Lancaster University Management School
Working Paper
2009/010

ESTAR model with multiple fixed points. Testing and Estimation

Ioannis A.Venetis, Ivan Paya and David Peel

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

ESTAR model with multiple fixed points. Testing and Estimation

Ioannis A. Venetis[a], Ivan Paya[b,1], David A. Peel[b]
[a]University of Patras, Department of Economics, University Campus, Rio, 26 504 Greece
[b]Lancaster University Management School, Lancaster, LA1 4YX, UK

Abstract

In this paper we propose a globally stationary augmentation of the Exponential Smooth Transition Autoregressive (ESTAR) model that allows for multiple fixed points in the transition function. An F-type test statistic for the null of nonstationarity against such globally stationary nonlinear alternative is developed. The test statistic is based on the standard approximation of the nonlinear function under the null hypothesis by a Taylor series expansion. The model is applied to the U.S real interest rate data for which we find evidence of the new ESTAR process.

Keywords: ESTAR, unit toot, real interest rates.
JEL classification: E43, C22, C52

[1]Corresponding author. Email address: i.paya@lancaster.ac.uk Tel: +44(0) 1524593504. Fax: +44(0) 1524594244.

Acknowledgements: We got valuable feedback from participants at the 14th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, 1st Workshop ESRC Seminar Series Nonlinear Economics (Keele University), Economic Analysis Department University of Alicante, and Lancaster University. Ivan Paya acknowledges financial support from the Spanish Ministerio de Educacion y Ciencia Research Project SEJ2005-02829/ECON.

1 Introduction

The exponential smooth transition autoregressive (ESTAR) process developed by Haggan and Ozaki (1981) has become a popular method for modelling a variety of relationships in macroeconomics and finance. Real exchange rates and purchasing power parity (PPP) deviations have been throroughly analysed using the ESTAR model (see e.g., Michael et al., 1997; Taylor et al., 2001; and Paya et al., 2003).[2] Empirical analysis of deviations from optimal money holdings have also been estimated using nonlinear ESTAR models (see Terasvirta and Eliasson, 2001; Sarno et al., 2003). Monetary policy rules where the central bank would follow the oportunistic approach to disinflation prposed by Orphanides and Wilcox (1996) have also been found to follow similar process than the ESTAR (see Bec et al., 2000). This type of model has also been used in finance. Symmetric deviations from arbitrage processes such as stock index futures have been reported to follow the process described by the ESTAR model (Monoyios and Sarno, 2002).

The standard ESTAR model is such that the transition function is bounded between zero and one depending on how far away the transition variable is away from a determined value, usually called "equilibrium".[3] For instance, in the case of the PPP, the further away the real exchange rate is from one (fixed equilibrium) the faster the real exchange rate would revert to such equilibrium.[4] However, many economic theories support the existence of multiple equilibria. For example, in the case of inflation, attempts by governments to finance substantial proportion of expenditure by seigniorage can lead to multiple inflationary equilibria (Cagan, 1956; Sargent and Wallace, 1973; Evans et al., 1996). Theoretical models suggest that, in these circumstances, inflation follows a non-linear process and that the stability characteristics depend on expectations formation. In the case of unemployment, shocks causing sharp cyclical swings in unemployment generate political reactions from public producing not merely fiscal and monetary (demand policy) responses but also changes in supply-side policy (affecting the equilibrium values of real variables or 'natural rates') (see Diamond, 1982; and Layard et al., 1991). With regard to monetary policy rules, some models suggest that once you take into account the zero bound on nominal interest rates, real interest rates might follow a number of equilibria (see Benhabib, Schmitt-Grohe, and

[2]The ESTAR functional form is even suggested explicitly in some economic models of real exchange rates (see Dumas, 1992; Sercu et al. 1995).

[3]The standard ESTAR model could also exhibit up until three equilibria. However, for this to be the case, you would need an explosive autoregressive process.

[4]See Paya and Peel (2006) for the case where the real exchange rate would follow an ESTAR model with time-vaying equilibrium.

Uribe, 1999).

In this paper we propose a new ESTAR type model that allows for multiple fixed points in the transition function. The purpose of this model is threefold: (i) it allows for multiple fixed points in a way that is parsimonious (stationary), (ii) it introduces up to 'k' points at which dynamics of the system might be similar in neighbourhood, and (iii) it allows data to determine if such possibilities exist and therefore generalises existing model.

The rest of the paper is organised as follows. Section 2 describes the k-ESTAR model. Section 3 presents the power of the Kapetanios, Shin and Snell (KSS) (2003) unit root test in the case where the alternative is generated by a k-ESTAR model. Section 4 develops a testing procedure to detect the presence of the k-ESTAR form when the null is a unit root. Section 5 examines the small sample properties of the test developed in section 4. Section 6 presents an empirical application using the US real interest rate, and Section 7 concludes.

2 The k-ESTAR model

The ESTAR model was introduced by Haggan and Ozaki (1981) and popularized by Granger and Terasvirta (1993) and Terasvirta (1994).[5] In this section we develop an extended version of the ESTAR model. In particular, we consider a nonlinear model of the form

$$y_t = \beta_0 + \sum_{j=1}^{p} \beta_j y_{t-j} + \left[\gamma_0 + \sum_{j=1}^{p} \gamma_j y_{t-j}\right] G(\alpha_k, \mathbf{r}; y_{t-d}) + u_t \tag{1}$$

with

$$\begin{aligned} G(\alpha_k, \mathbf{r}; y_{t-d}) &= \left[1 - \exp\left\{-f^2(\alpha_k, \mathbf{r}; y_{t-d})\right\}\right] \\ f(\alpha_k, \mathbf{r}; y_{t-d}) &= a_k\left(y_{t-d} - r_1\right)\left(y_{t-d} - r_2\right) \ldots \left(y_{t-d} - r_k\right) \end{aligned} \tag{2}$$

where u_t is a stationary and ergodic martingale difference sequence with variance σ^2, $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \beta_2, \cdots, \beta_p\right)'$, $\boldsymbol{\gamma} = \left(\gamma_0, \gamma_1, \gamma_2, \cdots, \gamma_p\right)'$, $\mathbf{r} = (r_1, r_2, \cdots, r_k)'$, α_k are unknown parameters and we make an implicit assumption that the location parameters satisfy $r_1 < r_2 < \ldots < r_k$. The variable y_{t-d} for $d \in 1, 2, \ldots, d_{\max}$ in function $G(\alpha_k, \mathbf{r}; y_{t-d})$ is the transition variable. Define the polynomials $\beta(L) = 1 - \sum_{j=1}^{p} \beta_j$ and $\gamma(L) = 1 - \sum_{j=1}^{p} \gamma_j$ as the "linear

[5] A survey of recent developments in ESTAR modelling can be found in van Dijk et al. (2002).

and nonlinear autoregressive polynomials". We are interested in the special case of a unit root $\beta(L)=0$ in the linear polynomial thus all our subsequent analysis is based on the restriction

$$\sum_{j=1}^{p}\beta_j = 1 \tag{3}$$

This is a generalized version of the ESTAR model employed by KSS which is nested in (1) for $k=1$, and our notation conforms as much as possible with the notation in KSS. We refer to the ESTAR model (1) with transition function (2) as the "$k-ESTAR$" model. Please note that the transition function $G(.)$ no longer admits the familiar U-shape of the $1-ESTAR$ model although it is bounded between 0 and 1. The smoothness or transition speed parameter α_k is one of the factors that determine the speed of transition between regimes $G(.)=0$ and $G(.)=1$ along with the distance of y_{t-d} from a specified location r_i (as in the typical ESTAR model).[6] However, notice that the $k-ESTAR$ model supports a much wider dynamic behavior since adjustment speed need not be symmetric around any location point depending on the number of the location points as well as their relative distance.

Similar geometric ergodicity and associated global stationarity conditions as those explained by KSS hold for model (1). Following Bhattacharya and Lee (1995, theorem 1) we assume $\left|\sum_{j=1}^{p}\left(\beta_j+\gamma_j\right)\right|<1$. Very general but difficult to verify conditions for geometric ergodicity and mixing properties of nonlinear autoregressive models are given in Liebscher (2005).

The novelty with representation (1) is that it allows for multiple endogenously determined "equilibria" where an equilibrium is considered to be any real valued fixed point y_* that solves

$$0=\beta_0+\left[\gamma_0+y\sum_{j=1}^{p}\gamma_j\right]\times G(\alpha_k,\mathbf{r};y) \tag{4}$$

When $y_{t-d}=r_1\vee r_2\vee\cdots\vee r_k$, the $k-ESTAR$ model allows for multiple "inner" regimes with $G(.)=0$ and (1) reduces to

$$\Delta y_t=\beta_0+u_t \tag{5}$$

behaving as a random walk process (with drift if $\beta_0\neq 0$). For $1-ESTAR$ models this case is consistent with the existence of an attractor (or "equilibrium") around which the series behaves as a random walk. For certain

[6] See Figure 1 for two transition functions $G(.)$ with different speeds of adjustment and same two "fixed" points.

parameter restrictions, the $k-ESTAR$ has one attractor that is a stable fixed point but allows for more than one "random walk points".

For example, we are interested in the cases where $\beta_0 = 0$ and $\sum_{j=1}^{p} \beta_j = 1$.[7] Then $y_0^* = -\frac{\gamma_0}{\sum_{j=1}^{p} \gamma_j}$ is a stable fixed point while $y_i^* = r_i$ $i = 1, ..., k$ represent positively neutral fixed points.[8] The same results hold if γ_0 is replaced with an r_i point in order to reduce the number of fixed points considered. For example if $\gamma_0 = r_1$ then r_1 is stable and the previous analysis hold true for all remaining r_i points $i \neq 1$.

In the "outer" regimes $[(y_{t-d} - r_1) \to -\infty$ and $(y_{t-d} - r_k) \to +\infty]$, function $G(.) \to 1$ and model (1) reduces to

$$\gamma(L)y_t = (\beta_0 + \gamma_0) + u_t \tag{6}$$

Depending on the magnitude of α_k and $\mathbf{r}$ it is possible to obtain (6) for values of y_{t-d} between the location points r_i as well.

In recent years new testing procedures have been developed in order to test the null of a unit root against nonlinear ESTAR alternatives (see Kapetanios et al., 2003). A natural step is then to find out whether those tests have power against the new $k-ESTAR$ model.

3 Small sample power of KSS t-test against k-ESTAR alternatives

An initial consideration is the small sample power of the t-test devised by KSS against the more elaborate $k-ESTAR$ model. The KSS $t-test$ is based on a finite Taylor approximation method of the nonlinear function and as such its power depends on the adequacy of the approximation under the alternative. The test is based on the $t-ratio$ of δ from the auxiliary regression

$$\Delta y_t = \delta y_{t-1}^3 + error \tag{7}$$

[7] Note also that when the latter hold the restriction $-2 < \sum_{j=1}^{p} \gamma_j < 0$ ensures ergodicity of the process.

[8] Following Bair and Haesbroeck (1997) further differentiation reveals that r_i^* are monotonously semistable from below as long as $r_i^* > -\frac{\gamma_0}{\sum_{j=1}^{p} \gamma_j}$ and monotonously semistable from above when $r_i^* < -\frac{\gamma_0}{\sum_{j=1}^{p} \gamma_j}$.

If the data is generated by a $k-ESTAR$ process then we expect the small sample power of the KSS $t-test$ to decrease. In a small scale experiment, we create series y_t based on the following DGP,

$$\Delta y_t = \gamma_1 y_{t-1}\left(1-\exp\{-a_k^2\left[(y_{t-1}-r_1)\dots(y_{t-1}-r_k)\right]^2\}\right)+\eta_t \tag{8}$$
$$y_0=0,\ t=1,...,T\quad \eta_t \sim N.I.D(0,1)$$

Different persistence profiles were examined using $\gamma_1=\{-1.5,-1,-0.5,-0.1\}$. For example, when $\gamma_1=-1$ and the process at $t-1$ is located far towards the outer regimes, it becomes $i.i.d$ and mean reverts to the full extend of y_{t-1}, that is $E(\Delta y_t|\, y_{t-1})=-y_{t-1}$, within one period, while for $\gamma_1=-1.5$ the series "overreacts" with $E(\Delta y_t|\, y_{t-1})=-1.5y_{t-1}$. As γ_1 approaches zero the series becomes progressively more persistent.

We consider a small sample size of $T=100$ where the first 150 observations are dropped to avoid initial condition effects and 50,000 replications are employed. An issue regarding the choice of parameter a_k (or a_k^2) in the simulation experiment arise. In general, the transition speed parameter a_k which affects the transition speed between fixed points is not scale free. In addition the parameter affects the persistence of the series with higher speeds implying less persistence. In the $1-ESTAR$ model employed by KSS this is resolved by setting $a_1^2(=\theta)=\{0.01,0.05,0.1,1\}$ after the observation that given γ,σ_η^2 (KSS, p.367) "*... the term $e^{-\theta y_{t-1}^2}$ measures the size of the largest root of the series at time t*". For comparison purposes we proceed in a similar way. In order to generate series of comparable persistence than in KSS but in a $k-ESTAR$ model, we use samples of $T=2,000$ and $2,000$ replications to generate y_t according to (8) with $\gamma=-1$, $\sigma_\eta^2=1$, $r_1=0$, $r_2=3$, $r_3=6$ and we search for a_k^2 values such that $\bar{\Xi}^*\approx\{0.95,0.80,0.5,0.25\}$ where

$$\Xi_t=\exp\{-a_3^2\left[(y_{t-1}-r_1)(y_{t-1}-r_2)(y_{t-1}-r_3)\right]^2\}$$

and $\bar{\Xi}^*$ denotes the average of the sample mean of Ξ_t across replications. The corresponding values of θ were

- $\theta=\{0.01,0.18,1.425,7.45\}$ for $r_1=0$,
- $\theta=\left\{\frac{0.01}{5.2},\frac{0.18}{4.35},\frac{1.425}{5.9},\frac{7.45}{7.855}\right\}$ for $r_1=0,r_2=3$ and
- $\theta=\left\{\frac{0.01}{80},\frac{0.18}{87.5},\frac{1.425}{154},\frac{7.45}{238}\right\}$ for $r_1=0,r_2=3,r_3=6$

For each replication, we estimate δ in (7) using OLS and we compare its $t-ratio$ value with the -2.22 critical value given in Table 1 of KSS. The rejection probabilities of the null hypothesis $H_0:\delta=0$ appear in table 1.

Not surprisingly, the table results confirm severe loss of power, especially for $k = 3$. In general the results show sensitivity of power to both transition speed and nonlinearity. As the transition speed magnitude decreases and the number of fixed points increase the loss of power increases rapidly. For example, when $k = 3$ and $\bar{\Xi}^* \approx 0.95$ with $\gamma = -1$ or $\gamma = -0.5$ the power of the test is 54.5% and 43% respectively. For values of γ that imply larger persistence, for example $\gamma = -0.1$, even with $\bar{\Xi}^* \approx 0.25$ the power is as low as 42.4%. Another finding is that the loss of power is not monotonic across persistence profiles. For moderate γ values the loss is smaller for $k = 1$ to $k = 2$ and then deteriorates significantly for $k = 3$.These findings conform with the orientation of the KSS test towards alternatives generated by the $1 - ESTAR$ model.

4 F-type testing procedure

In this section we develop an F-type test for the null hypothesis of unit root, $H_0 : a_k = 0$ in (1). Testing H_0 in (1) cannot be performed directly due to a well known identification problem (see Luukkonen et al. (1988), and Terasvirta (1994) for details). Following Luukkonen et al. (1988), the identification problem is circumvented by using a Taylor approximation of the nonlinear function $G(\alpha_k, .; .)$ around the null hypothesis.

Proposition 1 *In (1) let* $p \geq 1, k \geq 1, d \geq 1$ *and* $z_t = y_{t-d}$. *Also let* $\sum_{j=1}^{p} \beta_j = 1$. *Then, using a second order Taylor series approximation to the* $G(a_k, .; .)$ *function around* $a_k = 0$, *we obtain an auxiliary regression*

$$\Delta y_t = \lambda_0 + \sum_{j=1}^{p-1} \beta_j^{**} \Delta y_{t-j} + \sum_{j=1}^{p} \lambda_{1,j} y_{t-j} + \sum_{j=1}^{2k} \lambda_{2,j} z_t^j + \sum_{j=1}^{p} \sum_{s=1}^{2k} \lambda_{3,js} y_{t-j} z_t^s + error \tag{9}$$

If $r_i \neq 0$ *for all* i *then testing the null hypothesis of a unit root against the alternative of a "globally stationary" k-ESTAR process is equivalent with testing*

$$H_0 : \lambda_{1,j} = \lambda_{2,j} = \lambda_{3,js} = 0 \quad \text{for all } s, j$$

in (9). Constant λ_0 *is given by* $\lambda_0 = \beta_0 + a_k^2 \gamma_0 \delta_0$ *with* $\delta_0 = \prod_{i=1}^{k} r_i^2$ *while* $\lambda_{1,j} = a_k^2 \gamma_j \delta_0$.

If $r_i = 0$ *for a certain* i, *then testing the null hypothesis of a unit root against the alternative of a "globally stationary" k-ESTAR process is equivalent with testing*

$$H_0 : \lambda_{2,j} = \lambda_{3,js} = 0 \quad \text{for all } s, j = 2, ..., 2k$$

in the auxiliary regression

$$\Delta y_t = \beta_0 + \sum_{j=1}^{p-1} \beta_j^{**} \Delta y_{t-j} + \sum_{j=2}^{2k} \lambda_{2,j} z_t^j + \sum_{j=1}^{p} \sum_{s=2}^{2k} \lambda_{3,js} y_{t-j} z_t^s + error \qquad (10)$$

Proposition 1 is based on a second order Taylor series approximation of $G(\alpha_k, \mathbf{r}; .)$ around $\alpha_k = 0$.[9] If we differentiate with respect to a_k^2 then the usual first order approximation is enough and it yields identical results.

Equation (9) is heavily parameterized making the testing procedure cumbersome. If we only set $p = 1$, $k = 1$ a compact presentation of the auxiliary testing regression admits the form

$$\Delta y_t = \lambda_0 + \lambda_1 z_t + \lambda_2 z_t^2 + \lambda_3 y_{t-1} + \lambda_4 y_{t-1} z_t + \lambda_5 y_{t-1} z_t^2 + e_t$$

and the hypothesis of interest is translated into

$$H_0 : \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 0$$

However, we show in the appendix that not all regressors in (9) are necessary under the null hypothesis since they are asymptotically collinear leading to singular sample covariance matrices. Given Proposition 1, we identify the following testing procedure.

Proposition 2 *In (1) let $p \geq 1, k \geq 1, d \geq 1$ and $z_t = y_{t-d}$. Also let $\beta_0 = 0, \sum_{j=1}^{p} \beta_j = 1$. In order to test the null hypothesis of a unit root without drift against the alternative of a "globally stationary" k-ESTAR process estimate by least squares the following auxiliary regression*

$$\Delta y_t = \lambda_0 + \sum_{j=1}^{p-1} \beta_j^{**} \Delta y_{t-j} + b_1^* y_{t-1} + \sum_{j=2}^{2k} b_{2,j}^* y_{t-d}^j + b_3^* y_{t-1} y_{t-d}^{2k} + v_t \qquad (11)$$

and compute the F-type statistic

$$F_k = \frac{\hat{\mathbf{b}}_2^{*\prime} \left(X_2^{*\prime} M_1 X_2^* \right) \hat{\mathbf{b}}_2^*}{\hat{\sigma}_v^2} \qquad (12)$$

where $\hat{\mathbf{b}}_2^* = (\lambda_0, b_1^*, b_{2,2}^*, ... b_{2,2k}^*, b_3^*)'$ and $\hat{\sigma}_v^2$ the maximum likelihood estimator of the error variance. Under the null hypothesis $H_0 : a_k = 0$,

$$F_k \xrightarrow{d} G_{1*}'(W) G_{2*}^{-1}(W) G_{1*}(W)$$

[9] When we differentiate with respect to a_k we obtain $\left. \frac{\partial G}{\partial \alpha_k} \right|_{a_k=0} = 0$

where W denotes standard Brownian motion and G_{1*}, G_{2*} are functionals defined in the appendix. Under the alternative $H_1 : a_k > 0$ the F_k statistic is consistent since $F_k = O_p(T)$.
If $r_i = 0$ for a certain i, then we drop λ_0 and b_1^* from the auxiliray regression and the F statistic converges in distribution to the functional omitting the first two elements of $G_{1*}(W)$ and the first two rows and columns of $G_{2*}(W)$.

Asymptotic critical values for the F_k statistic regarding cases $k = 1, ..., 5$ computed via stochastic simulations are tabulated in Tables 2a and 2b.

For computational purposes the F_k statistic based on (11) can be easily calculated as follows: **(a)** estimate the unrestricted model (11) and keep the sum of squared residuals $SSR_U = \sum_t \hat{v}_{U,t}^2$ **(b)** estimate (11) under the restrictions implied by the null hypothesis and keep the sum of squared residuals $SSR_R = \sum_t \hat{v}_{R,t}^2$ **(c)** calculate the ratio $F_k = T\frac{SSR_R - SSR_U}{SSR_U}$ where T denotes the number of observations in the restricted regression and compare with the critical values reported in Tables 2a or 2b. This procedure facilitates comparison with the χ^2 version of the LM type statistics used in the case of stationary regressors (see van Dijk et al, 2002).

5 Small sample properties of the F test

5.1 Size simulations

We begin the analysis of the small sample properties of the F test developed above by reporting the results of Monte Carlo experiments investigating the size of the proposed test. The following random walk model was employed as a DGP:

$$\begin{aligned} y_t &= y_{t-1} + u_t,\ y_0 = 0,\ t = 1, ..., T \\ u_t &= \rho u_{t-1} + \eta_t,\ \eta_t \sim N.I.D(0,1) \end{aligned} \qquad (13)$$

We simulated series from this DGP with different parameter values $\rho = \{0.0, 0.5\}$, and computed the size of the F_k test for different values of $k = 1, 2, 3, 4$. The results are given in Tables 3a,3b for sample sizes $T = \{50, 100, 200\}$ with $50,000$ replications, and the three cases of the KSS test have been included for comparison purposes.

The F_k test resembles the familiar χ^2 test when under the null hypothesis the process is stationary. For this reason it may suffer from size problems when the number of restrictions is large and the time series is short. Indeed, from Table 3a we observe that the test is oversized for large values of k and as the sample size increases from $T = 50$ to $T = 200$ the F_k statistic turns to be conservative. Apparently, the size problems seem to reduce when the errors are autocorrelated as Table 3b shows. Still for $k = 4$ the test is oversized but in general the test size remains close to the nominal level.[10]

The small sample power simulation experiment is more demanding in its design. A similar procedure to the one reported in section 3 will be followed. Results are summarized in Tables 4a and 4b. In addition to the F_k tests for $k = 1, 2, 3, 4$ we compute rejection probabilities for the KSS t-test using raw, de-meaned and de-trended data (denoted by KSS_1, KSS_2, KSS_3 respectively).

The tests show some nontrivial power in all cases except for very small sample of $T = 50$ and highly persistent alternatives with $\gamma_1 = -0.5$. As expected, the KSS t-test is more powerful than the F_k test when $k = 1$ since it deals explicitly with one sided alternatives of stationarity and it involves estimation of a single parameter. To take a specific example, using $T = 100, \gamma_1 = -0.5, r_1 = 0$ and $\theta = 0.01$ (table 4b) the null of a unit root was correctly rejected in 69.7% of the trials by the KSS_1 test and in 24.5% of the trials by the F_1 test. However the performance of the F_1 test increases with both the sample size and the absolute value of θ.

Inspection of Tables 4a, and 4b reveals that the F_k tests have power irrespective of the number of fixed points present in the model. Thus, we cannot rely on the tests to distinguish the number of fixed points a priori. This is a consequence of the inadequacy of the Taylor approximation that offers a common polynomial structure under the alternative to be tested. Notice that for simplicity the auxiliary regression is derived from a second order Taylor approximation. If, for example $k = 2$ and we increase the Taylor expansion order then auxiliray regressions similar to the ones employed in cases $k = 3$ or $k = 4$ arise.

In general, as the number of fixed points increases the F_k tests perfom better with respect to KSS_1. For example, when $k = 3$ $(r_1 = 0, r_2 = 3, r_3 = 6)$, $T = 100, \gamma_1 = -1$ and $\theta = \frac{0.01}{80}$ (table 4b) the null of a unit root is rejected in 80.3% of the trials by F_3 and in 58.2% of the trials by KSS_1.

In addition, we observe that the KSS_2 t-ratio has increased power relative to both KSS_1 and KSS_3 as the number of fixed points increases from one to three. This is so because series created by (8) will not spend enough time around $r_1 = 0$ as the number of fixed points increases and will give the "impression" of a non-zero mean[11].

[10] When the errors in (13) are autocorrelated the lagged first differences Δy_{t-1} have been included in the right hand side of all auxiliary regressions.

[11] The exact moments of y_t generated by (8) are not known.

6 Empirical application: U.S. ex post real interest rate.

6.1 Linear and nonlinear unit root tests

We use the monthly ex-post U.S. real interest rate (y_t) for the period 1973-2005. Data for the nominal interest rate and CPI series are obtained from the IMF International Financial Statistics. We construct the ex-post real interest rate series (y_t) by subtracting the three month ahead inflation rate from the 3-month nominal bill rate ($y_t = r_t - (p_{t+3} - p_t)400$). We subject the series to the ADF test and the KSS test for a unir root against linear and 1-ESTAR globally stationary alternatives, respectively. Preliminary investigation based on the Ljung-Box statistic suggests that a unit root AR(10) model captures all autocorrelation producing residuals that are approximately white noise. Thus the maximum number of lags in the auxiliary regression

$$\Delta y_t = \sum_{i=1}^{p-1} \beta_i^{**} \Delta y_{t-i} + \delta \left(y_{t-1} \times y_{t-d}^2 \right) + error \tag{14}$$

was set to be $p-1 = 9$. The tests are based on the t-ratio of the OLS estimate of δ from the auxiliary regressions for delay lags $d = 1, 2, ..., 12$. The ADF test uses y_{t-1} instead of $y_{t-1} \times y_{t-d}^2$ in the right hand side. The final estimated auxiliary regression excludes insignificant augmentation terms Δy_{t-i}. The results appear in table 5. The KSS test has been calculated using the raw data (case 1), the demeaned data (case 2) and the detrended data (case 3). Asymptotic critical values are given by Kapetanios et al. (2003). In all cases the null hypothesis is not rejected for $d = 1$ but the KSS test rejects for higher values of d and in particular for $d = 6$. The qualitative decision of the KSS test was not altered in any of the three cases using 3, 12 or 24 lags in the auxiliary regression.

Subsequently, we apply the F_k-statistic (12) on the data. Theoretically any value of k can be employed but it seems reasonable to consider $k = 1, 2, 3, 4$. Larger powers induce near singular regressor matrices and are economically implausible. The auxiliary regression takes the form

$$\Delta y_t = \lambda_0 + \sum_{j=1}^{p-1} \beta_j^{**} \Delta y_{t-j} + b_1^* y_{t-1} + \sum_{j=2}^{2k} b_{2,j}^* y_{t-d}^j + b_3^* y_{t-1} y_{t-d}^{2k} + v_t \tag{15}$$

or

$$\Delta y_t = \sum_{j=1}^{p-1} \beta_j^{**} \Delta y_{t-j} + \sum_{j=2}^{2k} b_{2,j}^* y_{t-d}^j + b_3^* y_{t-1} y_{t-d}^{2k} + v_t \tag{16}$$

if one of the fixed points is assumed to be zero. Results appear in Tables 6a and 6b.

The tests reject the null of a unit root against $k-ESTAR$ alternatives for certain delay lag values, centered around $d = 6$. In fact most of the highest F_k values are obtained for $d = 6$. Hence in all subsequent models the delay parameter is chosen as to maximizee the value of the unit root tests and we set $z_t = y_{t-6}$.

6.2 Estimation and empirical results.

Once the transition variable $z_t = y_{t-6}$ have been selected, the next modelling stage is estimation of parameters in the $k - ESTAR$ model using NLS. The hypothesis of no ARCH in the disturbances was rejected by the standard residuals based LM tests and for this reason we tentatively assumed that the conditional variance follows a low order standard GARCH process.

$$y_t = \sum_{j=1}^{10} \beta_j y_{t-j} + \left[r_1 + \sum_{j=1}^{10} \gamma_j y_{t-j} \right] \left[1 - \exp\left\{ -a_2^2 \left((y_{t-6} - r_1)(y_{t-6} - r_2) \right)^2 \right\} \right] + u_t \tag{17}$$

The following restrictions have been imposed in the estimation $\sum_{j=1}^{10} \beta_j = 1$ and $\sum_{j=1}^{10} \gamma_j = -\sum_{j=1}^{10} \beta_j$ since they could not be rejected at the 5% significance level by the LR statistic. Estimation of (19) yielded two "equilibria" at levels $r_1 = 0$, and $r_2 = 5.75$.[12] Figure 2 displays the estimated transition function $G(.)$ against the transition variable y_{t-6}. Note that the series behaves very close to a random walk when its values are between the two "fixed" points r_1 and r_2. However, when the series is outside that "band" the speed of adjustment depends on the size of the deviation.

7 Conclusions

In this paper we have extended the popular nonlinear ESTAR model in a way that allows for multiple "fixed" points in the transition function, and we have named it the k-ESTAR model. This new feature has the potential to

[12] Please note that the p-values of the fixed point r_2 has been obtained through Monte Carlo. The fixed point $r_1 = 0$ also acts as an atractor whereas $r_2 = 5.75$ is semistable from below.

generate richer dynamics in the series than previously allowed in this type of models. In particular, it can be useful to model series that might exhibit multiple equilibria or multiple points where dynamics in their neighbourhood are complex. We develop an F-type test of the null of a unit root against a k-ESTAR alternative. Size and power of the test are analysed through simulations and it seems to outperform current nonlinear tests. We have estimated the new model for the US real interest rate data finding support for two equilibria in the series.

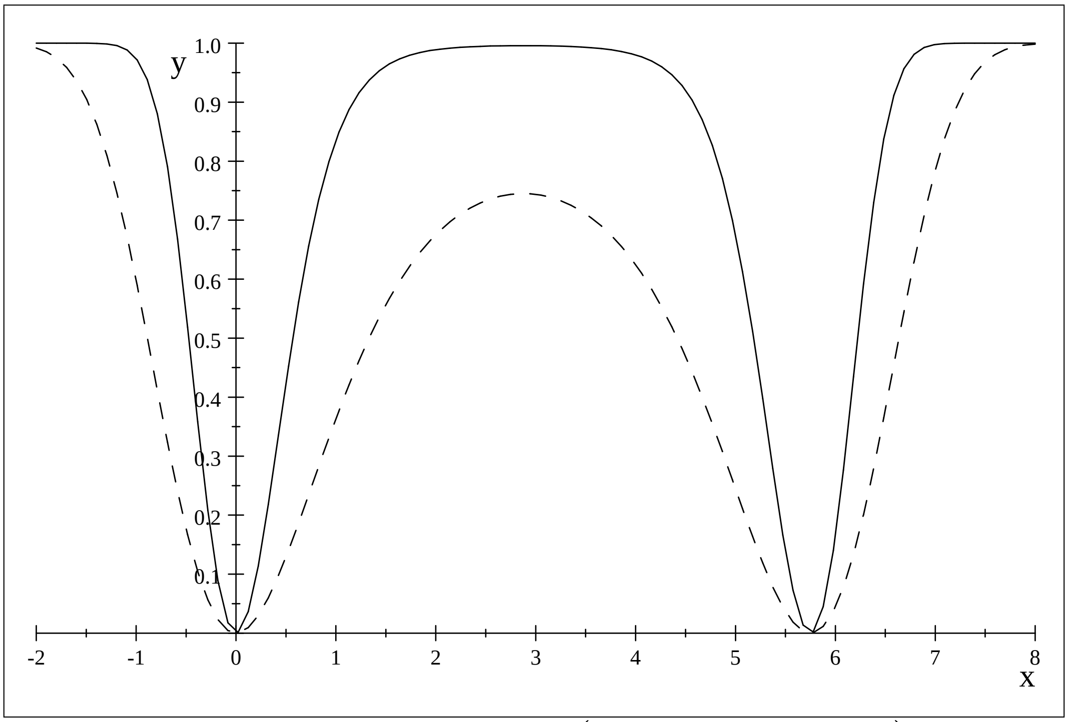

Figure 1. Transition functions $G_1 = \left(1 - e^{-0.08((x-0)(x-5.75))^2}\right)$ and $G_2 = \left(1 - e^{-0.02((x-0)(x-5.75))^2}\right)$

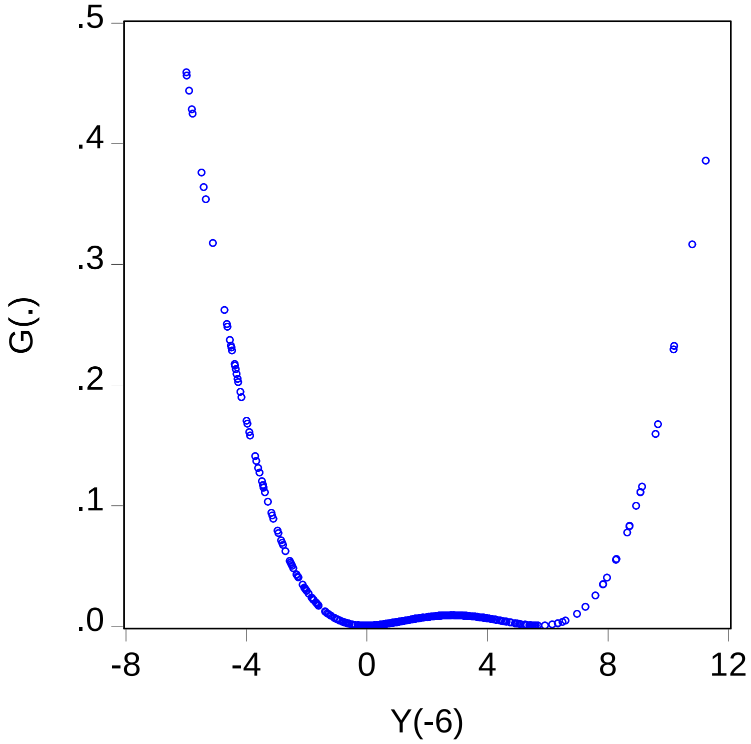

Figure 1: Figure 2. Estimated transition function $G(.)$ in (17) against transition variable y_{t-6}

.

Table 1.
Small sample ($T = 100$) power of the KSS test against $k - ESTAR$ alternatives

$\bar{\Xi}^{*}=$	0.95	0.80	0.50	0.25
$\gamma = -1.5$				
$r_1 = 0$	0.971	1	1	1
$r_1 = 0, r_2 = 3$	0.933	0.992	0.999	1
$r_1 = 0, r_2 = 3, r_3 = 6$	0.586	0.706	0.988	0.999
$\gamma = -1$				
$r_1 = 0$	0.903	1	1	1
$r_1 = 0, r_2 = 3$	0.890	0.986	0.997	0.999
$r_1 = 0, r_2 = 3, r_3 = 6$	0.545	0.608	0.942	0.998
$\gamma = -0.5$				
$r_1 = 0$	0.634	0.999	1	1
$r_1 = 0, r_2 = 3$	0.772	0.965	0.974	0.990
$r_1 = 0, r_2 = 3, r_3 = 6$	0.430	0.525	0.803	0.960
$\gamma = -0.1$				
$r_1 = 0$	0.127	0.519	0.515	0.508
$r_1 = 0, r_2 = 3$	0.282	0.465	0.484	0.487
$r_1 = 0, r_2 = 3, r_3 = 6$	0.164	0.218	0.311	0.424

Notes: To compute the rejection probabilities, the data under the alternative is generated by (8).

Table 2a.
Asymptotic critical values of F_k statistic

$Fractile(\%)$	10	5	1
$k = 1$	11.87	13.84	18.10
$k = 2$	15.44	17.63	22.12
$k = 3$	18.61	20.97	25.92
$k = 4$	20.37	22.77	27.69
$k = 5$	21.35	23.73	28.56

Note: Simulation was based on samples with size $T = 5,000$ and $50,000$ replications.

Table 2b.
Asymptotic critical values of F_k statistic when $r_i = 0$ for a certain i

$Fractile(\%)$	10	5	1
$k = 1$	8.09	9.79	13.48
$k = 2$	12.19	14.14	18.44
$k = 3$	15.70	17.92	22.44
$k = 4$	17.94	20.19	25.02
$k = 5$	19.32	21.61	26.44

Note: Simulation was based on samples with size $T = 5,000$ and $50,000$ replications.

Table 3a. The size of alternative tests

$T = 50, \rho = 0.0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
F_k	0.043	0.057	0.067	0.085
KSS case 1	0.042	0.042	0.041	0.039
KSS case 2	0.053	0.055	0.052	0.053
KSS case 3	0.058	0.058	0.057	0.059
$T = 100, \rho = 0.0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
F_k	0.036	0.038	0.040	0.048
KSS case 1	0.040	0.043	0.041	0.041
KSS case 2	0.052	0.050	0.053	0.050
KSS case 3	0.052	0.050	0.051	0.050
$T = 200, \rho = 0.0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
F_k	0.036	0.036	0.032	0.034
KSS case 1	0.043	0.044	0.042	0.044
KSS case 2	0.051	0.052	0.050	0.050
KSS case 3	0.050	0.051	0.051	0.053

Notes. Data generated by (13)

Table 3b. The size of alternative tests

$T = 50, \rho = 0.5$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
F_k	0.063	0.093	0.098	0.111
KSS case 1	0.047	0.049	0.048	0.048
KSS case 2	0.078	0.077	0.077	0.077
KSS case 3	0.097	0.101	0.100	0.098
$T = 100, \rho = 0.5$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
F_k	0.045	0.057	0.063	0.082
KSS case 1	0.044	0.044	0.043	0.045
KSS case 2	0.061	0.061	0.062	0.062
KSS case 3	0.072	0.072	0.073	0.072
$T = 200, \rho = 0.5$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
F_k	0.040	0.046	0.061	0.068
KSS case 1	0.045	0.044	0.045	0.044
KSS case 2	0.057	0.057	0.056	0.056
KSS case 3	0.062	0.062	0.063	0.061

Notes. Data generated by (13)

Table 4a. Power of alternative tests

$T=50, \gamma_1=-0.5$							
	F_1	F_2	F_3	F_4	KSS_1	KSS_2	KSS_3
$r_1=0, \theta=0.01$	0.095	0.086	0.098	0.154	0.251	0.119	0.092
$r_1=0, r_2=3, \theta=\frac{0.01}{5.2}$	0.312	0.255	0.254	0.316	0.429	0.265	0.201
$r_1=0, r_2=3, r_3=6, \theta=\frac{0.01}{80}$	0.034	0.035	0.034	0.040	0.027	0.029	0.024
$T=50, \gamma_1=-1$							
	F_1	F_2	F_3	F_4	KSS_1	KSS_2	KSS_3
$r_1=0, \theta=0.01$	0.171	0.139	0.147	0.208	0.476	0.187	0.131
$r_1=0, r_2=3, \theta=\frac{0.01}{5.2}$	0.541	0.460	0.444	0.494	0.579	0.450	0.349
$r_1=0, r_2=3, r_3=6, \theta=\frac{0.01}{80}$	0.047	0.053	0.052	0.056	0.034	0.045	0.039
$T=50, \gamma_1=-0.5$							
	F_1	F_2	F_3	F_4	KSS_1	KSS_2	KSS_3
$r_1=0, \theta=0.18$	0.650	0.498	0.450	0.514	0.939	0.641	0.436
$r_1=0, r_2=3, \theta=\frac{0.18}{4.35}$	0.589	0.495	0.501	0.557	0.708	0.595	0.416
$r_1=0, r_2=3, r_3=6, \theta=\frac{0.18}{87.5}$	0.281	0.302	0.311	0.387	0.321	0.333	0.232
$T=50, \gamma_1=-1$							
	F_1	F_2	F_3	F_4	KSS_1	KSS_2	KSS_3
$r_1=0, \theta=0.18$	0.965	0.935	0.879	0.905	0.998	0.958	0.875
$r_1=0, r_2=3, \theta=\frac{0.18}{4.35}$	0.838	0.856	0.864	0.879	0.823	0.885	0.790
$r_1=0, r_2=3, r_3=6, \theta=\frac{0.18}{87.5}$	0.435	0.572	0.621	0.676	0.362	0.615	0.479

Notes: To compute the rejection probabilities, the data under the alternative is generated by (8).

Table 4b. Power of alternative tests

$T = 100, \gamma_1 = -0.5$							
	F_1	F_2	F_3	F_4	KSS_1	KSS_2	KSS_3
$r_1 = 0, \theta = 0.01$	0.245	0.149	0.126	0.161	0.697	0.246	0.147
$r_1 = 0, r_2 = 3, \theta = \frac{0.01}{5.2}$	0.732	0.563	0.485	0.508	0.817	0.601	0.390
$r_1 = 0, r_2 = 3, r_3 = 6, \theta = \frac{0.01}{80}$	0.609	0.645	0.589	0.608	0.477	0.554	0.405
$T = 100, \gamma_1 = -1$							
	F_1	F_2	F_3	F_4	KSS_1	KSS_2	KSS_3
$r_1 = 0, \theta = 0.01$	0.540	0.316	0.255	0.295	0.930	0.495	0.286
$r_1 = 0, r_2 = 3, \theta = \frac{0.01}{5.2}$	0.906	0.834	0.776	0.779	0.916	0.847	0.687
$r_1 = 0, r_2 = 3, r_3 = 6, \theta = \frac{0.01}{80}$	0.709	0.839	0.803	0.809	0.582	0.779	0.651
$T = 100, \gamma_1 = -0.5$							
	F_1	F_2	F_3	F_4	KSS_1	KSS_2	KSS_3
$r_1 = 0, \theta = 0.18$	0.987	0.967	0.924	0.915	0.999	0.979	0.910
$r_1 = 0, r_2 = 3, \theta = \frac{0.18}{4.35}$	0.949	0.903	0.894	0.904	0.974	0.959	0.866
$r_1 = 0, r_2 = 3, r_3 = 6, \theta = \frac{0.18}{87.5}$	0.534	0.598	0.585	0.644	0.573	0.718	0.521
$T = 100, \gamma_1 = -1$							
	F_1	F_2	F_3	F_4	KSS_1	KSS_2	KSS_3
$r_1 = 0, \theta = 0.18$	1	1	0.999	0.999	1	0.999	0.999
$r_1 = 0, r_2 = 3, \theta = \frac{0.18}{4.35}$	0.990	0.996	0.998	0.998	0.991	0.997	0.990
$r_1 = 0, r_2 = 3, r_3 = 6, \theta = \frac{0.18}{87.5}$	0.671	0.896	0.920	0.944	0.661	0.902	0.819

Notes: To compute the rejection probabilities, the data under the alternative is generated by (8).

Table 5
Unit root test results for the U.S ex post real interest rate

	ADF	KSS_1	KSS_2	KSS_3
$d=1$	-2.214	-1.475	-2.149	-2.117
$d=2$		-1.325	-2.059	-2.130
$d=3$		-1.186	-1.780	-1.728
$d=4$		-1.918	-2.257	-2.149
$d=5$		-2.466^{**}	-2.907^{*}	-2.796
$d=6$		-2.776^{**}	-3.952^{***}	-3.875^{**}
$d=7$		-1.918	-3.231^{**}	-3.268^{*}
$d=8$		-0.516	-1.698	-1.677
$d=9$		-0.628	-0.908	-0.889
$d=10$		-1.101	-1.362	-1.260
$d=11$		-0.818	-0.630	-0.609
$d=12$		-1.269	-1.105	-1.106

Notes. The KSS_1, KSS_2, KSS_3 statistics are computed using the raw, demeaned and de-trended data in a regression model (14) with a maximum of nine augmentations, where the insignificant augmentation terms in a companion AR(9) model for Δy_t were excluded. The ADF statistic is based on demeaned data. d denotes delay lag. In all cases $*$, $**$ and $***$ denote significance at 10% , 5% and 1% level.

Table 6a. F_k statistic results

	$k=1$	$k=2$	$k=3$	$k=4$
$d=1$	6.710	14.304	17.116	18.866
$d=2$	7.309	8.883	12.689	14.310
$d=3$	5.177	10.380	11.385	33.201***
$d=4$	5.936	7.150	22.429**	25.688**
$d=5$	7.717	22.029**	22.315**	24.440**
$d=6$	12.607*	25.911***	26.884***	31.998***
$d=7$	9.954	11.679	24.670**	28.685***
$d=8$	6.430	9.135	18.382	19.774
$d=9$	6.457	9.189	10.613	11.141
$d=10$	6.507	7.069	9.351	18.547
$d=11$	8.838	10.682	10.144	12.069
$d=12$	6.644	9.725	10.796	16.751

Notes. Results of the F_k TEST statistic applied to model (15) using U.S ex post real interest rates. d denotes delay lag and k the number of fixed points in the $k-ESTAR$ model. In all cases $*$, $**$ and $***$ denote significance at 10% , 5% and 1% level.

Table 6b. F_k statistic results

	$k=1$	$k=2$	$k=3$	$k=4$
$d=1$	4.741	12.198*	16.337*	17.882
$d=2$	4.752	5.228	5.850	11.982
$d=3$	1.841	8.828	10.862	14.451
$d=4$	3.735	5.021	7.119	9.666
$d=5$	6.395	9.562	10.691	14.226
$d=6$	12.405**	14.709**	15.128	20.361**
$d=7$	8.880*	9.615	10.606	18.365*
$d=8$	1.241	6.343	6.861	7.616
$d=9$	0.561	2.138	6.652	8.002
$d=10$	1.546	4.159	5.276	6.181
$d=11$	2.159	4.148	6.729	7.468
$d=12$	1.878	7.268	9.022	12.927

Notes. Results of the F_k TEST statistic applied to model (16) using U.S ex post real interest rates. d denotes delay lag and k the number of fixed points in the $k-ESTAR$ model. In all cases $*$ and $**$ denote significance at 10% and 5% level.

APPENDIX

A) Proof of proposition 1.

Given a polynomial $\beta(z)$ of order p there exists an equivalent representation $\beta(z) = \beta(1)z + \beta^{**}(z)(1-z)$ where $\beta^{**}(z) = \beta^{*}(z) + \beta(1)$ is a polynomial of order $p-1$ and the coefficients of polynomial $\beta^{*}(z)$ are given from $\beta_j^{*} = -\sum\limits_{k=j+1}^{p} \beta_k$, $j = 0, ..., p-1$. Using this representation, model (1) is re-written as

$$\begin{aligned} \Delta y_t &= \beta_0 - \beta(1) y_{t-1} + \sum_{j=1}^{p-1} \beta_j^{**} \Delta y_{t-j} \\ &\quad + \left[\gamma_0 + \sum_{j=1}^{p} \gamma_j y_{t-j} \right] G(\alpha_k, \mathbf{r}; y_{t-d}) + u_t \end{aligned} \tag{18}$$

The second order Taylor series approximation of $G(\alpha_k) = G(\alpha_k, .; .)$ around $\alpha_k = 0$ is

$$G(\alpha_k) = G(0) + \left. \frac{\partial G}{\partial \alpha_k} \right|_{a_k=0} \alpha_k + \frac{1}{2} \left. \frac{\partial^2 G}{\partial \alpha_k^2} \right|_{a_k=0} \alpha_k^2 + R \Leftrightarrow$$

$$G(\alpha_k) = \alpha_k^2 \prod_{i=1}^{k} (y_{t-d} - r_i)^2 + R$$

since $G(0) = 0$ and $\left. \frac{\partial G}{\partial \alpha_k} \right|_{a_k=0} = 0$ (with R the remainder). Under the unit root assumption, $\beta(1) = 0$, we obtain

$$\begin{aligned} \Delta y_t &= \beta_0 + \sum_{j=1}^{p-1} \beta_j^{**} \Delta y_{t-j} \\ &\quad + a_k^2 \left[\gamma_0 + \sum_{j=1}^{p} \gamma_j y_{t-j} \right] \prod_{i=1}^{k} (y_{t-d} - r_i)^2 + e_t \end{aligned} \tag{19}$$

where $e_t = \left[\gamma_0 + \sum\limits_{j=1}^{p} \gamma_j y_{t-j} \right] R + u_t$.

Thus the null hypothesis of a unit root process against the globally stationary process generated by (1) is equivalent to testing

$$H_0 : \alpha_k^2 = 0 \tag{20}$$

in (19). Under the null hypothesis $e_t = u_t$ an F-type test can be constructed. However, it is clear that the approach results in overfitting even for moderate autoregressive polynomial orders p (assuming a reasonable value of k). The general auxiliary regression through which (20) will be tested can be written as

$$\begin{aligned}\Delta y_t &= \beta_0 + \sum_{j=1}^{p-1} \beta_j^{**} \Delta y_{t-j} \\ &+ a_k^2 \times \left[\gamma_0 \delta_0 + \delta_0 \sum_{j=1}^{p} \gamma_j y_{t-j} \right. \\ &\left. + \gamma_0 \sum_{j=1}^{2k} \delta_j y_{t-d}^j + \sum_{j=1}^{p} \sum_{s=1}^{2k} \gamma_j \delta_s y_{t-j} y_{t-d}^s \right] \\ &+ e_t\end{aligned}$$

where we have set $\prod_{i=1}^{k} (y_{t-d} - r_i)^2 = \delta_0 + \sum_{s=1}^{2k} \delta_s y_{t-d}^s$ with parameters δ_s being functions of the location parameters r_i. In particular, $\delta_0 = \prod_{i=1}^{k} r_i^2$ and $\delta_{2k} = 1$. In addition $\delta_s = 0$ for $s = 0, ..., 2k-1$ if $\mathbf{r} = \mathbf{0}$. Finally, if some $r_i = 0$ then $\delta_0 = 0$ and the auxiliary regression becomes

$$\Delta y_t = \beta_0 + \sum_{j=1}^{p-1} \beta_j^{**} \Delta y_{t-j} \tag{21}$$

$$+ a_k^2 \left[\gamma_0 \sum_{j=2}^{2k} \delta_j y_{t-d}^j + \sum_{j=1}^{p} \sum_{s=2}^{2k} \gamma_j \delta_s y_{t-j} y_{t-d}^s \right] + e_t$$

Note, that if we set

$$p = 1, k = 1, d = 1, \beta_0 = 0, \gamma_0 = 0, r_1 = 0$$

as in the 1-ESTAR model of KSSa, then we obtain

$$\Delta y_t = a_1^2 \gamma_1 y_{t-1}^3 + e_t$$

and the asymptotic stationarity conditions imply a test of $a_1^2 \gamma_1 = 0$ versus $a_1^2 \gamma_1 < 0$.

B) Simplifying auxiliary regression.

In subsequent analysis we always set $z_t = y_{t-d}$ where $d \leq p$ or $d > p$. In addition we set $\beta_0 = 0$ assuming random walk behavior without drift when $y_{t-d} = r_i$. We write (9) as a partitioned regression model,

$$\Delta Y = X_1 b_1 + X_2 b_2 + error \tag{22}$$

with X_1 the data matrix including the first $p-1$ regressors on the right hand side of (9),

$$(\Delta y_{t-1}, \Delta y_{t-2}, ..., \Delta y_{t-p+1})$$

that are stationary under the null hypothesis, while X_2 includes the $(p+1)(2k+1)$ (if $d > p$) or $p(2k+1)+1$ (if $d \leq p$) remaining regressors

$$\begin{aligned}
&(1, y_{t-1}, y_{t-2}, ..., y_{t-p},\\
&z_t, z_t^2, ..., z_t^{2k},\\
&y_{t-1}z_t, y_{t-1}z_t^2, ..., y_{t-1}z_t^{2k},\\
&\vdots\\
&y_{t-p}z_t, y_{t-p}z_t^2, ..., y_{t-p}z_t^{2k})
\end{aligned}$$

The set is modified accordingly by adding a column of ones if there is a constant in the auxiliary regression. Let $M_1 = I - X_1 (X_1'X_1)^{-1} X_1'$ be the orthogonal to X_1 projection matrix. The above presentation aims to conveniently expose the X_1 and X_2 data matrices structure. The proof of the following proposition shown in the appendix suggests that not all regressors in matrix X_2 are neccessary for the testing procedure. For example, under the null of non-stationarity and for finite orders p the regressors $y_{t-1}, y_{t-2}, ..., y_{t-p}$ are collinear asymptotically. The same conclusion is reached for regressors involving powers of the transition variable when $z_t = y_{t-d}$ and the cross-products $y_{t-j}z_t^s$.

Thus, we re-specify the auxiliary regression model into

$$\Delta Y = X_1 b_1 + X_2^* b_2^* + v \tag{23}$$

where X_2^* includes regressors $1, y_{t-1}, y_{t-d}^2, ..., y_{t-d}^{2k}, y_{t-1}y_{t-d}^{2k}$ while $v = u$ under the null hypothesis.

C) Proof of proposition 2.

Under the null hypothesis,

$$y_t = y_{t-1} + \eta_t \tag{24}$$

where the initial condition is set to $y_0 = 0$ although it may be any $O_p(1)$ random variable. The errors satisfy $\eta_t = \varphi(L)u_t = \sum_{j=0}^{+\infty} \varphi_j u_{t-j}$ where $\varphi(L) = \beta^{**^{-1}}(L)$ with $\varphi(1) \neq 0$ and $\sum_{j=0}^{+\infty} j \left|\varphi_j\right| < +\infty$ while u_t is a stationary and ergodic martingale difference sequence with variance σ_u^2. Then the following invariance principle holds

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \eta_t \overset{d}{\Rightarrow} B(r) = \sigma W(r)$$

where $\stackrel{d}{\Rightarrow}$ denotes weak convergence, $W(r)$ is standard Brownian motion $r \in [0,1]$ and $\sigma^2 = \sum_{j=-\infty}^{+\infty} E(\eta_0\eta_j) = \sigma_\eta^2 + 2\lambda_\eta$ is the long run variance of η_t with $\lambda_\eta = \sum_{j=1}^{+\infty} E(\eta_0\eta_j)$. Also let $\delta_\eta = \sigma_\eta^2 + \lambda_\eta$ denote the "one-sided" long run covariance of η_t. For brevity we will write $\int_0^1 B^k(r)dr$ as $\int_0^1 B^k$ and $\int_0^1 B^k(r)dB(r)$ as $\int_0^1 B^k dB$.

Using Hong and Phillips (2005) results, for k a positive integer, we have

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t}{\sqrt{T}}\right)^k = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_{t-1}}{\sqrt{T}}\right)^k + o_p(1) \rightarrow_d \int_0^1 B^k \tag{25}$$

$$\begin{aligned}\sum_{t=1}^{T}\left(\frac{y_t}{\sqrt{T}}\right)^k \frac{u_t}{\sqrt{T}} &\rightarrow {}_d\sigma_u\int_0^1 B^k dW + k\sigma_u^2\int_0^1 B^{k-1} \\ &= \sigma_u^{k+1}\varphi^k(1)\int_0^1 W^k dW + k\sigma_u^{2k-2}\varphi^{k-1}(1)\int_0^1 W^{k-1}\end{aligned} \tag{26}$$

and

$$\sum_{t=1}^{T}\left(\frac{y_{t-1}}{\sqrt{T}}\right)^k \frac{u_t}{\sqrt{T}} \rightarrow_d \sigma_u\int_0^1 B^k dW = \sigma_u^{k+1}\varphi^k(1)\int_0^1 W^k dW \tag{27}$$

In addition, for k_1, k_2 integers, we can substitute $y_{t-p} = y_{t-d} + \sum_{j=0}^{d-p-1} \Delta y_{t-p-j}$ for $d > p$ and use the binomial expansion and results (25), (26) to show that

$$\begin{aligned}&\frac{1}{T}\sum_{t=d+1}^{T}\left(\frac{y_{t-p}}{\sqrt{T}}\right)^{k_1}\left(\frac{y_{t-d}}{\sqrt{T}}\right)^{k_2} \\ &= \frac{1}{T^{1+\frac{k_1}{2}+\frac{k_2}{2}}}\sum_{t=d+1}^{T}\left(y_{t-d} + \sum_{j=0}^{d-p-1}\Delta y_{t-p-j}\right)^{k_1} y_{t-d}^{k_2} \\ &= \frac{1}{T}\sum_{t=d+1}^{T}\left(\frac{y_{t-d}}{\sqrt{T}}\right)^{k_1+k_2} + \frac{1}{T^{1+\frac{k_1}{2}+\frac{k_2}{2}}}\sum_{t=d+1}^{T}\left\{\sum_{s=1}^{k_1}\binom{k_1}{s} y_{t-d}^{k_1+k_2-s}\left(\sum_{j=0}^{d-p-1}\Delta y_{t-p-j}\right)^s\right\} \\ &= \frac{1}{T}\sum_{t=d+1}^{T}\left(\frac{y_{t-d}}{\sqrt{T}}\right)^{k_1+k_2} + o_p(1)\end{aligned} \tag{28}$$

thus

$$\frac{1}{T}\sum_{t=d+1}^{T}\left(\frac{y_{t-p}}{\sqrt{T}}\right)^{k_1}\left(\frac{y_{t-d}}{\sqrt{T}}\right)^{k_2} \rightarrow_d \int_0^1 B^{k_1+k_2} \tag{29}$$

The above generalizes to sample moments with more than two product terms.

In addition the cross product terms satisfy,

$$\begin{aligned}
\sum_{t=p+1}^{T}\left(\frac{y_{t-p}}{\sqrt{T}}\right)^{k}\frac{u_t}{\sqrt{T}} &= \frac{1}{T^{(k+1)/2}}\sum_{t=p+1}^{T} y_{t-p}^{k}u_t \\
&= \frac{1}{T^{(k+1)/2}}\sum_{t=p+1}^{T} y_{t-1}^{k}u_t \\
&+ \frac{1}{T^{(k+1)/2}}\sum_{t=p+1}^{T}\left\{\sum_{s=1}^{k}(-1)^{s}\binom{k}{s}y_{t-1}^{k-s}\left(\sum_{j=0}^{p-2}\Delta y_{t-j-1}\right)^{s}u_t\right\} \\
&= \frac{1}{T^{(k+1)/2}}\sum_{t=p+1}^{T} y_{t-1}^{k}u_t + o_p(1) \\
&\to_d \sigma_u\int_0^1 B^k dW
\end{aligned}$$

Hence we can consider the asymptotic behavior of the F-type statistic

$$F = \frac{1}{\hat{\sigma}_u^2}\left(\hat{b}_2 - b_2\right)'\left(X_2'M_1X_2\right)\left(\hat{b}_2 - b_2\right)$$

testing the null hypothesis $H_0 : Rb = c$ in (22) where $R = \left[\begin{array}{cc}\mathbf{0} & \mathbf{I}\end{array}\right]$, $c = \mathbf{0}$ and $b = \left(\begin{array}{cc} b_1 & b_2\end{array}\right)'$. The sampling error of $\hat{b}_2$ is given by the known formula,

$$\left(\hat{b}_2 - b_2\right) = \left(X_2'M_1X_2\right)^{-1}X_2'M_1u$$

where

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

and

$$X_2'M_1u = X_2'u - X_2'X_1\left(X_1'X_1\right)^{-1}X_1'u$$

Matrices $X_1'X_1$ and $X_1'u$ involve sums of ergodic and stationary series thus $\frac{1}{T}X_1'X_1 = O_p(1)$ and $\frac{1}{T}X_1'u = o_p(1)$.

Given the results in F1-F5, define the $p(2k+1)+1 \times p(2k+1)+1$ (case[13] $d \leq p$) normalization matrix D_T as

$$\begin{aligned}
D_T &= diag\Bigg(T^{-1/2}, \underbrace{T^{-1}, ..., T^{-1}}_{p\ times}, \\
&\underbrace{\left[T^{-3/2}, T^{-2}, ..., T^{-(k+\frac{1}{2})}, T^{-(k+1)}\right], ..., \left[T^{-3/2}, T^{-2}, ..., T^{-(k+\frac{1}{2})}, T^{-(k+1)}\right]}_{p\ times}\Bigg)
\end{aligned}$$

[13] We chose the case $d \leq p$ for simplification purposes. When $d > p$ the normalization matrix D_T is defined accordingly with dimensions $(2k+1)(p+1) \times (2k+1)(p+1)$

Then, under the null hypothesis,

$$
\begin{aligned}
D_T X_2' X_2 D_T &\rightarrow {}_d G_2(B) \\
D_T X_2' X_1 &= O_p(1)
\end{aligned}
$$

and

$$D_T X_2' u \rightarrow_d G_1(B)$$

where functionals $G_1(B)$ and $G_2(B)$ are given by

$$
G_1(B) = \sigma_u \left(W(1) \quad \underbrace{\int BdW \quad \cdots \quad \int BdW}_{p\ times} \quad \underbrace{\int B^2 dW \quad \int B^3 dW \quad \cdots \quad \int B^{2k+1} dW \quad \cdots \quad \int B^2 dW \quad \int B^3 dW \quad \cdots \quad \int B^{2k+1} dW}_{p\ times} \right)'
$$

and

$$
G_2(B) = \begin{bmatrix}
1 & \int B & \cdots & \int B & \int B^2 & \cdots & \int B^{2k+1} & \cdots & \int B^2 & \cdots & \int B^{2k+1} \\
\int B & \int B^2 & \cdots & \int B^2 & \int B^3 & \cdots & \int B^{2k+2} & \cdots & \int B^3 & \cdots & \int B^{2k+2} \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\
\int B & \int B^2 & \cdots & \int B^2 & \int B^3 & \cdots & \int B^{2k+2} & \cdots & \int B^3 & \cdots & \int B^{2k+2} \\
\int B^2 & \int B^3 & \cdots & \int B^3 & \int B^4 & \cdots & \int B^{2k+3} & \cdots & \int B^4 & \cdots & \int B^{2k+3} \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\
\int B^{2k+1} & \int B^{2k+2} & \cdots & \int B^{2k+2} & \int B^{2k+3} & \cdots & \int B^{4k+2} & \cdots & \int B^{2k+3} & \cdots & \int B^{4k+2} \\
\int B^2 & \int B^3 & \cdots & \int B^3 & \int B^4 & \cdots & \int B^{2k+3} & \cdots & \int B^4 & \cdots & \int B^{2k+3} \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\
\int B^{2k+1} & \int B^{2k+2} & \cdots & \int B^{2k+2} & \int B^{2k+3} & \cdots & \int B^{4k+2} & \cdots & \int B^{2k+3} & \cdots & \int B^{4k+2} \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\
\int B^2 & \int B^3 & \cdots & \int B^3 & \int B^4 & \cdots & \int B^{2k+3} & \cdots & \int B^4 & \cdots & \int B^{2k+3} \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\
\int B^{2k+1} & \int B^{2k+2} & \cdots & \int B^{2k+2} & \int B^{2k+3} & \cdots & \int B^{4k+2} & \cdots & \int B^{2k+3} & \cdots & \int B^{4k+2}
\end{bmatrix}
$$

Clearly, $G_2(B)$ is singular, hence asymptotically, under the null hypothesis of non-stationarity, some of the regressors are collinear carrying the same information.

In order to overcome this difficulty, we re-specify the auxiliary regression model (22) into

$$\Delta Y = X_1 b_1 + X_2^* b_2^* + v \tag{30}$$

where X_2^* includes regressors $1, y_{t-1}, y_{t-d}^2, ..., y_{t-d}^{2k}, y_{t-1}y_{t-d}^{2k}$ while $v = u$ under the null hypothesis. Under the alternative, (30) is a mispecified regression with $v = X_2^{**}b_2^{**} + u$ and X_2^{**} a data matrix including regressors other than $1, y_{t-1}, y_{t-d}^2, ..., y_{t-d}^{2k}, y_{t-1}y_{t-d}^{2k}$.

Based on our previous analysis it is seen that under the null, the $F-type$ statistic

$$F_k = \frac{1}{\hat{\sigma}_v^2}\left(\hat{b}_2^* - b_2^*\right)'\left(X_2^{*\prime}M_1X_2^*\right)\left(\hat{b}_2^* - b_2^*\right) \xrightarrow{d} G_{1*}'(W)G_{2*}^{-1}(W)G_{1*}(W)$$

where

$$G_{1*}(W) = \begin{pmatrix} W(1) & \int WdW & \int W^2dW & \int W^3dW & ... & \int W^{2k+1}dW \end{pmatrix}'$$

and

$$G_{2*}(W) = \begin{bmatrix} 1 & \int W & \int W^2 & & \int W^{2k+1} \\ \int W & \int W^2 & \int W^3 & \cdots & \int W^{2k+2} \\ \int W^2 & \int W^3 & \int W^4 & & \int W^{2k+3} \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ \int W^{2k+1} & \int W^{2k+2} & \cdots & \cdots & \int W^{4k+2} \end{bmatrix}$$

Under the alternative, y_t is asymptotically stationary, hence

$$\begin{aligned} \left(\hat{b}_2^* - b_2^*\right) &= \left(X_2^{*\prime}M_1X_2^*\right)^{-1}X_2^{*\prime}M_1X_2^{**}b_2^{**} + \left(X_2^{*\prime}M_1X_2^*\right)^{-1}X_2^{*\prime}M_1u \\ &= O_p(1) + O_p(T^{-1/2}) = O_p(1) \end{aligned}$$

and

$$\left(X_2^{*\prime}M_1X_2^*\right) = O_p(T)$$

As a result, $F_k = O_p(T)$ and the test statistic is consistent.

References

Bair, J., Haesbroeck, G., (1997). Monotonous stability for neutral fixed points. Bull. Belg. Math. Soc., 4, 639-646

Bec F, Ben Salem M, Collard F. 2000. Nonlinear Economic Policies: Pitfalls in the Lucas Critique Empirical Counterpart. Paper No 1401 in *Econometric Society World Congress 2000* Contributed Papers.

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

Bhattacharya, R., Lee, C., (1995). On geometric ergodicity of nonlinear autoregressive models. Statistics & Probability Letters, 22, 311-315

Cagan, P. (1956). The Monetary Dynamics of Hyperinflation, in Milton Friedman ed., Studies in the Quantity Theory of Money, Chicago: University of Chicago Press.

Christiano, L.J., Eichembaum, M., and C.L. Evans (1996). The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds. Review of Economics and Statistics 78, 16-34.

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

Diamond, P.A. (1982). Aggregate Demand Management in Search Equilibrium, Journal of Political Economy 90, 881–894.

Doornik, J.A. (2002), Object-Oriented Matrix Programming Using Ox, 3rd ed. London: Timberlake Consultants Press and Oxford: www. nuff. ox. ac. uk/ Users /Doornik.

Dumas, Bernard. (1992) “Dynamic Equilibrium and the Real Exchange Rate in a Spatially Separated World.” *Review of Financial Studies,* 5, 153-180.

Elliot, G., T.J., Rothenberg, and J.H. Stock (1996) "Efficient tests for an autoregressive unit root", Econometrica, 64, 813-836.

Eklund, B., (2003). A nonlinear alternative to the unit root hypothesis. Stockholm School of Economics, SSE/EFI Working Paper Series in Economics and Finance, No. 547, November 2003

Granger, C. W. J. and T. Terasvirta (1993). Modelling Nonlinear Economic Relationships. Oxford: Oxford University Press.

Haggan, V., Ozaki, T., (1981). Modelling nonlinear random vibrations using amplitude-dependent autoregressive time series model, Biometrica, 68, 189-196

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

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992) "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" Journal of Econometrics, 54, 159-178.

Layard, R., S.J.Nickell, and R. Jackman. Unemployment: macroeconomic performance and the labour market.OxfordUniversity Press.

Liebscher, E., (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. Journal of Time Series Analysis, Vol.26, No.5, 669-689

Luukkonen, R., P. Saikkonen and T. Terasvirta (1988). Testing linearity against smooth transition autoregressive models. Biometrika 75, 4919.

Micheal, P., R.A. Nobay and D. A. Peel (1997). Transactions costs and nonlinear adjustment in real exchange rates: An empirical investigation. Journal of Political Economy 105, 86279.

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

Ng, S. and P. Perron (2001) "Lag length selection and the construction of unit root tests with good size and power", Econometrica, 69, 1519-1554.

Orphanides A, Wilcox, DW. 1996. The Opportunistic Approach to Disinflation. No 96-24 working paper in Finance and Economics Discussion Series from *Board of Governors of the Federal Reserve System, U.S.*

Paya, I., I. A. Venetis, and D. A. Peel. (2003) “Further Evidence on PPP Adjusment Speeds: The Case of Effective Real Exchange Rates and the EMS.” *Oxford Bulletin of Economics and Statistics,* 65, 421-438.

Paya, I., and D. A. Peel (2006). A new analysis of the determinants of the real dollar-sterling exchange rate: 1871-1994. Journal of Money, Credit and Banking 38, 1971-1990.

Sargent, T.J. and N. Wallace (1973). Rational expectations and the dynamics of hyperinflation. International Economics Review 14, 328-350.

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

Sercu, Piet, Raman Uppal, and C. Van Hull. (1995) "The Exchange Rate in the Presence of Transactions Costs: Implications for Tests of Purchasing Power Parity." *Journal of Finance* 50, 1309-1319.

Taylor, M. P., D. A. Peel and L. Sarno (2001). Non-linear in real exchange rates: Towards a solution of the purchasing power parity puzzles. International Economic Review 42, 1015-42.

Terasvirta, T. (1994). Specification, estimation and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association 89, 20818.

Terasvirta, T. and A. Eliasson (2001) "Non-linear error correction and the UK demand for broad money, 1878-1993," *Journal of Applied Econometrics* 16, pp. 277-288.

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