Edinburgh School of Economics

Discussion Paper Series

Number 104

Unit Root Tests in Three-Regime SETAR Models

George Kapetanios (Queen Mary, University of London)
Yongcheol Shin (University of Edinburgh)

Date
November 2003

Published by
School of Economics
University of Edinburgh
30 -31 Buccleuch Place
Edinburgh EH8 9JT
+44 (0)131 650 8361
http://www.ed.ac.uk/schools-departments/economics

THE UNIVERSITY *of* EDINBURGH

Unit Root Tests in Three-Regime SETAR Models*

George Kapetanios
Department of Economics, Queen Mary, University of London

Yongcheol Shin
School of Economics, University of Edinburgh

This Version, November 2003

Abstract

This paper proposes a simple testing procedure to distinguish a unit root process from a globally stationary three-regime self-exciting threshold autoregressive process. Following the threshold cointegration literature we assume that the process follows the random walk in the corridor regime, and therefore we propose that the null of a unit root be tested by the Wald statistic for the joint significance of autoregressive parameters in both lower and upper regimes. We establish that when threshold parameters are known, the suggested Wald test has a well-defined asymptotic null distribution free of nuisance parameters. In the general case where threshold parameters are unknown *a priori*, we consider the three most commonly used summary statistics - average, exponential average and supremum. Assuming that the grid set for thresholds can be selected such that the corridor regime be of finite width both under the null and under the alternative, we can establish both stochastic equicontinuity and uniform convergence of the aforementioned summary statistics. Monte Carlo evidence clearly indicates that the proposed tests are more powerful than the Dickey-Fuller test that ignores the threshold nature under the alternative. We illustrate the usefulness of our proposed tests by examining stationarity of bilateral real exchange rates for the G7 countries.

JEL Classification: C12, C13, C32.
Key Words: Self-exciting Threshold Autoregressive Models, Unit Roots, Globally Stationary Processes, Threshold Cointegration, Monte Carlo Simulations, Real Exchange Rates, Transactions Costs, Dread of Depreciation.

*We are grateful to Frederique Bec, In Choi, James Davidson, Juan Dolado, Jayasri Dutta, Emmanuel Guerre, Inmoo Kim, Bruce Hansen, Taewhan Kim, Gary Koop, Stephen Leybourne, Paul Newbold, Peter Phillips, Pentti Saikkonen, Myunghwan Seo, Andy Snell, Rob Taylor and the seminar participants at the Econometric Study Group, Bristol 2001, Sungkunkwan University, University of Edinburgh, University of Nottingham, and the 57th ESEM Conference, Venice, Italy, 2002 for their helpful comments on an earlier version of this paper. Partial financial support from the ESRC (grant No. R000223399) is also gratefully acknowledged. The usual disclaimer applies.

1 Introduction

There has been increasing concern in econometrics that the information revealed by the analysis of a linear model in a single time series may be insufficient to give definitive inference on important hypotheses. In particular, the power of tests such as the Dickey-Fuller (1979, hereafter DF) unit root test or the Engle-Granger (1987) test for cointegration has been called into question. At the same time the stability of estimated parameters over the sorts of time horizons required to invoke the guidance of asymptotics in linear models has also come under suspicion. As a response to these problems, attention is turning to nonlinear dynamics to improve estimation and inference. Theoretical models of nonlinear adjustments have been proposed earlier by Hicks (1950) and others in the context of business cycle analysis. Also in the context of asset markets, the extent of arbitrage trading in response to return differentials is limited by the level of transaction costs. These costs may lead to a nonlinear relationship between the level of arbitrage activity and the size of the return differentials. Therefore, the level of arbitrage trading and hence the speed with which the returns differential reverts towards zero are an increasing function of the size of the returns differential itself. Sercu *et al.* (1995) and Michael *et al.* (1997) develop the theory suggesting that the larger the deviation from the purchasing power parity (PPP), the stronger the tendency for real exchange rates to move back to equilibrium. Some progress has already been made in other respects as well and now the applied macro time-series literature abounds with cases where departing from linearity has yielded significant gains in both prediction and inference. See for example Koop *et al.* (1996), Pesaran and Potter (1997), Kapetanios (1999), Taylor (2001), Ioannides *et al.* (2003) and Kapetanios *et al.* (2003a,b).

In particular, Balke and Fomby (1997) have popularised a joint analysis of nonstationarity and nonlinearity in the context of threshold cointegration. The threshold cointegrating process is defined as a globally stationary process such that it might follow a unit root in the middle regime, but it is dampened in outer regimes. Importantly, they have shown via Monte Carlo experiments that the power of the DF unit root tests falls dramatically with threshold parameters of a three-regime threshold autoregressive (TAR) model. Since then, there have been a few studies to address the joint issues of nonstationarity and nonlinearity, but mostly using univariate two regime TAR models. The first line of research follows the self-exciting TAR (SETAR) modelling approach where the lagged dependent variable is used as the transition variable. Enders and Granger (1998) have proposed an F-test for the null hypothesis of a unit root against an alternative of a stationary two-regime TAR process. Contrary to expectations, however, their simulation results show that the suggested F test is less powerful than the DF test that ignores the threshold nature under the alternative. See Berben and van Dijk (1999) for an extension.

There has also been an alternative line of studies in a two-regime TAR model. Caner and Hansen (2001) have considered tests for threshold nonlinearity when the underlying univariate process follows a unit root, then developed unit root tests when threshold nonlinearity is either present or absent. See also Gonzalez and Gonzalo (1998) and Shin and Lee (2001). This approach is critically different from the aforementioned SETAR-based approach; it allows only for the case where transition variables are stationary. Thus, the possibility of using the lagged dependent variable as the transition variable is excluded since it becomes nonstationary under the null. In this regard, this approach might be of reduced interest in the current case where we wish to analyse the global stationarity of the underlying long-run relationships such as PPP.

To bridge the two areas of nonstationarity and nonlinearity in the context of the threshold cointegration, we consider a more general three regime SETAR model. Unlike the previous studies using the two step-based approach proposed by Balke and Fomby (1997) and Lo and Zivot (2001), this paper aims to provide a direct test that would have more power against the alternative of globally stationary three regime SETAR processes. For further economic and econometric backgrounds in favor of three regime TAR models see Anderson (1997), Bec *et al.* (2001), Taylor (2001), Dutta and Leon (2002) and Bec *et al.* (2002).

Following the threshold cointegration literature, we impose that the processes follow a unit root in the corridor regime and thus propose that the null hypothesis of a (linear) unit root against the alternative of globally stationary three regime SETAR process be tested by the Wald test for the joint significance of autoregressive parameters under both lower and upper regimes. We establish that when threshold parameters are known, the suggested Wald test does have a well-defined asymptotic null distribution free of nuisance parameters. Moreover, in the special case where autoregressive parameters under both outer regimes are symmetric, the null of a unit root can now be tested by the Wald test for the significance of the common autoregressive parameter, and its asymptotic null distribution is shown to be equivalent to the distribution of the squared DF t-statistic. In the general case where threshold parameters are unknown *a priori*, this kind of test suffers from the well-known Davies (1987) problem since threshold parameters are not identified under the null. Following Hansen (1996), we consider the three most commonly used summary statistics - average, exponential average and supremum - over a grid set of possible threshold values. Under the maintained assumption that the corridor regime will be inactive, we do not need to estimate any parameter in the middle regime. This observation leads us to assume that the grid set for unknown thresholds can be selected such that the corridor regime be of finite width both under the null and under the alternative. Given this assumption, we can establish first stochastic equicontinuity of the Wald statistic and then uniform convergence of the aforementioned summary statistics.

The small sample performance of our suggested tests is compared to that of the DF test via Monte Carlo experiments. We find that both average and exponential average tests have reasonably correct size, but the supremum test tends to display significant size distortions in smal samples. As expected, both average and exponential average tests eventually dominate the power of the DF test as the threshold band widens. We illustrate the usefulness of our proposed tests by examining the stationarity of bilateral real exchange rates for the G7 countries (excluding France). In sum, our proposed (asymmetry) Wald tests reject the null three times out of five cases, while the DF test rejects the null only once. Given some support for the SETAR alternative, we estimate SETAR models and find that autoregressive parameters in outer regimes are significantly negative in all cases except for Canada. More interestingly, the speed of mean or range reversion is faster in the lower regime (depreciation) than in the upper regime (appreciation) for Germany, Italy and Japan. This raises further issue that empirical evidence may be available to bolster the hypothesis of asymmetric foreign exchange market interventions that countries may choose to resist depreciations more vigorously than appreciations, so-called "dread of depreciation," *e.g.* Dutta and Leon (2002).

The plan of the paper is as follows: Section 2 describes globally stationary three-regime TAR processes. Section 3 develops the Wald statistic that directly tests the null of unit root against the alternative of globally stationary three-regime SETAR processes, and presents the asymptotic theory. Section 4 investigates the small sample performance of the suggested test. Section 5 presents an empirical application to real exchange rates for the G7 countries. Section 6 concludes with further discussions. The appendix contains mathematical proofs.

2 Globally Stationary Three Regime Threshold Autoregressive Processes

Suppose that a univariate series y_t follows the three-regime self-exciting threshold autoregressive (SETAR) model:

$$y_t = \left\{ \begin{array}{lc} \phi_1 y_{t-1} + u_t & if\ y_{t-1} \le r_1 \\ \phi_0 y_{t-1} + u_t & if\ r_1 < y_{t-1} \le r_2 \\ \phi_2 y_{t-1} + u_t & if\ y_{t-1} > r_2 \end{array} \right\},\ t = 1, 2, ..., T, \tag{2.1}$$

where u_t is assumed to follow an *iid* sequence with zero mean, constant variance σ_u^2 and finite $4+\delta$ moments for some $\delta > 0$, r_1 and r_2 are threshold parameters and $r_1 < r_2$. Here, the lagged dependent variable is used as the transition variable with the delay parameter set to 1 for simplicity.[1] The intuitive appeal of the scheme in (2.1) is that it allows the speed of adjustment to vary asymmetrically with regimes.

Suppose that

$$\phi_0 \ge 1,\ |\phi_1|,\ |\phi_2| < 1. \tag{2.2}$$

The series are then locally nonstationary, but globally ergodic. Geometric ergodicity of the process is easily established using the drift condition proposed by Tweedie (1975). This condition states that a process is ergodic under regularity conditions that disturbances have positive densities everywhere if the process tends towards the center of its state space at each point in time. More specifically, an irreducible aperiodic Markov chain y_t is geometrically ergodic if there exists constants $\delta < 1$, $B, L < \infty$, and a small set C such that

$$E\left[\|y_t\| \mid y_{t-1} = y\right] < \delta \|y\| + L,\quad \forall y \notin C, \tag{2.3}$$

$$E\left[\|y_t\| \mid y_{t-1} = y\right] \le B,\quad \forall y \in C, \tag{2.4}$$

where $\|\cdot\|$ is a norm. The concept of the small set is the equivalent of a discrete Markov chain state in a continuous context. For more details see Tweedie (1975), Chan *et al.* (1985)

[1] In practice, there is likely to be little theoretical or prior guidance as to the value of the delay parameter d. We would suggest that d be chosen to maximise goodness of fit over $d = \{1, 2, ..., d_{\max}\}$, for example. In what follows, to clarify ideas and in keeping with empirical practice to date, we set $d = 1$.

and Balke and Fomby (1997). For the process y_t in (2.1) to be geometrically ergodic, we need the condition, $|\phi_1| < 1$ and $|\phi_2| < 1$. To prove this, define the small set $C = [r_1, r_2]$. Then, it is easily seen that the condition (2.4) is satisfied by the finiteness of $E(\|u_t\|)$. We thus need to prove (2.3), but it can be shown that

$$E\left[||y_t|| \mid y_{t-1} = y\right] \leq \max\left(|\phi_1|, |\phi_2|\right) \|y\| + L,$$

for all $y \notin C$ and for some finite L.[2]

We now consider the special case,

$$\phi_0 = \phi_1 = \phi_2 = 1. \tag{2.5}$$

In this case y_t reduces to a linear random walk process. Using Monte Carlo experiments based on the three regime SETAR model with $\phi_0 = 1$, $\phi_1 = \phi_2 < 1$, Pippenger and Goering (1993) show that the power of the DF test falls dramatically as the absolute value of common threshold parameter $r_1 = r_2$ increases. Assuming that y_t can be regarded as a known economic long-run relationship, then the threshold cointegration process is defined as a globally stationary three regime SETAR processes following a unit root in the middle regime, but being dampened in outer regimes. In this regard Balke and Fomby (1997) obtain similar findings in the context of threshold cointegration.

However, most studies applying threshold autoregressive or threshold error correction models adopt the two-step testing approach proposed by Balke and Fomby (1997). Here the first step determines the presence of unit root or cointegration using the standard linear unit root or cointegration tests, and if stationarity or cointegration is found, the second step tests whether or not threshold nonlinearity behavior is present. See for example Lo and Zivot (2001). While such linear tests will have power against nonlinear SETAR alternatives, it seems far more sensible to use a direct test that is designed to have power against the alternative of interest, *e.g.* the SETAR nonlinear dynamic adjustment. Only recently, has there been any attempt to address this joint testing issue. For example Enders and Granger (1998) test for unit root against two regime SETAR model alternatives. Despite this growing literature, there have been no attempts to develop unit root or cointegration tests in the context of three regime TAR-based models. In the next section we aim to fill this important gap in the literature by deriving a unit root test designed to have power against alternatives where the process is globally stationary and follows SETAR dynamics.

[2]Sufficient (but not necessary) conditions for geometric ergodicity might be similarly obtained for TAR processes with $p > 1$ and $d > 1$ by defining a Markov chain $\mathbf{y}_{-1} = \left(y_{t-1}, \ldots, y_{t-\max(p,d)}\right)'$. Though a formal proof is nontrivial and beyond the scope of this paper, we conjecture that the sufficient condition becomes that both lag polynomials of the outer regimes, denoted by $\phi_1(L)$ and $\phi_2(L)$, have roots outside the unit circle. See also Bec *et al.* (2001).

3 Testing the Null of a Unit Root Against the Alternative of Globally Stationary Three-Regime TAR Process

Following the maintained assumption in the literature, we now impose $\phi_0 = 1$ in (2.1), which implies that y_t follows a random walk in the corridor regime. Then, defining $1_{\{.\}}$ as a binary indicator function, (2.1) can be compactly written as

$$\Delta y_t = \beta_1 y_{t-1} 1_{\{y_{t-1} \leq r_1\}} + \beta_2 y_{t-1} 1_{\{y_{t-1} > r_2\}} + u_t, \tag{3.1}$$

where $\beta_1 = \phi_1 - 1$, $\beta_2 = \phi_2 - 1$, and $y_{t-1} 1_{\{y_{t-1} \leq r_1\}}$ and $y_{t-1} 1_{\{y_{t-1} > r_2\}}$ are orthogonal to each other by construction. We consider the (joint) null hypothesis of unit root as

$$H_0 : \beta_1 = \beta_2 = 0, \tag{3.2}$$

against the alternative hypothesis of threshold stationarity,

$$H_1 : \beta_1 < 0; \ \beta_2 < 0. \tag{3.3}$$

There have been a few attempts to develop unit root tests in the two regime TAR framework. First, Enders and Granger (1998) have addressed this issue using a two-regime TAR model with implicitly known threshold value,[3]

$$\Delta y_t = \left\{ \begin{array}{ll} \beta_1 y_{t-1} + u_t & if \ y_{t-1} \leq 0 \\ \beta_2 y_{t-1} + u_t & if \ y_{t-1} > 0 \end{array} \right\}, \ t = 1, 2, ..., T, \tag{3.4}$$

and suggested an F-statistic for $\beta_1 = \beta_2 = 0$ in (3.4). Despite the main aim to derive a more powerful test, their simulation evidence shows that the proposed F test is less powerful than the DF test that ignores the threshold nature of this two regime alternative. But they also provided simulation results showing that the F-test may have higher power than the DF test against the three regime asymmetric TAR models. Using consistent estimates of the threshold parameters under the alternative, Berben and van Dijk (1999) show that their proposed tests are more powerful than the DF test, especially when the adjustment is asymmetric.

We propose a more general approach based on a three-regime SETAR model, (3.1). Further assuming that the cointegrating parameters are known *a priori*, this approach can also be theoretically related to the analysis of threshold cointegration advanced by Balke and Fomby (1997). Lo and Zivot (2001) have also examined similar issues in a bivariate three regime TAR model, but only applied the two-regime-based Enders and Granger and Berben and van Dijk tests, assuming that the cointegrating parameters are known. Interestingly, it is found that these tests are more powerful than the standard cointegration test that totally ignores the three regime threshold nature of the alternative.

[3] In the case where the data has the non-zero mean, they suggest to use the de-meaned data, whereas for the processes with non-zero mean and non-zero linear trend, the de-meaned and de-trended series is used.

There has also been an alternative line of studies. Caner and Hansen (2001) have considered the following two-regime TAR model:

$$\Delta y_t = \boldsymbol{\theta}_1' \mathbf{x}_{t-1} 1_{\{\Delta y_{t-1} \leq r\}} + \boldsymbol{\theta}_2' \mathbf{x}_{t-1} 1_{\{\Delta y_{t-1} > r\}} + e_t,\ t = 1, 2, ..., T, \tag{3.5}$$

where $\mathbf{x}_{t-1} = (y_{t-1}, 1, \Delta y_{t-1}, ..., \Delta y_{t-k})'$, r is an unknown threshold parameter, and e_t is an *iid* error. They have first developed tests for threshold nonlinearity when y_t follows a unit root, and then unit root tests when the threshold nonlinearity is either present or absent. This approach clearly differs from our SETAR-based approach at least in two senses. *First*, they apply threshold nonlinearity explicitly to all parameters including an intercept, whereas we focus only on the TAR(1) parameter. *Second*, we use the lagged level of the series as the transition variable, as opposed to the difference of the series as used in Caner and Hansen (2001). Their approach would be useful in certain univariate contexts, *e.g.*, their empirical application to unemployment rates, but it may be of reduced interest for analysing the long-run economic relationship in the context of threshold cointegration. On the other hand our approach is theoretically more congruent when investigating the stationary nature of many economic relationships such as real exchange rates, real interest rates and the price-dividend ratio.

We now write (3.1) in matrix notation,

$$\Delta \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{3.6}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, and

$$\Delta \mathbf{y} = \begin{pmatrix} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_T \end{pmatrix};\ \mathbf{X} = \begin{pmatrix} y_0 1_{\{y_0 \leq r_1\}} & y_0 1_{\{y_0 > r_2\}} \\ y_1 1_{\{y_1 \leq r_1\}} & y_1 1_{\{y_1 > r_2\}} \\ \vdots & \vdots \\ y_{T-1} 1_{\{y_{T-1} \leq r_1\}} & y_{T-1} 1_{\{y_{T-1} > r_2\}} \end{pmatrix};\ \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}.$$

Then, the joint null hypothesis of linear unit root against the nonlinear threshold stationarity can be tested using the Wald statistic given by

$$\mathcal{W}_{(r_1, r_2)} = \hat{\boldsymbol{\beta}}' \left[Var\left(\hat{\boldsymbol{\beta}}\right)\right]^{-1} \hat{\boldsymbol{\beta}} = \frac{\hat{\boldsymbol{\beta}}' \left(\mathbf{X}'\mathbf{X}\right) \hat{\boldsymbol{\beta}}}{\hat{\sigma}_u^2}, \tag{3.7}$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator of $\boldsymbol{\beta}$, $\hat{\sigma}_u^2 \equiv \frac{1}{T-2} \sum_{t=1}^{T} \hat{u}_t^2$, and $\hat{u}_t$ are the residuals obtained from (3.1).

To derive the asymptotic null distribution of the Wald statistic, we first begin by considering the simple case of known and finite threshold parameters. In this case, it will be shown that the asymptotic null distribution of the Wald statistic does not depend on the values of r_1 and r_2. Thus, we consider the special case of $r_1 = r_2 = 0$, where the three regime SETAR model (3.1) reduces to the two regime model (3.4), which can be expressed as

$$\Delta \mathbf{y} = \mathbf{X}_0 \boldsymbol{\beta} + \mathbf{u}, \tag{3.8}$$

where

$$\mathbf{X} = \begin{pmatrix} y_0 1_{\{y_0 \leq 0\}} & y_0 1_{\{y_0 > 0\}} \\ y_1 1_{\{y_1 \leq 0\}} & y_1 1_{\{y_1 > 0\}} \\ \vdots & \vdots \\ y_{T-1} 1_{\{y_{T-1} \leq 0\}} & y_{T-1} 1_{\{y_{T-1} > 0\}} \end{pmatrix}.$$

The Wald statistic testing for $\boldsymbol{\beta} = \mathbf{0}$ in (3.8) is given by

$$\mathcal{W}_{(0)} = \frac{\hat{\boldsymbol{\beta}}' \left(\mathbf{X}_0' \mathbf{X}_0\right) \hat{\boldsymbol{\beta}}}{\hat{\sigma}_u^2}, \tag{3.9}$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator of $\boldsymbol{\beta}$, $\hat{\sigma}_u^2 \equiv \frac{1}{T-2} \sum_{t=1}^{T} \hat{u}_t^2$, and $\hat{u}_t$ are the residuals obtained from (3.4).

Theorem 3.1 *Consider the two-regime SETAR model (3.8). Then, the Wald statistic testing for $\boldsymbol{\beta} = \mathbf{0}$, defined by (3.9), has the following asymptotic null distribution:*

$$\mathcal{W}_{(0)} \Rightarrow \frac{\left\{\int_0^1 1_{\{W(s) \leq 0\}} W(s) dW(s)\right\}^2}{\int_0^1 1_{\{W(s) \leq 0\}} W(s)^2 ds} + \frac{\left\{\int_0^1 1_{\{W(s) > 0\}} W(s) dW(s)\right\}^2}{\int_0^1 1_{\{W(s) > 0\}} W(s)^2 ds}, \tag{3.10}$$

where $W(s)$ is a standard Brownian motion defined on $s \in [0,1]$.

This result is exactly the same as obtained for the F-test considered by Enders and Granger (1998), *i.e.* $F = \mathcal{W}_{(0)}/2$. This result is of limited use, but the next theorem shows that the limiting null distribution of the statistic $\mathcal{W}_{(r_1,r_2)}$ is in fact equivalent to that of $\mathcal{W}_{(0)}$.

Theorem 3.2 *Assuming that r_1 and r_2 are finite and given, and under the null hypothesis $\beta_1 = \beta_2 = 0$, the $\mathcal{W}_{(r_1,r_2)}$ statistic defined in (3.7) weakly converges to $\mathcal{W}_{(0)}$. Furthermore, under the alternative hypothesis $\beta_1 < 0$ and $\beta_2 < 0$, $\mathcal{W}_{(r_1,r_2)}$ diverges to infinity.*

This (null) distributional invariance is due to the well-established fact that the unit root process stays within the (fixed) corridor regime for a proportion of time which goes to zero at rate $T^{-1/2}$, *e.g.*, Feller (1957).

Asymptotic results are so far derived under the simplifying assumption that threshold parameters are known, and we now consider a general case with unknown threshold parameters. In such a case it is well-established that this test suffers from the Davies (1987) problem since unknown threshold parameters are not identified under the null. Most solutions to this problem involve integrating out unidentified parameters from the test statistics. This is usually achieved by calculating test statistics over a grid set of possible values of threshold parameters, r_1 and r_2, and then constructing the summary statistics. For stationary TAR models this problem has been studied in Tong (1990) and Hansen (1996). We consider the three most commonly used statistics: the supremum, the average and the exponential average of the Wald statistic defined respectively by

$$\mathcal{W}_{\text{sup}} = \sup_{i \in \Gamma} \mathcal{W}_{(r_1,r_2)}^{(i)}, \; \mathcal{W}_{\text{avg}} = \frac{1}{\#\Gamma} \sum_{i=1}^{\#\Gamma} \mathcal{W}_{(r_1,r_2)}^{(i)}, \; \mathcal{W}_{\text{exp}} = \frac{1}{\#\Gamma} \sum_{i=1}^{\#\Gamma} exp\left(\frac{\mathcal{W}_{(r_1,r_2)}^{(i)}}{2}\right), \tag{3.11}$$

where $\mathcal{W}^{(i)}_{(r_1,r_2)}$ is the Wald statistic obtained from the i-th point of the threshold parameters grid set, Γ and $\#\Gamma$ is the number of elements of Γ.

Unlike the stationary TAR models, the selection of the grid of threshold parameters needs more attention. The threshold parameters r_1 and r_2 usually take on the values in the interval $(r_1, r_2) \in \Gamma = [r_{\min}, r_{\max}]$ where $r_{\min}$ and $r_{\max}$ are picked so that $\Pr(y_{t-1} < r_{\min}) = \pi_1 > 0$ and $\Pr(y_{t-1} > r_{\max}) = \pi_2 < 1$. The particular choice for π_1 and π_2 is somewhat arbitrary, and in practice must be guided by the consideration that each regime needs to have sufficient observations to identify the underlying regression parameters. If we were to select the set Γ using the conventional quantile-based approach under which threshold parameters diverge under the null of a unit root and are bounded under the alternative, then the above asymptotic results will not hold.

However, since our approach assumes that the coefficient on the lagged dependent variable is set to zero in the corridor regime $(r_1 \leq y_{t-1} < r_2)$, we can assign arbitrarily small samples (relative to total sample) to the corridor regime since we do not need to estimate any parameters in the corridor regime. Notice also that the threshold parameters exist only under the alternative hypothesis in which the process is stationary and therefore bounded in probability. In this case only a finite grid search will be meaningful for further estimation. For a discussion on the construction of the grid in stationary threshold models that support our approach, see Tong (1990) and Chan (1993). This observation leads us to make an assumption that the grid for unknown threshold parameters should be selected such that the selected corridor regime be of finite width both under the null and under the alternative. Under this restriction, we can further establish that the theoretical results obtained in Theorems 3.1 and 3.2 do hold in the more general case with unknown threshold parameters as shown below. Noticing that a random walk process will stay within a corridor regime of finite width for $O_p\left(\sqrt{T}\right)$ periods only, then setting $\pi_1 = \bar{\pi} - c/T^{\delta}$ and $\pi_2 = \bar{\pi} + c/T^{\delta}$, where $\bar{\pi}$ is the sample quantile corresponding to zero and $\delta \geq 1/2$, guarantees that the grid set will be of finite width under the null hypothesis.[4] In practice, c can be chosen so as to give a reasonable coverage of each regime in samples of sizes usually encountered. For example, for $T = 100$ and $\delta = 1/2$, c can be set to 3 to give a 60% coverage of the sample for the grid.

Recently, in a similar context, Bec *et al.* (2002) develop an adaptive consistent unit root tests based on the symmetric three regime TAR model and propose an adaptive choice of the grid set which restricts the grid to remain bounded under the null (as we do) but unbounded under the alternative.[5] Unlike our approach, they do not impose that the autoregressive coefficient in the corridor regime is known in (3.1). Since the value of the autoregressive parameter of the corridor regime is likely to lie close to zero even under the alternative, esti-

[4]Further restrictions on the limits of the grid in the form of a minimum difference between the upper and lower bound may also be placed to guarantee that the grid width does not tend to zero asymptotically under the alternative hypothesis. For example, estimate a AR(1) model and obtain a consistent estimate of σ^2 under the null. We then use this and an assumed AR coefficient, say 0.99, to obtain the standard deviation of the process implied by a linear AR(1) model. This measure of spread can then be used to fix the minimum width of the thresholds grid.

[5]The assumption of an unbounded grid under the alternative made in Bec *et al.* (2002) does not seem to boost the power of the tests as also reported in their simulations. Since the threshold can meaningfully only take finite values under the alternative as the process is stationary, the likelihood of the model (and also the test statistic) is maximised only for finite thresholds.

mating this additional parameter and testing its (joint) significance will lead to a loss of the power of the test. Further, our approach not to estimate the inactive corridor regime simplifies considerably the asymptotic analysis. Finally, the symmetry assumption is sometimes too restrictive, see the discussion of the empirical section.

The pointwise convergence obtained in Theorem 3.2 is not sufficient for establishing uniform convergence of the supremum, the average and the exponential average of the Wald statistic. In addition, we need to prove the stochastic equicontinuity of $\mathcal{W}_{(r_1,r_2)}^{(i)}$ over the set Γ. For a definition of stochastic equicontinuity see for example Davidson (1994, p. 336).

Theorem 3.3 *Assuming that the grid set* Γ *is of finite width, the Wald statistic* $\mathcal{W}_{(r_1,r_2)}^{(i)}$ *is weakly stochastically equicontinuous, namely,* $\forall\epsilon$ *there exists* $\delta > 0$ *such that*

$$\underset{T\to\infty}{limsup} \Pr\left[\sup_{\mathbf{r}\in\Gamma} \sup_{\mathbf{r}'\in\mathbf{S}(\mathbf{r},\boldsymbol{\delta})} \left|\mathcal{W}_{\mathbf{r}}^{(i)} - \mathcal{W}_{\mathbf{r}'}^{(i)}\right| \geq \epsilon\right] < \epsilon, \tag{3.12}$$

where $\mathcal{W}_{\mathbf{r}}^{(i)}$ *is the Wald statistic obtained from the i-th point of* Γ, $\mathbf{r} = (r_1, r_2)$, $\mathbf{r}' = (r_1', r_2')$ *and* $S(\mathbf{r}, \delta)$ *is a sphere of radius* δ *centered around* $\mathbf{r}$.

Combining the stochastic equicontinuity established in Theorem 3.3 with pointwise convergence of $\mathcal{W}_{(r_1,r_2)}^{(i)}$ to $\mathcal{W}_{(0)}^{(i)}$ established in Theorem 3.2 we now establish the uniform convergence of $\mathcal{W}_{\text{sup}}$ and $\mathcal{W}_{\text{avg}}$ to $\mathcal{W}_{(0)}$ and of $\mathcal{W}_{\text{exp}}$ to $\exp\left(\mathcal{W}_{(0)}/2\right)$.

The previous results can be generalised threefold. *First*, processes with intercept and/or linear deterministic trend can be easily accommodated as follows: In the case where the data has the non-zero mean such that $z_t = \mu + y_t$, we use the de-meaned data $y_t = z_t - \bar{z}$ in (3.1), where $\bar{z}$ is the sample mean. In this case the asymptotic distribution is the same as (3.10) except that $W(s)$ is replaced by the de-meaned standard Brownian motion $\widetilde{W}(s)$ defined on $s \in [0, 1]$. Similarly, for the case with non-zero mean and non-zero linear trend, $z_t = \mu + \delta t + y_t$, we use the de-meaned and de-trended data $y_t = z_t - \hat{\mu} - \hat{\delta}t$ in (3.1), where $\hat{\mu}$ and $\hat{\delta}$ are the OLS estimators of μ and δ. Now the associated asymptotic distributions are such that $W(s)$ is replaced by the de-meaned and de-trended standard Brownian motion $\widehat{W}(s)$ defined on $s \in [0, 1]$. We refer to these three cases as Case 1: the zero mean process; Case 2: the process containing nonzero mean; Case 3: the process containing both nonzero mean and linear trend.[6] Table 1 presents selected fractiles of the asymptotic critical values, which have been tabulated using random walks of 5,000 observations and 50,000 replications.

Table 1. Asymptotic Critical Values of the $\mathcal{W}_{(0)}$ Statistic

	Case 1	Case 2	Case 3
90%	6.01	7.29	10.35
95%	7.49	9.04	12.16
99%	10.94	12.64	16.28

[6] Alternatively, we may consider the GLS detrending procedure advanced by Elliott *et al.* (1996). See Kapetanios and Shin (2003)

Second, we allow for the case where the errors in (3.1) are serially correlated. We simply follow Dickey and Fuller (1979), and consider the following augmented regression:[7]

$$\Delta y_t = \beta_1 y_{t-1} 1_{\{y_{t-1} \leq r_1\}} + \beta_2 y_{t-1} 1_{\{y_{t-1} > r_2\}} + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + u_t, \tag{3.13}$$

where $u_t \sim iid\,(0, \sigma_u^2)$.

Theorem 3.4 *The asymptotic null distribution of the Wald statistics testing for $\beta_1 = \beta_2 = 0$ in (3.13) is equivalent to that obtained under the case where the underlying disturbances are not serially correlated.*

Third, we consider a symmetric three-regime SETAR model compactly written as

$$\Delta y_t = \beta y_{t-1} 1_{\{|y_{t-1}| > r\}} + u_t, \tag{3.14}$$

where we impose $r_1 = r_2 = r$ and $\beta_1 = \beta_2 = \beta$. In this case we can consider the Wald test for $\beta = 0$ in (3.14), denoted by $\mathcal{W}_{(r)}$. Assuming that r is given, then it is also seen that the asymptotic null distribution of the $\mathcal{W}_{(r)}$ statistic is equivalent to the squared DF t-distribution. When this symmetry restriction holds, we expect that the $\mathcal{W}_{(r)}$ test would be more powerful. The same generalisations as mentioned above can be made to accommodate processes with intercept and/or linear deterministic trend, serially correlated errors as well as an unknown threshold parameter.

4 Monte Carlo Study

In this section we undertake a small-scale Monte Carlo investigation of the small sample size and power performance of the suggested tests in comparison with the DF test. In the first set of experiments we examine the size performance of the tests. Experiment 1(a) considers the random walk process:

$$y_t = y_{t-1} + u_t, \tag{4.1}$$

where $u_t \sim N(0, 1)$. Experiment 1(b) allows for serially correlated errors,

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{4.2}$$

[7]The augmentations may actually enter in a nonlinear way. In such cases, we would view the use of linear augmentation terms as a first order approximation to the underlying dynamics. In practice, it will also be interesting to investigate whether our proposed simple testing procedure can be generalised to a more general SETAR model, where all other coefficients on Δy_{t-j}, $j = 1, ..., p$, in (3.13) are also subject to the same SETAR adjustments. Although we do not pursue it here, it is straightforward to allow for the lag order p to tend to infinity and thus allow for general weakly dependent error processes for (3.1). Since nonlinear (stationary) processes have an infinite MA representation via the Wold decomposition (and hence an infinite AR representation), we expect that our test will be robust to the presence of further nonlinearities in (3.13).

where $\varepsilon_t \sim N(0,1)$ and $\rho = 0.3$ is considered. The next set of experiments examines the power performance of the tests, where the data is generated by

$$y_t = \begin{cases} \phi_1 y_{t-1} + u_t & if\ y_{t-1} \leq r_1 \\ y_{t-1} + u_t & if\ r_1 < y_{t-1} \leq r_2 \\ \phi_2 y_{t-1} + u_t & if\ y_{t-1} > r_2 \end{cases}, \ t = 1, 2, ..., T, \tag{4.3}$$

where $u_t \sim N(0,1)$. Experiment 2(a) considers the symmetric adjustment with $\phi_1 = \phi_2 = 0.9$, whereas asymmetric adjustments are examined in Experiment 2(b) with $\phi_1 = 0.85$ and $\phi_2 = 0.95$.

All experiments are carried out using the following statistics: the three version of summary (asymmetric) Wald statistics, $\mathcal{W}_{\text{sup}}$, $\mathcal{W}_{\text{avg}}$ and $\mathcal{W}_{\text{exp}}$, defined by (3.11), and their symmetric counterparts denoted by $\mathcal{W}^S_{\text{sup}}$, $\mathcal{W}^S_{\text{avg}}$ and $\mathcal{W}^S_{\text{exp}}$. For all power experiments, 200 initial observations are discarded to minimise the effect of initial conditions, and we also consider the standard DF t-statistic, DF, and the adaptive sup-based Wald statistic, $\mathcal{W}^{BGG}_{\text{sup}}$, proposed by Bec *et al.* (2002).[8] All experiments are based on 1,000 replications, and samples of 100 and 200 are considered. Empirical size and power of the tests are evaluated at the 5% nominal level. In all experiments we consider two cases: the process containing nonzero mean and the process containing both nonzero mean and linear trend. We select six different sets of threshold parameter values from 0.15 to 3.90 and -0.15 to -3.90, at steps of 0.75 and -0.75, respectively. For each sample the grid of either lower or upper threshold parameter comprises of eight equally spaced points between the 10% percentile (lower threshold) or 90% percentile (upper threshold) sample observation and the mean of the sample. For the symmetric tests the grid is also restricted to be symmetric.

As a benchmark, the upper half of Table 2 gives empirical size of the tests when the underlying DGP is the random walk process with serially uncorrelated errors. First of all, the supremum-based tests show substantial size distortions. But the tests based on the average and the exponential average seem to have more or less correct sizes. The lower half of Table 3 summarizes the results for the unit root processes with AR(1) serially correlated errors. To compute the test statistics we simply use the correct ADF(1) regression, see (3.13). Almost qualitatively similar results are observed as obtained previously. Again, the size distortion of the supremum tests is nonnegligible for all cases considered, and thus we do not consider their power performance in what follows.[9]

Table 2 about here

Table 3 presents relative power performance when autoregressive parameters in outer regimes are equal at 0.9. When the threshold band is relatively small, *e.g.*, $(r_1, r_2) = (-0.15, 0.15)$, the symmetric Wald and the DF tests are more powerful than the asymmetric Wald test. But, as shown by Pippenger and Goering (1993), the power of DF test decreases

[8] For simplicity we use our suggested grid set of finite width, not following their adaptive choice of the set. The 95% critical values for the three versions of the $\mathcal{W}^{BGG}_{\text{sup}}$ tests are 8.80, 11.31 and 14.61, respectively, which are obtained by stochastic simulation.

[9] Only when we carry out size experiment with $T = 2000$ using a grid of diminishing width given by e.g. $\pi_1 = \bar{\pi} \pm 2/\sqrt{T}$, where $\bar{\pi}$ is the sample quantile corresponding to zero, we find that the size of the supremum test improves dramatically as predicted by the asymptotic theory. But, this is of little practical relevance.

monotonically with the threshold values. On the other hand, the decrease in power of our suggested tests is much slower especially for the exponential average test, and the power of our tests eventually dominate the DF test as the threshold band gets wider. For example, looking at the demeaned processes with $(r_1, r_2) = (-3.15, 3.15)$ and $T = 200$, we find that the powers of the $\mathcal{W}_{\text{exp}}$, $\mathcal{W}_{\text{avg}}$, $\mathcal{W}^S_{\text{exp}}$, $\mathcal{W}^S_{\text{avg}}$ and DF tests are 0.772, 0.681, 0.737, 0.648 and 0.544, respectively. Despite our expectation that the symmetric Wald test is more powerful than the asymmetric one in this set-up, the overall powers of both tests are comparable.

Table 3 about *here*

Table 4 gives the results for asymmetric threshold autoregressive parameters set to 0.85 and 0.95, respectively. We find that all the tests are less powerful now than obtained in the symmetric case. The power loss is less significant for our suggested tests as the corridor regime widens, since the power loss of the DF test is faster. Also as expected, the asymmetric Wald test is now more powerful than the symmetric test as the threshold band gets larger.

Table 4 about *here*

Overall results suggest that both average and exponential average statistics have reasonably correct size and good power. Since the exponential average test seems to be more powerful, we recommend to use the exponential average tests in practice.[10] On the other hand, the power of the $\mathcal{W}^{BGG}_{\text{sup}}$ test is significantly lower than even the DF test for almost all experiments considered. As expected, the test proposed by Bec *et al.* (2002) does not perform well when the grid set of the thresholds are bounded under the alternative.[11]

5 Empirical Application to G7 Real Exchange Rates

Exchange rates affect both the relative price of goods and the return differential on assets. The first effect dictates purchasing power parity (PPP) and the second dictates uncovered interest parity, which are central to the study of international economics. However, at least for advanced economies, deviations form PPP are highly persistent and a substantial body of evidence suggests that the real exchange rate, measuring deviation from PPP, are indistinguishable from nonstationary time series. In this section we apply our proposed tests and examine whether the real exchange rates follow unit root processes. Since the real exchange rate embodies the long-run purchasing power parity relationship between nominal exchange rates, domestic and foreign prices, this test can be regarded as the univariate-based test for threshold cointegration, assuming that the cointegrating parameters are known. For the underlying theoretical background see Sercu *et al.* (1995), Balke and Fomby (1997), Michael *et al.* (1997), and Taylor (2001).

[10]Empirical results in the next section also seem to be consistent with this finding.

[11]Similar results are also found in Bec *et al.* (2002) when the grid set is chosen using the conventional quantile-based approach. But we note in passing that their test is designed to be more powerful under the adaptive choice of the grid set in which the upper bound tends to infinity under the alternative.

Quarterly data on US real exchange rates for the G7 countries were collected covering the period 1960Q1 to 2000Q4.[12] Following the Monte Carlo findings we consider only the average and the exponential average of both asymmetric and symmetric Wald tests, jointly with the DF tests. In practice, the number of augmentations in (3.13) must be selected prior to the test to accommodate possible serially correlated errors. We could propose that standard model selection criteria be used for this purpose because under the null of a linear model, the properties of these criteria are well understood at least in linear models, *e.g.*, Ng and Perron (1995). Here we choose four augmentations for the underlying regression as we have quarterly observations. Since all real exchange rates seem to be trending over the whole sample period, we use the detrended version of the tests. To construct the threshold parameter grid, we set the grid of either lower or upper threshold parameter comprises of eight equally spaced points between the 10% quantile (lower threshold) or 90% quantile (upper threshold) and the mean of the sample as described in the previous section.

Table 5 below presents the unit root test results. In sum, the DF test fails to reject the null hypothesis of a unit root for any of countries at the 5% significance level. On the other hand, our proposed asymmetry Wald tests reject the null three times out of five cases, namely for Germany, and Japan real exchange rates at the 5% significance level, and further for Italy at the 10% level, whereas the symmetry Wald test rejects the null two times for Germany and Japan at the 5% significance level.

Table 5 about *here*

Given the strength of evidence against the null and some support for the SETAR alternative we also obtain estimates of both asymmetric and symmetric SETAR autoregressive parameters in outer regimes along with the estimates of corresponding threshold parameters. The estimation results are also reported in Table 5. Although we cannot interpret the t-statistic as a significance test, we refer to it as "significant" if an asymptotic 95% confidence interval around the estimate excludes zero. We see that $\hat{\beta}_1$ and $\hat{\beta}_2$ in asymmetric SETAR models and $\hat{\beta}$ in symmetric SETAR models are "significant" in all cases except for Canada. These are consistent with similar findings by Michael *et al.* (1997) that nonlinear mean reversion (or more precisely range reversion) arising from fixed costs of transactions costs in currency markets is present for several industrial countries. Another interesting finding is that the speed of mean reversion is faster in the lower regime than in the upper regime for Germany, Italy and Japan. As high values of real exchange rates are defined as appreciation and vice versa, this implies that the data periods dominated by extreme depreciation may display substantially faster reversion towards their underlying equilibrium range than those characterised by extreme appreciation. This raises the issue that empirical evidence may be available to bolster the (alternative) hypothesis of asymmetric foreign exchange market interventions that countries may choose to resist depreciations more vigorously than appreciations, so-called "dread of depreciation," see Dutta and Leon (2002) for further details. This result clearly highlights the need for tests that are designed to deal with asymmetric SETAR models.

[12] The data have been obtained from the IFS database. Real exchange rates are calculated using the wholesale price index. But, the full data for France are not available, so we drop the French case.

6 Concluding Remarks

The investigation of nonstationarity in conjunction with the threshold autoregressive modelling has recently attracted attention in econometric study. It is clear that misclassifying a stable nonlinear process as nonstationary can be misleading both in impulse response and forecasting analysis. In this paper we have proposed unit root tests that are designed to have power against globally stationary three regime SETAR processes. Our proposed tests are shown to have better power than the DF test that ignores the three regime SETAR nature of the alternative. Although our tests are based on a univariate model, we have illustrated that it can also be used as a test of linear lack of cointegration against nonlinear threshold cointegration, assuming that the process under investigation can be regarded as a linear combination of the nonstationary variables with known cointegrating parameters.

There are further research issues. *First*, it might be possible to find an alternative testing procedure based on an arranged regression along similar lines to Tsay (1998) and Berben and van Dijk (1999), which is likely to boost the power of the tests. *Second*, a more general TAR(p) model could be adopted where all the parameters including intercepts are also subject to the same nonlinear scheme as in Caner and Hansen (2001). *Finally*, although our test is univariate, it could be extended to establish the existence of cointegrating equilibrium relationship such as those conjectured to govern real exchange rates. Recently, Kapetanios *et al.* (2003b) propose testing procedures to detect the presence of a cointegrating relationship that follows a globally stationary smooth transition autoregressive process in a nonlinear error correction model with an application to price-dividend relationship. In this regard, a cointegration test based on an error correction model subject to SETAR nonlinearity is currently under investigation.

Table 2. Size of Alternative Tests						
Experiment 1(a)						
	$\mathcal{W}_{\text{sup}}$	$\mathcal{W}_{\text{avg}}$	$\mathcal{W}_{\text{exp}}$	$\mathcal{W}^S_{\text{sup}}$	$\mathcal{W}^S_{\text{avg}}$	$\mathcal{W}^S_{\text{exp}}$
The process with nonzero mean						
$T = 100$	.161	.035	.051	.097	.033	.047
$T = 200$	.183	.041	.057	.108	.041	.052
The process with nonzero mean and linear trend						
$T = 100$	.125	.034	.045	.078	.030	.039
$T = 200$	.153	.036	.050	.089	.031	.044
Experiment 1(b)						
	$\mathcal{W}_{\text{sup}}$	$\mathcal{W}_{\text{avg}}$	$\mathcal{W}_{\text{exp}}$	$\mathcal{W}^S_{\text{sup}}$	$\mathcal{W}^S_{\text{avg}}$	$\mathcal{W}^S_{\text{exp}}$
The process with nonzero mean						
$T = 100$	.186	.037	.053	.105	.036	.048
$T = 200$	.186	.036	.054	.104	.036	.047
The process with nonzero mean and linear trend						
$T = 100$	.140	.032	.046	.083	.027	.038
$T = 200$	.150	.033	.050	.087	.031	.040

Table 3: Power of Alternative Tests for Experiment 2(a)								
	r_1	r_2	$\mathcal{W}_{\text{avg}}$	$\mathcal{W}_{\text{exp}}$	$\mathcal{W}^S_{\text{avg}}$	$\mathcal{W}^S_{\text{exp}}$	DF	$\mathcal{W}^{BGG}_{\text{sup}}$
The process containing nonzero mean								
$T = 100$	-0.15	0.15	.299	.361	.314	.374	.363	.295
	-0.90	0.90	.302	.385	.328	.377	.337	.278
	-1.65	1.65	.235	.315	.253	.311	.266	.223
	-2.40	2.40	.215	.283	.208	.266	.201	.178
	-3.15	3.15	.186	.280	.182	.240	.166	.160
	-3.90	3.90	.162	.228	.137	.188	.123	.135
$T = 200$	-0.15	0.15	.771	.824	.802	.853	.894	.792
	-0.90	0.90	.768	.815	.788	.834	.852	.779
	-1.65	1.65	.779	.830	.809	.847	.847	.739
	-2.40	2.40	.725	.799	.740	.803	.740	.630
	-3.15	3.15	.681	.772	.648	.737	.544	.500
	-3.90	3.90	.526	.663	.471	.595	.360	.309
The process containing nonzero mean and linear trend								
$T = 100$	-0.15	0.15	.176	.237	.168	.227	.216	.150
	-0.90	0.90	.182	.246	.184	.234	.188	.146
	-1.65	1.65	.157	.204	.139	.185	.159	.128
	-2.40	2.40	.129	.191	.122	.167	.148	.092
	-3.15	3.15	.124	.178	.109	.154	.127	.088
	-3.90	3.90	.117	.164	.098	.139	.103	.092
$T = 200$	-0.15	0.15	.553	.643	.530	.638	.651	.567
	-0.90	0.90	.529	.624	.518	.616	.635	.528
	-1.65	1.65	.506	.605	.500	.576	.560	.498
	-2.40	2.40	.439	.539	.414	.497	.441	.374
	-3.15	3.15	.355	.469	.329	.421	.334	.294
	-3.90	3.90	.267	.361	.220	.306	.235	.213

Table 4: Power of Alternative Tests for Experiment 2(b)

	r_1	r_2	$\mathcal{W}_{\text{avg}}$	$\mathcal{W}_{\text{exp}}$	$\mathcal{W}^S_{\text{avg}}$	$\mathcal{W}^S_{\text{exp}}$	DF	$\mathcal{W}^{BGG}_{\text{sup}}$
The process containing nonzero mean								
$T = 100$	-0.15	0.15	.226	.286	.232	.287	.250	.204
	-0.90	0.90	.230	.291	.234	.284	.250	.204
	-1.65	1.65	.214	.281	.210	.278	.240	.214
	-2.40	2.40	.202	.286	.198	.244	.181	.153
	-3.15	3.15	.178	.250	.151	.206	.131	.131
	-3.90	3.90	.140	.207	.117	.168	.117	.121
$T = 200$	-0.15	0.15	.628	.699	.626	.692	.692	.617
	-0.90	0.90	.615	.700	.597	.674	.674	.607
	-1.65	1.65	.622	.697	.593	.660	.611	.536
	-2.40	2.40	.597	.684	.575	.636	.551	.488
	-3.15	3.15	.518	.641	.472	.557	.426	.366
	-3.90	3.90	.428	.558	.378	.483	.323	.290
The process containing nonzero mean and linear trend								
$T = 100$	-0.15	0.15	.154	.204	.144	.181	.162	.133
	-0.90	0.90	.139	.193	.129	.173	.163	.132
	-1.65	1.65	.152	.213	.135	.193	.161	.121
	-2.40	2.40	.129	.181	.114	.151	.130	.109
	-3.15	3.15	.108	.151	.084	.122	.093	.077
	-3.90	3.90	.103	.150	.088	.124	.097	.076
$T = 200$	-0.15	0.15	.426	.512	.398	.487	.487	.395
	-0.90	0.90	.396	.480	.369	.447	.463	.396
	-1.65	1.65	.404	.491	.363	.443	.408	.346
	-2.40	2.40	.368	.454	.318	.398	.353	.317
	-3.15	3.15	.298	.388	.253	.336	.268	.238
	-3.90	3.90	.247	.343	.185	.267	.198	.184

Table 5: Test and Estimation Results for Three-Regime SETAR Models

	Canada	Germany	Italy	Japan	UK
$\mathcal{W}_{\text{avg}}$	2.47	12.5**	8.52	13.6**	8.89
$\mathcal{W}_{\text{exp}}$	3.85	815.3**	430.7*	1500**	105.1
$\mathcal{W}^S_{\text{avg}}$	1.65	10.8*	7.44	12.9*	8.48
$\mathcal{W}^S_{\text{exp}}$	2.33	325.2*	52.1	1132**	90.4
DF	-1.17	-2.99	-2.26	-3.23*	-2.73
Asymmetric Three-Regime SETAR					
$\hat{\beta}_1$	-0.022 (0.011)	-0.189 (0.052)	-0.283 (0.076)	-0.189 (0.050)	-0.125 (0.050)
$\hat{\beta}_2$	-0.011 (0.015)	-0.070 (0.035)	-0.114 (0.059)	-0.105 (0.045)	-0.116 (0.043)
$\hat{r}_1$	-0.088	-0.157	-0.200	-0.140	-0.043
$\hat{r}_2$	0.182	0.125	0.089	0.147	0.202
Symmetric Three-Regime SETAR					
$\hat{\beta}$	-0.011 (0.008)	-0.110 (0.029)	-0.156 (0.049)	-0.144 (0.035)	-0.139 (0.043)
$\hat{r}$	$\lvert 0.081 \rvert$	$\lvert 0.125 \rvert$	$\lvert 0.160 \rvert$	$\lvert 0.147 \rvert$	$\lvert 0.202 \rvert$

Note: * and ** indicate the rejection of the null of a unit root at the 10% and 5% significance level, respectively. The figures in $(\cdot)$ the below the estimates are their standard errors.

A Appendix

A.1 Proof of Theorem 3.1

Under the null, the $\mathcal{W}_{(0)}$ statistic defined in (3.9) can be expressed as

$$\begin{aligned}
\mathcal{W}_{(0)} &= \frac{1}{\hat{\sigma}_u^2}\hat{\boldsymbol{\beta}}'\left(\mathbf{X_0'X_0}\right)\hat{\boldsymbol{\beta}} = \frac{1}{\hat{\sigma}_u^2}\mathbf{u'X_0}\left(\mathbf{X_0'X_0}\right)^{-1}\mathbf{X_0'u} \\
&= \frac{1}{\hat{\sigma}_u^2}\left(\sum_{t=1}^{T} 1_{\{y_{t-1}\leq 0\}}y_{t-1}u_t, \sum_{t=1}^{T} 1_{\{y_{t-1}>0\}}y_{t-1}u_t\right) \\
&\times \begin{pmatrix} \sum_{t=1}^{T} 1_{\{y_{t-1}\leq 0\}}y_{t-1}^2 & 0 \\ 0 & \sum_{t=1}^{T} 1_{\{y_{t-1}>0\}}y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} 1_{\{y_{t-1}\leq 0\}}y_{t-1}u_t \\ \sum_{t=1}^{T} 1_{\{y_{t-1}>0\}}y_{t-1}u_t \end{pmatrix} \\
&= \frac{1}{\hat{\sigma}_u^2}\left(\frac{\left\{\sum_{t=1}^{T} 1_{\{y_{t-1}\leq 0\}}y_{t-1}u_t\right\}^2}{\sum_{t=1}^{T} 1_{\{y_{t-1}\leq 0\}}y_{t-1}^2} + \frac{\left\{\sum_{t=1}^{T} 1_{\{y_{t-1}>0\}}y_{t-1}u_t\right\}^2}{\sum_{t=1}^{T} 1_{\{y_{t-1}>0\}}y_{t-1}^2}\right).
\end{aligned}$$

Since the function $g_1(z) = z1_{\{z\leq 0\}}$ and $g_2(z) = z1_{\{z>0\}}$ are continuous, we can apply the continuous mapping theorem and can show that

$$\frac{1}{\sigma_u\sqrt{T}}1_{\{y_{t-1}\leq 0\}}y_{t-1} = 1_{\left\{\frac{1}{\sigma_u\sqrt{T}}y_{t-1}\leq 0\right\}}\frac{1}{\sigma_u\sqrt{T}}y_{t-1} \Rightarrow 1_{\{W(s)\leq 0\}}W(s), \tag{A.1}$$

where $W(s)$ is a standard Brownian motion defined on $s \in [0,1]$. Combining this result together with the following well-established result:

$$\frac{1}{\sigma_u\sqrt{T}}\sum_{t=1}^{[Ts]} u_t \Rightarrow W(s), \tag{A.2}$$

then it is straightforward to show that conditions of Theorem 2.2 in Kurz and Protter (1991) hold. More specifically, for processes $\{X_t\}^T \equiv X_T$ and $\{Y_t\}^T \equiv Y_T$, if (C1) X_T and Y_T are $\mathcal{F}$-adapted for some σ-field $\mathcal{F}$, (C2) $(X_T, Y_T) \Rightarrow (X, Y)$ and (C3) Y_T is a semi-martingale, then $\int X_T dY_T \Rightarrow \int X dY$. First, continuity of g_1 and g_2 imply (C1), secondly, (C2) has been shown in (A.1) and (A.2), and finally, $\sum_{t=1}^{[Ts]} u_t$ is clearly a semi-martingale and thus (C3) is trivially satisfied. Therefore, by this theorem on weak convergence of stochastic integrals we obtain

$$\frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}\leq 0\}}y_{t-1}u_t \Rightarrow \sigma_u^2\int_0^1 1_{\{W(s)\leq 0\}}W(s)dW(s),$$

$$\frac{1}{T^2}\sum_{t=1}^{T} 1_{\{y_{t-1}\leq 0\}}y_{t-1}^2 \Rightarrow \sigma_u^2\int_0^1 1_{\{W(s)\leq 0\}}W(s)^2 ds,$$

$$\frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}>0\}}y_{t-1}u_t \Rightarrow \sigma_u^2\int_0^1 1_{\{W(s)>0\}}W(s)dW(s),$$

$$\frac{1}{T^2}\sum_{t=1}^{T} 1_{\{y_{t-1}>0\}}y_{t-1}^2 \Rightarrow \sigma_u^2\int_0^1 1_{\{W(s)>0\}}W(s)^2 ds.$$

It is also easily seen that $\hat{\beta}$ is consistent and thus $\hat{\sigma}_u^2 \xrightarrow{p} \sigma_u^2$. Combining these results we obtain (3.10).

A.2 Proof of Theorem 3.2

To establish (pointwise) convergence in probability of $\mathcal{W}_{(r_1,r_2)}$ to $\mathcal{W}_{(0)}$ we need to show that

$$\frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}\le 0\}}\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2 - \frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}< r_1\}}\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2 \xrightarrow{p} 0, \tag{A.3}$$

$$\frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}> 0\}}\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2 - \frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}> r_2\}}\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2 \xrightarrow{p} 0, \tag{A.4}$$

$$\frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}\le 0\}}y_{t-1}u_t - \frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}< r_1\}}y_{t-1}u_t \xrightarrow{p} 0, \tag{A.5}$$

$$\frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}> 0\}}y_{t-1}u_t - \frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}> r_2\}}y_{t-1}u_t \xrightarrow{p} 0. \tag{A.6}$$

Considering for example (A.5), it can be shown that

$$\frac{1}{T}\sum_{t=1}^{T}\left[1_{\{y_{t-1}>0\}}\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2 - 1_{\{y_{t-1}>r_2\}}\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2\right] = \frac{1}{T}\sum_{t=1}^{T} 1_{\{0<y_{t-1}<r_2\}}\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2. \tag{A.7}$$

Standard analysis of random walks indicates that for finite r_1 and r_2, the number of nonzero terms in the summation in (A.7) is of order $\sqrt{T}$. Notice also that $\frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}\le r_1\}}\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2$ and $\frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}>r_2\}}\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2$ are bounded away from zero in probability, respectively, for any finite r_1 and r_2. As each of these terms is $O_p(1)$, the final expression in (A.7) tends to zero in probability. Similar analysis provides the desired result for other terms and thus proves the result.

To prove consistency we write

$$\mathcal{W}_{(r_1,r_2)} = \frac{\hat{\boldsymbol{\beta}}'\left(\mathbf{X}'\mathbf{X}\right)\hat{\boldsymbol{\beta}}}{\hat{\sigma}_u^2} = \frac{\left(\Delta\mathbf{y}'\mathbf{X}\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\Delta\mathbf{y}\right)}{\hat{\sigma}_u^2}. \tag{A.8}$$

Under the alternative, the process is stationary and thus it is easily seen that $\hat{\sigma}_u^2$ converges to nonzero constant. Further, assuming that u_t has an absolutely continuous density function and finite $4+\delta$ moments for some $\delta > 0$, and using Remark B of Chan (1993), we can prove that $T^{-1}\mathbf{XX}$ tends to a finite matrix. Therefore, we only need to show that $\Delta\mathbf{y}'\mathbf{X}$ diverges to infinity at rate T. For this we make the dependence of $\mathbf{X}$ on r_1 and r_2 explicit, say by $\mathbf{X}_{(r_1,r_2)}$. Denote the true value of the thresholds by r_1^0 and r_2^0. Expressing $\Delta\mathbf{y} = \mathbf{X}_{(r_1^0,r_2^0)}\boldsymbol{\beta} + \mathbf{u}$, it is sufficient to show that that $T^{-1}\mathbf{X}'_{(r_1^0,r_2^0)}\mathbf{X}_{(r_1,r_2)}$ has a finite nonzero probability limit. It is easily seen that this holds if we show either (i) the expectation of y_{t-1}^2 conditional on that $y_{t-1} < r$, $r < r_1^0$ and $r < r_1$ is nonzero or (ii) the expectation of y_{t-1}^2 conditional on that $y_{t-1} > r'$, $r' > r_2^0$ and $r' > r_2$ is nonzero where both r and r' are finite. But these are the variance of y_t conditional on the events $y_{t-1} < r$ and $y_{t-1} > r'$, respectively. These conditional variances have nonzero expectation unconditionally by stationarity and the finiteness of r and r'.

A.3 Proof of Theorem 3.3

We only consider the stochastic equicontinuity of $T^{-1}\sum_{t=1}^{T} 1_{\{y_{t-1}>r\}}y_{t-1}u_t$ because similar arguments can be applied to other terms. We assume that $r \in [-M, M]$ for some constant M. Following the definition of (weak) stochastic equicontinuity in (3.12), we have to prove that

$$\limsup_{T\to\infty} \Pr\left[\sup_{r}\sup_{r'\in S(\mathbf{r},\delta)}\left|\frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}>r\}}y_{t-1}u_t - \frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}>r'\}}y_{t-1}u_t\right| \ge \epsilon\right] < \epsilon, \tag{A.9}$$

where $S(r,\delta)$ is a sphere of radius δ centred at r. Assuming without loss of generality that $r' < r$, then the probability in (A.9) can be written as

$$\begin{aligned} & \limsup_{T\to\infty} \Pr\left[\sup_{r} \sup_{r'\in S(\mathbf{r},\delta)} \left|\frac{1}{T}\sum_{t=1}^{T} 1_{\{r'\leq y_{t-1}\leq r\}} y_{t-1}u_t\right| \geq \epsilon\right] \\ \leq \quad & \limsup_{T\to\infty} \Pr\left[\sup_{r} \sup_{r'\in S(\mathbf{r},\delta)} \frac{1}{T}\sum_{t=1}^{T} \left|1_{\{r'\leq y_{t-1}\leq r\}} u_t\right| |y_{t-1}| \geq \epsilon\right]. \end{aligned} \tag{A.10}$$

A standard result in random walk theory (see *e.g.*, Feller, 1957) is that a random walk crosses zero $O_{a.s.}(\sqrt{T})$ times. This implies that a random walk lies within a corridor of finite width for $O_{a.s}(\sqrt{T})$ periods. Therefore, $1_{\{r'\leq y_{t-1}\leq r\}}$ will take unity at most $\left[c\sqrt{T}\right]$ periods for some fixed constant c, where [.] denotes integer part, and zero otherwise.[13] Therefore, only $\left[c\sqrt{T}\right]$ terms in the summation in (A.10) are non-zero. In the cases where these terms are non zero, $|y_{t-1}|$ can be at most M. Taking the supremum over r and r' inside the summation in (A.10), it is easily seen that (A.10) holds if

$$\limsup_{T\to\infty} \Pr\left[\frac{M}{T}\sum_{i=1}^{\left[c\sqrt{T}\right]} |u_{t_i}| \geq \epsilon\right] < \epsilon, \tag{A.11}$$

where t_i denotes the subsequence of periods when the process lies within the finite corridor band. This is smaller than

$$\limsup_{T\to\infty} \Pr\left[\frac{M}{T}\sum_{i=1}^{\left[c\sqrt{T}\right]} \{|u_{t_i}| - E(|u_{t_i}|)\} + \frac{M}{T}\sum_{i=1}^{\left[c\sqrt{T}\right]} E(|u_{t_i}|) \geq \epsilon\right]. \tag{A.12}$$

By the finiteness of the second moment of u_t, $\frac{M}{T}\sum_{i=1}^{\left[c\sqrt{T}\right]} E(|u_{t_i}|)$ tends to zero. Hence, we concentrate on

$$\limsup_{T\to\infty} \Pr\left[\frac{M}{T}\sum_{i=1}^{\left[c\sqrt{T}\right]} \{|u_{t_i}| - E(|u_{t_i}|)\} \geq \epsilon\right]. \tag{A.13}$$

But by the law of large numbers, and using the assumption that u_t's are *iid*, we have

$$\limsup_{T\to\infty} \Pr\left[\frac{\sum_{i=1}^{\left[c\sqrt{T}\right]} \{|u_{t_i}| - E(|u_{t_i}|)\}}{cT^{1/2}} \geq \epsilon\right] = 0, \tag{A.14}$$

As the normalisation M/T in (A.13) is smaller than the normalisation $1/T^{1/2}$ needed for (A.14) to hold, hence (A.13) holds, which proves (A.9). Alternatively, using the law of the iterated logarithm (*e.g.*, Davidson, 1994, p. 408), it can also be shown that

$$\limsup_{T\to\infty} \frac{\sum_{i=1}^{\left[c\sqrt{T}\right]} \{|u_{t_i}| - E(|u_{t_i}|)\}}{T^{1/4}\ln(\ln(T^{1/2}))} = c_1,$$

where c_1 is a constant a.s. Since the normalisation M/T in (A.13) is smaller than the normalisation $1/T^{1/4}\ln(\ln(T^{1/2}))$ needed for the above result to hold, hence this will also prove the following strong equicontinuity condition:

$$\Pr\left[\limsup_{T\to\infty} \sup_{r} \sup_{r'\in S(\mathbf{r},\delta)} \left|\frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}>r\}} y_{t-1}u_t - \frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}>r'\}} y_{t-1}u_t\right| \geq \epsilon\right] = 0.$$

[13] A standard result in random walk theory is that a random walk crosses zero $O_{a.s.}(\sqrt{T})$ times. This implies that a random walk lies within a corridor of finite width for $O_{a.s}(\sqrt{T})$ periods. See Feller (1957).

A similar analysis provides a proof for stochastic equicontinuity of $T^{-1}\sum_{t=1}^{T} 1_{\{y_{t-1}>r\}} y_{t-1}^2$, for example. Given that $T^{-1}\sum_{t=1}^{T} 1_{\{y_{t-1}>r\}} y_{t-1}^2$ is almost surely bounded away from zero for all finite r, stochastic equicontinuity of the ratio of $\left(T^{-1}\sum_{t=1}^{T} 1_{\{y_{t-1}>r\}} y_{t-1}u_t\right)^2$ to $T^{-1}\sum_{t=1}^{T} 1_{\{y_{t-1}>r\}} y_{t-1}^2$ would be obtained.

A.4 Proof of Theorem 3.4

(3.13) can be written in the matrix form as

$$\Delta\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}, \tag{A.15}$$

where $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_p)'$, $\mathbf{Z} = (\Delta\mathbf{y}_{-1}, ..., \Delta\mathbf{y}_{-p})$, $\Delta\mathbf{y}_{-i} = (\Delta y_{-i+1}, ..., \Delta y_{T-i})$, $i = 1, ..., p$. Then,

$$\mathcal{W}_{(r_1,r_2)} = \frac{\hat{\boldsymbol{\beta}}'(\mathbf{X}'\mathbf{M}_T\mathbf{X})\hat{\boldsymbol{\beta}}}{\hat{\sigma}_u^2} = \frac{(\mathbf{u}'\mathbf{M}_T\mathbf{X})(\mathbf{X}'\mathbf{M}_T\mathbf{X})^{-1}(\mathbf{X}'\mathbf{M}_T\mathbf{u})}{\hat{\sigma}_u^2},$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator of $\boldsymbol{\beta}$, $\hat{\sigma}_u^2 \equiv \frac{1}{T-2}\sum_{t=1}^{T}\hat{u}_t^2$, $\hat{u}_t^2$ are the residuals obtained from (A.15), and $\mathbf{M}_T = \mathbf{I}_T - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is the $T \times T$ idempotent matrix. Defining the $T \times 1$ vectors, $\mathbf{x}_1 = \left[y_0 1_{\{y_0 \le r_1\}}, y_1 1_{\{y_1 \le r_1\}}, ..., y_{T-1} 1_{\{y_{T-1} \le r_1\}}\right]'$ and $\mathbf{x}_2 = \left[y_0 1_{\{y_0 > r_2\}}, y_1 1_{\{y_1 > r_2\}}, ..., y_{T-1} 1_{\{y_{T-1} > r_2\}}\right]'$, then,

$$\begin{aligned}\mathcal{W}_{(r_1,r_2)} &= \frac{1}{\hat{\sigma}_u^2}(\mathbf{u}'\mathbf{M}_T\mathbf{x}_1, \mathbf{u}'\mathbf{M}_T\mathbf{x}_2)\begin{pmatrix}\mathbf{x}_1'\mathbf{M}_T\mathbf{x}_1 & 0\\ 0 & \mathbf{x}_2'\mathbf{M}_T\mathbf{x}_2\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{x}_1'\mathbf{M}_T\mathbf{u}\\ \mathbf{x}_2'\mathbf{M}_T\mathbf{u}\end{pmatrix}\\ &= \frac{1}{\hat{\sigma}_u^2}\left\{\mathbf{u}'\mathbf{M}_T\mathbf{x}_1(\mathbf{x}_1'\mathbf{M}_T\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{M}_T\mathbf{u} + \mathbf{u}'\mathbf{M}_T\mathbf{x}_2(\mathbf{x}_2'\mathbf{M}_T\mathbf{x}_2)^{-1}\mathbf{x}_2'\mathbf{M}_T\mathbf{u}\right\}.\end{aligned}$$

Now, it is easily seen that

$$\frac{1}{T}\mathbf{x}_1'\mathbf{M}_T\mathbf{u} = \frac{1}{T}\mathbf{x}_1'\mathbf{u} + o_p(1),\ \frac{1}{T}\mathbf{x}_2'\mathbf{M}_T\mathbf{u} = \frac{1}{T}\mathbf{x}_2'\mathbf{u} + o_p(1),$$

$$\frac{1}{T^2}\mathbf{x}_1'\mathbf{M}_T\mathbf{x}_1 = \frac{1}{T^2}\mathbf{x}_1'\mathbf{x}_1 + o_p(1),\ \frac{1}{T^2}\mathbf{x}_2'\mathbf{M}_T\mathbf{x}_2 = \frac{1}{T^2}\mathbf{x}_2'\mathbf{x}_2 + o_p(1).$$

Hence,

$$\mathcal{W}_{(r_1,r_2)} = \frac{1}{\hat{\sigma}_u^2}\left\{\mathbf{u}'\mathbf{x}_1(\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{u} + \mathbf{u}'\mathbf{x}_2(\mathbf{x}_2'\mathbf{x}_2)^{-1}\mathbf{x}_2'\mathbf{u}\right\} + o_p(1). \tag{A.16}$$

Consider now the special case of $r_1 = r_2 = 0$. Along similar lines of logic, we have

$$\begin{aligned}\mathcal{W}_{(0)} &= \frac{(\mathbf{u}'\mathbf{M}_T\mathbf{X}_0)(\mathbf{X}_0'\mathbf{M}_T\mathbf{X}_0)^{-1}(\mathbf{X}_0'\mathbf{M}_T\mathbf{u})}{\hat{\sigma}_u^2}\\ &= \frac{1}{\hat{\sigma}_u^2}\left\{\mathbf{u}'\mathbf{x}_{01}(\mathbf{x}_{01}'\mathbf{x}_{01})^{-1}\mathbf{x}_{01}'\mathbf{u} + \mathbf{u}'\mathbf{x}_{02}(\mathbf{x}_{02}'\mathbf{x}_{02})^{-1}\mathbf{x}_{02}'\mathbf{u}\right\} + o_p(1),\end{aligned}$$

where $\mathbf{X}_0 = (\mathbf{x}_{01}, \mathbf{x}_{02})$, $\mathbf{x}_{01} = \left[y_0 1_{\{y_0 \le 0\}}, y_1 1_{\{y_1 \le 0\}}, ..., y_{T-1} 1_{\{y_{T-1} \le 0\}}\right]'$ and $\mathbf{x}_{02} = \left[y_0 1_{\{y_0 > 0\}}, y_1 1_{\{y_1 > 0\}}, ..., y_{T-1} 1_{\{y_{T-1} > 0\}}\right]'$. Furthermore,

$$\frac{1}{T}\mathbf{x}_{01}'\mathbf{u} = \frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}\le 0\}} y_{t-1}u_t \Rightarrow \sigma_u\sigma_{LR}\int_0^1 1_{\{W(s)\le 0\}} W(s)dW(s),$$

$$\frac{1}{T^2}\mathbf{x}_{01}'\mathbf{x}_{01} = \frac{1}{T^2}\sum_{t=1}^{T} 1_{\{y_{t-1}\le 0\}} y_{t-1}^2 \Rightarrow \sigma_{LR}^2\int_0^1 1_{\{W(s)\le 0\}} W(s)^2 ds,$$

$$\frac{1}{T}\mathbf{x}_{02}'\mathbf{u} = \frac{1}{T}\sum_{t=1}^{T} 1_{\{y_{t-1}>0\}} y_{t-1} u_t \Rightarrow \sigma_u \sigma_{LR} \int_0^1 1_{\{W(s)>0\}} W(s) dW(s),$$

$$\frac{1}{T^2}\mathbf{x}_{02}'\mathbf{x}_{02} = \frac{1}{T^2}\sum_{t=1}^{T} 1_{\{y_{t-1}>0\}} y_{t-1}^2 \Rightarrow \sigma_{LR}^2 \int_0^1 1_{\{W(s)>0\}} W(s)^2 ds,$$

where σ_{LR}^2 is the long-run variance of Δy_t. Using these results in (A.17), we obtain

$$\mathcal{W}_{(0)} \Rightarrow \frac{\left\{\int_0^1 1_{\{W(s)\leq 0\}} W(s) dW(s)\right\}^2}{\int_0^1 1_{\{W(s)\leq 0\}} W(s)^2 ds} + \frac{\left\{\int_0^1 1_{\{W(s)>0\}} W(s) dW(s)\right\}^2}{\int_0^1 1_{\{W(s)>0\}} W(s)^2 ds},$$

which is the same result as obtained in the case with serially uncorrelated errors. Next, using the same arguments as in the proof of Theorem 3.2, we can establish that for all finite r_1 and r_2, $\mathcal{W}_{(r_1,r_2)} \xrightarrow{p} \mathcal{W}_{(0)}$.

References

[1] Anderson, H. (1997) Transactions costs and non-Linear adjustment towards equilibrium in the US treasury bill market. *Oxford Bulletin of Economics and Statistics* 59, 465-84.

[2] Bec, F., Salem, M. B. and Carrasco, M. (2001) Tests of unit-root versus threshold specification with an application to the PPP. Unpublished manuscript, University of Rochester.

[3] Bec. F., Guay, A. and Guerre, E. (2002) Adaptive consistent unit root tests based on autoregressive threshold model. Unpublished manuscript, Université Paris 6.

[4] Balke, N. S. and Fomby, T.B. (1997) Threshold cointegration. *International Economic Review* 38, 627-45.

[5] Berben, R. and van Dijk, D. (1999) Unit root tests and asymmetric adjustment: a reassessment. Unpublished manuscript, Erasmus University Rotterdam.

[6] Caner, M. and Hansen, B. E. (2001) Threshold autoregression with a unit root. *Econometrica* 69, 1555-96.

[7] Chan, K.S. (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics* 21, 520-33.

[8] Chan, K.S., Petrucelli, J.D., Tong, H. and Woolford, S. W. (1985) A multiple threshold AR(1) model. *Journal of Applied Probability* 22, 267-79.

[9] Davidson J. (1994) *Stochastic Limit Theory.* Oxford University Press: Oxford.

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

[11] Dickey, D. A. and Fuller, W. A. (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427-31.

[12] Dutta, J. and Leon, H. (2002) Dread of depreciation: measuring real exchange rate interventions. Unpublished manuscript, University of Birmingham.

[13] Elliot, B. E., Rothenberg, T. J. and Stock, J. H. (1996) Efficient tests of the unit root hypothesis. *Econometrica* 64, 813–836.

[14] Enders, W. and Granger, C. W. J. (1998) Unit root test and asymmetric adjustment with an example using the term structure of interest rates. *Journal of Business and Economic Statistics* 16, 304-11.

[15] Engle, R. and Granger, C. W. J. (1987) Cointegration and error correction: representation, estimation and testing. *Econometrica* 55, 251-76.

[16] Feller, W. (1957) *An Introduction to Probability Theory and its Applications.* Wiley: New York.

[17] Gonzalez, M. and Gonzalo, J. (1998) Threshold unit root models. *Universidad Carlos III de Madrid, Discussion Paper.*

[18] Hansen, B. E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 414–30.

[19] Hicks, J. R. (1950) *A Contribution to the Theory of the Trade Cycle.* Oxford University Press: Oxford.

[20] Ioannides, C., Peel, D. and Peel, M. (2003) The time series properties of financial ratios: revisited. Forthcoming in *Journal of Business, Finance and Accounting.*

[21] Kapetanios, G. (1999) *Essays on the Econometric Analysis of Threshold Models.* Unpublished Ph.D. Thesis, University of Cambridge.

[22] Kapetanios, G. and Shin, Y. (2003) GLS detrending-based unit root tests in nonlinear STAR and SETAR framework. Unpublished manuscript, University of Edinburgh.

[23] Kapetanios, G., Shin, Y. and Snell, A. (2003a) Testing for a unit root in the nonlinear STAR framework. *Journal of Econometrics* 112, 359-79.

[24] Kapetanios, G., Shin, Y. and Snell, A. (2003b) Testing for cointegration in nonlinear STAR error correction models. Unpublished manuscript, University of Edinburgh.

[25] Koop, G., Pesaran, M. H. and Potter, S. M. (1996) Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* 74, 119-47.

[26] Kurtz, T. G. and Protter, P. (1992) Weak limit theorems for stochastic integrals and stochastic difference equations. *Annals of Probability* 19, 1035-70.

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

[28] Michael, P., Nobay, R. A. and Peel, D. A. (1997) Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation. *Journal of Political Economy* 105, 862-79.

[29] Ng, S. and Perron, P. (1995) Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90, 268-81.

[30] Pesaran, M.H. and Potter, S. M. (1997) A floor and ceiling model of US output. *Journal of Economic Dynamics and Control* 21, 661-95.

[31] Pippenger, M. K. and Goering, G. E. (1993) A note on the empirical power of unit root tests under threshold processes. *Oxford Bulletin of Economics and Statistics* 55, 473-81.

[32] Sercu, P., Uppal, R. and van Hulle, C. (1995) The exchange rate in the presence of transaction costs: implications for tests of purchasing power parity. *Journal of Finance* 50, 1309-19.

[33] Shin, D. W. and Lee, O. (2001) Test for asymmetry in possibly nonstationary time series data. *Journal of Business and Economic Statistics* 19, 233-44.

[34] Tong, H. (1990) *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press: Oxford.

[35] Taylor, A. (2001) Potential pitfalls for the PPP puzzle? Sampling and specification biases in mean reversion tests of the LOOP. *Econometrica* 69, 473-98.

[36] Tsay, R. S. (1998) Testing and modelling multivariate threshold models. *Journal of the American Statistical Association* 94, 1188-1202.

[37] Tweedie, R. L. (1975) Sufficient conditions for ergodicity and recurrence of Markov on a general state space. *Stochastic Processes Appl.* 3, 385-403.