Testing for Long Memory Against ESTAR Nonlinearities

Heri Kuswanto and Philipp Sibbertsen[1]

Institute of Statistics, Leibniz Universität Hannover

Abstract

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.

Keywords: directed-Wald test, ESTAR, long memory

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

[1]Corresponding Author:
Institute of Statistics
Leibniz Universität Hannover
Königsworther Platz 1
30167 Hannover
Germany
Email: Sibbertsen@statistik.uni-hannover.de

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. (2002) 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 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 2 introduces the theoretical framework. In section 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, and section 5 illustrates an empirical application to exchange rates. Section 6 concludes and all proofs are given in the appendix.

2 The Model

In this section we introduce the considered processes, namely long memory and nonlinear 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.

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:

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

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 u_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). Under (1), y_t is a long memory process with a certain degree of fractional integration d and the fractional difference operator is defined by

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{d\Gamma(j+d)}{\Gamma(1+d)\Gamma(j+1)}. \tag{2}$$

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 (1) can be written as an infinite moving average process in terms of u_t

$$y_t = \sum_{j=0}^{\infty} a_j u_{t-j}, \tag{3}$$

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

$$y_t - \sum_{j=1}^{\infty} b_j y_{t-j} = u_t. \tag{4}$$

By defining

$$z_t = \sum_{j=1}^{t} b_j y_{t-j}, \tag{5}$$

with $b_j = -\frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}$, we can write term (4) as

$$y_t = u_t - z_t. \tag{6}$$

Following Kapetanios and Shin (2003), we use (6) to derive the test statistic in the next section. Consider the scaled partial sum process

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

where u_t is defined by (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 $\text{var}(S_T) \sim cT^{2(d+1/2)}$ and $Y_d(r)$ is a fractional Brownian motion with $d \in (0, 1/2)$ or $d \in (1/2, 1)$ 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.

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:

$$y_t = \alpha_1 y_{t-1}\{1 - \exp(-\gamma y_{t-l}^2)\} + \alpha_2 y_{t-1} + \epsilon_t, \tag{7}$$

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].

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 the 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:

$$u_t \quad = \quad \alpha_1 F(y_{t-1}) + \alpha_2 y_{t-1} + \alpha_3 z_t + \epsilon_t, \tag{8}$$

$t = 1, ..., T$ with u_t and z_t are defined as in the previous section. The condition for the error ϵ_t is formalized in the following assumption

Assumption 1 :
Let ϵ_t be iid 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 (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

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

expansion to the transition function of the ESTAR model given in section 2.2 yields

$$u_t = \alpha_1 y_{t-1}^3 + \alpha_2 y_{t-1} + \alpha_3 z_t + \epsilon_t. \tag{9}$$

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

$$u_t = \alpha_1 y_{t-1}\{1 - \exp(-\gamma y_{t-1}^2)\} + \alpha_2 y_{t-1} + \alpha_3 z_t + \epsilon_t. \tag{10}$$

We test whether y_t is long memory under the null against the alternative of an ESTAR process using the models (9) and (10). Our null hypothesis is:

$$\mathbf{H_0} : \alpha_1 = \alpha_2 = \alpha_3 = 0 \tag{11}$$

and is tested against the alternative of:

$$\mathbf{H_1} : \alpha_1 \neq 0, \alpha_2 \neq 0, \alpha_3 = 1 \tag{12}$$

Under the null, we can also write

$$u_t = \epsilon_t, \tag{13}$$

which means that y_t is long memory with given d. Based on (4) and (6) where $u_t - z_t = y_t$, then under the alternative hypothesis we have the ESTAR model

$$y_t = \alpha_1 F(y_{t-1}) + \alpha_2 y_{t-1} + \epsilon_t \tag{14}$$

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

3.1 Test Statistic and Limit Distribution

We propose a standard Wald test to test the null (11). We know that model (10) with a given γ is linear in the parameter $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ and so is the model (9). Therefore, we can estimate the parameter α by OLS and obtain the least squares estimator

$$\hat{\alpha} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}(\mathbf{X}'\mathbf{U}) \tag{15}$$

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

$$W \quad = \quad \hat{\alpha}'[Var(\hat{\alpha})]^{-1}\hat{\alpha}. \tag{16}$$

The Wald test above is normally used under the condition that the parameter α is unrestricted under the alternative hypothesis. Since the parameter α_3 in (12) is restricted ($\alpha_3 = 1$), the classical Wald test is no longer an appropriate 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 $\mathbf{H_0} : \alpha_3 = 0$ and $\mathbf{H_1} : \alpha_3 = A$, with $A = 1$. Then, the directed-Wald statistic, DW is given by:

$$DW_{(c)} \quad = \quad (1+c)^{(-1/2)} exp\left(\frac{1}{2}\frac{c}{1+c}W\right)\mathbf{\Phi}(A, \frac{c}{1+c}\hat{\alpha_3}, \frac{c}{1+c}Var(\hat{\alpha_3})) \tag{17}$$

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, (17) reduces to:

$$DW_{(\infty)} \quad = \quad W + 2\log[\mathbf{\Phi}(A, \hat{\alpha_3}, Var(\hat{\alpha_3}))]. \tag{18}$$

Here, $\mathbf{\Phi}(.)$ is 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 ∞. Note that $\mathbf{\Phi}(.)$ is not necessary to be zero because of the unrestricted parameters α_1 and α_2 under the alternative.

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).

3.2 First Approach

In this section, we derive the limit distribution of the statistics (16) and (18), which is mainly characterized by the approximation y^3_{t-1} 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 y_t is long memory, $\hat{\alpha}_{\mathbf{OLS}}$ is a consistent estimator of α and converges to its true value with the rate of convergence diag$(T^{3(1/2+d)}, T^{(1/2+d)}, T^{(1/2+d)})$ when $0 < d < 0.5$ and diag$(T^{3(1-d)}, T^{(1-d)}T^{(1-d)})$ when $0.5 < d < 1$. Its asymptotic distribution is

$$\begin{bmatrix} T^{3(1/2+d)} & 0 & 0 \\ 0 & T^{(1/2+d)} & 0 \\ 0 & 0 & T^{(1/2+d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \\ \hat{\alpha_3} \end{bmatrix} \Rightarrow \mathbf{Q_1^{-1}Q_2} \quad \textbf{if } 0 \leq d < 0.5 \tag{19}$$

$$\begin{bmatrix} T^{3(1-d)} & 0 & 0 \\ 0 & T^{(1-d)} & 0 \\ 0 & 0 & T^{(1-d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \\ \hat{\alpha_3} \end{bmatrix} \Rightarrow \mathbf{Q_1^{-1}Q_2} \quad \textbf{if } 0.5 < d < 1 \tag{20}$$

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

$$\mathbf{Q_1} = \begin{bmatrix} \int_0^1 Y_d(r)^6 dr & \int_0^1 Y_d(r)^4 dr & \int_0^1 Y_d(r)^3 dZ_d(r) \\ \int_0^1 Y_d(r)^4 dr & \int_0^1 Y_d(r)^2 dr & \int_0^1 Y_d(r) dZ_d(r) \\ \int_0^1 Y_d(r)^3 dZ_d(r) & \int_0^1 Y_d(r) dZ_d(r) & \int_0^1 Z_d(r)^2 dr \end{bmatrix} \tag{21}$$

$$\mathbf{Q_2} = \begin{bmatrix} \int_0^1 Y_d(r)^3 dY(r) \\ \int_0^1 Y_d(r) dY(r) \\ \int_0^1 Z_d(r) dY(r) \end{bmatrix}, \tag{22}$$

where $Y_d(r)$ *and* $Y(r)$ *are fractional Brownian motion and standard Brownian motion respectively, 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. 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 y_t *is long memory, the limit distribution of the Wald statistic is*

$$W \Rightarrow \mathcal{W} \equiv \mathbf{Q_2'Q_1^{-1}Q_2} \tag{23}$$

and the limit distribution of the directed-Wald statistic with $\alpha_3 = A$ *under the alternative is given* *by*

$$DW \Rightarrow [\mathcal{W} + 2\log \mathbf{\Phi}(A, \mathcal{B}, \mathcal{V})] \tag{24}$$

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

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

$$DW = \sup_{\gamma \in \Gamma} DW_\gamma = \sup_{\gamma \in \Gamma} \left\{ \frac{1}{\hat{\sigma_\epsilon}^2} \{\hat{\alpha}'(X'X)\hat{\alpha}\} + 2\log[\mathbf{\Phi}(A, \hat{\alpha_3}, Var(\hat{\alpha_3}))] \right\}, \tag{25}$$

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

$$\begin{bmatrix} T^{(1/2+d)} & 0 & 0 \\ 0 & T^{(1/2+d)} & 0 \\ 0 & 0 & T^{(1/2+d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \\ \hat{\alpha_3} \end{bmatrix} \Rightarrow \mathbf{Q_1}(\gamma)^{\mathbf{-1}}\mathbf{Q_2}(\gamma) \quad \textbf{if } 0 \leq d < 0.5 \tag{26}$$

$$\begin{bmatrix} T^{(1-d)} & 0 & 0 \\ 0 & T^{(1-d)} & 0 \\ 0 & 0 & T^{(1-d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \\ \hat{\alpha_3} \end{bmatrix} \Rightarrow \mathbf{Q_1}(\gamma)^{\mathbf{-1}}\mathbf{Q_2}(\gamma) \quad \textbf{if } 0.5 < d < 1 \tag{27}$$

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

$$\mathbf{Q_1}(\gamma) = \begin{bmatrix} (1-2\mu_\gamma+\psi_\gamma)\int_0^1 Y_d(r)^2dr & (1-\mu_\gamma)\int_0^1 Y_d(r)^2d(r) & (1-\mu_\gamma)\int_0^1 Y_d(r)dZ_d(r) \\ (1-\mu_\gamma)\int_0^1 Y_d(r)^2dr & \int_0^1 Y_d(r)^2dr & \int_0^1 Y_d(r)dZ_d(r) \\ (1-\mu_\gamma)\int_0^1 Y_d(r)dZ_d(r) & \int_0^1 Y_d(r)dZ_d(r) & \int_0^1 Z_d(r)^2dr \end{bmatrix} \tag{28}$$

$$\mathbf{Q_2}(\gamma) = \begin{bmatrix} (1-\mu_\gamma)\int_0^1 Y_d(r)dY(r) \\ \int_0^1 Y_d(r)dY(r) \\ \int_0^1 Z_d(r)dY(r) \end{bmatrix}, \tag{29}$$

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, $Y_d(r)$, $Y(r)$ and $Z_d(r)$ are defined as in theorem 1.

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 y_t is long memory, we have for the Wald statistic

$$W_\gamma \Rightarrow \mathcal{W}_\gamma \equiv \mathbf{Q_2}(\gamma)'\mathbf{Q_1}(\gamma)^{\mathbf{-1}}\mathbf{Q_2}(\gamma) \tag{30}$$

and for the sup-directed-Wald statsitic for $\alpha_3 = A$ we have

$$\sup_{\gamma\in\Gamma} DW_\gamma \Rightarrow \sup_{\gamma\in\Gamma} [\mathcal{W}_\gamma + 2\log \mathbf{\Phi}(A, \mathcal{B}, \mathcal{V})] \tag{31}$$

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

$$\limsup_{T\to\infty} P_r \left[\sup_{\gamma\in\Gamma} \sup_{\gamma':|\gamma-\gamma'|<\delta} |W_\gamma^{(i)} - W_{\gamma'}^{(i)}| \geq \epsilon\right] < \epsilon. \tag{32}$$

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 (31), theorem 5 implies the stochastic equicontinuity of the directed-Wald test.

To apply our test in practice, the parameter d has to be estimated from the series in order to use the correct critical values. This can be done by using any consistent estimator such as the GPH-estimator of Geweke and Porter-Hudak (1983). This, however, does not change the size and power results given in the next section significantly.

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 1 and 2.

Table 1: Critical values of the test for the first approach

Sign. Level	d=0.1	d=0.2	d=0.3	d=0.4
10%	5.0394	5.1238	5.1987	5.2605
5%	6.3948	6.4467	6.4516	6.6240
1%	9.7008	9.8097	10.0080	10.0369
Sign. Level	d=0.6	d=0.7	d=0.8	d=0.9
10%	5.4100	6.0787	6.7656	7.2822
5%	6.8649	7.6995	8.3055	8.8008
1%	10.3098	11.3106	12.0454	12.4371

Table 2: Critical values of the test for the second approach

Sign. Level	d=0.1	d=0.2	d=0.3	d=0.4
10%	5.6884	5.7290	6.0615	6.2647
5%	7.3199	7.4107	7.4391	7.6236
1%	10.6943	10.8046	10.9267	11.7017
Sign. Level	d=0.6	d=0.7	d=0.8	d=0.9
10%	6.7130	7.0191	7.4604	7.9760
5%	8.0670	8.4256	9.0551	10.1568
1%	12.7475	13.4372	14.1835	14.8061

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 3 and 4 contain the results of the size experiment.

Table 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	0.076	0.060	0.069	0.060
	250	0.059	0.056	0.051	0.056
10%	100	0.134	0.116	0.112	0.113
	250	0.108	0.101	0.098	0.092
Sign. level	T	d=0.6	d=0.7	d=0.8	d=0.9
5%	100	0.049	0.049	0.050	0.047
	250	0.047	0.043	0.048	0.041
10%	100	0.094	0.100	0.093	0.096
	250	0.088	0.095	0.090	0.094

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

Table 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	0.072	0.070	0.068	0.065
	250	0.066	0.062	0.057	0.059
10%	100	0.133	0.129	0.133	0.122
	250	0.117	0.119	0.120	0.116
Sign. level	T	d=0.6	d=0.7	d=0.8	d=0.9
5%	100	0.059	0.053	0.048	0.050
	250	0.051	0.048	0.047	0.049
10%	100	0.120	0.124	0.098	0.011
	250	0.099	0.0101	0.095	0.098

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

The rejection rates in table 3 and 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

$$y_t = \alpha_1 y_{t-1}\{1 - \exp(-\gamma y_{t-1}^2)\} + \alpha_2 y_{t-1} + \epsilon_t \tag{33}$$

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 (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

[3] These parameter setting was extensively examined for example in Rothe and Sibberten (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 provides the estimated long memory parameter for our DGPs.

Table 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 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. The power experiments are evaluated under the 5% significant level.

Table 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	0.908	0.912	0.977	0.967	0.994	0.987	0.949	0.997
	0.05	0.960	0.969	0.995	0.998	0.967	0.999	0.733	0.896
	0.1	0.985	0.988	0.978	0.999	0.871	0.998	0.714	0.888
-1	0.01	0.897	0.898	0.963	0.967	0.983	0.994	0.979	0.994
	0.05	0.944	0.954	0.988	0.992	0.988	0.997	0.818	0.996
	0.1	0.970	0.981	0.995	0.995	0.937	0.999	0.614	0.980
-0.5	0.01	0.877	0.881	0.946	0.954	0.978	0.986	0.984	0.989
	0.05	0.913	0.920	0.974	0.983	0.990	0.996	0.955	0.998
	0.1	0.938	0.941	0.985	0.991	0.993	0.997	0.851	0.999
-0.1	0.01	0.838	0.865	0.922	0.931	0.966	0.976	0.980	0.992
	0.05	0.860	0.869	0.942	0.945	0.970	0.985	0.987	0.988
	0.1	0.868	0.878	0.950	0.953	0.974	0.989	0.988	0.989

Table 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	0.141	0.467	0.080	0.256	0.159	0.627	0.525	0.994
	0.05	0.489	0.915	0.696	0.996	0.916	0.997	0.995	1.000
	0.1	0.848	0.999	0.960	0.999	0.995	0.998	0.995	1.000
-1	0.01	0.145	0.544	0.043	0.119	0.078	0.308	0.441	0.928
	0.05	0.236	0.628	0.353	0.867	0.659	0.997	0.947	0.995
	0.1	0.466	0.920	0.718	0.998	0.945	0.998	0.990	0.995
-0.5	0.01	0.270	0.790	0.036	0.051	0.025	0.070	0.118	0.490
	0.05	0.096	0.305	0.078	0.242	0.188	0.739	0.539	0.996
	0.1	0.098	0.295	0.151	0.531	0.395	0.979	0.817	0.995
-0.1	0.01	0.556	0.972	0.132	0.192	0.006	0.232	0.039	0.043
	0.05	0.364	0.887	0.029	0.043	0.012	0.014	0.058	0.169
	0.1	0.255	0.768	0.012	0.025	0.013	0.016	0.071	0.243

The power of the test using the first approach can be seen in table 6 and 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 a decreasing sample size.

For non-stationary long memory a tendency can be observed. The power tend to increase with increasing d. However, several parameter settings lead to a very low power, especially for $\alpha_1 = -0.1$ and high d. This is, however, intuitive, since α_1 close to zero means that the process is getting more persistent (near unit root) 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 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	0.002	0.008	0.013	0.019	0.086	0.552	0.440	0.794
	0.05	0.253	0.739	0.579	0.944	0.910	0.932	0.903	0.926
	0.1	0.780	0.999	0.953	0.987	0.966	0.989	0.967	0.989
-1	0.01	0.004	0.008	0.002	0.006	0.031	0.168	0.241	0.736
	0.05	0.041	0.134	0.198	0.688	0.523	0.899	0.841	0.900
	0.1	0.262	0.721	0.543	0.897	0.823	0.900	0.896	0.921
-0.5	0.01	0.041	0.109	0.001	0.005	0.008	0.011	0.072	0.476
	0.05	0.002	0.007	0.015	0.043	0.115	0.597	0.427	0.897
	0.1	0.010	0.063	0.063	0.343	0.344	0.867	0.707	0.900
-0.1	0.01	0.335	0.820	0.025	0.045	0.0006	0.004	0.035	0.143
	0.05	0.117	0.423	0.002	0.006	0.001	0.005	0.009	0.023
	0.1	0.069	0.244	0.0002	0.003	0.004	0.005	0.051	0.221

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

The power in table 8 has a similar tendency as in table 7. However, it is clear that the directed-Wald test outperforms the KS test. The results in table 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 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.

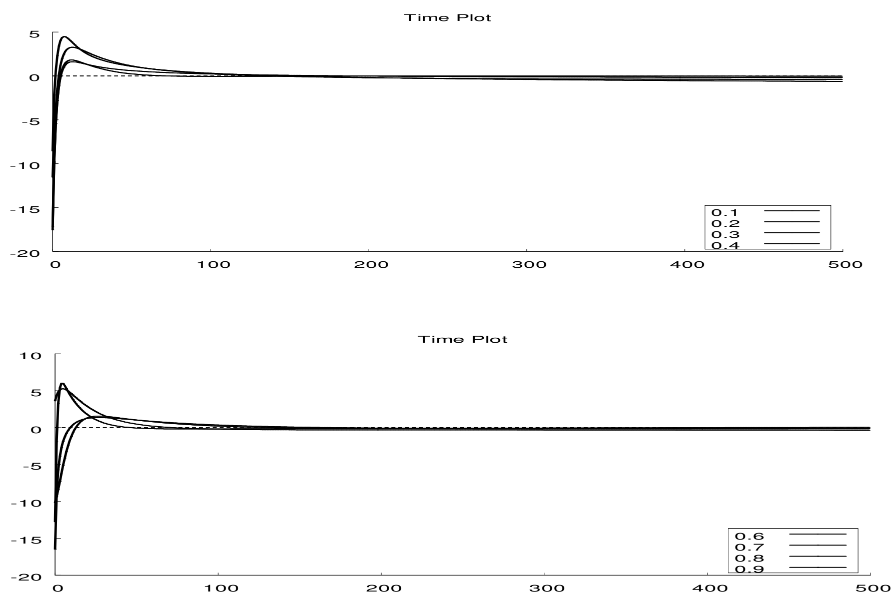

Figure 1: Plot of the test statistic depending on gamma

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 grid point 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 9 and 10 give the results of the power experiment for the second approach.

Table 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	0.998	1.000	0.964	1.000
	0.05	1.000	1.000	1.000	1.000	0.946	0.952	0.863	0.999
	0.1	0.999	1.000	0.979	1.000	0.852	0.876	0.818	0.993
-1	0.01	1.000	1.000	1.000	1.000	1.000	1.000	0.978	1.000
	0.05	1.000	1.000	1.000	1.000	0.997	1.000	0.838	1.000
	0.1	1.000	1.000	0.996	1.000	0.925	1.000	0.754	0.991
-0.5	0.01	1.000	1.000	1.000	1.000	0.998	1.000	0.988	1.000
	0.05	1.000	1.000	1.000	1.000	0.996	1.000	0.961	1.000
	0.1	1.000	1.000	0.999	1.000	0.993	1.000	0.894	0.999
-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	1.000
	0.1	1.000	1.000	1.000	0.999	1.000	1.000	0.998	1.000

Table 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	0.208	0.494	0.088	0.325	0.196	0.680	0.465	0.998
	0.05	0.528	0.993	0.782	0.999	0.916	1.000	0.997	1.000
	0.1	0.909	1.000	0.963	1.000	0.998	1.000	1.000	1.000
-1	0.01	0.216	0.634	0.050	0.206	0.063	0.382	0.330	0.955
	0.05	0.289	0.661	0.479	0.894	0.692	0.999	0.946	1.000
	0.1	0.517	0.993	0.817	1.000	0.987	1.000	0.991	1.000
-0.5	0.01	0.365	0.852	0.045	0.049	0.035	0.065	0.097	0.470
	0.05	0.166	0.401	0.085	0.338	0.180	0.821	0.553	0.998
	0.1	0.189	0.391	0.266	0.657	0.477	0.980	0.857	1.000
-0.1	0.01	0.634	1.000	0.246	0.298	0.033	0.040	0.041	0.052
	0.05	0.481	0.964	0.050	0.068	0.014	0.016	0.067	0.235
	0.1	0.300	0.856	0.029	0.034	0.016	0.019	0.085	0.322

Table 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 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. In general, the results of the power experiments tend to suggest that the power depends on the setting of parameters. In other words, it depends on the degree of the persistency of the process.

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 11, which contains also the results of our test.

Table 11: Empirical application: real exchange rates

RER	$\hat{d}$	Directed-Wald test 1st approach	Directed-Wald test 2nd approach	Neglected nonlin. test ANN	Neglected nonlin. test TLG
Australia	0.851	3.8824	8.3904**	11.4680*	11.6677*
Austria	0.850	2.8133	4.3439	2.3579	2.3553
Denmark	0.903	5.6021	5.6064	0.4437	0.4392
France	0.862	3.2222	4.7931	2.1239	2.1157
Germany	0.942	5.9584	7.1791	1.5767	1.6088
New Zealand	0.892	3.1336	5.9906	2.9766	2.9590
Netherland	0.900	2.9566	3.2666	3.1358	3.0728
US	0.931	5.5059	7.6043	0.5734	1.0531
Korea	0.524	24.1029*	33.8087*	10.6546*	10.5552*
Malaysia	0.856	7.4334**	13.4931*	8.6063*	8.6649*
Indonesia	0.628	28.0787*	39.5530*	11.2909*	11.0694*
Thailand	0.760	10.1556*	9.9240*	11.6930*	11.8100*
Philipina	0.821	1.7232	1.6681	5.4058	5.4071
Srilanka	0.940	11.8133*	11.9054*	4.9319	5.2819
No of rejection		5	6	5	5

Note: The (*) and (**) represent significance under the 5% and 10% level respectively. The critical values for the corresponding d can be obtained by interpolation from table 1 and 2 .

From this table we see that both directed-Wald tests give consistent results, with an exception for Australia where the second approximation can reject the null with 10% level of significance. Under 5% level of significance, the tests reject the null in 4 cases and 5 cases for first and second approach respectively . 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, Sri Lanka and Indonesia with the exception Australia. This finding is consistent with previous studies about real exchange rate behavior. Those found that nonlinear adjustments towards PPP hold more likely in

[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 for reasons of space. They are available from the authors upon request

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

$$\hat{u}_t = (1 - L)^{\hat{d}} y_t.$$

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 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 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.

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).

References

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., (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 & Economic Statistics*, 25(4), 447-461(15).

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

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.

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.

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.

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: Oxford University Press.

Davidson, J. and de Jong, R, M., (2000), "The functional centra 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.

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.

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.

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.

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

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 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: Oxford University Press.

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

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

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

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. 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.

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

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

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

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.

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.

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.

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

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.

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.

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

van Dijk, D., Franses, P. H., and Paap, R., (2002) "A nonlinear longmemory 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.

Appendix

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

Define the model:

$$u_t = \alpha_1 F(y_{t-1}) + \alpha_2 y_{t-1} + \alpha_3 z_t + \epsilon_t \tag{34}$$

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

$$\mathbf{U} = \mathbf{X}'\alpha + \epsilon \tag{35}$$

with $\mathbf{U} = \begin{bmatrix} u_1 \\ \vdots \\ u_t \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} F(y_0) & y_0 & z_1 \\ \vdots & & \\ F(y_{t-1}) & y_{t-1} & z_t \end{bmatrix}$, $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ 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 y_t is a long memory process, we have that

$$\sigma_T^{-1} y_{[Tr]} \Rightarrow Y_d(r), \tag{36}$$

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

$$Y_T(r) = T^{-1/2}\sigma^{-1} \sum_{t=1}^{[Tr]} \epsilon_t, \quad for \ \ 0 \leq r \leq 1, \tag{37}$$

where $Y_T(r)$ in (37) converges to a Brownian motion $Y(r)$, $\sigma^2 = \mathbf{E}\epsilon_t^2$.

From (35), the OLS estimate of α is

$$\hat{\alpha} = (\mathbf{X'X})^{-\mathbf{1}}(\mathbf{X'U}). \tag{38}$$

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

$$\begin{aligned}
\hat{\alpha} - \alpha &= (\mathbf{X'X})^{-\mathbf{1}}(\mathbf{X'}\epsilon) && (39)\\
\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \end{bmatrix} &= \begin{bmatrix} \sum_{t=1}^{T} x_t x_t & \sum_{t=1}^{T} x_t y_{t-1} & \sum_{t=1}^{T} x_t z_t \\ \sum_{t=1}^{T} x_t y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 & \sum_{t=1}^{T} z_t y_{t-1} \\ \sum_{t=1}^{T} x_t z_t & \sum_{t=1}^{T} y_{t-1} z_t & \sum_{t=1}^{T} z_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} x_t \epsilon_t \\ \sum_{t=1}^{T} y_{t-1} \epsilon_t \\ \sum_{t=1}^{T} z_t \epsilon_t \end{bmatrix}. && (40)
\end{aligned}$$

We begin with examining the asymptotic behavior of the terms in (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

$$z_t = \sum_{j=1}^{t} b_j y_{t-j} \tag{41}$$

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

$$\begin{aligned}
\sigma_T^{-1} z_{[Tr]} &= \sum_{i=1}^{T} \sigma_T^{-1} y_{t-i}(\beta_{t-i} - \beta_{t-i-1}) && (42)\\
Z_d(r) &= \int_0^r Y_d(s) d\beta(s - r). && (43)
\end{aligned}$$

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

$$T^{-(1+2d)} \sum_{t=1}^{T} z_t^2 \Rightarrow \sigma^2 \int_0^1 Z_d(r)^2 dr \tag{44}$$

$$T^{-(1+2d)} \sum_{t=1}^{T} y_{t-1} z_t \Rightarrow \sigma^2 \int_0^1 Y_d(r) dZ_d(r) \tag{45}$$

$$T^{-(1/2+d)} \sum_{t=1}^{T} z_t \epsilon_t \Rightarrow \sigma^2 \int_0^1 Z_d(r) dY(r) \tag{46}$$

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

$$T^{-2(1-d)} \sum_{t=1}^{T} z_t^2 \Rightarrow \sigma^2 \int_0^1 Z_d(r)^2 dr \tag{47}$$

$$T^{-2(1-d)} \sum_{t=1}^{T} y_{t-1} z_t \Rightarrow \sigma^2 \int_0^1 Y_d(r) dZ_d(r) \tag{48}$$

$$T^{-(1-d)} \sum_{t=1}^{T} z_t \epsilon_t \Rightarrow \sigma^2 \int_0^1 Z_d(r) dY(r). \tag{49}$$

Proof of Theorem 1

In this section, we derive the asymptotic distributions of the other terms in (40). We consider first the case of stationary long memory. By using the continuous mapping theorem, we have

$$T^{-3(1/2+d)} \sum_{t=1}^{T} y_{t-1}^3 \epsilon_t \Rightarrow \sigma^4 \int_0^1 Y_d(r)^3 dY(r) \tag{50}$$

$$T^{-6(1/2+d)} \sum_{t=1}^{T} y_{t-1}^6 \Rightarrow \sigma^6 \int_0^1 Y_d(r)^6 dr \tag{51}$$

$$T^{-4(1/2+d)} \sum_{t=1}^{T} y_{t-1}^3 z_t \Rightarrow \sigma^4 \int_0^1 Y_d(r)^3 dZ_d(r). \tag{52}$$

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

$$\begin{bmatrix} T^{-6(1/2+d)}\sum_{t=1}^{T} x_t x_t & T^{-4(1/2+d)}\sum_{t=1}^{T} x_t y_{t-1} & T^{-4(1/2+d)}\sum_{t=1}^{T} x_t z_t \\ T^{-4(1/2+d)}\sum_{t=1}^{T} x_t z_t & T^{-2(1/2+d)}\sum_{t=1}^{T} y_{t-1}^2 & T^{-2(1/2+d)}\sum_{t=1}^{T} y_{t-1} z_t \\ T^{-4(1/2+d)}\sum_{t=1}^{T} x_t z_t & {}^{-2(1/2+d)}\sum_{t=1}^{T} y_{t-1} z_t & T^{-(1+2d)}\sum_{t=1}^{T} z_t z_t \end{bmatrix} \Rightarrow \mathcal{A}_1 \mathbf{Q}_1 \mathcal{A}_1 \tag{53}$$

$$\begin{bmatrix} T^{-3(1/2+d)}\sum_{t=1}^{T} x_t \epsilon_t \\ T^{(-(1/2+d))}\sum_{t=1}^{T} y_{t-1} \epsilon_t \\ T^{(-(1/2+d))}\sum_{t=1}^{T} z_t \epsilon_t \end{bmatrix} \Rightarrow \mathcal{A}_2 \mathbf{Q}_2 \tag{54}$$

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

$$\mathbf{Q_1} = \begin{bmatrix} \int_0^1 Y_d(r)^6 dr & \int_0^1 Y_d(r)^4 dr & \int_0^1 Y_d(r)^3 dZ_d(r) \\ \int_0^1 Y_d(r)^4 dr & \int_0^1 Y_d(r)^2 dr & \int_0^1 Y_d(r) dZ_d(r) \\ \int_0^1 Y_d(r)^3 dZ_d(r) & \int_0^1 Y_d(r) dZ_d(r) & \int_0^1 Z_d(r)^2 dr \end{bmatrix} \tag{55}$$

$$\mathbf{Q_2} = \begin{bmatrix} \int_0^1 Y_d(r)^3 dY(r) \\ \int_0^1 Y_d(r) dY(r) \\ \int_0^1 Z_d(r) dY(r) \end{bmatrix}, \tag{56}$$

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

$$\mathcal{A}_1 = \begin{bmatrix} \sigma^3 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & \sigma \end{bmatrix} \tag{57}$$

and

$$\mathcal{A}_2 = \begin{bmatrix} \sigma^4 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}. \tag{58}$$

Then, from (51) and (52), for $0 < d < 0.5$, the rate of convergence of the OLS parameter is

$$\begin{bmatrix} T^{3(1/2+d)} & 0 & 0 \\ 0 & T^{(1/2+d)} & 0 \\ 0 & 0 & T^{(1/2+d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \\ \hat{\alpha_3} \end{bmatrix} \Rightarrow \mathbf{Q_1^{-1}Q_2} \quad \textbf{if } 0 \leq d < 0.5 \tag{59}$$

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

$$\begin{bmatrix} T^{3(1-d)} & 0 & 0 \\ 0 & T^{(1-d)} & 0 \\ 0 & 0 & T^{(1-d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \\ \hat{\alpha_3} \end{bmatrix} \Rightarrow \mathbf{Q_1^{-1}Q_2} \quad \textbf{if } 0.5 < d < 1 \tag{60}$$

Proof of Theorem 2

The Wald test for the null hypothesis $\alpha_1 = \alpha_2 = \alpha_3 = 0$ is

$$W = \frac{1}{\hat{\sigma}^2}(\epsilon'\mathbf{X})(\mathbf{X'X})^{-\mathbf{1}}(\mathbf{X}'\epsilon), \tag{61}$$

where $\hat{\sigma}^2$ is the variance of the OLS estimator of α and $\hat{\sigma}^2 \to_p \boldsymbol{\sigma^2}$. The asymptotic distribution of the Wald statistic (61) follows directly from the results of theorem 1

$$W \Rightarrow \mathcal{W} \equiv \left\{ \mathbf{Q_2' Q_1^{-1} Q_2} \right\}. \tag{62}$$

Let rewrite (34) as

$$\mathbf{U} = \mathbf{x}'\beta + \mathbf{z}'\alpha_{\mathbf{3}} + \epsilon_t \tag{63}$$

where $\mathbf{z} = [z_1, ..., z_t]$, $\mathbf{x} = [(y_0^3, y_0), ..., (y_{t-1}^3, y_{t-1})]'$ and $\beta = (\alpha_1, \alpha_2)$. The directed-Wald statistic with $c = \infty$ is

$$DW \Rightarrow \{W + 2\log[\mathbf{\Phi}(A, \hat{\alpha_3}, Var(\hat{\alpha_3}))]\}. \tag{64}$$

The information matrix of $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ is defined as (see Andrews (1998))

$$\mathbf{I} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{I}_2 \\ \mathbf{I}_2' & \mathbf{I}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{x}'\mathbf{x} & \mathbf{x}'\mathbf{z} \\ \mathbf{z}'\mathbf{x} & \mathbf{z}'\mathbf{z} \end{bmatrix} / \sigma^2. \tag{65}$$

For $\mathbf{M_x} = \mathbf{I_T} - \mathbf{x}(\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x}'$, the estimate of $\hat{\alpha_3}$ in (40) can be written as

$$\hat{\alpha_3} = (\mathbf{z}'\mathbf{M_x z})^{-1}\mathbf{z}'\mathbf{M_x}\epsilon \tag{66}$$

and the variance of $\hat{\alpha_3}$ is

$$Var(\hat{\alpha}_2) = \mathbf{z}'\mathbf{M_x z}/\sigma^2 \tag{67}$$

$$= \mathbf{I}_3 - \mathbf{I}_2\mathbf{I}_1^{-1}\mathbf{I}_2'. \tag{68}$$

From $\mathbf{Q}_1$ and $\mathbf{Q}_2$, the asymptotic distributions of $\hat{\alpha}_3$ is

$$\mathcal{B} \equiv \left(\int_0^1 Z_d(r)^2 dr \right)^{-1} \left(\int_0^1 Z_d(r) dY(r) \right) \tag{69}$$

Therefore, the limit distribution of the directed Wald test can be written as

$$DW \Rightarrow \{ \mathcal{W} + 2 \log \mathbf{\Phi}(A, \mathcal{B}, \mathcal{V}) \} . \tag{70}$$

where $\mathcal{V}$ is the asymptotic variance of $var(\hat{\alpha}_3)$. 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

$$W = \frac{1}{\hat{\sigma}^2} (\mathbf{U}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}(\mathbf{X}'\mathbf{U}). \tag{71}$$

Note that $\hat{\alpha_1}$ is dominant in (23), and $T^{-3}(X'X) = O_p(1)$. For y_t is a short-memory process under alternative, then $(U'X)$ diverges to infinity at rate $O_p(T^{3/2})$. Thereby, it is sufficient to show that the test statistic diverges to infinity with a rate of $O_p(T^{1/2})$ 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 (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

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

   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

$$T^{-(1/2+d)} \sum_{t=1}^{T} \epsilon_t y_{t-1} \Rightarrow \sigma^2 \int_0^1 Y_d(r) dY(r) \tag{72}$$

$$T^{-(1/2+d)} \sum_{t=1}^{T} \epsilon_t y_{t-1}\{\exp(-\gamma y_{t-1}^2)\} \Rightarrow \sigma^2 \mu_\gamma \int_0^1 Y_d(r) dY(r). \tag{73}$$

   Then we have

$$T^{-(1/2+d)} \sum_{t=1}^{T} \epsilon_t y_{t-1}\{1-\exp(-\gamma y_{t-1}^2)\} \Rightarrow \sigma^2(1-\mu_\gamma) \int_0^1 Y_d(r) dY(r) \tag{74}$$

   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

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

where

$$T^{-2(1/2+d)} \sum_{t=1}^{T} y_{t-1}^2 \quad \Rightarrow \quad \sigma^2 \int_0^1 Y_d(r)^2 dr \tag{75}$$

$$T^{-2(1/2+d)} \sum_{t=1}^{T} y_{t-1}^2 \{\exp(-\gamma y_{t-1}^2)\} \quad \Rightarrow \quad \sigma^2 \mu_\gamma \int_0^1 Y_d(r)^2 dr \tag{76}$$

$$T^{-2(1/2+d)} \{\sum_{t=1}^{T} y_{t-1}^2 \{\exp(-2\gamma y_{t-1}^2)\} \quad \Rightarrow \quad \sigma^2 \psi_\gamma \int_0^1 Y_d(r)^2 dr \tag{77}$$

$$T^{-2(1/2+d)} \sum_{t=1}^{T} y_{t-1}^2 \{1 - \exp(-\gamma y_{t-1}^2)\} \quad \Rightarrow \quad \sigma^2 (1 - 2\mu_\gamma + \psi_\gamma) \int_0^1 Y_d(r)^2 dr \tag{78}$$

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

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

   Then, the asymptotics of each element is

$$T^{-2(1/2+d)} \sum_{t=1}^{T} z_t y_{t-1} \quad \Rightarrow \quad \sigma^2 \int_0^1 Y_d(r) dZ_d(r) \tag{79}$$

$$T^{-2(1/2+d)} \sum_{t=1}^{T} z_t y_{t-1} \{\exp(-\gamma y_{t-1}^2)\} \quad \Rightarrow \quad \sigma^2 \mu_\gamma \int_0^1 Y_d(r) dZ_d(r) \tag{80}$$

   and therefore,

$$T^{-2(1/2+d)} \sum_{t=1}^{T} z_t y_{t-1} \{1 - \exp(-\gamma y_{t-1}^2)\} \Rightarrow \sigma^2 (1 - \mu_\gamma) \int_0^1 Y_d(r) dZ_d(r). \tag{81}$$

similarly, we can show that

$$T^{-2(1/2+d)}\sum_{t=1}^{T}y_{t-1}y_{t-1}\{1-\exp(-\gamma y_{t-1}^2)\}\Rightarrow\sigma^2(1-\mu_\gamma)\int_0^1 Y_d(r)^2d(r). \tag{82}$$

Combining these results, we obtain the following limit distribution

$$\begin{bmatrix} T^{-2(1/2+d)}\sum_{t=1}^{T}x_tx_t & T^{-2(1/2+d)}\sum_{t=1}^{T}x_ty_{t-1} & T^{-2(1/2+d)}\sum_{t=1}^{T}x_tz_t \\ T^{-2(1/2+d)}\sum_{t=1}^{T}x_tz_t & T^{-2(1/2+d)}\sum_{t=1}^{T}y_{t-1}^2 & T^{-2(1/2+d)}\sum_{t=1}^{T}y_{t-1}z_t \\ T^{-2(1/2+d)}\sum_{t=1}^{T}x_tz_t & {}^{-2(1/2+d)}\sum_{t=1}^{T}y_{t-1}z_t & T^{-2(1/2+d)}\sum_{t=1}^{T}z_tz_t \end{bmatrix} \Rightarrow \mathcal{A}_1\mathbf{Q}_1(\gamma)\mathcal{A}_1 \tag{83}$$

$$\begin{bmatrix} T^{-(1/2+d)}\sum_{t=1}^{T}x_t\epsilon_t \\ T^{(-(1/2+d))}\sum_{t=1}^{T}y_{t-1}\epsilon_t \\ T^{(-(1/2+d))}\sum_{t=1}^{T}z_t\epsilon_t \end{bmatrix} \Rightarrow \mathcal{A}_2\mathbf{Q}_2(\gamma) \tag{84}$$

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

$$\mathbf{Q_1}(\gamma)=\begin{bmatrix} (1-2\mu_\gamma+\psi_\gamma)\int_0^1 Y_d(r)^2dr & (1-\mu_\gamma)\int_0^1 Y_d(r)^2d(r) & (1-\mu_\gamma)\int_0^1 Y_d(r)dZ_d(r) \\ (1-\mu_\gamma)\int_0^1 Y_d(r)^2dr & \int_0^1 Y_d(r)^2dr & \int_0^1 Y_d(r)dZ_d(r) \\ (1-\mu_\gamma)\int_0^1 Y_d(r)dZ_d(r) & \int_0^1 Y_d(r)dZ_d(r) & \int_0^1 Z_d(r)^2dr \end{bmatrix} \tag{85}$$

$$\mathbf{Q_2}(\gamma)=\begin{bmatrix} (1-\mu_\gamma)\int_0^1 Y_d(r)dY(r) \\ \int_0^1 Y_d(r)dY(r) \\ \int_0^1 Z_d(r)dY(r) \end{bmatrix}, \tag{86}$$

with $\mathcal{A}_1$ is a matric variance covariance with diagonal of σ, and $\mathcal{A}_2$ has diagonal element of σ^2. Furthermore, for $0 < d < 0.5$, the rate of convergence of the OLS parameter is

$$\begin{bmatrix} T^{(1/2+d)} & 0 & 0 \\ 0 & T^{(1/2+d)} & 0 \\ 0 & 0 & T^{(1/2+d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \\ \hat{\alpha_3} \end{bmatrix} \Rightarrow \mathbf{Q_1}(\gamma)^{-\mathbf{1}}\mathbf{Q_2}(\gamma) \quad \textbf{if } 0 \leq d < 0.5 \tag{87}$$

and for $0.5 < d < 1$

$$\begin{bmatrix} T^{(1-d)} & 0 & 0 \\ 0 & T^{(1-d)} & 0 \\ 0 & 0 & T^{(1-d)} \end{bmatrix} \begin{bmatrix} \hat{\alpha_1} \\ \hat{\alpha_2} \\ \hat{\alpha_3} \end{bmatrix} \Rightarrow \mathbf{Q_1}(\gamma)^{-\mathbf{1}}\mathbf{Q_2}(\gamma) \quad \textbf{if } 0.5 < d < 1 \tag{88}$$

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 $\alpha_1 = \alpha_2 = \alpha_3 = 0$ is

$$W_\gamma \quad = \quad \frac{1}{\hat{\sigma}^2}(\epsilon'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}(\mathbf{X}'\epsilon), \tag{89}$$

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:

$$W_\gamma \Rightarrow \mathcal{W}_\gamma \equiv \left\{\mathbf{Q_2}(\gamma)'\mathbf{Q_1}(\gamma)^{-\mathbf{1}}\mathbf{Q_2}(\gamma)\right\}. \tag{90}$$

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

$$\sup_{\gamma\in\Gamma} DW_\gamma \Rightarrow \sup_{\gamma\in\Gamma} \left\{W_\gamma + 2\log[\mathbf{\Phi}(A, \hat{\alpha_2}, Var(\hat{\alpha_2}))]\right\}. \tag{91}$$

Now, let rewrite (34) as

$$\mathbf{U} = \mathbf{x}'\beta + \mathbf{z}'\alpha_{\mathbf{3}} + \epsilon_t \tag{92}$$

where $\mathbf{z} = [z_1, ..., z_t]$, $\mathbf{x} = [(y_0\{1-\exp(-\gamma y_0^2)\}, y_0), ..., (y_{t-1}\{1-\exp(\gamma y_{t-1}^2)\}, y_{t-1})]'$ and $\beta = (\alpha_1, \alpha_2)$. The directed-Wald statistic with $c = \infty$ is

$$DW_\gamma \Rightarrow \{W_\gamma + 2\log[\mathbf{\Phi}(A, \hat{\alpha_3}, Var(\hat{\alpha_3}))]\}\,. \tag{93}$$

Similar to the procedures applied for theorem 2, it is straightforward to show that the limit distribution of the sup-directed Wald is

$$\sup_{\gamma\in\Gamma} DW_\gamma \Rightarrow \sup_{\gamma\in\Gamma} \{\mathcal{W}_\gamma + 2\log \mathbf{\Phi}(A, \mathcal{B}, \mathcal{V})\}\,. \tag{94}$$

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

$$W_\gamma \ = \ \frac{1}{\hat{\sigma}^2}(\mathbf{U}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}(\mathbf{X}'\mathbf{U}). \tag{95}$$

$\hat{\alpha_1}$, $\hat{\alpha_2}$ and $\hat{\alpha_3}$ have the same convergence rate. Under alternative hypothesis that y_t is short memory, $(U'X)$ diverges to infinity at rate $O_p(T)$. Therefore, it is easy to show that the test statistic diverges to infinity with rate $O_p(T^{1/2})$, and it builds the consistency of the test.

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 Soe (2004). Therefore, we follow Soe (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 $\gamma^* \in [\gamma, \gamma']$. Now, we assume that ϵ_t is martingale difference sequence and thus, 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, $P(\sup_{|\gamma-\gamma'|\leq\delta} |G_T(\gamma) - G_T(\gamma')| > \epsilon) \to 0$.