2011

Unit root tests for ESTAR models

Heni Puspaningrum
University of Wollongong

Yan-Xia Lin
University of Wollongong, yanxia@uow.edu.au

Chandra Gulati
University of Wollongong

Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

15-11

Unit Root Tests for ESTAR Models

Heni Puspaningrum, Yan-Xia Lin and Chandra Gulati

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Unit Root Tests for ESTAR Models

Heni Puspaningrum, Yan-Xia Lin[1] and Chandra Gulati

Centre for Statistical and Survey Methodology
School of Mathematics and Applied Statistics, University of Wollongong,
Wollongong, NSW, 2522, AUSTRALIA

Abstract

Since the introduction of augmented Dickey-Fuller unit root tests, many new types of unit root tests have been developed. Developments in nonlinear unit root tests occurred to overcome poor performance of standard linear unit root tests for nonlinear processes. Venetis *et al.* (2009) developed a unit root test for the k-ESTAR(p) model where k is the number of equilibrium levels and p is the order of autoregressive terms. Their approach may cause singularity problem because some of the regressors might be collinear. To overcome the problem, they move collinear regressors into the error term. This paper extends the work of Venetis *et al.* (2009). Using a new approach given in this paper, the singularity problem can be avoided without worrying the issue of collinearity. For some cases, simulation results show that our approach is better than other unit root tests.

Keywords: ESTAR model, unit root test, augmented Dickey-Fuller test.

1 Introduction

Since the introduction of unit root tests in Fuller (1976) and then Dickey and Fuller (1979, 1981), many new types of unit root tests have been developed. Developments in nonlinear unit root tests occurred as the standard linear unit root tests performed poorly for nonlinear processes. For example, Pippenger and Goering (1993) showed that the power of the standard DF tests falls considerably when the true alternative is a threshold autoregressive (TAR) model. Other researchers have attempted to address similar issues in the context of a TAR model; see, for example Balke and Fomby (1997), Enders and Granger (1998), Berben and van Dijk (1999), Caner and Hansen (2001) and Lo and Zivot (2001).

The smooth transition autoregressive (STAR) process developed by Granger and Terasvirta (1993) has been a popular process for modelling economic and finance data due to its generality and flexibility. Nonlinear adjustment in a STAR model allows for smooth rather than discrete adjustment in a TAR model. In a STAR model with one equilibrium, adjustment takes place in every period but the speed of adjustment varies with the extent of the deviation from the equilibrium. A 1-STAR(p) model can be expressed as follows:

$$y_t = \theta_{1,0} + \sum_{j=1}^{p} \theta_{1,j}\, y_{t-j} + \left[\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j}\, y_{t-j}\right] G(\theta, e, y_{t-d}) + \epsilon_t, \qquad t = 1, 2, \ldots, T, \tag{1}$$

where $\{\epsilon_t\}$ is a stationary and ergodic martingale difference sequence with variance σ_ϵ^2; $d \geq 1$ is a delay parameter; $(\theta, e) \in \mathbb{R}^+ \times \mathbb{R}$ where $\mathbb{R}$ denotes the real space $(-\infty, \infty)$ and $\mathbb{R}^+$ denotes the positive real space $(0, \infty)$; e is an equilibrium; θ is the coefficient in the transition function; $\theta_{1,j}$, $j = 0, \ldots, p$, are autoregressive coefficients corresponding to linear term; $\theta_{2,j}$, $j = 0, \ldots, p$, are autoregressive coefficients corresponding to nonlinear term. The transition function $G(\theta, e, y_{t-d})$ determines the speed of adjustment to the equilibrium e. Two simple transition functions suggested by Granger and Terasvirta (1993) and Terasvirta (1994) are the logistic and exponential functions:

$$G(\theta, e, y_{t-d}) = \frac{1}{1 + \exp\{-\theta(y_{t-d} - e)\}} - \frac{1}{2}, \tag{2}$$

$$G(\theta, e, y_{t-d}) = 1 - \exp\{-\theta^2(y_{t-d} - e)^2\}. \tag{3}$$

If the transition function $G(\theta, e, y_{t-d})$ is given by (2), (1) is called a logistic smooth transition autoregressive (LSTAR) model. If the transition function $G(\theta, e, y_{t-d})$ is given by (3), (1) is called an exponential smooth transition autoregressive (ESTAR) model.

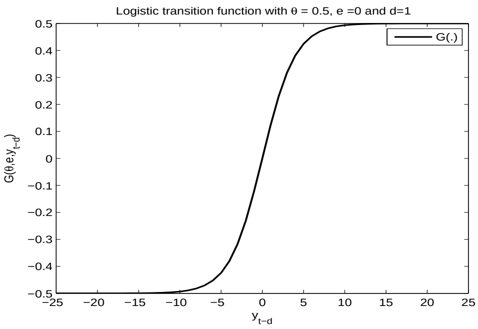
[1]Correspondence author, email: yanxia@uow.edu.au

The logistic transition function in (2) is bounded between -1/2 and 1/2 and implies asymmetric behavior of y_t depending upon whether it is above or below the equilibrium level (see Figure 1(a)). On the other hand, the exponential transition function in (3) is bounded between zero and unity and symmetrically inverse-bell shaped around equilibrium level e (see Figure 1(b)). These properties of an ESTAR model are more attractive in the present modelling context than a LSTAR model because it allows a smooth transition between regimes and symmetric adjustment of y_t for deviation above and below the equilibrium level e. The transition parameter θ in ESTAR models determines the speed of transition between the two extreme regimes, with lower absolute values of θ implying slower transition. The inner regime in ESTAR models corresponds to $y_{t-d} = e$, so that $G(\theta, e, y_{t-d}) = 0$ and (1) becomes a linear AR(p) model:

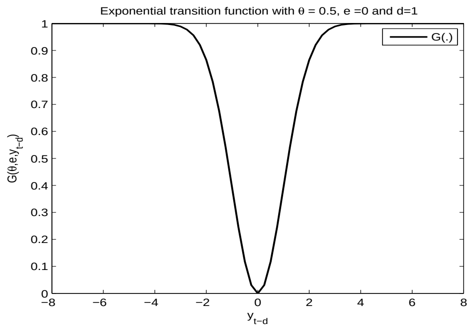
$$y_t = \theta_{1,0} + \sum_{j=1}^{p} \theta_{1,j}\ y_{t-j} + \epsilon_t, \qquad t = 1, 2, \ldots, T. \tag{4}$$

The outer regime of ESTAR models corresponds to $\lim_{(y_{t-d}-e)\to\pm\infty} G(\theta, e, y_{t-d}) = 1$, for a given θ, so that (1) becomes a different linear AR(p) model as follow:

$$y_t = \theta_{1,0} + \theta_{2,0} + \sum_{j=1}^{p} (\theta_{1,j} + \theta_{2,j})\, y_{t-j} + \epsilon_t, \qquad t = 1, 2, \ldots, T. \tag{5}$$



(a)



(b)

Figure 1: Plots of transition functions with $\theta = 0.5$, $e = 0$ and $d = 1$.

An ESTAR model has become a popular model to analyse some economic and finance data. Michael *et al.* (1997), Taylor *et al.* (2001) and Paya *et al.* (2003) used ESTAR models to analyse real exchange rate and purchasing power parity (PPP) deviations. Terasvirta and Elliasson (2001), and Sarno *et al.* (2002) used ESTAR models to analyse deviations from optimal money holding. Monoyios and Sarno (2002) found that symmetric deviations from arbitrage processes such as stock index futures follow ESTAR models. In economics and finance theories such as real exchange rate, PPP, and arbitrage processes, ESTAR models can be characterised by a unit root behaviour in the inner regime, but for large deviations, the process is mean reverting. Kapetanios *et al.* (2003) considered a unit root test for an ESTAR(1) model and applied their test to real interest rates and rejected the null hypothesis for several interest rates considered, whereas Augmented Dickey-Fuller (ADF) tests failed to do so.

Venetis *et al.* (2009) developed a unit root test for ESTAR models with the transition function involves k equilibriums:

$$G(\theta, \mathbf{e}, y_{t-d}) = 1 - \exp\left[-\theta^2 \left(\prod_{i=1}^{k}(y_{t-d} - e_i)\right)^2\right] \tag{6}$$

where $\mathbf{e} = (e_1, e_2, \ldots, e_k)'$.

As noted by Venetis *et al.* (2009), many economic theories support the existence of multiple equilibria. For example, in the case of inflation, attempts by governments to finance a substantial proportion of expenditure by seigniorage can lead to multiple inflationary equilibria (see Cagan, 1956 and Sargent and Wallace, 1973). In the case of unemployment, shocks from public produce not merely fiscal and monetary (demand policy) responses but also changes in supply-side policy affecting the equilibrium values of real variables or "natural rate" (see Diamond, 1982 and Layard *et al.*, 1991). With regard to monetary policy rules, some models suggest that real interest rates might follow a number of equilibria once the zero bound on the nominal interest rate is taken into account (see Benhabib *et al.*, 1999).

Even though Venetis *et al.* (2009) developed a unit root test for a more general form of ESTAR model but their approach might cause singularity problem because some of the regressors might be collinear. To overcome this problem, they moved collinear regressors into the error term. Even though the test under alternative hypothesis is consistent, but it may make a significant difference for some cases.

This paper extends the work of Kapetanios *et al.* (2003) by considering a unit root test for a k-ESTAR(p) model with a different approach to Venetis *et al.* (2009). By using a new approach given in this paper, singularity problems can be avoided without considering the issue of collinearity. For some cases, simulation results show that our approach is better than Venetis *et al.* (2009), Kapetanios *et al.* (2003) and the Augmented Dickey-Fuller (ADF) test of Dickey and Fuller (1979, 1981).

The rest of paper will be organised as follows: Section 2 explains the new unit root test derivation for a k-ESTAR(p) model while Section 3 explains the F-test procedure. As the asymptotic distribution of the test for a k-ESTAR(2) model does not contain a nuisance parameter while that for a k-ESTAR(p) model, $p > 2$, contains nuisance parameters, Section 4 will give further analysis of unit root test for a k-ESTAR(2) model and Section 5 will give further analysis of unit root test for a k-ESTAR(p) model, $p > 2$. Section 6 presents conclusions of this paper.

The following standard notation is used subsequently, i.e.: $\int W = \int_0^1 W(s)ds$ where $W(s)$ is the standard Brownian motion defined on $s \in [0, 1]$; "$\Rightarrow$" means convergence in distribution; "$\rightarrow$" means convergence in probability; $X_t = o_p(1)$ means that $X_t \rightarrow 0$ in probability as $t \rightarrow \infty$; $X_t = O_p(1)$ means that X_t is bounded in probability, i.e. for every $\varepsilon > 0$ there is an $M < \infty$ such that $P(|X_t| > M) < \varepsilon$ for all t.

2 A New Approach of Unit Root Test for a k-ESTAR(p) Model

In this section we develop a unit root test for a k-ESTAR(p) model with a slightly different approach compared to Venetis *et al.* (2009). Especially, we are interested in the case where y_t is a zero mean series[2] and has a unit root in the linear term, i.e. $\theta_{1,0} = 0$ and $\sum_{j=1}^{p} \theta_{1,j} = 1$ [3]. Consider a k-ESTAR(p) model in (1) and (6). An equilibrium level e_i can be defined as any real number y^* that solves the system

$$y^* = y^* + \left[\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j}\, y^*\right] G(\theta, \mathbf{e}, y^*)$$

or

$$0 = \left[\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j}\, y^*\right] G(\theta, \mathbf{e}, y^*). \tag{7}$$

One of solutions for (7) is $y_1^* = -\theta_{2,0}/\sum_{j=1}^{p} \theta_{2,j}$ where $\sum_{j=1}^{p} \theta_{2,j} \neq 0$. This solution is named as the first equilibrium e_1 with the other solutions[4] $y_i^* = e_i$, $i = 2, \cdots, k$ named further equilibriums. Note that when $\sum_{j=1}^{p} \theta_{2,j} \neq 0$ and $e_1 = -\theta_{2,0} \sum_{j=1}^{p} \theta_{2,j}$, then $e_1 = 0$ if only if $\theta_{2,0} = 0$. Therefore, if $\theta_{2,0} = 0$, one of the equilibriums should be zero.

Venetis *et al.* (2009) rearrange (1) to become

$$\Delta y_t \quad = \quad \theta_{1,0} + \sum_{j=1}^{p} \theta_{1,j}\, y_{t-1} + \sum_{j=1}^{p-1} \theta_{1,j}^{*}\, \Delta y_{t-j} + \left[\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j}\, y_{t-j}\right] G(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \tag{8}$$

where $t = 1, 2, \ldots, T$; $\theta_{1,j}^{*} = -\sum_{k=j+1}^{p} \theta_{1,k}, \quad j = 1, \cdots, (p-1)$.

[2]If y_t is not a zero mean series, we can de-mean the series so that the adjustment series will has a zero mean. This de-mean strategy was also applied in He and Sandberg (2005). Empirical examples in Venetis *et al.* (2009) and Monoyios and Sarno (2002) support the assumption is satisfied in practice.

[3]Note that when $\sum_{j=1}^{p} \theta_{1,j} = 1$ hold, the restriction $-2 < \sum_{j=1}^{p} \theta_{2,j} < 0$ ensures ergodicity of the process.

[4]Following Bair and Haesbroeck (1997) further differentiation reveals that e_i, $i = 2, 3, \cdots, k$, is monotonously semistable from below if $e_i > -\theta_{2,0}/\sum_{j=1}^{p} \theta_{2,j}$, and monotonously semistable from above if $e_i < -\theta_{2,0}/\sum_{j=1}^{p} \theta_{2,j}$.

Instead of rearrange (1) to become (8), we rearrange (1) to become

$$y_t = \theta_{1,0} + \sum_{j=1}^{p} \theta_{1,j}\ y_{t-1} + \sum_{j=1}^{p-1} \theta^*_{1,j}\ \Delta y_{t-j} + \left[\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j}\ y_{t-1} + \sum_{j=1}^{p-1} \theta^*_{2,j}\ \Delta y_{t-j}\right] G(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \tag{9}$$

where $t = 1, 2, \ldots, T$; $\theta^*_{i,j} = -\sum_{k=j+1}^{p} \theta_{i,k}$, $j = 1, \cdots, (p-1)$ and $i = 1, 2$.

Let $\theta_{1,0} = 0$ and $\sum_{j=1}^{p} \theta_{1,j} = 1$ meaning that $\{y_t\}$ has a unit root without a drift in the linear term[5]. Furthermore, without loss generality, assume that $\theta_{2,0} = 0$ so that $e_1 = 0$. Thus, (9) can be arranged to become

$$\Delta y_t = \sum_{j=1}^{p-1} \theta^*_{1,j}\ \Delta y_{t-j} + \left[\sum_{j=1}^{p} \theta_{2,j}\ y_{t-1} + \sum_{j=1}^{p-1} \theta^*_{2,j}\ \Delta y_{t-j}\right] G^*(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \qquad t = 1, 2, \ldots, T, \tag{10}$$

where

$$G^*(\theta, \mathbf{e}, y_{t-d})) = 1 - \exp\left\{-\theta^2 y_{t-d}^2 \left[\prod_{i=2}^{k} (y_{t-d} - e_i)\right]^2\right\}, \tag{11}$$

with $e_1 = 0$.

3 F-test Procedure

In this section, we develop a F-test for testing the null unit root hypothesis, $H_0 : \theta = 0$ against the alternative hypothesis of globally stationary k-ESTAR(p) model in (10). Like the 1-ESTAR(1) model in Kapetanios *et al.* (2003) and k-ESTAR(p) in Venetis *et al.* (2009), testing H_0 can not be done directly due to a well known identification problem. We use the same strategy in Venetis *et al.* (2009) to solve the problem by using a second order Taylor approximation to the nonlinear function around $\theta = 0$ in (11).

$$\begin{aligned} G^*(\theta, \mathbf{e}, y_{t-d})) &= 1 - \exp\left\{-\theta^2\ y_{t-d}^2 \left[\prod_{i=2}^{k} (y_{t-d} - e_i)\right]^2\right\} \\ &\approx \theta^2\ y_{t-d}^2 \left[\prod_{i=2}^{k} (y_{t-d} - e_i)\right]^2 + R \\ &= \theta^2\ y_{t-d}^2 \left(\delta_0 + \sum_{s=1}^{2(k-1)} \delta_s\ y_{t-d}^s\right) + R \\ &= \theta^2 \sum_{s=0}^{2(k-1)} \delta_s\ y_{t-d}^{s+2} + R, \end{aligned} \tag{12}$$

where R is the remainder, $\delta_0 = (\prod_{i=2}^{k} e_i)^2$ and $\delta_{2(k-1)} = 1$.

Substituting (12) into (10),

$$\Delta y_t = \sum_{j=1}^{p-1} \theta^*_{1,j}\ \Delta y_{t-j} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\ y_{t-1}\ y_{t-d}^{s+2} + \sum_{s=0}^{2(k-1)} \sum_{j=1}^{p-1} \gamma_{2,sj}\ y_{t-d}^{s+2} \Delta y_{t-j} + \epsilon^*_t, \tag{13}$$

where $\gamma_{1,s} = \theta^2 \delta_s \sum_{j=1}^{p} \theta_{2,j}$; $\gamma_{2,sj} = \theta^2 \delta_s \theta^*_{2,j}$; $s = 0, 1, \cdots, 2(k-1)$; $j = 1, 2, \cdots, (p-1)$ and $\epsilon^*_t = \epsilon_t + R\left[\sum_{j=1}^{p} \theta_{2,j}\ y_{t-j}\right]$. If $\theta = 0$, y_t is linear in y_{t-j}, $j = 1, 2, \cdots, p$ and $\epsilon^*_t = \epsilon_t$ since the remainder $R \equiv 0$.

[5]This paper only considers test for a unit root without a drift because in our next paper, we want to apply the test to a series in pair trading which is not possible if the series has a drift or trend.

Testing the null hypothesis of a unit root against the alternative of a globally stationary k-ESTAR(p) model is equivalent to testing,

$$H_0 : \gamma_{1,s} = \gamma_{2,sj} = 0, \qquad \text{for all } s \text{ and } j \text{ in (13)}$$

against

$$H_1 : \text{ Not all } \gamma_{1,s} \text{ and } \gamma_{2,sj} = 0.$$

Under the null hypothesis H_0, it follows for (13) that

$$\begin{aligned}
\Delta y_t &= \sum_{j=1}^{p-1} \theta^*_{1,j}\, \Delta y_{t-j} + \epsilon_t, && (14)\\
\left(1 - \sum_{j=1}^{p-1} \theta^*_{1,j}\, L^j\right) \Delta y_t &= \epsilon_t, \\
\Delta y_t &= \left(1 - \sum_{j=1}^{p-1} \theta^*_{1,j}\, L^j\right)^{-1} \epsilon_t, \\
&= \sum_{j=0}^{\infty} c_j\, \epsilon_{t-j}, \\
&= C(L)\epsilon_t = \eta_t, && (15)
\end{aligned}$$

where L is the lag operator, i.e. $Ly_t = y_{t-1}$. We assume that the sequence $\{\eta_t\}$ satisfies the following assumption:

Assumption 1 *Assumptions for* $\{\eta_t\}$ *:*

- $\eta_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j} = C(L)\epsilon_t$, *where* $\{\epsilon_t\}$ *is a stationary and ergodic martingale differences sequence (MDS) with natural filtration* $\mathcal{F}_t = \sigma(\{\epsilon_i\}_{-\infty}^{t})$, *variance* σ^2_ϵ, *and* $E|\epsilon_t|^{6+r} < \infty$ *for some* $r > 0$.
- $C(L) = \sum_{j=0}^{\infty} c_j L^j$ *is a one-sided moving average polynomial in the lag operator such that* $C(1) \neq 0$ *(no unit root),* $\sum_{j=0}^{\infty} c_j = C(1) < \infty$ *and* $\sum_{j=0}^{\infty} j^p |c_j|^p < \infty$ *(one-sumability and p-sumability) for* $p \geq 1$.

Following Phillips and Solo (1992), the Beveridge-Nelson (BN subsequently) decomposition (see Beveridge and Nelson, 1981) will be applied. We start with the BN lemma as follows:

Lemma 1 *(Lemma 2.1 in Phillips and Solo, 1992). Let* $C(L) = \sum_{j=0}^{\infty} c_j L^j$. *Then*

$$C(L) = C(1) - (1 - L)\widetilde{C}(L),$$

where $\widetilde{C}(L) = \sum_{j=0}^{\infty} \widetilde{c}_j L^j$ *and* $\widetilde{c}_j = \sum_{k=j+1}^{\infty} c_k$. *If* $p \geq 1$, *then*

$$\sum_{j=1}^{\infty} j^p |c_j|^p < \infty \Rightarrow \sum_{j=0}^{\infty} |\widetilde{c}_j|^p < \infty \text{ and } |C(1)| < \infty.$$

If $p < 1$, *then*

$$\sum_{j=1}^{\infty} j |c_j|^p < \infty \Rightarrow \sum_{j=0}^{\infty} |\widetilde{c}_j|^p < \infty.$$

Before we derive the F-test statistic for the unit root test for a k-ESTAR(p) model, we present the theorem below used in the F-test statistic derivation.

Theorem 1 *Assume that* $\{\eta_t\}_{t=1}^{\infty}$ *and* $\{\epsilon_t\}_{t=1}^{\infty}$ *satisfy Assumption 1. Let* $y_t = \sum_{i=0}^{t} \eta_i$, $t = 1, 2, \ldots, T$, *with* $y_0 = 0$. *Denote* $\lambda = \sigma_\epsilon C(1)$ *and* $\gamma_j = E(\eta_t \eta_{t-j}) = \sigma^2_\epsilon \sum_{s=0}^{\infty} c_s c_{s+j}$, $j = 0, 1, \ldots,$ *for all* t. *Then, under* H_0, *the following sums converge jointly.*

(a) $T^{-1} \sum_{t=p+1}^{T} \left(\frac{y_t}{\sqrt{T}}\right)^q \Rightarrow \lambda^q \int W^q$,

(b) $T^{-1}\sum_{t=p+1}^{T}\left(\frac{y_{t-1}}{\sqrt{T}}\right)^2\left(\frac{y_{t-d}}{\sqrt{T}}\right)^q \Rightarrow \lambda^{q+2}\int W^{q+2}$,

(c) $T^{-1}\sum_{t=p+1}^{T}\eta_{t-i}\eta_{t-j} \Rightarrow \gamma_{|j-i|}, \quad i,j=1,\ldots,(p-1)$,

(d) $T^{-1}\sum_{t=p+1}^{T}\left(\frac{y_{t-d}}{\sqrt{T}}\right)^q\eta_{t-i}\eta_{t-j} \Rightarrow \gamma_{|j-i|}\lambda^q\int W^q, \quad i,j=1,\ldots,(p-1)$,

(e) $T^{-1/2}\sum_{t=p+1}^{T}\left(\frac{y_{t-1}}{\sqrt{T}}\right)\left(\frac{y_{t-d}}{\sqrt{T}}\right)^q\frac{\eta_{t-i}}{\sqrt{T}} \Rightarrow 0, \quad i=1,\ldots,(p-1)$,

(f) $\sum_{t=p+1}^{T}\frac{y_{t-1}}{\sqrt{T}}\left(\frac{y_{t-d}}{\sqrt{T}}\right)^q\frac{\epsilon_t}{\sqrt{T}} \Rightarrow \sigma_\epsilon\lambda^{q+1}\int W^{q+1}dW$,

(g) $T^{-1/2}\sum_{t=p+1}^{T}\eta_{t-i}\epsilon_t \Rightarrow \sqrt{\gamma_0}\sigma_\epsilon W_i(1), \quad i=1,\ldots,(p-1)$,

(h) $\sum_{t=p+1}^{T}\left(\frac{y_{t-1}}{\sqrt{T}}\right)^q\frac{\eta_{t-i}\epsilon_t}{\sqrt{T}} \Rightarrow \sqrt{\gamma_0}\sigma_\epsilon\lambda^q\int W^q dW_i, \quad i=1,\ldots,(p-1)$,

as $T\to\infty$. $\int W^x = \int_0^1 W(s)^x ds$ *and* $\int W^x dW = \int_0^1 W(s)^x dW(s)$ *where* $W(s)$ *is the standard Brownian motion defined on* $s\in[0,1]$ *and* x *is an integer number.* W *is a standard Brownian motion corresponding to MDS* $\{\epsilon_t\}$ *and* W_i *a standard Brownian motion corresponding to MDS* $\{\eta_{t-i}\epsilon_t\}$, $i=1,2,\ldots,(p-1)$. *Note that* W *and* W_i *are independent as* $Cov(\epsilon_t,\eta_{t-i}\epsilon_t)=E(\eta_{t-i}\epsilon_t^2)=0$.

Proof: see Appendix A.1

Theorem 2 *Let us write (13) as a partitioned regression model,*

$$Y = X_1\,\mathbf{b}_1 + X_2\,\mathbf{b}_2 + \epsilon_t^* \tag{16}$$

where

$$\begin{aligned}
Y &= [\Delta y_{p+1},\Delta y_{p+2},\cdots,\Delta y_T]' \\
X_1 &= \big[(\Delta y_p,\Delta y_{p+1},\cdots,\Delta y_{T-1})',(\Delta y_{p-1},\Delta y_p,\cdots,\Delta y_{T-2)})',\cdots, \\
&\quad (\Delta y_2,\Delta y_3,\cdots,\Delta y_{T-(p-1)})'\big] \\
X_2 &= \big[(y_p y_{p+1-d}^2,\cdots,y_{T-1}y_{T-d}^2)',(y_p y_{p+1-d}^3,\cdots,y_{T-1}y_{T-d}^3)',\cdots, \\
&\quad (y_p y_{p+1-d}^{2k},\cdots,y_{T-1}y_{T-d}^{2k})',(y_{p+1-d}^2\Delta y_p,\cdots,y_{T-d}^2\Delta y_{T-1})',\cdots, \\
&\quad (y_{p+1-d}^2\Delta y_2,\cdots,y_{T-d}^2\Delta y_{T-(p-1)})',(y_{p+1-d}^3\Delta y_p,\cdots,y_{T-d}^3\Delta y_{T-1})',\cdots, \\
&\quad (y_{p+1-d}^3\Delta y_2,\cdots,y_{T-d}^3\Delta y_{T-(p-1)})',\cdots,(y_{p+1-d}^{2k}\Delta y_p,\cdots,y_{T-d}^{2k}\Delta y_{T-1})',\cdots, \\
&\quad (y_{p+1-d}^{2k}\Delta y_2,\cdots,(y_{T-d}^{2k}\Delta y_{T-(p-1)})'\big] \\
\mathbf{b}_1 &= \left(\theta_{11}^*,\theta_{12}^*,\ldots,\theta_{1(p-1)}^*\right)' \\
\mathbf{b}_2 &= \left(\gamma_1,\gamma_{21},\cdots,\gamma_{2(2k-2)},\gamma_{31},\cdots,\gamma_{3(p-1)},\gamma_{411},\cdots,\gamma_{4(2k-2)(p-1)}\right)'.
\end{aligned}$$

Under the null hypothesis of $H_0:\theta=0$, $\epsilon_t^*=\epsilon_t$, *an F-type test can be constructed. The F-type statistic to test the null hypothesis of a unit root without a drift against the alternative of a globally stationary k-ESTAR(p) model is*

$$F_{nl} = \frac{1}{\hat\sigma_{\epsilon^*}^2}\left(\hat{\mathbf{b}}_2-\mathbf{b}_2\right)'(X_2'M_1X_2)\left(\hat{\mathbf{b}}_2-\mathbf{b}_2\right) \tag{17}$$

where $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$ *is orthogonal to the* X_1 *projection matrix and* $\hat\sigma_{\epsilon^*}^2$ *is the maximum likelihood estimator of the error variance. The statistic*

$$F_{nl} \Rightarrow F_1'(W)F_2^{-1}(W)F_1(W), \text{ as } T\to\infty \tag{18}$$

where $F_1(W)$ *and* $F_2(W)$ *are described below.*

Let W denote a standard Brownian motion,

$$F_1(W) = \begin{bmatrix} \int W^3 dW \\ \vdots \\ \int W^{(2k+1)} dW \\ (\int W^2 dW_1 - W_1(1) \int W^2) \\ \vdots \\ (\int W^2 dW_{(p-1)} - W_{(p-1)}(1) \int W^2) \\ (\int W^3 dW_1 - W_1(1) \int W^3) \\ \vdots \\ (\int W^3 dW_{(p-1)} - W_{(p-1)}(1) \int W^3) \\ \vdots \\ (\int W^{2k} dW_1 - W_1(1) \int W^{2k}) \\ \vdots \\ (\int W^{2k} dW_{(p-1)} - W_{(p-1)}(1) \int W^{2k}) \end{bmatrix},$$

$$F_2(W) = \begin{bmatrix} F_{21}(W) & \mathbf{0} \\ \mathbf{0} & F_{22}(W) \end{bmatrix},$$

$$F_{21}(W) = \begin{bmatrix} \int W^6 & \cdots & \int W^{(2k+4)} \\ \vdots & \ddots & \vdots \\ \int W^{(2k+4)} & \cdots & \int W^{(4k+2)} \end{bmatrix},$$

$$F_{22}(W) = \begin{bmatrix} (\int W^4 - (\int W^2)^2)\,\mathbf{\Pi} & \cdots & (\int W^{2k+2} - \int W^2 \int W^{2k})\,\mathbf{\Pi} \\ \vdots & \ddots & \vdots \\ (\int W^{2k+2} - \int W^2 \int W^{2k})\,\mathbf{\Pi} & \cdots & (\int W^{4k} - (\int W^{2k})^2)\,\mathbf{\Pi} \end{bmatrix}$$

and

$$\mathbf{\Pi} = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-2} \\ \rho_1 & 1 & \cdots & \rho_{p-3} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix}_{(p-1)\times(p-1)}, \tag{19}$$

where $\int W^x = \int_0^1 W(s)^x ds$ and $\int W^x dW = \int_0^1 W(s)^x dW(s)$ where $W(s)$ is the standard Brownian motion defined on $s \in [0, 1]$ and x in an integer number. W is a standard Brownian motion corresponding to MDS $\{\epsilon_t\}$ and W_i a standard Brownian motion corresponding to MDS $\{\eta_{t-i}\epsilon_t\}$, $i = 1, 2, \ldots, (p-1)$. Note that W and W_i are independent as $Cov(\epsilon_t, \eta_{t-i}\epsilon_t) = E(\eta_{t-i}\epsilon_t^2) = 0$ and also note that $\{\eta_{t-i}\epsilon_t\}$ is MDS. ρ_i, $i = 1, \ldots, (p-2)$, are constants corresponding to correlation between Δy_t and Δy_{t-i}.

Proof: see Appendix A.2

4 Unit Root Test Analysis for a k-ESTAR(2) Model

In this section, a unit root test analysis for a k-ESTAR(2) model is considered. For this model, we can resolve the singularity problem in Venetis *et al.* (2009). For this model, the test does not involve the nuisance parameter $\mathbf{\Pi}$ in (19). We compare the performance of our approach and other unit root tests.

Now, consider (10) for a k-ESTAR(2) model,

$$\Delta y_t = \theta^*_{1,1}\,\Delta y_{t-1} + \left[(\theta_{2,1} + \theta_{2,2}) y_{t-1} + \theta^*_{2,1}\,\Delta y_{t-1}\right] G^*(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, t = 1, 2, \ldots, T \tag{20}$$

where $\theta > 0$ and $\{\epsilon_t\}$ is a stationary and ergodic martingale difference sequence with variance σ_ϵ^2 and $G^*(\theta, \mathbf{e}, y_{t-d})$ as in (11).

Recall the Taylor approximation for $G^*(\theta, \mathbf{e}, y_{t-d})$ around $\theta = 0$ in (12),

$$G^*(\theta, \mathbf{e}, y_{t-d}) \approx \theta^2 \sum_{s=0}^{2(k-1)} \delta_s y_{t-d}^{s+2} + R,$$

where R is the remainder, $\delta_0 = (\prod_{i=2}^{k} e_i)^2$ and $\delta_{2(k-1)} = 1$. Thus, (20) becomes,

$$\Delta y_t = \theta_{1,1}^* \; \Delta y_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s} \; y_{t-1} y_{t-d}^{s+2} + \sum_{s=0}^{2(k-1)} \gamma_{2,s} \; y_{t-d}^{s+2} \Delta y_{t-1} + \epsilon_t^* \tag{21}$$

where $\gamma_{1,s} = \theta^2 \delta_s (\theta_{2,1} + \theta_{2,2})$, $\gamma_{2,s} = \theta^2 \delta_s \theta_{2,1}^*$ and $\epsilon_t^* = \epsilon_t + R \left[\sum_{j=1}^{2} \theta_{2,j} \; y_{t-j}\right]$. If $\theta = 0$, y_t in (20) is linear in y_{t-1} and y_{t-2}, and $\epsilon_t^* = \epsilon_t$ since the remainder $R \equiv 0$.

Testing the null hypothesis of a unit root ($H_0 : \theta = 0$) against alternative of a globally stationary k-ESTAR(2) model is equivalent to testing,

$$H_0 : \gamma_{1,s} = \gamma_{2,s} = 0, \qquad \text{for all } s \text{ in (21) against its complement.}$$

Under the null hypothesis H_0, $\epsilon_t^* = \epsilon_t$, thus, (21) becomes

$$\Delta y_t = \theta_{1,1}^* \; \Delta y_{t-1} + \epsilon_t = \sum_{j=0}^{\infty} c_j \; \epsilon_{t-j} = C(L)\epsilon_t = \eta_t, \tag{22}$$

where L is the lag operator, i.e. $Ly_t = y_{t-1}$, and $\theta_{1,1}^* = -\theta_{1,2}$.

Following the results from Section 2, for $p = 2$, $\mathbf{\Pi}$ will become a constant 1, thus the F_{nl} statistic in (18) becomes

$$F_{nl} = (F_1(W))'(F_2(W))^{-1}F_1(W) \tag{23}$$

where

$$F_1(W) = \begin{bmatrix} \int W^3 dW \\ \vdots \\ \int W^{(2k+1)} dW \\ (\int W^2 dW_1 - W_1(1) \int W^2) \\ \vdots \\ (\int W^{2k} dW_1 - W_1(1) \int W^{2k}) \end{bmatrix},$$

$$F_2(W) = \begin{bmatrix} F_{21}(W) & \mathbf{0} \\ \mathbf{0} & F_{22}(W) \end{bmatrix},$$

$$F_{21}(W) = \begin{bmatrix} \int W^6 & \cdots & \int W^{(2k+4)} \\ \vdots & \ddots & \vdots \\ \int W^{(2k+4)} & \cdots & \int W^{(4k+2)} \end{bmatrix}$$

and

$$F_{22}(W) = \begin{bmatrix} (\int W^4 - (\int W^2)^2) & \cdots & (\int W^{2k+2} - \int W^2 \int W^{2k}) \\ \vdots & \ddots & \vdots \\ (\int W^{2k+2} - \int W^2 \int W^{2k}) & \cdots & (\int W^{4k} - (\int W^{2k})^2) \end{bmatrix}.$$

Note that even if the limit distribution of F_{nl} for a k-ESTAR(2) model in (23) does not depend on any nuisance parameters, special attention is needed for values of $\theta_{1,1}^*$ close to -1 or 1. Under the null hypothesis, y_t is a function of $\theta_{1,1}^*$, as is seen from (22). Thus, the time series Δy_t is then close to having a unit root or becoming nonstationary. In these situations the test may reject the null hypothesis too often.

In comparison with Venetis *et al.* (2009), denote the test statistic by F_{VPP} after the authors. Their approach will consider,

$$\Delta y_t \quad = \quad \theta_{1,1}^* \; \Delta y_{t-1} + [\theta_{2,1} \; y_{t-1} + \theta_{2,2} \; y_{t-2}] G^*(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \quad t = 1, 2, \ldots, T, \tag{24}$$

rather than (20). Substituting the Taylor approximation for $G^*(\theta, \mathbf{e}, y_{t-d})$ around $\theta = 0$ in (12) into

Table 1: Asymptotic critical values of F test statistics for k-ESTAR(2) models.

	Sig. Level					
	0.1	0.05	0.01	0.1	0.05	0.01
k		F_{nl}			F_{VPP}	
1	5.49	6.94	10.37	3.73	4.88	7.73
2	13.83	15.98	20.80	9.54	11.36	15.47
3	20.44	23.18	28.61	13.64	15.70	19.94
4	26.64	29.65	36.64	17.06	19.38	28.61

Note: Simulations were based on samples size T=10,000 and 50,000 replications.

(24) gives

$$\Delta y_t = \theta^*_{1,1}\ \Delta y_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\ y_{t-1}\ y^{s+2}_{t-d} + \sum_{s=0}^{2(k-1)} \gamma_{2,s}\ y_{t-2}\ y^{s+2}_{t-d} + \epsilon^*_t \tag{25}$$

rather than (21). Here ϵ^*_t is defined as in (21). Since asymptotically $y_{t-1}y^{s+2}_{t-d}$ are collinear with $y_{t-2}y^{s+2}_{t-d}$, for $s = 0, 1, \ldots, 2(k-1)$, Venetis *et al.* (2009) rearranged (25) to become

$$\Delta y_t = \theta^*_{1,1}\ \Delta y_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\ y_{t-1}\ y^{s+2}_{t-d} + \upsilon_t \tag{26}$$

where $\upsilon_t = \sum_{s=0}^{2(k-1)} \gamma_{2,s}\ y_{t-2}\ y^{s+2}_{t-d} + \epsilon^*_t$. Thus, they moved regressors to the error term and formed a new error term.

Using (26), the asymptotic F_{VPP} test statistics will be

$$F_{VPP} = (F_1(W))'(F_2(W))^{-1}F_1(W) \tag{27}$$

where

$$F_1(W) = \begin{bmatrix} \int W^3 dW \\ \int W^4 dW \\ \vdots \\ \int W^{(2k+1)} dW \end{bmatrix} \text{ and } F_2(W) = \begin{bmatrix} \int W^6 & \int W^7 & \cdots & \int W^{(2k+4)} \\ \int W^7 & \int W^8 & \cdots & \int W^{(2k+5)} \\ \vdots & \vdots & \ddots & \vdots \\ \int W^{(2k+4)} & \int W^{(2k+5)} & \cdots & \int W^{(4k+2)} \end{bmatrix}.$$

The expressions of $F_1(W)$ and $F_2(W)$ in F_{nl} and F_{VPP} are different. This is due to the fact that the F_{VPP} has some regressors added into the error term so that the expression of $F_1(W)$ and $F_2(W)$ in F_{nl}, i.e. our approach, have additional terms, compared to the F_{VPP}. The advantage of our approach is that it has solved the singularity problem for this case without the need to add regressors into the errors term. It will maximise the usage of the information of data.

Asymptotic critical values for F-type statistics from F_{nl} in (23) and F_{VPP} in (27) with $k = 1, \ldots, 4$ are obtained via stochastic simulations and presented in Table 1.[6]

As suggested by Venetis *et al.* (2009), for computational purposes F_{nl} and F_{VPP} can be easily calculated following the steps below:

1. Estimate the unrestricted model on (21) for F_{nl} or (26) for F_{VPP} and keep the sum of squared residuals SSR_U.

2. Estimate (22) as the restricted model implied by the null hypothesis and keep the sum of squared residuals SSR_R. Note that based on the null hypothesis, F_{nl} and F_{VPP} have the same restricted model.

3. Calculate the ratio $F = T(SSR_R - SSR_U)/SSR_U$ where T denotes the number of observations in the restricted model and then compare with the critical values in Table 1.

[6] Critical values for the VPP in this table are quite different to the values in Table 2b in Venetis *et al.* (2009) as they did not assume that $\theta_{2,0} = 0$ when $e_i = 0$ for a certain i, $i = 1, \ldots, k$.

4.1 Sufficient Conditions for Stationarity of a k-ESTAR(2) Model

For a k-ESTAR(2) model, we determine a set of sufficient conditions for parameter combinations corresponding to a stationary series. Knowing the conditions will be useful in doing the simulation study presented in the next subsection.

Let us rearrange y_t in (20) as follows,

$$y_t = (\theta_{1,1} + \theta_{2,1} G^*(\theta, \mathbf{e}, y_{t-d}))\, y_{t-1} + (\theta_{1,2} + \theta_{2,2}\, G^*(\theta, \mathbf{e}, y_{t-d}))\, y_{t-2} + \epsilon_t. \tag{28}$$

Considering (28) as an AR(2) process, the necessary stationarity conditions for this process (see Box and Jenkins, 1976, p. 58) are,

$$(\theta_{1,1} + \theta_{2,1}\, G^*(\theta, \mathbf{e}, y_{t-d})) + (\theta_{1,2} + \theta_{2,2}\, G^*(\theta, \mathbf{e}, y_{t-d})) < 1 \tag{29}$$

$$(\theta_{1,2} + \theta_{2,2}\, G^*(\theta, \mathbf{e}, y_{t-d})) - (\theta_{1,1} + \theta_{2,1}\, G^*(\theta, \mathbf{e}, y_{t-d})) < 1 \tag{30}$$

$$-1 < \theta_{12} + \theta_{2,2}\, G^*(\theta, \mathbf{e}, y_{t-d})) < 1 \tag{31}$$

Note that from (11), $0 < G^*(\theta, \mathbf{e}, y_{t-d}) < 1$, and under the null hypothesis of a unit root in the linear term, $\theta_{1,1} + \theta_{1,2} = 1$ (see the discussion in Section 2). Thus, from (29) we obtain,

$$(\theta_{2,1} + \theta_{2,2}) G^*(\theta, \mathbf{e}, y_{t-d}) < 0. \tag{32}$$

The stationarity conditions in (32) will be fulfilled if $(\theta_{2,1} + \theta_{2,2}) < 0$.

From (30), we have

$$(\theta_{1,2} - \theta_{1,1}) + (\theta_{2,2} - \theta_{2,1}) G^*(\theta, \mathbf{e}, y_{t-d}) < 1.$$

Under the null hypothesis of a unit root, $\theta_{1,1} + \theta_{1,2} = 1$. Thus,

$$\begin{aligned} ((1 - \theta_{1,1}) - \theta_{1,1}) + (\theta_{2,2} - \theta_{2,1}) G^*(\theta, \mathbf{e}, y_{t-d}) &< 1 \\ (\theta_{2,2} - \theta_{2,1}) G^*(\theta, \mathbf{e}, y_{t-d}) &< 2\theta_{1,1} \end{aligned} \tag{33}$$

The stationarity condition in (33) will be fulfilled if $0 \le (\theta_{2,2} - \theta_{2,1}) < 2\theta_{1,1}$ or $(\theta_{2,2} - \theta_{2,1}) \le 0 < 2\theta_{1,1}$ and $\theta_{1,1} \ge 0$.

To fulfill the stationarity condition in (31), $\theta_{1,2}$ should be $-1 < \theta_{1,2} < 1$, so that

$$-1 - \theta_{1,2} < \theta_{2,2} G^*(\theta, \mathbf{e}, y_{t-d}) < 1 - \theta_{1,2}.$$

Thus, the parameters should satisfy $-1 - \theta_{1,2} < \theta_{2,2} \le 0$ or $0 \le \theta_{2,2} < 1 - \theta_{1,2}$ to fulfill the stationarity conditions.

To summarise, one set of the sufficient stationarity conditions for (28) is

$$\begin{aligned} (\theta_{2,1} + \theta_{2,2}) &< 0 \\ \theta_{1,1} &\ge 0 \\ 0 \le (\theta_{2,2} - \theta_{2,1}) < 2\theta_{1,1} \text{ or } (\theta_{2,2} - \theta_{2,1}) &\le 0 \\ -1 < \theta_{1,2} &< 1 \\ -1 - \theta_{1,2} < \theta_{2,2} \le 0 \text{ or } 0 \le \theta_{2,2} &< 1 - \theta_{1,2}. \end{aligned} \tag{34}$$

4.2 Small Sample Properties of F_{nl} Test for a k-ESTAR(2) Model

In this subsection, small sample size and power performance of F_{nl} test for a k-ESTAR(2) model are undertaken using Monte Carlo experiments. For comparison sake, we include F_{VPP}, the augmented KSS test (denoted by AKSS; see Kapetanios *et al.*, 2003) and the augmented Dickey-Fuller test (denoted by ADF; see Fuller, 1976). For the AKSS test, we only consider Case 1 because based on Venetis *et al.* (2009), generally this case has more ability to detect the true model than the other cases.

The calculated F statistics from the F_{nl} and F_{VPP} are compared with the critical values in Table 1. The critical value for the t-test of AKSS test is -2.22 obtained from Table 1 in Kapetanios *et al.* (2003). The critical value for the t-test of ADF test is -1.95 obtained in Fuller (1976). For each experiment, the rejection probability (as a percentage) of the null hypothesis computed with the nominal sizes of the tests, which are set at 5%. The sample size is considered for $T = 50, 100, 200$ with the number of replications at 10,000.

4.2.1 The size of Alternative Tests

To obtain the size of the alternative tests, we generate samples from the null model, i.e.:

$$\Delta y_t = \theta^*_{1,1}\Delta y_{t-1} + \epsilon_t \tag{35}$$

where $\theta^*_{1,1} = -\theta_{12}$ and ϵ_t is drawn from the standard normal distribution. We take $\theta_{1,2} = \{-0.8, -0.5, -0.2, 0, 0.2, 0.5, 0.8\}$.

For computational purposes, the regression model in (35) becomes the restricted model for F_{nl} and F_{VPP}. Furthermore, the unrestricted models for F_{nl} and F_{VPP} are the regression models in (21) and (25) respectively. For the AKSS test and the ADF test, we include the lagged first difference (Δy_{t-1}) to overcome the autocorrelation [7], so that the regression model for the AKSS test is

$$\Delta y_t = \delta_1 y_{t-1} y^2_{t-d} + \delta_2 \Delta y_{t-1} + \epsilon_t \tag{36}$$

and the regression model for the ADF test is

$$\Delta y_t = \delta_1 y_{t-1} + \delta_2 \Delta y_{t-1} + \epsilon_t. \tag{37}$$

The null and alternative hypothesis for the AKSS test and ADF test are

$$H_0 : \delta_1 = 0 \quad vs \quad H_1 : \delta_1 < 0. \tag{38}$$

The calculated t-test statistics value for δ_1 is then compared with the critical values of the AKSS test and ADF test. The null hypothesis for the AKSS test and the ADF test conclude that y_t has a unit root without a drift. On the other hand, the alternative hypothesis for the AKSS test concludes that y_t is a globally stationary 1-ESTAR(1) model while the ADF test concludes that y_t is a stationary linear ARMA model. The size of the alternative tests are presented in Table 2.

As Venetis *et al.* (2009) noted, the F tests (F_{nl} and F_{VPP}) resemble the familiar $\mathcal{X}^2$ test when under the null hypothesis the process is stationary. For this reason, F_{nl} and F_{VPP} may suffer from size problems when the number of restrictions is large and the time series is short. As F_{nl} has a larger number of restrictions than F_{VPP}, the distortion becomes larger for F_{nl} than F_{VPP} for the same conditions. Table 2 shows that F_{nl} and F_{VPP} are oversized for large values of k and $\theta_{1,2} = -0.8$. If $\theta_{1,2}$ is close to -1, $\theta^*_{1,1}$ in (22) will be close to 1. It means that Δy_t will be close to has an explosive unit root. Generally, if the value of $\theta_{1,2}$ is close to 0 and the sample size increases from $T = 50$ to $T = 200$, F_{nl} and F_{VPP} have nominal size close to 5%. In addition to the F_{nl} and F_{VPP} tests for $k = 1, 2, 3, 4$, the rejection probabilities of the null hypothesis for the AKSS and the ADF tests are also computed in the last two columns in Table 2. For all cases, the rejection probabilities for the AKSS are less than 5%. Therefore, the AKSS test has more power to detect the null hypothesis than the other methods. It happens because this test involves fewer estimation parameters and deals with one sided alternatives of stationarity. On the other hand, even though the ADF test also involves fewer estimation parameters and deals with one sided alternatives of stationarity, the rejection probabilities for the ADF tests are close to or slightly higher than 5%. Even for $\theta_{1,2} = 0$ and $T = 200$, the rejection probability of the null hypothesis for the ADF test is 5.24%.

4.2.2 The Power of Alternative Tests

To evaluate the power of tests against the globally stationary k-ESTAR(2) model, samples were simulated from model in (20) with $d = 1$ and ϵ_t is drawn from a standard normal distribution. We calculate the rejection probabilities of the null hypothesis (percent) given that the y_t is an k-ESTAR(2) model. The simulation results are summarised in Tables 3-8.

The data for Tables 3-5 are simulated with $k = 1$, i.e. $e_1 = 0$. From the three tables, the rejection probabilities increase as k increases for the tests based on F_{nl} and F_{VPP} with $T = 50$. This may be due to large number of restrictions and short time series. Therefore, even though the rejection probabilities for the F_{nl} test with $k = 4$ are quite high (around 22%-31%), we do not recommend the results from very small sample. For $k = 1$, the rejection probabilities increase as T increases for the tests based on F_{nl} and F_{VPP}. Furthermore, for large sample ($T = 200$), the probabilities for $k = 1$ are the highest compared to other k. Apparently, with large samples, the F_{nl} and F_{VPP} tests are able to recognise the true number of equilibrium (for this case, $k = 1$). From the three tables, the F_{nl} test shows more power to detect the alternative compared to the other methods when $\theta_{1,2}$ close to 1. For

[7] Therefore we call the tests as the augmented KSS test and the augmented DF test

Table 2: The size of alternative tests (in percentage)

	F_{nl}				F_{VPP}				AKSS	ADF
	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4		
$\theta_{1,2} = -0.8$										
T=50	7.36	16.05	25.43	36.84	5.95	9.32	11.73	14.32	4.61	5.79
T=100	5.79	9.66	12.59	16.71	5.37	6.54	7.65	8.17	4.54	5.25
T=200	5.25	7.15	7.62	9.10	5.16	5.91	5.96	6.27	4.62	5.12
$\theta_{1,2} = -0.5$										
T=50	5.91	11.15	17.38	26.27	5.53	7.39	8.67	10.54	4.38	5.39
T=100	5.13	7.01	8.12	10.00	5.15	5.76	5.97	6.28	4.45	4.98
T=200	4.94	6.00	5.42	6.19	5.04	5.57	5.10	5.20	4.60	5.26
$\theta_{1,2} = -0.2$										
T=50	5.26	9.96	15.06	23.73	5.39	6.20	7.77	9.39	4.23	5.25
T=100	4.76	5.87	7.11	8.69	4.92	4.95	4.99	5.53	4.34	4.95
T=200	4.72	5.33	4.98	5.23	5.07	5.19	4.45	4.46	4.70	5.21
$\theta_{1,2} = 0$										
T=50	5.23	9.27	14.48	23.42	5.07	5.32	7.24	8.93	4.05	5.18
T=100	4.72	5.49	6.56	8.45	4.72	4.62	4.43	4.94	4.16	4.93
T=200	4.71	5.03	4.77	5.08	4.93	4.74	4.05	4.01	4.52	5.24
$\theta_{1,2} = 0.2$										
T=50	5.45	8.96	14.38	23.42	4.95	5.01	6.66	8.68	3.88	5.18
T=100	4.80	5.45	6.21	8.53	4.55	4.30	3.94	4.62	4.01	4.91
T=200	4.84	5.00	4.85	4.81	4.85	4.31	3.67	3.69	4.46	5.21
$\theta_{1,2} = 0.5$										
T=50	6.07	8.89	13.95	22.83	4.50	4.63	6.08	8.19	3.51	5.06
T=100	5.26	5.57	6.18	8.20	4.43	3.82	3.53	4.39	3.78	4.89
T=200	5.06	5.16	4.55	4.91	4.71	3.87	3.29	3.41	4.30	5.24
$\theta_{1,2} = 0.8$										
T=50	6.74	9.28	14.13	23.78	4.28	4.56	5.61	7.88	3.46	4.93
T=100	5.89	6.13	6.58	8.82	3.95	3.49	3.65	4.33	3.39	4.83
T=200	5.59	5.29	4.91	5.14	4.07	3.22	3.11	3.28	3.76	5.27

example, for $(\theta_{1,2}, \theta_{2,1}, \theta_{2,2}) = (0.9, 0, -0.9)$, $T = 200$ and $k = 1$, the F_{nl} test has power almost 60% while the F_{VPP}, the AKSS and the ADF have power around 32%, 31% and 41% respectively.

The data for Tables 6-8 are simulated with $k = 2$, i.e. $e_1 = 0, e_2 = 3$. Generally, the patterns are similar to $k = 1$. For small samples ($T = 50$), the rejection probabilities increase as k increases for the tests based on F_{nl} and F_{VPP}. For large samples ($T = 200$), for $(\theta_{2,1}, \theta_{2,2}) = (0, -0.9)$ in Table 6 and $(\theta_{2,1}, \theta_{2,2}) = (0.4, -0.9)$ in Table 8, the power for $k = 2$ are the highest compared to other k when we use the F_{nl} tests while the F_{VPP} tests still have the highest power with $k = 1$. Apparently, with large samples, the F_{nl} tests are more able to recognise the true number of equilibria (for this case, $k = 2$) compared to the F_{VPP} tests. Over the three tables, the F_{nl} test shows more power to detect the alternative than the competitors tests when $\theta_{1,2}$ is close to 1. For example, for $(\theta_{1,2}, \theta_{2,1}, \theta_{2,2}) = (0.9, 0.4, -0.5)$, $T = 200$ and $k = 2$, the F_{nl} test has power approximately 72% while the test based on F_{VPP}, AKSS and ADF are only around 5%, 4% and 19% respectively.

Table 3: The power of alternative tests (in percentage), $k=1$, $e_1=0$ and $\theta=0.01$.

	F_{nl}				F_{VPP}				AKSS	ADF
$(\theta_{1,2},\theta_{2,1},\theta_{2,2})$	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4		
(0, 0, -0.9)										
T=50	30.96	18.35	21.68	31.67	39.80	12.26	11.81	14.50	35.88	26.30
T=100	74.01	32.57	24.40	25.73	85.73	32.69	20.41	19.08	84.22	78.31
T=200	99.56	81.01	58.91	49.18	99.90	90.81	69.91	57.36	99.89	99.99
(0.2, 0, -0.9)										
T=50	21.03	14.18	18.64	28.51	24.27	7.99	8.98	11.95	21.38	16.45
T=100	52.76	21.12	17.08	19.04	66.32	17.16	12.10	11.83	63.68	52.86
T=200	96.33	56.92	37.04	30.67	98.93	64.55	39.77	31.31	98.74	99.13
(0.5, 0, -0.9)										
T=50	14.45	11.73	15.96	25.72	13.14	5.55	7.20	9.66	11.37	11.11
T=100	32.14	13.40	11.84	13.94	36.84	8.44	7.04	7.56	34.51	27.17
T=200	81.29	32.36	20.38	17.71	88.00	28.26	16.97	14.67	86.93	86.64
(0.7, 0, -0.9)										
T=50	12.96	11.39	16.68	25.24	9.64	4.77	6.32	8.64	8.38	9.43
T=100	25.01	11.68	10.58	12.94	24.23	6.04	5.67	5.90	22.47	18.82
T=200	66.93	24.00	15.51	13.69	67.23	15.58	10.80	9.46	65.80	67.32
(0.9, 0, -0.9)										
T=50	14.70	12.05	17.22	26.91	6.69	3.64	5.20	7.26	5.77	7.65
T=100	24.55	12.70	12.03	13.64	12.22	3.77	3.89	4.56	11.46	12.84
T=200	59.38	22.90	16.41	14.72	32.35	6.81	5.06	5.06	31.27	41.69

Table 4: The power of alternative tests (in percentage), $k=1$, $e_1=0$ and $\theta=0.01$.

	F_{nl}				F_{VPP}				AKSS	ADF
$(\theta_{1,2},\theta_{2,1},\theta_{2,2})$	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4		
(0, 0.4, -0.5)										
T=50	9.79	9.79	14.13	23.04	8.93	4.92	5.98	8.15	7.53	8.52
T=100	17.95	9.69	9.13	10.86	16.41	5.21	4.45	4.90	15.04	12.89
T=200	47.03	18.61	12.07	10.51	48.75	9.57	5.60	5.04	46.92	38.51
(0.2, 0.4, -0.5)										
T=50	8.21	9.22	13.88	22.66	6.64	4.06	5.55	7.72	5.77	7.44
T=100	13.13	7.64	7.82	9.06	10.08	3.55	3.59	4.04	9.11	9.64
T=200	31.09	12.87	8.90	8.00	26.91	5.31	3.79	3.71	25.45	20.66
(0.5, 0.4, -0.5)										
T=50	7.81	8.58	13.94	22.41	4.84	3.41	4.89	7.37	4.12	6.18
T=100	10.48	6.64	6.81	8.66	6.07	2.61	2.90	3.74	5.48	7.10
T=200	20.66	9.54	7.08	6.53	11.12	2.77	2.64	3.08	10.49	10.93
(0.7, 0.4, -0.5)										
T=50	8.09	8.88	13.69	22.54	4.09	3.42	5.21	7.35	3.47	5.44
T=100	10.71	6.55	7.35	9.22	4.19	2.08	2.80	3.63	3.87	6.19
T=200	19.95	9.35	7.03	6.63	6.01	1.88	2.35	2.44	5.63	7.55
(0.9, 0.4, -0.5)										
T=50	12.12	10.72	16.33	25.72	2.74	3.34	5.40	7.88	2.13	4.53
T=100	16.01	9.55	9.87	12.06	2.02	2.04	3.22	4.10	1.87	4.14
T=200	29.28	14.26	10.94	10.10	1.80	1.26	2.20	2.68	1.76	4.41

Table 5: The power of alternative tests (in percentage), $k=1$, $e_1=0$ and $\theta=0.01$.

$(\theta_{1,2},\theta_{2,1},\theta_{2,2})$	F_{nl} k=1	k=2	k=3	k=4	F_{VPP} k=1	k=2	k=3	k=4	AKSS	ADF
(0, 0.4, -0.9)										
T=50	21.28	14.29	18.27	28.21	23.81	7.96	8.60	10.99	20.66	16.38
T=100	52.82	21.66	17.49	19.17	65.04	16.71	10.61	10.96	62.60	52.67
T=200	95.93	57.26	36.76	31.32	98.58	62.72	35.05	27.43	98.43	98.94
(0.2, 0.4, -0.9)										
T=50	15.01	11.94	16.54	25.52	14.39	5.66	7.20	9.46	12.36	11.56
T=100	34.80	14.18	12.47	14.17	41.08	8.60	6.69	7.29	38.29	29.88
T=200	83.82	36.17	22.17	18.97	91.07	31.30	16.38	14.10	90.16	89.81
(0.5, 0.4, -0.9)										
T=50	11.33	10.47	14.67	23.73	8.06	4.33	5.97	8.97	6.92	8.47
T=100	21.13	10.12	9.90	11.37	18.98	4.64	4.47	5.03	17.43	15.76
T=200	57.20	20.26	13.00	12.05	61.43	10.19	7.25	6.74	59.49	55.33
(0.7, 0.4, -0.9)										
T=50	11.34	9.81	14.60	23.36	5.84	3.57	5.34	7.61	5.10	7.28
T=100	17.77	9.36	9.50	11.21	10.66	3.15	3.97	4.52	9.69	11.54
T=200	44.37	16.13	11.23	10.33	34.81	5.06	4.65	4.93	33.08	34.11
(0.9, 0.4, -0.9)										
T=50	14.88	12.02	17.56	26.86	3.39	3.30	5.53	8.08	3.04	5.92
T=100	22.92	11.96	12.14	13.85	3.84	2.53	3.14	4.13	3.44	6.84
T=200	47.91	20.84	15.55	13.58	8.38	2.47	2.75	3.40	7.72	16.31

Table 6: The power of alternative tests (in percentage), $k=2$, $e_1=0$, $e_2=3$ and $\theta=0.01$.

$(\theta_{1,2},\theta_{2,1},\theta_{2,2})$	F_{nl} k=1	k=2	k=3	k=4	F_{VPP} k=1	k=2	k=3	k=4	AKSS	ADF
(0, 0, -0.9)										
T=50	57.89	64.69	62.93	68.05	67.81	56.91	53.21	52.78	64.54	51.92
T=100	91.47	93.38	89.57	86.51	96.50	93.08	90.63	87.12	96.14	91.53
T=200	99.94	99.98	99.94	99.83	99.99	99.99	99.97	99.96	99.98	99.99
(0.2, 0, -0.9)										
T=50	46.72	50.93	50.41	57.34	52.36	41.09	37.70	38.74	49.23	38.67
T=100	77.40	82.05	75.50	72.13	87.90	77.64	73.57	68.54	86.80	76.10
T=200	99.34	99.73	99.07	97.83	99.88	99.63	99.40	98.76	99.88	99.76
(0.5, 0, -0.9)										
T=50	37.60	38.02	39.42	47.59	36.47	25.58	24.26	25.99	33.40	30.95
T=100	60.37	65.17	57.72	55.07	70.28	51.32	47.58	43.77	68.27	54.33
T=200	94.07	96.44	92.40	88.02	97.50	90.87	91.12	87.00	97.29	95.92
(0.7, 0, -0.9)										
T=50	35.83	34.58	37.06	45.00	26.85	18.53	18.24	20.38	24.39	26.49
T=100	55.52	58.64	51.31	49.85	56.23	35.19	32.36	30.64	54.53	46.71
T=200	89.44	92.28	87.50	81.52	92.59	71.87	73.71	68.13	92.20	89.75
(0.9, 0, -0.9)										
T=50	44.28	41.88	43.43	51.12	16.35	11.03	12.75	15.19	14.87	20.08
T=100	66.74	66.34	62.16	59.51	33.92	17.50	17.57	18.22	32.42	46.99
T=200	93.37	93.45	91.88	88.59	69.14	38.00	36.98	36.77	68.17	88.04

Table 7: The power of alternative tests (in percentage), $k = 2$, $e_1 = 0$, $e_2 = 3$ and $\theta = 0.01$.

	F_{nl}				F_{VPP}				AKSS	ADF
$(\theta_{1,2}, \theta_{2,1}, \theta_{2,2})$	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4		
(0, 0.4, -0.5)										
T=50	27.71	20.64	24.47	34.59	19.57	6.67	7.53	9.63	17.11	18.65
T=100	57.72	34.56	31.02	32.58	48.45	11.39	7.62	8.01	45.72	48.51
T=200	93.05	70.70	63.23	59.87	89.27	39.46	20.66	16.74	88.50	95.16
(0.2, 0.4, -0.5)										
T=50	22.01	17.14	21.71	31.02	11.70	4.77	6.19	8.49	10.17	12.28
T=100	42.61	26.06	24.19	25.51	27.33	5.76	5.32	6.10	25.20	29.04
T=200	82.71	54.69	48.92	46.03	72.41	16.69	9.62	9.08	71.03	74.78
(0.5, 0.4, -0.5)										
T=50	19.68	16.32	20.39	29.36	6.08	3.52	5.67	8.30	4.97	8.35
T=100	34.31	23.05	20.99	22.56	11.46	3.48	4.47	5.19	10.44	15.95
T=200	66.34	47.18	42.46	38.45	35.18	6.13	5.25	5.25	33.48	43.33
(0.7, 0.4, -0.5)										
T=50	21.52	17.42	21.83	31.40	3.98	3.27	5.78	8.43	3.28	6.06
T=100	37.15	26.86	24.26	25.13	6.02	2.72	3.98	5.29	5.39	10.44
T=200	65.76	52.44	47.30	42.56	16.43	3.98	4.40	5.14	15.55	30.72
(0.9, 0.4, -0.5)										
T=50	30.09	25.74	29.00	38.24	2.08	3.79	6.92	10.05	1.74	3.99
T=100	50.09	41.48	37.26	37.33	2.24	3.51	5.63	7.13	1.99	5.69
T=200	78.64	72.02	69.01	64.69	4.50	5.43	6.84	7.64	4.17	19.13

Table 8: The power of alternative tests (in percentage), $k = 2$, $e_1 = 0$, $e_2 = 3$ and $\theta = 0.01$ (continue).

	F_{nl}				F_{VPP}				AKSS	ADF
$(\theta_{1,2}, \theta_{2,1}, \theta_{2,2})$	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4		
(0, 0.4, -0.9)										
T=50	53.08	52.95	53.80	60.75	59.90	33.01	30.58	34.33	56.51	46.60
T=100	88.13	84.25	79.54	76.75	94.64	74.33	65.30	62.75	93.99	85.80
T=200	99.89	99.83	99.45	98.67	99.98	99.77	99.24	98.52	99.98	99.96
(0.2, 0.4, -0.9)										
T=50	39.96	41.15	42.63	50.49	39.45	19.28	19.52	23.50	36.15	34.17
T=100	69.42	69.06	63.01	60.72	80.34	44.90	39.49	38.32	78.55	65.04
T=200	98.65	97.97	95.79	92.89	99.59	95.63	90.28	86.56	99.55	98.68
(0.5, 0.4, -0.9)										
T=50	32.68	32.54	33.96	42.94	20.14	10.85	12.42	16.53	17.61	24.45
T=100	48.60	53.30	48.07	47.17	49.13	19.32	18.37	19.85	46.70	43.92
T=200	87.57	89.57	83.60	78.50	92.90	53.87	52.25	48.22	92.28	84.59
(0.7, 0.4, -0.9)										
T=50	33.65	32.49	35.20	43.73	10.84	7.80	10.09	13.51	9.54	17.65
T=100	46.56	52.95	47.85	46.85	28.49	11.88	13.11	15.43	26.33	39.06
T=200	77.65	86.19	81.50	75.98	73.33	26.06	30.67	30.03	71.93	70.84
(0.9, 0.4, -0.9)										
T=50	46.28	45.29	47.03	53.87	4.69	7.76	10.88	14.82	4.05	12.01
T=100	62.05	68.43	65.82	63.41	8.65	4.92	12.64	14.83	7.95	37.03
T=200	83.79	93.43	92.66	89.81	24.18	19.51	23.01	25.51	22.99	60.14

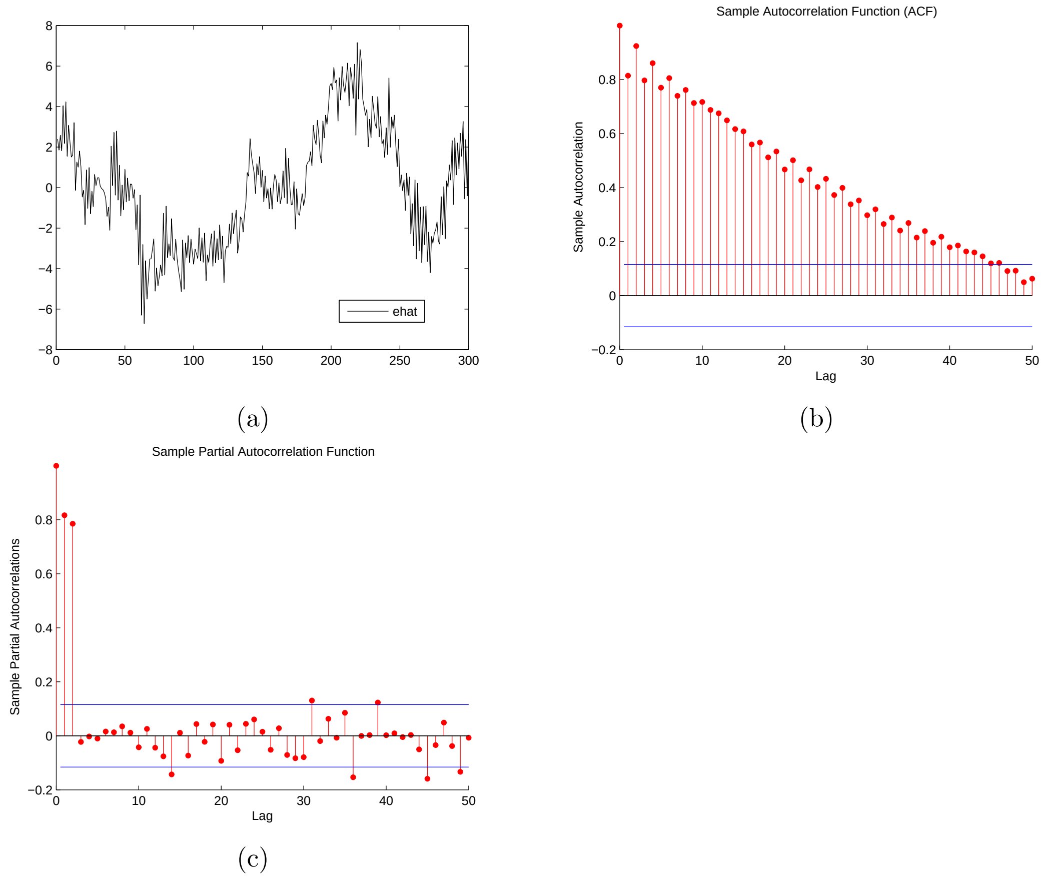


Figure 2: (a) Plot of ϵ_t; (b) ACF plot of ϵ_t; (c)PACF plot of ϵ_t.

4.3 Simulation Example

A series $\{\epsilon_t\}$ following a 1-ESTAR(2) model is generated by a simulation. The first 50 observations were discarded. Figure 2(a) shows the plot of $\{\epsilon_t\}$ while its autocorrelation function plot and partial autocorrelation function plot are shown in Figure 2(b) and (c).

$$\begin{aligned} \epsilon_t &= 0.1\epsilon_{t-1} + 0.9\epsilon_{t-2} + (0.4\epsilon_{t-1} - 0.5\epsilon_{t-2})\left[1 - \exp(-0.01\epsilon_{t-1}^2)\right] + \eta_t, \\ \eta_t &\sim N(0,1), \qquad \epsilon_{-1} = \epsilon_0 = 0, \end{aligned}$$

$t = 1, 2, \cdots, T$.

From Figure 2(a), $\{\epsilon_t\}$ seems stationary with equilibrium zero, even though for some periods, it takes quite a long time to return to zero. From the autocorrelation plot in Figure 2(b), the spikes are slowly falling to zero. From the partial autocorrelation plot in Figure 2(c), only the first two spikes are significant with the values of the spikes are around 0.8 indicating $p = 2$ is the appropriate lag for the models. However, to make a conclusion that the residuals series is a stationary series, more formal analysis is needed.

Table 9 reports summary statistics and the ADF unit root test for ϵ_t. It shows that the mean of $\hat{\epsilon}_t$ is virtually zero with variance around 7.85. The ADF unit root test statistics suggest that the series has a unit root (not stationary) on level but they are stationary on the first difference series. Using higher lags did not change the conclusion.

Table 10 reports linearity tests results for $\hat{\epsilon}_t$. The first linearity test employed is a RESET test (see Ramsey, 1969). The null hypothesis of linearity of the residuals from an AR(2) for $\hat{\epsilon}_t$is tested against the alternative hypothesis of general model misspecification involving a higher-order polynomial to represent a different functional form. Under the null hypothesis, the statistic is distributed as $\mathcal{X}^2(q)$ with q is equal to the number of higher-order terms in alternative model. Table 10 reports the result from applying the RESET test where the alternative model with a quadratic and a cubic terms are included. The null hypothesis cannot be rejected, suggesting that a linear AR(2) process for $\hat{\epsilon}_t$ is not misspecified.

Table 9: Summary Statistics and ADF unit root test

			Summary Statistics ϵ_t				
Minimum	-6.7124	Maximum	7.1589	Mean	-2.6666e-011	Variance	7.8493
			ADF unit root test				
ϵ_t	Lags		$\Delta\hat{\epsilon}_t$	Lags			
-1.8733	1		-16.114**	1			

Note: For the ADF test, ** superscript indicates significance at 1% level, based on critical values in Fuller (1976).

Table 10: Linearity tests on residuals $\hat{\epsilon}_t$

RESET Test	F(2,294) = 1.5296 [0.2183]				Lags used = 2		
Linearity tests based on Terasvirta (1994)							
d	LM^G		LM^3		LM^E		
1	2.1067 [0.0525]*		0.0431 [0.9578]		3.1592 [0.0145]**		
2	1.4760 [0.1862]		0.0091 [0.9909]		2.2246 [0.0664]*		
ESTAR unit root tests comparison							
d		F_{nl}			F_{VPP}		AKSS
	k=1	k=2	k=3	k=1	k=2	k=3	
1	14.45***	17.12**	20.61*	1.45	6.41	6.74	-1.19

Note: * and ** superscripts indicate significance at 5% and 1% level, respectively. The numbers in [] are the p-values.

The second linearity test is based on Terasvirta (1994). The test can also be used to discriminate between ESTAR or LSTAR models since the third-order terms disappear in the Taylor series expansion of the ESTAR transition function. We use this test to analyse whether the series $\hat{\epsilon}_t$ is a linear AR(2) model or a nonlinear ESTAR(2) or LSTAR(2) model. If $\hat{\epsilon}_t$ follows a LSTAR(2) model, the artificial regression will be as follow:

$$\hat{\epsilon}_t = \phi_{0,0} + \sum_{j=1}^{2} \left(\phi_{0,j}\ \hat{\epsilon}_{t-j} + \phi_{1,j}\ \hat{\epsilon}_{t-j}\hat{\epsilon}_{t-d} + \phi_{2,j}\ \hat{\epsilon}_{t-j}\hat{\epsilon}^2_{t-d} + \phi_{3,j}\ \hat{\epsilon}_{t-j}\hat{\epsilon}^3_{t-d}\right) + error \tag{39}$$

Keeping the delay parameter $d \leq 2$ fixed, testing the null hypothesis

$$H_0 : \phi_{1,j} = \phi_{2,j} = \phi_{3,j} = 0,$$

$\forall j \in \{1, 2\}$ against its complement is a general test (LM^G) of the hypothesis of linearity against smooth transition nonlinearity. Given that the ESTAR model implies no cubic terms in the artificial regression (i.e.: $\phi_{3,j} = 0$) if the true model is an ESTAR model, but $\phi_{3,j} \neq 0$ if the true model is an LSTAR model, testing the null hypothesis that

$$H_0 : \phi_{3,1} = \phi_{3,2} = 0,$$

against its complement provides a test (LM^3) of ESTAR nonlinearity against LSTAR-type nonlinearity. Moreover, if this restriction cannot be rejected at the chosen significance level, then a more powerful test (LM^E) for linearity against ESTAR-type nonlinearity is obtained by testing the null hypothesis

$$H_0 : \phi_{1,j} = \phi_{2,j} = 0 |\ \phi_{3j} = 0,$$

$\forall j \in \{1, 2\}$ against its complement. From this test, the statistics LM^G, LM^3 and LM^E with $d = 1$ are higher than the test statistics with $d = 2$, indicating that $d = 1$ is more preferred. Using $d = 1$, the LM^G test statistic is significant at 10% significant level, the LM^3 test statistic is not significant and the LM^E test statistic is significant at 5% significant level, indicating that $\hat{\epsilon}_t$ follows a nonlinear ESTAR(2) model [8].

[8] Note that this test uses ESTAR or LSTAR models with only one equilibrium. Therefore there is no k in the notation.

As in Table 9 the ADF unit root test does not confirm that ϵ_t is a stationary series, we use our ESTAR unit root test F_{nl} explained in Section 4. For comparison, we include the ESTAR unit root test of Venetis *et al.* (2009) denoted F_{VPP} and the augmented KSS test (denoted as AKSS, Kapetanios *et al.* (2003)). For the F_{nl} and F_{VPP} tests, we test for $k = 1, 2, 3$ and $d = 1$ with all at a 5% significant level. Following assumptions in Section 4 for k-ESTAR(2) unit root tests. The unrestricted regression for F_{nl} will be:

$$\Delta\hat{\epsilon}_t = \theta^*_{1,1}\ \Delta\hat{\epsilon}_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\ \hat{\epsilon}^{s+3}_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{2,s}\ \hat{\epsilon}^{s+2}_{t-1}\Delta\hat{\epsilon}_{t-1} + \eta^*_t \tag{40}$$

where $\theta^*_{1,1} = -\theta_{1,2}$, $\gamma_{1,s} = \theta\delta_s(\theta_{2,1}+\theta_{2,2})$, $\gamma_{2,s} = \theta\delta_s\theta^*_{2,1}$, $\theta^*_{2,1} = -\theta_{2,2}$ and $\eta^* = \eta_t + R\left[\sum_{j=1}^{2}\theta_{2,j}\ y_{t-j}\right]$ and R is the remainder.

The unrestricted regression for F_{VPP} will be:

$$\Delta\hat{\epsilon}_t = \theta^*_{1,1}\ \Delta\hat{\epsilon}_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\ \hat{\epsilon}^{s+3}_{t-1} + \eta^{**}_t \tag{41}$$

where $\eta^{**}_t = \sum_{s=0}^{2(k-1)} \gamma_{2,s}\ \hat{\epsilon}^{s+2}_{t-1}\Delta\hat{\epsilon}_{t-1} + \eta^*_t$.

Under the null hypothesis, the F_{nl} and F_{VPP} tests will:

$$\Delta\hat{\epsilon}_t = \theta^*_{1,1}\ \Delta\hat{\epsilon}_{t-1} + \eta_t. \tag{42}$$

This equation will be the restricted regression for the F_{nl} and F_{VPP} tests. The calculated F tests from the F_{nl} and F_{VPP} are compared with the critical values in Table 1.

For the AKSS test, we only consider for Case 1 because the mean of $\hat{\epsilon}_t$ is zero and there is no significant time trend in the series. For this test, the lagged first difference ($\Delta\hat{\epsilon}_{t-1}$) is included to overcome the autocorrelation, so that the regression model for the AKSS test is

$$\Delta\hat{\epsilon}_t = \delta_1\hat{\epsilon}^3_t + \delta_2\Delta\hat{\epsilon}_{t-1} + \eta_t. \tag{43}$$

The hypothesis for the AKSS test is

$$H_0 : \delta_1 = 0 \quad vs \quad H_1 : \delta_1 < 0. \tag{44}$$

Then, the calculated t-test for δ_1 in (43) is compared with the critical values of the AKSS test. The critical value for the t-test of AKSS test is -2.22 obtained from Table 1 in Kapetanios *et al.* (2003). The null hypothesis for the AKSS test concludes that $\hat{\epsilon}_t$ has a unit root without a drift. On the other hand, the alternative hypothesis for the AKSS test concludes that $\hat{\epsilon}_t$ is a globally stationary 1-ESTAR(2) model.

From the ESTAR tests results in Table 10, the F_{VPP} and AKSS tests cannot confirm that $\hat{\epsilon}_t$ is a stationary series but our test, the F_{nl} tests can identify that it is a nonlinear stationary ESTAR(2) model. As the most significant level is at $k = 1$, suggest that $\hat{\epsilon}_t$ follows a 1-ESTAR(2) model with $d = 1$, i.e.:

$$\hat{\epsilon}_t = \theta_{1,0} + \theta_{1,1}\ \hat{\epsilon}_{t-1} + \theta_{1,2}\ \hat{\epsilon}_{t-2} + (\theta_{2,0} + \theta_{2,1}\ \hat{\epsilon}_{t-1} + \theta_{2,2}\ \hat{\epsilon}_{t-2})\left(1 - \exp\left(-\theta^2(\hat{\epsilon}_{t-1} - e_1)^2\right)\right) + \eta_t, \tag{45}$$

where η_t is the error term and e_1 is the equilibrium point.

5 Unit Root Test Analysis for a k-ESTAR(p) model

Unlike the k-ESTAR(2) model, the F_{nl} test for k-ESTAR(p) model in (18) involves nuisance parameters $\mathbf{\Pi}$. To deal with this circumstance, we propose two methods. The first is a bootstrap method as an approximation to the asymptotic distribution of F_{nl}, and the second is approximation of critical values obtained by assuming $\mathbf{\Pi} = \mathbf{I}_{(p-1)\times(p-1)}$.

5.1 Bootstrap Method

A bootstrap approximation can be used to calculate critical values and p-values. For a review on bootstrapping time series, see Li and Maddala (1996) and for bootstrap applications as approximations of the asymptotic distributions of unit root test, see Caner and Hansen (2001) and Eklund (2003). Caner and Hansen (2001) analysed a unit root test for a threshold autoregressive (TAR) model involving a nuisance parameter function and suggested a bootstrap method to approximate the null distribution. Eklund (2003) analysed a unit root test for a 2-LSTAR(2) model. To overcome the problem of large inverse matrices, he followed the bootstrap method in Caner and Hansen (2001). Using the bootstrap method, Caner and Hansen (2001) and Eklund (2003) found fairly good results both in size and power tests. Having similar to STAR models and using a F test statistic as in Eklund (2003), in this section we also follow the bootstrap method in Eklund (2003).

Bootstrap method for k-ESTAR(p) models:

(B1) Calculate the F_{nl} statistic from the sample data based on (13) as an unrestricted model and (14) as a restricted model (see the calculation of the F_{nl} test statistic in Section 4 in the case of k-ESTAR(2) models).

(B2) Under the null hypothesis, y_t has a unit root as in (14), i.e.:

$$\Delta y_t = \sum_{j=1}^{p-1} \theta^*_{1,j}\ \Delta y_{t-j} + \epsilon_t, \qquad t = 1, \cdots, (T-p). \tag{46}$$

Let $\hat{\boldsymbol{\theta}}^* = (\hat{\theta}^*_{1,1}, \ldots, \hat{\theta}^*_{1,(p-1)})'$ and $N(\hat{\mu}_\epsilon, \hat{\sigma}^2_\epsilon)$ be the estimates of $\boldsymbol{\theta}^* = (\theta^*_{1,1}, \ldots, \theta^*_{1,(p-1)})$ and $N(\mu_\epsilon, \sigma^2_\epsilon)$ which is the distribution of the errors ϵ_t in (46) imposing the null hypothesis.

(B3) Let ϵ^b_t be a random draw from $N(\hat{\mu}_\epsilon, \hat{\sigma}^2_\epsilon)$ and generate the bootstrap time series

$$y^b_t = y^b_{t-1} + \sum_{j=1}^{p-1} \hat{\theta}^*_{1,j}\ \Delta y^b_{t-j} + \epsilon^b_t, \qquad t = 1, \cdots, (T-p). \tag{47}$$

Initial values for the resampling can be set to sample values of the de-meaned series. The distribution of the series y^b_t is called the bootstrap series distribution of the data. The F_{nl} test statistic is calculated from the resampled series y^b_t as in item (B1).

(B4) Repeating this resampling operation J times yields the empirical distribution of F^b_{nl}, which is the bootstrap distribution of F_{nl}, completely determined by $\hat{\boldsymbol{\theta}}^*$ and $N(\hat{\mu}_\epsilon, \hat{\sigma}^2_\epsilon)$. For a large number of independent F^b_{nl} tests, estimated from J resampled series, the bootstrap p-value, defined by $p^b = P(F^b_{nl} > F_{nl})$ can be approximated by the frequency of simulated F^b_{nl} that exceeds the observed value of F_{nl}.

5.2 Approximation of Critical Values Assuming $\boldsymbol{\Pi} = \mathbf{I}_{(p-1)\times(p-1)}$

Given the difficulty in obtaining the asymptotic null distribution of the test statistic, Eklund (2003) also suggested obtaining critical values by assuming the parameter in the null hypothesis equal zero. Under the null, his model is $\Delta y_t = \delta_1 \Delta y_{t-1} + \epsilon_t$. Assuming $\delta_1 = 0$ means that under the null, Δy_t are uncorrelated. Using the same argument, $\boldsymbol{\Pi} = \mathbf{I}$ means that under the null, Δy_t are uncorrelated as $(\rho_1, \ldots, \rho_{(p-2)})' = \mathbf{0}$. If $(\rho_1, \ldots, \rho_{(p-2)})'$ are not far from $\mathbf{0}$, these critical values may be a good approximation of critical values for the asymptotic null distribution. As an example, assuming $\boldsymbol{\Pi} = \mathbf{I}_{2\times 2}$, the critical values based on the asymptotic null distribution in (18) for k-ESTAR(3) models are tabulated in Table 11. We only consider k-ESTAR(3) models but for $p > 3$, we can follow the same procedure.

Table 11: Asymptotic critical values of F_{nl} test statistics for k-ESTAR(3) models with $\boldsymbol{\Pi} = \mathbf{I}_{2\times 2}$.

	Significance Level		
	0.1	0.05	0.01
k=1	7.124863	8.758735	12.306371
k=2	17.82701	20.35429	25.65715
k=3	26.86799	29.96162	36.30965

Note: Simulations were based on samples size T=10,000 and 50,000 replications.

5.3 Monte Carlo Experiments

Monte Carlo experiments are conducted for k-ESTAR(3) models to compare the power of the F_{nl} test to detect non-linearity with tests based on F_{VPP}, AKSS and ADF. We only consider k-ESTAR(3) models as for $p > 3$, we can follow the same procedure. Consider (9) for a k-ESTAR(3) model with $\theta_{1,0} = 0$, $\sum_{j=1}^{p} \theta_{1,j} = 1$ and $\theta_{2,0} = 0$ so that $e_1 = 0$ as follow,

$$y_t = \sum_{j=1}^{3} \theta_{1,j}\ y_{t-j} + \left(\sum_{j=1}^{3} \theta_{2,j}\ y_{t-j} \right) G^*(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t \tag{48}$$

where $\theta > 0$ and $\{\epsilon_t\}$ is a stationary and ergodic martingale difference sequence with variance σ_ϵ^2.

Using the same assumptions for (3), (48) can be arranged to become

$$\Delta y_t = \sum_{j=1}^{2} \theta_{1,j}^*\ \Delta y_{t-j} + \left(\sum_{j=1}^{3} \theta_{2,j} y_{t-1} + \sum_{j=1}^{2} \theta_{2,j}^*\ \Delta y_{t-j} \right) G^*(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \tag{49}$$

where $\theta_{i,1}^* = -(\theta_{i,2} + \theta_{i,3})$, $\theta_{i,2}^* = -\theta_{i,3}$, $i = 1, 2$.

Recalling the Taylor approximation for $G^*(\theta, \mathbf{e}, y_{t-d})$ around $\theta = 0$ in (12), (49) becomes,

$$\Delta y_t = \sum_{j=1}^{2} \theta_{1,j}^*\ \Delta y_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\ y_{t-1}\ y_{t-d}^{s+2} + \sum_{s=0}^{2(k-1)} \sum_{j=1}^{2} \gamma_{2,sj}\ y_{t-d}^{s+2}\ \Delta y_{t-j} + \epsilon_t^*, \tag{50}$$

where $\epsilon_t^* = \epsilon_t + R$. If $\theta = 0$, y_t in (48) is linear in term of y_{t-1}, y_{t-2} and y_{t-3} and $\epsilon_t^* = \epsilon_t$ since the remainder $R \equiv 0$.

Testing the null hypothesis of a unit root ($H_0 : \theta = 0$) against alternative of a globally stationary k-ESTAR(3) model is equivalent to testing,

$$H_0 : \gamma_{1,s} = \gamma_{2,sj} = 0, \qquad \text{for all } s \text{ and } j \ \text{ in (50) against its complement .}$$

5.3.1 The Size of Alternative Tests

In this simulation study, we want to know the probability that the proposed unit root test F_{nl} rejects H_0 with a pre-set significance level of 5% if the true underlying series is a linear unit root AR(3) model. If the size of alternative test is around 5% or less, it means the test is good in detecting the true underlying series. We also compare the results with the tests based on F_{VPP}, AKSS and ADF. To obtain the test sizes, we generate the null model of k-ESTAR(3) models, i.e.:

$$\Delta y_t = \theta_{1,1}^*\ \Delta y_{t-1} + \theta_{1,2}^*\ \Delta y_{t-2} + \epsilon_t \tag{51}$$

where $\theta_{1,1}^* = -(\theta_{1,2} + \theta_{1,3})$, $\theta_{1,2}^* = -\theta_{1,3}$ and ϵ_t is drawn from the standard normal distribution. In particular, we choose a broad range of parameter values for $\theta_{1,1}^*$ and $\theta_{1,2}^*$ so that Δy_t in (51) follows an AR(2) model. To fulfill the stationarity conditions of an AR(2) model, the parameters $\theta_{1,1}^*$ and $\theta_{1,2}^*$ should be: (i) $-1 < \theta_{1,2}^* < 1$, (ii) $\theta_{1,1}^* + \theta_{1,2}^* < 1$ and (iii) $\theta_{2,1}^* - \theta_{1,2}^* < 1$.

For computational purposes, the regression model in (51) becomes the restricted model for F_{nl} and F_{VPP}. The unrestricted model for F_{nl} is the regression models in (50) while the unrestricted model for F_{VPP} is:

$$\Delta y_t = \sum_{j=1}^{2} \theta_{1,j}^*\ \Delta y_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\ y_{t-1}\ y_{t-d}^{s+2} + \epsilon_t^*. \tag{52}$$

The bootstrap method is quite time consuming. Furthermore, when we apply the F_{nl} tests for $k > 1$, sometimes they fail due to singularity problems. This happens because under the null hypothesis, some nonlinear terms in (13) will be virtually zero. Therefore, for bootstrap method we only report results for $k = 1$. For the second method in Section 5.2 assuming $\mathbf{\Pi} = \mathbf{I}_{2\times 2}$, the F_{nl} statistic is compared to the critical values in Table 11.

For the F_{VPP} test, as it does not depend on p, the critical values for k-ESTAR(3) models are the same as the critical values for k-ESTAR(2) models in Table 1. For the AKSS test and the ADF test, we include Δy_{t-1} and Δy_{t-2} to overcome the autocorrelation in the error term, so that the regression model for the AKSS test is

$$\Delta y_t = \delta_1 y_{t-1} y_{t-d}^2 + \delta_2 \Delta y_{t-1} + \delta_3 \Delta y_{t-2} + error \tag{53}$$

and the regression model for the ADF test is

$$\Delta y_t = \delta_1 y_{t-1} + \delta_2 \Delta y_{t-1} + \delta_3 \Delta y_{t-2} + error. \tag{54}$$

The null and alternative hypothesis for the AKSS test and ADF test are

$$H_0 : \delta_1 = 0 \quad vs \quad H_1 : \delta_1 < 0. \tag{55}$$

Then, the calculated t-tests for δ_1 are compared with the critical values of the AKSS test and ADF test. The null hypothesis for the AKSS test and the ADF test conclude that y_t has a unit root without a drift. On the other hand, the alternative hypothesis for the AKSS test concludes that y_t is a globally stationary 1-ESTAR(1) model while the ADF test concludes that y_t is a stationary linear ARMA model. The sizes based on a 5% significant level are presented in Table 12.

In Table 12, F_{nl}^b denotes the F_{nl} test statistic with $k = 1$ using the bootstrap method described in Section 5.1. The data is generated from (51) with $T = 250$. The rejection of the null hypothesis percentages are based on critical values from 500 bootstrap series and then the simulations are repeated by 500 independent replications. For the other F_{nl} test statistics are based on the second method. For F_{nl}, F_{VPP}, AKSS and ADF tests, the data are generated from (51) with $T = 250$ and the rejection of the null hypothesis percentages are based on 10,000 independent replications.

Similar to the size of alternative test for k-ESTAR(2) model, for all cases, the rejection probabilities for the AKSS are less than or around 5%. It is followed by the ADF test with the rejection probabilities are close to or slightly higher than 5%. If we compare the results of F_{nl} tests using the bootstrap method in the second column and using the second method for $k = 1$ in the third column, generally the second method seems to produce better results as its values are close to or slightly higher than 5%. Furthermore, its highest value is 5.39 for parameter values (-0.9, -0.9) while the highest value from the F_{nl}^b is 6.4 for parameter values (0, -0.7). Comparing the F_{nl} tests and the F_{VPP} tests results, generally F_{VPP} tests are better than the F_{nl} tests. This is not surprising as the F_{VPP} tests involve less variables derived from the nonlinear term than the F_{nl} tests.

5.3.2 The Power of Alternative Tests

In this simulation study, we want to know the probability that the proposed unit root test F_{nl} rejects H_0 with a pre-set significance level if the true underlying series is a globally stationary nonlinear k-ESTAR(3) model. This probability can be defined as the power of alternative test. In this simulation study, we use a 5% significance level. We also compare the results with other tests, i.e. F_{VPP}, AKSS and ADF. To evaluate the power of tests against globally stationary k-ESTAR(3) model, samples from the model in (48) are generated with ϵ_t drawn from a standard normal distribution. The procedure to calculate the rejection probabilities of the null hypothesis is the same as the procedure in obtaining the size of the alternative tests. The simulation results are summarised in Tables 13 and 14.

The data for Table 13 are simulated with $k = 1$, i.e. $e_1 = 0$. From Table 13, comparing the F_{nl}^b test statistics and the F_{nl} test statistics for $k = 1$, we see that the two statistics have similar values. Thus, we conclude that the power of both methods are equal. The probabilities for the F_{nl} and F_{VPP} tests with $k = 1$ are the highest compared to other k. Apparently, with large samples, the tests are able to detect the true number of equilibrium (for this case, $k = 1$). The F_{nl} test shows more power to detect the alternative compared to the other methods when $\theta_{1,2}^*$ close to -1. For example, for $(\theta_{1,1}^*, \theta_{1,2}^*) = (-0.9, -0.9)$, $(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.5)$, and $k = 1$, the F_{nl} test can detect almost 30% while the F_{VPP}, the AKSS and the ADF are around 0.55%, 0.49% and 2.24% respectively.

The data for Table 14 are simulated with $k = 2$, i.e. $e_1 = 0$, $e_2 = 3$. Generally, the patterns are similar to $k = 1$. The results from the F_{nl}^b test and the F_{nl} test for $k = 1$, are still not much different. Generally, for $(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0, 0, -0.9)$ and $(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.9)$, the probabilities for $k = 2$ are the highest compared to other k when we use the F_{nl} tests while for $(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.5)$ the probabilities for $k = 1$ are the highest. The F_{VPP} tests still have the highest probabilities with $k = 1$ for all three combinations of $(\theta_{2,1}, \theta_{2,2}, \theta_{2,3})$. Apparently, the F_{nl} tests are more able to recognise the true number of equilibriums (for this case, $k = 2$) compared to the F_{VPP} tests. The F_{nl} test shows more power to detect the alternative compared to the other methods when $\theta_{1,2}^*$ close to -1. For example, for $(\theta_{1,1}^*, \theta_{1,2}^*) = (-0.9, -0.9)$, $(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.5)$, and $k = 2$,

Table 12: The size of alternative tests (in percentage)

$(\theta^*_{11}, \theta^*_{12})$	F^b_{nl}	F_{nl}			F_{VPP}			AKSS	ADF
		k=1	k=2	k=3	k=1	k=2	k=3		
(0,0)	5.1	4.89	5.11	5.21	5.14	5.00	4.34	4.66	4.97
(-0.2,-0.2)	5.4	4.94	5.12	5.12	5.0	3.96	3.61	4.53	4.98
(-0.5,-0.5)	5.2	5.17	4.67	4.77	4.25	3.03	3.07	3.75	5.00
(-0.7,-0.7)	5.6	5.19	4.85	4.97	3.76	2.59	2.50	3.33	4.98
(-0.9,-0.9)	4.9	5.39	5.73	5.83	3.23	2.12	2.23	2.80	5.03
(0.2,0.2)	4.8	4.85	5.29	5.74	5.35	6.01	5.42	4.88	4.93
(-0.3,0)	5.3	4.93	5.29	5.51	5.16	4.32	3.84	4.72	4.96
(-0.5,-0.2)	5.5	4.97	5.13	5.08	4.89	3.64	3.48	4.31	5.00
(-0.8,-0.5)	4.9	4.86	4.79	4.63	3.94	3.08	2.89	3.49	5.00
(-1,-0.7)	4.5	4.8	4.52	4.73	3.50	2.61	2.67	3.04	5.03
(-1.2,-0.9)	3.0	5.02	4.28	4.60	2.85	1.97	1.88	2.43	5.09
(-0.1,0.2)	5.6	4.84	5.29	5.58	5.30	5.31	4.54	4.77	4.99
(0.2,0.5)	4.8	5.11	5.82	7.50	5.54	6.47	6.85	4.95	4.72
(-0.7,0)	4.7	5.21	5.63	5.53	4.74	3.42	3.08	4.28	4.99
(-0.9,-0.2)	4.6	5.21	5.25	4.94	4.40	3.19	2.89	3.81	5.08
(-1.2,-0.5)	4.7	5.27	4.71	4.41	3.75	2.64	2.46	3.36	5.00
(-1.4,-0.7)	5.6	4.72	4.00	4.39	3.17	2.14	1.99	2.77	5.04
(-1.6,-0.9)	5.3	4.43	3.60	3.87	2.27	1.39	1.49	1.90	5.17
(-0.5,0.2)	4.7	5.05	6.01	6.10	5.08	4.14	3.67	4.69	5.02
(-0.2,0.5)	3.6	4.89	6.21	6.72	5.64	5.53	4.81	5.01	4.86
(0.3,0)	4.9	4.85	5.24	5.44	5.25	5.51	4.88	4.74	4.92
(0.1,-0.2)	5.3	4.90	5.12	5.15	5.12	4.39	4.20	4.58	4.92
(-0.2,-0.5)	6.2	5.18	4.83	4.92	4.59	3.33	3.03	4.10	4.98
(-0.4,-0.7)	5.7	5.13	4.74	4.91	4.03	2.94	2.61	3.54	5.00
(-0.6,-0.9)	5.1	4.95	4.77	5.05	3.09	2.16	2.08	2.69	5.06
(0.5,0.2)	4.7	5.05	5.60	6.83	5.39	6.36	6.78	4.82	4.84
(0.7,0)	4.6	5.0	5.72	6.71	5.48	6.00	6.48	4.83	4.89
(0.5,-0.2)	5.0	4.97	5.32	5.39	5.09	5.22	4.72	4.69	4.89
(0.2,-0.5)	5.9	5.01	4.97	5.01	4.61	3.84	3.48	4.17	4.91
(0,-0.7)	6.4	5.25	4.72	5.12	4.07	3.10	2.88	3.66	5.08
(-0.2,-0.9)	5.7	5.30	4.88	5.10	3.66	2.25	2.11	3.21	5.14

the F_{nl} test has power almost 85.7% while the F_{VPP}, the AKSS and the ADF are around 6.17%, 0.18% and 1.73% respectively.

6 Conclusion

This paper has extended the work of Kapetanios *et al.* (2003) and Venetis *et al.* (2009) by considering a unit root test for a k-ESTAR(p) model with a different approach. By using this approach, the singularity problem in Venetis *et al.* (2009) can be avoided. This approach will be able to enhance the power of test. However, for a k-ESTAR(p) model, $p > 1$, a problem with nuisance parameters emerges. To solve the problem, we suggest two methods, namely a bootstrap method and critical values approximation method assuming there is no autocorrelation in Δy_t. From Monte Carlo simulations for k-ESTAR(3) models, the bootstrap method is time consuming and if the underlying series is actually a linear unit root AR(3) model (under the null hypothesis), it may result in a singularity problem. Therefore, we favour to the critical values approximation method to the bootstrap method. For some cases, where the parameters are close to a unit root, simulation results show that our approach are better than the results from Venetis *et al.* (2009) , Kapetanios *et al.* (2003) and the Augmented Dickey-Fuller (ADF) tests Dickey and Fuller (1979, 1981) in term of identifying the nonlinearity.

Acknowledgement The authors would like to thank John Rayner for his comments.

Table 13: The power of alternative tests (in percentage), $\theta = 0.01$, $e_1 = 0$, $d = 1$ and $T = 250$.

$(\theta^*_{1,1}, \theta^*_{1,2})$	F^b_{nl}	F_{nl}			F_{VPP}			AKSS	ADF
		k=1	k=2	k=3	k=1	k=2	k=3		
$(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0, 0, -0.9)$									
(0,0)	99.8	99.62	84.53	66.23	100	96.28	83.20	100	100
(-0.2,-0.2)	90.2	89.30	43.57	29.19	97.08	53.93	30.67	96.67	98.47
(-0.5,-0.5)	53.2	49.89	18.51	13.84	64.17	13.82	8.97	62.49	60.04
(-0.7,-0.7)	38.4	36.03	14.60	11.54	39.27	77.75	5.52	37.75	32.75
(-0.9,-0.9)	40.0	41.61	19.21	14.72	20.42	4.19	3.18	19.31	20.27
$(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.9)$									
(0,0)	98.6	97.59	69.38	50.82	99.45	76.92	49.55	99.33	99.89
(-0.2,-0.2)	74.0	71.97	29.25	20.12	82.13	20.37	10.19	80.60	84.98
(-0.5,-0.5)	35.8	32.46	13.18	10.68	29.49	4.34	3.36	27.70	26.16
(-0.7,-0.7)	25.0	24.84	11.12	9.29	11.93	2.26	2.24	11.17	12.50
(-0.9,-0.9)	34.6	35.90	17.47	14.18	2.70	1.25	1.46	2.51	5.30
$(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.5)$									
(0,0)	70.0	67.00	35.98	26.09	60.26	10.16	5.21	58.22	63.75
(-0.2,-0.2)	31.6	30.98	14.09	10.75	15.20	2.94	2.37	13.97	13.56
(-0.5,-0.5)	15.0	16.72	8.27	6.80	3.90	1.41	1.63	3.65	5.71
(-0.7,-0.7)	14.6	15.59	8.06	7.26	2.04	1.02	1.34	1.91	4.19
(-0.9,-0.9)	26.8	27.77	14.46	12.67	0.55	0.87	1.34	0.49	2.24

Table 14: The power of alternative tests (in percentage), $\theta = 0.01$, $e_1 = 0$, $e_2 = 3$, $d = 1$ and $T = 250$.

$(\theta^*_{1,1}, \theta^*_{1,2})$	F^b_{nl}	F_{nl}			F_{VPP}			AKSS	ADF
		k=1	k=2	k=3	k=1	k=2	k=3		
$(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0, 0, -0.9)$									
(0,0)	100	99.98	99.98	99.96	99.98	99.98	99.98	99.98	100
(-0.2,-0.2)	98.8	98.26	99.11	98.47	99.79	98.34	98.36	99.78	99.59
(-0.5,-0.5)	77.4	79.22	90.24	85.68	90.44	74.66	75.24	89.81	85.72
(-0.7,-0.7)	69.2	71.82	85.99	81.49	76.69	54.04	53.60	75.54	73.22
(-0.9,-0.9)	100	83.99	93.62	93.17	53.82	27.44	27.38	52.55	75.31
$(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.9)$									
(0,0)	97.6	97.32	97.44	97.46	97.19	97.27	97.13	97.15	100
(-0.2,-0.2)	99.4	98.56	98.44	97.73	99.59	87.28	78.27	99.51	98.46
(-0.5,-0.5)	76.0	74.37	88.63	85.50	73.04	28.86	28.28	71.44	72.26
(-0.7,-0.7)	71.2	70.14	87.22	85.09	39.06	17.00	17.60	37.60	58.36
(-0.9,-0.9)	85.4	86.57	96.19	96.27	9.50	14.20	16.64	9.04	56.65
$(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.5)$									
(0,0)	100	99.76	98.81	98.22	97.38	82.69	66.58	97.09	99.87
(-0.2,-0.2)	92.8	92.66	79.78	80.03	64.49	15.72	9.13	62.73	81.79
(-0.5,-0.5)	70.2	72.17	62.62	62.42	11.43	4.39	4.26	10.70	32.67
(-0.7,-0.7)	68.6	72.05	67.61	65.01	2.62	3.52	4.39	2.33	12.61
(-0.9,-0.9)	85.6	85.27	85.68	84.15	0.26	6.17	8.21	0.18	1.73

A Appendices

A.1 Proof of Theorem 1

First, we prove that $y_t/\sqrt{T} \Rightarrow \lambda W(s)$ for $t \leq sT < t+1$ as $T \to \infty$. Since $\{\epsilon_t\}$ follows Assumption 1,

$$\frac{\sum_{i=1}^{t} \epsilon_i}{\sqrt{T}} \Rightarrow N(0, s\sigma_\epsilon^2) = \sigma_\epsilon W(s), \qquad t = 1, 2, \ldots, T, \tag{56}$$

where $W(s)$ is a standard Brownian motion with variance s, $s \in [0,1]$ (see Hong and Phillips, 2010).

Let $y_t = y_{t-1} + \eta_t$ where $\eta_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$ where $\{\eta_t\}$ and $\{\epsilon_t\}$ follow Assumption 1. Using the BN decomposition,

$$\eta_t = C(1)\epsilon_t + \widetilde{\epsilon}_{t-1} - \widetilde{\epsilon}_t, \tag{57}$$

where $\widetilde{\epsilon}_t = \widetilde{C}(L)\epsilon_t = \sum_{j=0}^{\infty} \widetilde{c}_j \epsilon_{t-j}$ and $\widetilde{c}_j = \sum_{k=j+1}^{\infty} c_k$.

From (57),

$$\frac{y_t}{\sqrt{T}} = \frac{\sum_{i=1}^{t} \eta_i}{\sqrt{T}} = C(1)\frac{\sum_{i=1}^{t} \epsilon_i}{\sqrt{T}} + \frac{\widetilde{\epsilon}_0}{\sqrt{T}} - \frac{\widetilde{\epsilon}_t}{\sqrt{T}}. \tag{58}$$

Using Markov's inequality [9],

$$P\left(\frac{\widetilde{\epsilon}_t^2}{T} > a\right) < \frac{E(\widetilde{\epsilon}_t^2)}{Ta} \to 0, \text{ as } T \to \infty, \tag{59}$$

for a positive real number a, because $E(\widetilde{\epsilon}_t^2) < \infty$. [10] Similar result happens for $\widetilde{\epsilon}_0$. Thus,

$$\begin{aligned} \frac{y_t}{\sqrt{T}} &= \frac{\sum_{i=1}^{t} \eta_i}{\sqrt{T}} = \frac{C(1)\sum_{i=1}^{t} \epsilon_i}{\sqrt{T}} + \frac{\widetilde{\epsilon}_0}{\sqrt{T}} - \frac{\widetilde{\epsilon}_t}{\sqrt{T}} \\ &\Rightarrow C(1)\sigma_\epsilon W(s) \qquad \text{by (56) and (59)} \\ &= \lambda W(s), \quad \text{as } T \to \infty. \end{aligned} \tag{60}$$

Given the result of (60), we start to prove Theorem 1.

(a) and (b): The proofs can be found in Venetis *et al.* (2009).

(c) Under H_0, $T^{-1}\sum_{t=p+1}^{T} \Delta y_{t-i}\Delta y_{t-j} = T^{-1}\sum_{t=p+1}^{T} \eta_{t-i}\eta_{t-j}$. Now, for given i, and j where $i, j = 1, \ldots, (p-1)$,

$$T^{-1}\sum_{t=p+1}^{T} \eta_{t-i}\eta_{t-j} \to E(\eta_{t-i}\eta_{t-j}) = \gamma_{|j-i|} \quad \text{as } T \to \infty. \tag{61}$$

If $i = j$,

$$T^{-1}\sum_{t=p+1}^{T} \eta_{t-i}^2 \to E(\eta_{t-i}^2) = \gamma_0 \quad \text{as } T \to \infty. \tag{62}$$

(d) Under H_0, $T^{-(q/2+1)}\sum_{t=p+1}^{T} y_{t-d}^q \Delta y_{t-i}\Delta y_{t-j} = T^{-(q/2+1)}\sum_{t=p+1}^{T} y_{t-d}^q \eta_{t-i}\eta_{t-j}$. Now, for given

[9] Markov's inequality: $P(|X| \geq a) \leq E(|X|)/a$ for given a random variable X and a positive real number a. However, we use the square of random variable instead of the absolute value.

[10] $E(\widetilde{\epsilon}_t^2) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \widetilde{c}_j^2 = \sigma_\epsilon^2 \sum_{j=0}^{\infty} |\widetilde{c}_j|^2 < \infty$ by Assumption 1 and Lemma 1.

i and j where $i, j = 1, \ldots, (p-1)$, and $i \geq j \geq d$,

$$
\begin{aligned}
& T^{-(q/2+1)} \sum_{t=p+1}^{T} y_{t-d}^{q} \eta_{t-i} \eta_{t-j} \\
&= T^{-(q/2+1)} \sum_{t=p+1}^{T} \left(y_{t-i-1} + \sum_{k=0}^{i-d} \eta_{t-d-k} \right)^{q} \eta_{t-i} \eta_{t-j} \\
&= T^{-(q/2+1)} \sum_{t=p+1}^{T} y_{t-i-1}^{q} \eta_{t-i} \eta_{t-j} \\
&\quad + T^{-(q/2+1)} \sum_{t=p+1}^{T} \left[\sum_{s=1}^{q} \binom{q}{s} y_{t-i-1}^{q-s} \left(\sum_{k=0}^{i-d} \eta_{t-d-k} \right)^{s} \eta_{t-i} \eta_{t-j} \right] \qquad (63) \\
&= \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q} \frac{(\eta_{t-i}\eta_{t-j} - E(\eta_{t-i}\eta_{t-j}))}{T} + \frac{E(\eta_{t-i}\eta_{t-j})}{T} \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q} \\
&\quad + T^{-(q/2+1)} \sum_{t=p+1}^{T} \left[\sum_{s=1}^{q} \binom{q}{s} y_{t-i-1}^{q-s} \left(\sum_{k=0}^{i-d} \eta_{t-d-k} \right)^{s} \eta_{t-i} \eta_{t-j} \right]. \qquad (64)
\end{aligned}
$$

Let $w_t = \left(\sum_{k=0}^{i-d} \eta_{t-d-k} \right)^{s} \eta_{t-i} \eta_{t-j}$. For fixed q and s,

$$
\begin{aligned}
& T^{-(q/2+1)} \sum_{t=p+1}^{T} y_{t-i-1}^{q-s} w_t \qquad (65) \\
&= T^{-s/2} \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q-s} \frac{w_t}{T} \\
&= T^{-s/2} \sum_{t=p+1}^{T} \left[\left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q-s} \frac{w_t - E(w_t)}{T} + \frac{E(w_t)}{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q-s} \right]. \qquad (66)
\end{aligned}
$$

Since $s \geq 1$,

$$
\frac{E(w_t)}{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q-s} \Rightarrow E(w_t) \lambda^{q-s} \int W^{q-s}
$$

and

$$
E(w_t) = E\left[\left(\sum_{k=0}^{i-d} \eta_{t-d-k} \right)^{s} \eta_{t-i} \eta_{t-j} \right] < \infty. \text{ [11]}
$$

Thus, for any constant $a > 0$, we have

$$
\begin{aligned}
P\left(\sum_{t=p+1}^{T} \frac{(w_t - E(w_t))^2}{T^2} > a \right) &\leq \frac{E\left[\sum_{t=p+1}^{T} (w_t - E(w_t))^2 \right]}{T^2 a}, \text{ Markov's inequality} \\
&= \frac{\sum_{t=p+1}^{T} E(w_t - E(w_t))^2}{T^2 a} \\
&= \frac{(T-p) Var(w_t)}{T^2 a} \\
&\leq \frac{Var(w_t)}{Ta} \\
&\to 0 \quad \text{as } T \to \infty \text{ and } Var(w_t) < \infty;
\end{aligned}
$$

[11] As w_t is a function of $\eta_{t-1}, \ldots, \eta_{t-(p-1)}$ which are stationary processes, then w_t is also a stationary process. As a stationary process, $E(w_t) < \infty$ and $Var(w_t) < \infty$.

and

$$\begin{aligned}
&\left|\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q-s}\frac{w_t-E(w_t)}{T}\right| \\
\leq\ & \sqrt{\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{2(q-s)}}\sqrt{\sum_{t=p+1}^{T}\frac{(w_t-E(w_t))^2}{T^2}},\ \text{Cauchy-Schwarz inequality} \\
\Rightarrow\ & \sqrt{\int W^{2(q-s)}}\sqrt{o_p(1)}=o_p(1),\quad \text{as } T\rightarrow\infty. \qquad (67)
\end{aligned}$$

Therefore,

$$T^{-(q/2+1)}\sum_{t=p+1}^{T}\left[\sum_{s=1}^{q}\binom{q}{s}y_{t-i-1}^{q-s}\left(\sum_{k=0}^{i-d}\eta_{t-d-k}\right)^{s}\eta_{t-i}\eta_{t-j}\right]=o_p(1). \qquad (68)$$

As $E(\eta_{t-i}\eta_{t-j})=\gamma_{|j-i|}<\infty$, the first term of (64) converges to $o_p(1)$. Thus,

$$T^{-(q/2+1)}\sum_{t=p+1}^{T}y_{t-d}^{q}\eta_{t-i}\eta_{t-j}\Rightarrow\gamma_{|j-i|}\lambda\int W^q,\qquad T\rightarrow\infty.$$

If $d>i\geq j$, recalling $y_{t-d}=y_{t-i-1}-\sum_{k=1}^{d-i-1}\eta_{t-i-k}$, the same result is obtained as follow:

$$\begin{aligned}
&T^{-(q/2+1)}\sum_{t=p+1}^{T}y_{t-d}^{q}\eta_{t-i}\eta_{t-j} \\
=\ & T^{-(q/2+1)}\sum_{t=p+1}^{T}\left(y_{t-i-1}-\sum_{k=1}^{d-i-1}\eta_{t-i-k}\right)^{q}\eta_{t-i}\eta_{t-j} \\
=\ & T^{-(q/2+1)}\sum_{t=p+1}^{T}y_{t-i-1}^{q}\eta_{t-i}\eta_{t-j} \\
& +T^{-(q/2+1)}\sum_{t=p+1}^{T}\left[\sum_{s=1}^{q}(-1)^s\binom{q}{s}y_{t-i-1}^{q-s}\left(\sum_{k=1}^{d-i-1}\eta_{t-i-k}\right)^{s}\eta_{t-i}\eta_{t-j}\right] \\
=\ & \sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q}\frac{\eta_{t-i}\eta_{t-j}-E(\eta_{t-i}\eta_{t-j})}{T}+\frac{E(\eta_{t-i}\eta_{t-j})}{T}\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q} \\
& +T^{-(q/2+1)}\sum_{t=p+1}^{T}\left[\sum_{s=1}^{q}(-1)^s\binom{q}{s}y_{t-i-1}^{q-s}\left(\sum_{k=1}^{d-i-1}\eta_{t-i-k}\right)^{s}\eta_{t-i}\eta_{t-j}\right] \\
\Rightarrow\ & \gamma_{|j-i|}\lambda^q\int W^q \qquad (69)
\end{aligned}$$

as the first and the last term of (69) are $o_p(1)$.

(e) Under H_0,

$$T^{-((q+1)/2+1)}\sum_{t=p+1}^{T}y_{t-1}y_{t-d}^{q}\Delta y_{t-i}=T^{-((q+1)/2+1)}\sum_{t=p+1}^{T}y_{t-1}y_{t-d}^{q}\eta_{t-i}.$$

Now, for given d and i, $i = 1, \ldots, (p-1)$, and if $d \leq i$,

$$\begin{aligned}
& T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{q} \eta_{t-i} \\
= \quad & T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} \left(y_{t-i-1} + \sum_{j=1}^{i} \eta_{t-j} \right) \left(y_{t-i-1} + \sum_{j=0}^{i-d} \eta_{t-d-j} \right)^{q} \eta_{t-i} \\
= \quad & T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} y_{t-i-1}^{q+1} \eta_{t-i} + T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} \left[y_{t-i-1}^{q} \eta_{t-i} \sum_{j=1}^{i} \eta_{t-j} \right] \\
& + T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} \left[\sum_{s=1}^{q} \binom{q}{s} y_{t-i-1}^{q-s+1} \left(\sum_{j=0}^{i-1} \eta_{t-j-1} \right)^{s} \eta_{t-i} \right] \\
& + T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} \left[\sum_{j=1}^{i} \eta_{t-j} \sum_{s=1}^{q} \binom{q}{s} y_{t-i-1}^{q-s} \left(\sum_{j=0}^{i-1} \eta_{t-j-1} \right)^{s} \eta_{t-i} \right]
\end{aligned} \tag{70}$$

Now, we need to show that all terms in (70) are $o_p(1)$.

(i) As $E(\eta_t \eta_s) \neq 0$ for $t \neq s$, the BN decomposition in (57) is used as follows (see Hong and Phillips, 2010),

$$\begin{aligned}
& \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q+1} \frac{\eta_{t-i}}{\sqrt{T}} \\
= \quad & \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q+1} \frac{\epsilon_{t-i} C(1)}{\sqrt{T}} - \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q+1} \left(\frac{\widetilde{\epsilon}_{t-i} - \widetilde{\epsilon}_{t-i-1}}{\sqrt{T}} \right).
\end{aligned}$$

By noting that

$$
\begin{aligned}
&\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q+1}\left(\frac{\widetilde{\epsilon}_{t-i}-\widetilde{\epsilon}_{t-i-1}}{\sqrt{T}}\right)\\
&= \left(\frac{y_{T-i}}{\sqrt{T}}\right)^{q+1}\frac{\widetilde{\epsilon}_{T-i}}{\sqrt{T}}-\left(\frac{y_{T-i}}{\sqrt{T}}\right)^{q+1}\frac{\widetilde{\epsilon}_{T-i}}{\sqrt{T}}+\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q+1}\left(\frac{\widetilde{\epsilon}_{t-i}-\widetilde{\epsilon}_{t-i-1}}{\sqrt{T}}\right)\\
&= \left(\frac{y_{T-i}}{\sqrt{T}}\right)^{q+1}\frac{\widetilde{\epsilon}_{T-i}}{\sqrt{T}}\\
&\quad -\sum_{t=p+1}^{T}\left[\left(\frac{y_{t-i}}{\sqrt{T}}\right)^{q+1}-\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q+1}\right]\frac{\widetilde{\epsilon}_{t-i}}{\sqrt{T}}-\left(\frac{y_{p-i}}{\sqrt{T}}\right)^{q+1}\frac{\widetilde{\epsilon}_{p-i}}{\sqrt{T}}\\
&= o_p(1)-\sum_{t=p+1}^{T}\left[\left(\frac{y_{t-i}}{\sqrt{T}}\right)^{q+1}-\left(\frac{y_{t-i}-\eta_{t-i}}{\sqrt{T}}\right)^{q+1}\right]\frac{\widetilde{\epsilon}_{t-i}}{\sqrt{T}}-o_p(1) \quad \text{by (59)}\\
&\approx -(q+1)\sum_{t=p+1}^{T}\left(\frac{y_{t-i}}{\sqrt{T}}\right)^{q}\frac{\eta_{t-i}\widetilde{\epsilon}_{t-i}}{T}\\
&= -(q+1)\sum_{t=p+1}^{T}\left(\frac{y_{t-i}}{\sqrt{T}}\right)^{q}\left(\frac{\eta_{t-i}\widetilde{\epsilon}_{t-i}-E(\eta_{t-i}\widetilde{\epsilon}_{t-i})}{T}\right)\\
&\quad -(q+1)\sum_{t=p+1}^{T}\left(\frac{y_{t-i}}{\sqrt{T}}\right)^{q}\frac{E(\eta_{t-i}\widetilde{\epsilon}_{t-i})}{T}\\
&= o_p(1)-(q+1)\sum_{t=p+1}^{T}\left(\frac{y_{t-i}}{\sqrt{T}}\right)^{q}\frac{E(\eta_{t-i}\widetilde{\epsilon}_{t-i})}{T} \quad \text{similar way with (67)}\\
&\Rightarrow -(q+1)\Lambda_{\eta\eta}\lambda^q\int W^q \qquad (71)
\end{aligned}
$$

where $E(\eta_{t-i}\widetilde{\epsilon}_{t-i})=\sum_{h=1}^{\infty}E(\eta_0\eta_h)=\Lambda_{\eta\eta}$. [12]

Therefore,

$$
\begin{aligned}
&\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q+1}\frac{\eta_{t-i}}{\sqrt{T}}\\
&\Rightarrow \int(\lambda W)^{(q+1)}\lambda dW-\left(-(q+1)\Lambda_{\eta\eta}\lambda^q\int W^q\right) \qquad \text{by (60) and (71)}\\
&= \lambda^{(q+2)}\int W^{(q+1)}dW+(q+1)\Lambda_{\eta\eta}\lambda^q\int W^q,
\end{aligned}
$$

and

$$
T^{-1/2}\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q+1}\frac{\eta_{t-i}}{\sqrt{T}}=o_p(1).
$$

[12]For η_t and ϵ_t satisfying Assumption 1, we have

$$
E(\eta_{t-i}\widetilde{\epsilon}_{t-i})=E\left(\sum_{j=0}^{\infty}c_j\epsilon_{t-i-j}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}c_k\epsilon_{t-i-j}\right)=\sum_{j=0}^{\infty}E(\epsilon_{t-i-j}^2)c_j\sum_{k=j+1}^{\infty}c_k=\sigma_\epsilon^2\sum_{j=0}^{\infty}c_j\sum_{k=j+1}^{\infty}c_k.
$$

$$
\sum_{h=1}^{\infty}E(\eta_0\eta_h)=\sum_{h=1}^{\infty}E\left(\sum_{j=0}^{\infty}c_j\epsilon_{-j}\sum_{j=0}^{\infty}c_j\epsilon_{h-j}\right)=\sum_{j=0}^{\infty}E(\epsilon_{-j}^2)c_j\sum_{h=1}^{\infty}c_{h+j}=\sigma_\epsilon^2\sum_{j=0}^{\infty}c_j\sum_{k=j+1}^{\infty}c_k.
$$

Therefore, $E(\eta_{t-i}\widetilde{\epsilon}_{t-i})=\sum_{h=1}^{\infty}E(\eta_0\eta_h)=\sigma_\epsilon^2\sum_{j=0}^{\infty}c_j\sum_{k=j+1}^{\infty}c_k\le\sigma_\epsilon^2\sum_{j=0}^{\infty}|c_j|\sum_{k=j+1}^{\infty}|c_k|\le\sigma_\epsilon^2(\sum_{j=0}^{\infty}|c_j|)^2<\infty$.

(ii) For given $i = 1, \ldots, (p-1)$,

$$\begin{aligned}
& T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} \left[y_{t-i-1}^{q} \eta_{t-i} \sum_{j=1}^{i} \eta_{t-j} \right] \\
= & \sum_{j=1}^{i} T^{-1/2} \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q} \frac{\eta_{t-i}\eta_{t-j}}{T} \\
= & \sum_{j=1}^{i} T^{-1/2} \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q} \left[\frac{\eta_{t-i}\eta_{t-j} - E(\eta_{t-i}\eta_{t-j})}{T} + \frac{E(\eta_{t-i}\eta_{t-j})}{T} \right] \\
\Rightarrow & \; o_p(1) \text{ as } T \to \infty.
\end{aligned} \tag{72}$$

Using similar method used in (67), the first term of (72) can be shown t be $o_p(1)$ and

$$\sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q} \frac{E(\eta_{t-i}\eta_{t-j})}{T} \Rightarrow \gamma_{|j-i|} \lambda^{q} \int W^{q}.$$

Therefore, the last term of (72) will be

$$T^{-1/2} \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q} \frac{E(\eta_{t-i}\eta_{t-j})}{T} \Rightarrow o_p(1).$$

Using similar procedure, the last two terms of (70) are also $o_p(1)$. Therefore, (e) is hold.

(f) For a given $d \geq 1$,

$$\begin{aligned}
& T^{-(q+2)/2} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{q} \epsilon_t \\
= & \; T^{-(q+2)/2} \sum_{t=p+1}^{T} y_{t-1} \left(y_{t-1} - \sum_{j=0}^{d-1} \eta_{t-d-j} \right)^{q} \epsilon_t \\
= & \; T^{-(q+2)/2} \sum_{t=d+1}^{T} y_{t-1}^{(q+1)} \epsilon_t \\
& + T^{-(q+2)/2} \sum_{t=p+1}^{T} \left[\sum_{s=1}^{q} (-1)^s \binom{q}{s} y_{t-1}^{q-s+1} \left(\sum_{j=0}^{d-1} \eta_{t-d-j} \right)^{s} \epsilon_t \right] \\
= & \sum_{t=p+1}^{T} \left(\frac{y_{t-1}}{\sqrt{T}} \right)^{(q+1)} \frac{\epsilon_t}{\sqrt{T}} + o_p(1) \qquad \text{similar way with (63)} \\
\Rightarrow & \int (\lambda W)^{(q+1)} \sigma_\epsilon dW \qquad \text{by (56) and (60)} \\
= & \; \sigma_\epsilon \lambda^{(q+1)} \int W^{(q+1)} dW.
\end{aligned}$$

(g) For any fixed i where $i = 1, \ldots, (p-1)$, under H_0, $\sum_{t=p+1}^{T} \Delta y_{t-i}\epsilon_t = \sum_{t=p+1}^{T} \eta_{t-i}\epsilon_t$.

$$E\left(\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \eta_{t-i}\epsilon_t \right) = \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} E(\eta_{t-i}) E(\epsilon_t) = 0$$

as η_{t-i} and ϵ_t are independent and

$$\begin{aligned}
&Var\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\eta_{t-i}\epsilon_t\right)\\
&= \frac{1}{T}\sum_{t=p+1}^{T}E(\eta_{t-i}^2)E(\epsilon_t^2) \qquad \text{as } \eta_{t-i} \text{ and } \epsilon_t \text{ are independent}\\
&= \frac{(T-p)}{T}\gamma_0\sigma_\epsilon^2 \qquad \text{as } \eta_{t-i}\epsilon_t \text{ are identic for each } t\\
&\to \gamma_0\sigma_\epsilon^2 \qquad \text{as } T\to\infty.
\end{aligned}$$

Since $E(\eta_{t-i}\epsilon_t|\epsilon_{t-1},\epsilon_{t-2},\ldots)=0$, $\{\eta_{t-i}\epsilon_t\}$ is MDS. Using Central Limit Theorem,

$$\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\eta_{t-i}\epsilon_t \Rightarrow N(0,\gamma_0\sigma_\epsilon^2)=\sqrt{\gamma_0}\sigma_\epsilon W_i(1). \tag{73}$$

Note that,

$$\begin{aligned}
&Cov(\eta_{t-i}\epsilon_t,\eta_{t-j}\epsilon_t) \text{ for } i,\, j=1,\ldots,(p-1),\, i\neq j, \text{ for all } t\\
&= E(\eta_{t-i}\eta_{t-j}\epsilon_t^2)\\
&= E(\eta_{t-i}\eta_{t-j})E(\epsilon_t^2) \qquad \text{as } (\eta_{t-i}\eta_{t-j}) \text{ and } (\epsilon_t^2) \text{ are independent}\\
&= \gamma_{|j-i|}\sigma_\epsilon^2.
\end{aligned}$$

Therefore, there is correlation between $W_i(1)$ and $W_j(1)$. Furthermore,

$$\begin{aligned}
&Cov(\eta_{t-i}\epsilon_t,\eta_{s-j}\epsilon_s) \qquad \text{for } i,\, j=1,\ldots,(p-1),\, i\neq j, \text{ for all } t\neq s\\
&= E(\eta_{t-i}\eta_{s-j}\epsilon_t\epsilon_s)\\
&= E(\eta_{t-i}\eta_{s-j})E(\epsilon_t)E(\epsilon_s) \qquad \text{as } (\eta_{t-i}\eta_{s-j}),\ (\epsilon_t) \text{ and } (\epsilon_s) \text{ are independent}\\
&= 0.
\end{aligned}$$

(h) Under H_0, $T^{-(q+1)/2}\sum_{t=p+1}^{T}y_{t-d}^q\Delta y_{t-i}\epsilon_t=\sum_{t=i+1}^{T}y_{t-d}^q\eta_{t-i}\epsilon_t$. Now, for given d, and i where $i=1,\ldots,(p-1)$, $d\geq 1$,

$$\begin{aligned}
&T^{-(q+1)/2}\sum_{t=p+1}^{T}y_{t-d}^q\eta_{t-i}\epsilon_t\\
&= T^{-(q+1)/2}\sum_{t=p+1}^{T}\left(y_{t-1}-\sum_{j=0}^{d-1}\eta_{t-d-j}\right)^q\eta_{t-i}\epsilon_t\\
&= T^{-(q+1)/2}\sum_{t=p+1}^{T}y_{t-1}^q\eta_{t-i}\epsilon_t\\
&\quad +T^{-(q+1)/2}\sum_{t=p+1}^{T}\left[\sum_{s=1}^{q}(-1)^s\begin{pmatrix} q\\ s\end{pmatrix}y_{t-1}^{q-s}\left(\sum_{j=0}^{d-1}\eta_{t-d-j}\right)^s\eta_{t-i}\epsilon_t\right]\\
&= \sum_{t=p+1}^{T}\left(\frac{y_{t-1}}{\sqrt{T}}\right)^q\left(\frac{\eta_{t-i}\epsilon_t}{\sqrt{T}}\right)+o_p(1) \qquad \text{similar way with (63)} \qquad (74)\\
&\Rightarrow \int(\lambda W)^q(\sqrt{\gamma_0}\sigma_\epsilon)dW_i \qquad \text{by (56) and (73)} \qquad (75)\\
&= \sqrt{\gamma_0}\sigma_\epsilon\lambda^q\int WdW_i
\end{aligned}$$

Note that as $\{\eta_{t-i}\epsilon_t\}$ is MDS and $\sum_{t=1}^{T}(\eta_{t-i}\epsilon_t/\sqrt{T})^2<\infty$, then based on Theorem 2.1 in

Hansen 1992, the first term of (74) will converge to (75).

A.2 Proof of Theorem 2

Consider the asymptotic distribution of the F-test statistics in (17) by testing the null hypothesis $H_0 : R\mathbf{b} = 0$ where $R = [\mathbf{0}\ \ \mathbf{I}]$ and $\mathbf{b} = (\mathbf{b}_1\ \ \mathbf{b}_2)'$. Under the null hypothesis of $H_0 : \theta = 0$, $\epsilon_t^* = \epsilon_t$, so that the sampling error of $\hat{\mathbf{b}}_2$ is given by

$$\hat{\mathbf{b}}_2 - \mathbf{b}_2 = (X_2'M_1X_2)^{-1}X_2'M_1\epsilon$$

where

$$X_2'M_1X_2 = X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2$$

and

$$X_2'M_1\epsilon = X_2'\epsilon - X_2'X_1(X_1'X_1)^{-1}X_1'\epsilon$$

Let

$$D_T = diag\Bigg(T^{-4/2}, T^{-5/2}, \cdots, T^{-(2k+2)/2}, \underbrace{T^{-3/2}, \cdots, T^{-3/2}}_{p-1}, \underbrace{T^{-4/2}, \cdots, T^{-4/2}}_{p-1}, \cdots, \underbrace{T^{-(2k+1)/2}, \cdots, T^{-(2k+1)/2}}_{p-1}\Bigg).$$

Thus, the F_{nl} statistics in (17) becomes

$$F_{nl} = \frac{1}{\hat{\sigma}^2_{\epsilon^*}}(D_TX_2'M_1\epsilon)'(D_TX_2'M_1X_2D_T)^{-1}(D_TX_2'M_1\epsilon). \tag{76}$$

In the following, we derive the asymptotic distribution of F_{nl}. Firstly, we consider the asymptotic distributions of $D_TX_2'M_1X_2D_T$ and $D_TX_2'M_1\epsilon$.

(i) Rewrite $D_TX_2'M_1X_2D_T$ as following

$$D_TX_2'M_1X_2D_T = D_TX_2'X_2D_T - \frac{1}{\sqrt{T}}D_TX_2'X_1\left(\frac{1}{T}X_1'X_1\right)^{-1}X_1'X_2D_T\frac{1}{\sqrt{T}}. \tag{77}$$

Let us define $\frac{1}{T}X_1'X_1 =$

$$\begin{bmatrix} \frac{1}{T}\sum_{t=p+1}^T(\Delta y_{t-1})^2 & \frac{1}{T}\sum_{t=p+1}^T\Delta y_{t-1}\Delta y_{t-2} & \cdots & \frac{1}{T}\sum_{t=p+1}^T\Delta y_{t-1}\Delta y_{t-(p-1)} \\ \frac{1}{T}\sum_{t=p+1}^T\Delta y_{t-1}\Delta y_{t-2} & \frac{1}{T}\sum_{t=p+1}^T(\Delta y_{t-2})^2 & \cdots & \frac{1}{T}\sum_{t=p+1}^T\Delta y_{t-2}\Delta y_{t-(p-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{T}\sum_{t=p+1}^T\Delta y_{t-1}\Delta y_{t-(p-1)} & \frac{1}{T}\sum_{t=p+1}^T\Delta y_{t-2}\Delta y_{t-(p-1)} & \cdots & \frac{1}{T}\sum_{t=p+1}^T(\Delta y_{t-(p-1)})^2 \end{bmatrix}.$$

Under the null hypothesis and by using the results in Theorem 1,

$$\frac{1}{T}X_1'X_1 \Rightarrow \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-3} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{bmatrix} = \gamma_0 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-2} \\ \rho_1 & 1 & \cdots & \rho_{p-3} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{p-2} & \gamma_{p-3} & \cdots & 1 \end{bmatrix} = \gamma_0\mathbf{\Pi} \tag{78}$$

where $\rho_i = \gamma_i/\gamma_0$, for $i = 1, \ldots, (p-2)$.

Let

$$D_TX_2'X_2D_T = \begin{bmatrix} A_1 & A_2 \\ A_2' & A_3 \end{bmatrix}$$

where

$$A_1 = \begin{bmatrix} \frac{1}{T^4}\sum_{t=p+1}^T y_{t-1}^2y_{t-d}^4 & \cdots & \frac{1}{T^{(2k+6)/2}}\sum_{t=p+1}^T y_{t-1}^2y_{t-d}^{2k+2} \\ \vdots & \ddots & \vdots \\ \frac{1}{T^{(2k+6)/2}}\sum_{t=p+1}^T y_{t-1}^2y_{t-d}^{2k+2} & \cdots & \frac{1}{T^{(4k+4)/2}}\sum_{t=p+1}^T y_{t-1}^2y_{t-d}^{4k} \end{bmatrix},$$

$$A_2 = \left[\begin{array}{cccc} A_{21} & A_{22} & \cdots & A_{2(2k-1)} \end{array} \right]$$

with

$$A_{2i} = \left[\begin{array}{ccc} \frac{1}{T^{(6+i)/2}} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{i+3} \Delta y_{t-1} & \cdots & \frac{1}{T^{(6+i)/2}} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{i+3} \Delta y_{t-(p-1)} \\ \frac{1}{T^{(7+i)/2}} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{i+4} \Delta y_{t-1} & \cdots & \frac{1}{T^{(7+i)/2}} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{i+4} \Delta y_{t-(p-1)} \\ \vdots & \ddots & \vdots \\ \frac{1}{T^{(2k+4+i)/2}} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{i+2k+1} \Delta y_{t-1} & \cdots & \frac{1}{T^{(2k+4+i)/2}} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{i+2k+1} \Delta y_{t-(p-1)} \end{array} \right],$$

and

$$A_3 = \left[\begin{array}{ccccc} A_{31} & A_{32} & A_{33} & \cdots & A_{3(2k-1)} \\ A_{32} & A_{33} & A_{34} & \cdots & A_{3(2k)} \\ A_{33} & A_{34} & A_{35} & \cdots & A_{3(2k+1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{3(2k-1)} & A_{3(2k)} & A_{3(2k+1)} & \cdots & A_{3(4k-3)} \end{array} \right]_{[(p-1)(2k-1)] \times [(p-1)(2k-1)]}$$

with

$$A_{3i} = \left[\begin{array}{ccc} \frac{1}{T^{(5+i)/2}} \sum_{t=p+1}^{T} y_{t-d}^{i+3} (\Delta y_{t-1})^2 & \cdots & \frac{1}{T^{(5+i)/2}} \sum_{t=p+1}^{T} y_{t-d}^{i+3} \Delta y_{t-1} \Delta y_{t-(p-1)} \\ \vdots & \ddots & \vdots \\ \frac{1}{T^{(5+i)/2}} \sum_{t=p+1}^{T} y_{t-d}^{i+3} \Delta y_{t-1} \Delta y_{t-(p-1)} & \cdots & \frac{1}{T^{(5+i)/2}} \sum_{t=p+1}^{T} y_{t-d}^{i+3} (\Delta y_{t-(p-1)})^2 \end{array} \right].$$

Under the null hypothesis and by using the results in Theorem 1,

$$A_1 \Rightarrow \left[\begin{array}{cccc} \lambda^6 \int W^6 & \lambda^7 \int W^7 & \cdots & \lambda^{(2k+4)} \int W^{(2k+4)} \\ \lambda^7 \int W^7 & \lambda^8 \int W^8 & \cdots & \lambda^{(2k+5)} \int W^{(2k+5)} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda^{(2k+4)} \int W^{(2k+4)} & \lambda^{(2k+5)} \int W^{(2k+5)} & \cdots & \lambda^{(4k+2)} \int W^{(4k+2)} \end{array} \right]_{(2k-1)\times(2k-1)}, \tag{79}$$

$$A_2 \Rightarrow \mathbf{0}_{(2k-1)\times(p-1)(2k-1)}$$

and

$$A_3 \Rightarrow \left[\begin{array}{cccc} \mathbf{\Pi}\gamma_0 \lambda^4 \int W^4 & \mathbf{\Pi}\gamma_0 \lambda^5 \int W^5 & \cdots & \mathbf{\Pi}\gamma_0 \lambda^{(2k+2)} \int W^{(2k+2)} \\ \mathbf{\Pi}\gamma_0 \lambda^5 \int W^5 & \mathbf{\Pi}\gamma_0 \lambda^6 \int W^6 & \cdots & \mathbf{\Pi}\gamma_0 \lambda^{(2k+3)} \int W^{(2k+3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Pi}\gamma_0 \lambda^{(2k+2)} \int W^{(2k+2)} & \mathbf{\Pi}\gamma_0 \lambda^{(2k+3)} \int W^{(2k+3)} & \cdots & \mathbf{\Pi}\gamma_0 \lambda^{(4k)} \int W^{(4k)} \end{array} \right].$$

Let us define

$$\frac{1}{\sqrt{T}} D_T X_2' X_1 = \left[\begin{array}{c} B_1 \\ B_2 \end{array} \right]$$

where

$$B_1 = \left[\begin{array}{ccc} \frac{1}{T^{5/2}} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{2} \Delta y_{t-1} & \cdots & \frac{1}{T^{5/2}} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{2} \Delta y_{t-(p-1)} \\ \frac{1}{T^{6/2}} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{3} \Delta y_{t-1} & \cdots & \frac{1}{T^{6/2}} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{3} \Delta y_{t-(p-1)} \\ \vdots & \ddots & \vdots \\ \frac{1}{T^{(2k+3)/2}} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{2k} \Delta y_{t-1} & \cdots & \frac{1}{T^{(2k+3)/2}} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{2k} \Delta y_{t-(p-1)} \end{array} \right]$$

and

$$B_2 = \left[\begin{array}{c} B_{21} \\ \vdots \\ B_{2(2k-1)} \end{array} \right]$$

with

$$B_{2i} = \begin{bmatrix} \frac{1}{T^{(3+i)/2}} \sum_{t=p+1}^{T} y_{t-d}^{i+1} (\Delta y_{t-1})^2 & \cdots & \frac{1}{T^{(3+i)/2}} \sum_{t=p+1}^{T} y_{t-d}^{i+1} \Delta y_{t-1} \Delta y_{t-(p-1)} \\ \frac{1}{T^{(3+i)/2}} \sum_{t=p+1}^{T} y_{t-d}^{i+1} \Delta y_{t-2} \Delta y_{t-1} & \cdots & \frac{1}{T^{(3+i)/2}} \sum_{t=p+1}^{T} y_{t-d}^{i+1} \Delta y_{t-2} \Delta y_{t-(p-1)} \\ \vdots & \ddots & \vdots \\ \frac{1}{T^{(3+i)/2}} \sum_{t=p+1}^{T} y_{t-d}^{i+1} \Delta y_{t-(p-1)} \Delta y_{t-1} & \cdots & \frac{1}{T^{(3+i)/2}} \sum_{t=p+1}^{T} y_{t-d}^{i+1} (\Delta y_{t-(p-1)})^2 \end{bmatrix}.$$

Under the null hypothesis and by using the results in Theorem 1,

$$\begin{aligned} B_1 &\Rightarrow \mathbf{0}_{(2k-1)\times(p-1)}, \\ B_2 &\Rightarrow \begin{bmatrix} \mathbf{\Pi}\gamma_0\lambda^2 \int W^2 \\ \vdots \\ \mathbf{\Pi}\gamma_0\lambda^{2k} \int W^{2k} \end{bmatrix}_{(p-1)(2k-1)\times(p-1)}. \end{aligned}$$

Thus,

$$\frac{1}{\sqrt{T}} D_T X_2' X_1 \Rightarrow \begin{bmatrix} \mathbf{0}_{(2k-1)\times(p-1)} \\ \mathbf{\Pi}\gamma_0\lambda^2 \int W^2 \\ \vdots \\ \mathbf{\Pi}\lambda^{2k} \int W^{2k} \end{bmatrix}. \tag{80}$$

$$\begin{aligned} & \frac{1}{\sqrt{T}} D_T X_2' X_1 \left(\frac{X_1' X_1}{T} \right)^{-1} X_1' X_2 D_T \frac{1}{\sqrt{T}} \\ \Rightarrow & \begin{bmatrix} \mathbf{0}_{(2k-1)\times(p-1)} \\ \mathbf{\Pi}\gamma_0\lambda^2 \int W^2 \\ \vdots \\ \mathbf{\Pi}\gamma_0\lambda^{2k} \int W^{2k} \end{bmatrix} \frac{1}{\gamma_0} \mathbf{\Pi}^{-1} \begin{bmatrix} \mathbf{0}_{(p-1)\times(2k-1)} & \mathbf{\Pi}\gamma_0\lambda^2 \int W^2 & \cdots & \mathbf{\Pi}\gamma_0\lambda^{2k} \int W^{2k} \end{bmatrix} \\ = & \begin{bmatrix} \mathbf{0}_{(2k-1)\times(2k-1)} & \mathbf{0}_{(2k-1)\times(p-1)} & \cdots & \mathbf{0}_{(2k-1)\times(p-1)} \\ \mathbf{0}_{(p-1)\times(2k-1)} & \gamma_0\lambda^4 (\int W^2)^2 \mathbf{\Pi} & \cdots & \gamma_0\lambda^{2k+2} \int W^2 \int W^{2k} \mathbf{\Pi} \\ \mathbf{0}_{(p-1)\times(2k-1)} & \gamma_0\lambda^5 \int W^2 \int W^3 \mathbf{\Pi} & \cdots & \gamma_0\lambda^{2k+3} \int W^3 \int W^{2k} \mathbf{\Pi} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{(p-1)\times(2k-1)} & \gamma_0\lambda^{2k+2} \int W^2 \int W^{2k} \mathbf{\Pi} & \cdots & \gamma_0\lambda^{4k} (\int W^{2k})^2 \mathbf{\Pi} \end{bmatrix}. \end{aligned}$$

Thus,

$$\begin{aligned} D_T X_2' M_1 X_2 D_T &= D_T X_2' X_2 D_T - \frac{1}{\sqrt{T}} D_T X_2' X_1 \left(\frac{X_1' X_1}{T} \right)^{-1} X_1' X_2 D_T \frac{1}{\sqrt{T}} \\ &\Rightarrow \begin{bmatrix} C_1 & \mathbf{0} \\ \mathbf{0} & C_2 \end{bmatrix} \\ &= \triangle F_2(W) \triangle \end{aligned} \tag{81}$$

where C_1 is the asymptotic distribution of A_1 in (79), C_2 is the asymptotic distribution of $A_3 - B_2$,

$$\triangle = diag\left(\lambda^3, \lambda^4, \cdots, \lambda^{2k+1}, \underbrace{\sqrt{\gamma_0}\lambda^2, \cdots, \sqrt{\gamma_0}\lambda^2}_{p-1}, \cdots, \underbrace{\sqrt{\gamma_0}\lambda^{2k}, \cdots, \sqrt{\gamma_0}\lambda^{2k}}_{p-1} \right) \tag{82}$$

and $F_2(W)$ is defined in Theorem 2.

(ii) $D_T X_2' M_1 \epsilon$ can be written as

$$D_T X_2' M_1 \epsilon = D_T X_2' \epsilon - \frac{1}{\sqrt{T}} D_T X_2' X_1 \left(\frac{1}{T} X_1' X_1 \right)^{-1} \frac{1}{\sqrt{T}} X_1' \epsilon. \tag{83}$$

The first term of (83) is

$$D_T X_2' \epsilon = \begin{bmatrix} E_1 \\ E_{21} \\ \vdots \\ E_{2(2k-1)} \end{bmatrix}$$

where

$$E_1 = \begin{bmatrix} T^{-2} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^2 \epsilon_t \\ \vdots \\ T^{-(2k+2)/2} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{2k} \epsilon_t \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma_\epsilon \lambda^3 \int W^3 dW \\ \vdots \\ \sigma_\epsilon \lambda^{(2k+1)} \int W^{(2k+1)} dW \end{bmatrix}, \tag{84}$$

$$E_{2i} = \begin{bmatrix} T^{-(1+i/2)} \sum_{t=p+1}^{T} y_{t-2}^{i+1} \Delta y_{t-1} \epsilon_t \\ \vdots \\ T^{-(1+i/2)} \sum_{t=p+1}^{T} y_{t-1}^{i+1} \Delta y_{t-(p-1)} \epsilon_t \end{bmatrix} \Rightarrow \begin{bmatrix} \sqrt{\gamma_0} \sigma_\epsilon \lambda^{i+1} \int W^{i+1} dW_1 \\ \vdots \\ \sqrt{\gamma_0} \sigma_\epsilon \lambda^{i+1} \int W^{i+1} dW_{(p-1)} \end{bmatrix}$$

For the second term of (83),

$$\frac{1}{\sqrt{T}} X_1' \epsilon \quad = \quad \begin{bmatrix} T^{-1/2} \sum_{t=p+1}^{T} \Delta y_{t-1} \epsilon_t \\ T^{-1/2} \sum_{t=p+1}^{T} \Delta y_{t-2} \epsilon_t \\ \vdots \\ T^{-1/2} \sum_{t=p+1}^{T} \Delta y_{t-(p-1)} \epsilon_t \end{bmatrix} \Rightarrow \begin{bmatrix} \sqrt{\gamma_0} \sigma_\epsilon W_1(1) \\ \sqrt{\gamma_0} \sigma_\epsilon W_2(1) \\ \vdots \\ \sqrt{\gamma_0} \sigma_\epsilon W_{(p-1)}(1) \end{bmatrix} \tag{85}$$

Thus, gathering (80), (78) and (85),

$$\begin{aligned} & \tfrac{1}{\sqrt{T}} D_T X_2' X_1 \left(\tfrac{X_1' X_1}{T} \right)^{-1} X_1' \epsilon \tfrac{1}{\sqrt{T}} \\ \Rightarrow & \begin{bmatrix} \mathbf{0} \\ \mathbf{\Pi} \gamma_0 \lambda^2 \int W^2 \\ \vdots \\ \mathbf{\Pi} \gamma_0 \lambda^{2k} \int W^{2k} \end{bmatrix} \tfrac{1}{\gamma_0} \mathbf{\Pi}^{-1} \begin{bmatrix} \sqrt{\gamma_0} \sigma_\epsilon W_1(1) \\ \sqrt{\gamma_0} \sigma_\epsilon W_2(1) \\ \vdots \\ \sqrt{\gamma_0} \sigma_\epsilon W_{(p-1)}(1) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ D_1 \\ \vdots \\ D_{(2k-1)} \end{bmatrix} \end{aligned} \tag{86}$$

where

$$D_i = \begin{bmatrix} \sqrt{\gamma_0} \sigma_\epsilon \lambda^{i+1} \int W^{i+1} W_1(1) \\ \vdots \\ \sqrt{\gamma_0} \sigma_\epsilon \lambda^{i+1} \int W^{i+1} W_{(p-1)}(1) \end{bmatrix}$$

Thus,

$$\begin{aligned} D_T X_2' M_1 \epsilon &= D_T X_2' \epsilon - \frac{1}{\sqrt{T}} D_T X_2' X_1 \left(\frac{X_1' X_1}{T} \right)^{-1} X_1' \epsilon \frac{1}{\sqrt{T}} \\ &= \begin{bmatrix} G_1 \\ G_{21} \\ \vdots \\ G_{2(2k-1)} \end{bmatrix} \Rightarrow \sigma_\epsilon \triangle F_1(W) \end{aligned} \tag{87}$$

where G_1 has the same asymptotic distribution with E_1 in (84),

$$G_{2i} = E_{2i} - D_i \Rightarrow \begin{bmatrix} \sqrt{\gamma_0} \sigma_\epsilon \lambda^{i+1} \left(\int W^{i+1} dW_1 - W_1(1) \int W^{i+1} \right) \\ \vdots \\ \sqrt{\gamma_0} \sigma_\epsilon \lambda^{i+1} \left(\int W^{i+1} dW_{(p-1)} - W_{(p-1)}(1) \int W^{i+1} \right) \end{bmatrix},$$

$\triangle$ and $F_1(W)$ are defined in (82) and Theorem 2 respectively.

Thus, under H_0, the asymptotic distribution of F_{nl} can be determined using the following results,

$$
\begin{aligned}
F_{nl} &= \frac{1}{\hat{\sigma}^2_{\epsilon^*}}\left(\hat{\mathbf{b}}_2-\mathbf{b}_2\right)'(X_2'M_1X_2)\left(\hat{\mathbf{b}}_2-\mathbf{b}_2\right) \\
&= \frac{1}{\hat{\sigma}^2_{\epsilon^*}}(D_TX_2'M_1\epsilon)'(D_TX_2'M_1X_2D_T)^{-1}(D_TX_2'M_1\epsilon) \\
&\Rightarrow \frac{1}{\hat{\sigma}^2_{\epsilon^*}}\left(\sigma_\epsilon\triangle F_1(W)\right)'\left(\triangle F_2(W)\triangle\right)^{-1}\left(\sigma_\epsilon\triangle F_1(W)\right) \\
&= \frac{\sigma^2_\epsilon}{\hat{\sigma}^2_{\epsilon^*}}\left(F_1(W)\right)'\triangle\triangle^{-1}\left(F_2(W)\right)^{-1}\triangle^{-1}\triangle\left(F_1(W)\right) \\
&= (F_1(W))'(F_2(W))^{-1}F_1(W). \qquad (88)
\end{aligned}
$$

The final result in (88) is obtained because under H_0, $\epsilon^*=\epsilon$ and $\widehat{\sigma}^2_{\epsilon^*}$ is a consistent estimator of $\sigma^2_{\epsilon^*}$. Thus, $\widehat{\sigma}^2_{\epsilon^*}$ is also a consistent estimator of σ^2_{ϵ}.

References

Bair, J. and Haesbroeck, G. (1997) Monotonous stability for neutral fixed points. *Bulletin Belgium Mathematics Society*, **4**, 639–646.

Balke, N. and Fomby, T. (1997) Threshold cointegration. *International Economics Review*, **38**, 627–645.

Benhabib, J., Schmitt-Grohe, S. and Uribe, M. (1999) The perils of Taylor rules. CEPR Discussion Papers 2314.

Berben, R. and van Dijk, D. (1999) Unit root tests and asymmetric adjustment: A reassessment. Unpublish manuscript, Tinbergen Institute, Erasmus University of Rotterdam.

Beveridge, S. and Nelson, C. (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics*, **7**, 151–174.

Box, G. and Jenkins, G. (1976) *Time Series Analysis: Forecasting and Control.* Holden-Day Inc.

Cagan, P. (1956) The monetary dynamics of hyperinflantion. In *Studies in the Quantity Theory of Money* (ed. M. Friedman), 25–117. Chicago: University of Chicago Press.

Caner, M. and Hansen, B. (2001) Threshold autoregression with a near unit root. *Econometrica*, **69**, 1555–1596.

Diamond, P. (1982) Aggregate demand management in search equilibrium. *Journal of Political Economy*, **90**, 881–894.

Dickey, D. and Fuller, W. (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**, 427–431.

Dickey, D. and Fuller, W. (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, **49**, 1057–1072.

Eklund, B. (2003) A nonlinear alternative to the unit root hypothesis. SSE/EFI Working Paper Series in Economics and Finance 547, Department of Economic Statistics, Stockholm School of Economics, Stockholm, Sweden.

Enders, W. and Granger, C. (1998) Unit root tests and asymmetric adjustment with an example using the trem structure of interest rates. *Journal of Business and Economics Statistics*, **16**, 304–311.

Fuller, W. (1976) *Introduction to Statistical Time Series.* New York: John Wiley.

Granger, C. and Terasvirta, T. (1993) *Modelling Nonlinear Economic Relationship.* Oxford: Oxford University Press.

He, C. and Sandberg, R. (2005) Dickey–Fuller type of tests against nonlinear dynamic models. SEE/EFI Working Paper Series in Economics and Finance 580, Department of Economic Statistics, Stockholm School of Economics, Stockholm, Sweden.

Hong, S. and Phillips, P. (2010) Testing linearity in cointegrating relations with an application to purchasing power parity. *Journal of Business and Economic Statistics*, **28**, 96–104.

Kapetanios, G., Shin, Y. and Snell, A. (2003) Testing for a unit root in the nonlinear STAR framework. *Journala of Econometrics*, **112**, 359–379.

Layard, R., Nickell, S. and Jackman, R. (1991) *Unemployment: Macroeconomic Performance and the Labour Market*. Oxford: Oxford University Press.

Li, H. and Maddala, G. (1996) Bootstrapping time series models. *Econometric Reviews*, **15**, 115–158.

Lo, M. and Zivot, E. (2001) Threshold cointegration and nonlinear adjustment to the law of one price. *Macroeconomic Dynamics*, **5**, 533–576.

Michael, P., Nobay, R. and Peel, D. (1997) Transactions costs and non-lienar adjustment in real exchange rates: An empirical investigation. *Journal of Political Economy*, **105**, 862–879.

Monoyios, M. and Sarno, L. (2002) Mean reversion in stock index futures markets: A nonlinear analysis. *The Journal of Futures Markets*, **22**, 285–314.

Paya, I., Venetis, I. and Peel, D. (2003) Further evidence on PPP adjustment speeds: the case of effective real exchange rates and the EMS. *Oxford Bulletin of Economics and Statistics*, **65**, 421–438.

Phillips, P. and Solo, V. (1992) Asymptotics for linear processes. *The Annals of Statistics*, **20**, 971–1001.

Pippenger, M. and Goering, G. (1993) A note on the empirical power of unit root tests under threshold processes. *Oxford Bulletin of Economics and Statistics*, **55**, 473–481.

Ramsey, J. (1969) Tests for specification errors in classical least-squares regression analysis. *Journal of the Royal Statistical Analysis, Series B,*, **13**, 350–371.

Sargent, T. and Wallace, N. (1973) Rational expectations and the dynamics of hyperinflantion. *International Economics Review*, **14**, 328–350.

Sarno, L., Taylor, M. and Peel, D. (2002) Nonlinear equilibrium correction in US real money balances, 1869–1997. *Journal of Money, Credit and Banking*, **35**, 787–799.

Taylor, M., Peel, D. and Sarno, L. (2001) Nonlinear in real exchange rates: Towards a solution of 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.

Terasvirta, T. and Elliasson, A. (2001) Nonlinear error correction and the UK demand for broad money, 1878–1993. *Journal of Applied Econometrics*, **16**, 277–288.

Venetis, I., Paya, I. and Peel, D. (2009) ESTAR model with multiple fixed points: Testing and estimation. Working paper, Department od Economics, Lancaster University Management School, Lancaster, UK.