Centre of Excellence
for Science and Innovation Studies

CESIS Electronic Working Paper Series

Paper No. 184

Testing for Unit Root against LSTAR model

– wavelet improvements under GARCH distortion

Yushu Li* and Ghazi Shukur**

(Centre for Labour Market Policy (CAFO),
Växjö University*,Jönköping International Business School**)

August 2009

The Royal Institute of Technology
Centre of Excellence for Science and Innovation Studies (CESIS)
http://www.cesis.se

Testing for Unit Root against LSTAR Model:
Wavelet Improvement under GARCH Distortion

By

Yushu Li[1] and Ghazi Shukur[2]

[1] Center for Labor Market Policy Research (CAFO) Department of Economic and Statistics, Växjö University, Sweden
[2] Jönköping International Business School, and Center for Labor Market Policy Research (CAFO) Department of Economic and Statistics, Växjö University, Sweden

Abstract

In this paper, we propose a Nonlinear Dickey-Fuller F test for unit root against first order Logistic Smooth Transition Autoregressive LSTAR (1) model with time as the transition variable. The Nonlinear Dickey-Fuller F test statistic is established under the null hypothesis of random walk without drift and the alternative model is a nonlinear LSTAR (1) model. The asymptotic distribution of the test is analytically derived while the small sample distributions are investigated by Monte Carlo experiment. The size and power properties of the test have been investigated using Monte Carlo experiment. The results have shown that there is a serious size distortion for the Nonlinear Dickey-Fuller F test when GARCH errors appear in the Data Generating Process (DGP), which lead to an over-rejection of the unit root null hypothesis. To solve this problem, we use the Wavelet technique to count off the GARCH distortion and to improve the size property of the test under GARCH error. We also discuss the asymptotic distributions of the test statistics in GARCH and wavelet environments. Finally, an empirical example is used to compare our test with the traditional Dickey-Fuller F test.

Keywords: Unit root Test, Dickey-Fuller F test, STAR model, GARCH (1, 1), Wavelet method, MODWT

JEL: C 32

I. Introduction

Empirical studies show that many economic variables display nonlinear features, such as the business cycles of production, investment and unemployment rates, where the economic behaviors change when certain variables lie in different regions (see Granger and Teräsvirta, 1993). To capture such nonlinear features, several nonlinear models have been introduced. Haggan, Heravi and Priestley (1984) were the first to present a family of "state dependent" models, including threshold autoregressive (TAR), exponential autoregressive (EAR) and smooth transition autoregressive (STAR) models, (see also Simon, 1999). Among them, STAR models allow nonlinear structures between the data regimes to be described with a smooth regime transition function. They are of particular interest in macroeconomics which always contains mass of economic agents, where even if the decisions are made discretely, the aggregated behaviors will show smooth regime changes (see Teräsvirta, 1994). There are two main STAR models: logistic STAR (LSTAR) and exponential STAR (ESTAR); the former contains TAR as a limit case. These models have wide applications; see for example, Teräsvirta and Anderson (1992), who applied the models to industrial production for 13 OECD countries and Europe. Hall, Skalin and Teräsvirta (2001) used nonlinear LSTAR to describe the most turbulent period in El Niño event. Arango and Gonzalez (2001) found evidence of STAR representations in annual inflation in Colombia.

However, before applying nonlinear models, testing linearity against nonlinearity is essential, especially for the forecast analysis (see Teräsvirta, 1994; Teräsvirta, Dijk, and Medeiros, 2003; and Wahlström, 2004). Among the tests, unit root tests against nonlinear model need cautious consideration; we know that the unit root tests in linear models, such as Dickey-Fuller (1979), Phillips-Perron (1998) lack power when the alternative model shows non-linearity. In nonlinear cases, Enders, Walter and Granger (1998), Berben and Dijk (1999), Caner and Hansen (2001) performed tests for unit root against TAR, and showed that several series are better described by the TAR models. Kapetanios, Shin, and Snell (2003) proposed a unit root test against ESTAR model; Eklund (2003) proposed tests against LSTAR with transition variables being the lagged dependent variables. Later, He and Sandberg (2006) proposed the nonlinear Dickey-Fuller ρ and t test statistics with time as the transition variable. In this paper we first derive Nonlinear Dickey-Fuller F test of unit root against LSTAR models with time as the transition variable, we also investigate the size and power property of the test under independent normal distributed ($n.i.d.$) error.

We next investigate the size property of the Nonlinear Dickey-Fuller F test when the error in the DGP shows conditional heteroskedasticity. The conditional heteroskedasticity was first mentioned in Engle (1982), who observed that many financial time series show apparent clustering in volatility although the overall series are stationary. ARCH models were designed to model this volatility and future variance forecast. As ARCH models require large order of lags when modeling persistence shocks, Bolleslev (1986) proposed GARCH models that can be represented by ARCH (∞). To estimate the parameters of ARCH/GARCH models, the Least Squares Estimate (LSE) and Maximum Likelihood Estimate (MLE) are used; the latter is more efficient under certain moment conditions, (see Li, Ling and McAleer, 2002).

However, Li, Ling and McAleer (2002) mentioned that ARCH/GARCH models are mainly employed to model the conditional variance without paying enough attention to the specification of the conditional mean, and any misspecification may lead to inconsistent estimates. Thus it is important to specify the conditional mean function at the outset. Specification tests are needed and the unit root test under ARCH/GARCH error has attracted much attention. Although Pantula (1988), Ling and Li (1997b) showed that the Dickey-Fuller tests could still be employed with ARCH/GARCH errors, Peter and Veloce (1988), Kim and Schmidt (1993) showed that they are generally not robust in the near integrated situation in GARCH error. Cook (2006) extended the study to the modified Dickey-Fuller unit root tests and showed that over-sizing was observed especially when the GARCH process exhibits a high degree of volatility.

Therefore, to improve the test property, numerous studies pay attention to deriving unit root test based on Maximum Likelihood Estimation (MLE), which jointly estimates the parameters of unit root model and the GARCH error model. Among them, Seo (1999) derived a t -statistic under the 8^{th} order moment condition. The distribution is a mixture of the Dickey-Fuller t distribution and the standard normal. Ling and Li (1998) derived unit root tests by MLE with GARCH error under the 4^{th} order moment condition, and later the 2^{nd} order moment condition in Ling and Li (2003).Those distributions are function of bivariate Brownian motion. Sjölander (2007) also presented an ADF-Best test by estimating GARCH error and model parameters before the unit root test.

However, the MLE is not a perfect solution to the GARCH error problem. Charles and Darné (2008) pointed out that when using MLE, Seo's (1999) conclusion is based on the ARCH parameter α being superior to GARCH parameter β in 8 of the 10 GARCH (1, 1) processes. As van Dijk, Franses and Lucas (1999); Poon and Granger (2003) showed that the estimated GARCH (1,1) models are mostly in a situation where $\beta > \alpha$ (The definition of β and α please refer to Section IV.), Charles and Darné (2008) re-examined Seo's Monte Carlo experiments with $0.8 < \alpha + \beta < 1$ and $\beta > \alpha$, and they showed that the empirical size and power of the Dickey-Fuller test is generally better than Seo's test. Moreover, in our LSTAR model, if we use MLE method, the estimated dimensional parameter space is larger than the linear case and it will be numerically quite complicated to obtain. Thus in this paper, we consider an alternative to MLE method to improve the unit root test under GARCH error.

We apply the wavelet method, which has been widely used after its theoretic foundation in 1980s (see Grossmann and Morelet, 1984 and Mallat, 1989), such as in signal smoothing and spectrum analysis where Chiann and Morettin (1998) showed how wavelet capture signals in different scales by wavelet spectrum decomposition. In economics, Schleicher (2002) found that since economic behaviors take place at different frequencies, the wavelet method can catch landscape characteristics in addition to the microscopic detail in economic areas. In this paper, we use the wavelet method to count off the finest local behavior of the series in the form of conditional heteroskedasticity in GARCH errors, whose information is caught by the highest scale in wavelet coefficients. The same logic can be found in Schleicher (2002), who pointed out that lower scales hold most of the energy of the unit root process and that non-lasting disturbances are captured by the higher scale coefficients. This logic is also reflected in Fan and Gençay (2006), who stated that the spectrum of a unit root process is infinite at frequency 0. They proposed a unit root test on the perspective of the frequency domain as the test is the ratio of the energy of the low frequency scale to the total energy of the time series.

Here our Nonlinear Dickey-Fuller F test statistic is in a time domain where we use the scaling coefficient directly in the test statistics; in this way, the asymptotic distribution of the test statistics will not be influenced under the wavelet environment. We use Maximal Overlap Discrete Wavelet Transform (MODWT) as it has no restriction on the sample size and LA (8) wavelet filter as it has better band pass character. For more information about the MODWT methods and LA filter, we refer to Vidakovic (1998), Percival and Walden (2000), and to Gençay Selcuk and Whicher (2001b).

The paper is organized as follows. Section II presents the LSTAR model, the procedure for testing unit root against the LSTAR alternatives, the asymptotic properties of the test statistics and the finite sample distribution of the test. Section III investigates the size and power property of the test, and offers an empirical example. Section IV shows the size distortion of the test statistics under GARCH (1, 1) error. Section V presents the wavelet size improvement of the small samples and the asymptotical distribution. Section VI presents another empirical example to illustrate the wavelet improvement of over-rejection in the linear case. Concluding remarks can be found in the final section. All proofs of theorems in this paper are given in the Appendix.

II. Model, Test procedure, The Nonlinear Dickey-Fuller F test

In STAR models, the main difference between LSTAR and ESTAR models is that the LSTAR model can catch the asymmetric feature of a process in two extreme states: when the economic contractions are always more violent, and when expansions are more stationary and persistent. We only consider the nonlinear LSTAR models in two cases; the first case does not contain a drift term and the second case does.

Case 1: $y_t = \pi_{11} y_{t-1} + \pi_{21} y_{t-1} F(t,\gamma,c) + u_t$.

(1)

Case 2: $y_t = \pi_{10} + \pi_{11} y_{t-1} + (\pi_{20} + \pi_{21} y_{t-1}) F(t,\gamma,c) + u_t$.

(2)

The transition function $F(t,\gamma,c)$ in (1) and (2) is defined as follows:

$$F(t;\gamma,c) = \frac{1}{(1+\exp\{-\gamma(t-c)\})} - \frac{1}{2}.$$

The transition function here differs from that of the LSTAR model presented in Teräsvirta (1994), where the transition variable is defined as the lag values. The model in Teräsvirta

(1994) depicts the situation in which regime change depends on the deviation of the lagged observations while our model, as the same situation in He and Sandberg (2006), implies that the equilibrium regimes switch as the time evolves. In the transition function, γ determines the speed of transition from one extreme regime to another at time c, the larger the γ is, the steeper the transition function will be, leading to a faster transition speed. In Figure 1, we set c fixed as halfway time in three cases where γ = 20, 10, 5. Then the smooth transition function Y is a bounded continuous non-decreasing transition function in t and t is from 1 to 44.

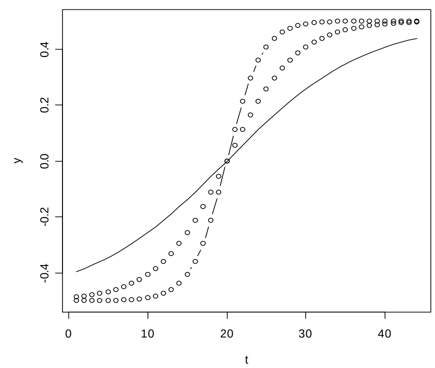

Figure 1. Logistic smooth transition functions: γ =20 (dashed-dotted line), γ =10 (dotted line), γ =5 (solid line)

Meanwhile, set c fixed, as $\gamma \to \infty$, the function turns into a step function of t and the model becomes a two regimes threshold autoregressive model (TAR). When setting t and c fixed, the situation when $\gamma \to 0$ leads the resulting model to be linear. Therefore, the linear hypothesis is equivalent to the hypothesis that $\gamma = 0$.

Our goal is to test the null hypothesis of a random walk without drift against the nonlinear LSTAR (1) model. The null hypothesis can be expressed as the following parameter restrictions:

Case1: $H_0^* : \gamma = 0, \pi_{11} = 1$.

Case2: $H_0 : \gamma = 0, \pi_{10} = 0, \pi_{11} = 1$.

Since γ=0 will lead to an identification problem under the null hypotheses (see Teräsvirta, 1994), we follow the approach used by Luukkonen, Saikkonen and Teräsvirta (1988), also He and Sandberg (2006), by applying a Taylor expansion of the $F(t,\gamma,c)$ with γ around. However, we should keep in mind that the first-order expansion will lead to low power if the transition takes place only in the drift (see Luukkonen, Saikkonen and Teräsvirta (1988)). Therefore, the third-order Taylor expansion is more robust in power. The first- and third-order Taylor expansions are as follows:

$$F_1(t;\gamma,c)=\frac{\gamma(t-c)}{4}+o(\gamma). \tag{3}$$

$$F_3(t;\gamma,c)=\frac{\gamma(t-c)}{4}+\frac{\gamma^3(t-c)^3}{48}+o(\gamma^3). \tag{4}$$

Substituting equations (3) and (4) into the models in equations (1) and (2), after merging the terms, we obtain the following auxiliary regressions:

$$Case1\,\&\,Order1: y_t=(y_{t-1}s_{1t})'\varphi_1+u_{1t}^*.$$
$$Case1\,\&\,Order3: y_t=(y_{t-1}s_{3t})'\varphi_3+u_{3t}^*.$$
$$Case2\,\&\,Order1: y_t=s_{1t}'\lambda_1+(y_{t-1}s_{1t})'\varphi_1+u_{1t}.$$
$$Case2\,\&\,Order3: y_t=s_{3t}'\lambda_3+(y_{t-1}s_{3t})'\varphi_3+u_{3t}.$$

The parameters are defined as follows:

$$s_{1t}=\begin{pmatrix}1\\ t\end{pmatrix},\lambda_1=\begin{pmatrix}\lambda_{10}\\ \lambda_{11}\end{pmatrix},\varphi_1=\begin{pmatrix}\varphi_{10}\\ \varphi_{11}\end{pmatrix};\ s_{3t}=\begin{pmatrix}1\\ t\\ t^2\\ t^3\end{pmatrix},\lambda_3=\begin{pmatrix}\lambda_{30}\\ \lambda_{31}\\ \lambda_{32}\\ \lambda_{33}\end{pmatrix},\varphi_3=\begin{pmatrix}\varphi_{30}\\ \varphi_{31}\\ \varphi_{32}\\ \varphi_{33}\end{pmatrix}$$

For details of the merging procedure, please refer to the Appendix. Note that the unit root test in the nonlinear time series model is a joint test of both unit root and linearity. The corresponding auxiliary null hypotheses are:

$$H_{0m}^*:\varphi_{m0}=1,\varphi_{mj}=0,\ j\geq 1; m=1,3.$$
$$H_{0m}:\lambda_{mi}=0,\forall i;\varphi_{m0}=1,\varphi_{mj}=0,\ j\geq 1; m=1,3.$$

We should keep in mind that under the null hypothesis, u_{mt}^* and u_{mt} are equal to u_t with $m = 1,3$.

Following the above auxiliary regressions and null hypothesis, we now derive the unit root test statistics and investigate their distribution properties. Here we assume that the error terms in equations (1) and (2) are independent identically distributed ($i.i.d.$). In the distribution form, the $W(.)$ represents a standard Brownian motion on [0, 1].

Assumption 1: *Let* $\{u_t\}$ *is* $i.i.d.$ *with* $E(u_t) = 0, Var(u_t) = \sigma_u^2$, *and* $E(u_t^4) < \infty$.

Under *Assumption 1*, we derive two theorems that will be used to deduce the Nonlinear Dickey-Fuller F test statistic distribution. We first consider the case that does not contain intercept.

Theorem1: *Assume that the following models* $y_t = (y_{t-1} s_{mt})' \varphi_m + u_{mt}^*$ *hold, and assume that* $(u_{mt}^*)_{t=1}^{\infty}$ *fulfills Assumption 1. Then for* $m = 1,3$, *we have the following:*

$$\hat{\psi}_m^* - \psi_m^* \xrightarrow{p} 0, \quad \Upsilon_m^* (\hat{\psi}_m^* - \psi_m^*) \xrightarrow{L} (\Psi_m^*)^{-1} \Pi_m^*, \quad (s_m^*)^2 \Upsilon_m^* (\sum x_{mt}^* (x_{mt}^*)')^{-1} \Upsilon_m^* \xrightarrow{L} \sigma^2 (\Psi_m^*)^{-1}. \textit{With}$$

parameter restrictions as follows:

$\Upsilon_1^* = diag\{T_1^*\}, \ T_1^* = [T \quad T^2]$,

$\Upsilon_3^* = diag\{T_3^*\}, \ T_3^* = [T \quad T^2 T^3 \ T^4], \ \hat{\psi}_m^* = (\hat{\varphi}_m), \ \psi_m^* = (\varphi_m)$,

$x_{mt}^* = [y_{t-1} s_{mt}], \ \Psi_m^* = \sigma_u^2 C_m, \ \Pi_m^* = E_m, \ C_m = [c_{ij}]_{(m+1)*(m+1)}, \quad c_{ij} = \int_0^1 r^{i+j-2} W(r)^2 dr$,

$$E_m = [e_i]_{(m+1)*1}, \quad e_i = \frac{\left(W(1)^2 - (i-1)\int_0^1 r^{i-2} W(r)^2 dr - 1/i \right)}{2}.$$

Based on *Theorem1*, under the null hypothesis $H0: \ R_m^* \psi_m^* = r_m^*$ we have the following Nonlinear Dickey-Fuller F test statistic:

$$\begin{aligned} F_m^* &= (\hat{\psi}_m^* - \psi_m^*)' R_m^{*'} \{(s_m^*)^2 R_m^* [\sum x_{mt}^* x_{mt}^{*'}]^{-1} R_m^{*'}\}^{-1} R_m^* (\hat{\psi}_m^* - \psi_m^*)/2 \\ &= (\hat{\psi}_m^* - \psi_m^*)' R_m^{*'} \Upsilon_m^* * \{(s_m^*)^2 \Upsilon_m^* R_m^* [\sum x_{mt}^* x_{mt}^{*'}]^{-1} R_m^{*'} \Upsilon_m^*\}^{-1} * \Upsilon_m^* R_m^* (\hat{\psi}_m^* - \psi_m^*)/2 \\ &\xrightarrow{L} [(\Psi_m^*)^{-1} \Pi_m^*]' [\sigma_u^2 (\Psi_m^*)^{-1}]^{-1} [(\Psi_m^*)^{-1} \Pi_m^*]/2 = (\Pi_m^*)' (\Psi_m^*)^{-1} \Pi_m^* / 2\sigma_u^2. \end{aligned}$$

Where: $R_m^* = I_{m+1}$, $(r_1^*)' = [1\ 0]$, $(r_3^*)' = [1\ 0\ 0\ 0]$.

For proof of *Theorem 1* we refer to the Appendix.

The following *Theorem 2* is for the second case that contains a smooth transition function both in drift and dynamics.

Theorem2: *Assume that the following models* $y_t = s_{mt}{}'\lambda_m + (y_{t-1}s_{mt})'\varphi_m + u_{mt}$ *hold, and assume that* $(u_{mt})_{t=1}^{\infty}$ *fulfills Assumption 1, then for* $m = 1,3$*, we have the following:*

$$\hat{\psi}_m - \psi_m \xrightarrow{p} 0, \qquad \Upsilon_m(\hat{\psi}_m - \psi_m) \xrightarrow{L} \Psi_m^{-1}\Pi_m, \quad s_m^2\Upsilon_m(\sum x_{mt}x_{mt}')^{-1}\Upsilon_m \xrightarrow{L} \sigma_u{}^2\Psi_m^{-1}.$$

Where the parameters are defined as follows:

$\Upsilon_1 = diag\{T_1\}$, $T_1 = [T^{1/2}\ T^{3/2}\ T\ \ T^2]$,

$\Upsilon_3 = diag\{T_3\}$, $T_3 = [T^{1/2}\ T^{3/2}\ T^{5/2}\ T^{7/2}\ T\ \ T^2\ T^3\ T^4]$,

$$\hat{\psi}_m = \begin{pmatrix} \hat{\lambda}_m \\ \hat{\varphi}_m \end{pmatrix},\ \psi_m = \begin{pmatrix} \lambda_m \\ \varphi_m \end{pmatrix},\ x_{mt} = \begin{bmatrix} s_{mt} \\ y_{t-1}s_{mt} \end{bmatrix},\ \Psi_m = \begin{bmatrix} A_m & \sigma_u B_m \\ \sigma_u B_m' & \sigma_u^2 C_m \end{bmatrix},\ \Pi_m = \begin{bmatrix} \sigma_u D_m \\ \sigma_u^2 E_m \end{bmatrix},$$

$$A_m = \left[a_{ij}\right]_{(m+1)*(m+1)},\ B_m = \left[b_{ij}\right]_{(m+1)*(m+1)},\ C_m = \left[c_{ij}\right]_{(m+1)*(m+1)},\ D_m = \left[d_i\right]_{(m+1)*1},\ E_m = \left[e_i\right]_{(m+1)*1},$$

$$a_{ij} = T^{-(i+j-1)}\sum_{t=1}^{T} t^{i+j-2},\quad b_{ij} = \int_0^1 r^{i+j-2}W(r)dr,\quad c_{ij} = \int_0^1 r^{i+j-2}W(r)^2 dr,$$

$$d_i = W(1) - (i-1)\int_0^1 r^{i-2}W(r)dr,\quad e_i = \frac{\left(W(1)^2 - (i-1)\int_0^1 r^{i-2}W(r)^2 dr - 1/i\right)}{2}.$$

Based on *Theorem 2*, under the null hypothesis $H0$: $R_m\psi_m = r_m$, we have the Nonlinear Dickey-Fuller F test statistic as follows:

$$\begin{aligned} F_m &= (\hat{\psi}_m - \psi_m)' R_m{}'\{s_m^2 R_m(\sum x_{mt}x_{mt}')^{-1}R_m{}'\}^{-1}R_m(\hat{\psi}_m - \psi_m)/2 \\ &= (\hat{\psi}_m - \psi_m)' R_m{}'\Upsilon_m * \{s_m^2\Upsilon_m R_m(\sum x_{mt}x_{mt}')^{-1}R_m{}'\Upsilon_m\}^{-1} * \Upsilon_m R_m(\hat{\psi}_m - \psi_m)/2 \\ &\xrightarrow{L} (\Psi_m^{-1}\Pi_m)'\{\sigma_u{}^2\Psi_m^{-1}\}^{-1}(\Psi_m^{-1}\Pi_m)/2 = \Pi_m{}'\Psi_m^{-1}\Pi_m/2\sigma_u{}^2. \end{aligned}$$

Where: $R_m = I_{2*(m+1)}$, $r_1' = [0\ 0\ 1\ 0]$, $r_3' = [0\ 0\ 0\ 0\ 1\ 0\ 0\ 0]$.

For proof of *Theorem 2* we refer to the Appendix.

To find out the finite-sample distributions of the test, we generate data from the model $y_t = y_{t-1} + u_t$ where $u \sim n.i.d.(0,1)$ with desired sample sizes. To get the asymptotic distributions for the Nonlinear Dickey-Fuller F test, we let T=100,000 simulate a Brownian motion $W(r)$, and the number of Monte Carlo replication is 10000. Here we only report the critical value table of the model: Case 2 and Order 1, as we only use this table in the following part of the paper.

Table 1. Critical values for the Nonlinear D-F F test; Case 2 and Order 1

T	99%	97.5%	95%	90%	10%	5%	2.5%	1%
25	1.0155	1.2389	1.4778	1.8214	6.2068	7.1836	8.0585	9.2705
50	1.1417	1.4198	1.6750	2.0125	6.7796	7.7654	8.5874	9.8294
100	1.2223	1.4928	1.7714	2.1522	6.8958	7.9696	8.8991	10.2894
250	1.3098	1.5791	1.8621	2.2335	7.2886	8.3162	9.3327	10.6171
500	1.2647	1.5526	1.8212	2.2124	7.3014	8.4495	9.4674	10.7152
∞	1.6440	1.8515	2.0155	2.3750	7.5639	8.2807	8.5646	8.6974

Size and power property of the test, Empirical Example

We again use the Monte Carlo method to investigate the size and power properties of our test statistics. We choose Case 2 and Order 1 as the alternative model since we want to take into account the dynamics in drift and to avoid high parameter dimension which shows in third-order Taylor expansion. The investigation of size has been carried out when DGP follows a unit root process with $u_t \sim n.i.d.(0,1)$, with 5% nominal size (the critical value is from Table 1). For sample sizes T=25, 50, 100, 250, 500. The estimated sizes are presented in the following table.

Table 2. Size property for the Nonlinear D-F F test; Case 2 & Order 1

T	25	50	100	250	500
Size (T)	0.049	0.049	0.057	0.050	0.047

To judge the reasonability of the results, the estimated size of the test should lay between the approximate 95% confidence intervals of the actual size 5%. With replication number equal to 10000, the approximate 95% confidence interval for the estimated size is: $0.05 \pm 1.96\sqrt{\frac{0.05(1-0.05)}{10000}} = (0.0457, 0.0543)$.

Table 2 shows that at the 95% confidence level, the Nonlinear Dickey-Fuller F test has almost an unbiased size, while when T = 100 the test tends to have a slight over-rejection.

In what follows we examine the power property of the Nonlinear Dicey-Fuller F test, with the LSTAR model being Case 2 and Order 1. As we are more interested in the variation of the dynamic parameters, we set the drift parameter stable with π_{10}=0, π_{20}=1. We also set the transition speed parameter γ=1, which is thought to be a reasonable starting value for the iterative nonlinear least squares estimation in Wahlström (2004). We set the transition time c as $T/2$. Thus the changing parameters are sample size T, dynamical parameters π_{11}, π_{21}. We also impose the Lagrange stability condition $\pi_{11} + \pi_{21} \in (0,1)$ with $\pi_{11} + \pi_{21} = 0.9$ to ensure the stable trajectories (see He and Sandberg (2005), Tong (1990)). To observe the power changes with the seriousness of the nonlinear dynamics impact which is measured by π_{21}, we set π_{21} six different values from 0.8 to 0.1. The designated values for the changing parameters are as follows:

$T \in (25, 50, 100, 250, 500)$, $\pi_{11} \in (0.1, 0.3, 0.4, 0.5, 0.6, 0.8)$, $\pi_{21} \in (0.8, 0.6, 0.5, 0.4, 0.3, 0.1)$

Table 3. Power property for the Nonlinear D-F F test; Case 2 and Order 1

T	π_{11}=0.8 π_{21}=0.1	π_{11}=0.6 π_{21}=0.3	π_{11}=0.5 π_{21}=0.4	π_{11}=0.4 π_{21}=0.5	π_{11}=0.3 π_{21}=0.6	π_{11}=0.1 π_{21}=0.8
25	0.123	0.247	0.321	0.45	0.592	0.836
50	0.15	0.396	0.636	0.773	0.921	0.994
100	0.154	0.816	0.972	0.998	1	1
250	0.727	0.94	1	1	1	1
500	0.97	1	1	1	1	1

The power properties with some parameter combinations are presented in Table 3. The table shows that for small sample 25, 50, 100, the power depends mainly on the proportions of the linear and nonlinear parts; the higher the nonlinear part, the better power it is. This can be explained in the form of the Taylor expansion, where with the decrease of π_{11} follows increase of π_{21}, which lead to the decrease of the proportion of y_{t-1}. Moreover, with the increases in the sample size and the nonlinear proportion, the test shows better power property as well.

We now show an empirical example to compare our Nonlinear Dickey-Fuller F test to the traditional Dickey-Fuller F test. We use unemployment rates in 10 OECD counties[1] from 1955 to 1999 for empirical illustration. We do not apply any data transformation before the test (such as log transform or seasonal adjust); Maraca (2005) has already observed that most data transformations will result in a loss of nonlinearity. By using the Dickey-Fuller F test, we find that unit root hypothesis was rejected only in 1 series: UK, while using the Nonlinear Dickey-Fuller F test, the unit root hypothesis was rejected in 4 series: Germany, Japan, France and UK, which shows that the traditional Dickey-Fuller F test have less power when the variables shows nonlinearity. For detailed procedure, we use the data below from France as an example:

[1] Austria, Denmark, Finland, France, Germany, Japan, Netherland, Norway, Sweden, UK

Figure 2. France's unemployment rate, 1955- 1999

In Figure 2, we can see that the time series shows an obvious data break around 1975, and the whole series is divided into two data periods smoothly. Thus we suppose that a STAR model, which contains a smooth transition function, should be a good choice. However, although we suppose the time series shows nonlinearity, we fit the data with an order one autoregressive model and test its unit root with traditional Dickey-Fuller F test first. We take this step into consideration because we will compare the result to the result we obtain by using our Nonlinear Dickey-Fuller F test in the latter case.

In the first step, a linear AR (1) process is built by OLS regression, and time t is from 1956 to 1999: $y_t = 0.2274 + 0.9971 y_{t-1} + u_t$ with sum of residue squares 15.04.
The traditional Dickey-Fuller F test statistic is:

$$F = (b_T - \beta)' \Upsilon_T' \{ s_T^2 \Upsilon_T (\sum x_t x_t')^{-1} \Upsilon_T' \}^{-1} \Upsilon_T (b_T - \beta)/2$$

Where: $b_T = \begin{bmatrix} 0.4710 \\ 0.9353 \end{bmatrix}$, $\beta = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\Upsilon_T = \begin{bmatrix} T^{1/2} & 0 \\ 0 & T \end{bmatrix}$, $\sum x_t x_t' = \begin{bmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{bmatrix}$

The calculated F statistic is 2.7538, after we compare it to the critical value 4.86 in Table B.7 in Hamilton, J.D. (1994), the null hypothesis of a unit root is not rejected at the 5% level which means the data can be viewed as random walk without drift. However, from the graph we can see that for the unemployment rate data, the series are initially at an equilibrium state but after around 1975 there shows another state, and the whole series shows a nonlinear structure. Therefore, the Dickey-Fuller F test may be not valid with its linearity assumption. We need to test the data's linearity. *Chow test* is used to test the series linearity against a single break. The first period is from 1955 to 1975, the second period is from 1976 to 1999 and we obtain the F_{Chow} statistic: F_{Chow} =0.2369. At the 20% critical value, we reject the null hypothesis and accept that there is a data break around 1975. Therefore we test the unit root by our Nonlinear Dickey-Fuller F test and we use the auxiliary regression model Case 2 & Order 1. By OLS regression, take t from 2 to 45 (correspond to year 1956 to year 1999), we arrive at the following model:

$$y_t = -0.5518 + 0.0595t + 0.0834 y_{t-1} - 0.06 t y_{t-1} + u_t \ .$$

With sum of residue squares is 11.584, the Nonlinear Dickey-Fuller F test statistic value for this specification is:

$$F_{NL} = (\hat{\psi} - \psi)' R' \{ s^2 R (\sum x_t x_t')^{-1} R' \}^{-1} R (\hat{\psi} - \psi) / 2$$

The calculated F_{NL} statistic is 9.3687, after we compare it to 7.7654 in Table 1, the null hypothesis of a unit root process for the series is rejected at the 5% level.
Then we consider building a LSTAR model:

$$y_t = \pi_{10} + \pi_{11} y_{t-1} + (\pi_{20} + \pi_{21} y_{t-1}) F(t; \gamma, c) + \mu_t .$$

Then by Nonlinear Least Square (NLS) regression, we get the following model, and t is from 2 to 45 (correspond to the year from 1956 to 1999)

$$y_t = 2.3243 + 0.3797 y_{t-1} + (2.5870 + 0.6153 y_{t-1})[\frac{1}{(1 + \exp\{-0.3408(t - 20.6567)\})} - \frac{1}{2}] + u_t$$ Her

e we notice that the estimation of c is 20.6567, which shows data that the break occurs around year 1975. The economical explanation for this break may related to the OPEC energy price rising in 1975 when the oil price raised 10% , which brought a huge shock to the economic field, including the job market. Therefore there are two different states of the unemployment rates: before and after the oil price change.

Teräsvirta (1994) pointed out it is not easy to get an accurate estimate of γ. When the true parameter is relatively large, we need huge observation in the neighbourhood of c. To solve the problem, Teräsvirta (1994) advised rescaling and here we set the starting values of γ as 1, which Wahlström (2004) has found reasonable. Anyway, here we accept the estimated γ and the sum of residual squares for LSTAR model is 9.535904, which is the lowest of the three models. Thus the LSTAR model best fits the nonlinear character of the unemployment rate in France.

From the above example, we can see that the Nonlinear Dickey-Fuller F test has better power property as it rejects the unit root hypothesis when the traditional Dickey-Fuller F test does not. In the following part we will concentrate on the size property of the test statistics when the data is influenced by GARCH (1, 1) error.

Size property under GARCH (1, 1) error

We turn here to the question of how the size property will be affected when GARCH errors appear. In the linear case, unit root tests such as Dickey-Fuller (1979), Phillips-Perron (1998) have asymptotical distribution which is invariant to heteroskedasticity. Furthermore, Pantula (1988), Ling and Li (1997b) both derived the asymptotic distribution by LSE of the unit root with ARCH/GARCH errors, and they noticed that the Dickey-Fuller test can still be used. Our nonlinear Dickey-Fuller test is based on LSE, and we assume the asymptotic distribution will be invariant to GARCH. Therefore we concentrate on the small sample property of the test under the GARCH error.

The GARCH (1, 1) is the most frequently used process due to its simplicity and robustness in measuring volatility (Engle, 2001). Actually, higher order of GARCH (p, q) models tends to overcomplicate the model and to be inconvenient to use. Thus in this paper, we only concentrate on how the test property will be influenced when the error of the DGP exhibits GARCH (1, 1). The model is as follows:

$$y_t = y_{t-1} + u_t,$$
$$u_t = \eta_t \sqrt{h_t}, h_t = w + \alpha u_{t-1}^2 + \beta h_{t-1}; \eta_t \sim i.i.d.(0,1), w = 1 - \alpha - \beta.$$

For the unit root based on the LSE in the linear case, Cook (2006) observed that size distortions of GARCH error are mainly caused by the volatility parameter α, rather than the persistence parameters $\alpha+\beta$. Hence, our experiment design in the nonlinear case will include both the cases where $\alpha \geq \beta$ and $\alpha < \beta$ in the following three situations: weak GARCH effect: $\alpha+\beta=$ 0.3; medium GARCH effect: $\alpha+\beta$=0.75; high GARCH effect: $\alpha+\beta$=0.95. We first check if there is a size distortion when the GARCH (1, 1) error is weak. The size property is as follows:

Table 4. Size property for the Nonlinear D-F F test under GARCH (1, 1) with $\alpha+\beta=$ 0.3; Case 2 and Order 1

T	α =0 β=0.3	α =0.05 β=0.25	α =0.1 β=0.2	α =0.15 β=0.15	α =0.2 β=0.1	α =0.25 β=0.05
25	0.0463	0.0548	0.0622	0.0742	0.0795	0.0815
50	0.0473	0.054	0.0586	0.0616	0.0704	0.0799
100	0.0504	0.0507	0.0539	0.0516	0.0603	0.0624
250	0.0503	0.0508	0.0534	0.056	0.06	0.0598
500	0.0522	0.0529	0.0524	0.0525	0.0522	0.0558

Table 4 shows that the size distortion is not serious in most of the cases. Especially in the larger samples, most of the size lies within the 95% confidence interval: $(0.0457, 0.0543)$. For sample sizes of 25, 50, 100, there is still some size distortions, and we can see that this over-rejection problem becomes more serious with the increase of the volatility parameter α, which is the same conclusion in Cook (2006). Thus from Table 4, we can see that when $\alpha+\beta$ is low, the Nonlinear Dickey-Fuller F test is mostly robust to the GARCH (1, 1) in large samples such as 250 and 500. Now, we consider the size property under the situation of medium GARCH effect with $\alpha+\beta=$ 0.75. The results based on Monte Carlo experiment are presented in the following table:

Table 5. Size property for the Nonlinear D-F F test under GARCH (1, 1) with $\alpha+\beta$ =0.75; Case 2 and Order 1

T	α =0.25 β=0.5	α =0.3 β=0.45	α =0.35 β=0.4	α =0.4 β=0.35	α =0.45 β=0.3	α =0.5 β=0.25
25	0.0839	0.0889	0.0958	0.0964	0.1135	0.1036
50	0.0795	0.0881	0.094	0.0998	0.1143	0.1134
100	0.0766	0.0803	0.0895	0.1027	0.1065	0.1039
250	0.0746	0.0782	0.0831	0.0909	0.0967	0.1014
500	0.0629	0.0707	0.0786	0.0831	0.0893	0.0937

Table 5 shows that when $\alpha+\beta$ = 0.75, there exists much more serious size distortion than in Table 4. However we can still notice the existence of two common characteristics in both tables: the size distortion is more serious as α increases and less serious in larger samples. This can be interpreted as follows: when GARCH effect is not high, the size distortion is mainly due to the volatility parameter α. Now, we investigate the size property in the situation of high GARCH effect where $\alpha+\beta$ = 0.95. The results are in the following Table 6. The shows a serious over-rejection problem. However, in contrast to Tables 4 and 5, we see that when $\alpha+\beta$ = 0.95, the size distortion is more severe when the sample size increases and the seriousness of size distortion does not have an obvious relationship with the increase of α, which is a common characteristic when $\alpha+\beta$ is equal to 0.3 and 0.75. As Table 5 and Table 6 show serious size distortion, we will try to solve this over-rejection problem in our wavelet environment in the next section.

Table 6. Size property for the Nonlinear D-F F test under GARCH (1, 1) with $\alpha+\beta$ =0.95; Case 2 and Order 1

T	α =0.35 β=0.6	α =0.4 β=0.55	α =0.45 β=0.5	α =0.5 β=0.45	α =0.55 β=0.4	α =0.6 β=0.35
25	0.0663	0.0701	0.0733	0.0743	0.0718	0.0799
50	0.0793	0.0806	0.0758	0.0795	0.0791	0.0781
100	0.0871	0.0884	0.0869	0.087	0.0829	0.0845
250	0.1029	0.1005	0.1003	0.0999	0.092	0.091
500	0.1077	0.1138	0.1013	0.0933	0.0975	0.0938

Wavelet improvement of size distortion under GARCH error

In this section we use the wavelet method to solve the problem of over-rejection under GARCH error. The process is simple: first we generate a new table of critical values where the DGP is the first level boundary wavelet scale coefficients get by Maximal Overlap Discrete Wavelet Transform (MODWT). Next, we apply the test using these scale coefficients instead of the original series. The logic behind this method is that, after wavelet decomposition, the wavelet's high frequency coefficients which contain short time volatility information brought by GARCH (1, 1) error are counted off. Those scale coefficients contain all the non stationary information when the original time series follows a unit root process, while the scale coefficients are still stationary when the original time series is stationary. Thus when conducting the unit root test, we use the scale coefficients $V_t = \sum_{l=0}^{L-1} g_l y_{t-l \bmod T}$ instead of the original data, while g_l is the scaling filter satisfying: $\sum_l g_l = 1,\ \sum_l g_l^2 = 1/2,\ \sum_l g_l g_{l+2*n} = 0$

Assumption 2: *$\{w_t\}$ is a linear process which can be defined as follows:*

$w_t = \psi(L)u_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}, \psi(1) \neq 0$ *, and* $\sum_{j=0}^{\infty} j\left|\psi_j\right| < \infty$.

Theorem 3: *Under Assumptions 1 and 2, the asymptotical distributions of the Nonlinear Dickey-Fuller F test statistics will not be influenced when we use wavelet scale coefficients $V_t = \sum_{l=0}^{L-1} g_l y_{t-l \bmod T}$ instead of original series y_t with $y_t = y_{t-1} + u_t$, where u_t fulfills Assumption 1.*

For a detailed proof of *Theorem 3,* please see the Appendix.

For small sample distributions, the Monte Carlo experiment gives a new critical value table as follows:

Table 7. Critical values for the wavelet improved Nonlinear D-F F test; Case 2 and Order 1.

T	99%	97.5%	95%	90%	10%	5%	2.5%	1%
25	0.1370	0.2244	0.3300	0.4931	3.8075	4.6933	5.5415	6.5950
50	0.1728	0.2685	0.3906	0.5634	3.9543	4.7773	5.6739	6.9212
100	0.2752	0.4195	0.5906	0.8380	4.5434	5.3926	6.2244	7.3566
250	0.5901	0.8151	1.0061	1.3096	5.3819	6.2949	7.1325	8.2737
500	0.9199	1.1309	1.3491	1.6709	5.8924	6.8326	7.8232	8.9899

We can see that as the sample size increase, the critical values will approach the one we obtain from Table 1, which also implies that the distribution will not be influenced asymptotically. We now investigate the size property of wavelet improved test when the nominal size is 5%, with $u_t \sim n.i.d.(0,1)$.The result is:

Table 8. Size property for the wavelet improved Nonlinear D-F F test; Case 2 and Order 1

T	25	50	100	250	500
Size (T)	0.0464	0.0492	0.0525	0.0468	0.0496

Table 8 shows, at 95% confidence level, that the wavelet improved Nonlinear Dickey-Fuller F test is unbiased when error term $u_t \sim n.i.d.(0,1)$. Next we investigate the size property of the wavelet improved test when the original DGP suffered from GARCH (1, 1) error. As Section IV shows that the distortion is not serious when $\alpha + \beta = 0.3$, we study the situations when $\alpha + \beta = 0.75$ and 0.95.

Table 9. Size property for Nonlinear D-F F test in wavelet under GARCH (1, 1) with $\alpha + \beta$ =0.75; Case 2 and Order 1.

T	α =0.25 β=0.5	α =0.3 β=0.45	α =0.35 β=0.4	α =0.4 β=0.35	α =0.45 β=0.3	α =0.5 β=0.25
25	0.0499	0.0588	0.0548	0.0619	0.0604	0.0569
50	0.054	0.0624	0.0644	0.0618	0.063	0.0669
100	0.0635	00.0672	0.0724	0.0738	0.0729	0.0762
250	0.0632	0.0712	0.0688	0.0722	0.0705	0.0763
500	0.0539	0.0588	0.0619	0.0634	0.0684	0.067

We compare Table 9 to Table 5; where no wavelet method is applied, we see that although there are only a few unbiased size results in Table 9, the over-rejection problem get improved in each grid. Moreover, the size property of the wavelet improved test under GARCH (1, 1) error with $\alpha + \beta$ =0.95 is as follows:

Table 10. Size property for Nonlinear D-F F test in wavelet under GARCH (1, 1) with $\alpha + \beta$ =0.95; Case 2 and Order 1

T	α =0.35 β=0.6	α =0.4 β=0.55	α =0.45 β=0.5	α =0.5 β=0.45	α =0.55 β=0.4	α =0.6 β=0.35
25	0.0463	0.438	0.0452	0.0453	0.0484	0.0494
50	0.0524	0.0533	0.0529	0.0467	0.0516	0.0472
100	0.0513	0.0532	0.054	0.053	0.0491	0.058
250	0.0582	0.0608	0.0594	0.0604	0.0574	0.0562
500	0.0738	0.0719	0.0743	0.0639	0.0637	0.0603

The over-rejection problem of the size is obviously improved when Table 10 is compared to Table 6, especially when the sample size is small, and where the wavelet improved size is almost unbiased and lies between the 95% confidence interval. However, when the sample size increases, there is still some size distortion, which may due to the aggregate influence of the GARCH effect in large sample sizes.

Conclusions

In this paper we first propose a nonlinear Dickey-Fuller F test against LSTAR (1) model with time as the transition variable. The asymptotic distribution of the Nonlinear Dickey-Fuller F test statistic is derived while distributions of finite samples are obtained by Monte Carlo simulations. The size of the test statistics is unbiased and the power shows good property in larger samples. We also use an empirical example to compare the nonlinear Dickey-Fuller F with traditional Dickey-Fuller F test.

Technically speaking, the Nonlinear Dickey-Fuller F test is not very innovative as it is mainly an addition of the nonlinear Dickey-Fuller ρ and t test proposed by He and Sandberg (2006). The main point of this paper is to show that our test suffered from serious size distortion under medium and high GARCH (1, 1) error. To resolve the problem, we use wavelet method as the wavelet scale coefficient can maintain the unit root information while count the GARCH effort off. We show that by using the wavelet method, the asymptotic distribution is not influenced, while the over-rejection problem in small sample size is improved.

References

Andrew A.W. (1986). "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory,* Vol. 2, No. 1, pp. 107-131.

Arango L.E. and Gonzalez A. (2001). "Some Evidence of Smooth Transition Nonlinearity in Colombian Inflation," *Applied Economics*, Vol. 33, No.2, pp. 155-162.

Bera, A. K. and Higgins M. L. (1993). "ARCH Models: Properties, Estimation and Testing," *Journal of Economic Surveys*, Vol.7, No. 4, pp.307-366.

Berben R.P. and Van Dijk. (1999). "Unit Roots and Asymetric Adjustment - a Reassessment," *Econometric Institute Report*, Vol. 101.

Bhargava A. (1986*).* "On the Theory of Testing for Unit Roots in Observed Time Series," *The Review of Economic Studies*, Vol. 53, No. 3 (Jul., 1986), pp. 369-384.

Bollerslev T. (1986). "Generalized Autorregressive Conditional Heteroskedasticity," *Journal of Econometrics,* Vol. 31, pp.307-327.

Bollerslev. and Engle. (1993). "Common Persistence in Conditional Variances," *Econometrica,* Vol.61,pp.167-186.

Caner M. and Hansen B.E. (2001). "Threshold Autoregression with a Unit Root," *Econometrica,* Econometric Society, Vol. 69(6), pp. 1555-1596.

Charles A. and Darné O. (2008). "A Note on Unit Root Tests and GARCH Errors: A Simulation Experiment," *Communications in Statistics - Simulation and Computation,* Vol. 37, Issue 2, pp. 314 – 319.

Chiann C. and Morettin P.A. (1998). "A Wavelet Analysis for Time Series," *Journal of Nonparametric Statistics,* Vol. 10, pp.1-46.

Cook S. (2006). "The Robustness of Modified Unit Root Tests in the Presence of GARCH," *Quantitative Finance*, Vol. 6, issue 4, pp.359-363.

Daubechies I. (1992). *Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics.* Vol. 61, SIAM, Philadelphia.

Dickey D.A. and Fuller W.A. (1979). "Distributions of the Estimators for Autoregressive Time Series with a Unit root," *Journal of American Statistical Association*, Vol. 74, pp.427–431.

Eklund B. (2003). "Testing the Unit Root Hypothesis against the Logistic Smooth Transition Autoregressive Model," Working Paper Series in Economics and Finance from Stockholm School of Economics, No. 546.

Enders, Walter and Granger C.W.J. (1998). "Unit-Root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates," *Journal of Business & Economic Statistics*, Vol. 16(3), pp. 304-311.

Engle R.F. (1982). "Autorregressive Conditional Heteroskedasticity with Estimates of United Kingdom Inflation," *Econometrica,* Vol. 50, pp.987-1008.

Engle R.F. (2001). "The Use of ARCH/GARCH Models in Applied Econometrics," *The Journal of Economic Persipectives*. Vol. 15, No. 4, pp. 157-168.

Fan Y.Q. and Gençay R. (2006). "Unit Root and Cointegration Tests with Wavelets," Canadian Econometric Study Group Meeting, McGill University, Montreal, September 29-30.

Gençay R., Selçuk F. and Whitcher, B. (2001b). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics.* Academic Press, San Diego.

Granger C.W.J. and Teräsvirta T. (1993). *Modelling Nonlinear Economic Relationships.* Oxford University Press, New York.

Gregory C.C. (1960**).** "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica*. Vol. 28, No.3, pp.591-605.

Grossmann A. and Morelet J. (1984). "Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shap," *SIAM Journal on Mathematical Analysis,* Vol. 15, pp. 723-736.

Hall A.D.,Skalin J. and Terasvirta T.(2001). "A Nonlinear Time Series Model of El Niño,"*Environmental Modelling and Software*, Vol.16, Number 2, pp. 139-146.

Haggan V., Heravi S.M, and Prietsley N.B. (1984) "A Study of the Application of State Dependent Models in Non-linear Time Series Analysis," *Journal of the Time Series Analysis,* Vol.5, pp. 69-102.

Hamilto J.D. (1994). *Time series analysis*, Princeton University Press, Princeton NJ.

Harvey D.I. and Mills T.C..(2002). "Unit Roots and Double Smooth Transitions," *Journal of Applied Statistics,* Vol. 29, pp. 675-683.

He C. and Strandberg, R. (2006). "Dickey-Fuller Type of Tests against Non-linear Dynamic Models," *Oxford Bulletin of Economics & Statistics*, Vol. 68, No. s1, 2006, pp. 835-86.

Kapetanios G., Shin Y. and Snell A. (2003). "Testing for a Unit Root in the Nonlinear STAR Framework," *Journal of Econometrics,* Elsevier, Vol. 112(2), pp. 359-379.

Kim and Schmidt P. (1993). "Unit Root Tests with Conditional Heteroskedasticity," *Journal of Econometrics* 59 pp. 287–300.

Leybourne S., Newbold P. and Vougas D. (1998). "Unit Roots and Smooth Transitions," *Journal of Time Series Analysis,* Vol.19, pp. 83-92.

Li W.K., Ling S.Q. and Michael M. (2002). "Recent Theoretical Results for Time Series Models with GARCH Errors," *Journal of Economic Surveys*, Vol. 16 Issue 3, pp. 245-269.

Ling S.Q. and Li W. K. (2003). "Asymptotic Inference for Unit root Processes with GARCH (1, 1) errors*," Econometric Theory*, Vol. 19, Issue 04, pp. 541-564.

Ling S.Q. and Li, W.K. (1998). "Limiting Distributions of Maximum Likelihood Estimators for Unstable Autoregressive Moving-Average Time Series with General Autoregressive Heteroscedastic Errors," *Annals of Statistics*, Vol. 26 (1), pp.84-125.

Ling S.Q., Li W. K. and McAleer M. (2003). "Estimation and Testing for Unit Root Processes with GARCH (1, 1) Errors: Theory and Monte Carlo Evidence," *Econometric Reviews,* Vol. 22, Issue 2, pp. 179 – 202.

Luukkonen R.,Saikkonen P. and Teräsvirta. T. (1988). "Testing Linearity against Smooth Transition Autoregressive Models," *Biometrika,* Vol. 75, pp. 491-499.

Mallat S.G.. (1989). "A Theory for Multiresolution Signal Decomposition: the Waveletrepresentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions*. Vol. 11, Issue. 7, pp. 674-693.

Marrocu E. (2005). "An Investigation of the Effects of Data Transformation on Nonlinearity," *Empirical Economics,* Vol. 31, No. 4, pp. 801-820.

Neumann M. H. (1996). "Spectral Density Estimation via Nonlinear Wavelet Methods for Stationary Non-Gaussian Time Series," *Journal of Time Series Analysis,* Vol. 17, pp. 601–33.

Pantula S. G. (1988). "Estimation of Autoregressive Models with ARCH Errors," *The Indian Journal of Statistics,* Vol. 50, pp.119-138.

Percival D.B. (1995). "On Estimation of the Wavelet Variance," *Biometrika***,** Vol. 82, No. 3, pp. 619-631.

Percival D.B. and Walden A.T. (2000). *Wavelet Methods for Time Series Analysis.* Cambridge Press, Cambridge.

Peters T.A. and Veloce W. (1988). "Robustness of Unit Root Tests in ARMA Models with Generalized ARCH Errors," Unpublished manuscript, Brock University.

Phillips P.C.B. (1987). "Time Series Regression with a Unit Root," *Econometrica*, Vol. 55, pp. 277-301.

Phillips P.C.B. and Perron P. (1988). "Testing for Unit Root in Time Series Regression," *Biometrika,* Vol. 75, pp. 335-346.

Poon S.H. and Granger C.W.J. (2003). "Forecasting Volatility in Financial Markets: a Review," *Journal of Economic Literature,* Vol. 41, pp. 478-539.

Schleicher C. (2002). "An Introduction to Wavelets for Economists," Working paper 2002-3.

Seo B. (1999). "Distribution Theory for Unit Root Tests with Conditional Heteroskedasticity," *Journal of Econometrics*, Vol. 91, Issue 1, pp. 113-144.

Simon M. P. (1999). "Nonlinear Time Series Modelling: an Introduction," *Journal of Economic Surveys*, Blackwell Publishing, Vol. 13(5), pp. 505-528.

Sjölander P. (2007). "A new Test for Simultaneous Estimation of Unit Roots and GARCH Risk in the Presence of Stationary Conditional Heteroscedasticity Disturbance," PhD Thesis, Jönköping International Business School.

Skalin J. and Teräsvirta T. (2002). "Modeling Asymmetries and Moving Equilibrium in Unemployment Rates," *Macroeconomic Dynamics,* Vol. 6, pp. 202-241.

Teräsvirta T. (1994). "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models,"*Journal of the American Statistical Association,*Vol. 89, pp. 208-218.

Teräsvirta T., Van Dilk. and Medeiros M. C. (2003). "Smooth Transition Autoregressions, Neural networks, and Linear Models in Forecasting Macroeconomic Time Series: A re-examination," *International Journal of Forecasting*, 2005, pp. 755-774.

Tong H. (1990). *Nonlinar time series*. Oxford Science Publicatons.

Wahlstrom S. (2004). "Comparing Forecasts form LSTAR and Linear Autoregressive Models".

Valens C. (1999). "A Really Friendly Guide to Wavelets".

Van Dijk, Franses P.H. and Lucas A. (1999). "Testing for ARCH in the presence of additive outliers," *Journal of Applied Econometrics,* Vol. 14, pp. 539-562.

Vidakovic B. (1999). *Statistical Modeling by Wavelets.* New York: John Wiley and Sons.

Appendix

Merging procedure to get the auxiliary regressions

Here we only illustrate the procedure in Case and & Order 1, as the other cases calculate in the same way. For Case 2 the original LSTAR(1) model with drift is:

Case 2: $y_t = \pi_{10} + \pi_{11} y_{t-1} + (\pi_{20} + \pi_{21} y_{t-1}) F(t, \gamma, c) + u_t$.

(2)

The order 1 Taylor expansion of the transition function is:

$$F_1(t; \gamma, c) = \frac{\gamma(t-c)}{4} + o(\gamma). \qquad (3)$$

Thus substitute equation (3) to (2), we have:

$$\begin{aligned} y_t &= \pi_{10} + \pi_{11} y_{t-1} + (\pi_{20} + \pi_{21} y_{t-1})\left(\frac{\gamma(t-c)}{4} + o(\gamma)\right) + u_t \\ &= \pi_{10} + \pi_{11} y_{t-1} + \frac{\pi_{20}\gamma}{4} t + \frac{\pi_{21}\gamma}{4} y_{t-1} t - \frac{\pi_{20}\gamma c}{4} - \frac{\pi_{21}\gamma c}{4} y_{t-1} + o(\gamma) + u_t \\ &= \underbrace{(\pi_{10} - \frac{\pi_{20}\gamma c}{4})}_{\lambda_{10}} + \underbrace{\frac{\pi_{20}\gamma}{4}}_{\lambda_{11}} t + \underbrace{(\pi_{11} - \frac{\pi_{21}\gamma c}{4})}_{\varphi_{10}} y_{t-1} + \underbrace{\frac{\pi_{21}\gamma}{4}}_{\varphi_{11}} y_{t-1} t + o(\gamma) + u_t \end{aligned}$$

Proofs of Theorems

To prove *Theorem 2,* we use *Lemma A1* given below:

Lemma A1. If $\{u_t\}_{t=1}^{\infty}$ satisfy *Assumption1*, $\{v_t\}_{t=1}^{\infty}$ satisfy *Assumtpin2*, and $\xi_t = \xi_{t-1} + u_t$ with $P(\xi_0 = 0) = 1$, then as T→∞

$$T^{-(p+\frac{q}{2}+1)} \sum_{t=1}^{T} t^p \xi_{t-1}^q \xrightarrow{d} \lambda^q \int_0^1 r^p W(r)^q dr$$

$$T^{-(p+1)} \sum_{l=0}^{L-1} t^p v_t v_{t-h} \xrightarrow{a.s} \frac{\gamma_h}{p+1}$$

$$T^{-(p+1/2)} \sum_{t=0}^{T} t^p v_{t-h} \xrightarrow{d} \lambda W(1) - p\lambda \int_0^1 r^{p-1} W(r) dr$$

$$T^{-(v+\frac{1}{2})} \sum_{t=1}^{T} t^v u_t \xrightarrow{d} \sigma_u W(1) - v\sigma_u \int_0^1 r^{v-1} W(r) dr$$

$$T^{-(p+1)} \sum_{t=1}^{T} t^p \xi_{t-1} u_t \xrightarrow{d} \frac{\lambda \sigma_u \left(W(1)^2 - p\int_0^1 r^{p-1} W(r)^2 dr - \frac{1}{p+1} \right)}{2}$$

$$T^{-(p+1)} \sum_{t=1}^{T} t^p \xi_{t-1} v_{t-h} \xrightarrow{d} \begin{cases} (a): \dfrac{\lambda^2 W(1)^2 - p\lambda^2 \int_0^1 r^{p-1} W(r)^2 dr - \frac{\gamma_0}{p+1}}{2}, h = 0 \\ (b): \dfrac{\lambda^2 W(1)^2 - p\lambda^2 \int_0^1 r^{p-1} W(r)^2 dr - \frac{\gamma_0}{p+1}}{2} + \dfrac{\sum_{s=0}^{h-1} \gamma_s}{p+1}, h > 0 \end{cases}$$

$$T^{-(p+1/2)}\sum_{l=0}^{L-1} t^p u_t v_{t-h} \xrightarrow{d} N(0, \frac{\gamma_0 \sigma_u^2}{p+1}), h > 0$$

Proof of *Lemma A1* please refer to He and Sandberg (2006)

Proof of *Theorem 2*

From OLS, we have

$$\hat{\psi}_m - \psi_m = \left[\sum_{t=1}^{T} x_{mt} x_{mt}^{'}\right]^{-1} \left[\sum_{t=1}^{T} x_{mt} u_{mt}\right]$$

$$\Upsilon_m (\hat{\psi}_m - \psi_m) = \left\{\Upsilon_m^{-1} \left[\sum_{t=1}^{T} x_{mt} x_{mt}^{'}\right] \Upsilon_m^{-1}\right\}^{-1} \left\{\Upsilon_m^{-1} \sum_{t=1}^{T} x_{mt} u_{mt}\right\}$$

As

$$\Upsilon_m^{-1} \left[\sum_{t=1}^{T} x_{mt} x_{mt}^{'}\right] \Upsilon_m^{-1} = \begin{bmatrix} \left[T^{-(i+j-1)} \sum_{t=1}^{T} t^{i+j-2}\right]_{(m+1)*(m+1)} & \left[T^{-(i+j-\frac{1}{2})} \sum_{t=1}^{T} t^{i+j-2} y_{t-1}\right]_{(m+1)*(m+1)} \\ \left[T^{-(i+j-\frac{1}{2})} \sum_{t=1}^{T} t^{i+j-2} y_{t-1}\right]_{(m+1)*(m+1)} & \left[T^{-(i+j)} \sum_{t=1}^{T} t^{i+j-2} y_{t-1}^2\right]_{(m+1)*(m+1)} \end{bmatrix}$$

$$= \begin{bmatrix} \left[\alpha_{ij}\right]_{(m+1)*(m+1)} & \left[\beta_{ij}\right]_{(m+1)*(m+1)} \\ \left[\beta_{ij}\right]_{(m+1)*(m+1)} & \left[\delta_{ij}\right]_{(m+1)*(m+1)} \end{bmatrix}$$

$$\beta_{ij} = T^{-(i+j-\frac{1}{2})} \sum_{t=1}^{T} t^{i+j-2} y_{t-1} = T^{-(i+j-2+\frac{1}{2}+1)} \sum_{t=1}^{T} t^{i+j-2} y_{t-1} \xrightarrow{d} \sigma_u \int_0^1 r^{i+j-2} W(r) dr = \sigma_u b_{ij}$$

(From *Lemma A1* where $p = i + j - 2, q = 1, \lambda = \sigma_u$)

$$\delta_{ij} = T^{-(i+j)} \sum_{t=1}^{T} t^{i+j-2} y_{t-1}^2 = T^{-(i+j-2+\frac{1}{2}+1)} \sum_{t=1}^{T} t^{i+j-2} y_{t-1}^2 \xrightarrow{d} \sigma_u^2 \int_0^1 r^{i+j-2} W(r)^2 dr = \sigma_u^2 c_{ij}$$

(From *Lemma A1* where $p = i + j - 2, q = 2, \lambda = \sigma_u$)

From the above equations, we can prove that

$$\Upsilon_m^{-1} \left[\sum_{t=1}^{T} x_{mt} x_{mt}^{'}\right] \Upsilon_m^{-1} \xrightarrow{d} \Psi_m$$

As

$$\Upsilon_m^{-1} \left[\sum_{t=1}^{T} x_{mt} u_{mt}\right] = \begin{bmatrix} \left[T^{-(i-\frac{1}{2})} \sum_{t=1}^{T} t^{i-1} u_t\right]_{(m+1)*1} \\ \left[T^{-i} \sum_{t=1}^{T} t^{i-1} y_{t-1} u_t\right]_{(m+1)*1} \end{bmatrix} = \begin{bmatrix} \left[\eta_i\right]_{(m+1)*1} \\ \left[\theta_i\right]_{(m+1)*1} \end{bmatrix}$$

$$\eta_i = T^{-(i-\frac{1}{2})}\sum_{t=1}^{T} t^{i-1}u_t = T^{-(i-1+\frac{1}{2})}\sum_{t=1}^{T} t^{i-1}u_t \xrightarrow{d} \sigma_u\left(W(1)-(i-1)\int_0^1 r^{i-2}W(r)dr\right) = \sigma_u d_i$$

(From *Lemma A1* where $\nu = i-1$)

$$\theta_i = T^{-i}\sum_{t=1}^{T} t^{i-1}y_{t-1}u_t = T^{-(i-1+1)}\sum_{t=1}^{T} t^{i-1}y_{t-1}u_t \xrightarrow{d} \frac{\sigma_u^2\left(W(1)-(i-1)\int_0^1 r^{i-2}W(r)^2 dr - \frac{1}{i}\right)}{2} = \sigma_u^2 e_i$$

(From *Lemma A1* where $p = i-1, \lambda = \sigma_u$)

From the above equations, we can prove that

$$\Upsilon_m^{-1}\left[\sum_{t=1}^{T} x_{mt}u_{mt}\right] \xrightarrow{d} \Pi_m$$

From the above, we prove that

$$\Upsilon_m(\hat{\psi}_m - \psi_m) = \left\{\Upsilon_m^{-1}\left[\sum_{t=1}^{T} x_{mt}x_{mt}^{'}\right]\Upsilon_m^{-1}\right\}^{-1}\left\{\Upsilon_m^{-1}\left[\sum_{t=1}^{T} x_{mt}u_{mt}\right]\right\} \xrightarrow{d} \Psi_m^{-1}\Pi_m$$

It is also easy to show that $s^2 \to \sigma_u^2$, then we get

$$s_m^2\Upsilon_m(\sum x_{mt}x_{mt}^{'})^{-1}\Upsilon_m \xrightarrow{L} \sigma_u{}^2\Psi_m^{-1}$$

Then *Theorem 2* is proved

As *Theorem 1*, we can combine the character of partitioned matrices and it is easy to get proved

Proof of *Theorem 3*

The wavelet scale coefficient after MODWT transform of original DGP is $V_t = \sum_{l=0}^{L-1} g_l y_{t-l \bmod T}$

where $\sum_{l=0}^{L-1} g_l = 1$ and $y_t = y_{t-1} + u_t, u_t$ fulfill *Assumption 1*. As we are proving the asymptotic property, we let $t > L-1$, Thus we have

$$V_t = \sum_{l=0}^{L-1} g_l y_{t-l} = (\sum_{l=0}^{L-1} g_l)y_{t-(L-1)} + \{(\sum_{l=0}^{L-2} g_l)u_{t-(L-2)} + (\sum_{l=0}^{L-3} g_l)u_{t-(L-3)} + ... + g_0 u_t\} = y_{t+1-L} + w_t,$$

with w_t fulfill *Assumption 2.*

Let $\varsigma_{ij} = T^{-(i+j-1/2)}\sum_{t=0}^{T} t^{i+j-2}V_{t-1} = T^{-(i+j-1/2)}\sum_{t=0}^{T} t^{i+j-2}(y_{t-L} + w_{t-1})$

As From *Lemma A1,* we have $\sum_{t=0}^{T} t^n w_{t-1} = O_p(T^{n+1/2})$, and we have: $T^{-(i+j-1/2)}\sum_{t=0}^{T} t^{i+j-2}w_{t-1}$

$= T^{-(n+3/2)}\sum_{t=0}^{T} t^n w_t \xrightarrow{d} 0$, thus ς_{ij} converge to the same distribution of β_{ij}

Let $\vartheta_{ij} = T^{-(i+j)}\sum_{t=0}^{T} t^{i+j-2}V_{t-1}^2 \; = T^{-(i+j)}\sum_{t=0}^{T} t^{i+j-2}(y_{t-L} + w_{t-1})^2$

$$= T^{-(i+j)}\sum_{t=0}^{T} t^{i+j-2}(y_{t-L}^2 + 2y_{t-L}w_{t-1} + w_{t-1}^2)$$

As From *Lemma A1,* we have $\sum_{t=0}^{T} t^n y_{t-1} w_{t-h} = O_p(T^{n+1}), \sum_{t=0}^{T} t^n w_{t-h}^2 = O_p(T^{n+1})$, thus ϑ_{ij} converge to the same distribution of δ_{ij}

Let $\kappa_i = T^{-i} \sum_{t=0}^{T} t^{i-1} V_{t-1} u_t = T^{-i} \sum_{t=0}^{T} t^{i-1} (y_{t-L} + w_{t-1}) u_t$

As from *Lemma A1,* we have $\sum_{t=0}^{T} t^n w_{t-1} u_t = O_p(T^{n+1/2})$, thus κ_i converge to the same distribution of θ_i .

Therefore *Theorem 3* is proven.