	0.120	39.935*	0.345	150.018*
0.9	0.173	58.814*	0.480	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 (2005) for details).

Similar to the other nonlinear models, periodogram which is generated from the Markov switching model does not show much difference than those of the true long memory process (see figure 2.6 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.

Table 2.9: Rejection probabilities of the V/S test for Markov switching processes

	$T = 250$					$T = 600$				
$\phi_1 = -\phi_2$	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 2.9 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 $T = 600$. This is in contrast to our findings before and very likely due to the mean shifting property of the Markov switching model.

Our results show that the GPH estimator fails to distinguish between long memory and standard short memory models used in the economic practice. It seems to be a useful research strategy not to simply rely on the GPH estimator but also to apply the V/S test to the data. The V/S test used with the optimal lag length q_5 proves to be quiet robust against nonlinear alternatives and seems therefore to be a suitable choice when the researcher is in doubt whether the long memory in his data might be spuriously caused by some nonlinearities.

2.5 Testing linearity

In this section we apply a general linearity test, namely the Neural Network test of Teräsvirta et al. (1993), 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 $T = 100, 500, 1000$ 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 misspecification of the model. It is also not robust against outliers. Figure 2.1 shows the rejection probabilities of this test when the true DGP is long memory. If the DGP is a pure long memory processes (Figure 2.1(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.

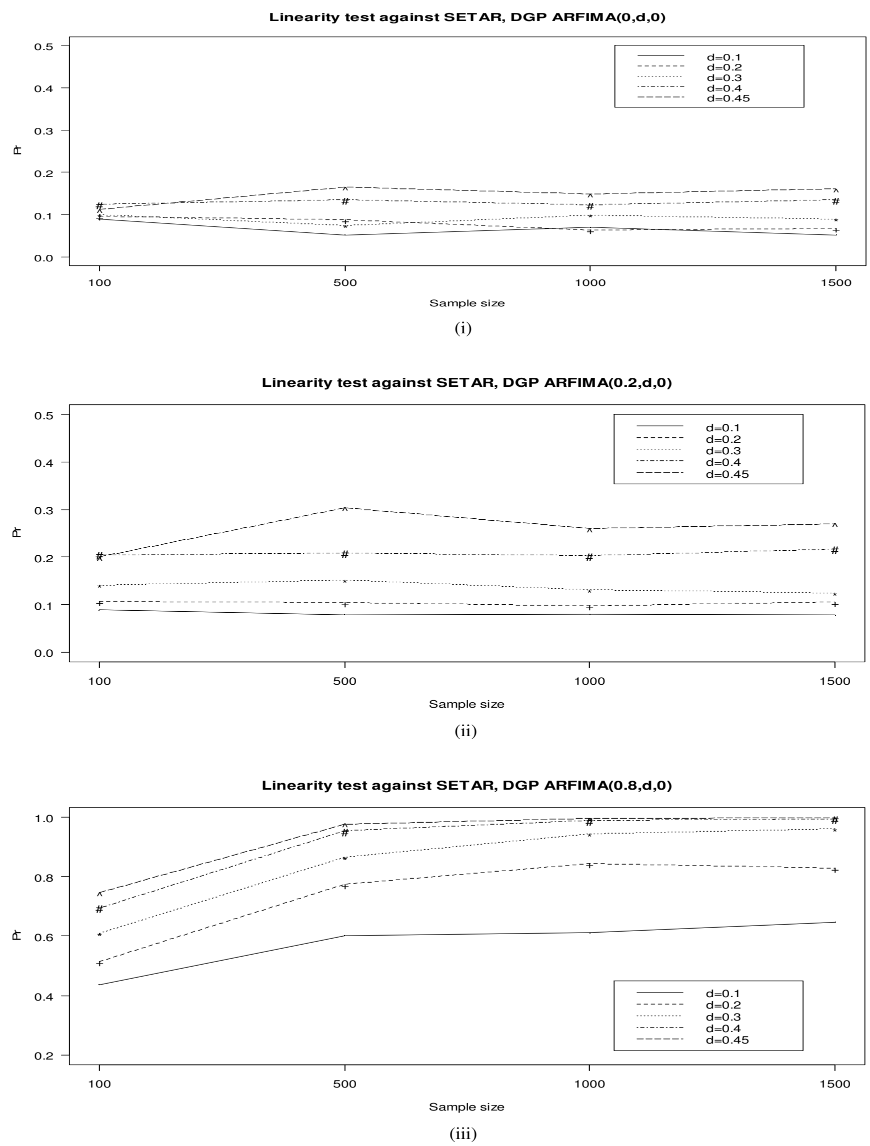


Figure 2.1: Rejection probabilities of linearity test against SETAR model (i) DGP is ARFIMA(0,d,0),(ii) DGP is ARFIMA(0.2,d,0) and (iii) DGP is ARFIMA(0.8,d,0)

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 2.2 below presents the results of the test.

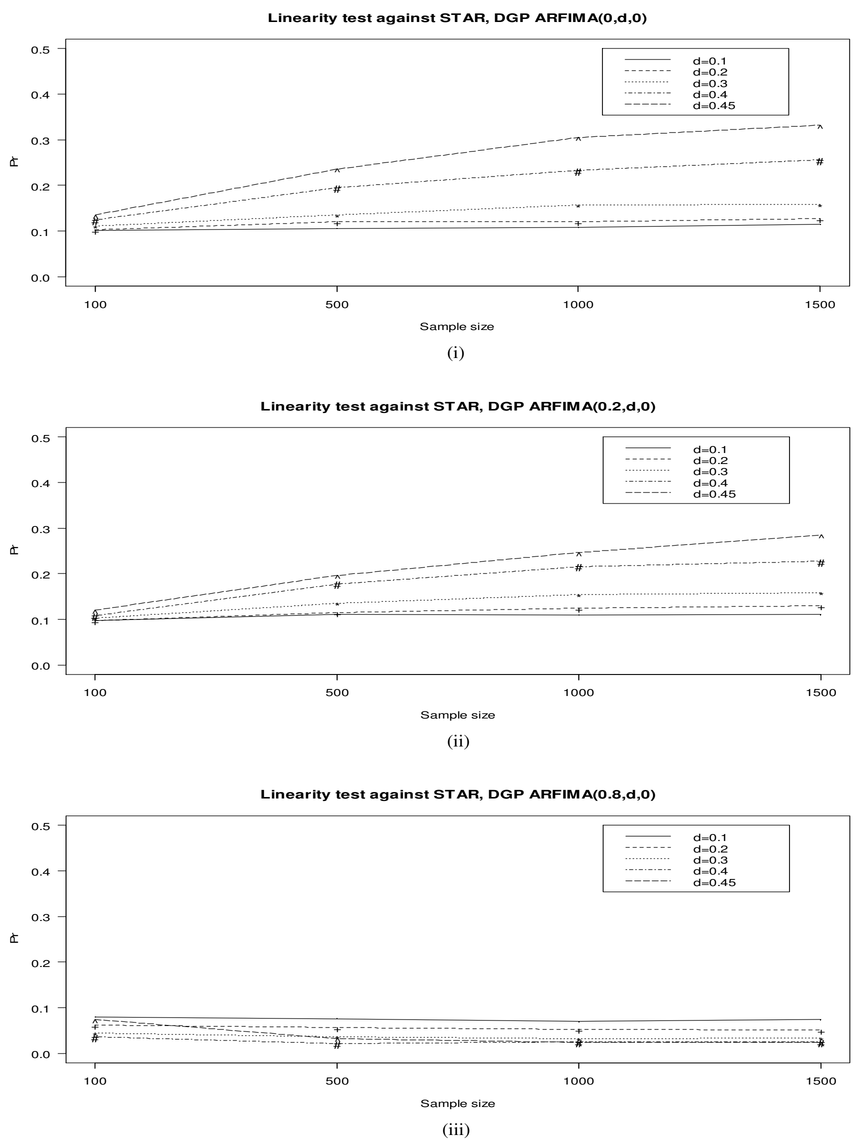
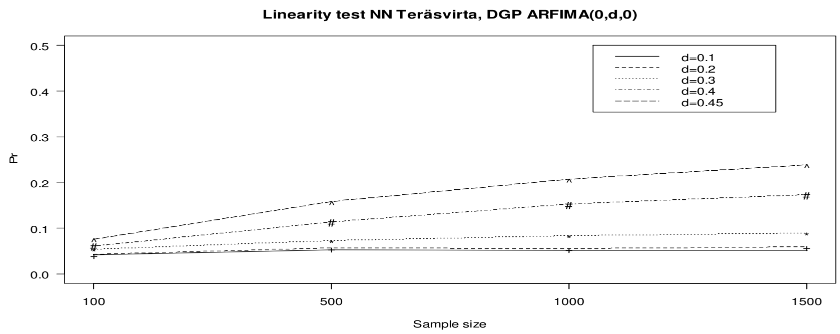


Figure 2.2: Rejection probabilities of linearity test against STAR model (i) DGP is ARFIMA(0,d,0),(ii) DGP is ARFIMA(0.2,d,0) and (iii) DGP is ARFIMA(0.8,d,0)

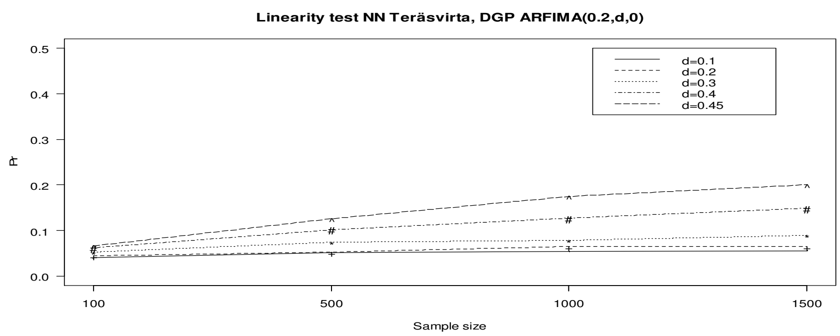
If the DGP is a pure long memory process, the results are similar to those of Andersson et al. (1999). The rejection probability increases with the value of the memory parameter and with the sample sizes. The same results are obtained for an ARFIMA $(\phi, d, 0)$ - process with a small autoregressive 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. (1993). 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.

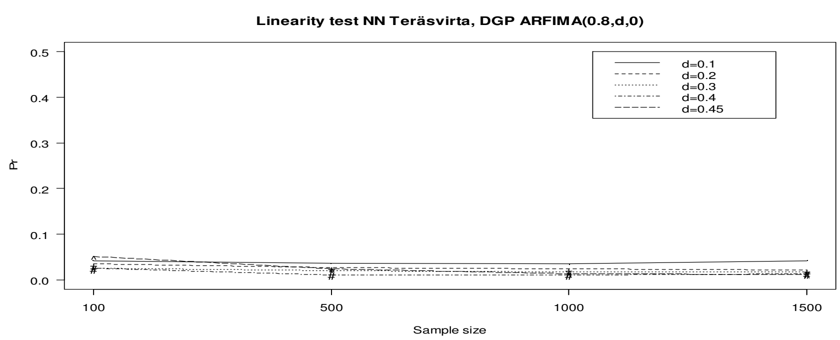
Let consider Figure 2.3 for the results of this test. The results are similar to those of the STAR test considered before. For a pure long memory DGP as well as for an ARFIMA $(\phi, d, 0)$ - process with a small autoregressive parameter ($\phi = 0.2$), the values of the rejection probability increase 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)

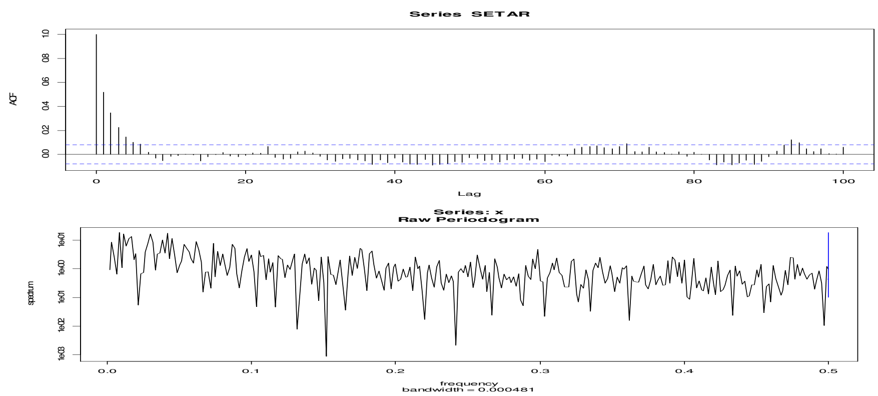
Figure 2.3: Rejection probabilities of linearity test against Neural Network model (i) DGP is ARFIMA(0,d,0),(ii) DGP is ARFIMA(0.2,d,0) and (iii) DGP is ARFIMA(0.8,d,0)

2.6 Conclusion

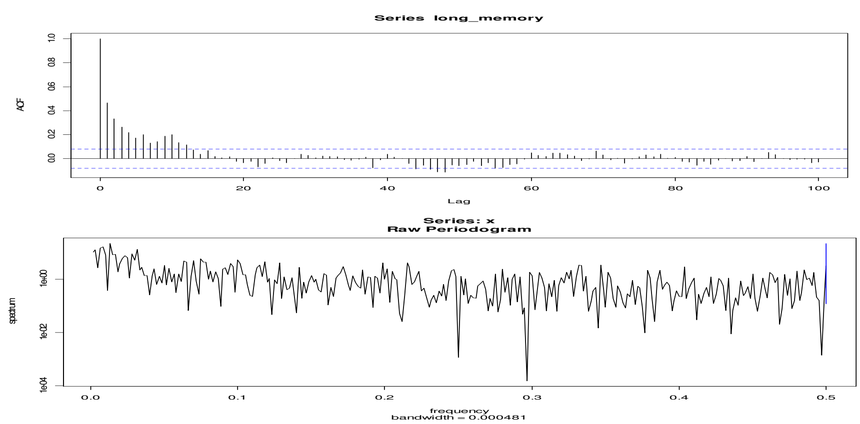
In this paper we show by Monte Carlo that popular nonlinear models such as TAR, STAR and Markov switching models can easily be misspecified as long memory. We estimate the memory parameter for various specifications of the above models and find that the GPH estimator is positively biased indicating long-range 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 behavior. 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.

Appendix

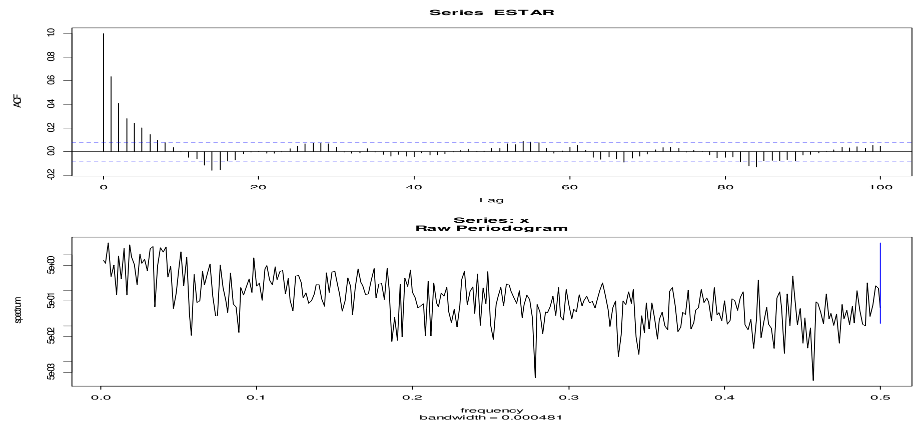


(i)

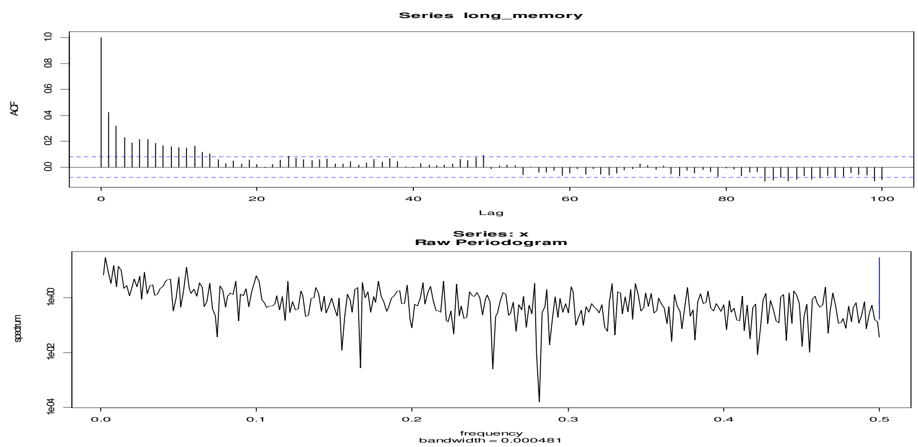


(ii)

Figure 2.4: Sample periodograms and ACF plots (i) SETAR process (ii) Long memory with $d = 0.341$

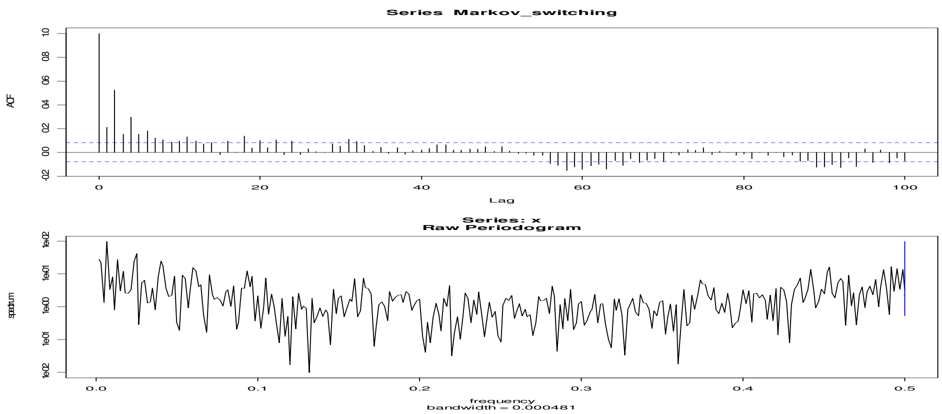


(i)

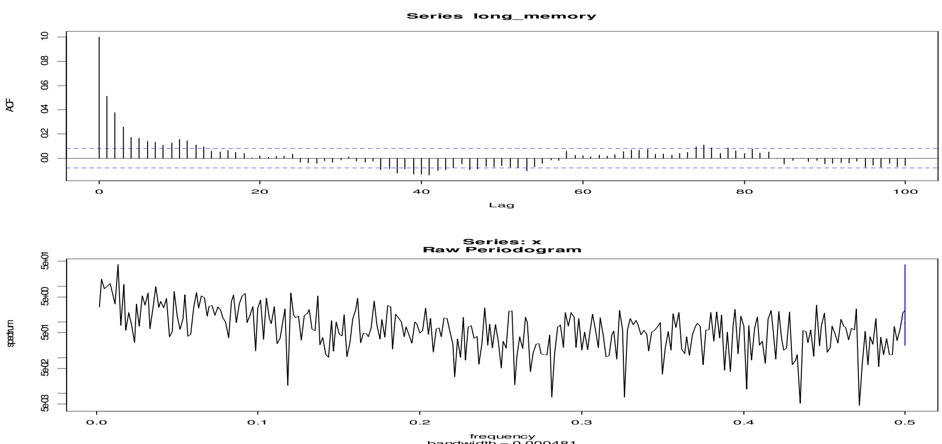


(ii)

Figure 2.5: Sample periodograms and ACF plots (i) ESTAR process (ii) Long memory with $d = 0.342$



(i)



(ii)

Figure 2.6: Sample periodograms and ACF plots (i) Markov switching process (ii) Long memory with $d = 0.391$

Chapter 3

A STUDY ON "SPURIOUS LONG MEMORY IN NONLINEAR TIME SERIES MODELS"[1]

3.1 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 are 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 (2004b) 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.

---

[1]This chapter is co-authored with Prof. Dr. Philipp Sibbertsen, Leibniz Universität Hannover Germany. It was originally published in *Applied Mathematical Science* (2008), 2(55), 2713-2734. Publication within this thesis with kind permission of Hikari. ltd.

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 special cases such as random level shift, AR(1) and random walk 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 renewal regime switching scheme. Parke (1999) introduces an error duration representation for fractional integration. Gourieoux and Jasiax (2001) study how processes with infrequent regime switching, which is binary process may generate a long memory effect in the autocorrelation function. Other 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 3.2 discusses some basic characteristics of long memory processes, section 3.3 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 3.4 and section 3.5 concludes.

3.2 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 y_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 y_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 y_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 y_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 y_t be a stationary process and denote by $\sigma_y(T)$ the variance of the partial sums $S_T = \sum_{t=1}^{T} y_t$. If the variance $\sigma_y(T)$ has the following asymptotic behavior*

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

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

In order to give a more severe understanding of the definitions above, figure 3.1 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.

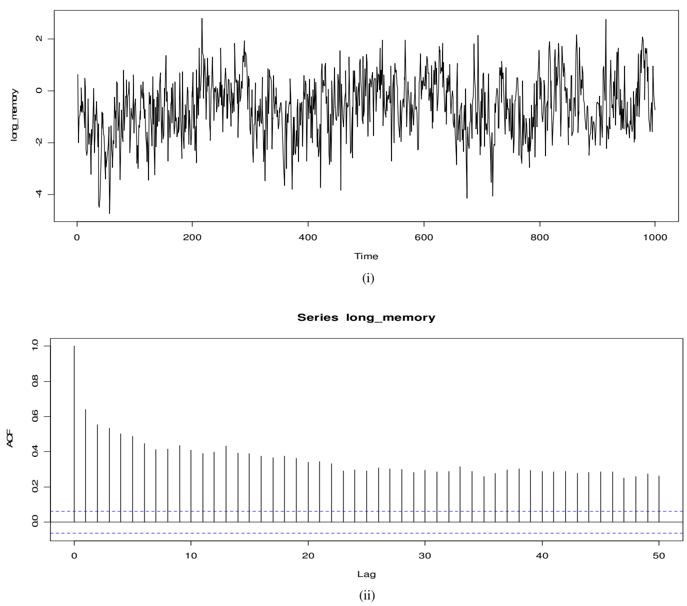


Figure 3.1: Long memory process with d=0.4. (i) time series plot (ii) autocorrelation function

3.3 *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:

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

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:

$$(1-B)^d = \sum_{j=0}^{\infty} \frac{d\Gamma(j+d)}{\Gamma(1+d)\Gamma(j+1)}, \tag{3.1}$$

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).

GPH estimator proposed by Geweke and Porter-Hudak (1983) is the most popular method to estimate the memory parameter. Let us consider the following J periodogram ordinates

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

where $\lambda_j = 2\pi j/T$, $j = 1, ..., J$ and J is the bandwidth frequency, which is a positive integer smaller than T.

The GPH method estimates the memory parameter by least square estimator from a linear regression model of $\{\log I_j\}$ on a constant and the regressor x_j defined as

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

In other words, the linear regression can be written as

$$\log I_j = \log C_f - 2dx_j + \log \xi_j \tag{3.2}$$

where x_j is 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. From this, the GPH estimator is given by

$$\hat{d}_p = \frac{-0.5 \sum_{j=1}^{J} (x_j - \bar{x}) \log I_j}{\sum_{j=1}^{J} (x_j - \bar{x})^2} \tag{3.3}$$

where $\bar{x}$ is the mean of the Fourier frequency x defined as $\bar{x} = \frac{1}{J} \sum_{j=1}^{J} x_j$.

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. The term in (3.3) can be arranged as follows:

$$d_p = \hat{d} + \frac{\sum_{j=1}^{J} (x_j - \bar{x}) \log \hat{I}_j / I_j}{\sum_{j=1}^{J} (x_j - \bar{x})^2} \tag{3.4}$$

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:

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

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 (2005) 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.

3.4 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 3.2 and 3.3 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.

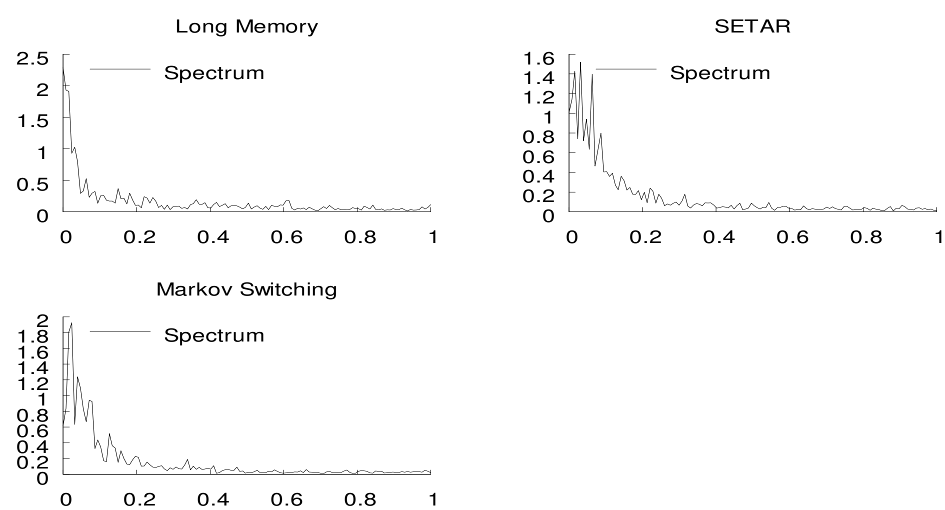


Figure 3.2: Plot spectrum of long memory, threshold and Markov switching process

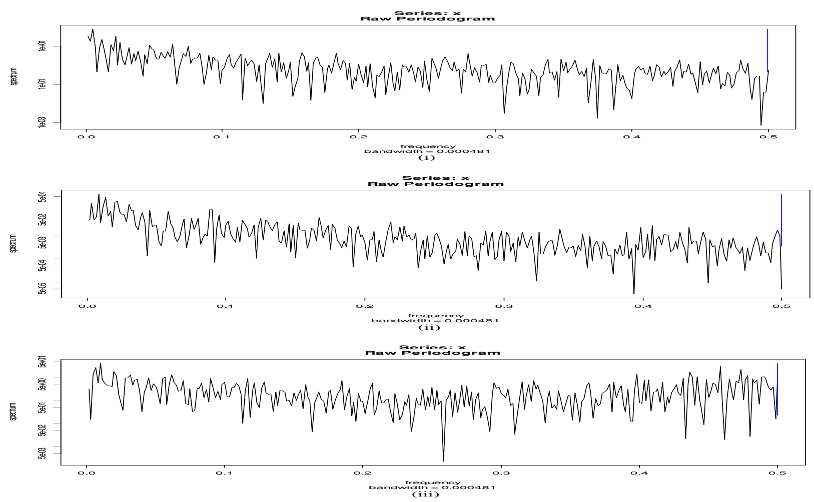


Figure 3.3: Plot periodogram of (i) long memory process (ii) Markov switching process (iii) Threshold process

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 nonlinear processes. Firstly, a simulation study applies the GPH estimator with the original bandwidth frequency proposed by GPH, which is $J = T^{0.5}$.

3.4.1 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)):

$$y_t = \begin{cases} \mu_1 + \phi_1 y_{t-1} + \sigma_1 \epsilon_t & \text{if } s_t = 1 \\ \mu_2 + \phi_2 y_{t-1} + \sigma_2 \epsilon_t & \text{if } s_t = 2 \end{cases} \tag{3.5}$$

The model above can be written as:

$$y_t = \mu_{s_t} + \phi_{s_t} y_{t-1} + \sigma_{s_t} \epsilon_t \tag{3.6}$$

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:

$$\mathbf{P} = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix} \tag{3.7}$$

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 y_0 is drawn from any stationary distribution, then y_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 1: *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 to identify long memory. Guégan and Rioublanc (2003) derived the autocovariance function for model (3.5). 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 (3.7).

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 y_t follows the theorem below:

Theorem 2: *Let y_t be the process defined in (3.5), by assuming that the assumption $(1) - (5)$ hold, then the autocovariance function $\gamma(\tau)$ of the process y_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 (3.5) 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 (3.5) 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 3.1 we generate a data set following model (3.5) 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 3.1: GPH estimator for Markov switching process (3.5) 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 indicates 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 3.1 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 3.2: GPH estimator for the Markov switching process (3.5) 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 3.2 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 (3.5) 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 (3.5) 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, $\forall_t$:

$$y_t = \begin{cases} \mu_1 + \epsilon_t & \text{if } s_t = 1 \\ \mu_2 + \epsilon_t & \text{if } s_t = 2 \end{cases} \tag{3.8}$$

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 (3.8) is called as mean switching model where y_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 (3.8) 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 (3.8) 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 1: *The autocorrelation function $\rho(\tau)$ of the process y_t defined by (3.8) is* *equal to*

$$\rho(\tau) = \frac{(\mu_1 - \mu_2)^2(1-p_{11})(1-p_{22})r^\tau}{(2-p_{11}-p_{22})^2[\pi_1\mu_1^2 + \pi_2\mu_2^2 + 1 - (\pi_1\mu_1 + \pi_2\mu_2)^2]} \quad (3.9)$$

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 (3.8) matches those of long memory processes under certain conditions. The following proposition holds:

Proposition 1: *Assume that (a)* $\mu_1 \neq \mu 2$ *and that (b)* $p_{11} = 1 - C_1T^{-\delta_1}$ *and* $p_{22} = 1 - C_2T^{-\delta_2}$, *with* $\delta_1, \delta_2 > 0$ *and* $0 < C_1, C_2 < 1$, *then the variances of the partial sums of* y_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 Inoue (2001) determine that the variance of the partial sums of y_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 3.3: GPH estimator for Markov switching process (3.8) 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 3.4: GPH estimator for Markov switching process (3.8) 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, the GPH estimator is biased towards long memory for $p_{11} = p_{22} = 0.8$ with a higher μ. 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 (2005) 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 3: *Consider the Markov switching process in (3.8), 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*

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

Proof: see Smith (2005)

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

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

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 3.5: 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 3.5 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 3.3 and 3.4, but using $\gamma = 0.8$.

Table 3.6: GPH estimator for Markov switching processes (3.8) 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 3.7: GPH estimator for Markov switching processes (3.8) 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 (3.3) with (3.6) and table (3.4) with (3.7) 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. Moreover, 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.

3.4.2 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:

$$y_t = F_1(y_{t-1}, \Phi)(1 - I(y_{t-l} > c) + F_2(y_{t-1}, \Phi)(I(y_{t-l} > c)) + \epsilon_t, \tag{3.10}$$

where the functions F_1 and F_2 are autoregressive processes depending on the past values of y_t and ϵ_t. The process ϵ_t is white noise and I an indicator function. The model (3.10) becomes a STAR model and the regime changes smoothly by setting the indicator function to a continuous function, $G(y_{t-l}, \gamma, c)$. If the function F_1 and F_2 are short memory, then the process in (3.10) 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:

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1} y_{t-1} + \epsilon_t & \text{if } y_{t-1} \leq c \\ \phi_{0,2} + \phi_{1,2} y_{t-1} + \epsilon_t & \text{if } y_{t-1} > c \end{cases} \tag{3.11}$$

The delay parameter in the model above is set to be one. Chan (1993) and van Dijk, et al. (2002) define the stationary conditions of (3.11) 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 chapter 2 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 3.8 as $\phi_{1,1} = 1$, $\phi_{1,2} = 0.1$, and c is set to be zero.

Table 3.8: 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 3.9: 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 3.10: 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 $J = T^{0.8}$. Using $J = T^{0.5}$ might give different result. However, we examine only $J = T^{0.8}$ due to reasons mentioned in the previous subsection.

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

$$y_t = \begin{cases} (1-B)^{-d}\epsilon_t^{(1)} & \text{if } y_{t-1} \le c \\ \epsilon_t^{(2)} & \text{if } y_{t-1} > c \end{cases} \tag{3.12}$$

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 (3.12) can be expressed as

$$\gamma(\tau) \sim \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\tau^{2d-1}, \ as\ \tau \to +\infty \tag{3.13}$$

which is not summable and the spectrum has the following representation

$$f(\lambda) \sim C\lambda^{-2d}+, \; as \; \lambda \to 0 \tag{3.14}$$

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.

3.5 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.

Chapter 4

TESTING FOR LONG MEMORY AGAINST ESTAR NONLINEARITIES[1]

4.1 Introduction

Long memory and nonlinear time series have both been extensively applied in empirical studies on the business cycle and other macroeconomic time series leading to different economic implications. However, several studies provided theoretical evidence that long memory can easily be confused with nonlinear regime-switching processes. Granger and Ding (1996) pointed out that a number of processes can be mistaken as long memory although providing only nonlinear features. Granger and Teräsvirta (1999) demonstrated that a simple nonlinear time series model can mimic linear properties whereas Andersson et al. (1999) found that on the other hand linear time series can mimic nonlinear properties as well. Diebold and Inoue (2001) proved analytically that stochastic regime switching can easily be confused with long memory. Davidson and Sibbertsen (2005) argued that the aggregation of processes with structural breaks converge to long memory.

Recently, several approaches to combine the two phenomena long memory and short memory nonlinearity appeared. For instance, van Dijk et al. (2000) and Smallwood (2005) developed the FI-STAR model, a joint model which covers long memory and nonlinear STAR processes. Tsay and Härdle (2008) propose a new Markov switching

[1] Co-author: Prof. Dr. Philipp Sibbertsen, Leibniz Universität Hannover Germany

model which allows the switching autoregressive component in the regimes to have a certain degree of fractional integration. Another Markov switching long memory model was examined by Heildrup and Nielsen (2006). Meanwhile, Goldman and Tsurumi (2006) developed the TARFIMA model, a simultaneous model which contains of threshold autoregression and long memory.

These aforementioned models are joint models combining long memory and a specific nonlinear process. Less attention has been addressed to the issue of distinguishing between long memory and nonlinearities. This is of interest when it is not clear to the practitioner whether a data set under investigation contains long memory or has a short memory nonlinear structure or whether both is present. Especially when the underlying nonlinear structure is unknown it is difficult to apply a specific joint model. Another major drawback is that long memory and nonlinear models are non-nested, which complicates the analysis. To the best of our knowledge, Kapetanios and Shin (2003) is the only approach which tries to solve the problem of distinguishing between the two phenomena, by assuming that the memory parameter is known. Unfortunately, the results of their simulation study indicated that the test does not have sophisticating power properties.

Baillie and Kapetanios (2007) provide a framework of simultaneously modeling long memory and nonlinearity. They, furthermore, suggest tests on neglected nonlinearity in the sense that they test whether a given long memory process has an additional nonlinear component. The problem is that a neglected nonlinearity component artificially creates a strongly biased estimate for the memory parameter and, therefore, falsely indicates long memory. Using this biased estimate in their testing framework decreases the power of their tests significantly. Therefore, a test barely considering the problem of distinguishing between these two phenomena is still of interest.

In this paper we suggest a test which is able to distinguish between long memory and a specific nonlinear time series process, namely ESTAR-processes. By using the basic idea of Kapetanios and Shin (2003), we propose a new test which is basically developed by using a standard Wald statistic and provide the adjusted critical values for our testing problem. The hypothesis is defined to be long memory under the null against ESTAR under the alternative. Since this involves a restricted parameter under the alternative, using a standard Wald test is inappropriate. Therefore, we suggest a directed-Wald statistic proposed by Andrews (1998) to overcome this problem. Furthermore, we consider two different approaches to develop the test statistic. The results indicate that the supremum statistic of the second approach is more powerful than the standard approach.

The paper is organized as follows. Section 4.2 introduces the theoretical framework. In section 4.3 the test statistics and their asymptotic distributions are derived. A simulation study showing the finite sample properties of our tests is given in section 4.4, and section 4.5 illustrates an empirical application to exchange rates. Section 4.6 concludes and all proofs are given in the appendix.

4.2 The model

In this section we introduce the considered processes, namely long memory and non-linear ESTAR processes. The model specification which is used to construct the test statistic refers to Kapetanios and Shin (2003) (denoted by KS hereafter). We propose two tests derived from two different approaches. The first one is similar to KS which uses a Taylor expansion to obtain an auxiliary regression on which the test is based. However, our regression differs from that in KS to some extents. For the second approach, we consider a model with an unidentified parameter in the transition function and take a supremum statistic.

4.2.1 Long memory process

Fractional integration (FI) models were first introduced by Granger and Joyeux (1980). Our work is based on the following simple model:

$$(1-L)^d u_t = \phi(L)^{-1}\epsilon_t = y_t, \tag{4.1}$$

where $t = 1, \ldots, T$, L is the lag operator, ϵ_t is an iid error term with variance σ^2 and finite fourth moments and y_t is a short memory process such as a stationary invertible ARMA (p,q) whose partial sums converge to a Brownian motion $Y(r)$ (see de Jong and Davidson, 2000). u_t is a long memory process with a certain degree of fractional integration d and the fractional difference operator is defined by

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{d\Gamma(j+d)}{\Gamma(1+d)\Gamma(j+1)}. \tag{4.2}$$

The value of d is $0 < d < 1/2$ for a stationary long memory process and $1/2 < d < 1$ for a non-stationary long memory process.

Model (4.1) can be written as an infinite moving average process in terms of y_t

$$u_t = \sum_{j=0}^{\infty} a_j y_{t-j}, \tag{4.3}$$

where $a_j = \frac{\Gamma(d+1)}{\Gamma(j+1)\Gamma(d-j+1)}(-1)^j$. Equivalently, it can be written as an infinite autoregressive process

$$u_t = \sum_{j=1}^{\infty} b_j u_{t-j} + y_t \tag{4.4}$$

with $b_j = -\frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}$. By defining

$$z_t = \sum_{j=1}^{t} b_j u_{t-j}, \tag{4.5}$$

we can write term (4.4) as

$$u_t = y_t + z_t \ \ \textbf{or} \ \ y_t = u_t - z_t. \tag{4.6}$$

Following Kapetanios and Shin (2003), we use (4.6) to derive the test statistic in the next section. Consider the scaled partial sum process

$$S_{[rT]} = \sum_{t=1}^{[rT]} u_t, \qquad r \in (0,1],$$

where u_t is defined by (4.1) with $d \neq 1/2$, we have that (see Marinucci and Robinson (1999))

$$c^{-1/2} T^{-(d+1/2)} S_{[rT]}(r) \Rightarrow Y_d(r),$$

where c is a constant such that $\mathrm{var}(S_T) \sim cT^{2(d+1/2)}$ and $Y_d(r)$ is a fractional Brownian motion with $d \in (0, 1/2)$ or $d \in (1/2, 3/2)$ respectively. "$\Rightarrow$" denotes weak convergence in distribution. A detailed discussion regarding the fractional Brownian motion can be found in Mandelbrot and Van Ness (1968). Beran (1994) gives an overview over the concept of long memory.

4.2.2 ESTAR model

Exponential Smooth Transition Autoregressive (ESTAR) models were introduced by Granger and Teräsvirta (1993). A survey of recent developments in ESTAR modeling can be found in van Dijk et al. (2002). A simple ESTAR model can be written as:

$$y_t = \alpha_1 y_{t-1}\{1 - \exp(-\gamma y_{t-l}^2)\} + \alpha_2 y_{t-1} + \epsilon_t, \tag{4.7}$$

where y_t is a stationary process and α_1, α_2 and γ are unknown parameters. The parameter γ controls the degree of nonlinearity and determines the speed of transition between the two extreme regimes, and y_{t-l} in the transition function is the transition variable with lag $l \geq 1$. As frequently applied in the literature, we set the delay parameter l equal to 1, therefore $y_{t-l} = y_{t-1}$[2].

[2]Taylor, Peel and Sarno (2001) provide an overview about the motivation to choose "l" equal to one related to empirical applications.

4.3 Testing long memory against ESTAR

As we pointed out in the previous section, we apply two different approaches to develop the test statistic. The first approach applies a first order Taylor expansion to the transition function of ESTAR model in order to obtain an auxiliary regression. This approach is standard when considering tests for ESTAR processes. The second approach allows the parameter γ in the transition function to be unidentified by applying a supremum statistic. By using this approach, we expect that the test has a higher power, since we do not use linear approximations for non-linear processes.

Let us write the general model specification as:

$$u_t = \alpha_1 F(y_{t-1}) + \alpha_2 z_t + \epsilon_t, \tag{4.8}$$

$t = 1, ..., T$ with u_t and z_t are defined as in the previous section. The error ϵ_t is allowed to be a general stationary process such as a stationary strong mixing process. The following assumption formalize the condition for ϵ_t.

Assumption 1 :
Let ϵ_t be a stationary strong mixing sequence with $\mathbf{E}\epsilon_t^2 = \sigma^2$ for all t and $\sup_t \|\epsilon_t\|_4 < \infty$.

This assumption is necessary to derive the limit distribution of our test in the next section and to have a consistent estimator of the error variance, that is $\sigma_T^2 \rightarrow_p \sigma^2$ with $\sigma_T^2 = (1/T)\sum_{t=1}^{T} \epsilon_t^2$. The test statistic is derived from (4.8) by using two different approaches which depend on $F(y_{t-1})$. In other words, we define the model as follows. For the first approach, applying a first order Taylor expansion to the transition function $F(y_{t-1})$ yields

$$u_t = \alpha_1 y_{t-1}^3 + \alpha_2 z_t + \epsilon_t. \tag{4.9}$$

In this case, $F(y_{t-1}) = y_{t-1}^3$. For the second approach, $F(y_{t-1})$ is originally defined as $y_{t-1}\{1 - \exp(-\gamma y_{t-1}^2)\}$, which yields

$$u_t = \alpha_1 y_{t-1}\{1 - \exp(-\gamma y_{t-1}^2)\} + \alpha_2 z_t + \epsilon_t. \tag{4.10}$$

We test the null of long memory against the alternative of an ESTAR process using the models (4.9) and (4.10). Our null hypothesis is:

$$\mathbf{H_0} \ : \ \alpha_1 = \alpha_2 = 0 \tag{4.11}$$

and is tested against the alternative of:

$$\mathbf{H_1} \ : \ \alpha_1 \neq 0, \alpha_2 = 1 \tag{4.12}$$

Under the null, we can also write

$$u_t = \epsilon_t, \tag{4.13}$$

which is a simple long memory model and under the alternative hypothesis we have the ESTAR model

$$y_t = \alpha_1 F(y_{t-1}) + \epsilon_t \tag{4.14}$$

with the corresponding function $F(y_{t-1})$. We discuss the test statistic and its limit distribution in the following subsections.

4.3.1 Test statistic and limit distribution

We propose a standard Wald test to test the null (4.11). We know that model (4.10) with a given γ is linear in the parameter $\alpha = (\alpha_1, \alpha_2)'$ and so is the model (4.9). Therefore, we can estimate the parameter α by OLS and obtain the least square estimator

$$\hat{\alpha} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}(\mathbf{X}'\mathbf{U}) \tag{4.15}$$

with $\mathbf{U} = [u_1, u_2, ..., u_t]'$, $\mathbf{X} = \begin{bmatrix} F(y_0) & z_1 \\ \vdots & \\ F(y_{t-1}) & z_t \end{bmatrix}$. The Wald statistic for the null of $\alpha = \mathbf{0}$ is

$$W = \hat{\alpha}'[Var(\hat{\alpha})]^{-1}\hat{\alpha}. \quad (4.16)$$

The Wald test above is normally used under the condition that the parameter α is unrestricted under the alternative hypothesis. Since the parameter α_2 in (4.12) is restricted ($\alpha_2 = 1$), the classical Wald test is no longer an optimal test. To overcome this problem, we use a directed-Wald statistic proposed by Andrews (1998). This test is designed for testing hypotheses with one or more restricted parameters under the alternative.

Let us define that $\mathbf{H_0} : \alpha_2 = 0$ and $\mathbf{H_1} : \alpha_2 = A$, with $A = 1$. Then, the directed-Wald statistic, DW is given by:

$$DW_{(c)} = (1+c)^{(-1/2)} exp\left(\frac{1}{2}\frac{c}{1+c}W\right) \mathbf{\Phi}(A, \frac{c}{1+c}\hat{\alpha_2}, \frac{c}{1+c}Var(\hat{\alpha_2})) \quad (4.17)$$

with c being a scalar relative weight given to alternatives that are close to the null against alternatives that are away from the null. Andrews (1998) provides a procedure for choosing the value of c and presents a simulation study for several values of c. This suggests that the power of the directed-Wald test does not vary much with c, for $c \neq 0$, which implies that the choice of c is not crucial. Since Andrews (1998) found that $c = \infty$ is optimal for all cases, we set $c = \infty$ for our directed-Wald test and therefore, (4.17) reduces to:

$$DW_{(\infty)} = W + 2\log[\mathbf{\Phi}(A, \hat{\alpha_2}, Var(\hat{\alpha_2}))]. \quad (4.18)$$

Here, $\mathbf{\Phi}(.)$ represents a normal probability density function defined as $\mathbf{\Phi}(A, \mu, \sigma^2) =$

$P(V \in A)$, where $V \sim N(\mu, \sigma^2)$. For notational simplicity, we suppress the subscript ∞.

Let us now discuss the asymptotic distribution of both approaches. The asymptotic distribution is derived from the continuous mapping theorem of Kurtz and Protter (1991), the functional central limit theorem and weak convergence to stochastic integrals of Davidson and de Jong (2000) and theorem 30.13 of Davidson (1994).

4.3.2 First approach

In this section, we derive the limit distribution of the statistics (4.16) and (4.18), which is mainly characterized by the approximation y_{t-1}^3 of $F(y_{t-1})$. By using assumption 1, we obtain theorem 1 below (all proofs are given in the appendix):

Theorem 1: *Under the null hypothesis that u_t is long memory, $\hat{\alpha}_{\mathbf{OLS}}$ is a consistent estimator of $\alpha = 0$ and converges to its true value with the rate of convergence $diag(T^{3/2}, T^{(1/2+d)})$ when $0 < d < 0.5$ and $diag(T^{3/2}, T^{(1-d)})$ when $0.5 < d < 1$. Its asymptotic distribution is*

$$\begin{bmatrix} T^{3/2} & 0 \\ 0 & T^{(1/2+d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} \Rightarrow \mathbf{Q_1^{-1} Q_2} \quad \textbf{if } 0 \leq d < 0.5 \tag{4.19}$$

$$\begin{bmatrix} T^{3/2} & 0 \\ 0 & T^{(1-d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} \Rightarrow \mathbf{Q_1^{-1} Q_2} \quad \textbf{if } 0.5 < d < 1 \tag{4.20}$$

with $\mathbf{Q_1}$ and $\mathbf{Q_2}$ defined as

$$\mathbf{Q_1} = \begin{bmatrix} \int_0^1 Y(r)^6 dr & \int_0^1 Y(r)^3 dZ_d(r) \\ \int_0^1 Y(r)^3 dZ_d(r) & \int_0^1 Z_d(r)^2 dr \end{bmatrix} \tag{4.21}$$

$$\mathbf{Q_2} = \begin{bmatrix} \int_0^1 Y(r)^3 dY^*(r) \\ \int_0^1 Z_d(r) dY^*(r) \end{bmatrix}, \tag{4.22}$$

where $Y(r)$ and $Y^*(r)$ are standard Brownian motions and $Z_d(r)$ is a function of the fractional Brownian motion as defined in the appendix.

Theorem 1 shows that the OLS estimator has a nonstandard limit distribution which depends on the value of d. The convergence rate of the estimator differs between stationary and non-stationary long memory. The asymptotic distribution of the Wald and directed-Wald statistic follows directly from theorem 1:

Theorem 2: *Under the null that u_t is long memory, the limit distribution of the Wald statistic is*

$$W \Rightarrow \mathcal{W} \equiv \mathbf{Q_2'Q_1^{-1}Q_2} \tag{4.23}$$

and the limit distribution of the directed-Wald statistic with $\alpha_2 = A$ under the alternative is given by

$$DW \Rightarrow [\mathcal{W} + 2\log \mathbf{\Phi}(A, \mathcal{B}, \mathcal{V})] \tag{4.24}$$

with $\mathcal{B}$ and $\mathcal{V}$ as defined in the appendix. Under the alternative, the statistic diverges implying the consistency of the test.

4.3.3 Second approach

Similar to the procedures applied for the first approach, we need to define $F(y_{t-1})$ as $F(y_{t-1}, \gamma) = y_{t-1}\{1 - \exp(-\gamma y_{t-1}^2)\}$. The test statistic is nonstandard since γ is unidentified under the null. To overcome this problem, Davies (1987) proposed a supremum statistic, which maximizes the test with respect to the nuisance parameter. Andrews and Ploberger (1994) showed that the supremum test is optimal. Other tests using sup-Wald statistics can be found among others in White (1982) and Carrasco (2002).

For γ in the range of Γ, our directed-Wald test can be written as

$$DW = \sup_{\gamma \in \Gamma} DW_{\gamma} = \sup_{\gamma \in \Gamma} \left\{ \frac{1}{\hat{\sigma_{\epsilon}}^2} \left\{ \hat{\alpha}'(X'X)\hat{\alpha} \right\} + 2\log[\mathbf{\Phi}(A, \hat{\alpha_2}, Var(\hat{\alpha_2}))] \right\}, \quad (4.25)$$

where $\hat{\sigma_{\epsilon}}^2$ is the error variance of the OLS estimator and $\gamma = [\underline{\gamma}, \overline{\gamma}] \in R^+$ is such that $0 < \underline{\gamma} < \gamma < \overline{\gamma}$.

To derive the limit distribution of the test statistic for model (4.10), we need an additional assumption:

Assumption 2: *Suppose that $F(y_{t-1}, \gamma)$ is continuously differentiable with respect to γ and $\sup_t \| \sup_{\gamma \in \Gamma} |F'(y_{t-1}, \gamma)| \|_2 < \infty$ where $F'(y_{t-1}, \gamma) = \frac{\partial F(y_{t-1}, \gamma)}{\partial \gamma}$.*

Assumption 2 is necessary to assure stochastic equicontinuity implying weak convergence. More details about this assumption can be found in Park and Shintani (2005). They discuss further conditions for the transition function, including differentiability with respect to γ. By using assumption 1 and 2, we obtain the following theorem.

Theorem 3: *Under the null that u_t is long memory, $\hat{\alpha}_{\mathbf{OLS}}$ is a consistent estimator of $\alpha = 0$ and converges to its true value with the rate diag$(T^{1/2}, T^{(1/2+d)})$ when $0 < d < 0.5$ and diag$(T^{1/2}, T^{(1-d)})$ when $0.5 < d < 1$. Its limit distribution is*

$$\begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{(1/2+d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \end{bmatrix} \Rightarrow \mathbf{Q_1}(\gamma)^{\mathbf{-1}}\mathbf{Q_2}(\gamma) \quad \textbf{if } 0 \leq d < 0.5 \qquad (4.26)$$

$$\begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{(1-d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \end{bmatrix} \Rightarrow \mathbf{Q_1}(\gamma)^{\mathbf{-1}}\mathbf{Q_2}(\gamma) \quad \textbf{if } 0.5 < d < 1 \qquad (4.27)$$

with $\mathbf{Q_1}(\gamma)$ *and* $\mathbf{Q_2}(\gamma)$ *are*

$$\mathbf{Q_1}(\gamma) = \begin{bmatrix} (1-2\mu_\gamma+\psi_\gamma)\int_0^1 Y(r)^2 dr & (1-\mu_\gamma)\int_0^1 Z_d(r)dY(r) \\ (1-\mu_\gamma)\int_0^1 Z_d(r)dY(r) & \int_0^1 Z_d(r)^2 dr \end{bmatrix} \tag{4.28}$$

$$\mathbf{Q_2}(\gamma) = \begin{bmatrix} (1-\mu_\gamma)\int_0^1 Y(r)dY^*(r) \\ \int_0^1 Z_d(r)dY^*(r) \end{bmatrix}, \tag{4.29}$$

where μ_γ *and* ψ_γ *are defined as* $\mu_\gamma = \mathbf{E}\{\exp(-\gamma y_{t-1}^2)\}$ *and* $\psi_\gamma = \mathbf{E}\{\exp(-2\gamma y_{t-1}^2)\}$ *respectively,* $Y(r)$ *and* $Y^*(r)$ *are standard Brownian motions and* $Z_d(r)$ *is a function of the fractional Brownian motion as defined in the appendix.*

Theorem 3 differs from theorem 1 regarding to $\mathbf{Q_1}$ and $\mathbf{Q_2}$ which depend on γ. The limit distribution of the Wald and directed-Wald statistic are given in theorem 4.

Theorem 4: *Under the null that* u_t *is long memory, we have for the Wald statistic*

$$W_\gamma \Rightarrow \mathcal{W}_\gamma \equiv \mathbf{Q_2}(\gamma)'\mathbf{Q_1}(\gamma)^{\mathbf{-1}}\mathbf{Q_2}(\gamma) \tag{4.30}$$

and for the sup-directed-Wald statsitic for $\alpha_2 = A$ *we have*

$$\sup_{\gamma\in\Gamma} DW_\gamma \Rightarrow \sup_{\gamma\in\Gamma} [\mathcal{W}_\gamma + 2\log \mathbf{\Phi}(A, \mathcal{B}, \mathcal{V})] \tag{4.31}$$

with $\mathcal{B}$ *and* $\mathcal{V}$ *as defined in the appendix. Under the alternative, the statistic diverges implying the consistency of the test.*

The pointwise convergence derived above is not sufficient for establishing uniform stochastic convergence of the limit distribution of the sup-Wald test. Therefore, we need to prove stochastic equicontinuity, a condition that $\forall\epsilon$, there exists a $\delta > 0$ such that

$$\lim\sup_{T\to\infty} P_r\left[\sup_{\gamma\in\Gamma}\sup_{\gamma':|\gamma-\gamma'|<\delta} |W_\gamma^{(i)} - W_{\gamma'}^{(i)}| \geq \epsilon\right] < \epsilon. \tag{4.32}$$

The proof of the stochastic equicontinuity condition is obtained under the assumption that $\gamma \in \Gamma$.

Theorem 5 *Under assumption 2, the test statistic* $\sup W_\gamma$ *is stochastically equicontinuous over* Γ*.*

Since the directed-Wald test contains the Wald test as a special case for a certain weight defined by the second term in (4.31), theorem 5 implies the stochastic equicontinuity of the directed-Wald test.

4.4 Monte Carlo

In this section, we carry out a Monte Carlo simulation to study the size and power properties of the test in finite sample sizes. We showed that the optimal test has a nonstandard distribution, therefore the critical values have to be simulated, which is done by generating long-memory series of length 5000 to which the test is applied. The number of replications is 10000.

Particularly for the second approach, the value of γ is set to be in the interval $\gamma \in (0.01, 2.5)$. The supremum is obtained by a grid search with steps of 0.01. A large γ leads to a flat transition function.

The critical values of the tests are given in table 4.1 and 4.2.

Table 4.1: Critical values of the test for the first approach

Sign. Level	d=0.1	d=0.2	d=0.3	d=0.4
90%	2.7142	2.7613	2.8309	2.8814
95%	3.6190	3.7288	3.9006	4.0122
99%	6.2231	6.4725	6.6618	6.7120
Sign. Level	d=0.6	d=0.7	d=0.8	d=0.9
90%	2.9104	3.0629	3.3821	4.0680
95%	4.2318	4.7160	5.1209	5.5649
99%	6.7710	7.1023	8.0160	9.5548

Table 4.2: Critical values of the test for the second approach

Sign. Level	d=0.1	d=0.2	d=0.3	d=0.4
90%	3.5076	3.5720	3.6016	3.6213
95%	4.7613	4.8242	4.8506	4.9001
99%	7.5855	8.0604	8.0640	8.1321
Sign. Level	d=0.6	d=0.7	d=0.8	d=0.9
90%	3.7478	3.9201	4.1683	4.3124
95%	4.9909	5.2931	5.4705	5.9658
99%	7.8707	8.8037	9.1337	11.3395

We first study the size of the test when the data generating process has stationary long memory with $d = 0.1, 0.2, 0.3, 0.4$ and non-stationary long memory with $d = 0.6, 0.7, 0.8, 0.9$. For each experiment, we do 1000 replications and compute the rejection probability for the 5% and the 10% significance level. The sample sizes are 100 and 250. Tables 4.3 and 4.4 contain the results of the size experiment.

Table 4.3: Size of the directed-Wald test for the first approach

Sign. level	T	d=0.1	d=0.2	d=0.3	d=0.4
5%	100	0.092	0.085	0.058	0.067
	250	0.065	0.072	0.065	0.061
10%	100	0.148	0.157	0.111	0.120
	250	0.120	0.126	0.140	0.123
Sign. level	T	d=0.6	d=0.7	d=0.8	d=0.9
5%	100	0.070	0.050	0.046	0.041
	250	0.063	0.046	0.048	0.037
10%	100	0.115	0.105	0.088	0.094
	250	0.116	0.107	0.093	0.086

Note: the data generating process is long memory as in (1)

Table 4.4: Size for the directed-Wald test for the second approach

Sign. level	T	d=0.1	d=0.2	d=0.3	d=0.4
5%	100	0.071	0.083	0.077	0.085
	250	0.069	0.071	0.057	0.069
10%	100	0.131	0.148	0.136	0.136
	250	0.128	0.119	0.120	0.121
Sign. level	T	d=0.6	d=0.7	d=0.8	d=0.9
5%	100	0.055	0.043	0.042	0.035
	250	0.055	0.048	0.050	0.055
10%	100	0.098	0.100	0.090	0.091
	250	0.105	0.094	0.104	0.121

Note: the data generating process is long memory as in (1)

The rejection rates in table 4.3 and 4.4 do not differ significantly and have a similar tendency. For a higher sample size ($T = 250$), the rejection rate converges to the nominal size. Stationary long memory tends to over-rejection, whereas non-stationary long memory tends to under-rejection. Nevertheless, the values are very close to the nominal size, which indicates that the tests are correctly sized in general, although there are little size distortions for the smaller sample size.

In the second experiment, we study the power of the test. The data generating process is an ESTAR process

$$y_t = \alpha_1 y_{t-1}\{1 - \exp(-\gamma y_{t-1}^2)\} + \alpha_2 y_{t-1} + \epsilon_t \tag{4.33}$$

with $\epsilon_t \sim N(0, 1)$. The error term follows a standard normal distribution. We discard the first 100 observations to minimize the effect of initial values. We set the parameter $\alpha_2 = 1$ and $\alpha_1 \in \{-1.5, -1, -0.5, -0.1\}$ with various γ and $\gamma \in \{0.01, 0.05, 0.1\}$ [3].

Before we do the power experiment, we have to check whether the aforementioned parameter settings for the ESTAR process (4.33) are of interest in the sense that they can be mistaken as long memory. Therefore, the memory parameter is estimated by means of any consistent estimators. In this paper, we use the GPH estimator proposed by Geweke and Porter-Hudak (1983). The following table provides the estimated long memory parameter for our DGPs.

---

[3] These parameter setting was extensively examined for example in Rothe and Sibbertsen (2006), Kapetanios et al. (2003). Imposing $\alpha_2 = 1$ leads to a globally stationary ESTAR process, which has a unit root process in one regime.

Table 4.5: Mean and confidence interval of the estimated $\hat{d}$ by GPH

parameters		$T = 100$	$T = 250$
α_1	γ	mean($\hat{d}$)	mean($\hat{d}$)
-1.5	0.01	0.6586 [0.6022;0.7150]	0.4448 [0.4197;0.4699]
	0.05	0.2977 [0.2289;0.3566]	0.1481 [0.1201;0.1760]
	0.1	0.1398 [0.0807;0.1990]	0.0684 [0.0427;0.0941]
-1	0.01	0.7069 [0.6518;0.7619]	0.5135 [0.4879;0.5392]
	0.05	0.4196 [0.3664;0.4279]	0.2153 [0.1880;0.2426]
	0.1	0.2650 [0.2078;0.3221]	0.1163 [0.0896;0.1431]
-0.5	0.01	0.8226 [0.7678;0.8775]	0.6332 [0.6089;0.6576]
	0.05	0.5943 [0.5403;0.6483]	0.3637 [0.3374;0.3900]
	0.1	0.4494 [0.3925;0.5063]	0.2569 [0.2295;0.2842]
-0.1	0.01	0.9280 [0.8789;0.9851]	0.8708 [0.8452;0.8964]
	0.05	0.8543 [0.7974;0.9112]	0.7113 [0.6842;0.7384]
	0.1	0.8018 [0.7511;0.8652]	0.6352 [0.6090;0.6615]

Table 4.5 shows the mean of the estimated fractional integration order as well as the 95% confidence interval. All values are obtained by 1000 replications. We see from the table that under those parameter settings ESTAR processes generate spurious long memory, where the order of fractional integration lies either in the stationary or in the non-stationary region.

The following table presents the power results for the directed-Wald test. As we pointed out in the previous section the memory parameter is assumed to be known or given.

Table 4.6: Power experiment for the Wald test for the first approach

parameters		$d = 0.1$		$d = 0.2$		$d = 0.3$		$d = 0.4$	
α_1	γ	$T = 100$	$T = 250$	$T = 100$	$T = 250$	$T = 100$	$T = 250$	$T = 100$	$T = 250$
-1.5	0.01	1.000	1.000	0.999	1.000	0.989	1.000	0.949	1.000
	0.05	0.903	0.994	0.749	0.975	0.573	0.828	0.324	0.573
	0.1	0.384	0.728	0.230	0.380	0.119	0.169	0.066	0.087
-1	0.01	1.000	1.000	0.999	1.000	0.997	1.000	0.986	0.923
	0.05	0.999	1.000	0.956	1.000	0.815	0.986	0.629	0.963
	0.1	0.877	0.999	0.732	0.961	0.440	0.781	0.245	0.474
-0.5	0.01	1.000	1.000	1.000	1.000	1.000	1.000	0.995	1.000
	0.05	1.000	1.000	1.000	1.000	0.991	1.000	0.927	0.999
	0.1	0.998	1.000	0.991	1.000	0.945	0.999	0.790	0.978
-0.1	0.01	1.000	1.000	1.000	1.000	1.000	1.000	0.997	1.000
	0.05	1.000	1.000	1.000	1.000	1.000	1.000	0.994	1.000
	0.1	1.000	1.000	1.000	1.000	0.998	1.000	0.993	1.000

Table 4.7: Power experiment for the Wald test for the first approach

parameters		$d = 0.6$		$d = 0.7$		$d = 0.8$		$d = 0.9$	
α_1	γ	$T = 100$	$T = 250$	$T = 100$	$T = 250$	$T = 100$	$T = 250$	$T = 100$	$T = 250$
-1.5	0.01	0.557	0.881	0.224	0.493	0.064	0.095	0.022	0.030
	0.05	0.073	0.102	0.072	0.067	0.062	0.128	0.099	0.220
	0.1	0.088	0.139	0.124	0.255	0.148	0.327	0.168	0.411
-1	0.01	0.657	0.956	0.344	0.693	0.087	0.175	0.017	0.022
	0.05	0.162	0.313	0.058	0.092	0.037	0.048	0.042	0.142
	0.1	0.047	0.071	0.056	0.073	0.092	0.156	0.111	0.247
-0.5	0.01	0.839	0.990	0.479	0.856	0.151	0.325	0.014	0.046
	0.05	0.478	0.801	0.185	0.378	0.045	0.054	0.017	0.027
	0.1	0.253	0.564	0.076	0.161	0.036	0.031	0.035	0.062
-0.1	0.01	0.931	1.000	0.719	0.974	0.329	0.668	0.056	0.115
	0.05	0.876	0.997	0.607	0.926	0.213	0.470	0.027	0.053
	0.1	0.846	0.994	0.541	0.887	0.176	0.411	0.030	0.049

The power of the test using the first approach can be seen in table 4.6 and 4.7 for a length of $T = 100$ and $T = 250$. For stationary long memory, we see that the test has satisfying power properties. The power tends to decrease by an increasing value of d and decreasing sample size.

For non-stationary long memory a tendency can be observed. The power decreases strongly with increasing d. It is satisfying only for a few parameter settings. For $d = 0.8$ and $d = 0.9$ for instance, the test has rather low power. This is, however, natural, since a higher d means that the process is getting more persistent and nonlinear processes are easily confused with highly persistent processes.

Note that the first approach can be seen as a modified version of the test statistic proposed by Kapetanios and Shin (2003). The standard Wald type test of KS is denoted as $STAR_2$ in the paper and will be compared with the directed-Wald statistic. This is done only for non-stationary long memory.

Table 4.8: Power for the KS test

parameters		$d = 0.6$		$d = 0.7$		$d = 0.8$		$d = 0.9$	
α_2	γ	$T = 100$	$T = 250$	$T = 100$	$T = 250$	$T = 100$	$T = 250$	$T = 100$	$T = 250$
-1.5	0.01	0.306	0.719	0.083	0.245	0.010	0.028	0.007	0.007
	0.05	0.013	0.027	0.005	0.009	0.024	0.042	0.037	0.127
	0.1	0.029	0.051	0.043	0.094	0.056	0.150	0.094	0.253
-1	0.01	0.413	0.873	0.124	0.432	0.021	0.059	0.007	0.006
	0.05	0.066	0.133	0.012	0.018	0.015	0.010	0.024	0.062
	0.1	0.013	0.018	0.012	0.013	0.019	0.056	0.056	0.149
-0.5	0.01	0.616	0.974	0.264	0.635	0.050	0.133	0.007	0.010
	0.05	0.228	0.624	0.051	0.155	0.059	0.006	0.002	0.010
	0.1	0.113	0.300	0.017	0.053	0.021	0.003	0.014	0.024
-0.1	0.01	0.816	0.999	0.517	0.914	0.143	0.412	0.029	0.046
	0.05	0.715	0.938	0.316	0.793	0.071	0.244	0.009	0.023
	0.1	0.656	0.979	0.265	0.705	0.053	0.482	0.005	0.015

Note: the critical values of the test are provided in table 1 of Kapetanios and Shin (2003)

The power in table 4.8 has a similar tendency as in table 4.7. However, it is clear that the directed-Wald test outperforms the KS test. The results in table 4.8 are consistent to the simulation results of KS, which also show that the test has low power. Therefore, this suggests, that provided the new critical values in table 4.1, the directed-Wald test has a higher power.

Now, let us consider the power of the test when using the second approach. The gamma defined above ($\gamma \in \{0.01, 0.05, 0.1\}$) are the true values of gamma of the data generating process. To apply our test, we need to define a certain range for gamma and the statistic will be the supremum over the defined range. We use a grid of $\gamma \in (0.01, 2.5)$ with a step size of 0.01.

Before we proceed, we show a sample plot of the statistic for the respective gammas. The idea of this plot is to see whether the defined interval for γ is correctly specified.

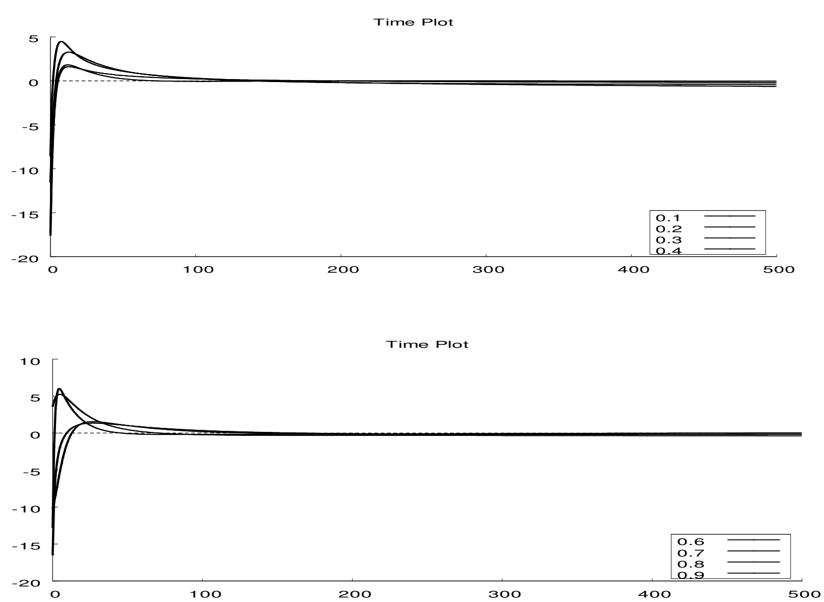


Figure 4.1: Plot of the test statistic depending on gamma

The plot is the result of the test applied to an ESTAR process with $\alpha_1 = -1.5$, $\alpha_2 = 1$ and $\gamma = 0.1$ for several d values. In the figure, there are 500 grid points at the x-axis, which represents the interval of γ, meaning that the 1st point corresponds to $\gamma = 0.01$ and the 500th point corresponds to $\gamma = 5$. Moreover, we standardize the value of the statistic in order to have a figure which covers all d. We see from the figure that the supremum is achieved in the interval 0 to 100, and it is very close to the true value of γ.

Table 4.9 and 4.10 give the results of the power experiment for the second approach.

Table 4.9: Power experiment for the Wald test for the second approach

parameters		$d = 0.1$		$d = 0.2$		$d = 0.3$		$d = 0.4$	
α_1	γ	$T = 100$	$T = 250$	$T = 100$	$T = 250$	$T = 100$	$T = 250$	$T = 100$	$T = 250$
-1.5	0.01	1.000	1.000	1.000	1.000	1.000	1.000	0.995	1.000
	0.05	1.000	1.000	1.000	1.000	0.954	0.952	0.735	0.995
	0.1	0.973	0.975	0.802	0.816	0.484	0.876	0.191	0.396
-1	0.01	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.05	1.000	1.000	1.000	1.000	0.997	1.000	0.949	0.963
	0.1	0.998	1.000	0.990	0.994	0.996	0.999	0.638	0.861
-0.5	0.01	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.05	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.998
	0.1	1.000	1.000	0.994	0.999	0.998	1.000	0.973	0.990
-0.1	0.01	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.05	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.999
	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000

Table 4.10: Power experiment for the Wald test for the second approach

parameters		$d = 0.6$		$d = 0.7$		$d = 0.8$		$d = 0.9$	
α_1	γ	$T = 100$	$T = 250$	$T = 100$	$T = 250$	$T = 100$	$T = 250$	$T = 100$	$T = 250$
-1.5	0.01	0.720	0.979	0.294	0.666	0.069	0.150	0.030	0.026
	0.05	0.100	0.186	0.042	0.042	0.081	0.147	0.180	0.439
	0.1	0.078	0.097	0.146	0.284	0.250	0.253	0.433	0.877
-1	0.01	0.851	0.997	0.457	0.837	0.111	0.265	0.034	0.020
	0.05	0.268	0.595	0.063	0.100	0.035	0.042	0.090	0.211
	0.1	0.071	0.093	0.039	0.072	0.093	0.187	0.210	0.480
-0.5	0.01	0.921	1.000	0.618	0.941	0.202	0.467	0.029	0.051
	0.05	0.629	0.961	0.390	0.538	0.053	0.098	0.025	0.035
	0.1	0.430	0.801	0.102	0.254	0.036	0.027	0.045	0.102
-0.1	0.01	0.962	1.000	0.807	0.955	0.408	0.775	0.083	0.146
	0.05	0.946	1.000	0.698	0.974	0.277	0.662	0.064	0.092
	0.1	0.928	1.000	0.633	0.958	0.235	0.543	0.103	0.077

Table 4.9 shows, that the test has considerable power for stationary long memory process and is more powerful compared to the results for the first approach. The power reaches almost 1 for all parameter settings. In line with the results of the first approach, the power tends to decrease by increasing the value of d and decreasing the sample size. However, we can see that for all cases, the power of the second approach is much higher than the first one and of course, higher than using the standard Wald type test. This result is due to the fact that the first approach uses a linearization of the nonlinear transition function. The price for such linearizations is a power loss.

4.5 Empirical application

In this section we apply the directed-Wald test to real exchange rate data. Many studies found evidence of mean reversion in real exchange rates which can be either due to long memory (see Diebold et al. (1991), Cheung (1993)) or a nonlinear ESTAR behavior (see Taylor et al. (2001) and references therein). We examine real exchange rates of several countries against the Japanese YEN. The choice of this case is motivated by a previous study of Cheung and Lai (2001). They investigated the Japanese YEN based real exchange rates of several developed countries and found a confusion between long memory and nonlinear processes. Another study about JPY bilateral real exchange rates is Chortareas and Kapetanios (2004).

We consider the bilateral real exchange rates of 22 countries against the Japanese YEN. We use quarterly data spanning from 1970Q1 to 1998Q4, which is the same period as considered in Baillie and Kapetanios (2007). Our data is taken from Datastream. These countries are: Austria, Canada, Belgium, Denmark, France, Italy, Malaysia, Korea, New Zealand, Netherland, Portugal, Spain, Sweden, Switzerland, UK, Indonesia, Thailand, the Philippines, Sri Lanka, Germany, Australia and the US.

Initially we have to show that long memory as well as nonlinear ESTAR can be detected in the considered rates. To do this, we apply two tests which are frequently used in empirical studies. We use the HML test proposed by Harris et al. (2008) for testing long memory. Moreover, we apply the nonlinear unit root test proposed by Kapetanios et al. (2003) to identify the nonlinearity, which is basically an ESTAR process. Furthermore, we need to estimate the memory parameter prior to applying the directed-Wald test. To do this, we estimate the memory parameter by means of the GPH estimator. Our initial study shows that there are only 14 cases for which our test is needed[4]. These 14 cases are given in table 4.11, which contains also the results of our test.

[4]This means that only 14 real exchange rates show a long memory behavior regarding the HML test and a nonlinear ESTAR behavior regarding the KSS test. The others have either long memory or they are ESTAR or none of them. We omit the results of the tests as well as the estimations of the long memory parameter for reasons of space. They are available from the authors upon request

Table 4.11: Empirical application: real exchange rates

	Directed-Wald test		Neglected nonlin. test	
RER	1st approach	2nd approach	ANN	TLG
Australia	4.6262	4.6261	11.4680*	11.6677*
Austria	4.0508	4.0507	2.3579	2.3553
Denmark	3.6634	3.6633	0.4437	0.4392
France	6.5467*	6.5464*	2.1239	2.1157
Germany	0.9218	0.9401	1.5767	1.6088
New Zealand	3.4046	3.4045	2.9766	2.9590
Netherland	0.7580	1.0005	3.1358	3.0728
US	2.9123	2.9122	0.5734	1.0531
Korea	61.5143*	61.5141*	10.6546*	10.5552*
Malaysia	18.2643*	18.4827*	8.6063*	8.6649*
Indonesia	74.4015*	76.7220*	11.2909*	11.0694*
Thailand	24.5955*	24.6735*	11.6930*	11.8100*
Philipina	6.1224*	6.1746*	5.4058	5.4071
Srilanka	7.5312*	7.4870*	4.9319	5.2819
No of rejection	7	7	5	5

Note: The (*) represents significance under the 5% level

From this table we see that both directed-Wald tests give consistent results and the values of their test statistics do not differ significantly. The test rejects the null in 7 cases at the 5% level of significance. It suggests that the real exchange rates of the corresponding countries can better be explained as ESTAR than as long-memory processes. We also note the interesting finding that long memory appears more likely in the real exchange rates of developed countries. Meanwhile, ESTAR is mostly found in developing countries, such as Malaysia, Korea, Thailand, Philippines, Sri Lanka and Indonesia with the exception France. This finding is consistent with previous studies about real exchange rate behavior. Those found that nonlinear adjustments towards PPP hold more likely in developing countries, due transportation costs and trade barriers. More details about the sources of nonlinearities in real exchange rates of developing countries can be found in Bahmani-Oskooee et al. (2008), Sarno and

Taylor (2001a), Sarno and Taylor (2001b) and Taylor (2003). Another empirical evidence about the existence of nonlinear ESTAR in developing countries can be found in Ceratto and Sarantis (2006). Therefore, our results provide an alternative solution to the puzzle of Cheung and Lai (2001).

In addition to the results from our test, we also apply the test of Baillie and Kapetanios (2007), denoted BK hereafter, to the real exchange rates to show the consistency of our results. The test is intended to detect any neglected nonlinearity in long memory time series by suggesting a simultaneous model such as FI-STAR, FI-GARCH or TARFIMA. Although our test is not directly comparable to the BK test, it is interesting to have the results for comparison. For this test, the nonlinear ANN test of Lee et al. (1993) or TLG test of Teräsvirta et al. (1993) is applied to the short memory component $\hat{y}_t$ after filtering the long memory u_t by using the estimated $\hat{d}$, such that

$$\hat{y}_t = (1-L)^{\hat{d}} u_t.$$

Following BK, the third order Taylor expansion is used for the TLG and ANN model and the delay parameter is set to be one. The long memory parameter is estimated by maximum likelihood. The results of the test can be seen in the two last columns of table 4.11. We see that both the ANN and TGL approach give a consistent result to our findings.

From the table, we see also that the results of the directed-Wald test and the BK test are almost the same. Both are able to detect nonlinearities in the developing countries. However, given the fact that nonlinear adjustments toward PPP are the case for most developing countries, the BK test fails to detect the nonlinearity for Philippines and Sri Lanka at the 5% significance level. Moreover, the neglected nonlinearities detected by the BK test do not imply an ESTAR specification since the basic model used is a neural network model. As we pointed out above, this test suggests to model long memory and nonlinear structures in a simultaneous model and do not consider the problem of distinguishing between the two phenomena.

4.6 Conclusions

In this paper we derive Wald-type tests to distinguish between long memory and ESTAR-type nonlinearities. We test the null hypothesis of a either stationary or non-stationary long-memory process against the alternative of an ESTAR process. Tests in the ESTAR framework have so far been based on two ideas. The first is linearizing the transition function of the ESTAR process by means of a Taylor expansion and the second is to overcome the problem of unidentified parameters by using a supremum statistics. Therefore, we derive the limit distribution for our Wald-type test under both situations showing that the supremum statistics has better power properties than the test based on the Taylor expansion. This is in line with previous findings in the literature. As our testing problem has a restricted parameter under the alternative we cannot use a standard Wald test but have to apply a directed-Wald test to overcome this problem. We derive the limit distribution of this test and show that it has fine size and satisfying power properties.

We apply our test to real exchange rates of several countries and find that mainly developing countries show an ESTAR behavior. This finding is also in line with the results of Baillie and Kapetanios (2007).

Appendix

In this part, we first describe the general outline of the proof of the theorems for both approaches. Define the model:

$$u_t = \alpha_1 F(y_{t-1}) + \alpha_2 z_t + \epsilon_t \tag{4.34}$$

with $F(y_{t-1}) = y_{t-1}^3$ for the first approach and $F(y_{t-1}, \gamma) = F(y_{t-1}) = y_{t-1}\{1 - \exp(-\gamma y_{t-1}^2)\}$ for the second. In matrix form, the model can be written as

$$\mathbf{U} = \mathbf{X}'\alpha + \epsilon \tag{4.35}$$

with $\mathbf{U} = \begin{bmatrix} u_1 \\ \vdots \\ u_t \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} F(y_0) & z_1 \\ \vdots & \\ F(y_{t-1}) & z_t \end{bmatrix}$, $\alpha = (\alpha_1, \alpha_2)'$ and $\epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_t \end{bmatrix}$.

We show pointwise convergence in distribution of the test statistic by first examining the asymptotic distribution of the OLS estimator.

Let us revisit the standard result of Kurtz and Protter (1991) about the continuous mapping theorem, that for processes ${X_t}^T \equiv X_T$ and ${Y_t}^T \equiv Y_T$ the following holds:

1. $X_T Y_T$ are $\mathcal{F}_t-$ adapted to some σ field $\mathcal{F}_t$
2. $(X_T, Y_T) \Rightarrow (X, Y)$
3. If Y_T is a semi martingale then $\int X_T dY_T \Rightarrow \int X dY$.

Also, from the fractional functional central limit theorem and standard functional central limit theorem, since u_t is a long memory process, we have that

$$\sigma_T^{-1} u_{[Tr]} \Rightarrow Y_d(r), \tag{4.36}$$

where $Y_d(r)$ is a fractional Brownian motion, $[rT]$ is the largest integer less than or equal to rT and $\sigma_T^2 = \mathbf{E}(\sum_{t=1}^{T} u_t)^2$. For iid ϵ_t we define that

$$Y_T(r) = T^{-1/2}\sigma^{-1}\sum_{t=1}^{[Tr]}\epsilon_t, \quad for \;\; 0 \le r \le 1, \tag{4.37}$$

where $Y_T(r)$ in (4.37) converges to a Brownian motion $Y^*(r)$ with $Y^*(r) = Y(1)$ when (4.37) is evaluated at $r = 1$, $\sigma^2 = \mathbf{E}\epsilon_t^2$.

From (4.35), the OLS estimate of α is

$$\hat{\alpha} = (\mathbf{X'X})^{-1}(\mathbf{X'U}). \tag{4.38}$$

Under the null and with $x_t = F(y_{t-1})$,

$$\hat{\alpha} - \alpha = (\mathbf{X'X})^{-1}(\mathbf{X'}\epsilon) \tag{4.39}$$

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{T} x_t x_t & \sum_{t=1}^{T} x_t z_t \\ \sum_{t=1}^{T} x_t z_t & \sum_{t=1}^{T} z_t z_t \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} x_t \epsilon_t \\ \sum_{t=1}^{T} z_t \epsilon_t \end{bmatrix}. \tag{4.40}$$

We begin with examining the asymptotic behavior of the terms in (4.40) which contain z_t. The other terms will be considered later. In this case, the proof is similar to KS. Let us summarize it in brief. z_t is defined by

$$z_t = \sum_{j=1}^{t} b_j u_{t-j} \tag{4.41}$$

Define the function $b(r) = b_{[Tr]}$ for $r \in (0,1)$, and its cumulative sum as $\beta(r) = \int_0^r b(s)ds$. The β can be expressed also as $\beta_t = \sum_{i=0}^{t}\beta_i$, then

$$\sigma_T^{-1} z_{[Tr]} = \sum_{i=1}^{T} \sigma_T^{-1} u_{t-i}(\beta_{t-i} - \beta_{t-i-1}) \tag{4.42}$$

$$Z_d(r) = \int_0^r Y_d(s) d\beta(s-r). \tag{4.43}$$

Therefore, for stationary long memory with $0 < d < 0.5$, we have

$$T^{-(1+2d)}\sigma_T^{-2}\sum_{t=1}^{T} z_t^2 \;\Rightarrow\; \int_0^1 Z_d(r)^2 dr \tag{4.44}$$

$$T^{-(1/2+d)}\sigma_T^{-1}\sum_{t=1}^{T} z_t\epsilon_t \;\Rightarrow\; \sigma\int_0^1 Z_d(r)dY^*(r) \tag{4.45}$$

and for non-stationary long memory, the rates of convergence are

$$T^{-2(1-d)}\sigma_T^{-2}\sum_{t=1}^{T} z_t^2 \;\Rightarrow\; \int_0^1 Z_d(r)^2 dr \tag{4.46}$$

$$T^{-(1-d)}\sigma_T^{-1}\sum_{t=1}^{T} z_t\epsilon_t \;\Rightarrow\; \sigma\int_0^1 Z_d(r)dY^*(r). \tag{4.47}$$

Proof of Theorem 1

In this section, we derive the asymptotic distributions of the other terms in (4.40). We consider first the case of stationary long memory. By using the continuous mapping theorem, we have

$$T^{-3/2}\sum_{t=1}^{T} y_{t-1}^3\epsilon_t \;\Rightarrow\; \sigma\int_0^1 Y(r)^3 dY^*(r) \tag{4.48}$$

$$T^{-3}\sum_{t=1}^{T} y_{t-1}^6 \;\Rightarrow\; \sigma\int_0^1 Y(r)^6 dr \tag{4.49}$$

$$T^{-(2+d)}\sum_{t=1}^{T} y_{t-1}^3 z_t \;\Rightarrow\; \sigma\int_0^1 Y(r)^3 dZ_d(r). \tag{4.50}$$

By combining these results, the asymptotic distribution can be written as

$$\begin{bmatrix} T^{-3}\sum_{t=1}^{T} x_t x_t & T^{-(2+d)}\sum_{t=1}^{T} x_t z_t \\ T^{-(2+d)}\sum_{t=1}^{T} x_t z_t & T^{-(1+2d)}\sum_{t=1}^{T} z_t z_t \end{bmatrix} \Rightarrow \mathcal{A}_1 \mathbf{Q}_1 \mathcal{A}_1 \tag{4.51}$$

$$\begin{bmatrix} T^{-3/2}\sum_{t=1}^{T} x_t \epsilon_t \\ T^{(-(1/2+d))}\sum_{t=1}^{T} z_t \epsilon_t \end{bmatrix} \Rightarrow \mathcal{A}_2 \mathbf{Q}_2 \tag{4.52}$$

with $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are defined as

$$\mathbf{Q}_1 = \begin{bmatrix} \int_0^1 Y(r)^6 dr & \int_0^1 Y(r)^3 dZ_d(r) \\ \int_0^1 Y(r)^3 dZ_d(r) & \int_0^1 Z_d(r)^2 dr \end{bmatrix} \tag{4.53}$$

$$\mathbf{Q}_2 = \begin{bmatrix} \int_0^1 Y(r)^3 dY^*(r) \\ \int_0^1 Z_d(r) dY^*(r) \end{bmatrix}. \tag{4.54}$$

$\mathcal{A}_1$ and $\mathcal{A}_2$ are the corresponding variance-covariance matrices defined as

$$\mathcal{A}_1 = \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix} \tag{4.55}$$

and

$$\mathcal{A}_2 = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}. \tag{4.56}$$

Then, from (4.51) and (4.52), for $0 < d < 0.5$, the rate of convergence of the OLS parameter is

$$\begin{bmatrix} T^{3/2} & 0 \\ 0 & T^{1/2+d} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \end{bmatrix} \Rightarrow \mathbf{Q_1^{-1} Q_2}. \tag{4.57}$$

By similar arguments, it is straightforward to show that for $0.5 < d < 1$

$$\begin{bmatrix} T^{3/2} & 0 \\ 0 & T^{1-d} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \end{bmatrix} \Rightarrow \mathbf{Q_1^{-1} Q_2}. \tag{4.58}$$

Proof of Theorem 2

The Wald test for the null hypothesis $\alpha_1 = \alpha_2 = 0$ is

$$W \;=\; \frac{1}{\hat{\sigma}^2}(\epsilon'\mathbf{X})(\mathbf{X'X})^{-\mathbf{1}}(\mathbf{X}'\epsilon), \tag{4.59}$$

where $\hat{\sigma}^2$ is the variance of the OLS estimator of α and $\hat{\sigma}^2 \to_p \sigma^{\mathbf{2}}$. The asymptotic distribution of the Wald statistic (4.59) follows directly from the results of theorem 1

$$W \Rightarrow \mathcal{W} \equiv \left\{\mathbf{Q_2'Q_1^{-1}Q_2}\right\}. \tag{4.60}$$

The directed-Wald statistic with $c = \infty$ is

$$DW \Rightarrow \{W + 2\log[\mathbf{\Phi}(A, \hat{\alpha_2}, Var(\hat{\alpha_2}))]\}. \tag{4.61}$$

The information matrix of $\alpha = (\alpha_1, \alpha_2)'$ is defined as (see Andrews (1998))

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}_1 & \mathcal{I}_2 \\ \mathcal{I}_2' & \mathcal{I}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{x'x} & \mathbf{x'z} \\ \mathbf{z'x} & \mathbf{z'z} \end{bmatrix} /\sigma^2. \tag{4.62}$$

For $M_x = I_T - x(x'x)^{-1}x'$, the estimate of $\hat{\alpha_2}$ in (4.40) can be written as

$$\hat{\alpha}_2 = (\mathbf{z'M_xz})^{-1}\mathbf{z'M_x}\epsilon \tag{4.63}$$

and the variance of $\hat{\alpha_2}$ is

$$Var(\hat{\alpha}_2) \;=\; \mathbf{z'M_xz}/\sigma^2 \tag{4.64}$$

$$\;=\; \mathcal{I}_3 - \mathcal{I}_2\mathcal{I}_1^{-1}\mathcal{I}_2'. \tag{4.65}$$

From $\mathbf{Q}_1$ and $\mathbf{Q}_2$, the asymptotic distributions of $\hat{\alpha}_2$ and $Var(\hat{\alpha}_2)$ are

$$\mathcal{B} \equiv \left\{(q1_{22} - q1_{12}q1_{11}q1_{12})^{-1}(q2_{21} - q1_{12}q1_{11}q2_{11})\right\} \tag{4.66}$$

$$\mathcal{V} \equiv \left\{q1_{22} - \frac{q1_{12}q1_{12}}{q1_{11}}\right\}^{-1} \tag{4.67}$$

respectively, with $q1_{ij}$ and $q2_{ij}$ are the elements of the matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ at the i-th row and j-th column. Therefore, the limit distribution of the directed Wald test can be written as

$$DW \Rightarrow \{\mathcal{W} + 2\log \mathbf{\Phi}(A, \mathcal{B}, \mathcal{V})\}\,. \tag{4.68}$$

To prove the consistency of the test under the alternative, it is sufficient to examine only the first term of the directed Wald statistic, which is the standard Wald statistic, since the factor $\mathbf{\Phi}$ is a weight which convergence to a constant. Let us write

$$W \;=\; \frac{1}{\hat{\sigma}^2}(\mathbf{U}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}(\mathbf{X}'\mathbf{U}). \tag{4.69}$$

Note that $\hat{\alpha_1}$ is dominant in (4.23) with a rate of $O_p(T)$, and $T^{-1}(X'X) = O_p(1)$. Since u_t is a long-memory process, then $(U'X)$ diverges to infinity at rate $O_p(T^{2+d})$ when $0 < d < 0.5$ and $O_p(T^{5/2-d})$ when $0.5 < d < 1$. Thereby, it is sufficient to show that the test statistic diverges to infinity with a rate of $O_p(T^{2(1+2d)})$ when $0 < d < 0.5$ and $O_p(T^{2(1-d)})$ when $0.5 < d < 1$ and the test is consistent.

Proof of Theorem 3

The second approach was originally applied by Kilic (2003) for testing of a unit root against ESTAR. The limit distribution of the exponential term is mainly derived from theorem 4.17 of White (1984). In line with the asymptotics of the first approach, we consider the asymptotics for the case of a stationary long memory. In this part, we continue to derive the asymptotic distribution of the remaining terms in (4.40).

1. **First term**: $\sum_{t=1}^{T} x_t \epsilon_t = \sum_{t=1}^{T} \epsilon_t y_{t-1}\{1-\exp(-\gamma y_{t-1}^2)\}$
   This term can be written as

$$\begin{aligned} & T^{-1/2}\sum_{t=1}^{T} \epsilon_t y_{t-1}\{1-\exp(-\gamma y_{t-1}^2)\} \\ = \ & T^{-1/2}\left\{\sum_{t=1}^{T} \epsilon_t y_{t-1} - \sum_{t=1}^{T} \epsilon_t y_{t-1}\{\exp(-\gamma y_{t-1}^2)\}\right\}. \end{aligned}$$

   By application of the continuous mapping theorem and weak convergence of stochastic integrals (see also Chan and Wei (1988), Caceres and Nielsen (2007)), the asymptotics of each element is

$$T^{-1/2}\sum_{t=1}^{T} \epsilon_t y_{t-1} \Rightarrow \sigma^2 \int_0^1 Y(r)dY^*(r) \quad (4.70)$$

$$T^{-1/2}\sum_{t=1}^{T} \epsilon_t y_{t-1}\{\exp(-\gamma y_{t-1}^2)\} \Rightarrow \sigma^2 \mu_\gamma \int_0^1 Y(r)dY^*(r). \quad (4.71)$$

   Then we have

$$T^{-1/2}\sum_{t=1}^{T} \epsilon_t y_{t-1}\{1-\exp(-\gamma y_{t-1}^2)\} \Rightarrow \sigma^2(1-\mu_\gamma)\int_0^1 Y(r)dY^*(r) \quad (4.72)$$

   with $\mu_\gamma = \mathbf{E}\{\exp(-\gamma y_{t-1}^2)\}$.

2. **Second term**: $\sum_{t=1}^{T} x_t x_t = \sum_{t=1}^{T} y_{t-1}^2\{1-\exp(-\gamma y_{t-1}^2)\}^2$
   By employing a similar procedure as for the first term, let us write

$$\begin{aligned} & T^{-1}\sum_{t=1}^{T} y_{t-1}^2\{1-\exp(-\gamma y_{t-1}^2)\}^2 \\ = \ & T^{-1}\left\{\sum_{t=1}^{T} y_{t-1}^2 - 2\sum_{t=1}^{T} y_{t-1}^2\{\exp(-\gamma y_{t-1}^2)\}\right\} \\ + \ & T^{-1}\left\{\sum_{t=1}^{T} y_{t-1}^2\{\exp(-2\gamma y_{t-1}^2)\}\right\}, \end{aligned}$$

where

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T} y_{t-1}^2 &\Rightarrow \sigma^2\int_0^1 Y(r)^2 dr && (4.73)\\
T^{-1}\sum_{t=1}^{T} y_{t-1}^2\{\exp(-\gamma y_{t-1}^2)\} &\Rightarrow \sigma^2\mu_\gamma\int_0^1 Y(r)^2 dr && (4.74)\\
T^{-1}\{\sum_{t=1}^{T} y_{t-1}^2\{\exp(-2\gamma y_{t-1}^2)\} &\Rightarrow \sigma^2\psi_\gamma\int_0^1 Y(r)^2 dr && (4.75)\\
T^{-1}\sum_{t=1}^{T} y_{t-1}^2\{1-\exp(-\gamma y_{t-1}^2)\} &\Rightarrow \sigma^2(1-2\mu_\gamma+\psi_\gamma)\int_0^1 Y(r)^2 dr && (4.76)
\end{aligned}
$$

with $\psi_\gamma = \mathbf{E}\{\exp(-2\gamma y_{t-1}^2)\}$.

3. **Third term**: $\sum_{t=1}^{T} x_t z_t = \sum_{t=1}^{T} z_t y_{t-1}\{1-\exp(-\gamma y_{t-1}^2)\}$

   Again, this term can be written as

$$
\begin{aligned}
& T^{-(1+d)}\sum_{t=1}^{T} z_t y_{t-1}\{1-\exp(-\gamma y_{t-1}^2)\}\\
= \;& T^{-(1+d)}\left\{\sum_{t=1}^{T} z_t y_{t-1} - \sum_{t=1}^{T} z_t y_{t-1}\{\exp(-\gamma y_{t-1}^2)\}\right\}.
\end{aligned}
$$

   Then, the asymptotics of each element is

$$
\begin{aligned}
T^{-(1+d)}\sigma_T^{-1}\sum_{t=1}^{T} z_t y_{t-1} &\Rightarrow \sigma\int_0^1 Z_d(r)dY(r) && (4.77)\\
T^{-(1+d)}\sigma_T^{-1}\sum_{t=1}^{T} z_t y_{t-1}\{\exp(-\gamma y_{t-1}^2)\} &\Rightarrow \sigma\mu_\gamma\int_0^1 Z_d(r)dY(r) && (4.78)
\end{aligned}
$$

   and therefore,

$$
T^{-(1+d)}\sigma_T^{-1}\sum_{t=1}^{T} z_t y_{t-1}\{1-\exp(-\gamma y_{t-1}^2)\} \Rightarrow \sigma(1-\mu_\gamma)\int_0^1 Z_d(r)dY(r). \quad (4.79)
$$

Combining these results, we obtain the following limit distribution

$$\begin{bmatrix} T^{-1}\sum_{t=1}^{T} x_t x_t & T^{-(1+d)}\sum_{t=1}^{T} x_t z_t \\ T^{-(1+d)}\sum_{t=1}^{T} x_t z_t & T^{-(1+2d)}\sum_{t=1}^{1} z_t z_t \end{bmatrix} \Rightarrow \mathcal{A}_1 \mathbf{Q}_1(\gamma)\mathcal{A}_1 \tag{4.80}$$

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} x_t \epsilon_t \\ T^{-(1/2+d)}\sum_{t=1}^{T} z_t \epsilon_t \end{bmatrix} \Rightarrow \mathcal{A}_2 \mathbf{Q}_2(\gamma) \tag{4.81}$$

with $\mathbf{Q}_1(\gamma)$ and $\mathbf{Q}_2(\gamma)$ are defined as

$$\mathbf{Q}_1(\gamma) = \begin{bmatrix} (1-2\mu_\gamma+\psi_\gamma)\int_0^1 Y(r)^2 dr & (1-\mu_\gamma)\int_0^1 Z_d(r)dY(r) \\ (1-\mu_\gamma)\int_0^1 Z_d(r)dY(r) & \int_0^1 Z_d(r)^2 dr \end{bmatrix} \tag{4.82}$$

$$\mathbf{Q}_2(\gamma) = \begin{bmatrix} (1-\mu_\gamma)\int_0^1 Y(r)dY^*(r) \\ \int_0^1 Z_d(r)dY^*(r) \end{bmatrix} \tag{4.83}$$

with $\mathcal{A}_1$ and $\mathcal{A}_2$ defined in (4.55) and (4.56) respectively. Furthermore, for $0 < d < 0.5$, the rate of convergence of the OLS parameter is

$$\begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{(1/2+d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \end{bmatrix} \Rightarrow \mathbf{Q_1}(\gamma)^{\mathbf{-1}}\mathbf{Q_2}(\gamma) \tag{4.84}$$

and for $0.5 < d < 1$

$$\begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{1-d} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \end{bmatrix} \Rightarrow \mathbf{Q_1}(\gamma)^{\mathbf{-1}}\mathbf{Q_2}(\gamma). \tag{4.85}$$

Proof of Theorem 4

The proof of theorem 4 is similar to the proof of theorem 2. The only difference is that the Wald statistic depends on the nuisance parameter γ. The Wald test for the null hypothesis $\alpha_1 = \alpha_2 = 0$ is

$$W_\gamma \;=\; \frac{1}{\hat{\sigma}^2}(\epsilon'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}(\mathbf{X}'\epsilon), \tag{4.86}$$

where $\hat{\sigma}^2$ is the variance of the OLS estimator of α and $\hat{\sigma}^2 \rightarrow_p \boldsymbol{\sigma^2}$. The limit distribution of the Wald test follows directly from the results of theorem 3:

$$W_\gamma \Rightarrow \mathcal{W}_\gamma \equiv \left\{\mathbf{Q_2}(\gamma)'\mathbf{Q_1}(\gamma)^{-\mathbf{1}}\mathbf{Q_2}(\gamma)\right\}. \tag{4.87}$$

Under $c = \infty$, the directed-Wald statistic is

$$\sup_{\gamma\in\Gamma} DW_\gamma \Rightarrow \sup_{\gamma\in\Gamma}\left\{W_\gamma + 2\log[\boldsymbol{\Phi}(A, \hat{\alpha_2}, Var(\hat{\alpha_2}))]\right\}. \tag{4.88}$$

Similar to theorem 2, we define the information matrix of $\alpha = (\alpha_1, \alpha_2)'$ by

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}_1 & \mathcal{I}_2 \\ \mathcal{I}_2' & \mathcal{I}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{x}'\mathbf{x} & \mathbf{x}'\mathbf{z} \\ \mathbf{z}'\mathbf{x} & \mathbf{z}'\mathbf{z} \end{bmatrix} /\sigma^2. \tag{4.89}$$

For $M_x = I_T - x(x'x)^{-1}x'$, the estimate of $\hat{\alpha_2}$ in (4.40) can be written as

$$\hat{\alpha}_2 = (\mathbf{z}'\mathbf{M_x}\mathbf{z})^{-1}\mathbf{z}'\mathbf{M_x}\epsilon \tag{4.90}$$

and the variance of $\hat{\alpha_2}$ is

$$Var(\hat{\alpha}_2) \;=\; \mathbf{z}'\mathbf{M_x}\mathbf{z}/\sigma^2 \tag{4.91}$$

$$\;=\; \mathcal{I}_3 - \mathcal{I}_2\mathcal{I}_1^{-1}\mathcal{I}_2'. \tag{4.92}$$

From $\mathbf{Q}_1(\gamma)$ and $\mathbf{Q}_2(\gamma)$, the asymptotic expressions of $\hat{\alpha}_2$ and $Var(\hat{\alpha}_2)$ are

$$\mathcal{B} \equiv \left\{(q1_{22} - q1_{12}q1_{11}q1_{12})^{-1}(q2_{21} - q1_{12}q1_{11}q2_{11})\right\} \tag{4.93}$$

$$\mathcal{V} \equiv \left\{q1_{22} - \frac{q1_{12}q1_{12}}{q1_{11}}\right\}^{-1} \tag{4.94}$$

respectively, with $q1_{ij}$ and $q2_{ij}$ being the elements of the matrices $\mathbf{Q}_1(\gamma)$ and $\mathbf{Q}_2(\gamma)$ in the i-th row and j-th column. We see that $\mathcal{B}$ and $\mathcal{V}$ depend on γ through q. However, the term $2\log \mathbf{\Phi}(A, \mathcal{B}, \mathcal{V})$ is a weight which converges to a constant for any given γ. Therefore, the limit distribution of the sup-directed Wald is

$$\sup_{\gamma\in\Gamma} DW_{\gamma} \Rightarrow \sup_{\gamma\in\Gamma} \left\{\mathcal{W}_{\gamma} + 2\log \mathbf{\Phi}(A, \mathcal{B}, \mathcal{V})\right\}. \tag{4.95}$$

In line with the proof of theorem 2, we need to examine the Wald statistic

$$W_{\gamma} \;=\; \frac{1}{\hat{\sigma}^2}(\mathbf{U'X})(\mathbf{X'X})^{-\mathbf{1}}(\mathbf{X'U}). \tag{4.96}$$

Again, $\hat{\alpha_1}$ is dominant in (4.30) with rate of $O_p(T)$, and $T^{-1}(X'X) = O_p(1)$. Since u_t is a long memory process, $(U'X)$ diverges to infinity at rate $O_p(T^{1+d})$ when $0 < d < 0.5$ and $O_p(T^{3/2-d})$ when $0.5 < d < 1$. The test statistic diverges to infinity with rate $O_p(T^{1+2d})$ when $0 < d < 0.5$ and $O_p(T^{2(1-d)})$ when $0.5 < d < 1$ and the test is consistent.

Proof of Theorem 5

This section proves the stochastic equicontinuity of the supremum Wald test. This condition is necessary to ensure the weak convergence $G_T(\gamma) \Rightarrow G(\gamma)$. We defined that $\gamma \in \Gamma \in R^{+}$. We examine only one fraction of the statistic which contains γ and the stochastic equicontinuity will be proved for the following term. Similar arguments can be applied for the others. Let us define

$$G_T(\gamma) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t F(y_{t-1}, \gamma),$$

with $F(y_{t-1}, \gamma) = y_{t-1}\{1 - \exp(-\gamma y_{t-1}^2)\}$. The term $G_T(\gamma)$ above is similar to $U_n(v)$ of Seo (2004). Therefore, we follow Seo (2004) in proving the stochastic equicontinuity.

Let us define $F(\gamma) = F(y_{t-1}, \gamma)$. By using assumption 1 and 2, we have

$$\begin{aligned}
P(\sup_{|\gamma-\gamma'|\leq\delta} |G_T(\gamma) - G_T(\gamma')| > \epsilon) &\leq \frac{1}{\epsilon}\mathbf{E} \sup_{|\gamma-\gamma'|\leq\delta} |G_T(\gamma) - G_T(\gamma')| \\
&= \frac{1}{\epsilon}\mathbf{E} \sup_{|\gamma-\gamma'|\leq\delta} |\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t (F(\gamma) - F(\gamma'))| \\
&= \frac{1}{\epsilon}\mathbf{E} \sup_{|\gamma-\gamma'|\leq\delta} |\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t F^{'}(\gamma^*)(\gamma - \gamma')| \\
&\leq \frac{\delta}{\epsilon}\mathbf{E} \sup_{\gamma\in\Gamma} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} |\epsilon_t||F^{'}(\gamma)| \\
&\leq \frac{\delta}{\epsilon} \sup_{t} \| \sup_{\gamma\in\Gamma} |F^{'}(\gamma)| \|_2 \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \|\epsilon_t\|_2,
\end{aligned}$$

where $\gamma^* \in [\gamma, \gamma']$. By using Burkholder's inequality, it can be shown that $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \|\epsilon_t\|_2 \leq c_1 \sup_t \|\epsilon_t\|_2$, where $c_1 = 36\sqrt{2}$. For $T \to \infty$ and small δ, $P(\sup_{|\gamma-\gamma'|\leq\delta} |G_T(\gamma) - G_T(\gamma')| > \epsilon) \to 0$.

Chapter 5

A NEW SIMPLE TEST AGAINST SPURIOUS LONG MEMORY USING TEMPORAL AGGREGATION

5.1 Introduction

Let x_t be a linear long memory process characterized mainly by the following condition

$$\rho_k \sim C_\rho(k)k^{2d-1}, \quad \textbf{as } k \to \infty \tag{5.1}$$

for $d \in (0, 0.5)$. We consider an aggregated long memory process defined as

$$y_t = \sum_{j=0}^{m-1} x_{mt-j} = \sum_{j=0}^{m-1} B^j x_{mt} \tag{5.2}$$

where B is backshift operator and m denotes the aggregation level. Chambers (1998), Man and Tiao (2001) and Souza (2008) show that if x_t satisfies (5.1) with $d < 0.5$, then its aggregation process y_t also satisfies (5.1) with the same fractional integration order d. This condition implies invariance of the memory parameter to aggregation.

Spurious long memory can arise in many cases, especially in stock market data. It still has been highly debated whether the observed long memory is real or a spurious phenomena. Many studies found long memory in the volatility of stock returns (Heimstra and Jones (1997), Henry (2002), Tolvi (2003) among others). Lobato and Savin (1998) and the references therein discuss the real and spurious long memory properties of stock market data. They investigated major causes of spurious long memory, such as aggregation, nonstationarity and regime switching. By using the LM type test of Lobato and Robinson (2003), they estimated the memory parameter and tested the significance of the parameter to conclude whether the observed memory is real or spurious. However, it is well known that several processes are able to create spurious long memory by generating a certain degree of fractional integration (see Granger and Ding (1996), Granger and Teräsvirta (1999) among others). Therefore, developing a test which is able to distinguish long memory from spurious processes is still of interest, which may lead to a proper model choice.

The fact that the memory parameter does not change with aggregation can be used as a means to distinguish long memory from spurious processes. Ohanissian, Russell and Tsay (2008) estimate the memory parameter across several aggregation levels and propose a Wald type test to distinguish these two phenomena. They show that the test is able to detect several spurious processes in the alternative with considerable power. Their results are based on the simulation study by examining very large numbers of observations, meaning that it has a good performance for high frequency data and our initial study shows that the test looses the power significantly under small and finite sample sizes. Furthermore, they use the GPH method of Geweke and Porter-Hudak (1983) to estimate the memory parameter and the theoretical properties of the test have been well investigated. However, Teles et al. (1999) proved that using the GPH estimator of aggregated series for testing long memory has very serious consequences on the power of the test which may lead to the wrong conclusions, especially by using a bandwidth frequency of $T^{0.5}$.

In this paper, we propose a new test against spurious long memory based on the invariance principle, in line with the basic idea of Ohanissian et al. (2008). Our test calculates a value for every pair of aggregation levels and takes the maximum among the values. This testing procedure has been previously applied by Beran and Terrin (1996) for testing for changes in the memory parameter. Moreover, we estimate the long memory parameter by applying the semi-parametric local Whittle maximum likelihood instead of the GPH estimator. This estimation method has been proved to have the smallest bias with a minimum standard deviation (Souza (2007)).

This paper is organized as follows. Section 5.2 discusses the main result including the proposed test and its asymptotic distribution. Section 5.3 presents the results of simulation study to assess the test performance in finite sample size. The empirical application, ie. the case of German stock returns is given in section 5.4. The proof is given in the appendix.

5.2 Main result

A stationary ARFIMA(p, d, q) process x_t has the following representation:

$$\phi(B)(1-B)^d x_t = \theta(B)\epsilon_t \quad t = 1, ..., N \tag{5.3}$$

where B is the backshift operator, $\phi(B)$ and $\theta(B)$ are the AR and MA polynomials respectively and ϵ_t is a white noise process. The spectral density of (5.3) satisfies

$$f_x(\omega) = C_f(\omega)|\omega|^{-2d} \;\; \textbf{as} \;\; \omega \to 0 \tag{5.4}$$

We aggregate the process x_t by a level of aggregation m following (5.2), with $m = 2, ..., M$. Under the aggregated series y_t, the series length becomes $n = N/m$. Note that $m = 1$ corresponds to the original series x_t. The spectral density of y_t with memory parameter d satisfies

$$f_y(\lambda) \sim m^{2d+1} C_{f_x}(\omega)|\lambda|^{-2d}, \;\; \textbf{as}\ \lambda \to 0 \tag{5.5}$$

where $\lambda = 2\pi jm/N = \omega m$ and the periodogram of y_t is given by

$$I_{y^{(m)}}(\lambda_j) = \frac{1}{2\pi n}\left|\sum_{j=1}^{n}(y_j - \bar{y})\exp^{ij\lambda_j}\right|^2, \;\; \bar{y} = \sum_{j=1}^{n} y_j/n \tag{5.6}$$

Our statistic is constructed based on the semi-parametric local Whittle estimator proposed by Robinson (1995). Let us consider the Gaussian objective function for the original series x_t:

$$Q(G,d) = \frac{1}{l}\sum_{j=1}^{l}\left[\log(G\omega_j^{-2d}) + \frac{\omega_j^{2d}}{G}I_x(\omega_j)\right] \tag{5.7}$$

by which discrete averaging is evaluated over a small bandwidth frequency $l < N$. As G can be estimated by $\hat{G} = \frac{1}{l}\sum_1^l \omega_j^{2d} I_x(\omega_j)$, the memory parameter d can be estimated by minimizing the following objective function

$$\mathcal{Q}(d) = \log\left(\frac{1}{l}\sum_1^l \omega_j^{2d} I_x(\omega_j)\right) - 2d\frac{1}{l}\sum_1^l \log\omega_j \tag{5.8}$$

Souza (2007) discusses consistency of the estimator for aggregated series. It is worthwhile to summarize it as follows. Under the following regularity conditions:

1. As $\lambda \to 0+$, $f(\omega) \sim G_0 \in (0,\infty)$ and $-0.5 < \Delta_1 \leq d \leq \Delta_2 < 0.5$.

2. In a neighborhood $(0,\delta)$ of the origin, $f(\omega)$ is differentiable and

$$\frac{d}{d\omega}\log f(\omega) = O(\omega^{-1})\ as\ \omega \to 0+$$

3. $$x_t - \mathbf{E}[x_0] = \sum_{j=0}^{\infty}\alpha_j\epsilon_{t-j}, \quad \sum_{j=0}^{\infty}\alpha_j^2 < \infty$$

   where $\mathbf{E}(\epsilon_t|F_{t-1}) = 0, \mathbf{E}(\epsilon_t^2|F_{t-1}) = 1\ a.s.,\ t = 0, \pm1, ...,$ in which F_t is the σ-field of events generated by $\epsilon_s, s \leq t$, and there exists a random variable ϵ such that $\mathbf{E}(\epsilon^2) < \infty$ and for all $\eta > 0$ and some $K > 0, P(|\epsilon_t| > \eta) \leq KP(|\epsilon| > \eta)$.

4. As $N \to \infty, \frac{1}{l} + \frac{l}{N} \to 0$

5. $f(\omega)$ is bounded above and $f'(\omega)$ exists and is finite in the vicinity of the non-zero Nyquist frequencies[1].

[1]Nyquist frequency is the frequency with the sampling rate of $2\pi/N$

6. For some $\beta \in (0,2]$, as $\omega \to 0+$, $f(\omega) \sim G_0 \omega^{-2d}(1+O(\omega^{\beta}))$, where $G_0 \in (0,\infty)$ and $-0.5 < \Delta_1 \le d \le \Delta_2 < 0.5$.

7. In a neighborhood $(0,\delta)$ of the origin, $\alpha(\omega)$ is differentiable and

$$\frac{d}{d\omega}\alpha(\omega) = O\left(\frac{|\alpha(\omega)|}{\omega}\right), \quad as \;\; \omega \to 0+$$

where $\alpha(\omega) = \sum_{j=0}^{\infty} \alpha_j e^{ij\omega}$

8. Condition 3 holds and also $\mathbf{E}(\epsilon_t^3|F_{t-1}) = \mu_3, a.s., \mathbf{E}(\epsilon_t^4) = \mu_4, t = 0, \pm 1, ...$ for finite constant μ_3 and μ_4.

9. As $N \to \infty$, there exists a β satisfying Condition 6 such that

$$\frac{1}{l} + \frac{l^{1+2\beta}(\log l)^2}{N^{2\beta}} \to 0$$

If condition 1 to 5 hold for x_t, then it builds the consistency of the local Whittle estimator for aggregated time series y_t. Also, if condition 5 to 9 hold for x_t, then the local Whittle estimator for y_t is asymptotically normal such that

$$\sqrt{l}(\hat{d} - d) \xrightarrow{D} N(0, 1/4) \tag{5.9}$$

The readers are referred to Souza (2007) for the proof and the details of these conditions.

Now, consider two objective functions for two aggregated series $y^{(m_1)}$ and $y^{(m_2)}$ as follows:

$$\begin{aligned} \mathcal{Q}(n_1, d) &= \log\left(\frac{1}{l}\sum_{1}^{l} \lambda_j^{2d} I_{y^{(m_1)}}(\lambda_j)\right) - 2d\frac{1}{l}\sum_{1}^{l} \log \lambda_j \\ \mathcal{Q}(n_2, d) &= \log\left(\frac{1}{l}\sum_{1}^{l} \lambda_j^{2d} I_{y^{(m_2)}}(\lambda_j)\right) - 2d\frac{1}{l}\sum_{1}^{l} \log \lambda_j \end{aligned}$$

where $\mathcal{Q}(n_1, d)$ and $\mathcal{Q}(n_2, d)$ denote the objective function of the aggregated series y_t with level m_1 and m_2 respectively. From this, the local Whittle estimator $\hat{d}$ is defined by

$$\hat{d}^{(m_1)} = \textbf{argmin}\ \mathcal{Q}(n_1; \hat{d}), \quad \hat{d}^{(m_2)} = \textbf{argmin}\ \mathcal{Q}(n_2; \hat{d})$$

We will test the constancy of the estimated memory parameter among several aggregation levels to prove the invariance principle of the memory parameter to aggregation. The null hypothesis we attempt to test is that

$$\mathbf{H_0} : d^{(m_1)} = d^{(m_2)} = ... = d^{(m_M)}$$

The alternative hypothesis is, therefore, defined as any violation of the equalities in $\mathbf{H_0}$, i.e at least one pair of aggregated levels, m_i and m_j, $d^{(m_i)} \neq d^{(m_j)}$ where $i \neq j$.

In this paper, the idea of the test is similar to testing for a change in the long memory parameter ((see Beran and Terrin (1996), Horváth and Shao (1999), Lee and Lee (2007)). To test the constancy of the long memory parameter between two aggregated levels $\{m_1 \neq m_2\}$, we propose the following statistic

$$z_{m_1,m_2} = \sqrt{n_1 + n_2} \left\{ \frac{n_1 n_2}{(n_1 + n_2)^2} \right\} \left(\hat{d}^{(m_1)} - \hat{d}^{(m_2)} \right).$$

The calculation of z_{m_1,m_2} involves two levels of aggregated series for all combinations of the paired m. It means that for any choice of M aggregation level, we have ${}_MC_2$ values of z. In this case, M is chosen such that the aggregated series can still be used for estimating the long memory parameter. The maximum value is proposed as the statistical test. Therefore, to test the constancy of the parameter d among several aggregation levels, we suggest the statistic

$$\chi_n = \max_{1 \leq i,j \leq M} |z_{m_i,m_j}|, \quad i \neq j$$

The asymptotic distribution of the proposed test statistic is given in the following theorem.

Theorem 1: *Assume* $0 < d < 0.5$ *and the condition 6, 7, 8 and 9 are satisfied, then by the asymptotic normality of* $\hat{d}$ *we have for* $m_1 \neq m_2$

$$z_{m_1,m_2} \xrightarrow{D} \sigma V(t)$$

in $\mathbf{D}[0,1]$ *as* $T \to \infty$ *and* $V(t), 0 \leq t \leq 1$ *is a Brownian bridge and* $\xrightarrow{D}$ *denotes convergence in distribution. Hence, the statistic* χ_n *converges to*

$$\chi_n \xrightarrow{D} \sigma \sup_{0 \leq t \leq 1} |V(t)|, \quad i \neq j$$

and the variance σ^2 *is given by*

$$\sigma^2 = \mathbf{E}(\epsilon_0^4 - \sigma_\epsilon^4) \left(\sum_{j=0}^{\infty} a_j c_j \right)^2 + \sigma_\epsilon^2 \sum_{l=1}^{\infty} \left(\sum_{j=0}^{\infty} \{a_j c_{j+l} + c_j a_{j+l}\} \right)^2 .$$

From the theorem above, we reject the null hypothesis for large values of χ_n. In principle, it is possible to generate the critical values from a sequence of Brownian bridges $V(t)$ and variances σ^2 as written in the theorem. However, it seems that σ^2 has a very complicated form which leads to some difficulties. To avoid this, the critical values will be determined by using the simulated sampling distribution of χ_n.

5.3 Simulation

This section carries out simulation studies to obtain the critical values, as well as to assess the test performance in finite samples. As we pointed out above, the critical values are obtained by using the simulated sampling distribution of $\max_{1 \leq i,j \leq M} |z_{m_i,m_j}|$. It is done by generating samples of length 50000 and 10000 replications. The aggregation levels are set to be $m = 2, 3, 4, 6, 8, 12$, which are commonly used in empirical applications as suggested by Teles et al. (1999, 2008). In the latter work, they studied the effect of the use of aggregate time series on the Dickey-Fuller test for unit root, and a new unit root test based on aggregate time series was developed.

Table 5.1: Quantile of the asymptotic distribution

d	sign. level	Aggregation level (m)					
		2	3	4	6	8	12
0.1	90%	0.4542	0.5549	0.6030	0.6872	0.7586	0.7588
	95%	0.5258	0.6456	0.7133	0.7740	0.8248	0.8407
	99%	0.7098	0.8253	0.8708	0.9034	0.9518	1.0595
0.2	90%	0.5095	0.6351	0.6579	0.7232	0.7454	0.7780
	95%	0.5909	0.7203	0.7650	0.8108	0.8454	0.8516
	99%	0.8253	0.8708	0.9177	0.9756	0.9784	0.9967
0.3	90%	0.5164	0.6572	0.7160	0.7298	0.7802	0.7953
	95%	0.6347	0.7617	0.8025	0.8354	0.8723	0.8849
	99%	0.8195	0.9282	0.9832	1.0579	1.0718	1.0534
0.4	90%	0.6111	0.6981	0.7530	0.8083	0.8226	0.8390
	95%	0.7037	0.8213	0.8495	0.8982	0.9517	0.9179
	99%	0.8732	0.9930	1.0313	1.0782	1.0995	1.0796

Table 5.1 provides the critical values of the test for $d = 0.1, 0.2, 0.3, 0.4$. We see that the critical value increases with d and m through the constant σ in theorem 1.

A size experiment is done by evaluating the performance of the test in finite samples. In this case, we generate 1000 time series length of 5000. The rejection rate is calculated based on the critical values in table 5.1. The data generating process (DGP) is a pure stationary long memory with degree of fractional integration $d = 0.1, 0.2, 0.3, 0.4$. Therefore, the DGP does not account for short range dependencies. The model can be rewritten as

$$(1 - B)^d x_t = \epsilon_t \quad t = 1, ..., N.$$

Table 5.2 presents the mean and standard deviation of the estimated long memory parameter for several aggregation levels. It is useful to assess the performance of the local Whittle estimator.

Table 5.2: Invariance of memory parameter to aggregation

d	Aggregation level (m)					
	2	3	4	6	8	12
0.1	0.1002 (0.0209)	0.1053 (0.0253)	0.1016 (0.0262)	0.1021 (0.0377)	0.1045 (0.0377)	0.0999 (0.047)
0.2	0.2061 (0.0227)	0.2030 (0.0275)	0.2070 (0.0296)	0.2007 (0.0372)	0.2112 (0.0412)	0.2131 (0.0490)
0.3	0.3058 (0.0254)	0.3084 (0.0283)	0.3106 (0.0309)	0.3138 (0.0370)	0.3155 (0.0370)	0.3160 (0.0400)
0.4	0.4088 (0.0220)	0.4143 (0.0244)	0.4135 (0.0308)	0.4163 (0.0336)	0.4160 (0.0405)	0.4238 (0.0527)

Note: The Data Generating Process (DGP) is ARFIMA(0,d,0)

As expected, the estimated memory parameters are very close to the original value. For instance, under ARFIMA(0,0.1,0), the estimated memory parameters range from 0.0999 to 0.1053. Also, under the DGP ARFIMA(0,0.2,0), the estimated memory parameters range from 0.2007 to 0.2131, and so they do for ARFIMA with $d = 0.3$ and $d = 0.4$. It indicates that the local Whittle estimator is a good approximation for our test. In line with Souza (2007), the standard deviation of the estimated memory parameter increases with the aggregation level. The following table presents the result of size experiment.

Table 5.3: Size experiment

d	nom. size	Aggregation level (m)					
		2	3	4	6	8	12
0.1	0.05	0.059	0.042	0.041	0.049	0.053	0.044
	0.1	0.106	0.086	0.103	0.086	0.093	0.096
0.2	0.05	0.058	0.037	0.052	0.043	0.057	0.051
	0.1	0.090	0.088	0.098	0.078	0.103	0.085
0.3	0.05	0.055	0.032	0.047	0.054	0.056	0.045
	0.1	0.101	0.088	0.070	0.097	0.101	0.078
0.4	0.05	0.042	0.054	0.050	0.046	0.048	0.046
	0.1	0.090	0.101	0.092	0.087	0.098	0.094

Note: The Data Generating Process (DGP) is ARFIMA(0,d,0)

From table 5.3, it is obvious that the rejection rate is very close to the nominal value although some values indicate size distortions, meaning that the test is correctly sized under the null of long memory.

The power experiment is carried out by generating several processes which are able to create spurious long memory, ie. Markov switching, STOP-BREAK and random level shift processes. These models can be described as follows:

- Markov-switching process,

$$x_t = \begin{cases} \phi_1 x_{t-1} + \epsilon_t & \text{if } s_t = 1 \\ \phi_2 x_{t-1} + \epsilon_t & \text{if } s_t = 2 \end{cases}$$

  with $\epsilon_t \sim N(0,1)$ and the state transition probability p_{00} and p_{11}.

- STOP-BREAK process,

$$x_t = \mu_t + \epsilon_t, \quad \mu_t = \mu_{t-1} + \frac{\epsilon_{t-1}^2}{\gamma + \epsilon_{t-1}^2} \epsilon_{t-1}$$

with $\epsilon_t \sim N(0,1)$.

- Stationary random level shift process,

$$x_t = \mu_t + \epsilon_t, \quad \mu_t = (1 - j_t)\mu_{t-1} + j_t \varepsilon_t$$

with j_t is IID Bernoulli(p), ε_t and ϵ_t are short memory process with mean 0 and variance $\sigma^2_{\varepsilon_t}$

- Nonstationary random level shift process,

$$x_t = \mu_t + \epsilon_t, \quad \mu_t = \mu_{t-1} + j_t \varepsilon_t$$

with j_t is IID Bernoulli(p), ε_t and ϵ_t are short memory process with mean 0 and variance $\sigma^2_{\varepsilon_t}$

These models are strong candidates which can easily mislead the properties of long memory (Granger and Ding (1996), Diebold and Ineoue (2001), Granger and Hyung (2004), Sibbertsen (2004b), Banerjee and Urga (2005)). We call them model 1, model 2, model 3 and model 4 respectively hereafter. Basically, they are short memory processes with zero integration order. Therefore, any degree of fractional integration more than zero observed from these processes are spurious results. For each model, the considered parameters as well as the result of the power experiment can be seen in table 5.4, 5.5, 5.6 and 5.7.

In this part, we generate data with two different sample sizes, $N = 2000$ and $N = 5000$ with 1000 replications. Note that for $N = 2000$, it is considered a very small sample in practice, especially in the context of volatility modeling. Meanwhile, $N = 5000$ is a reasonable sample size for this case. Moreover, aggregating 5000 sample size with level of 12 results on big enough samples required to estimate the memory parameter. In the table, we present the mean value of the fractional integration order obtained from a sample size of 5000. A smaller bias is observed for a smaller sample size. However, we omit the results for the reason of space.

Table 5.4: Power experiment

Model 1									
	$p_{00} = p_{11} = 0.90$			$p_{00} = p_{11} = 0.90$			$p_{00} = p_{11} = 0.90$		
	$\phi_1 = -\phi_2 = 0.8$			$\phi_1 = -\phi_2 = 0.5$			$\epsilon_1 = N(1,1), \epsilon_2 = N(-1,1)$		
m	mean(d)	reject freq.		mean(d)	reject freq.		mean(d)	reject freq.	
		N=2000	N=5000		N=2000	N=5000		N=2000	N=5000
1	0.3470 (0.0251)	-	-	0.1031 (0.0193)	-	-	0.3281 (0.0154)	-	-
2	0.2115 (0.0319)	0.989	1.000	0.0450 (0.0225)	0.799	0.949	0.2712 (0.0211)	0.207	0.878
3	0.1567 (0.0318)	0.994	1.000	0.0366 (0.0271)	0.666	0.940	0.2419 (0.0285)	0.432	0.995
4	0.1178 (0.0376)	0.998	1.000	0.0253 (0.0306)	0.731	0.940	0.1759 (0.0340)	0.680	0.995
6	0.0988 (0.0370)	0.999	1.000	0.0200 (0.0359)	0.722	0.901	0.1299 (0.0337)	0.795	1.000
8	0.0610 (0.0455)	1.000	1.000	0.0120 (0.0407)	0.653	0.870	0.0943 (0.0437)	0.808	1.000
12	0.0411 (0.0495)	1.000	1.000	0.0072 (0.0518)	0.567	0.854	0.0679 (0.0456)	0.852	1.000

Note: The third model specification has parameter $\phi_1 = -\phi_2 = 0$

Table 5.5: Power experiment

Model 2									
	$\gamma = 180$			$\gamma = 90$			$\gamma = 40$		
m	mean(d)	reject freq.		mean(d)	reject freq.		mean(d)	reject freq.	
		N=2000	N=5000		N=2000	N=5000		N=2000	N=5000
1	0.2290 (0.0587)	-	-	0.3409 (0.0571)	-	-	0.4709 (0.0554)	-	-
2	0.2842 (0.0590)	0.608	0.989	0.4055 (0.0655)	0.809	0.985	0.5660 (0.0645)	0.999	1.000
3	0.3353 (0.0658)	0.794	1.000	0.4589 (0.0667)	0.928	0.996	0.6276 (0.0696)	0.998	1.000
4	0.3577 (0.0712)	0.735	1.000	0.5025 (0.0713)	0.957	0.995	0.6708 (0.0738)	1.000	1.000
6	0.4005 (0.0831)	0.779	1.000	0.5586 (0.07613)	0.985	1.000	0.7407 (0.0689)	1.000	1.000
8	0.4458 (0.0781)	0.823	1.000	0.6003 (0.0797)	0.987	1.000	0.7815 (0.0602)	1.000	1.000
12	0.5011 (0.0865)	0.828	1.000	0.6817 (0.0852)	0.987	1.000	0.8417 (0.0569)	1.000	1.000

Table 5.6: Power experiment

Model 3									
		$p = 0.001$			$p = 0.01$			$p = 0.1$	
m	mean(d)	reject freq.		mean(d)	reject freq.		mean(d)	reject freq.	
		N=2000	N=5000		N=2000	N=5000		N=2000	N=5000
1	0.2596 (0.0919)	-	-	0.4931 (0.0738)	-	-	0.6581 (0.2208)	-	-
2	0.3370 (0.1134)	0.553	0.951	0.5845 (0.0777)	0.955	1.000	0.7238 (0.2788)	-	-
3	0.3747 (0.1223)	0.625	0.965	0.6419 (0.0899)	0.986	1.000	0.8070 (0.2468)	-	-
4	0.4047 (0.1275)	0.647	0.963	0.6881 (0.0940)	0.985	1.000	0.8022 (0.2980)	-	-
6	0.4606 (0.1596)	0.664	0.968	0.7563 (0.0925)	0.992	1.000	0.8770 (0.2527)	-	-
8	0.4926 (0.1659)	0.668	0.978	0.8106 (0.09004)	0.983	1.000	0.8797 (0.2782)	-	-
12	0.5976 (0.1617)	0.634	0.981	0.8554 (0.1097)	0.990	1.000	0.8747 (0.3131)	-	-

Table 5.7: Power experiment

Model 4									
		$p = 0.001$			$p = 0.01$			$p = 0.1$	
m	mean(d)	reject freq.		mean(d)	reject freq.		mean(d)	reject freq.	
		N=2000	N=5000		N=2000	N=5000		N=2000	N=5000
1	0.2802 (0.0911)	-	-	0.4927 (0.0681)	-	-	0.7185 (0.0567)	-	-
2	0.3374 (0.1109)	0.553	0.941	0.5950 (0.0723)	0.963	1.000	0.8266 (0.0486)	-	-
3	0.3875 (0.1048)	0.560	0.964	0.6496 (0.0712)	0.996	1.000	0.8806 (0.0407)	-	-
4	0.4064 (0.1277)	0.618	0.972	0.7110 (0.0713)	0.999	1.000	0.9124 (0.0361)	-	-
6	0.4587 (0.1445)	0.683	0.973	0.7656 (0.0692)	1.000	1.000	0.9483 (0.0347)	-	-
8	0.5258 (0.1254)	0.640	0.975	0.8118 (0.0664)	1.000	1.000	0.9665 (0.0368)	-	-
12	0.5757 (0.1398)	0.626	0.980	0.8744 (0.0607)	1.000	1.000	0.9827 (0.0470)	-	-

Dealing with the ability of the processes to resemble long memory, we see that all data generating processes are able to generate fractional integration orders which lie in the long memory range. It can be seen from the mean values of the long memory parameter under $m = 1$, which corresponds to the original series. Therefore, the examined parameters are correctly specified. However, the point of consideration in this paper is not focused on whether the models are able to create spurious long memory or not, since it has been proved in the aforementioned references. Through the power experiment, we assess the behavior of the estimated memory parameter to aggregation and the ability of our test to specify these models into their class, which is spurious long memory. Since our test involves a pair of aggregation levels, thus we cannot obtain any value for $m = 1$. We denote it with "-" in the table.

Let us consider Markov switching processes in table 5.4. The choice of the transition probabilities mainly refers to previous works which found that the higher the transition probability p_{ii}, the longer the process is expected to remain in state i and the process becomes more persistent. Under this condition, the process will easily be confused with long memory (see chaper 2 and chapter 3 for intensive simulation results). The first two parameter settings in model 1 are general Markov switching processes and the last is Markov switching with iid regimes (MS-IID) and therefore, $\phi_1 = -\phi_2 = 0$. From table 5.4, under the defined parameter settings, the test is able to specify the Markov switching processes as spurious long memory process with high power. Only two cases have power lower than 0.5. The power increases with sample size and shows no monotonic tendency regarding the level of aggregation. However, we can see that most cases have higher power with higher aggregation level.

Now, we discuss the results for model 2. The STOP-BREAK model was introduced by Engle and Smith (1999). Similar results as for Markov switching are observed for this case. Under the three different parameter settings defined in table 5.5, the test is able to detect the model as spurious long memory with satisfying power, both in small and medium samples. Especially for $N = 5000$, the power reaches almost one for all cases. For random level shift processes, either stationary or nonstationary, the test also performs very well. Under small probabilities for the Bernoulli distribution, the estimated fractional integration parameters are biased towards stationary long memory. For $p = 0.1$, the memory parameter is biased towards nonstationary long memory. It indicates that a higher probability leads to a more persistent process. Since our test is derived under stationary long memory, therefore, this case (nonstationary long memory with $d \geq 0.5$) is out of consideration and the power of the test cannot be presented. The considered random level shift processes in this paper were firstly introduced by Chen and Tiao (1990). Further conditions about the possibility of these models to resemble long memory have been investigated by Breidt and Hsu (2002).

Our results in this experiment are consistent with the test proposed by Ohanissian et al. (2008). Their test is also able to distinguish long memory from the spurious processes with extremely high power. However, as we pointed out before, their test is applicable to high frequency data and looses the power significantly in finite samples. Therefore, our test fills this gap by having good performance in finite sample size.

5.4 Empirical application

The dataset used in this study consists of daily absolute and squared returns for 9 German stock price series, listed in the DAX30. The examined cases are Allianz, BASF, BAYER, BMW, Commerz Bank, Continental, Deutsche Bank, Siemens and Volskwagen (VW) spanning from the period of January 1973 to December 2007. Therefore, we have 9132 observations for each stock. Several previous studies found long memory in the volatility of German stock returns (Sibbertsen (2004a), Gurgul and Wojtowiczh (2006)), based on the fact that several estimation procedures such as GPH, the Whittle estimator or Wavelet estimator give a fractional integration order within the long memory range. Again, it becomes crucial since several processes are able to create spurious long memory by having a certain degree of fractional integration as discussed in the previous section. Hassler and Olivares (2007) independently study the daily absolute returns of the German stock price index DAX and found a significant break in mean, which might be one source of the spurious long memory.

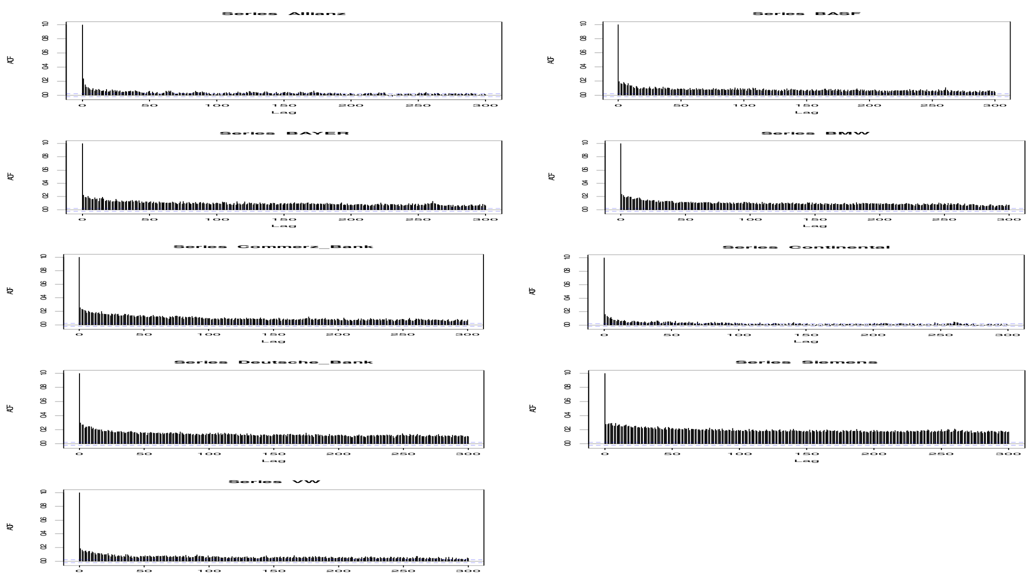


Figure 5.1: ACF plot of absolute returns

Figure 5.1 and 5.2 depict the autocorrelation function (ACF) of absolute and squared returns of the considered stocks respectively. We plot the autocorrelations up to 300 lags. The figures show that the autocorrelations of both absolute and squared returns are strongly correlated until long lags.

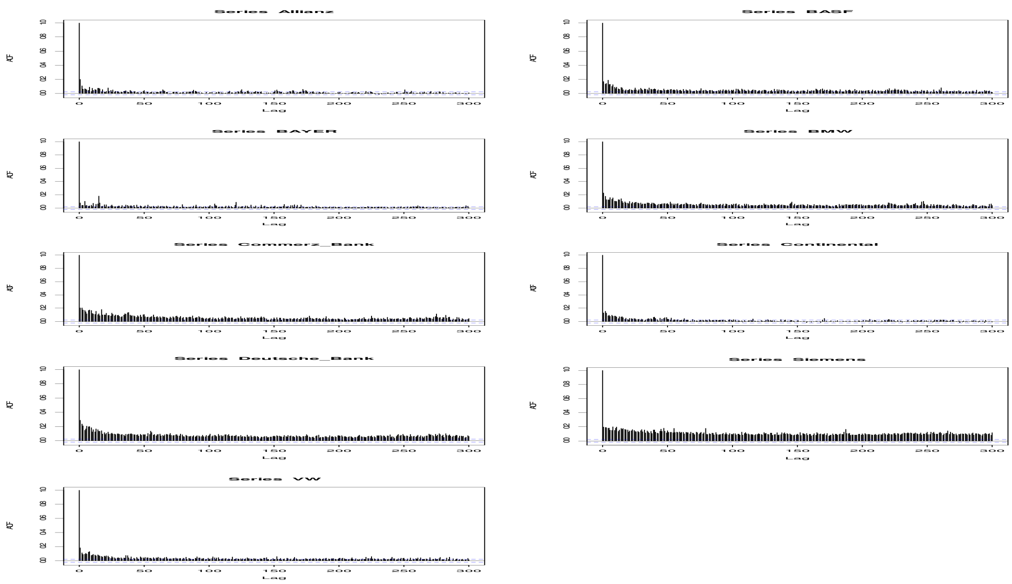


Figure 5.2: ACF plot of squared returns

They decay slowly with hyperbolic rates and show the property of long memory. Again, having this property does not provide enough evidence that the processes are long memory. In chapter 2, we demonstrated that several nonlinear processes under specific parameter settings may produce a similar feature of the autocorrelation function as under long memory. This similarity holds also for the spectrum of both processes. Therefore, using only this information may lead to the wrong conclusions.

We apply our test as a formal procedure to detect whether the long memory which can be observed in German stock returns is real or spurious. The results of the test are presented in table 5.8 and table 5.9, for absolute and squared returns respectively. In the tables, we provide the estimated long memory parameter of the aggregated series under several aggregated levels m. The value in the last column is the statistic $|\lambda_n|$ obtained from applying the test with $m = 12$. This choice is based on the simulations which suggest that the test tends to have more power for high aggregation levels.

Table 5.8 below presents the results of the test for absolute returns.

Table 5.8: Test for absolute returns

stock	m							$\|\lambda_n\|$
	1	2	3	4	6	8	12	
Allianz	0.1959	0.2170	0.2363	0.2426	0.2587	0.2883	0.3272	1.0166*
BASF	0.2365	0.2945	0.3201	0.3201	0.2982	0.3070	0.3475	1.7279*
BAYER	0.2491	0.2880	0.3373	0.3640	0.3872	0.3963	0.4189	1.9809*
BMW	0.2437	0.3015	0.3569	0.3730	0.3894	0.3942	0.4050	2.3434*
Commerz Bank	0.2705	0.3142	0.3534	0.3795	0.3982	0.4335	0.4806	1.8642*
Continental	0.2060	0.2280	0.2460	0.2455	0.2499	0.2763	0.3068	0.8276**
Deutsche Bank	0.2701	0.3398	0.3936	0.3986	0.3966	0.3898	0.4367	2.5551*
Siemens	0.2951	0.3480	0.3766	0.3404	0.4323	0.4709	0.5167	2.3127*
VW	0.2278	0.2829	0.3097	0.3473	0.3440	0.3582	0.3623	2.0418*

Note: the asterisks * and ** refer to the significance levels of 5% and 10% respectively

From the table, we see that for the 5% level of significance the test rejects almost all cases, except for Continental. Since we have under the alternative hypothesis that there is a violation to the invariant condition of the estimated memory parameters, to reject the null hypothesis means that the observed long memory is spurious. Continental is the only case which seems to have real long memory. It is quiet natural if we look at the d values under several aggregation levels, they are very close to each other. For this, we are only able to reject the null of long memory by 10% level of significance. Now we analyze the results for squared returns, which are given in the following table

Table 5.9: Test for squared returns

stock	m							$\|\lambda_n\|$
	1	2	3	4	6	8	12	
Allianz	0.1470	0.1713	0.2020	0.2244	0.2467	0.2675	0.2763	1.6631*
BASF	0.2378	0.2673	0.2783	0.2629	0.2319	0.2389	0.2739	0.8362**
BAYER	0.1422	0.1501	0.1912	0.2911	0.3087	0.2995	0.2601	2.4222*
BMW	0.1994	0.2486	0.3128	0.3268	0.3227	0.3290	0.3191	2.3460*
Commerz Bank	0.2385	0.3029	0.3362	0.3622	0.3646	0.3801	0.4076	2.1151*
Continental	0.2028	0.2290	0.2615	0.2646	0.2555	0.2762	0.3037	1.2861*
Deutsche Bank	0.2326	0.3109	0.3698	0.3631	0.3505	0.3281	0.3399	2.8387*
Siemens	0.2469	0.2842	0.3215	0.3812	0.4020	0.4202	0.4285	2.2959*
VW	0.1757	0.2454	0.2724	0.3049	0.2987	0.2991	0.2968	2.2086*

Note: the asterisks * and ** refer to the significance levels of 5% and 10% respectively.

In line with the result for absolute returns, the test rejects the null of real long memory. By 5% level of significance, it fails to reject the null only for BASF case. Therefore, we may say that long memory observed in most of the German stock returns is a spurious process, both in absolute and squared returns. The existence of this spurious process could be the result of non-stationarity, regime switching, mean shift, aggregation, etc. These results thus give new evidence about the behavior of German stock returns dealing with long memory.

5.5 *Conclusion*

This paper contributes to the literature on spurious long memory tests by providing a simple procedure to detect the spurious long memory based on the invariance principle of the estimated memory parameter under several aggregation levels. The test performs well in finite sample size. The empirical application gives evidence of spurious long memory in the absolute and squared German stock returns.

Appendix

This session gives the proof of theorem 1. We start the proof by showing that the following holds

$$\mathcal{Q}(n) - \sigma W(n) = O(n^{1/2-\varepsilon}) \ \ a.s \qquad << A1 >>$$

where $\{W(t), 0 \le t < \infty\}$ is a Wiener process and $\varepsilon > 0$. By theorem 1.1 of Horváth and Shao (1999), condition $<< A1 >>$ is satisfied if we can show that there exists $\varsigma > 0$, $\tau > 0$, $\vartheta > 0$ satisfying $\varsigma + \tau > 1/2$ and $\vartheta + 2\varsigma > 1$, such that

$$(i).\ a(k) = O(|k|^{-\frac{1}{2}-\varsigma}),\ \ (ii).\ b(k) = O(|k|^{-\frac{1}{2}-\vartheta}),\ \ (iii).\ c(k) = O(|k|^{-\frac{1}{2}-\tau}) \quad (5.10)$$

where $c(k) = b(o)a(k) + 2\sum_{j=1}^{\infty} b(j)a(k-j)$.

Suppose that the original series x_t has the following infinite moving average representation:

$$x_t = \sum_{i=1}^{\infty} \alpha_i \epsilon_{t-i} \qquad (5.11)$$

where ϵ_t is mean zero independent, identical distributed random variable and having variance σ_ϵ^2. Now, equation (5.2) can be written as

$$y_t = \sum_{j=0}^{m-1} B^j \sum_{i=1}^{\infty} \alpha_i \epsilon_{t-i} \qquad (5.12)$$

$$= \sum_{i=0}^{\infty} a_i \epsilon_{t-i} \qquad (5.13)$$

where $a_i = \displaystyle\sum_{j=i-m+1}^{i} \alpha_j$ and $\alpha_j = 0$ for $j < 0$. Before we proceed (i), we first need to show that a_i converges in mean square. This condition has been previously examined by Teles et al. (1999). Nevertheless, let us describe it in brief here since it is very important for the test.

Let us define

$$(1+B)^d = \sum_{j=0}^{\infty} \varphi_j B^j \tag{5.14}$$

where $\varphi_j = \begin{pmatrix} d \\ j \end{pmatrix} = \dfrac{\Gamma(d+1)}{\Gamma(j+1)\Gamma(d-j+1)}$ and satisfies $\varphi_j \sim \dfrac{\Gamma(d+1)}{2\pi}(-1)^{j-d-1/2} j^{-(d+1)}$.
From the definition of aggregated long memory y_t in (2), we have

$$\begin{aligned}(1+B+...+B^{m-1})^d &= \prod_{j=1}^{m-1}(1+\zeta_j B)^d \quad (5.15)\\ &= \prod_{j=1}^{m-1}\left[\sum_{k=0}^{\infty}\varphi_k \zeta_j^k B^k\right] \quad (5.16)\end{aligned}$$

therefore, for $d > -0.5$, $\prod_{j=1}^{m-1}\left[\sum_{k=0}^{\infty}|\varphi_k \zeta_j^k|^2\right] < \infty$ and this implies that $\sum_{i=0}^{\infty} a_i^2 < \infty$, which is the basic condition allowing the statistic using aggregated long memory. Moreover, from equation (5.6), it implies

$$a_k \sim L(k)k^{2d-1} \tag{5.17}$$

as $k \to \infty$ for some L slowly varying at infinity.

To examine (ii), let us define $b(k) = \frac{1}{4\pi^2}\int_{\pi}^{-\pi} e^{ik\lambda} f^{-1}(\lambda, d)d\lambda$ and assume that $f(\lambda, d)$ and $f^{-1}(\lambda, d)$ are continuous at all λ and d (Tsay and Chan (2005)) such that

$$\frac{\partial f^{-1}(\lambda, d)}{\partial d} = O(|\lambda|^{-2d}) \approx O(|\lambda|^{-2d}) \tag{5.18}$$

Recall the covariance of y as follows

$$\mathbf{E} y_j y_k = \sigma_x^2 r(j-k) = \sigma_x^2 \int_{-\pi}^{\pi} e^{i(j-k)\lambda} d\lambda. \tag{5.19}$$

Define a Toeplitz matrix $\mathcal{R}_{n\mathbf{x}n}$ with the j,k-th entry $r(j-k)$ and a matrix $\mathcal{A}_{n\mathbf{x}n}$ with the j,k-th entry $b(j-k)$. Then, by assumption 1 and Parsevals relation, $\mathcal{A}_{n\mathbf{x}n}$ can be defined as an inverse of the covariance matrix $\mathcal{R}_{n\mathbf{x}n}$ (Fox and Taqqu (1986), Bleher (1981))

$$\mathcal{R}\left(\frac{1}{4\pi^2}f^{-1}(\lambda,d)\right). \tag{5.20}$$

By this relation, we intend to get the asymptotic of $b(k)$. Furthermore, by proposition 1 of Souza (2008), the autocovariance of y_t is given by

$$\gamma_y(k) \sim m^2\sigma_x^2 C_\rho(k)k^{2d-1} + O(k^{2d-3}), \quad \textbf{as} \;\; k \to \infty \tag{5.21}$$

From this, it is sufficient to show that

$$r(k) \sim L(\lambda)|k|^{2d-1}, \quad \textbf{as} \;\; k \to \infty \tag{5.22}$$

and therefore for $0 < \delta < 1/2 - d$

$$|b(k)| = O(|k|^{\delta-1}), \quad \textbf{as} \;\; k \to \infty \tag{5.23}$$

Further details about the autocovariance function of y_t, the readers are referred to Souza (2008).

From (5.18) and (5.24), it is sufficient to have as $n \to \infty$,

$$\begin{aligned} |c(k)| &= O(L(k)k^{2d-1} + O(L(k)k^{2d-1+\delta})\beta(\delta,d)) \quad (5.24)\\ &= O(L(k)k^{2d-1+\delta}) \quad (5.25)\end{aligned}$$

where $\beta(\delta,d)$ is beta function defined as $\beta(\delta,d) = \int_0^1 y^{\delta-1}(1-y)^{2d-1}dy$.

Now, the condition $<< A1 >>$ is satisfied and we can define a sequence of Brownian bridges $V_n(t), 0 \le t \le 1$ such that

$$\max_{0\le s_1,s_2\le 1} T^{1/2}s_1s_2\left|\left\{\frac{\mathcal{Q}(n_1,d)}{n_1} - \frac{\mathcal{Q}(n_2,d)}{n_2}\right\}\right| \xrightarrow{D} \sup_{0\le t\le 1}\sigma|V(t)| \tag{5.26}$$

and

$$\sup_{0\le s_1,s_2\le 1}|T^{1/2}s_1s_2\{\hat{d}^{(m_1)} - \hat{d}^{(m_2)}\} - \sigma V_n(t)| = O_p(T^{-1/2}) \tag{5.27}$$

with $s_1 = \dfrac{n_1}{n_1+n_2}$, $s_2 = \dfrac{n_2}{n_1+n_2}$ and $T = n_1 + n_2$ and theorem 1 is proved.

REFERENCES

Agiakloglou, C., Newbold, P., and Wohar, M., (1993), "Bias in an estimator of the fractional difference parameter", *Journal of Time Series Analysis*, 14, 235-46.

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

Andersson, M. K., Eklund, B. and Lyhagen, J., (1999), "A simple linear time series model with misleading nonlinear properties", *Economics Letters*, 65(3), 281-284.

Andrews, D. W.K., (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation", *Econometrica*, 59, 817-858.

Andrews, D. W. K., (1998), "Hypothesis testing with a restricted parameter space", *Journal of Econometrics*, 84(1), 155-199.

Andrews, D. W. K. and Ploberger, W., (1994), "Optimal test when nuisance parameteris present only under the alternative", *Econometrica*, 62, 1383-1414.

Bahmani-Oskooee, M., Kutan, A. M., and Zhou, S., (2008), "Do real exchange rates follow a nonlinear mean reverting process in developing countries?", *Southern Economic Journal*, 74(4), 1049-1062.

Baillie, R. T. and Kapetanios, G., (2007), "Testing for neglected nonlinearity in long memory models", *Journal of Business and Economic Statistics*, 25(4), 447-461(15).

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

Beran, J., (1994), "Statistics for long memory processes", Chapman & Hall, New

York.

Beran, J. and Terrin, N., (1996), "Testing for a change of the long memory parameter", *Biometrika*, 83(3), 627-638.

Bleher, M., (1981), "Inversion of Teoplz matrices", *Trans. Moscow Math. Soc.*, Issue 2, 201-229.

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

Caceres, C. and Nielsen, B., (2007), "Convergence to stochastic integrals with non-linear integrands", Nuffield College Economics Working Paper.

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

Cerrato, M., and Sarantis, N., (2006), "Nonlinear mean reversion in real exchange rates: evidence from developing and emerging market economies", *Economics Bulletin*, 6(7), 1-14.

Chambers, M.J., (1998), "Long memory and aggregation in macroeconomic time series", *International Economic Review*, 39(4), 1053-1072.

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

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

Chan, N. H. and Wei, C. Z., (1988), "Limiting distribution of least square estimates of unstable autoregressive processes", *The Annals of Statistics*, 16(1), 367-401.

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

Cheung, Y. W., (1993), "Long memory in foreign-exchange rates", *Journal of Business and Economic Statistics*, 11(1), 93-101.

Cheung, Y.-W. and Lai, K. S., (2001), "Long memory and nonlinear mean reversion in Japanese Yen-based real exchange rates", *Journal of International Money and Finance*, 20, 115-132.

Choi, 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.

Chortareas, G. and Kapetanios, G., (2004), "The Yen real exchange rate may be stationary after all: evidence from non-linear unit-root tests", *Oxford Bulletin of Economics and Statistics*, 66(1), 113-131.

Davidson, J. , (1994), "Stochastic limit theory", Oxford University Press, Oxford.

Davidson, J. , (2002), "Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes", *Journal of Econometrics*, 106, 243-269.

Davidson, J. and de Jong, R, M., (2000), "The functional central limit theorem and week convergence to stochastic integral II, fractional integrated process", *Econometric Theory*, 16, 643-666.

Davidson, J. and Sibbertsen, P., (2005), "Generating schemes for long memory processes: regimes, aggregation and linearity", *Journal of Econometrics*, 128, 253-282.

Davidson, J. and Sibbertsen, P., (2009), "Test of bias in log-periodogram regression", *Economics Letters*, 102, 83-86

Davies, R., (1987), "Hypothesis testing when a nuisance parameter is present only under the alternative", *Biometrika*, 74(1), 33-43.

de Jong, R. M., and Davidson, J., (2000), "The functional central limit theorem and week convergence to stochastic integral I, The weakly dependent processes", *Econometric Theory*, 16.

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

Diebold, F. X. and Inoue, C. A., (2001), "Long memory and regime switching", *Journal of Econometrics*, 105(1), 131-159.

Diebold, F. X., Husted, S. and Rush, M., (1991), "Real exchange rates under the gold standard", *Journal of Political Economy*, 99(6), 1252-1271.

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

Engle, R. and Smith, A., (1999), "Stochastic permanent breaks", *Review of Economics and Statistics*, 81, 553-574.

Fox, R. and Taqqu, M.S., (1987), "Central limit theorem for quadratic forms in random variables having long-range dependence", *Probab. Th. Rel. Fields*, 74, 213-240.

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.

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.

Goldman, E. and Tsurumi, H., (2006), "Bayesian comparison of long memory and threshold nonlinearity in time series models", *Mimeo.*

Gourieroux, C., (1997), "ARCH models and financial applications", Springer, New York.

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

Granger, C. W. J. and Ding, Z., (1996), "Varieties of long memory models", *Journal of Econometrics*, 73(1), 61-77.

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

Granger, C. W. J. and Joyeux, R., (1980), "An introduction to long mmeory time series and fractionally differencing", *Journal of Time Series Analysis*, 1, 15-29.

Granger, C. W. J. and Teräsvirta, T., (1993), "Modelling nonlinear economic relationships", Oxford University Press, Oxford.

Granger, C. W. J. and Teräsvirta, T., (1999), "A simple nonlinear time series model with misleading linear properties", *Economics Letters*, 62(2), 161-165.

Grugul, H. and Wojtowicz, T., (2006), "Long memory on the German stock exchange", *Czech Journal of Economics and Finance*, 56(09-10), 447-468.

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

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

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

Haldrup, N. and Nielsen, M. O., (2006), "A regime switching long memory model for electricity prices", *Journal of Econometrics*, 135(1-2), 349-376.

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

Harris, D., McCabe, B. P. M. and Leybourne, S. J., (2008), "Testing for long memory", *Econometric Theory*, 24, 143-175.

Hassler, U. and Olivares, M., (2007), "Long memory and structural change: new evidence from German stock market returns", Working Paper.

Henry, O.T., (2002), "Long memory in stock returns: some international evidence", *Applied Financial Economics*, 12, 725-729.

Hiemstra, C. and Jones, J.D., (1997), "Another look at long memory in common stock returns", *Journal of Empirical Finance*, 4, 373-401.

Horváth, L. and Shao, Qi-M., (1999), "Limit theorems for quadratic forms with applications to Whittle's estimate", *The Annals of Applied Probability*, 9(1), 146-187.

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

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.

Kapetanios, G. and Shin, Y., (2003), "Testing for nonstationary long memory against nonlinear ergodic models", Univ. of London Queen Mary Economics Working Paper No. 500.

Kapetanios, G., Shin, Y., Snell, A., (2003), "Testing for a unit root in the nonlinear STAR framework", *Journal of Econometrics*, 112(2), 359-379.

Kilic, R., (2003), "A testing procedure for a linear unit root against ESTAR process", *Oxford Bulletin of Economics and Statistics*, revise and re-submit.

Kurtz, T. G. and Protter, P., (1991), "Weak limit theorem for stochastic integrals and stochastic differential equations", *Annals of Probability*, 19(3), 1035-1070.

Lee, T. and Lee, S., (2007), "Test for parameter change in linear processes based on Whittle's estimator", *Communication in Statistics-Theory and Methods*, 36, 2129-2141.

Lee, T. H., White, H., and Granger, C. J. W., (1993), "Testing for neglected nonlinearity in time series models: a comparison of neural network methods and alternative tests", *Journal of Econometrics*, 56, 269-290.

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

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

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

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

Lobato, I.N and Savin, N., (1998), "Real and spurious long-memory properties of stock-market data", *Journal of Business and Economic Statistics*, 16, 261-268.

Lobato, I.N and Robinson, P.M., (2003), "A nonparametric test for I(0)", *Review of Economic Studies*, 65(3), 475-495.

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

Man, K.S. and Tiao, G.C., (2006), "Aggregation effect and forecasting temporal aggregates of long memory processes", *International Journal of Forecasting*, 22(2), 267-281.

Mandelbrot, B. B. and van Ness, J. W., (1968), "Fractional Brownian motions, fractional noises and application", *SIAM Review*, 10(4), 422-437.

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.

Marinucci, D. and Robinson, P. M., (1999), "Alternative forms of fractional Brownian motion", *Journal of Statistical Planning and Inference*, 80, 111-122.

Ohanissian, A., Russell, J.R. and Tsay, R.S., (2008), "True or spurious long memory? a new test", *Journal of Business and Economic Statistics*, 26(2), 161-175.

Park, J. Y. and Shintani, M., (2005), "Testing for unit root against transitional autoregressive models", Department of Economics, Vanderbilt University Working Papers No. 05010.

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

Petrucelli, J.D. and Davies, N., (1986), "A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series", *Biometrika*, 73(3), 687-694.

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

Robinson, P.M., (1995), "Gaussian semiparametric estimation of long range dependence", *Annals of Statistics*, 23, 1630-1661.

Rothe, C. and Sibbertsen, P., (2006), "Phillips-Perron-type unit root tests in the nonlinear ESTAR framework," *AStA Advances in Statistical Analysis*, 90(3), 439-456.

Saikkonen, P. and Luukkonen, R., (1988), "Lagrange multiplier tests for testing nonlinearities in time series models", *Scandinavian Journal of Statistics*, 15, 55-68.

Sarno, L. and Taylor, M. P., (2001a), "Official intervention in the foreign exchange market: Is it effective and, if so, how does it work?", *Journal of Economic Literature*, 39, 839-68.

Sarno, L. and Taylor, M. P., (2001b), "The microstructure of the foreign exchange market: A selective survey of the literature", Princeton, NJ: Princeton University.

Seo, B., (2004) "Testing for nonlinear adjustment in smooth transition vector correction models", Econometric Society 2004 Far Eastern Meetings No. 749.

Sibbertsen, P., (2004a), "Long-memory in volatilities of German stock returns", *Empirical Economics*, 29, 477-488.

Sibbertsen, P., (2004b), Long-memory versus structural change: an overview, *Statistical Papers*, 45, 465-515.

Smallwood, A. D., (2005) "Joint test for non-linearity and long memory: the case of purchasing power parity", *Studies in Nonlinear Dynamics and Econometrics*, 9(2), article 7.

Smith, A., (2005), "Level shift and the illusion of long memory in economic time series", *Journal of Business and Economic Statistics*, 23(3), 321-335.

Souza, L.R., (2007), "Temporal aggregation and bandwidth selection in estimating long memory", *Journal of Time Series Analysis*, 28(5), 701-722.

Souza, L.R., (2008), "Spectral properties of temporally aggregated long memory processes", *Brazilian Journal of Probability and Statistics*, 22(2), 135-155.

Taylor, M. P., (2003), "Purchasing power parity", *Review of International Economics*, 11, 436-52.

Taylor, M. P., Peel, D. A. and Sarno, L., (2001), "Nonlinear mean-reversion in real exchange rates: toward a solution to the purchasing power parity puzzles", *International Economic Review*, 42(4), 1015-1042.

Teles, P., Wei, W.W.S. and Crato, N., (1999), "The use of aggregate series in testing for long memory", *Bulletin of International Statistical Institute*, 22nd Session Book 3, 341-342.

Teles, P., Wei, W.W.S. and Hodges, E.M., (2008), "Testing unit root based on aggregate time series", *Communication in Statistics-Theory and Methods*, 37, 365-590.

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., (1993), "Power of the neural network linearity test", *Journal of Time Series Analysis*, 14(2), 209-220.

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

Tolvi, J., (2003), "Long memory in a small stock market", *Economics Bulletin*, 7(3), 1-13.

Tsai, H. and Chan, K.S., (2005), "Quasi maximum likelihood estimation for a class of continuous time-long memory processes", *Journal of Time Series Analysis*, 26(5), 691-713.

Tsay, W. J. and Härdle, W., (2008), "A generalized ARFIMA process with Markov-switching fractional differencing parameter", *Journal of Statistical Computation and Simulation*, forthcoming.

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.

van Dijk, D., Franses, P. H., and Paap, R., (2002) "A nonlinear long memory model with an application to US unemployment", *Journal of Econometrics*, 110, 135-

165.

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

White, H., (1982), "Maximum likelihood estimation of misspecified models", *Econometrica*, 50, 1-26.

White, H., (1984), "Asymptotic theory for econometricians", Academic Press Inc.

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