Edinburgh School of Economics

Discussion Paper Series

Number 60

Testing for a Linear Unit Root against Nonlinear Threshold Stationarity

George Kapetanios (NIESR)
Yongcheol Shin (University of Edinburgh)

Date
July 2000

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

Testing for a Linear Unit Root against Nonlinear Threshold Stationarity

George Kapetanios
NIESR

Yongcheol Shin
Department of Economics, University of Edinburgh

July 2000

Abstract

In this paper we propose a direct testing procedure to detect the presence of linear unit root against geometrically ergodic process defined by the self-exciting threshold autoregressive (SETAR) model with three regimes. Assuming that the process follows the random walk in the corridor regime, the null can be tested by the Wald test for the joint significance of the threshold autoregressive parameters under both lower and upper regimes. We prove that the suggested Wald test does not depend on unknown threshold values under the null at least asymptotically. We also derive its analytic asymptotic null distribution. Monte Carlo evidence clearly indicates that the exponential average of the Wald statistic is more powerful than the standard Dickey-Fuller test that ignores the threshold nature under the alternative.

JEL Classification: C12, C13, C32.
Key Words: Self-exciting Threshold Autoregressive Model, Exponentially Ergodic Process, Unit Roots, Threshold Cointegration, Wald Tests, Critical Values, Monte Carlo Simulations.

1 Introduction

The investigation of nonstationarity in economics and econometrics has assumed great significance in 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) unit root test or the Engel-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.

With regards to dynamics, however, economic theory has little to say about whether or not dynamic adjustment is linear or nonlinear but the presumption of linearity has always been made for the sake of convenience and computational simplicity. However nowadays, mathematical and computational simplicity are no longer prerequisites for successful execution and communication of research and such presumptions may be dispensed with. This is particularly the case when we consider that nonlinear models usually nest linear ones as special cases so that provided the nonlinear extension is reasonably parsimonious, invoking nonlinear dynamic adjustment should be a dominant modelling strategy. Of course this argument presumes the existence of the relevant statistical theory for estimation and inference.

Much 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. For example, Koop *et al.* (1996) use nonlinear threshold dynamics to fit asymmetries in the business cycle. They 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. See also Granger and Teräsvirta (1993), Balke and Fomby (1997), Michael et al. (1997), Caner and Hansen (1998), Tsay (1998) and Hansen and Seo (2000) for some examples of the growing literature examining the interplay between (non-) stationarity and nonlineariy.

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. See Sercu et al. (1995), With these arguments in mind, Kapetanios *et al.* (2000) incorporate nonlinear dynamics into a time series model explaining OECD real interest rates in the form of a smooth transition autoregressive (STAR) model. The STAR model has the property that the degree of persistence in a stationary time series may vary according to the proximity of the series to its equilibrium value. Kapetanios *et al.* (2000) find that an LM test of the null of a unit root in a linear AR process against an alternative of a STAR process is able to reject the null and hence solve the so-called "real interest rate puzzle" where despite theory to the contrary, rates appear to be nonstationary (see Rose (1988)).

In this paper we aim to bridge the two areas of nonstationarity and nonlinearity using the alternative threshold autoregressive (TAR) model. Theoretical models of nonlinear adjustments have been proposed earlier by Hicks (1950) and others in the context of (business) cycle analysis. However, much of the recent enthusiasm for these schemes is generated mainly by empirical findings. In particular, many researchers have found that the adjustment of output and employment to shocks appears to be asymmetric rather than symmetric as the linear model would impose (see Beaudry and Koop (1993)). The threshold model is capable of accommodating such asymmetric adjustment. For similar but more complex types of switching model see also Tong (1990), Pesaran and Potter(1997) and Kapetanios (1999).

Recently, Balke and Fomby (1997) have popularised a joint analysis of nonstationarity and nonlinearity in the context of threshold cointegration. They suggest the two step approach for testing for threshold cointegration. The first step determines the presence of cointegration using the Engle and Granger (1987) test based on the linear model. Then, the second step involves testing whether threshold behavior is present. Using Monte Carlo experiments based on the TAR model with three regimes Balke and Fomby (1997) have shown that the power of standard DF test falls dramatically with threshold parameters. See also Pippenger and Goering (1993).

As acknowledged by Balke and Fomby (1997), however, it would be ideal to test the null of no linear cointegration directly against the alternative of threshold cointegration. Enders and Granger (1998) have provided the direct test of the linear unit root against threshold stationarity in the context of the TAR model with two regimes. Contrary to main expectation to obtain a more powerful test, the simulation evidence presented shows that the suggested F test has less power than standard (linear) DF test that ignores the threshold nature of the two regime alternative. Berben and van Dijk (1999) have reconsidered this approach and argued that the low power of the Enders and Granger test is likely due to the use of biased estimates of threshold parameter under the alternative. They then have suggested alternative tests based on the use of consistent estimates of threshold parameters under the alternative, which are shown to be more powerful. Lo and Zibot (1999) have considered the testing for threshold cointegration in a multivariate TAR model with three regimes, then applied these univariate-based tests assuming that the cointegrating parameters are known. Their Monte Carlo evidence suggests that these tests are more powerful than the linear cointegration test that ignores the threshold nature of the alternative. But, the approach taken by Lo and Zibot (1999) is an extension of Balke and Fomby's two step approach, and therefore, testing for cointegration and testing for nonlinearity must be carried out separately.

This paper develops the direct testing procedure for the null of linear nonstationarity against the alternative of a geometrically ergodic process defined by the threshold autoregressive process with three regimes in a univariate framework. Assuming that the process follows the random walk in the corridor regime, the null can be tested by the Wald test for the joint significance of the threshold autoregressive parameters under both lower and upper regimes. It is well-established that this kind of test suffers from the Davies (1987) problem since threshold parameters are not identified under the null. But, we prove that the suggested test does not depend on unknown threshold values under the null at least asymptotically.

We also derive its analytic asymptotic null distribution, which is a functional of standard Brownian motions. This finding may be useful in practice, since most test problem for the case with threshold parameters unidentified under the null involves some sort of simulation or bootstrapping techniques to evaluate its distribution, see for example Hansen (1996) and Hansen and Seo (2000).

The finite performance of the suggested Wald tests is evaluated via a small-scale Monte Carlo experiments and compared to that of the liner DF tests. More specifically we consider three most common summary statistics - average, supremum and exponential average of the Wald statistic. We find that both the average and the exponential average of Wald statistics have reasonably correct sizes, but the supremum of the Wald test tends to have significant size distortions. Turning to the power performance, the exponential average of the Wald statistic is more powerful than the standard DF test for most cases considered, especially when the threshold nonlinearity is asymmetric under the alternative.

The plan of the paper is as follows: Section 2 overviews the univariate threshold autoregressive model and discusses nonlinear geometrically ergodic processes. Section 3 develops the nonlinear unit root testing procedure, and derives the asymptotic theory. Section 4 investigates the finite sample performance of the new testing procedures via Monte Carlo simulations. Section 5 discusses further issues on testing for cointegration in error correction framework, and concludes. Appendix contains mathematical proofs for theoretical results.

2 Threshold Autoregressive Process: Overview

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

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

where u_t is an *iid* sequence with zero mean and constant variance σ_u^2, r_1 and r_2 are unknown threshold values. For a general discussion on SETAR model see Tong (1990). Here the lagged dependent variable is used as the threshold variable with delay parameter set to 1 without loss of generality. 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. Theoretical models of nonlinear adjustment have been proposed earlier by Hicks (1950).

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.[1] For further details and an application see Chan et al (1985), Balke and Fomby (1997) and Kapetanios (1998). Consider now the special case,

$$\phi_0 = \phi_1 = \phi_2 = 1. \tag{2.3}$$

In this case y_t reduces to a linear random walk process. In this paper we aim to develop the direct testing procedure designed to distinguish the linear nonstationary process defined by (2.3) from the nonlinear (asymptotically) stationary process defined by (2.2).

The justification for using the level of the series as the threshold variables, as opposed to the difference of the series as used elsewhere, e.g. Hansen and Seo (2000), is as follow: If the model follows a (linear) random walk process, regime switches do not alter the nonstationary nature of the data such that the specification of the switch does not matter. If the model is asymptotically stationary, then there is a less compelling reason to suggest that the difference of the series should regulate regime switches. But, as suggested by Enders and Granger (1998), there might be some practical situations in practice to use growth rates as threshold variables.

Balke and Fomby (1997) have analysed and popularised similar problems in the context of threshold cointegration. They suggest the two step approach for testing for threshold cointegration. The first step approach determines the presence of cointegration using the Engle and Granger (1987) approach. The second step involves determining whether threshold behavior is present. Utilising a bivariate threshold vector error correction model Lo and Zibot (1999) have extended the Balke and Fomby's two step approach for testing for threshold cointegration to a multivariate setting, but also developed a systematic strategy for inference, estimation and specification test to examine nonlinear adjustments based on TAR models with three regimes. As mentioned earlier, however, it would be ideal to test the null of no linear cointegration directly against the alternative of threshold cointegration.

Recently, a few attempts have been made to address and/or develop the direct test of the linear unit root against threshold stationarity. Enders and Granger (1998) have addressed this problem in the context of the simpler TAR models with two regimes and implicitly known threshold value,[2]

$$y_t = \left\{ \begin{array}{ll} \phi_1 y_{t-1} + u_t & if\ y_{t-1} < 0 \\ \phi_2 y_{t-1} + u_t & if\ y_{t-1} \geq 0 \end{array} \right\},\ t = 1, 2, ..., T, \tag{2.4}$$

and suggested to use the two step-based F-statistic testing for $\phi_1 = \phi_2 = 0$. Although main aim is to derive a more powerful test,[3] the simulation evidence shows that the suggested

[1] More specifically the condition states that an irreducible aperiodic Markov chain y_t is geometrically ergodic if there exists constants $\rho > 1$, $B < \infty$, a nonnegative measurable function g and a small set C such that

$$\rho E\left[g(y_t) \mid y_{t-1} = y\right] < g(y), \quad \forall y \notin C,$$
$$E\left[g(y_t) \mid y_{t-1} = y\right] \leq B, \quad \forall y \in C.$$

For more details see Tweedie (1975).

[2] They have also considered alternative linear attractors of the unconditional mean or the linear trend line.

[3] 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 standard DF test falls dramatically with absolute values of common threshold parameter $r_1 = r_2$. See also Balke and Fomby (1997).

F test has less power than standard (linear) DF test that ignores the threshold nature of the two regime alternative. But they also provided the simulation results showing that the F-test may have higher power than the DF test against the three regime asymmetric TAR models.[4] Berben and van Dijk (1999) have reconsidered the approach put forward by Enders and Granger (1998), and argued that the low power of the Enders and Granger test is likely due to the use of biased estimates of threshold parameter under the alternative. Then, they suggest a more powerful test based on the use of consistent estimates of threshold parameters under the alternative. Lo and Zibot (1999) have applied the univariate-based test suggested by Enders and Granger (1998) and Berben and van Dijk (1999) to testing for threshold cointegration with three regimes in a multivariate setting, assuming that the cointegrating parameters are known. Via small-scale Monte Carlo experiments they find that these tests are more powerful than the standard linear cointegration test that totally ignores the threshold nature of the three regime alternative.[5]

3 Nonlinear Unit Root Test Against Threshold Stationarity

In this section we develop the direct unit root test against the alternative of threshold stationarity based on the SETAR model with three regimes (2.1). There are at least two main motivations. First, the three regimes TAR model is more theoretically oriented as explained in section 2. Secondly, although the test based on the two regimes TAR model can be used against the TAR models with three regimes, it would be more desirable to derive a direct test designed to have a power against the SETAR process with three regimes under the alternative, see (2.1).

Following the maintained assumption in literature (e.g. Balke and Fomby (1997) and Lo and Zibot (1999)), we now impose $\phi_0 = 1$ in (2.1), implying that y_t follows random walk in the corridor regime. Then, using the DF transformation and defining $I_A(\cdot)$ as a binary indicator function, (2.1) can be compactly written as

$$\Delta y_t = \beta_1 y_{t-1} I_{(-\infty, r_1)}(y_{t-1}) + \beta_2 y_{t-1} I_{(r_2, \infty)}(y_{t-1}) + u_t, \tag{3.5}$$

where $\beta_1 = \phi_1 - 1$, $\beta_2 = \phi_2 - 1$, and by construction $y_{t-1} I_{(-\infty, r_1)}(y_{t-1}) x_{1t}$ and $y_{t-1} I_{(r_2, \infty)}(y_{t-1})$ are orthogonal each other. We then consider the following (joint) null hypothesis of linear unit root

$$H_0 : \beta_1 = \beta_2 = 0, \tag{3.6}$$

against the alternative hypothesis of nonlinear threshold stationarity[6]

$$H_1 : \beta_1 < 0; \ \beta_2 < 0. \tag{3.7}$$

[4]But, they consider only the stationary corridor regime.

[5]Lo and Zibot (1999) have conjectured that the Berben and van Dijk's test should also have power against three regime TAR alternatives. But there is no discussion on the size performance of the tests in Berben and van Dijk (1999) and Lo and Zibot (1999) especially in the presence of serially correlated errors.

[6]The case where $\phi_1 > 0$ or $\phi_2 > 0$ 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.

More compactly, we write (3.5) in matrix notation,

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

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

$$\Delta \mathbf{y} = \begin{pmatrix} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_T \end{pmatrix}; \ \mathbf{X} = \begin{pmatrix} y_0 I_{(-\infty, r_1)}(y_0) & y_0 I_{(r_2, \infty)}(y_0) \\ y_1 I_{(-\infty, r_1)}(y_1) & y_1 I_{(r_2, \infty)}(y_1) \\ \vdots & \vdots \\ y_{T-1} I_{(-\infty, r_1)}(y_{T-1}) & y_{T-1} I_{(r_2, \infty)}(y_{T-1}) \end{pmatrix}; \ \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}.$$

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

$$\mathcal{W}_{(r_1, r_2)} = \frac{\hat{\boldsymbol{\beta}}' (\mathbf{X}'\mathbf{X}) \hat{\boldsymbol{\beta}}}{\hat{\sigma}_u^2}, \tag{3.9}$$

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 $\hat{u}_t^2$, $t = 1, 2, ..., T$, are the residuals obtained from (3.5).

It is well-established that this kind of test suffers from the Davies (1987) problem since threshold parameters r_1 and r_2 are not identified under the null. Most solutions to this problem involve some sort of integrating out unidentified parameters from the test statistic. In the context of threshold autoregressive models the problem has been investigated in Tong (1990) and Hansen (1996). This is usually achieved by calculating test statistics for a grid of possible values of r_1, r_2, and then constructing a summary statistic. The three most common statistics are average, supremum and exponential average statistics defined respectively by

$$\mathcal{W}_{(r_1, r_2)}^{\text{sup}} = \sup_{i \in \#\Gamma} \mathcal{W}_{(r_1, r_2)}^{(i)}, \tag{3.10}$$

$$\mathcal{W}_{(r_1, r_2)}^{ave} = \frac{1}{\#\Gamma} \sum_{i=1}^{\#\Gamma} \mathcal{W}_{(r_1, r_2)}^{(i)}, \tag{3.11}$$

$$\mathcal{W}_{(r_1, r_2)}^{\text{exp}} = \frac{1}{\#\Gamma} \sum_{i=1}^{\#\Gamma} exp\left(\frac{\mathcal{W}_{(r_1, r_2)}^{(i)}}{2} \right), \tag{3.12}$$

where $\mathcal{W}_{(r_1, r_2)}^{(i)}$ 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. In particular, Andrews and Ploberger (1994) have shown that the exponential average test is most powerful in the case where some nuisance parameters are unidentified under the null hypothesis.

We first consider the special case of $r_1 = r_2 = 0$. In this case the three regime SETAR model reduces to the two regime one with known threshold values. Then, (3.8) can be expressed as

$$\Delta \mathbf{y} = \mathbf{X}_0 \boldsymbol{\beta} + \mathbf{u}, \tag{3.13}$$

where

$$\mathbf{X}_0 = \begin{pmatrix} y_0 I_{(-\infty,0)}(y_0) & y_0 I_{(0,\infty)}(y_0) \\ y_1 I_{(-\infty,0)}(y_1) & y_1 I_{(0,\infty)}(y_1) \\ \vdots & \vdots \\ y_{T-1} I_{(-\infty,0)}(y_{T-1}) & y_{T-1} I_{(0,\infty)}(y_{T-1}) \end{pmatrix}.$$

The Wald statistic testing for $\boldsymbol{\beta} = \mathbf{0}$ in (3.13) is now given by

$$\mathcal{W}_{(0,0)} = \frac{\hat{\boldsymbol{\beta}}' \left(\mathbf{X}_0' \mathbf{X}_0\right) \hat{\boldsymbol{\beta}}}{\hat{\sigma}_u^2}, \tag{3.14}$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator of $\boldsymbol{\beta}$ in (3.13), $\hat{\sigma}_u^2 \equiv \frac{1}{T-2}\sum_{t=1}^{T} \hat{u}_t^2$, and $\hat{u}_t^2$ are the residuals obtained from (3.13). In this special case we obtain the following asymptotic distributional result:

Theorem 3.1 *Consider the threshold autoregressive model (3.13) with threshold values $r_1 = r_2 = 0$. Then, the Wald statistic testing for $\boldsymbol{\beta} = \mathbf{0}$ has the following asymptotic null distribution:*

$$\mathcal{W}_{(0,0)} \Rightarrow \frac{\left\{\int_0^1 I_{(-\infty,0)}\left[W(s)\right] \times W(s) dW(s)\right\}^2}{\int_0^1 I_{(-\infty,0)}\left[W(s)\right] \times W(s)^2 ds} + \frac{\left\{\int_0^1 I_{[0,\infty)}\left[W(s)\right] \times W(s) dW(s)\right\}^2}{\int_0^1 I_{[0,\infty)}\left[W(s)\right] \times W(s)^2 ds}, \tag{3.15}$$

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

Proof. See the Appendix. ■

In general such a perfect information about threshold values in TAR model is not feasible, and therefore, the above results is of limited use. However, next theorem shows that the limiting null distribution of our suggested test statistic $\mathcal{W}_{(r_1,r_2)}$ defined by (3.9) is equivalent to that of $\mathcal{W}_{(0,0)}$.

Theorem 3.2 *Under the null $\beta_1 = \beta_1 = 0$, $\mathcal{W}_{(r_1,r_2)}$ converges in probability pointwise to $\mathcal{W}_{(0,0)}$.*

Proof. See the Appendix. ■

The above theorem shows that the asymptotic null distribution of the Wald statistics $\mathcal{W}_{(r_1,r_2)}$ does not depend on the values of nuisance parameters r_1 and r_2.[7] This distributional invariance is due to the random walk nature of the process under the null. More specifically, under the null the random walk process stays within the corridor regime for a proportion of time which goes to zero at rate $T^{-1/2}$ as proved in the Appendix.

To evaluate this theoretical finding we have carried out stochastic simulations and summarised the result in Figure 1, which plots the proportion of time spent in the corridor

[7] The pointwise convergence result is sufficient for the validity of the average and exponential average of the Wald tests. Unfortunately, we do not provide an analytic proof for uniform convergence. Consequently, weak convergence result cannot be established for the supremum of the Wald test. But, simulation results in Section 4 clearly indicate that the supremum of the Wald test may not be valid due to substantial size distortions.

regime by the random process for varying widths of r_1 and r_2. As predicted by theory, this proportion does down at rate $T^{1/2}$ as the sample size increases.

Figure 1 about *here*

We also present the empirical distribution of $\mathcal{W}_{(r_1,r_2)}$ for different values of r_1 and r_2 in Figure 2, which clearly shows that the distribution of the test statistic is invariant to the width of the corridor.[8]

Figure 2 about *here*

We establish the consistency of the test in the following theorem:

Theorem 3.3 *Under the alternative hypothesis $\beta_1 < 0, \beta_2 < 0$, $\mathcal{W}_{(r_1,r_2)}$ diverges to infinity.*

Proof. See the Appendix. ■

Models with intercept and/or linear deterministic trend can be accommodated as follows: First, 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 (3.5), where $\bar{z}$ is the sample mean. In this case the asymptotic distribution is the same as (3.15) 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, $x_t = \mu + \delta t + y_t$, we use the de-meaned and de-trended data $y_t = z_t - \hat{\mu} - \hat{\delta} t$ in (3.5), 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 will refer to the demeaned and demeaned and detrended tests as $\widetilde{\mathcal{W}}_{(r_1,r_2)}$ and $\widehat{\mathcal{W}}_{(r_1,r_2)}$, respectively.

There are a few remarks. The Wald statistic $\mathcal{W}_{(0,0)}$ has well-defined limiting distribution, since threshold values are known to be zero. This result is clearly related to the F-test considered by Enders and Granger (1998), i.e. $F = \mathcal{W}_{(0,0)}/2$. Although they have tabulated critical values of the F statistic, they have not formally derived its asymptotic distribution. Theorem 3.1 provides its theoretical validity. Table 1 below presents the 95% and 99% fractiles of asymptotic null critical values.[9]

Table 1 about here

We also note that Berben and van Dijk (1999) have derived the asymptotic distribution of the F statistic under the assumption of drifting thresholds, in which case the samples are rearranged according to the order statistics of the threshold variable y_{t-1}. Based on this they have tabulated critical values for various values of $(\tau_1, \tau_2) \in (0,1)$. Such an extension might help improve the finite sample performance of the Wald tests. See also Tsay (1998).

We extend to the more general case where the errors in (3.5) are serially correlated. Following Dickey and Fuller (1979) we then consider the following nonlinear $ADF\,(p)$ regression:

$$\Delta y_t = \beta_1 y_{t-1} I_{(-\infty,r_1)}(y_{t-1}) + \beta_2 y_{t-1} I_{[r_2,\infty)}(y_{t-1}) + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + u_t, \tag{3.16}$$

[8]These figure have been constructed using stochastic simulations with 50,000 replications and a sample size of 10,000. We have generated random walk processes with zero initial value and standard normal errors and used them to estimate the relevant proportions and test statistics.

[9]Notice that this result is the same as given in Table 1 for $T = 1,000$ of Enders and Granger (1998).

where $u_t \sim iid\,(0, \sigma_u^2)$. The statistic for testing the null hypothesis $\beta_1 = \beta_2 = 0$ in this set up is again given by the Wald statistic with thresholds r_1 and r_2 in (3.16).

Theorem 3.4 *The asymptotic null distribution of the Wald statistics testing for $\beta_1 = \beta_2 = 0$ in (3.16) are equivalent to that obtained under the case where the underlying disturbances are not serially correlated.*

Proof. See the Appendix. ■

4 Monte Carlo Study

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

$$y_t = y_{t-1} + u_t, \tag{4.17}$$

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

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{4.18}$$

where $\varepsilon_t \sim N\,(0, 1)$ and $\rho = 0.3$ is considered.

In the next two sets of experiments we investigate the power performance of the tests. For these experiments the data is generated by

$$y_t = \begin{cases} \phi_1 y_{t-1} + u_t & if\ y_{t-1} < r_1 \\ \phi_0 y_{t-1} + u_t & if\ r_1 \leq y_{t-1} \leq r_2 \\ \phi_2 y_{t-1} + u_t & if\ y_{t-1} > r_2 \end{cases}, \ t = 1, 2, ..., T, \tag{4.19}$$

where $u_t \sim N\,(0, 1)$. In Experiment 2 we set $\phi_0 = 1$. We consider the symmetric adjustments in Experiment 2(a) with $\phi_1 = \phi_2 = 0.9$, whereas we examine asymmetric adjustments in Experiment 2(b) with $\phi_1 = 0.85$ and $\phi_2 = 0.95$. Experiment 3 investigates the power against geometric ergodic process with explosive corridor regime with $\phi_0 = 1.3$. More specifically, we consider symmetric adjustments in Experiment 3(a) with $\phi_1 = \phi_2 = 0.9$, whereas we examine asymmetric adjustments in Experiment 3(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, $\mathcal{W}^{\text{sup}}_{(r_1,r_2)}$, $\mathcal{W}^{\text{ave}}_{(r_1,r_2)}$ and $\mathcal{W}^{\text{exp}}_{(r_1,r_2)}$, defined by (3.10) - (3.12), and the standard linear DF test. The normal errors are obtained using the GAUSS random number generator. For all 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 250 are considered. Empirical size and power of the tests are evaluated at the 5% nominal level. In all experiments we consider both the demeaned and the demeaned and detrended process. To control for the grid of threshold parameters we select six different sets of values from 0.15 to 3.90 and -0.15 to -3.90, respectively, at step of 0.75.[10] 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.

The results for the first set of experiments are summarized in Tables 2 and 3. As a benchmark, Table 2 gives empirical size of the tests when the underlying DGP is the simple random walk process with serially uncorrelated errors. All tests except for the $\mathcal{W}^{\text{sup}}_{(r_1,r_2)}$ tend to have correct sizes. Table 3 summarizes the results when the DGP follows the unit root process with AR(1) serially correlated errors. Here to compute the test statistics we use the correct ADF(1) regression with one augmentation, see (3.16). We now find qualitatively similar results to those obtained previously. Since the $\mathcal{W}^{\text{sup}}_{(r_1,r_2)}$ test seems to overreject significantly even for $T = 250$, hence we do not consider its power performance in what follows.

Table 2 about here

Table 3 about here

The results of the power of the test for the second set of experiments are summarized in Tables 4 and 5. Table 4 presents the relative power performance when the threshold autoregressive parameter in outer regimes is symmetric at 0.9. When the threshold band is relatively small, the DF test is slightly more powerful than the Wald tests. But, as expected and shown by Pippenger and Goering (1993), the power of DF test decreases significantly with threshold values. The decrease in power of the Wald tests is much slower especially for relatively large sample, e.g. $T = 250$. Therefore, the power of the Wald tests eventually dominates the DF test. In addition the power of the $\widetilde{\mathcal{W}}^{\text{exp}}_{(r_1,r_2)}$ test is much more favorable than the test $\widetilde{\mathcal{W}}^{\text{ave}}_{(r_1,r_2)}$. For example, when $(r_1, r_2) = (-3.9, 3.9)$ and $T = 250$, the power of $\widetilde{\mathcal{W}}^{\text{exp}}_{(r_1,r_2)}$, $\widetilde{\mathcal{W}}^{\text{ave}}_{(r_1,r_2)}$ and DF tests are 0.831, 0.736 and 0.515, respectively. Table 5 gives the results for asymmetric threshold autoregressive parameters set to 0.85 and 0.95, respectively. When compared to the symmetric case, we find that all of the tests are now more powerful. The power gain is much more significant for the Wald tests especially for wide corridor regimes, which also confirms finding by Enders and Granger (1998). In particular, for the demeaned data the $\widetilde{\mathcal{W}}^{\text{exp}}_{(r_1,r_2)}$ is now remarkably more powerful than the DF test for all cases considered. Again as threshold band widens, the power loss of the DF test is much faster.

Table 4 about here

Table 5 about here

Finally, the results for the third set of experiments with explosive corridor regime are summarized in Tables 6 and 7. Unlike linear models where explosive behaviour is of little economic interest, in nonlinear models there is no strong *a priori* reason to rule out the

[10]We also find via simulation that the processes have spent at least 5% of the time in each of the outer regimes even for the largest threshold parameter values considered.

explosive behaviour in corridor regime. As discussed in Section 2, such processes are still compatible with global geometric ergodicity.[11] As before, all tests are more powerful for asymmetric case than for symmetric case. Interestingly, the power of the Wald tests increases almost monotonically as the threshold band widens, while the power of the DF test first decreases and then increases. For almost all cases considered the power of the Wald tests, especially the $\widetilde{\mathcal{W}}^{\exp}_{(r_1,r_2)}$ test, dominates that of the DF test.

Table 6 about here

Table 7 about here

Overall the suggested Wald statistics have reasonably good size and power properties. In particular, simulation results clearly recommends the use of the $\widetilde{\mathcal{W}}^{\exp}_{(r_1,r_2)}$ test in finite samples.

5 Concluding Remarks and Further Issues

The investigation of nonstationarity in conjunction with the threshold autoregressive modelling has recently assumed a prominent role in econometric study. It is clear that mistaking a stable nonlinear process for a nonstationary one can be misleading both in impulse response and forecasting analysis. In this paper we have developed the simple direct unit root test which is designed to have the power against the geometrically ergodic process defined explicitly in the context of TAR model with three regimes. The tests proposed here are shown to have better power performance than the standard Dickey Fuller tests that ignores the threshold nature of the three regime alternative. Although our suggested test is based on the univariate model, it can also be regarded as direct tests of linear cointegration against nonlinear threshold cointegration, assuming that the process under investigation can be regarded as a linear combination of the underlying nonstationary variables (or error correction term) with known cointegrating parameters, e.g. real exchange rates.

There a couple of further research agenda. First, an extension of this paper to testing the null of linear no cointegration directly against the alternative of threshold cointegration in the multivariate regression context with unknown cointegrating parameters would be useful, but the consequent theoretical developments will require further studies. In this case main difficulty lies in the fact that both cointegrating parameters and threshold parameters are not identified under the null, which complicates the inference problem further. Second, it might be possible to find an alternative testing procedure based on the arranged regression along similar lines to Tsay (1998) and Berben and van Dijk (1999), which is likely to boost the power of the tests.

[11] The question whether or not the threshold process is subject to explosive behaviour in corridor regime is an empirical one. Then, we also need to address the estimation of such a model and the extent of small sample biases that may afflict explosive autoregressive coefficients.

Table1. Asymptotic null critical values

	$\mathcal{W}_{(r_1,r_2)}$	$\widetilde{\mathcal{W}}_{(r_1,r_2)}$	$\widehat{\mathcal{W}}_{(r_1,r_2)}$
95%	7.49	9.04	12.16
99%	10.94	12.64	16.28

Table 2. Size of alternative tests for Experiment 1(a) (*iid* errors)

Demeaned case				
T	$\widetilde{\mathcal{W}}^{\text{ave}}_{(r_1,r_2)}$	$\widetilde{\mathcal{W}}^{\text{exp}}_{(r_1,r_2)}$	$\widetilde{\mathcal{W}}^{\text{sup}}_{(r_1,r_2)}$	DF
100	.041	.062	.182	.061
250	.060	.065	.205	.052
Demeaned and detrended case				
T	$\widehat{\mathcal{W}}^{\text{ave}}_{(r_1,r_2)}$	$\widehat{\mathcal{W}}^{\text{exp}}_{(r_1,r_2)}$	$\widetilde{\mathcal{W}}^{\text{sup}}_{(r_1,r_2)}$	DF
100	.035	.051	.113	.090
250	.039	.052	.163	.055

Table 3. Size of alternative tests for Experiment 2(a) (serially correlated errors)

Demeaned case				
T	$\widetilde{\mathcal{W}}^{\text{ave}}_{(r_1,r_2)}$	$\widetilde{\mathcal{W}}^{\text{exp}}_{(r_1,r_2)}$	$\widetilde{\mathcal{W}}^{sup}_{(r_1,r_2)}$	ADF
100	.037	.057	.189	.046
250	.039	.059	.219	.052
Demeaned and detrended case				
T	$\widehat{\mathcal{W}}^{\text{ave}}_{(r_1,r_2)}$	$\widehat{\mathcal{W}}^{\text{exp}}_{(r_1,r_2)}$	$\widetilde{\mathcal{W}}^{sup}_{(r_1,r_2)}$	ADF
100	.024	.042	.148	.059
250	.038	.055	.151	.034

Insert Table 4 here

Insert Table 5 here

Insert Table 6 here

Insert Table 7 here

Insert Figure 1 here

Insert Figure 2 here

A Appendix

This Appendix provides mathematical proofs of Theorems of the paper.

A.1 Proof of Theorem 1.

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

$$\mathcal{W}_{(0,0)} = \frac{1}{\hat{\sigma}_u^2}\hat{\boldsymbol{\beta}}'\left(\mathbf{X}_0'\mathbf{X}_0\right)\hat{\boldsymbol{\beta}} = \frac{1}{\hat{\sigma}_u^2}\mathbf{u}'\mathbf{X}_0\left(\mathbf{X}_0'\mathbf{X}_0\right)^{-1}\mathbf{X}_0'\mathbf{u}.$$

Hence,

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

Since the function $g_1(z) = I_{(-\infty,0)}(z)z$ is are continuous, by the continuous mapping Theorem we obtain

$$I_{(-\infty,0)}(y_{t-1})y_{t-1} = I_{(-\infty,0)}\left(\frac{1}{\sigma_u\sqrt{T}}y_{t-1}\right)\frac{1}{\sigma_u\sqrt{T}}y_{t-1} \Rightarrow I_{(-\infty,0)}\left[W(s)\right]W(s).$$

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

$$\frac{1}{T}\sum_{t=1}^{T} I_{(-\infty,0)}\left(y_{t-1}\right)y_{t-1}u_t \Rightarrow \sigma_u^2\int_0^1 I_{(-\infty,0)}\left[W(s)\right]\times W(s)dW\left(s\right),$$

$$\frac{1}{T^2}\sum_{t=1}^{T} I_{(-\infty,0)}\left(y_{t-1}\right)y_{t-1}^2 \Rightarrow \sigma_u^2\int_0^1 I_{(-\infty,0)}\left[W(s)\right]\times W(s)^2 ds.$$

Similarly, we have

$$\frac{1}{T}\sum_{t=1}^{T} I_{(0,\infty)}\left(y_{t-1}\right)y_{t-1}u_t \Rightarrow \sigma_u^2\int_0^1 I_{(0,\infty)}\left[W(s)\right]\times W(s)dW\left(s\right),$$

$$\frac{1}{T^2}\sum_{t=1}^{T} I_{(0,\infty)}\left(y_{t-1}\right) y_{t-1}^2 \Rightarrow \sigma_u^2 \int_0^1 I_{(0,\infty)}\left[W(s)\right] \times W(s)^2 ds.$$

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 (3.15).

A.2 Proof of Theorem 2.

Here we need to show that

$$\mathcal{W}_{(r_1,r_2)} \xrightarrow{p} \mathcal{W}_{(0,0)}, \tag{A.1}$$

uniformly over for $(r_1, r_2) \in \mathbf{R}$ where $\mathbf{R}$ is a closed, bounded subset of $\mathbb{R}^2$. First, pointwise convergence in probability of $\mathcal{W}_{(r_1,r_2)}$ to $\mathcal{W}_{(0,0)}$ requires:

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

$$\frac{1}{T}\sum_{t=1}^{T} I\left\{_{[0,\infty)}(y_{t-1})\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2 - I_{[r_2,\infty)}(y_{t-1})\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2\right\} \xrightarrow{p} 0, \tag{A.3}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\left\{I_{(-\infty,0)}(y_{t-1})y_{t-1}u_t - I_{(-\infty,r_1)}(y_{t-1})y_{t-1}u_t\right\} \xrightarrow{p} 0. \tag{A.4}$$

$$\frac{1}{T}\sum_{t=1}^{T}\left\{I_{[0,\infty)}(y_{t-1})y_{t-1}u_t - I_{[r_2,\infty)}(y_{t-1})y_{t-1}u_t\right\} \xrightarrow{p} 0. \tag{A.5}$$

Considering for example the term in (A.4),

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T}\left[I_{[0,\infty)}(y_{t-1})\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2 - I_{[r_2,\infty)}(y_{t-1})\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2\right] \\ = & \frac{1}{T}\sum_{t=1}^{T} I_{[0,r_2)}(y_{t-1})\left(\frac{1}{\sqrt{T}}y_{t-1}\right)^2. \end{aligned} \tag{A.6}$$

Standard analysis of random walks (see, for example, Feller (1957)) indicates that for $(r_1, r_2) \in \mathbf{R}$, the number of nonzero terms in the summation in (A.6) is of order $\sqrt{T}$, asymptotically. As each of these terms is $O_p(1)$, the final expression in (A.6) tends to zero in probability. Similar analysis provides the desired result for other terms.

A.3 Proof of Theorem 3.

Under the alternative of $\beta_1 < 0$ and $\beta_2 < 0$, we have

$$\begin{aligned} \mathcal{W}_{(r_1,r_2)} &= \frac{\hat{\boldsymbol{\beta}}'\left(\mathbf{X}'\mathbf{X}\right)\hat{\boldsymbol{\beta}}}{\hat{\sigma}_u^2} = \frac{\left(\Delta\mathbf{y}'\mathbf{X}\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\Delta\mathbf{y}\right)}{\hat{\sigma}_u^2} \\ &= \frac{1}{\hat{\sigma}_u^2}\left\{\boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta} + 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{u} + \left(\mathbf{u}'\mathbf{X}\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{u}\right)\right\}. \end{aligned}$$

It is easily seen that $\boldsymbol{\beta}'(\mathbf{X}'\mathbf{X})\boldsymbol{\beta}$ is the leading term, which diverges to infinity at rate T, since $\mathbf{X}$'s are asymptotically stationary. Notice also that even if the corridor regime is explosive consistency holds by the same argument.

A.4 Proof of Theorem 4.

(3.16) 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 $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_p)'$ and

$$\underset{T\times p}{\mathbf{Z}} = (\Delta\mathbf{y}_{-1}, ..., \Delta\mathbf{y}_{-p}),\ \Delta\mathbf{y}_{-i} = (\Delta y_{-i+1}, ..., \Delta y_{T-i}),\ i = 1, ..., p.$$

Then, the Wald statistic is given by

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

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.7), 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,

$$\mathbf{x}_1 = \begin{pmatrix} y_0 I_{(-\infty,r_1)}(y_0) \\ y_1 I_{(-\infty,r_1)}(y_1) \\ \vdots \\ y_{T-1} I_{(-\infty,r_1)}(y_{T-1}) \end{pmatrix};\ \mathbf{x}_2 = \begin{pmatrix} y_0 I_{(r_2,\infty)}(y_0) \\ y_1 I_{(r_2,\infty)}(y_1) \\ \vdots \\ y_{T-1} I_{(r_2,\infty)}(y_{T-1}) \end{pmatrix};$$

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,

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

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

$$\mathcal{W}_{(0,0)} = \frac{(\mathbf{u}'\mathbf{M}_T\mathbf{X}_0)(\mathbf{X}_0'\mathbf{M}_T\mathbf{X}_0)^{-1}(\mathbf{X}_0'\mathbf{M}_T\mathbf{u})}{\hat{\sigma}_u^2} \qquad \text{(A.9)}$$
$$\frac{1}{\hat{\sigma}_u^2}\left\{\mathbf{u}'\mathbf{x}_{01}(\mathbf{x}_{01}'\mathbf{x}_{01})^{-1}\mathbf{x}_{01}'\mathbf{u} + \mathbf{u}'\mathbf{x}_{02}(\mathbf{x}_{02}'\mathbf{x}_{02})^{-1}\mathbf{x}_{02}'\mathbf{u}\right\} + o_p(1),$$

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

$$\mathbf{x}_{01} = \begin{pmatrix} y_0 I_{(-\infty,0)}(y_0) \\ y_1 I_{(-\infty,0)}(y_1) \\ \vdots \\ y_{T-1} I_{(-\infty,0)}(y_{T-1}) \end{pmatrix}; \ \mathbf{x}_{02} = \begin{pmatrix} y_0 I_{(0,\infty)}(y_0) \\ y_1 I_{(0,\infty)}(y_1) \\ \vdots \\ y_{T-1} I_{(0,\infty)}(y_{T-1}) \end{pmatrix};$$

Furthermore,

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

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

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

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

where σ_{LR}^2 is the long-run variance of Δy_t. Using these results in (A.9), we obtain

$$\mathcal{W}_{(0,0)} \Rightarrow \frac{\left\{\int_0^1 I_{(-\infty,0)}[W(s)] \times W(s)dW(s)\right\}^2}{\int_0^1 I_{(-\infty,0)}[W(s)] \times W(s)^2 ds} + \frac{\left\{\int_0^1 I_{[0,\infty)}[W(s)] \times W(s)dW(s)\right\}^2}{\int_0^1 I_{[0,\infty)}[W(s)] \times W(s)^2 ds},$$

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 2 we can establish that

$$\mathcal{W}_{(r_1,r_2)} \xrightarrow{p} \mathcal{W}_{(0,0)}.$$

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] Beaudry, P. and G. Koop (1993), "Do Recessions Permanently Change Output?" *Journal of Monetary Economics, 31*, 149-164.

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

[5] Caner, M. and B.E. Hansen (1998), "Threshold Autoregression with a Near Unit Root," unpublished manuscript, University of Wisconsin.

[6] 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.

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

[8] 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.

[9] 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.

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

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

[12] Granger, C.W.J. and T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships*. Oxford University Press: Oxford.

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

[14] Hansen, B.E. and B. Seo (2000), "Testing for Threshold Cointegration," unpublished manuscript, University of Wisconsin.

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

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

[17] Kapetanios, G., Y. Shin and A. Snell (2000), "Testing for a Unit Root against Nonlinear STAR models," unpublished manuscript, University of Edinburgh (www.ed.ac.uk/˜shiny/).

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

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

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

[21] 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.

[22] 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.

[23] 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.

[24] Rose, A.K. (1988), "Is the Real Interest Rate Stable?" *Journal of Finance*, 43, 1095-1112.

[25] 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.

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

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

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