Unit Root Tests in Three-Regime SETAR Models*

George Kapetanios
Bank of England

Yongcheol Shin
Department of Economics, University of Edinburgh

September 2001

Abstract

This paper proposes a simple direct testing procedure to distinguish a linear unit root process from a globally stationary three-regime self-exciting threshold autoregressive process. Following the threshold cointegration literature and thus assuming that the process follows the random walk in the corridor regime, the null of a unit root can be tested by the Wald test for the joint significance of threshold autoregressive parameters under the lower and the upper regime. We derive the asymptotic null distribution of the Wald statistic, and show that it does not depend on unknown fixed threshold values. Monte Carlo evidence clearly indicates that the exponential average of the Wald statistic is more powerful than the Dickey-Fuller test that ignores the threshold nature under the alternative.

JEL Classification: C12, C32, F31.
Key Words: Self-exciting Threshold Autoregressive Model, Unit Roots, Geometrically Ergodic Process, Threshold Cointegration, Wald Tests, Monte Carlo Simulations.

*We are grateful to James Davidson, Gary Koop, Peter Phillips, Andy Snell and the seminar participants at the Econometric Study Group, Bristol 2001 for their helpful comments on an earlier version of this paper. Partial financial support from the ESRC (grant No. R000223399) is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

The investigation of nonstationarity in economics and econometrics has assumed great significance over the past two decades. There has been increasing concern in macroeconomics 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 economic hypotheses. In particular, the power of tests such as the Dickey-Fuller (1979, 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 large T (number of time periods) asymptotics in linear models has also come under suspicion. As a response to these problems, macroeconomists are increasingly turning to nonlinear dynamics to improve estimation and inference from macro data. In particular, the issue of nonlinearity in economic phenomena and its econometric investigation have recently assumed a more pronounced role.

Some progress has already been made in this respect and now the applied macro/time-series literature abounds with cases where departing from linearity has yielded significant gains in both prediction and inference. In particular, Koop *et al.* (1996) use nonlinear threshold dynamics to fit asymmetries in the business cycle, and find that such asymmetries are a significant feature of real activity and that the threshold model dominates the linear alternative both within sample and in prediction. For more types of switching models see Tong (1990), Pesaran and Potter (1997) and Kapetanios (1999).

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, and 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. For example, Sercu *et al.* (1995) and Michael, Nobay and Peel (1997) have analysed real exchange rates, and developed the theory suggesting that the larger the deviation from the purchasing power parity, the stronger the tendency for real exchange rates to move back to equilibrium.

Balke and Fomby (1997) have recently popularised a joint analysis of nonstationarity and nonlinearity in the context of threshold cointegration. The threshold cointegrating process is defined as globally stationary such that it might follow a unit root or be explosive in the middle regime, but it is geometrically ergodic in the outer regimes. Most importantly, they have shown via Monte Carlo experiments that the power of standard unit root tests falls dramatically with threshold parameters of a three-regime TAR model. [See also Pippenger and Goering (1993).] Unfortunately, they have not derived a test that would be more powerful against these globally stationary TAR processes. Instead they suggest a two step approach. The first step determines the presence of cointegration using the standard cointegration test, whereas the second step tests whether or not threshold behavior is present, once cointegration is found.

Since then, there have been a few studies to address the two issues of nonstationarity and nonlinearity jointly, mostly using a univariate two regime TAR model. The first line of research follows the self-exciting TAR (SETAR) modelling approach where the lagged dependent variable is used as the threshold 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 test is less powerful than the DF test that ignores the threshold nature under the alternative. Berben and van Dijk (1999) have claimed that the low power of the Enders and Granger test is likely to be due to the use of biased estimate of threshold parameter under the alternative. They suggested other tests, which use consistent estimates of threshold parameters under the alternative, and are shown to be more powerful than the DF test, especially when the adjustment is asymmetric.

There has also been an alternative line of studies using a more general two-regime TAR model. Caner and Hansen (2000) have considered tests for threshold nonliearity when the underlying univariate process follows a unit root, and then unit root tests when the threshold nonlinearity is either present or absent. See also Gonzalez and Gonzalo (1998) for the asymptotic distribution of tests for threshold effects when the underlying processes follow both I(0) and I(1). This approach is critically different from the aforementioned SETAR-based approach; it allows only for stationary threshold variable. Thus, the possibility of using the lagged dependent variable as the threshold variable one is excluded since it becomes nonstationary under the null. In this regard this approach is of reduced interest when one wishes to analyse the threshold cointegration.

To bridge the two areas of nonstationarity and nonlinearity in the context of the threshold cointegration, we consider a three regime SETAR model rather than the two regime TAR model which has been the focus of much of earlier work. Clearly, approach is theoretically more sensible in terms of the speed of convergence arguments for investigating some economic hypotheses such as the PPP hypothesis and the stationarity of real interest rates. Lo and Zivot (1999) have examined similar issue in a multivariate three regime TAR model, but have not provided such a direct test that would have more power against stationary three regime TAR processes. This paper on the other hand provides such a direct test to distinguish the null of unit root from the alternative of a globally stationary three regime SETAR process.

Following threshold cointegration literature, *e.g.*, Balke and Fomby (1997) and thus assuming that the process follows the unit root in the corridor regime, the null hypothesis can be tested by the Wald test for the joint significance of autoregressive parameters under both lower and upper regimes. We then show that the suggested Wald test does not depend on threshold parameter values under the null asymptotically when threshold parameters are known. In this case its asymptotic null distribution (divided by 2) is shown to be equivalent to the distribution of the F-statistic as obtained for the two regime TAR model by Enders and Granger (1998). Moreover, in the special case where the autoregressive parameters under both lower and upper regimes are symmetric, the null hypothesis of a unit root can now be tested by the Wald test for the significance of the common autoregressive parameter. Its asymptotic null distribution is equivalent to the distribution of the squared DF t-statistic as obtained for the linear model. Furthermore, in our approach the coefficient on the lagged dependent variable is set to zero in the corridor regime and thus no parameters need to be identified in the corridor regime. This observation leads us to assume that even the grid for unknown thresholds should be of finite width. We therefore establish that the previous theoretical finding extends to a general case with unknown threshold parameters.

However, when threshold parameters are unknown, this kind of test suffers from the Davies (1987) problem since threshold parameters are not identified under the null. Following Andrews and Ploberger (1994) and Hansen (1996), we consider the three most commonly used summary statistics - average, supremum and exponential average of the statistics. Then, the small sample performance of these tests is compared to that of the DF test via a small-scale Monte Carlo experiment. We find that both the average and the exponential average tests have reasonably correct size and good power, but the supremum test tends to display significant size distortions. Both average and exponential average tests eventually dominates the power of the DF test as threshold band widens. Since the exponential average test is more powerful than the average test in most cases, we recommend use of the exponential average test, which is consistent with Andrews and Ploberger (1994)'s earlier finding in other context.

We illustrate the usefulness of our proposed tests by examining the stationarity of real exchange rates for the G7 countries as a way of testing the validity of PPP. We find that our proposed tests are able to reject the null hypothesis of a unit root in a number of cases whereas the DF test fails.

The plan of the paper is as follows: Section 2 describes globally stationary TAR processes in the context of the associated three-regime SETAR model. Section 3 develops the Wald statistic that directly tests the null of unit root against the alternative of globally stationary TAR processes, and presents the asymptotic theory. Section 4 investigates the small sample performance of the suggested tests via Monte Carlo simulations. Section 5 presents an empirical illustration. Section 6 discusses further issues and concludes. The appendix contains the 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:

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

where u_t is an *iid* sequence with zero mean and constant variance σ_u^2, r_1 and r_2 are threshold parameters. Here, the lagged dependent variable is used as the threshold variable with the delay parameter set to 1 without loss of generality.[1] This characterization may be relevant in various economic phenomena where relatively small shocks do not trigger a mean-reverting mechanism whereas relatively large shocks do. The intuitive appeal of the scheme in (2.1) is that it allows the speed of adjustment to vary with regimes.

[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 $d = 1$.

Suppose now that

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

Then, the series can be 'locally' nonstationary, but globally ergodic. The 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 satisfied by assuming a disturbance with positive density 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

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

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

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) 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 $y \in C^c$ 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 simply to a linear unit root process. Using Monte Carlo experiments based on the symmetric SETAR model (i.e. $\phi_0 = 1$, $\phi_1 = \phi_2 < 1$) Pippenger and Goering (1993) have first shown that the power of the standard DF test falls dramatically with absolute values of common threshold parameter $r_1 = r_2$. Balke and Fomby (1997) have obtained similar finding, and suggested the two step approach for testing for threshold cointegration. (Here y_t could be a known economic long-run relationships such as PPP.) The first step approach determines the presence of cointegration using the linear cointegration test, *e.g.*, the Engle and Granger (1987) test. The second step then involves determining whether or not threshold behavior is present, once cointegration is found. Utilizing a bivariate threshold vector error correction model, Lo and Zivot (1999) have extended the Balke and Fomby's two step approach for testing for threshold cointegration to a multivariate setting. However, they have not provided a direct test that would be more powerful against the globally ergodic alternative defined by (2.2).

In next section we derive such a direct testing procedure to distinguish between the linear nonstationary process defined by (2.5) and the nonlinear (asymptotically) stationary process defined by (2.2).

---

[2] Sufficient (but not necessary) conditions for geometric ergodicity might be similarly obtained for TAR processes with higher lag order p and longer delay parameter d by defining a Markov chain $\mathbf{y}_{-1} = \left(y_{t-1}, \ldots, y_{t-\max(p,d)}\right)'$ and carrying out similar steps. The condition then becomes that both lag polynomials, denoted by $\phi_1(L)$ and $\phi_2(L)$, have roots outside the unit circle.

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

Following the maintained assumption in the literature [see Balke and Fomby (1997) and Lo and Zivot (1999)], we now impose $\phi_0 = 1$ in (2.1), implying that y_t follows a random walk in the corridor regime. Then, using the DF transformation and defining $1_{\{.\}}$ as a binary indicator function, (2.1) can be compactly written as

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

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

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

against the alternative hypothesis of threshold stationarity,[3]

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

There have been a few attempts to develop the test of the unit root against threshold stationarity directly, but in a less general context. First, Enders and Granger (1998) have addressed this issue using a two-regime TAR model with implicitly known threshold value,

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

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 (only with stationary corridor regime). Berben and van Dijk (1999) have argued that the low power of the Enders and Granger test is due to the use of biased estimates of the threshold parameter under the alternative, and suggested an alternative test based on the use of consistent estimates of threshold parameters under the alternative. In particular, they have estimated the threshold parameter by drifting thresholds defined as a linear combination of the maximum and the minimum of y_{t-1} after the samples are rearranged according to the order statistics of the threshold variable, y_{t-1}, and tabulated critical values of the F-statistic for various values of the drifting threshold. Then, they have shown via simulations that their suggested test is more powerful than the DF test.

[3] The case where $\beta_1 > 0$ or $\beta_2 > 0$ is not of immediate economic interest. In such cases the nonlinear model will not be ergodic and thus the identifiability of thresholds and parameters of interest cannot be guaranteed.

In this section we propose a more general approach based on a three-regime SETAR model. Assuming that cointegrating parameters are known *a priori*, this approach is theoretically more relevant to analyse threshold cointegration advanced by Balke and Fomby (1997). Lo and Zivot (1999) have also examined similar issues in a bivariate three regime TAR model, but have not provided such tests that would have more power against the globally stationary three regime TAR process. Instead they have applied the two-regime-based Enders and Granger and Berben and van Dijk tests for threshold cointegration, assuming that the cointegrating parameters are known, and found that these tests are more powerful than the standard cointegration test that totally ignores the three regime threshold nature of the alternative.

There have also been an alternative line of studies in a similar context. Caner and Hansen (2000) have considered a more general two-regime TAR model set-up,

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

where $\mathbf{x}_{t-1} = (y_{t-1}, 1, \Delta y_{t-1}, ..., \Delta y_{t-k})'$, r is unknown threshold, and e_t is an *iid* error, then developed tests for threshold nonliearity when y_t follows a unit root, and unit root tests when the threshold nonlinearity is either present or absent. This approach differs from ours. First, threshold nonlinearity is explicitly applied to all parameters including an intercept, whereas we focus on the TAR(1) parameter(s) only. [See below for our treatment of nonzero intercept and linear trend coefficient, and of serially correlated errors.] Second, more importantly, we use the lagged level of the series as the threshold variable, as opposed to the difference of the series as used in Caner and Hansen (2000). Their approach would be useful in certain contexts, e.g. an empirical application to the unemployment rate, but clearly it is of reduced interest for analysing threshold cointegration. On the other hand our approach using the lagged level as a threshold variable is theoretically more sensible when investigating the stationary nature of some economic cointegrating relationshipssuch as the PPP and real interest rates, assuming that the cointegrating parameters are known *a priori*.

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

$$\Delta \mathbf{y} = \begin{pmatrix} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_T \end{pmatrix}; \ \mathbf{X} = \begin{pmatrix} y 1_{\{y_0 \le r_1\}} & y_0 1_{\{y_0 > r2\}} \\ y_1 1_{\{y_1 \le r_1\}} & y_1 1_{\{y_1 > r_2\}} \\ \vdots & \vdots \\ y_{T-1} 1_{\{y_{T-1} \le r_1\}} & y_{T-1} 1_{\{y_{T-1} > r_2\}} \end{pmatrix}; \ \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}.$$

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

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

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 begin to consider the simple case that threshold parameters are given. (A more general case for unknown thresholds will be examined below.) 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

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

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 (3.4) with zero threshold value. Then, the Wald statistic testing for $\boldsymbol{\beta} = \mathbf{0}$, defined by (3.9), has the following asymptotic null distribution:*

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

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.* $F = \mathcal{W}_{(0)}/2$, though they have provided only critical values of the F statistic, but have not formally derived its asymptotic distribution. In general this result is of limited use, but as next theorem shows, 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 given, and under the null hypothesis $\beta_1 = \beta_2 = 0$, the $\mathcal{W}_{(r_1, r_2)}$ statistic defined in (3.7) converges in probability 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). To confirm this theoretical finding we have also carried out stochastic simulations with 50,000 replications and a sample size of 10,000 random walk processes. Figure 1 plots the empirical distribution of $\mathcal{W}_{(r_1,r_2)}$ for different values of r_1 and r_2, which clearly shows that the distribution of the test statistic is invariant to the width of the corridor.

Figure 1 about *here*

The previous results can be generalised as follows: 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 the 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. Table 1 presents selected fractiles of the asymptotic critical values, which have been tabulated using 5,000 random walks and 50,000 replications.

Table 1 about here

Second, we allow for the case where the errors in (3.1) are serially correlated. Here we simply follow Dickey and Fuller (1979), and consider the following augmented (nonlinear) regression:[4]

$$\Delta y_t = \beta_1 y_{t-1} 1_{\{y_{t-1} \leq r_1\}} + \beta_2 y_{t-1} 1_{\{y_{t-1} > r_2\}} + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + u_t, \tag{3.11}$$

where $u_t \sim iid\,(0, \sigma_u^2)$. We now provide the asymptotic null distribution of the Wald statistic for testing the null hypothesis $\beta_1 = \beta_2 = 0$ in (3.11) in next theorem.

Theorem 3.3 *The asymptotic null distribution of the Wald statistics testing for $\beta_1 = \beta_2 = 0$ in (3.11) is equivalent to that obtained under the case where the underlying disturbances are not serially correlated.*

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

$$\Delta y_t = \beta y_{t-1} I_{(|r|,\infty)}\,(y_{t-1}) + u_t, \tag{3.12}$$

where we impose $r_1 = r_2$ and $\beta_1 = \beta_2 = \beta$. In this case we can consider the Wald test for $\beta = 0$ in (3.12), denoted by $\mathcal{W}_{(r)}$.[5] Assuming that r is given, then it is now easily seen

[4] Alternatively, nonparametric corrections can be used to accommodate serial correlation as popularised by Phillips and Perron (1988).

[5] Alternatively, the one-sided t-test might be more desirable, which would be more powerful. But, this causes some computational problems for calculating the summary statistics discussed below. So we focus on the Wald test here.

that the asymptotic null distribution of the $\mathcal{W}_{(r)}$ statistic is equivalent to of 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 as well as serially correlated errors.

So far all asymptotic results are derived under the unrealistic assumption that threshold parameters are given, and we now consider a general case with unknown thresholds. First of all, it is well-established that this kind of test now suffers from the Davies (1987) problem since unknown threshold parameters are not identified under the null. Most solutions to this problem involve some sort of integrating out unidentified parameters from the test statistics. This is usually achieved by calculating test statistics for a grid of possible values of threshold parameters, r_1 and r_2, and then constructing the summary statistics. This problem for stationary TAR models has been studied in Tong (1990) and Hansen (1996). As also advanced by Andrews and Ploberger (1994), the three most commonly used statistics are the supremum, the average and the exponential average of the Wald statistic defined respectively by

$$\mathcal{W}^{\sup}_{(r_1,r_2)} = \sup_{i \in \#\Gamma} \mathcal{W}^{(i)}_{(r_1,r_2)}, \tag{3.13}$$

$$\mathcal{W}^{\text{avg}}_{(r_1,r_2)} = \frac{1}{\#\Gamma} \sum_{i=1}^{\#\Gamma} \mathcal{W}^{(i)}_{(r_1,r_2)}, \tag{3.14}$$

$$\mathcal{W}^{\exp}_{(r_1,r_2)} = \frac{1}{\#\Gamma} \sum_{i=1}^{\#\Gamma} exp\left(\frac{\mathcal{W}^{(i)}_{(r_1,r_2)}}{2} \right), \tag{3.15}$$

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

Unlike the case for stationary TAR models, the selection of the grid of threshold parameters is now more complicated. 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 $\Pr(y_{t-1} < r_1) = \pi_1 > 0$ and $\Pr(y_{t-1} > r_2) = \pi_2 < 1$. It is typical to treat π_1 and π_2 symmetrically so that $\pi_1 = 1 - \pi_2$ which imposes the restriction that each of the lower and the upper regime has no less than π_1% of the total sample. The particular choice for π_1 is somewhat arbitrary, and in practice must be guided by the consideration that each regime needs to have sufficient observations to identify the regression parameters. But, notice that our approach assumes that the coefficient on the lagged dependent variable is set to zero in the corridor regime, following the threshold cointegration literature, *e.g.*, Balke and Fomby (1997). This implies that we could assign arbitrarily small samples to the corridor regime since we do not have to identify any parameters here. It would be equally argued that under the null the borders of data-dependent grids would grow without bound at least asymptotically. If so, estimation of the three-regime model is no longer plausible even under the alternative of stationarity. Hence, to avoid this anomaly, the set Γ must be restricted such that the sample size assigned to the corridor regime would be as arbitrarily small relative to the total sample size as possible. This observation leads us to make the assumption that the grid for unknown threshold should be of finite width, under which the theoretical arguments made in Theorems 3.1 and 3.2 do hold.

However, the pointwise convergence result we have obtained is not sufficient for deriving the asymptotic distribution of the supremum, the average and the exponential average of the Wald statistic. Unfortunately, we do not provide an analytic proof for uniform convergence. But, a weak convergence result for the average and the exponential average of the statistic may be obtained by restricting Γ to be a countable set. Then, it can be easily seen that the distribution of the average and exponential average follows from Theorem 3.2, because elements of the threshold parameter grid are enumerated. Since for each element of this enumerable set the test statistic converges in probability to $\mathcal{W}_{(0)}$, both average and exponential average statistics also converge to $\mathcal{W}_{(0)}$ and its exponential function, respectively. Sepcifically, we need to show that $\lim_{\#\Gamma,T\to\infty}\mathcal{W}^{\text{avg}}_{(r_1,r_2)}$ converges weakly to the probability limit of $\mathcal{W}_{(0)}$, which is given by $\lim_{\#\Gamma\to\infty}(1/\#\Gamma)\sum_{i=1}^{\#\Gamma}\lim_{T\to\infty}\mathcal{W}^{i}_{(r_1,r_2)}$ But, $\lim_{T\to\infty}\mathcal{W}^{i}_{(r_1,r_2)}$ does not depend on i for all $\#\Gamma$ and thus can be taken outside the summation. Therefore, the result follows. Similarly for the exponential average test. As there always exist countable sets that are dense in any real interval, this restriction is not very problematic and is of course of little relevance in finite samples. We further note that conditional on a given sample of finite or infinite size the estimates of the thresholds can only be identified within a countable set, becasue if we rank observations in ascending order then any real number between two consecutively ranked observations gives the same likelihood as the smaller (or larger depending on the definition of the indicator function) of the two observations.

In fact for all practical testing situations the standard methods for selecting the grid of threshold values may be used. This weak convergence argument for the average and the exponential average test will be evaluated mainly via Monte Carlo experiments in next section. As seen below, both average and exponential average tests have reasonably correct size and good power. A similar heuristic argument can also be made for the supremum tests, though simulation results clearly indicate that the supremum test may not be useful due to substantial size distortions in small samples.

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 the error term u_t is serially uncorrelated and drawn from the standard normal distribution. Experiment 1(b) allows for serially correlated errors,

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

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 we examine asymmetric adjustments 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 Wald statistics, $\mathcal{W}^{\sup}_{(r_1,r_2)}$, $\mathcal{W}^{\text{avg}}_{(r_1,r_2)}$ and $\mathcal{W}^{\exp}_{(r_1,r_2)}$, defined by (3.13) - (3.15), their symmetric counterparts denoted by $\mathcal{W}^{\sup}_{(r)}$, $\mathcal{W}^{\text{avg}}_{(r)}$ and $\mathcal{W}^{\exp}_{(r)}$, and the DF t-test. For all power experiments, 200 initial observations are discarded to minimise the effect of initial conditions. 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 three cases: Case 1: the zero mean process; Case 2: the process containing nonzero mean; Case 3: 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.[6] For each sample the grid of either lower or upper threshold parameter comprises of 8 equally spaced points between the minimum (lower threshold) or maximum (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, 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 $\mathcal{W}^{\sup}_{(r_1,r_2)}$ and the $\mathcal{W}^{\sup}_{(r)}$ tests show substantial size distortions. But the tests based on the average and the exponential average seem to have more or less correct sizes, though the average test is slightly undersized.

Table 2 about here

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 (3.11). Almost qualitatively similar results are observed here as obtained previously. Again, the size distortion of the supremum tests is nonnegligibel for all cases considered, and we thus do not consider their power performance in what follows.

Table 3 about here

Next, turning to the power comparison, Table 4 presents the relative power performance when the threshold autoregressive parameters in outer regimes are equal at 0.9. When the threshold band is relatively small, e.g. $(r_1, r_2) = (-0.15, 0.15)$, then 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 monotonically with the threshold values.

[6] We also find via simulation that the processes have spent at least 10% of the time in each of the outer regimes even for the largest threshold parameter values considered.

On the other hand, the decrease in power of our suggested tests is much slower especially for the exponential average, and the power of our suggested tests eventually dominate the DF test as the threshold band is wider. For example, looking at Case 2 (the demeaned processes) with $(r_1, r_2) = (-3.9, 3.9)$ and $T = 200$, we find that the powers of the $\mathcal{W}^{\exp}_{(r_1,r_2)}$, $\mathcal{W}^{\text{avg}}_{(r_1,r_2)}$, $t^{\exp}_{(r)}$, $t^{\text{avg}}_{(r)}$ and DF tests are 0.659, 0.533, 0.603, 0.481 and 0.352, respectively. Despite expectation that the symmetric Wald test is more powerful than the asymmetric Wald test, overall power for both tests are comparable, unless one is interested in the pedagogical Case 1. Though the power of the test is not size-adjusted, we may conclude that the exponential average test is more powerful than the average test.

Table 4 *here*

Table 5 gives the results for asymmetric threshold autoregressive parameters set to 0.85 and 0.95, respectively. We find that all the tests are more powerful now than obtained in the symmetric case. The power gain is much more significant for our suggested tests as the corridor regime widens, since the power loss of the DF test is much faster. Also as expected, the asymmetric Wald test is now more powerful than the symmetric test as the threshold band gets larger.

Table 5 *here*

Overall results suggest that both the average and the exponential average statistics have reasonably correct size and reasonable power. But, since the exponential average test is more powerful than the average test, we recommend to use the exponential average tests, which is consistent with Andrews and Ploberger (1994)'s findings.[7]

5 Empirical Illustration

In this section we apply our proposed tests and examine whether the real exchange rates follow unit root or are globally stationary threshold autoregressive processes. Considering that the real exchange rate is possibly the long-run cointegrating 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 and that the adjustment towards such long-run relationship can only be activated when the deviation from this equilibrium exceeds certain threshold values. See Sercu *et al.* (1995), Michael, Nobay and Peel (1997) and Balke and Fomby (1997).

Quarterly data on real exchange rates for the G7 countries were collected covering the period 1960Q1 to 2000Q4.[8] Following the Monte Carlo findings we consider only the average and the exponential average of both asymmetric and symmetric Wald tests, jointly with

[7] We have carried out another set of experiments with explosive corridor regime (with $\phi_0 = 1.1$ and 1.3), and obtained qualitatively similar results. We have also considered the bootstrap-based test as suggested by Hansen (1996), but find that such tests are less powerful than our suggested tests. The detailed results will be available upon request.

[8] 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.

the DF tests. In practice, the number of augmentations must be selected prior to the test to accommodate possible serially correlated errors. We would 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 and suggested. Here we simply choose the four augmentations in the underlying regression to match with quarterly observations. Considering that all real exchange rates seem to be trending over the whole sample periods, we use the detrended version of the tests. To construct the threshold parameter gird, 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.

Table 6 below presents the test results, which clearly demonstrate the empirical worth of our approach. In sum, the DF tests fail to reject the null hypothesis of a unit root for any of countries at the conventional 5% significance level, whereas our proposed tests clearly reject the null for the bilateral DM/USD and JPY/USD real exchange rates. At the 10% significance level, the $\mathcal{W}^{\exp}_{(r_1,r_2)}$ test is also able to reject the unit root for Italy.

6 Concluding Remarks

The investigation of nonstationarity in conjunction with the threshold autoregressive modelling has recently assumed a prominent role 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 the direct unit root test that is designed to have power against the geometrically ergodic process defined explicitly in the context of three regime SETAR. Our proposed tests are shown to have better power than the DF tests that ignores the three regime SETAR nature of the alternative. Although our suggested test is based on the univariate model, we have illustrated that it can also be used as a test of linear no 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, an extension to testing the null of linear no cointegration against the alternative of threshold cointegration in the multivariate regression context with unknown cointegrating parameters would be useful. In this case, both cointegrating parameters and threshold parameters are not identified under the null, and therefore inference would be more complicated. Second, 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. Third, a more general TAR(p) model could be adopted where all the parameters including coefficients on lagged first differences of the dependent variable are also subject to the same nonlinear scheme as in Caner and Hansen (2000).

Table 1 : Asymptotic Critical Values of the $\mathcal{W}_{(r_1,r_2)}$ 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

Table 2: Size of Alternative Tests for Experiment 1(a)

	$\mathcal{W}^{\sup}_{(r_1,r_2)}$	$\mathcal{W}^{\text{avg}}_{(r_1,r_2)}$	$\mathcal{W}^{\exp}_{(r_1,r_2)}$	$\mathcal{W}^{\sup}_{(r)}$	$\mathcal{W}^{\text{avg}}_{(r)}$	$\mathcal{W}^{\exp}_{(r)}$	DF
Case 1: zero mean process							
$T = 100$	.298	.041	.078	.287	.095	.130	.070
$T = 200$	.310	.045	.083	.286	.094	.129	.067
Case 2: the process containing nonzero mean							
$T = 100$	.161	.035	.051	.097	.033	.047	.045
$T = 200$	.183	.041	.057	.108	.041	.052	.049
Case 3: the process containing nonzero mean and linear trend							
$T = 100$	.125	.034	.045	.078	.030	.039	.054
$T = 200$	.153	.036	.050	.089	.031	.044	.050

Table 3: Size of Alternative Tests for Experiment 1(b)

	$\mathcal{W}^{\sup}_{(r_1,r_2)}$	$\mathcal{W}^{\text{avg}}_{(r_1,r_2)}$	$\mathcal{W}^{\exp}_{(r_1,r_2)}$	$\mathcal{W}^{\sup}_{(r)}$	$\mathcal{W}^{\text{avg}}_{(r)}$	$\mathcal{W}^{\exp}_{(r)}$	DF
Case 1: zero mean process							
$T = 100$	.323	.043	.088	.297	.089	.132	.062
$T = 200$	.315	.041	.084	.288	.091	.129	.068
Case 2: the process containing nonzero mean							
$T = 100$	.186	.037	.053	.105	.036	.048	.048
$T = 200$	.186	.036	.054	.104	.036	.047	.043
Case 3: the process containing nonzero mean and linear trend							
$T = 100$	.140	.032	.046	.083	.027	.038	.054
$T = 200$	.150	.033	.050	.087	.031	.040	.046

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

	r_1	r_2	$\mathcal{W}^{\text{avg}}_{(r_1,r_2)}$	$\mathcal{W}^{\text{exp}}_{(r_1,r_2)}$	$\mathcal{W}^{\text{avg}}_{(r)}$	$\mathcal{W}^{\text{exp}}_{(r)}$	DF
Case 1: zero mean process							
$T = 100$	-0.15	0.15	.438	.513	.765	.797	.792
	-0.90	0.90	.418	.496	.737	.770	.781
	-1.65	1.65	.402	.481	.756	.782	.747
	-2.40	2.40	.357	.437	.707	.739	.644
	-3.15	3.15	.283	.401	.657	.702	.490
	-3.90	3.90	.228	.343	.584	.642	.329
$T = 200$	-0.15	0.15	.888	.914	.977	.985	.999
	-0.90	0.90	.908	.935	.984	.990	1.00
	-1.65	1.65	.900	.928	.985	.989	.998
	-2.40	2.40	.906	.936	.994	.996	.998
	-3.15	3.15	.856	.908	.992	.994	.987
	-3.90	3.90	.728	.821	.979	.984	.943
Case 2: the process containing nonzero mean							
$T = 100$	-0.15	0.15	.273	.353	.283	.363	.330
	-0.90	0.90	.296	.363	.311	.370	.350
	-1.65	1.65	.248	.327	.259	.326	.295
	-2.40	2.40	.218	.304	.221	.281	.237
	-3.15	3.15	.171	.262	.166	.233	.164
	-3.90	3.90	.153	.245	.139	.207	.140
$T = 200$							
	-0.15	0.15	.766	.827	.797	.840	.876
	-0.90	0.90	.771	.836	.800	.859	.869
	-1.65	1.65	.763	.817	.795	.847	.826
	-2.40	2.40	.761	.827	.776	.836	.748
	-3.15	3.15	.676	.764	.649	.731	.560
	-3.90	3.90	.533	.659	.481	.603	.352
Case 3: the process containing nonzero mean and linear trend							
$T = 100$	-0.15	0.15	.171	.235	.162	.213	.196
	-0.90	0.90	.180	.255	173	.228	.194
	-1.65	1.65	.168	.212	.161	.209	.173
	-2.40	2.40	.132	.188	.122	.167	.138
	-3.15	3.15	.101	.151	.096	.133	.110
	-3.90	3.90	.113	.166	.096	.140	.116
$T = 200$	-0.15	0.15	.549	.642	.541	.629	.668
	-0.90	0.90	.529	.617	.509	.612	.636
	-1.65	1.65	.518	.597	.509	.599	.586
	-2.40	2.40	.470	.569	.458	.537	.457
	-3.15	3.15	.360	.464	.312	.397	.319
	-3.90	3.90	.265	.368	.230	.317	.233

Table 5: Power of Alternative Tests for Experiment 2(b)

	r_1	r_2	$\mathcal{W}^{\text{avg}}_{(r_1,r_2)}$	$\mathcal{W}^{\text{exp}}_{(r_1,r_2)}$	$\mathcal{W}^{\text{avg}}_{(r)}$	$\mathcal{W}^{\text{exp}}_{(r)}$	DF
Case 1: zero mean process							
$T = 100$	-0.15	0.15	.715	.771	.931	.948	.974
	-0.90	0.90	.720	.778	.944	.956	.979
	-1.65	1.65	.688	.750	.942	.957	.951
	-2.40	2.40	.642	.733	.942	.953	.904
	-3.15	3.15	.484	.625	.876	.903	.733
	-3.90	3.90	.352	.535	.757	.805	.501
$T = 200$	-0.15	0.15	.998	.999	1.00	1.00	1.00
	-0.90	0.90	.998	.999	.999	.999	1.00
	-1.65	1.65	.993	.997	1.00	1.00	1.00
	-2.40	2.40	.994	.996	1.00	1.00	1.00
	-3.15	3.15	.988	.993	1.00	1.00	1.00
	-3.90	3.90	.955	.978	.998	.998	.998
Case 2: the process containing nonzero mean							
$T = 100$	-0.15	0.15	.533	.611	.557	.635	.652
	-0.90	0.90	.564	.638	.578	.646	.655
	-1.65	1.65	.492	.590	.517	.581	.511
	-2.40	2.40	.424	.536	.400	.496	.339
	-3.15	3.15	.322	.460	.271	.400	.229
	-3.90	3.90	.253	.385	.210	.298	.173
$T = 200$	-0.15	0.15	.982	.990	.985	.994	.995
	-0.90	0.90	.975	.986	.982	.995	.999
	-1.65	1.65	.970	.978	.976	.990	.988
	-2.40	2.40	.971	.983	.966	.976	.962
	-3.15	3.15	.944	.970	.920	.956	.865
	-3.90	3.90	.834	.913	.782	.865	.579
Case 3: the process containing nonzero mean and linear trend							
$T = 100$	-0.15	0.15	.346	.432	.340	.428	.415
	-0.90	0.90	.376	.462	.368	.451	.437
	-1.65	1.65	.286	.371	.276	.355	.300
	-2.40	2.40	.226	.317	.209	.288	.221
	-3.15	3.15	.182	.268	.147	.211	.158
	-3.90	3.90	.154	.227	.136	.198	.142
$T = 200$	-0.15	0.15	.878	.929	.882	.941	.958
	-0.90	0.90	.884	.921	.886	.934	.942
	-1.65	1.65	.871	.922	.871	.920	.896
	-2.40	2.40	.824	.883	.794	.846	.758
	-3.15	3.15	.715	.820	.639	.741	.558
	-3.90	3.90	.465	.618	.386	.518	.324

Table 6: Unit Root Tests Against the Three-Regime SETAR+

	$\mathcal{W}^{avg}_{(r_1,r_2)}$	$\mathcal{W}^{exp}_{(r_1,r_2)}$	$\mathcal{W}^{avg}_{(r)}$	$\mathcal{W}^{exp}_{(r)}$	DF
Germany	12.46^{**}	815.3^{**}	10.75^{*}	325.2^{*}	-2.99
Japan	13.61^{**}	1500.6^{**}	12.92^{**}	1132.2^{**}	-3.23^{*}
Italy	8.52	430.7^{*}	7.44	52.1	-2.26
UK	8.89	105.1	8.48	90.4	-2.73
Canada	2.47	3.85	1.65	2.33	-1.17

$^{+}$: Real exchange rates for each country are measured with respect to US dollars, and the test is conducted over the period 1960Q1 to 2000Q4 using the underlying regressions (*e.g.*, (3.11)) with determinstic trends and four augmentations . * and ** indicate the 10% and 5% significance level, respectively.

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

$$\mathcal{W}_{(0)} = \frac{1}{\hat{\sigma}_u^2}\hat{\boldsymbol{\beta}}'\left(\mathbf{X_0'X_0}\right)\hat{\boldsymbol{\beta}} = \frac{1}{\hat{\sigma}_u^2}\mathbf{u'X_0}\left(\mathbf{X_0'X_0}\right)^{-1}\mathbf{X_0'u}.$$

Hence,

$$\begin{aligned}
\mathcal{W}_{(0)} &= \frac{1}{\hat{\sigma}_u^2}\left(\sum_{t=1}^{T} 1_{\{y_{t-1}\leq 0\}}y_{t-1}u_t, \sum_{t=1}^{T} 1_{\{y_{t-1}>0\}}y_{t-1}u_t\right) \\
&\times \begin{pmatrix} \sum_{t=1}^{T} 1_{\{y_{t-1}\leq 0\}}y_{t-1}^2 & 0 \\ 0 & \sum_{t=1}^{T} 1_{\{y_{t-1}>0\}}y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} 1_{\{y_{t-1}\leq 0\}}y_{t-1}u_t \\ \sum_{t=1}^{T} 1_{\{y_{t-1}>0\}}y_{t-1}u_t \end{pmatrix} \\
&= \frac{1}{\hat{\sigma}_u^2}\left(\frac{\left\{\sum_{t=1}^{T} 1_{\{y_{t-1}\leq 0\}}y_{t-1}u_t\right\}^2}{\sum_{t=1}^{T} 1_{\{y_{t-1}\leq 0\}}y_{t-1}^2} + \frac{\left\{\sum_{t=1}^{T} 1_{\{y_{t-1}>0\}}y_{t-1}u_t\right\}^2}{\sum_{t=1}^{T} 1_{\{y_{t-1}>0\}}y_{t-1}^2}\right).
\end{aligned}$$

Since the function $g_1(z) = 1_{\{z\leq 0\}}z$ and $g_2(z) = 1_{\{z>0\}}z$ are continuous, by the continuous mapping theorem we obtain

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

Combining this result together with the following well-established result:

$$\frac{1}{\sigma_u\sqrt{T}}\sum_{t=1}^{T} u \Rightarrow W(s),$$

then it is straightforward to show that the conditions of Theorem 2.2 in Kurz and Potter (1991) hold. By this theorem on weak convergence of stochastic integrals we now obtain

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

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

Similarly, we have

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

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

Using these results it is easily seen that $\hat{\beta}$ is consistent and thus so $\hat{\sigma}_u^2 \xrightarrow{p} \sigma_u^2$. Combining all of 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

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

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

$$\frac{1}{T}\sum_{t=1}^{T}\left\{1_{\{y_{t-1}\leq 0\}}y_{t-1}u_t - 1_{\{y_{t-1}<r_1\}}y_{t-1}u_t\right\} \xrightarrow{p} 0. \tag{A.3}$$

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

Considering for example (A.3), it can be shown that

$$\frac{1}{T}\sum_{t=1}^{T}\left[1_{\{y_{t-1}>0\}}\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2 - 1_{\{y_{t-1}>r_2\}}\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2\right] = \frac{1}{T}\sum_{t=1}^{T}1_{\{0<y_{t-1}<r_2\}}\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2. \tag{A.5}$$

Standard analysis of random walks indicates that for finite r_1 and r_2, the number of nonzero terms in the summation in (A.5) is of order $\sqrt{T}$. As each of these terms is $O_p(1)$, the final expression in (A.5) tends to zero in probability. Similar analysis provides the desired result for other terms and thus for (**??**).

To prove consistency we write

$$\mathcal{W}_{(r_1,r_2)} = \frac{\hat{\boldsymbol{\beta}}'\left(\mathbf{X}'\mathbf{X}\right)\hat{\boldsymbol{\beta}}}{\hat{\sigma}_u^2} = \frac{\left(\Delta\mathbf{y}'\mathbf{X}\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\Delta\mathbf{y}\right)}{\hat{\sigma}_u^2}. \tag{A.6}$$

Under the alternative of $\beta_1 < 0$ and $\beta_2 < 0$, the process is stationary. Thus, it can be shown that $\hat{\sigma}_u^2$ converges to some nonzero constant, and $T^{-1}\mathbf{X}\mathbf{X}$ tends to a finite matrix. Therefore, if we show that $\Delta\mathbf{y}'\mathbf{X}$ diverges to infinity at rate T, then the theorem is proved. For the purposes of this proof only 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. Rewriting $\Delta\mathbf{y}$ in terms of $\mathbf{X}$, it is sufficient to show that $\mathbf{X}'_{(r_1^0,r_1^0)}\mathbf{X}_{(r_1,r_2)}$ diverges to infinity at rate T or equivalently that $1/T\mathbf{X}'_{(r_1^0,r_1^0)}\mathbf{X}_{(r_1,r_2)}$ has a nonzero probability limit. It is easily seen that this holds if we show either that (i) the expectation of y_{t-1}^2 conditional on $y_{t-1} < r$, $r < r_1^0$ and $r < r_1$ is non zero or that (ii) the expectation of y_{t-1}^2 conditional on $y_{t-1} > r'$, $r' > r_2^0$ and $r' > r_2$ is non zero where r, r' are finite. But these quantities are the variances of y_t conditional on the events $y_{t-1} < r$ and $y_{t-1} > r'$ respectively. These conditional variances have non-zero unconditional expectations by stationarity and the finiteness of r, r'. Similar arguments could in principle be used to show consistency of the tests for explosive alternatives.

A.3 Proof of Theorem 3.3

(3.11) can be written in the matrix form as

$$\Delta\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}, \tag{A.7}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_p)'$, $\mathbf{Z} = (\Delta\mathbf{y}_{-1}, ..., \Delta\mathbf{y}_{-p})$, $\Delta\mathbf{y}_{-i} = (\Delta y_{-i+1}, ..., \Delta y_{T-i})$, $i = 1, ..., p$. Then, the Wald statistic is given by

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

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.7), and $\mathbf{M}_T = \mathbf{I}_T - \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$ is the $T \times T$ idempotent matrix. Defining the $T \times 1$ vectors,

$$\mathbf{x}_1 = \begin{pmatrix} y_0 1_{\{y_0<r_1\}} \\ y_1 1_{\{y_1<r_1\}} \\ \vdots \\ y_{T-1} 1_{\{y_{T-1}<r_1\}} \end{pmatrix}; \ \mathbf{x}_2 = \begin{pmatrix} y_0 1_{\{y_0>r_2\}} \\ y_1 1_{\{y_1>r_2\}} \\ \vdots \\ y_{T-1} 1_{\{y_{T-1}>r_2\}} \end{pmatrix};$$

then,

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

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,

$$\mathcal{W}_{(r_1,r_2)} = \frac{1}{\hat{\sigma}_u^2}\left\{\mathbf{u}'\mathbf{x}_1\left(\mathbf{x}_1'\mathbf{x}_1\right)^{-1}\mathbf{x}_1'\mathbf{u} + \mathbf{u}'\mathbf{x}_2\left(\mathbf{x}_2'\mathbf{x}_2\right)^{-1}\mathbf{x}_2'\mathbf{u}\right\} + o_p(1). \tag{A.8}$$

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

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

where $\mathbf{X}_0 = \left(\mathbf{x}_{01}, \mathbf{x}_{02}\right)$ and

$$\mathbf{x}_{01} = \begin{pmatrix} y_0 1_{\{y_0\leq 0\}} \\ y_1 1_{\{y_1\leq 0\}} \\ \vdots \\ y_{T-1} 1_{\{y_{T-1}\leq 0\}} \end{pmatrix}; \ \mathbf{x}_{02} = \begin{pmatrix} y_0 1_{\{y_0>0\}} \\ y_1 1_{\{y_1>0\}} \\ \vdots \\ y_{T-1} 1_{\{y_{T-1}>0\}} \end{pmatrix};$$

Furthermore,

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

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

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

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

where σ_{LR}^2 is the long-run variance of Δy_t. Using these results in (**??**), 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 argument as in the proof of Theorem 3.2, we can establish that

$$\mathcal{W}_{(r_1,r_2)} \xrightarrow{p} \mathcal{W}_{(0)} \quad \text{for all finite} \;\; r_1, r_2.$$

References

[1] Andrews, D.W.K. and W. Ploberger (1994), "Optimal Tests when a Nuisance Parameter is Present only under the Alternative," *Econometrica*, 62, 1383-1414.

[2] Balke, N.S. and T.B. Fomby (1997), "Threshold Cointegration," *International Economic Review*, 38, 627-645.

[3] Berben, R. and D. van Dijk (1999), "Unit Root Tests and Asymmetric Adjustment: A Reassessment," unpublished manuscript, Erasmus University Rotterdam.

[4] Caner, M. and B.E. Hansen (2000), "Threshold Autoregression with a Unit Root," forthcoming in *Econometrica*.

[5] Chan, K.S., J.D. Petrucelli, H. Tong and S.W. Woolford (1985), "A Multiple Threshold AR(1) Model," *Journal of Applied Probability*, 22, 267-279.

[6] Davies, R.B. (1987), "Hypothesis Testing When a Nuisance Parameter is Present Under the Alternative," *Biometrika*, 74, 33-43.

[7] Dickey, D.A. and W.A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.

[8] Enders, W. and C.W.J. Granger (1998), "Unit Root Test and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates," *Journal of Business and Economic Statistics*, 16, 304-311.

[9] Engle, R. and C.W.J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251-276.

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

[11] Gonzalez, M. and J. Gonzalo (1998), "Threshold Unit Root Models," *Universidad Carlos III de Madrid, Discussion Paper.*räsvirta (1993), *Modelling Nonlinear Economic Relationships*. Oxford University Press: Oxford.

[12] Hansen, B.E. (1996), "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis," *Econometrica*, 64, 414–430.

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

[14] Koop, G., M.H. Pesaran and S. Potter (1996), "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics*, 74, 119-147.

[15] Kurtz, T.G. and P. Potter (1991), "Weak Limit Theorems for Stochastic Integrals and Stochastic Difference Equations," *Annals of Probability*, 19, 1035-1070.

[16] Lo, M.C. and E. Zivot (1999), "Threshold Cointegration and Nonlinear Adjustment to the Law of One Price," unpublished manuscript, University of Washington.

[17] Michael, 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, 862-879.

[18] Pesaran, M.H. and S.M. Potter (1997), "A Floor and Ceiling Model of US Output," *Journal of Economic Dynamics and Control*, 21, 661-695.

[19] Pippenger, M.K. and G.E. Goering (1993), "A Note on the Empirical Power of Unit Root Tests under Threshold Processes," *Oxford Bulletin of Economics and Statistics*, 55, 473-481.

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

[21] Sercu, P., R. Uppal and C. Van Hulle (1995), "The Exchange Rate in the Presence of Transaction Costs: Implications for Tests of Purchasing Power Parity," *Journal of Finance*, 50, 1309-1319.

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

[23] Tsay, R.S. (1998), "Testing and Modelling Multivariate Threshold Models," *Journal of the American Statistical Association*, 94, 1188-1202.

[24] Tweedie, R.L. (1975), "Sufficient Conditions for Ergodicity and Recurrence of Markov on a General State Space," *Stochastic Processes Appl.*, 3, 385-403.