UNIVERSITÉ PARIS 1 – PANTHÉON - SORBONNE –

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers

A test for a new modelling : The Univariate MT-STAR Model

Peter Martey ADDO, Monica BILLIO, Dominique GUEGAN

2011.83

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://centredeconomiesorbonne.univ-paris1.fr/bandeau-haut/documents-de-travail/
ISSN : 1955-611X

A test for a new modelling : The Univariate MT-STAR Model

Peter Martey ADDO*

Université Paris1 Panthéon-Sorbonne, MSE-CES UMR8174 , 106-113 boulevard de l'hopital, 75013, Paris, France, email: pkaddo2000@yahoo.com

Monica BILLIO

Department of Economics, Ca'Foscari University of Venice, 30121, Venice, Italy, email: billio@unive.it

Dominique GUÉGAN

Université Paris1 Panthéon-Sorbonne, MSE-CES UMR8174 , 106-113 boulevard de l'hopital, 75013, Paris, France, email: dguegan@univ-paris1.fr

Abstract

In ESTAR models it is usually quite difficult to obtain parameter estimates, as it is discussed in the literature. The problem of properly distinguishing the transition function in relation to extreme parameter combinations often leads to getting strongly biased estimators. This paper proposes a new procedure to test for the unit root in a nonlinear framework, and contributes to the existing literature in three separate directions. First, we propose a new alternative model – the MT-STAR model – which has similar properties as the ESTAR model but reduces the effects of the identication problem and can also account for cases where the adjustment mechanism towards equilibrium is not symmetric. Second, we develop a testing procedure to detect the presence of a nonlinear stationary process by establishing the limiting non-standard asymptotic distributions of the proposed test-statistics. Finally, we perform Monte Carlo simulations to assess the small sample performance of the test and then to highlight its power gain over existing tests for a unit root.

Keywords: Nonlinearity, Exponential smooth transition autoregressive model, Unit roots, Globally stationary nonlinear processes, Monte Carlo simulations, Real exchange rates
JEL: C12, C22, C58

1. Introduction

Nonlinear time series models like smooth transition autoregressive (STAR) models (Terasvirta (1994)) have been successfully applied to explaining the behavior of various macro-economic time series, such as output, exchange rates, and (un)employment at different phases of the business cycle. In particular, Exponential Smooth Transition Autoregressive (ESTAR) models have

*Correspondence to: Peter Martey Addo, Université Paris1 Panthéon-Sorbonne, MSE-CES UMR8174 , 106-113 boulevard de l'hopital, 75013, Paris, France, Email: pkaddo2000@yahoo.com

Preprint submitted to Elsevier *December 13, 2011*

been used for modeling real exchange rates and real interest rates, where the presence of a unit root cannot be rejected using conventional linear unit root tests (Abuaf and Jorion (1990), Taylor M.P and Sarno (2001), Lothian and Taylor (1996), Engel (2000), O'Connell (1998), Dickey and Fuller (1979) and Phillips and Perron (1988)). This has lead[1] researchers to develop new testing procedures to detect the presence of nonlinear mean reversion against nonstationarity. Papers that develop such tests include Kapetanios G and Snell (2003), Pascalau (2007), Dieu-Hang and Kompas (2010), Bec and Carrasco (2004).

Recently, Kapetanios G and Snell (2003) propose a popular type test, denoted KSS test, to detect the presence of a particular kind of nonlinear stationary dynamics using an Exponential Smooth Transition Autoregressive(ESTAR) model, which is originally proposed by Haggan and Ozaki (1981). Unfortunately, the results of their simulation study indicated that the test does not have sophisticating power properties. In this paper, we are mainly concerned with the limitations of this exponential STAR (ESTAR) model and develop a new procedure to test for a unit root in a nonlinear framework. In particular, we develop a test, called QEM test, which enables us to distinguish between a linear nonstationary process and a specific new nonlinear globally stationary STAR process. In view of this objective, we introduce a new STAR model named MT-STAR model, which has similar properties as the ESTAR model but reduces the effects of the identication problem (Donauer S and Sibbertsen (2010)) and can also account for cases where the adjustment mechanism towards equilibrium is asymmetric.

The nonlinear structure of the ESTAR model leads to unidentified parameter occuring for certain combinations of the transition parameter and the error term variance. As discussed in the literature (Luukkonen R and Terasvirta (1988), Haggan and Ozaki (1981),Lutkepohl and Kratzig (2004)), it is usually difficult to obtain good parameter estimates in the ESTAR models. In the ESTAR setting, very small values of error term variance leads to an unidentified transition parameter, making nearly impossible to obtain a consistent estimate of transition parameter. In a seminal paper, Donauer S and Sibbertsen (2010) address this so-called identification problem of this ESTAR model − the problem of properly distinguishing the transition function in relation to extreme parameter combinations − and proposed an alternative model to the ESTAR model namely the TSTAR model. It is noteworthy that these two competitive models, ESTAR and TSTAR, can be very useful in modeling adjustment process, which is a growing part of the econometrics literature. However, these two models are limited by the assumption of symmetric adjustment in the transition to equilibrium. The adjustment towards equilibrium might not be the same for a given degree of positive or negative deviation from equilibrium. The main concern about the assumption of symmetric adjustment when using the ESTAR model, similarly the TSTAR, is that if the adjustment towards equilibrium is asymmetric, the alternative hypothesis in the ESTAR model will be mis-specified and test based on the ESTAR model might not be valid.

Given this limitation, this paper, we first introduce a new STAR model, the MT-STAR model, based on a more general new smooth transition function which nests the T-STAR model by Donauer S and Sibbertsen (2010) and also accounts for cases where the adjustment towards equilibrium is not necessarily symmetric. For practical purposes, we focus on a particular case of this new transition function and then develop a linearity test plus a unit root test for this new STAR model. Regarding the new unit root test, the pair of hypothesis is defined to be

[1]There are now reasons to believe that the exchange rate is not in fact driven by a linear stochastic process (Dumas (1992), Sercu P and Hulle (1995))

linear nonstationary process (unit root) under the null against nonlinear globally stationary MT-STAR process under the alternative. The results indicates that the new test statistic has more sophisticated power properties than some other unit root test proposed by Kapetanios G and Snell (2003), Pascalau (2007), Dieu-Hang and Kompas (2010) and Dickey and Fuller (1979).

The paper is organized as follows. Section 2 presents notations, existing unit root tests and drawbacks of using ESTAR model in modelling adjustment processes. After introducing new modelling, MT-STAR model, specification in more detail in section 3, we discussed in section 4 the linearity and unit root test associated with this new modelling. The non-standard limiting distribution of the QEM test statistic is derived and consistency of the test is proven. We also show that the limiting distribution remains unchanged if we account for potential serial correlation in the innovation terms by augmenting the test regression with lags of the dependent variable. In section 5, a Monte Carlo study is used to compare finite sample properties of the QEM test and existing alternative unit root tests under a variety of conditions. The new QEM test is correctly sized and quite often superior in terms of power. In particular, it exhibits higher power compared to KSS when the data generating process is nonlinear with asymmetric adjustment to the long-run equilibrium. Finally, we illustrate an empirical application to a monthly real efective exchange rate time series for the Euro. The results indicate that the new QEM test yields new evidence on the stationarity of the EU real effective exchange rate which suggests the validity of Purchasing Power Parity (PPP). Proofs are given in the appendix.

2. State of the Art

2.1. Overview of STAR models and Notation

In general, univariate STAR(p) models, $p \geq 1$, for a process $(y_t)_t$ are given for all t by

$$y_t = [\Psi\omega_t]\cdot[1 - G(y_{t-d}, \gamma, c)] + [\Theta\omega]\cdot G(y_{t-d}, \gamma, c) + \varepsilon_t \tag{1}$$

where $d \leq p$, $\Psi = (\psi_0, \psi_1, \cdots, \psi_p)$, $\Theta = (\vartheta_0, \vartheta_1, \cdots, \vartheta_p)$, and $\omega_t = (1, y_{t-1}, \cdots, y_{t-p})'$. The equation (1) can be reparametrized as:

$$y_t = [\Psi\omega_t] + [\Phi\omega_t]\cdot G(y_{t-d}, \gamma, c) + \varepsilon_t, \quad t \geq 1, \tag{2}$$

where $\Phi = (\varphi_0, \varphi_1, \cdots, \varphi_p) = (\vartheta_0 - \psi_0, \vartheta_1 - \psi_1, \cdots, \vartheta_p - \psi_p)$. The process $(\varepsilon_t)_t$ assumed to be a martingale difference sequence with respect to the history of the time series up to time $t-1$, denoted as $\Omega_{t-1} = \{y_{t-1}, \cdots, y_{t-p}\}$, i.e., $E[\varepsilon_t|\Omega_{t-1}] = 0$. For computational reasons, we restrict[2] the conditional variance of the process $(\varepsilon_t)_t$ as constant, $E[\varepsilon_t^2|\Omega_{t-1}] = \sigma^2$.

The transition function $G(\cdot\,; \gamma, c) : \mathbb{R} \to [0, 1]$ which models the regime-switching behavior depends on three parameters: γ which controls the degree of nonlinearity, the threshold c, the delay d which can be chosen to maximize goodness of fit over $d = \{1, 2, \cdots, d_{max}\}$ (Kapetanios G and Snell (2003)). In practice this last parameter is often chosen equal to 1, therefore $y_{t-d} = y_{t-1}$ in (1) and (2). In $\Psi(\cdot)$ and $\Phi(\cdot)$ the parameter p is generally determined using AIC. Now, in literature two main functions are used:

[2] Extension of the STAR model could be to allow for possibly asymmetric autoregressive conditional heteroscedasticity.

- The Logistic function

$$G(y_{t-d};\gamma,c) = (1+\exp\{-\gamma(y_{t-d}-c)\})^{-1}, c\in\mathbb{R}, \gamma>0, \quad (3)$$

the resultant model is called the logistic STAR [LSTAR] model.

- The exponential function

$$G(y_{t-d};\gamma,c) = 1-\exp\{-\gamma(y_{t-d}-c)^2\}, c\in\mathbb{R}, \gamma>0, \quad (4)$$

the resultant model called the ESTAR model.

2.2. An Overview on Unit Root Tests

We present an overview of some existing unit tests against STAR nonlinear alternatives. We will present the ADF-type unit root test against ESTAR named KSS test by Kapetanios G and Snell (2003), W_{nl} tests by Dieu-Hang and Kompas (2010) and the Modified Wald Type test,τ, (Kruse (2008)). Other tests are the F_{NL} test by Pascalau (2007) and the famous Dickey and Fuller (1979) test. We will consider these tests under the Monte Carlo Study, in order to compare their performance with the new test we propose in the next section.

1. The KSS test. This test has been established for testing the ESTAR model defined as:

$$\Delta y_t = \alpha y_{t-1} + \varphi y_{t-1}(1-e^{(-\gamma(y_{t-1}-c)^2)}) + \varepsilon_t \quad (5)$$

where $\varepsilon_t \sim iid(0,\sigma^2)$. Kapetanios G and Snell (2003) show that the ESTAR model under the restriction $\alpha = 0$, is globally stationary if $-2 < \varphi < 0$ although it is locally non-stationary in the sense that it contains a partial unit root when $y_{t-1} = c$ holds. More specifically, make the restriction $c = 0$ and replace the model(5) by

$$\Delta y_t = \varphi y_{t-1}(1-e^{(-\gamma y_{t-1}^2)}) + \varepsilon_t \quad . \quad (6)$$

Applying a first-order Taylor approximation to (5) leads to the auxiliary regression

$$\Delta y_t = \beta_1 y_{t-1}^3 + u_t \ where \quad \beta_1 = \varphi\gamma \quad (7)$$

with u_t being a noise term depending on ε_t, φ and the remainder of the Taylor expansion. The unit root hypothesis against globally stationary ESTAR model corresponds to

$$H_0 : \beta_1 = 0 \quad versus \quad H_1 : \beta_1 < 0 \quad . \quad (8)$$

Given a sample $y_1, \cdots, y_T$, $\hat{\beta}_1$ is the OLS estimate of the auxiliary regression (7), and then the corresponding statistic for the test (8) is :

$$KSS \equiv \frac{\hat{\beta}_1}{\sqrt{Var(\hat{\beta}_1)}} = \frac{\sum_{t=1}^T y_{t-1}^3 \Delta y_t}{\sqrt{\hat{\sigma}^2 \sum_{t=1}^T y_{t-1}^6}} \quad (9)$$

where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^T(\Delta y_t - \hat{\beta}_1 y_{t-1}^3)^2$ is the usual estimator of the error variance.
The limiting distribution of the KSS statistic under the null is then given by

$$KSS \xrightarrow{d} \frac{\frac{1}{4}W(1)^4 - \frac{3}{2}\int_0^1 W(r)^2 dr}{(\int_0^1 W(r)^6 dr)^{1/2}}$$

where $W(r)$ be the Brownian motion defined on $r \in [0,1]$

2. The τ test. This test has been built to test another approximation of (5), where $\alpha = 0$ and $c \neq 0$. Then we get

$$\Delta y_t = \beta_1 y_{t-1}^3 + \beta_2^2 y_{t-1} + u_t \tag{10}$$

where $\beta_1 = \gamma\varphi$ and $\beta_2 = -2c\gamma\varphi$. The following test called τ (Kruse (2008)) corresponds to the following assumptions:

$$H_0 : \beta_1 = \beta_2 = 0 \quad against \quad H_1 : \beta_1 < 0, \beta_2 \neq 0 \tag{11}$$

The standard Wald test statistics for this hypothesis is inappropriate since in this testing problem one parameter is one-sided under H_1 while the other one is two-sided (see Kruse (2008)), thus a modified Wald test is built upon the one-sided parameter (β_1) and the transformed two-sided parameter, say $\beta_2^{\perp}$, that are stochastically independent (Abadir and Distaso (2007)). A simpler and more intuitive way to formulate the test statistic for (11) is

$$\tau = t^2_{\beta_2^{\perp}=0} + 1(\hat{\beta}_1 < 0) t^2_{\beta_1=0} \tag{12}$$

where $t^2_{\beta_2^{\perp}=0}$ is a squared t-statistic for the hypothesis $\beta_2^{\perp} = 0$ with $\beta_2^{\perp}$ being orthogonal to β_1 and $t^2_{\beta_1=0}$ is a squared t-statistic for the hypothesis $H_1 = 0$, the one-sidedness under H_1 is obtained by the multiplied indicator function (see Abadir and Distaso (2007), Kruse (2008) for details). The limiting distribution of the τ statistic (11) under the null is then given by

$$\tau \xrightarrow{d} \mathcal{A}(W(r)) + \mathcal{B}(W(r))$$

where $\mathcal{A}$ and $\mathcal{B}$ are functions[3] of the Brownian motion $W(r)$. (see Kapetanios G and Snell (2003))

3. The F_{NL} test. This test has been built to test unit root against nonlinear globally stationary LSTAR process (Pascalau (2007)) defined as

$$\Delta y_t = \phi y_{t-1}(1 + e^{-\gamma((z-c)^2)})^{-1} + \varepsilon_t$$

with the auxiliary regression

$$\Delta y_t = \beta_1 y_{t-1}^2 + \beta_2 y_{t-1}^3 + \beta_3 y_{t-1}^4 + u_t, \; where \quad u_t \sim i.i.d.(0, \sigma^2)$$

For testing unit root against nonlinear globally stationary LSTAR process, the author considers the pair of hypotheses;

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \; against \quad H_1 : \beta_1 + \beta_2 + \beta_3 < 0$$

---

[3]

$$\mathcal{A}(W(r)) = \left\{ \frac{(\frac{1}{3}W(1)^3 - \int_0^1 W dr)((\int_0^1 W^4 dr)^{\frac{3}{2}}(\int_0^1 W^6 dr) - (\int_0^1 W^4 dr)^{\frac{1}{2}}(\int_0^1 W^5 dr)^2)}{(\int_0^1 W^4 dr)^2(\int_0^1 W^6 dr) - (\int_0^1 W^4 dr)(\int_0^1 W^5 dr)^2} \right\}^2$$

$$\mathcal{B}(W(r)) = \frac{\left\{(\int_0^1 W^4 dr)(\frac{1}{4}W(1)^4 - \frac{3}{2}\int_0^1 W^2 dr) - (\frac{1}{3}W(1)^3 - \int_0^1 W dr)(\int_0^1 W^5 dr)\right\}^2}{(\int_0^1 W^4 dr)^2(\int_0^1 W^6 dr) - (\int_0^1 W^4 dr)(\int_0^1 W^5 dr)^2}$$

where we denote $\int_0^1 (W(r))^n dr = \int_0^1 W^n dr$

This is equivalent to testing:

$$H_0 : R\beta = r$$

against the alternative

$$H_1 : R\beta \neq r$$

with

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \beta = (\beta_1, \beta_2, \beta_3)' \, and \; r = (0,0,0)'.$$

The corresponding test statistic is

$$F_{NL} = (R(\hat{\beta} - \beta))' \Big[\hat{\sigma}^2 (R \sum_{t=1}^{T} x_t x_t')^{-1} R'\Big]^{-1} (R(\hat{\beta} - \beta)) \tag{13}$$

where $x_t = (y_{t-1}^2, y_{t-1}^3, y_{t-1}^4)'$ and $\hat{\sigma}^2$ is the variance of the above auxiliary regression.
The limiting distribution of the F_{NL} statistic under the null is $F_{NL} \longrightarrow_d v'Q^{-1}v \quad with,$

$$v = \begin{bmatrix} \frac{1}{3}W(1)^3 - \int_0^1 W(r)dr \\ \frac{1}{4}W(1)^4 - \frac{2}{3}\int_0^1 W(r)^2 dr \\ \frac{1}{5}W(1)^5 - 2\int_0^1 W(r)^3 dr \end{bmatrix}$$

and

$$Q = \begin{bmatrix} \int_0^1 W(r)^4 dr & \int_0^1 W(r)^5 dr & \int_0^1 W(r)^6 dr \\ \int_0^1 W(r)^5 dr & \int_0^1 W(r)^6 dr & \int_0^1 W(r)^7 dr \\ \int_0^1 W(r)^6 dr & \int_0^1 W(r)^7 dr & \int_0^1 W(r)^8 dr \end{bmatrix}$$

where $W(r)$ be the Brownian motion defined on $r \in [0,1]$ (see Pascalau (2007))

4. The W_{nl} test. This test is used to test an extension of model (5), proposed by Dieu-Hang and Kompas (2010), say

$$\Delta y_t = \varphi y_{t-1}(1 - e^{-\gamma((y_{t-1}-c)^2 + k(y_{t-1}-c))}) + \varepsilon_t \quad k \in \mathbb{R}$$

with the auxiliary regression

$$\Delta y_t = \beta_1 y_{t-1}^3 + \beta_2 y_{t-1}^2 + u_t, \; where \quad u_t \sim i.i.d.(0, \sigma^2)$$

where $\beta_1 = \varphi\gamma$ and $\beta_2 = \varphi\gamma k$. For testing linear unit root process against the nonlinear globally stationary M-ESTAR process, the authors consider the pair of hypotheses;

$$H_0 : \beta_1 = \beta_2 = 0 \quad against \quad H_1 : \; at\ least\ one\ \beta_i \neq 0 \quad i = 1,2 \quad . \tag{14}$$

They introduce the following Wald statistic

$$W_{nl} = (R(\hat{\beta} - \beta))' \Big[\hat{\sigma}^2 (R \sum_{t=1}^{T} x_t x_t')^{-1} R'\Big]^{-1} (R(\hat{\beta} - \beta)) \tag{15}$$

where $x_t = (y_{t-1}^3, y_{t-1}^2)'$,

$$R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \beta = (\beta_1, \beta_2)'$$

and $\hat{\sigma}^2$ is the variance of the above auxiliary regression.
The limiting distribution of the W_{nl} statistic under the null is

$$W_{nl} \longrightarrow_d \frac{(\int_0^1 W^2 dW)^2 \int_0^1 W^6 dr - 2\int_0^1 W^2 dW \int_0^1 W^3 dW \int_0^1 W^5 dr + (\int_0^1 W^3 dW)^2 \int_0^1 W^4 dr}{\int_0^1 W^4 dr \int_0^1 W^6 dr - (\int_0^1 W^5 dr)^2}$$

where we denote $\int_0^1 (W(r))^n dr = \int_0^1 W^n dr \quad \int_0^1 (W(r))^n dW = \int_0^1 W^n dW$ and W is the Brownian motion.

2.3. Testing for non-linearity

The nonlinear structure of the ESTAR model leads to the presence of unidentified parameters occuring for certain combinations of γ and σ^2 , the transition parameter and the error term variance, respectively. In the ESTAR setting, very small values of σ^2 leads to an unidentified γ, making it nearly impossible to obtain a consistent estimate of γ. Donauer S and Sibbertsen (2010) address this so-called identication problem of the ESTAR, the problem of properly distinguishing the transition function in relation to extreme parameter combinations, by showing that the variance of the conditional Maximum likelihood estimator $\hat{\gamma}$ tends to infinity as σ^2 vanishes (see Lemma 2.4 of). Thus proposed a new type of nonlinear model formulation named T-STAR using an alternative transition function to (4) which is

$$T(y_{t-d}, \gamma, c) = 1 - (1 + (y_{t-d} - c)^2)^{-\gamma}, \gamma > 0, \tag{16}$$

which shares same properties as (4) and also reduces the identication problem associated with the ESTAR models. In addition the authors propose a linearity and a unit root test for the new model. In this paper, we extend this last work, we first introduce a new model which takes into account asymmetric adjustments, and then develop a linearity and a unit root test for our model.

3. A new Model: The MT-STAR model

In this section, We extend the work of by introducing the possibility of asymmetric adjustment towards equilibrium. In view of this objective, we introduce a more general smooth transition function which nests the T-STAR model and also accounts for cases where the adjustment towards equilibrium is not necessarily symmetric.

3.1. A general MT-STAR(n,p) Model

We define the univariate MT-STAR(n, p) model of order p for a process $(y_t)_t$ to have the following representation:

$$\begin{cases} y_t = [\Psi\omega_t] \cdot [1 - G_n(y_{t-d}, \gamma, c)] + [\Theta\omega] \cdot G_n(y_{t-d}, \gamma, c) + \varepsilon_t \\ G_n(z; \gamma, c) = 1 - (1 + f_n(z; \gamma, c))^{-\gamma}, \quad n \in \mathbb{N} \\ f_n(z; \gamma, c) = (\prod_{i=1}^n (z - c_i))^2 + k(\prod_{i=1}^n (z - c_i)), \; c_1 \leq c_2 \leq \cdots \leq c_n, k \in \mathbb{R}. \end{cases} \tag{17}$$

where $G_n(\cdot;\gamma,c)$ is the general nth order smooth transition function, n is the degree of polynomial of the transition function, $d \leq p$, $\Psi = (\psi_0, \psi_1, \cdots, \psi_p)$, $\Theta = (\vartheta_0, \vartheta_1, \cdots, \vartheta_p)$, and $\omega_t = (1, y_{t-1}, \cdots, y_{t-p})'$. The equation (17) can be reparametrized as:

$$y_t = [\Psi\omega_t] + [\Phi\omega_t]\cdot G_n(y_{t-d}, \gamma, c) + \varepsilon_t, \quad t \geq 1, \tag{18}$$

where $\Phi = (\varphi_0, \varphi_1, \cdots, \varphi_p) = (\vartheta_0 - \psi_0, \vartheta_1 - \psi_1, \cdots, \vartheta_p - \psi_p)$. We called the model equation (18) a generalized MT-STAR model of order p, denoted MT-STAR (n, p).

Assuming $n = 1$ in equations (18) and denote $G(\cdot,\gamma,c) = G_1(\cdot,\gamma,c)$, we get the following model

$$\begin{cases} y_t = [\Psi\omega_t] + [\Phi\omega_t]\cdot G(y_{t-d}, \gamma, c) + \varepsilon_t \\ G(y_{t-d}, \gamma, c) = 1 - (1 + (y_{t-d} - c)^2 + k(y_{t-d} - c))^{-\gamma} \end{cases} \tag{19}$$

- The MT-STAR model (19) incorporates the component $k(y_{t-d} - c)$ to the transition function (16), which makes the functional form of $G(y_{t-d};\gamma,c)$ not necessarily symmetric.
  - For $k = 0$, the function $G(y_{t-d};\gamma,c)$ in (19) induces a nonmonotonic change which is symmetric around $y_{t-d} = c$. Suppose $\gamma \to \infty$ then $G(\cdot) \to 1 - \mathcal{I}_c$, where $\mathcal{I}_c$ is the indicator function at c, which corresponds to a single abrupt break only at $y_{t-d} = c$. This creates same behavior as the transition function $T(y_{t-d};\gamma,c)$ defined in (16) proposed by .
  - For $k \neq 0$, the transition function $G(y_{t-d};\gamma,c)$ is asymmetric around the equilibrium c of y_t, and we get the new modeling (19).
- In addition, the transition function $G(y_{t-d};\gamma,c)$ has the following interesting nice properties:
  - if $\gamma = 0$ then $G(y_{t-d};\gamma,c) = 0$ and we are back to the linear error correction framework, the process $(y_t)_t$ follows an AR modeling in our representation.
  - If $\gamma > 0$, $G(y_{t-d};\gamma,c)$ approximates to 0 when y_{t-1} is near c and approximates to 1 when y_{t-1} approaches $\pm\infty$. So apart from the property of symmetry, the transition function $G(y_{t-d};\gamma,c)$ has the same properties as function $T(y_{t-d};\gamma,c)$ defined in (16).

Hence, by not imposing the assumption of symmetry, the function $G(y_{t-d};\gamma,c)$ in the MT-STAR (19) is more general than the transition function $T(y_{t-d};\gamma,c)$ in the TSTAR model of Donauer S and Sibbertsen (2010). As a result, in modeling a nonlinear adjustment process, which we usually do not know in advance whether it is symmetric or not, using MT-STAR model would be more appropriate for applications than using the TSTAR and the ESTAR models whose definition appear too restrictive in the applications. The MT-STAR model can therefore be seen as a modification for the TSTAR and also as an alternative model to the M-ESTAR model (see Dieu-Hang and Kompas (2010)), applicable to the same situations. The existence and uniqueness of a stationary distribution for the process $\{y_t, t \geq 1\}$ (19) is guaranteed by geometric ergodicity (see R.L.Tweedie (1975), Tjøstheim (1986), Fan and Yao (2003)) as long it satifies the condition

$$|\psi_i + \varphi_i| < 1 \quad \forall i \in \{1, 2, \cdots, p\}$$

3.2. MT-STAR(1,1) modelling

For application reasons, We restrict to the univariate MT-STAR (1,1) model of order 1 with delay, $d = 1$, whose fair representation is

$$\begin{cases} y_t = \psi y_{t-1} + \varphi y_{t-1} G(y_{t-1};\gamma,c) + \varepsilon_t \\ G(y_{t-1};\gamma,c) = 1 - (1 + (y_{t-1} - c)^2 + k(y_{t-1} - c))^{-\gamma} \end{cases} . \tag{20}$$

This process $(y_t)_t$ is geometrically ergodic [4] as soon as $|\psi + \varphi| < 1$. We can now distinguish two classes of modellings with respect of the process $(\varepsilon_t)_t$

1. If $(\varepsilon_t)_t$ is a strong white noise $(0, \sigma^2)$, the model (20) can be reparameterised as

$$\Delta y_t = \beta y_{t-1} + \varphi y_{t-1} G(y_{t-1};\gamma,c) + \varepsilon_t \tag{21}$$

   where $\beta = \psi - 1$, $\Delta y_t = y_t - y_{t-1}$ and geometrically ergodic when $|\beta + \varphi| < 0$. If the smoothness parameter γ approaches zero, the MT-STAR model becomes a linear AR(1) model, i.e. $\Delta y_t = \beta y_{t-1} + \varepsilon_t$ that is stationary if $-2 < \beta < 0$.
2. If the process $(\varepsilon_t)_t$ follows an $AR(p)$ with the form :

$$\varepsilon_t = \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t. \tag{22}$$

   with $(u_t)_t$ a strong white noise $(0, \sigma^2)$ then the model (21) becomes

$$\Delta y_t = \beta y_{t-1} + \varphi y_{t-1} G(y_{t-1};\gamma,c) + \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t. \tag{23}$$

   This representation (23) is interesting for testing purposes (Dickey and Fuller (1979) and Said and Dickey (1984)) and it includes (21) as a particular case. The effective determination of the speed of adjustment in the MT-STAR model arises for $\gamma > 0$. The adjustment coefficient changes smoothly from β when y_{t-1} is at equilibrium to $\beta + \varphi$ when y_{t-1} is far from c. As such, the speed of adjustment depends on the magnitude of deviations from equilibrium. This makes sense in that many economic models predict that underlying process tend to display mean reverting behavior for large deviations from equilibrium but might follow a unit root or even explosive behavior when the deviations are small. In the setting of the MT-STAR, it implies that as soon as $\beta \geq 0$, we must have $\varphi < 0$ and $-2 < \beta + \varphi < 0$ since under these conditions the process displays a unit root or an explosive behavior when deviations from equilibrium are small but displays mean reverting behavior for large deviations.
   Additionally, following the practice in the literature (e.g., Balke and Fomby (1997) for threshold autoregressive models and Kapetanios G and Snell (2003) for ESTAR models)

[4] Let the process $\{y_t\}_{t\in\mathbb{Z}}$ be a functional coefficient autoregressive model (Chen and R.S.Tsay (1993)) and as such, a homogeneous Markov chain with state space $\mathbb{R}$ equipped with Borel σ-algebra. By Theorem 8.1 of Fan and Yao (2003), the process is geometrically ergodic as long as the condition in Theorem 8.1 is fulfilled. Then the existence and uniqueness of a stationary distribution of $\{y_t, t \geq 1\}$ is guaranteed by geometric ergodicity (see R.L.Tweedie (1975), Tjøstheim (1986)). Geometric ergodicity follows from general conditions for ergodicity of nonlinear time series which is satisfied as soon as $|\psi + \varphi| < 1$. (see Chan and Tong (1985), Fan and Yao (2003))

one can impose $\beta = 0$ implying the process $(y_t)_t$ follows a unit root process in the first regime. When $\beta = 0$, the MT-STAR model (23) becomes:

$$\Delta y_t = \varphi y_{t-1}[1 - (1 + (y_{t-1} - c)^2 + k(y_{t-1} - c))^{-\gamma}] + \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t \tag{24}$$

It is globally stationary for $-2 < \varphi < 0$ and is locally non-stationary when $y_{t-1} = c$ since it will contain a unit root.

4. Linearity and Non Stationarity Tests

In this section, we develop a testing procedure based on the MT-STAR model which does not impose the restriction of symmetric adjustment around equilibrium. This is due to the issue that the standard linear ADF test and the KSS test on the nonlinear ESTAR framework might not be powerful if the true stationary process is stationary but in a nonlinear and asymmetric fashion. In that purpose, we first develop a test for linearity and then proceed to propose a unit root test for this model.

4.1. Test for Linearity

4.1.1. A general methodology based on model (19)

Testing Linearity against STAR modelling constitutes a first step towards building STAR models. Thus, testing linearity is important as a preliminary stage of modelling with non-linear models. The null hypothesis of linearity can be expressed as null of Φ parameters in model (19). Many nonlinear models are only identified when the alternative hypothesis holds (the model is genuinely nonlinear) but not when the null hypothesis is valid. Since the parameters of an unidentified model cannot be estimated consistently, testing linearity before fitting any of these models is an unavoidable step in nonlinear modeling.
The procedure is based on relationship (19). Under linearity, we will have only one regime, and no transition between two regimes. As such,we test

$$H_0 : \Phi = 0_{(1\times p)} \quad vs. \quad H_1 : at \quad least \quad one \quad \varphi \neq 0; i = 1, \cdots, p \tag{25}$$

which is equivalent to testing

$$H_0 : \gamma = 0 \quad vs. \quad H_1 : \gamma > 0 \quad . \tag{26}$$

The MT-STAR model (19) reduces under the null to a linear AR(p) model of order p given any of the hypothesis (25) or (26). When $\gamma = 0$, H_1 is not identified, given that the vector Φ and c can take on any value without changing the value of the likelihood function when $\gamma = 0$ and vice versa. As such, we approximate G in (19) by the following representation [5]:

$$G_h(\cdot) = \sum_{n=1}^{h}(-1)^n \frac{\gamma(\gamma+1)(\gamma+2)\dots(\gamma+n-1)}{n!}[(y_{t-d}-c)^2+k(y_{t-d}-c)]^n + O(.), \quad \gamma > 0. \tag{27}$$

[5] We employ the approach proposed by Donauer S and Sibbertsen (2010)

After expanding the terms $[(y_{t-d}-c)^2+k(y_{t-d}-c)]^n, n=1,\cdots,h$, and making some rearrangements, we obtain the auxiliary regression model for y_t for a fixed $d \leq p$ and $h \in \mathbb{N}$,

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{p} \delta_{j,0} y_{t-j} + \sum_{j=1}^{p} \delta_{j,1} y_{t-j} y_{t-d} + \sum_{j=1}^{p} \delta_{j,2} y_{t-j} y_{t-d}^2 + \cdots + \sum_{j=1}^{p} \delta_{j,2h} y_{t-j} y_{t-d}^{2h} + \xi_t \quad (28)$$

The error terms ξ_t in the regression (28) are not more ε_t. After approximating (19) by (28), to test linearity against nonlinearity, we only need to test the nullity of parameters $\delta_{j,\cdot}$. We do this using a simple F -test. Consequently the properties of the error term under the null and thus the asymptotic distribution of the classical F -test remains unaffected. We illustrate in details now this results with a MT-STAR(1, 1) model defined in (20) as:

$$y_t = \psi_1 y_{t-1} + \varphi_1 y_{t-1}[1 - (1 + (y_{t-1}-c)^2 + k(y_{t-1}-c))^{-\gamma}] + \varepsilon_t \quad (29)$$

where $c \neq 0$. To test linearity for process $(y_t)_t$, we approximate G in (29) by G_3 given by (27) with $d = 1$ and $h = 3$. The process $(y_t)_t$ in (29) becomes

$$y_t = \psi_1 y_{t-1} + \varphi_1 y_{t-1}[\gamma x - \frac{1}{2}\gamma(\gamma+1)x^2 + \frac{1}{6}\gamma(\gamma+1)(\gamma+2)x^3] + u_t \quad (30)$$

where $x = [(y_{t-1}-c)^2 + k(y_{t-1}-c)]$. An auxiliary regression of (30) is given by

$$y_t = \psi_1 y_{t-1} + \delta_{1,0} y_{t-1} + \delta_{1,1} y_{t-1}^2 + \delta_{1,2} y_{t-1}^3 + \delta_{1,3} y_{t-1}^4 + \delta_{1,4} y_{t-1}^5 + \delta_{1,5} y_{t-1}^6 + \delta_{1,6} y_{t-1}^7 + u_t \quad (31)$$

where
$\delta_{1,0} = \varphi_1[m_1(c^2-kc) + m_2(c^4-2kc^3+kc^2) + m_3(c^6-3kc^5+3k^2c^3+k^3c^3)]$, $\delta_{1,1} = \varphi_1[m_1(k-2c) + m_2(6kc^2-4c^3-2kc) + m_3(-6c^5+15kc^4+9k^2c^2+3k^3c^2)]$, $\delta_{1,2} = \varphi_1[m_1 + m_2(6c^2-6kc+k) + m_3(15c^4-30kc^3-9k^2c-3ck^3)]$, $\delta_{1,3} = \varphi_1[m_2(2k-4c) + m_3(-20c^3+30c^2k+3k^2+k^3)]$, $\delta_{1,4} = \varphi_1[m_2 + m_3(15c^2-15kc)]$, $\delta_{1,5} = \varphi_1[m_3(3k-6c)]$, $\delta_{1,6} = \varphi_1[m_3]$, $m_1 = \gamma$, $m_2 = -\frac{1}{2}\gamma(\gamma+1)$, $m_3 = \frac{1}{6}\gamma(\gamma+1)(\gamma+2)$.

Now to test linearity against nonlinear MT-STAR, we test

$$H_0 : \delta_{1,1} = \cdots = \delta_{1,6} = 0 \quad H_1 : \textit{at least one } \delta_{1,i} \neq 0; i = 1, \cdots, 6. \quad (32)$$

We suggest to use the F-version of the LM test statistics since it has better size properties than the χ^2 variants, which may be heavily oversized in small samples (D van Dijk and Franses (2002)). This can be performed with the following steps:

1. Estimate the model under the null hypothesis of linearity by regressing y_t on y_{t-1}. Compute the residuals ϵ_t and the sum of squared residuals $SSR_0 = \sum_{t=0}^{T} \epsilon_t^2$
2. Estimate the auxiliary regression of y_t on y_{t-1} and $y_{t-1}y_{t-1}^i$ for $i = 1, \cdots, 6$. Compute the residuals u_t and the sum of squared residuals $SSR_1 = \sum_{t=0}^{T} u_t^2$
3. Under the null (32) the test statistic is

$$F = \frac{(SSR_0 - SSR_1)/12}{SSR_1/(T-14)} \quad (33)$$

which is approximately Fisher (F) distributed with 12 numerator degrees of freedom and and $T - 14$ denominator degrees of freedom.

Remark 1. *It is noteworthy that if we approximate the transition function G in* (27) *by setting* $h = 1$ *and letting* $k = 0$ *:*

- *if* $c = 0$ *, we obtain the auxiliary regression* (7) *on which the KSS test Kapetanios G and Snell (2003), introduced in section2, was obtained as a special case.*
- *if* $c \neq 0$ *, we obtain the auxiliary regression* (10) *used to drive the modified Wald form test statistic (Kruse (2008))*

4.2. Test for Non Stationarity in the MT-STAR framework

Assuming that linearity is rejected, we focus on non stationarity. We propose a testing procedure to distinguish between the null of a unit root process[6] and an alternative of a nonlinear globally stationary MT-STAR process, defined in (24). This testing procedure corresponds to:

$$\begin{cases} H_0 : \gamma = 0 \quad \varphi = 0 \quad (y_t)_t \; follows\ a\ linear\ unit\ root\ process \\ H_1 : \gamma \neq 0 \quad \varphi \neq 0 \quad (y_t)_t \; is\ a\ nonlinear\ globally\ stationary\ MT-STAR\ process \end{cases} \tag{34}$$

We will differentiate three cases for testing procedures: a process $(y_t)_t$ defined in (24); a centered process $(z_t)_t = y_t - \mu$; and a demeaned and detrended process, $(w_t)_t = y_t - \mu - \alpha t$. Now in the following, we restrict to $c = 0$ and $\psi = 1$ in the model (24). Thus the process $(y_t)_t$ becomes:

$$\Delta y_t = \varphi y_{t-1} G(y_{t-1}; \gamma, c) + \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t \tag{35}$$

which is same as

$$\Delta y_t = \varphi y_{t-1}[1 - (1 + y_{t-1}^2 + k y_{t-1})^{-\gamma}] + \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t \quad . \tag{36}$$

Now, in order to avoid the presence of nuisance parameters under the null hypothesis, we approximate the smooth transition function $G(y_{t-1}, \gamma, c)$ in model (36) by (27) with $d = 1$ and $h = 3$. We specify now the three cases.

1. We assume that $(y_t)_t$ is a mean zero process. This yields to the auxiliary regression:

$$\Delta y_t = \delta_{1,2} y_{t-1}^3 + \delta_{1,4} y_{t-1}^5 + \delta_{1,6} y_{t-1}^7 + \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t \tag{37}$$

$u_t \sim iid(0, \sigma^2)$. Suppose that initial sample is of the size $T+p+1$ so there are T observations in the regression and we test the hypothesis

$$H_0 : \delta_{1,2} = \delta_{1,4} = \delta_{1,6} = 0 \quad vs. \quad H_1 : at\ least\ one\ \delta_{1,i} \neq 0; i = 2, 4, 6. \tag{38}$$

which is equivalent to testing the null

$$H_0 : R\beta = r \ against\ H_1 : R\beta \neq r \tag{39}$$

[6] The process $\{X_t\}$ is a unit root process if it satisfies $(1 - L)X_t = U_t$, where $\{U_t\}$ is a mean zero covariance stationary process with short memory, L being the backward shift operator.

where

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}_{3\times(p+3)} = \begin{pmatrix} I_3 & 0_{3\times p} \end{pmatrix}, \beta = \begin{pmatrix} \delta_{1,2} \\ \delta_{1,4} \\ \delta_{1,6} \\ \rho_1 \\ \rho_2 \\ \cdots \\ \rho_p \end{pmatrix}_{(p+3)\times 1} \quad r = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}_{3\times 1}$$

For testing we use the statistic

$$\tilde{F}_{NL} = (R(\hat{\beta}-\beta))'\left[\hat{\sigma}_T^2 R(\sum_{t=1}^{T} X_t X_t')^{-1} R'\right]^{-1} R(\hat{\beta}-\beta) \tag{40}$$

where $X_t = \begin{pmatrix} y_{t-1}^3 \\ y_{t-1}^5 \\ y_{t-1}^7 \\ \Delta y_{t-1} \\ \Delta y_{t-2} \\ \cdots \\ \Delta y_{t-p} \end{pmatrix}_{(p+3)\times 1}$ $\hat{\beta}$ is the OLS estimator for the parameter β, and $\hat{\sigma}_T^2$ is the variance of Δy_t in (37). Under the null of $\beta = \mathbf{0}$ then (40) becomes

$$\tilde{F}_{NL} = \frac{\hat{\beta}'[Var(\hat{\beta})]^{-1}\hat{\beta}}{3} \tag{41}$$

Before providing the limiting non-standard distribution of this $\tilde{F}_{NL}$ test statistic (41), we introduce a technical result.

Proposition 1. *If $(y_t)_t$ is a linear single unit root process of the form:*

$$\Delta y_t = \rho_1 \Delta y_{t-1} + \rho_2 \Delta y_{t-2} + \cdots + \rho_p \Delta y_{t-p} + u_t, \quad u_t \sim i.i.d(0, \sigma^2)$$

with $(\gamma_j)_j = Cov(\Delta y_t, \Delta y_{t-j})$, denoting $\Psi(1) = (1-\rho_1-\rho_2-\cdots-\rho_p)^{-1}$, then the following results hold:

(a) $\frac{1}{T^{\frac{n}{2}+1}} \sum_{t=1}^{T} y_{t-1}^n \longrightarrow_d [\Psi(1)]^n \sigma^n \int_0^1 [W(r)]^n dr$

(b) $\frac{1}{T^{\frac{n+1}{2}}} \sum_{t=1}^{T} y_{t-1}^n u_t \longrightarrow_d [\Psi(1)]^n \sigma^{n+1} \int_0^1 [W(r)]^n dW(r)$

(c) $\frac{1}{T} \sum_{t=1}^{T} \Delta y_{t-i} \cdot \Delta y_{t-j} \longrightarrow_d \gamma_{|i-j|}$

(d) $\frac{1}{T^{\frac{n}{2}+1}} \sum_{t=1}^{T} y_{t-1}^n \cdot \Delta y_{t-j} \longrightarrow_d 0$

(e) $\frac{1}{T^{\frac{1}{2}}} \sum_{t=1}^{T} \Delta y_{t-i} u_t \longrightarrow_d N(0, \sigma^2 \gamma_0), \quad i \geq 1$

The proof of this proposition is postponed in the Appendix. Let us introduce now the main theorem.

Theorem 1. *Let be the process $(y_t)_t$ defined in (37), and $y_1, \cdots, y_T$ a sample with size T. If $\hat{\beta}$ is a consistent estimator of β and converges to its true value with rate of convergence*

$$diag(T^2, T^3, T^4, T^{1/2}, T^{1/2}, \cdots, T^{1/2})_{(p+3)\times(p+3)}$$

then the $\tilde{F}_{NL}$ *test statistic given in* (41) *testing that* $(y_t)_t$ *is unit root under the null* (39) *has the following asymptotic distribution:*

$$\tilde{F}_{NL} \longrightarrow_d v' Q^{-1} v \tag{42}$$

with

$$v = \begin{bmatrix} \frac{1}{4}W(1)^4 - \frac{3}{2}\int_0^1 W(r)^2 dr \\ \frac{1}{6}W(1)^6 - \frac{5}{2}\int_0^1 W(r)^4 dr \\ \frac{1}{8}W(1)^8 - \frac{7}{2}\int_0^1 W(r)^6 dr \end{bmatrix}$$

and

$$Q = \begin{bmatrix} \int_0^1 W(r)^6 dr & \int_0^1 W(r)^8 dr & \int_0^1 W(r)^{10} dr \\ \int_0^1 W(r)^8 dr & \int_0^1 W(r)^{10} dr & \int_0^1 W(r)^{12} dr \\ \int_0^1 W(r)^{10} dr & \int_0^1 W(r)^{12} dr & \int_0^1 W(r)^{14} dr \end{bmatrix}$$

where $W(r)$ *is the standard Brownian motion defined on* $r \in [0,1]$*. Under the alternative the test is consistent.*

The proof of this proposition is postponed in the Appendix.

Corollary 1. *If the process* $(y_t)_t$ *in* (21) *restrict to* $c = 0$ *and* $\psi = 1$*, then*

$$\Delta y_t = \delta_{1,2} y_{t-1}^3 + \delta_{1,4} y_{t-1}^5 + \delta_{1,6} y_{t-1}^7 + u_t \tag{43}$$

with $u_t \sim i.i.d(0, \sigma^2)$*, then the test statistic* $\tilde{F}_{NL}$ (41) *is computed for* $\beta = (\delta_{1,2}, \delta_{1,4}, \delta_{1,6})'$ *and it has the same asymptotic distribution given in Theorem 1.*

2. We introduce now the mean process $(y_t^\mu)_t$:

$$y_t^\mu = y_t - \frac{1}{T}\sum_{k=1}^{T} y_k$$

   and the following process

$$\Delta y_t^\mu = \varphi y_t^\mu [1 - (1 + (y_t^\mu)^2 + k y_t^\mu)^{-\gamma}] + u_t, \quad u_t \sim i.i.d(0, \sigma^2). \tag{44}$$

   For this model, the asymptotic distribution of $\tilde{F}_{NL}$ test statistic (41) has the same form as in Theorem 1 replacing the Brownian motion $W(r)$ by the de-mean Brownian motion

$$\bar{W}(r) = W(r) - \int_0^1 W(r) dr \quad .$$

3. If we consider now the process $(y_t^\gamma)_t$, defined as

$$y_t^\gamma = y_t - \hat{\mu} - \hat{\alpha} t \tag{45}$$

   then we test the model

$$\Delta y_t^\gamma = \varphi y_t^\gamma [1 - (1 + (y_t^\gamma)^2 + k y_t^\gamma)^{-\gamma}] + u_t \quad . \tag{46}$$

For this process the asymptotic distribution of the test statistic $\tilde{F}_{NL}$ given in (41) is the same as in the Theorem 1 replacing the Brownian motion $W(r)$ by the demeaned and detrended Brownian motion[7]

$$\tilde{W}(r) = W(r) + (6r - 4)\int_0^1 W(r)dr + (6 - 12r)\int_0^1 rW(r)dr \quad .$$

The theoretical power of the test based on $\tilde{F}_{NL}$ is not known, but we investigate the empirical size and power of this new test using Monte Carlo experiments in the next section. This new $\tilde{F}_{NL}$ statistic given in (41) will be referred to as the QEM test.

5. Monte Carlo Study

In this section, we carry out a Monte Carlo simulation to study the size and power properties of the new QEM test for finite sample sizes. In particular, we compare the test with some existing proposed unit root tests under the STAR framework. Thus, we carry out a study considering the tests given in section 2.2: that we list now:

1. KSS test (Kapetanios G and Snell (2003)) given in (42)
2. F_{NL} test (Pascalau (2007)) given in (13). This is a proposed test for a unit root in the asymmetric nonlinear smooth transition framework.
3. W_{nl} test of Dieu-Hang and Kompas (2010) given in (12). The M-ESTAR model is a modification of the ESTAR to account for cases where the adjustments to equilibrium are not necessarily symmetric.
4. Dickey and Fuller (1979) unit root tests.

We distinguish the three cases of datasets: Case 1, Case 2 and Case 3 refer to the underlying model with the raw data, the de-meaned data and the de-trended data, respectively. We provide in Table 1, the asymptotic critical values for the QEM test based on our distribution given in Theorem 1 taking into account the three cases of datasets defined below. We set the sample size to $T = 10000$ and the number of replications to 1000000.

We calibrate each test, assuming that we have a random walk under the null.
In Table 2 present the size of alternative tests for different sample size at a nominal level, $\alpha = 5\%$ using $50,000$ replications. The size of our proposed QEM test appear to be properly sized for a nominal of 1% for any sample size considered. However, for a nominal level of 5% or 10% there exist some mild distortions in small samples sizes $T < 500$ but the test approaches its nominal level as the sample size increases. This does not invalidate the use of our proposed test in small samples since a small size of test implies the real type I error is smaller than the nominal error and as such we are willing to accept. Unreported results for F_{NL} test for Case 3 test indicates that in small samples it tends to over-reject the null hypothesis when the true process has a unit root.

We evaluate the empirical powers of the five previous tests. We generate a dataset which follows (DGP):

$$y_t = \psi y_{t-1} + \varphi y_{t-1}(1 - (1 + y_{t-1}^2 + k y_{t-1})^{-\gamma}) + \epsilon_t \tag{47}$$

[7]Derivations of $\tilde{W}(r)$ are given in the proof of Theorem 5.1 of Stock and Watson (1988b) and in Park and Phillips (1988); the result can also be derived from Theorem 2.1 of Durlauf and Phillins (1988). As Park and Phillips (1988) demonstrate, $\tilde{W}(r)$ can be thought of as detrended Brownian motion, the residual of the projection of W onto $(1, t)$

Asymptotic Critical values of QEM statistic			
Fractile (α)	Case1	Case2	Case3
1.00%	4.721747	5.474641	6.601956
5.00%	3.459391	4.130903	5.132345
10.00%	2.885569	3.514505	4.443986

Table 1: Asymptotic critical values of QEM statistic

The Size of Alternative Test for Unit Root [in %]

	Case 2(demeaned data)				
	$\alpha = 5\%$				
T	QEM	KSS	W_{nl}	F_{NL}	DF
50	3.792	4.986	3.764	4.726	6.408
100	3.53	4.732	4.096	4.696	5.636
150	3.688	4.868	4.624	4.916	5.522
200	3.682	4.922	4.546	4.836	5.288
500	4.21	4.88	4.79	4.766	5.07
1000	4.708	5.068	5.08	4.84	5.218
5000	5.118	4.896	4.994	4.772	5.108
10000	4.956	4.85	4.862	4.702	5.056

Table 2: The Size of Alternative Tests [in %]

where $\epsilon_t \sim N(0, \sigma^2)$.
We choose a broad range of parameter values:

$\varphi = \{-0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, -0.05\}$,
$\gamma = \{0.5, 0.8, 1\}$, $\psi = 1$,
$\sigma = \{1, 0.1\}$, $k = \{0, 1, 4\}$
$T = \{50, 100, 150, 200, 500, 1000\}$

The values of φ are choosen to satisfy the condition $-2 < \varphi < 0$ for global stationarity. The values k are considered to account for both symmetic and asymmetric adjustments. Thus, k is the measure of the magnitude of asymmetry in the adjustment process. In particular, the value of $k = 4$ is considered to illustrate the powers of alternative tests when the DGP is highly asymmetric.
The parameter φ indicates the difference between the regimes. As this parameter approaches zero, the DGP approximates a linear process and it is expected that DF will perform very well. We examine the power of the various tests considering two cases of error variance $\sigma = \{1, 0.1\}$. We report[8] simulation results for Case 2 [9] at $\alpha = 5\%$, $\gamma = 0.8$, $T = \{100, 200, 500, 1000\}$ for $\sigma = \{1, 0.1\}$, $k = \{0, 4\}$.

5.1. Results

In general, the empirical power increases as soon as the sample size T increases for any values of k, γ and σ. The empirical power for each of the tests, we considered, increases as the value of the transition variable γ increases. When the DGP is approximately linear, corresponding to φ very small or large γ, it does not appear to be much power gain using a test different than a simple DF one. This is not a surprising finding because as γ grows large, the model becomes approximately linear. Considering the range of φ values from -0.05 (corresponding nearly to

[8]The results for the other Cases , T, γ, k and *sigma* as well as all other unreported results are available from authors upon request

[9]Most economic variables exhibits deterministic component. Other results for Case 1 and Case 3 can be provided if necessary

stay in one regime. DGP approximates to a linear process) to −0.7 (corresponding the case in which there is switching in the regimes), we observe that the empirical power increases with an increase in the magnitude of φ except when DGP is with $\sigma = 0.1$ and $k = 4$. Thus, power of tests decreases when the difference between the regimes, φ, is very small, corresponding to a small value of φ, except for DGP with $\sigma = 0.1$ and $k = 4$. The results indicate a good overall performance of the unit root test in all sample sizes considered especially as sample size increases. The ability to distinguish between a unit root process and a globally stationary MT-STAR model increases if either the difference between the regimes becomes larger or even faster if the sample size increases.

A general finding is that our suggested QEM test is relatively more powerful for both symmetric $k = 0$ and asymmetric $k \neq 0$ DGP when $\sigma = 1$ regardless of the values γ, φ, and T. For instance, in table 4, fixing $\varphi = -0.2$ and $T = 200$, QEM test records higher values for any k, compared to the other test considered in this paper.

Interestingly, in the region of the null, where the series is relatively more persistent, corresponding to the relatively small value of γ and/or φ, the QEM test performs best relative to the KSS test for $T \geq 500$ and $\sigma = 1$. For example, when looking at Table 4 with $T = 500$, $\varphi = -0.05$ and for any k, KSS test exhibit lower power relative to QEM test. Considering that most economic time series are likely to be highly persistent or stay near unit root, this might be a useful finding at least empirically (Kapetanios G and Snell (2003)). The new QEM test performs better for the demeaned and/or de-trended data than KSS, F_{NL} and W_{nl} tests for any k and for all sample size. For instance, for $k = 4$, where the DGP is highly asymmetic, QEM test is more powerful than the other tests for any sample size. This result supports our prior arguments that the KSS test might be unable to detect the presence of a globally stationary process if the adjustment is asymmetic.

We examine the behavior of the newly developed QEM unit root test under very small variances of the innovation term. Under small error variance, say $\sigma = 0.1$, the power of all the tests considered deteriorates compared to the case of white noise distribution. However, for sample size $T \geq 500$, we get nearly the same power properties for all the tests whatever the values of k, φ and γ we consider. In particular, for small values of error variances, $\sigma = 0.1$, and varying values of k, (from 0 to 4) which corresponds to the degree of asymmetry, the power of KSS test deteriorates as soon as k increases. In table 4, for any φ and T, KSS test records lower power compared to the QEM, F_{NL} and W_{nl} tests. In particular, for $k = 4$, the QEM, F_{NL} and W_{nl} tests exhibit good power properties for all values of T.

6. Empirical Illustration

We now propose a real exercise to compare the accuracy of our proposed tests on real data. In this exercise, we considered three datasets : monthly real effective exchange rate CPI deflated of the euro over the period from October,1980 to October,2011; eight bilateral real exchange rates relative to the euro over the period from January 1999 to November 2011, and five normalised real exchange rates relative to the US dollar over the period from January 1973 to June 2008. We focus here on the problem of Purchasing Power Parity (PPP). Unit root tests have become a very popular tool in the literature that is concerned with testing validity of the Purchasing Power Parity (PPP) which counts to one of the most important parities in international macroeconomics. The findings of a unit root in real exchange rates by Meese and Rogoff (1988) subsequently shifted the interest in modeling real exchange rates to non-linear models.

The Purchasing Power Parity (PPP) holds if and only if the real exchange rate is stationary. As

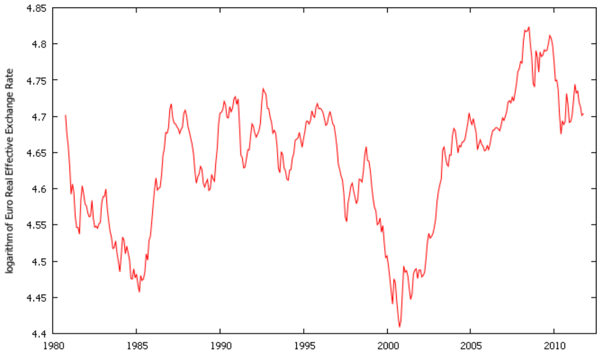

Figure 1: Logarithm of real effective exchange rate (October, 1980 to October, 2011)

such, real exchange rate should not behave like a unit root process but rather be non-linear and globally stationary process to support PPP. Henceforth, testing the unit root hypothesis means testing the non-validity of the PPP theory. Since linear unit root tests like the ones of Dickey and Fuller (1979) andPhillips and Perron (1988) often fail to reject the null hypothesis of non-stationarity when being applied to real exchange rate data, researchers tend to use nonlinear unit root tests where the specific model that is true under the alternative is congruent with economic models of financial markets.

6.1. Application to Real Effective Exchange Rates of Euro

We apply our proposed QEM unit root tests and other suggested existing unit root tests (KSS, W_{nl} and F_{NL}) against nonlinear alternatives. We also considered the famous DF tests against linear alternatives to the monthly real effective exchange rate CPI deflated[10] time series for the Euro. The data[11] spans from 1980:10 to 2011:10 implying 373 observations. The logged time series is depicted in Figure 1. It is observed from the figure that no linear trend can be seen in the data but the mean appears to be highly significant. As such, we demean the data in a first step. In the next step we estimate the test regressions with a lag length ($\hat{p} = 1$) chosen accordingly to the Bayesian information criterion. We obtain that the KSS test = -2.676300 and τ test= 7.480098, failed to reject the null hypothesis of unit root at the ten percent level suggesting that PPP does not hold. Furthermore, the unit root tests against linear alternatives by Dickey and Fuller (1979) (DF) do not provide any evidence against the null hypothesis. However, the QEM test = 11.269512, W_{nl} = 14.569502 and F_{NL} = 11.224548 rejects the null hypothesis at one percent level of significance. This yields new evidence on the stationarity of the EU real effective exchange rate which suggests the validity of PPP. It is worth noting that, the time series data has possibly an asymmetric adjustment to equilibrium since the only STAR type tests that rejects the null hypothesis accounts for such adjustments. The false non-rejections of the null by

[10]The effective exchange rates (EERs) of the euro are geometrically weighted averages of the bilateral exchange rates of the euro against the currencies of the euro area's main trading partners. For additional information, see the "Daily nominal effective exchange rate of the euro" section of the ECB's website

[11]Data source: ECB's website http://www.ecb.int/stats/services/downloads/html/index.en.html and code: EXR.M.Z08.EUR.ERC0.A

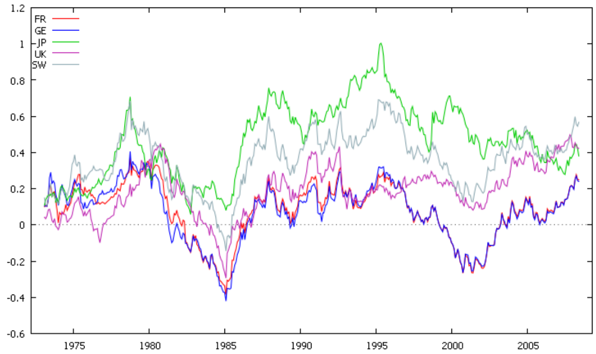

Figure 2: Time series plot of the normalised real exchange rates over the period from January 1973 to June 2008.

KSS test leading to rejecting a nonlinear adjustment process for real exchange rates could be due to extremely small error variances of the process. We conclude that nonlinearities with potential adjustment to long-run equilibrium being asymmetic over time are present in the data. The new QEM unit root test yields new evidence on the stationarity of the EU real effective exchange rate which suggests the validity of PPP.

6.2. Application to Five Bilateral real exchange rate series relative to US Dollar

We apply our proposed test procedure with other tests previously discussed on the same dataset used by Buncic (2009). Our objective is to verify if indeed the ESTAR model was appropriate for modelling the real exchange rates considered by Buncic (2009). The data [12] consists of five real exchange rates relative to the US dollar corresponding to the UK, Japan, German, France and Switzerland, from January 1973 to June 2008. These real exchange rates are constructed in the standard way as $q_t \equiv log(\frac{CPI_t^{home}}{CPI_t^{US} S_t})$, where S_t is the home currency price of one US dollar. We employ the linearity test proposed in our work and the one proposed by Terasvirta (1994) on each of the real exchange rate. We obtain that the only time series that exhibits clear nonlinearity is that of the German real exchange rate series. This results shows that the real exchange rates considered in this exercise might follow a nonlinear model like ESTAR model. In order to test for unit root, we first consider the underlying time series without demeaning and then perform the unit root testing procedure. We obtain QEM = 5.012076 and KSS = −2.788665 for the French series, which are significant at 1% and 5% level suggesting that PPP holds. For the German series, the test statistics QEM = 4.238476, KSS = −2.720900 are significant at 5% and F_{NL} = 3.126625 significant at 10% level. Furthermore, no test statistics fail to reject the non-validity of PPP at 5% for the Japan, UK, and Switzerland real exchange rate series. As one can see from figure 2, no linear trend can be seen for each series but the mean appears to be highly significant. As such, we demean each series and then perform the unit root tests. The results shows that no test statistics provide support that the French, German, Japan and Switzerland real exchange rate series are nonlinear globally stationary process. However, for the UK series we obtain QEM = 6.133468, which is significant at 1% level suggesting PPP holds. These findings

[12]The data can be downloaded from http://www.mathstat.unisg.ch/buncic/data/rer_data.xls

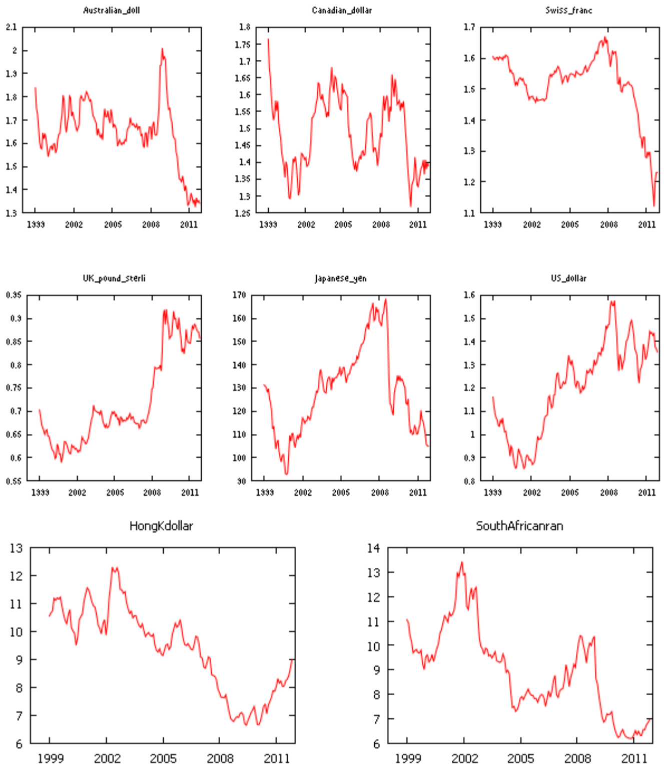

Figure 3: Bilateral real exchange rates over the period from January 1999 to November 2011

provide no support for the use of ESTAR model to forecast such real exchange rates, particularly for Japan, UK and Switzerland. Hence, our findings are consistent with the results of Buncic (2009) on no forecast gained by ESTAR model over linear autoregressive model.

6.3. Application to Bilateral real exchange rate series relative Euro

We now consider eight bilateral monthly exchange rates of Australian dollar, Canadian dollar, Swiss franc, UK pound sterling, Japanese yen, US dollar, Hong Kong dollar and SouthAfrican rand relative to the Euro. Our data is taken from European Central Bank [13] and spans from 1999:01 to 2011:11 implying 155 observations. Applying linearity tests on the exchange rates displayed in figure 3, we obtain that linearity is rejected for the Canadian and UK exchange rates based on our proposed linearity test procedure. However, the Canadian, Switzerland, UK and JP exchange rates rejects the null of linearity based on tests developed by Terasvirta (1994). On performing the unit root tests with no transformation on the exchange rates, we obtain the test statistics QEM = 4.554032, W_{nl} = 5.098192, F_{NL} = 4.113948, and τ = 10.196384 significant at 5% level suggesting PPP holds for Canadian exchange rate. We obtain evidence of

[13] Data source: ECB's website http://www.ecb.int/stats/services/downloads/html/index.en.html

possible asymmetric adjustment to long-run equilibrium for the UK exchange rate series since QEM = 5.0490912 and F_{NL} = 3.2378842 are significant at 1% and 10% respectively. Thus, the bilateral exchange rate series for Australian, Switzerland, Japan, US, Hong Kong, SouthAfrican, behavior like linear unit root process and might not yield better forecast ability when modelled with ESTAR model over linear autoregressive models. Furthermore, when we perform the unit root tests on the demeaned or detrended real exchange rate series depending on the evolution of the time series, we obtain no evidence that PPP holds for the Australian, Switzerland, Japan, US and Hong Kong. However, for the case of SouthAfrican, we obtain F_{NL} = 4.063662, and for the UK series the test statistics W_{nl} = 5.959871, τ = 11.919741, KSS = −3.267500 are significant at 10% level. We still obtain evidence of PPP on Canadian series with QEM = 6.359354, W_{nl} = 10.136174, KSS = −4.383255, F_{nl} = 6.874602 and τ = 20.272348 significant at 1% level.

7. Conclusion

This paper extends the work of Donauer S and Sibbertsen (2010) by introducing the possibility of asymmetric adjustment towards equilibrium. Based on the work of Terasvirta (1994) and Kapetanios G and Snell (2003), the present paper proposes a new unit root test, called QEM test, that has power against nonlinear but globally stationary alternatives, where the adjustment is smooth over time. The paper derives the asymptotic limit distributions of this new test and empirical application shows that the new test should be used together with earlier tests proposed by Terasvirta (1994) and Kapetanios G and Snell (2003) to distinguish whether the adjustment to long-run equilibrium is either symmetric or asymmetric over time. In this paper, we assessed the power performance of the QEM test using numerical approach. A possible extension of this work will be to establish the theoritical power of the test by allowing the alternative to be contiguous to the null hypothesis (see D. Guégan (1992)).

8. AppendixA

8.1. Functional Central Limit Theorem (Recall)

Suppose $u_t \sim i.i.d(0, \sigma^2)$, $t = 1, 2, \cdots, T$ and

$$X_T(r) = \frac{1}{T} \sum_{t=1}^{\lfloor T_r \rfloor} u_t$$

where $X_T(r)$ is a variable constructed from the sample mean of the first r^{th} fraction of random variables u_t, $r \in [0, 1]$ and where $\lfloor T_r \rfloor$ denotes the largest integer that is less than or equal to T_r. Thus $X_T(r)$ is a step function in r defined as follows:

$$X_T(r) = \begin{cases} 0 & if\ 0 \leq r < 1/T \\ u_1/T & if\ 1/T \leq r < 2/T \\ (u_1 + u_2)/T & if\ 2/T \leq r < 3/T \\ \cdots & \cdots \\ (u_1 + u_2 + \cdots + u_T)/T & if\ r = 1 \end{cases}$$

Then the sequence of stochastic functions $\left\{ \frac{\sqrt{T} X_T(\cdot)}{\sigma} \right\}_{T=1}^{\infty}$ converges in distribution to a standard Brownian motion $W(\cdot)$, denotes: $\frac{\sqrt{T} X_T(\cdot)}{\sigma} \xrightarrow{d} W(\cdot)$

8.2. Proof of Proposition 1

Proof. We want to prove (a), (b), (c), (d), (e) of proposition 1. First let us introduce the $AR(p+1)$ process:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p} + \phi_{p+1} y_{t-p-1} + u_t \quad where\ u_t \sim i.i.d(0, \sigma^2)$$

This process can be written in the form:

$$y_t = \rho y_{t-1} + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta y_{t-2} + \cdots + \alpha_p \Delta y_{t-p} + u_t$$

where: $\rho = \phi_1 + \phi_2 + \cdots + \phi_{p+1}$, $\alpha_i = -(\phi_{i+1} + \phi_{i+2} + \cdots + \phi_{i+p+1})$.
Suppose now that $(y_t)_t$ contains a single unit root: that is one of the roots is 1 in the equation :

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_{p+1} z^{p+1} = 0$$

thus

$$1 - \phi_1 - \cdots - \phi_{p+1} = 0 \quad and \quad \phi_1 + \phi_2 + \cdots + \phi_{p+1} = 1 \quad implies\ that\ \rho = 1$$

Thus, under the null hypothesis of a unit root we have:

$$\Delta y_t = \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta y_{t-2} + \cdots + \alpha_p \Delta y_{t-p} + u_t$$

thus

$$(1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p) \Delta y_t = u_t \tag{48}$$

a. Suppose that process $(y_t)_t$ has only one unit root and all other roots are outside the unit circle, then Δy_t is stationary. From equation (48) we have:

$$\Delta y_t = (1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p)^{-1} u_t = \Psi(L) \cdot u_t$$

$$where\ \Psi(L) = (1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p)^{-1}$$

Next, define $\forall t\ \varepsilon_t = \Delta y_t$, this implies that $\varepsilon_t = \Psi(L) \cdot u_t, \quad u_t \sim (0, \sigma^2)$, and

$$y_t = y_0 + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t \tag{49}$$

The Beveridge-Nelson decomposition implies that:

$$y_t = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t + y_0 = \Psi(1)\Big[u_1 + u_2 + \cdots + u_t\Big] + \eta_t - \eta_0 + y_0 \tag{50}$$

where $\Psi(1) = (1 - \alpha_1 - \alpha_2 - \cdots - \alpha_p)^{-1}$ and $(\eta_t)_t$ is a stationary process with zero mean of the form:

$$\eta_t = \sum_{j=0}^{\infty} \lambda_j u_{t-j} \quad where \quad \sum_{j=0}^{\infty} |\lambda_j| < \infty$$

We construct a variable $Z_T(r)$ from the sample mean of the first r^{th} fraction of random variables ε_t, $r \in [0, 1]$:

$$Z_T(r) = \frac{1}{T} \sum_{t=1}^{\lfloor T_r \rfloor} \varepsilon_t \tag{51}$$

$$\sqrt{T} Z_T(r) = \sqrt{T} \frac{1}{T} \sum_{t=1}^{\lfloor T_r \rfloor} \varepsilon_t = \sqrt{T} \cdot \frac{1}{T} \Big(\Psi(1) \sum_{t=1}^{\lfloor T_r \rfloor} u_t + \eta_{\lfloor T_r \rfloor} - \eta_0\Big)$$

$$= \Psi(1) \sqrt{T} X_T(r) + \frac{1}{\sqrt{T}} \Big(\eta_{\lfloor T_r \rfloor} - \eta_0\Big)$$

From the functional Central Limit theorem and continuous mapping theorem:
$\sqrt{T} X_T(\cdot) \rightarrow_d \sigma W(\cdot) \quad \Rightarrow \quad \Psi(1) \sqrt{T} X_T(r) \rightarrow_d \Psi(1) \cdot \sigma W(r)$, with $(\eta_t)_t$ a zero mean stationary process, thus we have: $\frac{1}{\sqrt{T}}\Big(\eta_{\lfloor T_r \rfloor} - \eta_0\Big) \longrightarrow_p 0$ (*see Hamilton*, 1994, *example*17.2). It follows that:

$$\sqrt{T} Z_T(\cdot) \rightarrow_d \Psi(1) \cdot \sigma W(\cdot) \tag{52}$$

The stochastic function $Z_T(r)$ in (51) can be expressed in the form of a step function in r as follows:

$$Z_T(r) = \begin{cases} 0 & if\ 0 \le r < 1/T \\ \varepsilon_1/T & if\ 1/T \le r < 2/T \\ (\varepsilon_1 + \varepsilon_2)/T & if\ 2/T \le r < 3/T \\ \cdots & \cdots \\ (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_T)/T & if\ r = 1 \end{cases} \tag{53}$$

$$thus \quad \sqrt{T}(Z_T(r) + \frac{y_0}{T}) = \begin{cases} \frac{y_0}{\sqrt{T}} & if\ 0 \le r < 1/T \\ \frac{y_1}{\sqrt{T}} & if\ 1/T \le r < 2/T \\ \frac{y_2}{\sqrt{T}} & if\ 2/T \le r < 3/T \\ \dots & \dots \\ \frac{y_T}{\sqrt{T}} & if\ r = 1 \end{cases}$$

The stochastic function $F_T(r) = \left[\sqrt{T}(Z_T(r) + \frac{y_0}{T}) \right]^n$ is then expressed in the form:

$$F_T(r) = \begin{cases} \frac{y_0^n}{T^{\frac{n}{2}}} & if\ 0 \le r < 1/T \\ \frac{y_1^n}{T^{\frac{n}{2}}} & if\ 1/T \le r < 2/T \\ \frac{y_2^n}{T^{\frac{n}{2}}} & if\ 2/T \le r < 3/T \\ \dots & \dots \\ \frac{y_T^n}{T^{\frac{n}{2}}} & if\ r = 1 \end{cases} \tag{54}$$

and

$$\int_0^1 F_T(r)dr = \frac{y_0^n}{T^{\frac{n}{2}}} \cdot \frac{1}{T} + \frac{y_1^n}{T^{\frac{n}{2}}} \cdot \frac{1}{T} + \frac{y_2^n}{T^{\frac{n}{2}}} \cdot \frac{1}{T} + \dots + \frac{y_{T-1}^n}{T^{\frac{n}{2}}} \cdot \frac{1}{T} = \frac{1}{T^{(n/2+1)}} \sum_{t=1}^{T} y_{t-1}^n$$

From (52) we have:

$$F_T(\cdot) = \left[\sqrt{T}(Z_T(\cdot) + \frac{y_0}{T}) \right]^n \longrightarrow_d [\Psi(1)]^n \sigma^n [W(\cdot)]^n \tag{55}$$

thus

$$\int_0^1 F_T(r)dr \rightarrow_d [\Psi(1)]^n \sigma^n \int_0^1 [W(r)]^n dr \Rightarrow \frac{1}{T^{(n/2+1)}} \sum_{t=1}^{T} y_{t-1}^n \rightarrow_d [\Psi(1)]^n \sigma^n \int_0^1 [W(r)]^n dr$$

b. Using the Beveridge-Nelson decomposition in (50):

$$y_t = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t + y_0 = \Psi(1)\left[u_1 + u_2 + \dots + u_t\right] + \eta_t - \eta_0 + y_0$$

$$\Rightarrow \Delta y_t = y_t - y_{t-1} = \varepsilon_t = \Psi(1)u_t + \eta_t - \eta_{t-1} \tag{56}$$

$$\frac{1}{T^{(n+1)/2}} \sum_{t=1}^{T} y_{t-1}^n \varepsilon_t = \frac{1}{T^{(n+1)/2}} \sum_{t=1}^{T} y_{t-1}^n \Big(\Psi(1)u_t + \eta_t - \eta_{t-1}\Big)$$

$$= \frac{1}{T^{(n+1)/2}} \Psi(1) \sum_{t=1}^{T} y_{t-1}^n u_t + \frac{1}{T^{(n+1)/2}} \sum_{t=1}^{T} y_{t-1}^n \Big(\eta_t - \eta_{t-1}\Big)$$

$$\Rightarrow \frac{1}{T^{(n+1)/2}} \Psi(1) \sum_{t=1}^{T} y_{t-1}^n u_t = \frac{1}{T^{(n+1)/2}} \sum_{t=1}^{T} y_{t-1}^n \varepsilon_t - \frac{1}{T^{(n+1)/2}} \sum_{t=1}^{T} y_{t-1}^n \Big(\eta_t - \eta_{t-1}\Big) \tag{57}$$

Denote $F_T(r) = F_{T\lfloor T_r \rfloor}$, from the function $F_T(r)$ in (54) we have: $F_{T\lfloor T_r \rfloor} = \frac{y_{T\lfloor T_r \rfloor}^n}{T^{n/2}}$
Denote $Z_T(r) = Z_{T\lfloor T_r \rfloor}$, from the function $F_T(r)$ in (53) we have

$$Z_{T\lfloor T_r \rfloor} = \frac{\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{\lfloor T_r \rfloor}}{T} \Rightarrow \Delta Z_{T,\lfloor T_r \rfloor} = Z_{T,\lfloor T_r \rfloor} - Z_{T,\lfloor T_r \rfloor - 1} = \frac{\varepsilon_{\lfloor T_r \rfloor}}{T} \tag{58}$$

Now let us consider the first element of the RHS of (57):

$$\frac{1}{T^{(n+1)/2}}\sum_{t=1}^{T} y_{t-1}^n\varepsilon_t = \sum_{t=1}^{T}\frac{y_{t-1}^n}{T^{n/2}}\cdot\sqrt{T}\frac{\varepsilon_t}{T} = \sum_{\lfloor T_r\rfloor=1}^{T} F_{T,\lfloor T_r\rfloor-1}\cdot\sqrt{T}\frac{\varepsilon_{\lfloor T_r\rfloor}}{T} = \sum_{\lfloor T_r\rfloor=1}^{T} F_{T,\lfloor T_r\rfloor-1}\cdot\sqrt{T}\Delta Z_{T,\lfloor T_r\rfloor}$$

We have the stochastic integral:

$$\sum_{\lfloor T_r\rfloor=1}^{T} F_{T,\lfloor T_r\rfloor-1}\cdot\sqrt{T}\Delta Z_{T,\lfloor T_r\rfloor} = \int_0^1 F_T(r)\cdot\sqrt{T}dZ_T(r) = \int_0^1 F_T(r)\cdot d\sqrt{T}Z_T(r)$$

$$\Rightarrow\quad \frac{1}{T^{(n+1)/2}}\sum_{t=1}^{T} y_{t-1}^n\varepsilon_t = \int_0^1 F_T(r)\cdot d\sqrt{T}Z_T(r)$$

From (53) and (55), and by convergence to stochastic integrals we have:

$$\int_0^1 F_T(r)\cdot d\sqrt{T}Z_T(r) \xrightarrow{d} \int_0^1[\Psi(1)]^n\sigma^n[W(\cdot)]^n d\Psi(1)\cdot\sigma W(r) = [\Psi(1)]^{n+1}\sigma^{n+1}\int_0^1[W(r)]^n dW(r)$$

$$\Rightarrow\quad \frac{1}{T^{(n+1)/2}}\sum_{t=1}^{T} y_{t-1}^n\varepsilon_t \xrightarrow{d} [\Psi(1)]^{n+1}\sigma^{n+1}\int_0^1[W(r)]^n dW(r) \tag{59}$$

Next, Consider the second element on RHS of (57):

$$\frac{1}{T^{(n+1)/2}}\sum_{t=1}^{T} y_{t-1}^n(\eta_t-\eta_{t-1}) = \sum_{t=1}^{T}\frac{y_{t-1}^n}{T^{n/2}}\cdot\frac{\eta_t-\eta_{t-1}}{T^{1/2}} = \sum_{\lfloor T_r\rfloor=0}^{T-1} F_T(r)\cdot\frac{\eta_{\lfloor T_r\rfloor+1}-\eta_{\lfloor T_r\rfloor}}{\sqrt{T}}$$

From (55): $F_T(\cdot) = \left[\sqrt{T}(Z_T(\cdot)+\frac{y_0}{T})\right]^n \longrightarrow_d [\Psi(1)]^n\sigma^n[W(\cdot)]^n$

$$\Rightarrow\quad \sum_{\lfloor T_r\rfloor=0}^{T-1} F_T(r)\cdot\frac{\eta_{\lfloor T_r\rfloor+1}-\eta_{\lfloor T_r\rfloor}}{\sqrt{T}} \xrightarrow{d} [\Psi(1)]^n\sigma^n[W(r)]^n\sum_{\lfloor T_r\rfloor=0}^{T-1}\frac{\eta_{\lfloor T_r\rfloor+1}-\eta_{\lfloor T_r\rfloor}}{\sqrt{T}}$$

$$\Rightarrow\quad \sum_{\lfloor T_r\rfloor=0}^{T-1} F_T(r)\cdot\frac{\eta_{\lfloor T_r\rfloor+1}-\eta_{\lfloor T_r\rfloor}}{\sqrt{T}} \xrightarrow{d} [\Psi(1)]^n\sigma^n[W(r)]^n\Big[\sqrt{T}(\frac{1}{T}\sum_{t=1}^{T}\eta_t)-\sqrt{T}(\frac{1}{T}\sum_{t=1}^{T}\eta_{t-1})\Big]$$

η_t is a stationary process with zero mean of the form: $\eta_t = \sum_{j=0}^{\infty}\lambda_j u_{t-j}$ where $\sum_{j=0}^{\infty}|\lambda_j| < \infty$. Applying the central limit theorem for stationary process we have:

$$\Big[\sqrt{T}(\frac{1}{T}\sum_{t=1}^{T}\eta_t)-\sqrt{T}(\frac{1}{T}\sum_{t=1}^{T}\eta_{t-1})\Big]\xrightarrow{d} 0 \tag{60}$$

$$\Rightarrow\quad \sum_{\lfloor T_r\rfloor=0}^{T-1} F_T(r)\cdot\frac{\eta_{\lfloor T_r\rfloor+1}-\eta_{\lfloor T_r\rfloor}}{\sqrt{T}} \xrightarrow{d} 0.$$

It follows that:

$$\frac{1}{T^{(n+1)/2}}\sum_{t=1}^{T} y_{t-1}^n(\eta_t-\eta_{t-1}) \xrightarrow{d} 0 \tag{61}$$

From (57), (59), and (61) we obtain:

$$\frac{1}{T^{(n+1)/2}}\Psi(1)\sum_{t=1}^{T}y_{t-1}^{n}u_t \xrightarrow{d} [\Psi(1)]^{n+1}\sigma^{n+1}\int_0^1[W(r)]^n dW(r)$$

$$\Rightarrow \quad \frac{1}{T^{(n+1)/2}}\sum_{t=1}^{T}y_{t-1}^{n}u_t \xrightarrow{d} [\Psi(1)]^{n}\sigma^{n+1}\int_0^1[W(r)]^n dW(r)$$

c. From (56) : $\Delta y_t = y_t - y_{t-1} = \varepsilon_t = \Psi(1)u_t + \eta_t - \eta_{t-1}$ where η_t is a zero mean stationary process, $u_t \sim i.i.d(0,\sigma^2)$

$$\Rightarrow \quad E(\Delta y_t) = E[\Psi(1)u_t + \eta_t - \eta_{t-1}] = \Psi(1)E(u_t) + E(\eta_t) - E(\eta_{t-1}) = 0$$

$$\Rightarrow \quad \frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-i}\cdot\Delta y_{t-j} = \frac{1}{T}\sum_{t=1}^{T}[\Delta y_{t-i} - E(\Delta y_{t-i})][\Delta y_{t-j} - E(\Delta y_{t-j})]$$

y_t has a single unit root $\Rightarrow \Delta y_t$ is stationary with: $E(\Delta y_t) = 0, Var(\Delta y_t)$ is finite and $Cov(\Delta y_t, \Delta y_{t-s}) = \gamma_s$ only depends on s. By law of large number:

$$\frac{1}{T}\sum_{t=1}^{T}[\Delta y_{t-i} - E(\Delta y_{t-i})][\Delta y_{t-j} - E(\Delta y_{t-j})] \xrightarrow{p} Cov(\Delta y_{t-i}, \Delta y_{t-j}) = \gamma_{|i-j|}$$

$$\Rightarrow \quad \frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-i}\cdot\Delta y_{t-j} \xrightarrow{p} \gamma_{|i-j|}$$

d.

$$\frac{1}{T^{n/2+1}}\sum_{t=1}^{T}y_{t-1}^{n}\cdot\Delta y_{t-j} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{y_{t-1}^{n}}{T^{n/2}}\cdot\frac{\Delta y_{t-j}}{\sqrt{T}} = \frac{1}{\sqrt{T}}\sum_{\lfloor T_r\rfloor=0}^{T-1}F_T(r)\cdot\frac{\Delta y_{\lfloor T_r\rfloor+1-j}}{\sqrt{T}}$$

where we have $F_T(r) \longrightarrow_d [\Psi(1)]^n\sigma^n[W(r)]^n$

$$\Rightarrow \quad \frac{1}{T^{n/2+1}}\sum_{t=1}^{T}y_{t-1}^{n}\cdot\Delta y_{t-j} \xrightarrow{d} [\Psi(1)]^n\sigma^n[W(r)]^n\frac{1}{T}\sum_{\lfloor T_r\rfloor=0}^{T-1}\Delta y_{\lfloor T_r\rfloor+1-j}$$

$$\Rightarrow \quad \frac{1}{T^{n/2+1}}\sum_{t=1}^{T}y_{t-1}^{n}\cdot\Delta y_{t-j} \xrightarrow{d} [\Psi(1)]^n\sigma^n[W(r)]^n\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-j}$$

With Δy_t is a zero mean stationary process, applying law of large number for a covariance stationary process it follows that: $\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-j} \xrightarrow{p} 0$

$$\Rightarrow \quad \frac{1}{T^{n/2+1}}\sum_{t=1}^{T}y_{t-1}^{n}\cdot\Delta y_{t-j} \xrightarrow{d} 0 \quad \Rightarrow \quad \frac{1}{T^{n/2+1}}\sum_{t=1}^{T}y_{t-1}^{n}\cdot\Delta y_{t-j} \xrightarrow{p} 0$$

e. Consider sequence: $\Delta y_{t-i}\cdot u_t\ \forall i \geq 1$
Notice that u_t is uncorrelated to Δy_{t-i} i.e. $E(\Delta y_{t-i}\cdot u_t) = E(\Delta y_{t-i})\cdot E(u_t) = 0$, then the conditional mean:

$$E(\Delta y_{t-i}\cdot u_t/\Delta y_{t-i-1}\cdot u_t) = E(\Delta y_{t-i}/\Delta y_{t-i-1}\cdot u_t)\cdot E(u_t/\Delta y_{t-i-1}\cdot u_t)$$

$$= E(\Delta y_{t-i}/\Delta y_{t-i-1} \cdot u_t) \cdot E(u_t) = 0$$

Hence Δy_{t-i} is a martingale difference sequence.

$$Var(\Delta y_{t-i} \cdot u_t) = E[(\Delta y_{t-i} \cdot u_t)^2] = E[(\Delta y_{t-i})^2 \cdot u_t^2] = E[(\Delta y_{t-i})^2] \cdot E[u_t^2] = \sigma^2 E[(\Delta y_{t-i})^2]$$

since Δy_t is a zero mean stationary process
$\Rightarrow \quad E[(\Delta y_{t-i})^2] = Var(\Delta y_{t-i}) = Var(\Delta y_t) = \gamma_0$
$\Rightarrow \quad Var(\Delta y_{t-i} \cdot u_t) = \sigma^2 \gamma_0$
Thus, the martingale difference sequence $\Delta y_{t-i} \cdot u_t$ satisfies the usual central limit theorem :

$$\sqrt{T}\Big(\frac{1}{T}\sum_{t=1}^{T} \Delta y_{t-i} u_t\Big) \xrightarrow{d} N(0, \sigma^2 \gamma_0)$$

$$\Rightarrow \quad \frac{1}{T^{1/2}}\sum_{t=1}^{T} \Delta y_{t-i} u_t \xrightarrow{d} N(0, \sigma^2 \gamma_0)$$

□

8.3. Proof of Theorem 1

Proof. Let $\tilde{D}_T$ be a 3×3 diagonal matrix of the form:

$$\tilde{D}_T = \begin{pmatrix} T^2 & 0 & 0 \\ 0 & T^3 & 0 \\ 0 & 0 & T^4 \end{pmatrix}$$

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}_{3\times(p+3)}$$

$\beta = (\delta_{1,2}, \delta_{1,4}, \delta_{1,6}, \rho_1, \rho_2, \cdots, \rho_p)'_{(p+3)\times 1}$ and $X_t = (y_{t-1}^3, y_{t-1}^5, y_{t-1}^7, \Delta y_{t-1}, \cdots, \Delta y_{t-p})'_{(p+3)\times 1}$ with $\hat{\beta}$ being the OLS estimator for the parameter β, and $\hat{\sigma}_T^2$ the variance of Δy_t in (37).

Next, Consider the scaling matrix:

$$D_T = \begin{pmatrix} T^2 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & T^3 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & T^4 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & T^{1/2} & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & T^{1/2} & \dots & 0 \\ \dots & & \dots & & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & T^{1/2} \end{pmatrix}_{(p+3)\times(p+3)}$$

The matrices $\tilde{D}_T$ and D_T are symmetric and diagonal and $\tilde{D}_T R = R D_T$
Therefore, we obtain:

$$\tilde{F}_{NL} = (\hat{\beta} - \beta)' R' \tilde{D}_T \cdot \tilde{D}_T^{-1} \Big[\hat{\sigma}_T^2 R(\sum_{t=1}^{T} X_t X_t')^{-1} R'\Big]^{-1} \tilde{D}_T^{-1} \cdot \tilde{D}_T R(\hat{\beta} - \beta)$$

$$= (\hat{\beta}-\beta)'R'\tilde{D}_T\Big[\hat{\sigma}^2_T\tilde{D}_T R(\sum_{t=1}^{T} X_t X_t')^{-1}R'\tilde{D}_T\Big]^{-1}\tilde{D}_T R(\hat{\beta}-\beta)$$

since $\tilde{D}_T$ is symmetric : $\tilde{D}_T = \tilde{D}_T'$

$$\Rightarrow \quad \tilde{F}_{NL} = (\hat{\beta}-\beta)'(\tilde{D}_T R)'\Big[\hat{\sigma}^2_T\tilde{D}_T R(\sum_{t=1}^{T} X_t X_t')^{-1}(\tilde{D}_T R)'\Big]^{-1}\tilde{D}_T R(\hat{\beta}-\beta)$$

and since $\tilde{D}_T R = RD_T$ we have :

$$\tilde{F}_{NL} = (\hat{\beta}-\beta)'(RD_T)'\Big[\hat{\sigma}^2_T RD_T(\sum_{t=1}^{T} X_t X_t')^{-1}(RD_T)'\Big]^{-1}RD_T(\hat{\beta}-\beta)$$

$$\Rightarrow \quad \tilde{F}_{NL} = [(RD_T)(\hat{\beta}-\beta)]'\Big[\hat{\sigma}^2_T RD_T(\sum_{t=1}^{T} X_t X_t')^{-1}D_T R'\Big]^{-1}RD_T(\hat{\beta}-\beta)$$

$$\Rightarrow \quad \tilde{F}_{NL} = [(RD_T)(\hat{\beta}-\beta)]'\Big\{\hat{\sigma}^2_T R\Big[D_T^{-1}(\sum_{t=1}^{T} X_t X_t')D_T^{-1}\Big]^{-1}R'\Big\}^{-1}RD_T(\hat{\beta}-\beta)$$

with

$$D_T^{-1} = \begin{pmatrix} T^{-2} & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & T^{-3} & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & T^{-4} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & T^{-1/2} & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & T^{-1/2} & \dots & 0 \\ \dots & & \dots & & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & T^{-1/2} \end{pmatrix}_{(p+3)\times(p+3)}$$

where $X_t = (y^3_{t-1}, y^5_{t-1}, y^7_{t-1}, \Delta y_{t-1}, \cdots, \Delta y_{t-p})'_{(p+3)\times 1}$. It follows that: $D_T^{-1}(\sum_{t=1}^{T} X_t X_t')D_T^{-1} =$

$$\begin{pmatrix} T^{-4}\sum_{t=1}^T y^6_{t-1} & T^{-5}\sum_{t=1}^T y^8_{t-1} & T^{-6}\sum_{t=1}^T y^{10}_{t-1} & T^{-5/2}\sum_{t=1}^T y^3_{t-1}\Delta y_{t-1} & \dots & T^{-5/2}\sum_{t=1}^T y^3_{t-1}\Delta y_{t-p} \\ T^{-5}\sum_{t=1}^T y^8_{t-1} & T^{-6}\sum_{t=1}^T y^{10}_{t-1} & T^{-7}\sum_{t=1}^T y^{12}_{t-1} & T^{-\frac{7}{2}}\sum_{t=1}^T y^5_{t-1}\Delta y_{t-1} & \dots & T^{-\frac{7}{2}}\sum_{t=1}^T y^5_{t-1}\Delta y_{t-p} \\ T^{-6}\sum_{t=1}^T y^{10}_{t-1} & T^{-7}\sum_{t=1}^T y^{12}_{t-1} & T^{-8}\sum_{t=1}^T y^{14}_{t-1} & T^{-9/2}\sum_{t=1}^T y^7_{t-1}\Delta y_{t-1} & \dots & T^{-9/2}\sum_{t=1}^T y^7_{t-1}\Delta y_{t-1} \\ T^{-5/2}\sum_{t=1}^T y^3_{t-1}\Delta y_{t-1} & T^{-\frac{7}{2}}\sum_{t=1}^T y^5_{t-1}\Delta y_{t-1} & T^{-9/2}\sum_{t=1}^T y^7_{t-1}\Delta y_{t-1} & \frac{1}{T}\sum_{t=1}^T(\Delta y_{t-1})^2 & \dots & \frac{1}{T}\sum_{t=1}^T \Delta y_{t-1}\Delta y_{t-p} \\ \dots & & \dots & & & \\ T^{-5/2}\sum_{t=1}^T y^3_{t-1}\Delta y_{t-p} & T^{-\frac{7}{2}}\sum_{t=1}^T y^5_{t-1}\Delta y_{t-p} & T^{-9/2}\sum_{t=1}^T y^7_{t-1}\Delta y_{t-p} & \frac{1}{T}\sum_{t=1}^T \Delta y_{t-1}\Delta y_{t-p} & \dots & \frac{1}{T}\sum_{t=1}^T(\Delta y_{t-p})^2 \end{pmatrix} \quad (62)$$

Under the null hypothesis, $(y_t)_t$ is a unit root process that statisfies the condition of proposition 1. Hence applying the proposition, it follows that: $D_T^{-1}(\sum_{t=1}^{T} X_t X_t')D_T^{-1} \xrightarrow{d}$

$$\xrightarrow{d} \begin{pmatrix} \Psi^6\sigma^6\int_0^1 W(r)^6 dr & \Psi^8\sigma^8\int_0^1 W(r)^8 dr & \Psi^{10}\sigma^{10}\int_0^1 W(r)^{10} dr & 0 & 0 & 0 & \dots & 0 \\ \Psi^8\sigma^8\int_0^1 W(r)^8 dr & \Psi^{10}\sigma^{10}\int_0^1 W(r)^{10} dr & \Psi^{12}\sigma^{12}\int_0^1 W(r)^{12} dr & 0 & 0 & 0 & \dots & 0 \\ \Psi^{10}\sigma^{10}\int_0^1 W(r)^{10} dr & \Psi^{12}\sigma^{12}\int_0^1 W(r)^{12} dr & \Psi^{14}\sigma^{14}\int_0^1 W(r)^{14} dr & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{p-1} \\ 0 & 0 & 0 & \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{p-2} \\ 0 & 0 & 0 & \gamma_2 & \gamma_1 & \gamma_0 & \dots & \gamma_{p-3} \\ \dots & \dots & \dots & \dots & & \dots & \dots & \\ 0 & 0 & 0 & \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \dots & \gamma_0 \end{pmatrix}$$

$$\Rightarrow \quad D_T^{-1}(\sum_{t=1}^{T} X_t X_t')D_T^{-1} \xrightarrow{d} \begin{pmatrix} Q_{3\times 3} & 0 \\ 0 & V_{p\times p} \end{pmatrix} \quad (63)$$

with $Q_{3\times 3} = \begin{pmatrix} \Psi^6\sigma^6 \int_0^1 W(r)^6 dr & \Psi^8\sigma^8 \int_0^1 W(r)^8 dr & \Psi^{10}\sigma^{10} \int_0^1 W(r)^{10} dr \\ \Psi^8\sigma^8 \int_0^1 W(r)^8 dr & \Psi^{10}\sigma^{10} \int_0^1 W(r)^{10} dr & \Psi^{12}\sigma^{12} \int_0^1 W(r)^{12} dr \\ \Psi^{10}\sigma^{10} \int_0^1 W(r)^{10} dr & \Psi^{12}\sigma^{12} \int_0^1 W(r)^{12} dr & \Psi^{14}\sigma^{14} \int_0^1 W(r)^{14} dr \end{pmatrix}$,

and $V_{p\times p} = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{p-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \dots & \gamma_{p-3} \\ \dots & \dots & \dots & & \dots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \dots & \gamma_0 \end{pmatrix}$

$$\Rightarrow \quad R\Big[D_T^{-1}(\textstyle\sum_{t=1}^T X_t X_t')D_T^{-1}\Big]^{-1} R' \xrightarrow{d} (I_3 \quad 0_{3\times p}) \begin{pmatrix} Q_{3\times 3}^{-1} & 0 \\ 0 & V_{p\times p}^{-1} \end{pmatrix} \begin{pmatrix} I_3 \\ 0_{p\times 3} \end{pmatrix}$$

$$\Rightarrow \quad R\Big(D_T^{-1}(\textstyle\sum_{t=1}^T X_t X_t')D_T^{-1}\Big)^{-1} R' \xrightarrow{d} Q_{3\times 3}^{-1}$$

$$\Rightarrow \quad \Big[R\Big(D_T^{-1}(\textstyle\sum_{t=1}^T X_t X_t')D_T^{-1}\Big)^{-1} R'\Big]^{-1} \xrightarrow{d} Q_{3\times 3}$$

when have: $\sum_{t=1}^T X_t u_t = \begin{pmatrix} \sum y_{t-1}^3 u_t \\ \sum y_{t-1}^5 u_t \\ \sum y_{t-1}^7 u_t \\ \sum \Delta y_{t-1} u_t \\ \sum \Delta y_{t-2} u_t \\ \dots \\ \sum \Delta y_{t-p} u_t \end{pmatrix}_{(p+3)\times 1} \quad \Rightarrow D_T^{-1} \sum_{t=1}^T X_t u_t = \begin{pmatrix} \frac{1}{T^2} \sum y_{t-1}^3 u_t \\ \frac{1}{T^3} \sum y_{t-1}^5 u_t \\ \frac{1}{T^4} \sum y_{t-1}^7 u_t \\ \frac{1}{T^{1/2}} \sum \Delta y_{t-1} u_t \\ \frac{1}{T^{1/2}} \sum \Delta y_{t-2} u_t \\ \dots \\ \frac{1}{T^{1/2}} \sum \Delta y_{t-p} u_t \end{pmatrix}$

Thus, we have showed that the inner product $\sum_{t=1}^T X_t X_t'$ of the regressor matrix from (37) including the additional regressors is asymptotically block diagonal. Applying proposition 1, Under the null hypothesis of a unit root, it follows that :

$$D_T^{-1} \sum_{t=1}^T X_t u_t \quad \xrightarrow{d} \quad \begin{pmatrix} \Psi^3\sigma^4 \int_0^1 W^3 dW \\ \Psi^5\sigma^6 \int_0^1 W^5 dW \\ \Psi^7\sigma^8 \int_0^1 W^7 dW \\ N(0, \sigma^2\gamma_0) \\ \dots \\ N(0, \sigma^2\gamma_0) \end{pmatrix}_{(p+3)\times 1} = \begin{pmatrix} A_{3\times 1} \\ B_{p\times 1} \end{pmatrix} \tag{64}$$

where $A_{3\times 1} = \begin{pmatrix} \Psi^3\sigma^4 \int_0^1 W^3 dW \\ \Psi^5\sigma^6 \int_0^1 W^5 dW \\ \Psi^7\sigma^8 \int_0^1 W^7 dW \end{pmatrix}$ and $B_{p\times 1} = \begin{pmatrix} N(0, \sigma^2\gamma_0) \\ \dots \\ N(0, \sigma^2\gamma_0) \end{pmatrix}$

Using $D_T(\hat{\beta} - \beta) = [D_T^{-1}(\sum_{t=1}^T X_t X_t')D_T^{-1}]^{-1}[D_T^{-1}(\sum_{t=1}^T X_t u_t)]$ and from 63 and 64 we have :

$$D_T(\hat{\beta} - \beta) \xrightarrow{d} \begin{pmatrix} Q_{3\times 3}^{-1} & 0 \\ 0 & V_{p\times p}^{-1} \end{pmatrix} \begin{pmatrix} A_{3\times 1} \\ B_{p\times 1} \end{pmatrix} = \begin{pmatrix} Q_{3\times 3}^{-1} A_{3\times 1} \\ V_{p\times p}^{-1} B_{p\times 1} \end{pmatrix}_{(p+3)\times 1}$$

$$\Rightarrow \quad RD_T(\hat{\beta} - \beta) \xrightarrow{d} (I_3 \quad 0_{3\times p}) \begin{pmatrix} Q_{3\times 3}^{-1} A_{3\times 1} \\ V_{p\times p}^{-1} B_{p\times 1} \end{pmatrix}_{(p+3)\times 1}$$

$$\Rightarrow \quad RD_T(\hat{\beta} - \beta) \xrightarrow{d} Q_{3\times 3}^{-1} A_{3\times 1} \tag{65}$$

Using $\left[R\left(D_T^{-1}(\sum_{t=1}^T X_tX_t')D_T^{-1}\right)^{-1}R'\right]^{-1} \xrightarrow{d} Q_{3\times3}$ and 65 we have:

$$\tilde{F}_{NL} = [(RD_T)(\hat{\beta}-\beta)]'\left\{\hat{\sigma}_T^2 R\left[D_T^{-1}(\sum_{t=1}^T X_tX_t')D_T^{-1}\right]^{-1}R'\right\}^{-1}RD_T(\hat{\beta}-\beta)$$

$$\Rightarrow \quad \tilde{F}_{NL} \xrightarrow{d} \frac{1}{\sigma^2}\left\{(Q_{3\times3}^{-1}A_{3\times1})' \cdot Q_{3\times3} \cdot Q_{3\times3}^{-1}A_{3\times1}\right\}$$

$$\Rightarrow \quad \tilde{F}_{NL} \xrightarrow{d} \frac{1}{\sigma^2}\left\{A_{3\times1}{}' Q_{3\times3}^{-1} \cdot A_{3\times1}\right\} \quad (since\ Q_{3\times3}\ is\ symmetric)$$

It follows that: $\tilde{F}_{NL} \longrightarrow_d v'Q^{-1}v \quad with,$

$$v = \begin{bmatrix} \frac{1}{4}W(1)^4 - \frac{3}{2}\int_0^1 W(r)^2dr \\ \frac{1}{6}W(1)^6 - \frac{5}{2}\int_0^1 W(r)^4dr \\ \frac{1}{8}W(1)^8 - \frac{7}{2}\int_0^1 W(r)^6dr \end{bmatrix}$$

and

$$Q = \begin{bmatrix} \int_0^1 W(r)^6dr & \int_0^1 W(r)^8dr & \int_0^1 W(r)^{10}dr \\ \int_0^1 W(r)^8dr & \int_0^1 W(r)^{10}dr & \int_0^1 W(r)^{12}dr \\ \int_0^1 W(r)^{10}dr & \int_0^1 W(r)^{12}dr & \int_0^1 W(r)^{14}dr \end{bmatrix}$$

□

8.4. Proof of Corollary 1

Proof. We consider the asymptotic behavior of the least squares estimator $\hat{\beta} = (\hat{\delta}_1^2, \hat{\delta}_1^4, \hat{\delta}_1^6)'$ under the null $\Delta y_t = u_t$ and the OLS estimator $\hat{\beta}$ can be written as:

$$\hat{\boldsymbol{\beta}}_T = \begin{pmatrix} \sum_{t=1}^T y_{t-1}^6 & \sum_{t=1}^T y_{t-1}^8 & \sum_{t=1}^T y_{t-1}^{10} \\ \sum_{t=1}^T y_{t-1}^8 & \sum_{t=1}^T y_{t-1}^{10} & \sum_{t=1}^T y_{t-1}^{12} \\ \sum_{t=1}^T y_{t-1}^{10} & \sum_{t=1}^T y_{t-1}^{12} & \sum_{t=1}^T y_{t-1}^{14} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T y_{t-1}^3 u_t \\ \sum_{t=1}^T y_{t-1}^5 u_t \\ \sum_{t=1}^T y_{t-1}^7 u_t \end{pmatrix}$$

The following results are needed:

$$\frac{1}{T^4}\sum_{t=1}^T y_{t-1}^6 = \frac{1}{T}\sum_{t=1}^T (\frac{1}{\sqrt{T}}y_{t-1})^6 = \frac{1}{T}\sum_{t=1}^T \int_{\frac{k-1}{T}}^{\frac{k}{T}} (\frac{1}{\sqrt{T}}Y_T(r)dr)^6 \longrightarrow_d \sigma^6 \int_0^1 W(r)^6dr$$

where $\frac{k-1}{T} \le t < \frac{k}{T}$, $Y_T(r) = \frac{S_{k-1}}{\sqrt{T}} = \frac{\sum_{t=1}^{k-1} u_t}{\sqrt{T}}$
similarly, we have:

$$\frac{1}{T^5}\sum_{t=1}^T y_{t-1}^8 \longrightarrow_d \sigma^8 \int_0^1 W(r)^8dr$$

$$\frac{1}{T^6}\sum_{t=1}^T y_{t-1}^{10} \longrightarrow_d \sigma^{10} \int_0^1 W(r)^{10}dr$$

$$\frac{1}{T^7}\sum_{t=1}^T y_{t-1}^{12} \longrightarrow_d \sigma^{12} \int_0^1 W(r)^{12}dr$$

$$\frac{1}{T^8}\sum_{t=1}^{T} y_{t-1}^{14} \longrightarrow_d \sigma^{14}\int_0^1 W(r)^{14}dr$$

Now using directly the continuous mapping theorem, Itô's formula, and the weak convergence of stochastic integrals one obtains a general result for $i \in \mathbb{N}_{>0}$

$$\frac{1}{T^{\frac{(i+1)}{2}}}\sum_{t=1}^{T} y_{t-1}^{i}u_t \longrightarrow_d \int_0^1 W^i(r)dW(r) = \sigma^{i+1}\Big\{\frac{1}{i+1}\int_0^1 W(1)^{(i+1)} - \frac{i}{2}\int_0^1 W(r)^{i-1}dr\Big\}$$

and we obtain:

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^{3}u_t \longrightarrow_d \int_0^1 W^3(r)dW(r) = \sigma^4\Big\{\frac{1}{4}\int_0^1 W(1)^4 - \frac{3}{2}\int_0^1 W(r)^2dr\Big\}$$

$$\frac{1}{T^3}\sum_{t=1}^{T} y_{t-1}^{5}u_t \longrightarrow_d \int_0^1 W^5(r)dW(r) = \sigma^6\Big\{\frac{1}{6}\int_0^1 W(1)^6 - \frac{5}{2}\int_0^1 W(r)^4dr\Big\}$$

$$\frac{1}{T^4}\sum_{t=1}^{T} y_{t-1}^{7}u_t \longrightarrow_d \int_0^1 W^7(r)dW(r) = \sigma^8\Big\{\frac{1}{8}\int_0^1 W(1)^8 - \frac{7}{2}\int_0^1 W(r)^6dr\Big\}$$

It is straightforward that the estimators have different convergence rates. Thus, the least squares estimators need to be scaled using the following scaling matrix: $D_T = diag(T^2, T^3, T^4)$. Denote $\hat{\beta}_T = (\hat{\delta}_1^2, \hat{\delta}_1^4, \hat{\delta}_1^6)'$ and $X_T = (y_{t-1}^3, y_{t-1}^5, y_{t-1}^7)'$ Then, we have that:

$$D_T(\hat{\beta}_T - \beta) = [D_T^{-1}(\sum_{t=1}^{T} X_tX_t')D_T^{-1}]^{-1}[D_T^{-1}(\sum_{t=1}^{T} X_tu_t)]$$

After some algebra one gets that

$$D_T(\hat{\beta}_T - \beta) \longrightarrow_L \frac{1}{\sigma^2}(\Gamma Q\Gamma)^{-1}(\Gamma v)$$

where

$$v = \begin{bmatrix} \frac{1}{4}W(1)^4 - \frac{3}{2}\int_0^1 W(r)^2dr \\ \frac{1}{6}W(1)^6 - \frac{5}{2}\int_0^1 W(r)^4dr \\ \frac{1}{8}W(1)^8 - \frac{7}{2}\int_0^1 W(r)^6dr \end{bmatrix}$$

and

$$Q = \begin{bmatrix} \int_0^1 W(r)^6dr & \int_0^1 W(r)^8dr & \int_0^1 W(r)^{10}dr \\ \int_0^1 W(r)^8dr & \int_0^1 W(r)^{10}dr & \int_0^1 W(r)^{12}dr \\ \int_0^1 W(r)^{10}dr & \int_0^1 W(r)^{12}dr & \int_0^1 W(r)^{14}dr \end{bmatrix}$$

and $\Gamma = diag(1, \sigma^2, \sigma^4)$

Our test statistic has then the following representation:

$$\tilde{F}_{NL} = (\hat{\beta}_T - \beta)'(RD_T)'[\hat{\sigma}_T^2 D_T R(\sum_{t=1}^{T} X_tX_t')^{-1}D_T R']^{-1}RD_T(\hat{\beta}_T - \beta)$$

with $R = \mathbb{I}_3$, and has the limiting distribution:

$$\tilde{F}_{NL} \longrightarrow_L ((\Gamma Q \Gamma)^{-1}(\Gamma v))'((\Gamma Q \Gamma)^{-1})^{-1}((\Gamma Q \Gamma)^{-1}(\Gamma v)) = v' Q^{-1} v$$

By the law of large numbers it is easy to show that under the null as $T \to \infty$

$$\hat{\sigma}_T^2 = \frac{1}{T-4}\sum_{t=1}^{T}(\Delta y_t - \delta_1^2 y_{t-1}^3 - \delta_1^4 y_{t-1}^5 - \delta_1^6 y_{t-1}^7)^2 \quad \xrightarrow{P} \quad \sigma_T^2$$

Under the alternative Δy_t and $y_{t-1}^i, \forall i \in \mathbb{N}_{>0}$ are $I(0)$ and thus it is readily seen that

$$\frac{1}{T}\sum_{t=1}^{T}\Delta y_t = O_P(1); \quad \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^i = O_P(1),$$

are bounded in probability. Furthermore the innovation process $(u_t)_t$ is by assumption $I(0)$ and thus

$$\frac{1}{T}\sum_{t=1}^{T} u_t = O_P(1)$$

as well. For the OLS estimate $\hat{\beta}$ we have

$$(O_P(T))^{-1}O_P(T^2) = (TO_P(1))^{-1}T^2O_P(1) = \frac{1}{T}T^2O_P(1) = TO_P(1) = O_P(T)$$

Hence, the $\tilde{F}_{NL}$ statistic diverges to infinity at the rate $O_P(D_T)$

□

Table 4: The Power of Alternative Tests against the hypothesis of global MTSTAR stationarity [in %]																
Case 2 [$\gamma = 0.8$]																
	$\sigma_\epsilon = 1$								$\sigma_\epsilon = 0.1$							
	k=0				k=4				k=0				k=4			
	QEM	KSS	W_{nl}	F_{NL}	QEM	KSS	W_{nl}	F_{NL}	QEM	KSS	W_{nl}	F_{NL}	QEM	KSS	W_{nl}	F_{NL}
$\varphi = -0.7$																
T=100	100	99.8	99.9	99.8	99.6	94.71	98.9	99.6	15.7	22.9	22.6	18.5	86	52.9	89.8	89.9
T=200	100	100	100	100	99.7	93.78	99.5	99.7	53.3	72.6	76	67.4	87.9	52.5	89.9	89.9
T=500	100	100	100	100	99.6	91.05	99.6	99.7	100	99.9	100	100	89.9	52.1	89.9	90
T=1000	100	100	100	100	99.6	88.88	99.6	99.7	100	100	100	100	90.2	51.8	89.7	89.5
$\varphi = -0.6$																
T=100	99.9	99.2	99.3	99.1	99.8	96.36	98.8	99.7	13.5	19.4	15.3	18.2	87.4	57	90.9	91.1
T=200	100	100	100	100	99.8	96.31	99.6	99.8	44.4	64.3	67.4	58.1	89.3	56.7	91.6	91.5
T=500	100	100	100	100	99.8	94.8	99.8	99.8	100	99.9	100	100	90.8	55	91.3	91.3
T=1000	100	100	100	100	99.8	93.19	99.8	99.8	100	100	100	100	91	53.8	90.9	98.9
$\varphi = -0.5$																
T=100	99.3	97.4	97.4	96.6	99.8	97	98.5	99.3	12.2	17	16.2	13.4	89.4	60.7	91.6	92.5
T=200	100	100	100	100	99.9	97.97	99.7	99.9	35.4	54.5	56.6	47.2	90.8	60.9	92.8	92.7
T=500	100	100	100	100	99.9	97.23	99.9	99.9	99.9	99.7	100	99.9	92	58.4	92.8	92.7
T=1000	100	100	100	100	99.9	96.38	99.9	99.9	100	100	100	100	92.1	56.3	92.3	92.2
$\varphi = -0.4$																
T=100	95.1	92	91.2	88.7	99.6	95.02	96.3	96.8	10.1	13.8	13.1	10.8	91.9	61.8	93.3	94.4
T=200	100	99.9	100	100	99.9	98.71	99.7	99.9	26.5	42.8	44.3	35.7	92.8	66.7	94.4	94.5
T=500	100	100	100	100	99.9	98.7	99.9	99.9	99.3	98.8	99.8	99.5	94	64.1	94.7	94.6
T=1000	100	100	100	100	100	98.32	99.9	100	100	100	100	100	94	61.3	94.3	94.3
$\varphi = -0.3$																
T=100	77.4	78	75.8	70.6	95	85.91	87.4	85.7	8.59	11.5	10.8	9.03	93	51.1	93.8	95.4
T=200	100	99.2	99.4	99.2	99.9	97.55	99.2	99.6	19.5	31.2	31.4	25	95.1	70.8	96.1	96.5
T=500	100	100	100	100	100	98.88	99.9	100	94.6	96.1	98.4	97	96.3	72.9	96.8	96.7
T=1000	100	100	100	100	100	98.68	100	100	100	100	100	100	96.3	68.7	96.6	96.5
$\varphi = -0.2$																
T=100	42.9	51.2	47.8	41.5	67	61.5	62.4	58	7.07	9.29	8.53	7.18	82.3	23.7	79.6	84
T=200	96.5	92.5	93	91.2	98.1	88.89	92.9	93.4	13.2	20.1	19.8	15.8	97.1	54.7	97.4	98
T=500	100	100	100	100	99.9	98.46	99.9	99.9	73.3	85.1	89.2	83.7	98.1	83.3	98.3	98.3
T=1000	100	100	100	100	100	98.53	99.9	100	100	99.9	100	100	98.4	81.7	98.5	98.5
$\varphi = -0.1$																
T=100	14.1	20.4	18.5	14.9	19.7	23.92	22.9	19.7	5.47	7.17	6.54	6	25.6	6.64	27.6	25.1
T=200	46.4	54.4	52.7	46.4	55.9	52.62	54.6	51.1	8.49	11.5	11.1	9.11	72.9	17	76.4	73.8
T=500	99.8	97.4	98.1	97.7	99.4	93.57	96.1	95.8	31.7	48.8	50	40.9	99.4	78.9	99.4	99.6
T=1000	100	100	100	100	99.9	98.52	99.8	99.9	96	96.2	98.4	97	99.6	95.1	99.7	99.7
$\varphi = -0.05$																
T=100	6.63	9.55	8.56	7.12	7.78	10.38	9.59	8.05	4.45	5.97	5.34	5.02	6.68	4.55	7.73	6.77
T=200	15.1	22.5	21	17.1	16.9	22.64	22	18.4	6.26	8.39	7.87	6.79	18.6	8.56	25.7	20.8
T=500	71.5	71.3	71.4	66.6	70.1	67.71	68.1	63.5	15	22.1	21.6	17.1	86.2	46.4	86	83.7
T=1000	99.9	97.3	98.2	97.8	99.7	95.83	97.1	96.9	55.1	70.2	72.9	64.8	99.9	91.5	99.5	99.7

References

Abadir, K., Distaso, W., 2007. Testing joint hypotheses when one of the alternatives is one-sided. Journal of Econometrics 140, 695–718.

Abuaf, N., Jorion, P., 1990. Purchasing power parity in the long run. Journal of Finance 45, 157–174.

Balke, N., Fomby, T., 1997. Threshold cointegration. International Economics Reviews 38, 627–647.

Bec, F., M. B. S., Carrasco, M., 2004. Detecting mean reversion in real exchange rates from a multiple regime star model, rCER Working Papers 509, University of Rochester - Center for Economic Research (RCER).

Buncic, D., 2009. Understanding forecast failure of estar models of real exchange rates. EERI Research Paper Series, Economics and Econometrics Research Institute (EERI) 18.

Chan, K., Tong, H., 1985. On the use of the deterministic lyapounov function for the ergodicity of stochastic difference equations. Advances in Applied Probability 17, 666–678.

Chen, R., R.S.Tsay, 1993. Functional-coefficient autoregressive models. American Statistical Association 88, 298–308.

D. Guégan, T. P., 1992. Power of score tests against bilinear time series analysis. Statistica Sinica 2,1, 157–171.

D van Dijk, T. T., Franses, P., 2002. Smooth transition autoregressive models -a survey of recent developments. Econometric Reviews 21, 1–47.

Dickey, D., Fuller, W., 1979. Distribution of the estimators for autoregressive time series. Journal of the American Statistical Association 74, 427–431.

Dieu-Hang, T., Kompas, T., September 2010. A modification of the estar model and testing for a unit root in a nonlinear framework, discussion Paper, Crowford school of Economics and Government, Australian National University.

Donauer S, F. H., Sibbertsen, P., May 2010. Identification problems in estar models and a new model, discussion Paper No. dp-444, Universitat Hannover.

Dumas, B., 1992. Dynamic equilibrium and real exchange rate in a spatially separated world. The Review of Financial Studies 5(2), 153–80.

Durlauf, S., Phillins, P., 1988. Trends versus random walks in time series analvsis. Econometrica 56, 1333–1354.

Engel, C., 2000. Long-run ppp may not hold after all. Journal of International Economics 57, 243–73.

Fan, J., Yao, Q., 2003. Nonlinear time series: nonparametric and parametric methods. Springer-Verlag.

Haggan, V., Ozaki, T., 1981. Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. Biometrika 68, 198–196.

Kapetanios G, Y. S., Snell, A., 2003. Testing for a unit root in the nonlinear star framework. Journal of Econometrics 112, 359–379.

Kruse, R., Apr 2008. A new unit root test against estar based on a class of modified statistics, discussion Paper No. dp-398, Leibniz University of Hannover.

Lothian, J., Taylor, M. P., 1996. Real exchange rate behavior of purchasing power parity under the current float. Journal of Political Economy 104, 488–510.

Lutkepohl, H., Kratzig, H., 2004. Applied Time Series Econometrics. Cambridge University Press, Cambridge.

Luukkonen R, P. S., Terasvirta, T., 1988. Testing linearity against smooth transition autoregressive models. Biometrica 75, 491–499.

Meese, R., Rogoff, K., 1988. Was it real? the exchange rate-interest differential relation over the modern floating rate period. Journal of Finance 43, 933–948.

O'Connell, P. G. J., 1998. The overvaluation of purchasing power parity. Journal of International Economics 44, 1–19.

Park, J., Phillips, P., 1988. Statistical inference in regressions with integrated processes: Part i. Econometric Theory 4, 468–497.

Pascalau, R., August 2007. Testing for a unit root in the asymmetric nonlinear smooth transition framework, working Paper, University of Alabama, Department of Economics, Finance and Legal Studies, USA.

Phillips, P., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335–346.

R.L.Tweedie, 1975. Sufficient conditions for ergodicity and recurrence of markov chain on a general state space. Stochastic Processes and Their Applications 3, 385–402.

Said, S. E., Dickey, W. A., 1984. Testing for a unit root in autoregressive moving average models of unknown order. Biometrika 71, 599–607.

Sercu P, R. U., Hulle, C., 1995. The exchange rate in the presence of transaction costs: Implications of test of purchasing power parity. The Journal of Finance 50, 1309–1319.

Stock, J., Watson, M., 1988b. Testing for common trends. Journal ofthe American Statistical Association 83, 1097–1107.

Taylor M.P, D. P., Sarno, L., 2001. Nonlinear mean-reversion in real exchange rates: Toward a solution to the purchasing power parity puzzles. International Economic Review 42, 1015–1042.

Terasvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association 89, 208–218.

Tjøstheim, D., 1986. Estimation in nonlinear time series models. Stochastic Processes and their Applications 21, 251–273.