Correlation robust threshold unit root tests

Robert M. de Jong* Chien-Ho Wang† Youngsoo Bae‡

September 19, 2007

Abstract

This paper proposes a three-regime threshold unit root (TUR) test that is robust against errors that are not i.i.d., but have fading memory properties. We consider a test statistic that is obtained by optimizing over the threshold parameter, which is assumed to be unknown and unidentified under the null hypothesis. Similar to the Phillips-Perron test, a bandwidth-type sequence is used to remove the consequences of correlation in the errors for the limit distribution of the test statistic, and the limit distribution of our test statistic does therefore not depend on nuisance parameters. The test is applied to test for purchasing power parity.

JEL numbers: C22; C32

Keywords: unit root test, threshold unit root model, nonlinearity

*Corresponding author: Robert M. de Jong, Department of Economics, Ohio State University, 429 Arps Hall, Columbus, OH 43210, USA, Email dejong@econ.ohio-state.edu.

†Department of Economics, National Taipei University, Taipei City 104, Taiwan, Email wangchi3@mail.ntpu.edu.tw.

‡Department of Economics, LeHigh University, 621 Taylor Street, Bethlehem, PA 18015, USA, email yob206@lehigh.edu

1 Introduction

Economic time series data often show some sudden changes, as a result of an external shock. It is generally believed that linear time series models cannot capture such a structural change. This is the central criticism of unit root tests based on linear time series models, such as the Dickey and Fuller (1979) and Phillips and Perron (1988) tests. One statistical model that attempts to capture such a sudden structural change in different regimes is the threshold autoregressive model developed by Tong (1990). The threshold autoregressive model captures regime switching based on the lagged values of the variables. This is a very attractive property for economists, and the topic of unit root testing within the framework of the threshold autoregressive model has received quite some attention over recent years. A central analytical issue with tests for the unit root hypothesis in this setting is the fact that a Davies problem typically occurs in this setting; i.e., parameter(s) can be unidentified under the null hypothesis. The two solutions that have been proposed to tackle this issue are (i) use a Taylor series expansion approach and (ii) optimize the test statistic over the parameters that are unidentified under the null hypothesis.

The first papers to investigate unit root structure in threshold autoregressive model was González and Gonzalo (1997). They used the threshold autoregressive model as their framework. González and Gonzalo consider a t-type test in the context of a threshold unit root model, and they assume the threshold value to be known and fixed, which means there is no Davies problem in their setting. Of course, generally in economics time series, threshold values can be unknown. Berben and van Dijk (1999) considered unknown threshold values and the test statistic that was suggested optimizes over the parameters that are unidentified under the null, but their paper appears to be analytically incomplete. Caner and Hansen (2001) used the lagged difference of the dependent variable as the threshold variable, and assume the threshold value to be unknown. In their setting, the threshold variable is stationary, and we should expect that for many interesting time series, one would want the level of the process to be the threshold value, rather than the lagged difference. Kapetanios and Shin (2002) proposed threshold unit root tests against nonlinear stationary alternatives. They used a Taylor series expansion approach for their paper. The drawback of this approach is that in their setting, the Taylor series expansion will not become more accurate in some asymptotic sense.

Bec et al. (2004) consider LM type test statistics that optimize over a space that grows with sample size, which is achieved by optimizing over a space that is bounded by percentiles of y_t. Their test statistic is pivotal. Bec et al. (2006) consider tests similar to those in Bec et al. (2004). They thoroughly analyze the properties of pivotal test statistics obtained both by optimization over a fixed and a sample size dependent parameter space. The only paper to consider non-i.i.d. disturbances appears to be Seo (2006). Seo (2006) considers a statistic that optimizes over the unidentified parameter using a fixed (i.e. not sample size dependent) parameter space. The limit distribution of Seo (2006)'s statistic depends on various nuisance parameters, including the short- and long-run variance parameters, and he proposes to use the residual-based block bootstrap for obtaining critical values.

After writing this paper, the authors have learned that Shintani and Park (2005) have simultaneously proposed a test that is based on an ADF approach and that is also capable of dealing with correlated errors. This paper takes a different approach that is more reminiscent of the approach taken in the Phillips-Perron test.

In this paper, we will propose a test statistic that optimizes over the parameter that is unidentified under the null, explicitly allows for weakly dependent errors, and is asymptotically pivotal. Therefore, we view our paper as the natural follow-up to Bec et al. (2004), Bec et al. (2006), and Seo (2006). We obtain a pivotal test statistic by using a bandwidth-type sequence k_n, which is reminiscent of the Phillips-Perron test. The only condition needed on this bandwidth-type sequence for establishing the limit distribution under the null will be that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. We consider the threshold unit root model

$$\Delta y_t = \begin{cases} u_t & \text{if } c_1 \leq y_{t-1} \leq c_2 \\ \varphi_1(y_{t-1} - c_1) + u_t & \text{if } y_{t-1} < c_1 \\ \varphi_2(y_{t-1} - c_2) + u_t & \text{if } y_{t-1} > c_2 \end{cases} \tag{1}$$

where $\varphi_1, \varphi_2 \in (-2, 0)$. We will assume that the threshold values c_1 and c_2 are not known a priori. That is, we assume that an AR(1) model with intercept holds in the stationary outer regimes, but we assume that the response function is continuous and has no jumps at c_1 and/or c_2. The methodology of this paper however should be easily transferrable to a range of other threshold models.

This paper is organized as follow. In Section 2, we will derive the asymptotic results that will allow us to construct our test statistic. Using the results of Section 2, Section 3 establishes the limit behavior of the t-statistics obtained from a regression of Δy_t on $(y_{t-1}-c_1)I(y_{t-1}<c_1)$ and $(y_{t-1}-c_2)I(y_{t-1}>c_2)$, optimized over a set of possible values for c_1 and c_2. In Section 3, we conduct a simulation experiment to illustrate the performance of our test. Section 4 contains an empirical application of our test. The conclusion will be found in Section 5. Critical values for our test are tabulated in Appendix 1. All proofs are in Appendix 2.

2 Main results

The reasoning behind our test is as follows. We can rewrite the model of Equation (1) as

$$\Delta y_t = \varphi_1(y_{t-1}-c_1)I(y_{t-1}<c_1) + \varphi_2(y_{t-1}-c_2)I(y_{t-1}>c_2) + u_t. \tag{2}$$

Therefore, under the null hypothesis $\varphi_1 = \varphi_2 = 0$, c_1 and c_2 will be unidentified. Therefore, we suggest to consider a modification in the spirit of the Phillips-Perron test of

$$\inf_{c_1,c_2} \hat{t}_{H_0:\varphi_1=0}(c_1,c_2) \quad \text{and} \quad \inf_{c_1,c_2} \hat{t}_{H_0:\varphi_2=0}(c_1,c_2), \tag{3}$$

where $\hat{t}_{H_0:\varphi_1=0}(c_1,c_2)$ and $\hat{t}_{H_0:\varphi_2=0}(c_1,c_2)$ denotes the t-statistics that would be obtained from a regression of Δy_t on $(y_{t-1}-c_1)I(y_{t-1}<c_1)$ and $(y_{t-1}-c_2)I(y_{t-1}>c_2)$ if c_1 and c_2 were known. Below, we will concentrate on $\hat{t}_{H_0:\varphi_2=0}(c_1,c_2)$, but we note that the treatment of $\hat{t}_{H_0:\varphi_1=0}(c_1,c_2)$ is completely analogous. A crucial aspect of determining the behavior of $\hat{t}_{H_0:\varphi_2=0}(c_1,c_2)$ will be to find the weak convergence limit of a modification $H_n^*(\cdot)$ of

$$G_n(x_2) = n^{-1/2}\sum_{t=2}^{n} u_t(n^{-1/2}y_{t-1}-x_2)I(n^{-1/2}y_{t-1}>x_2) \tag{4}$$

under the null hypothesis; this process will be crucial for determining the analytical behavior of the the numerator of $\hat{t}_{H_0:\varphi_2=0}(c_1,c_2)$.

Finding the limit process for $G_n(\cdot)$ is likely to be complicated and may result in a limit distribution that involves a nuisance correlation parameter. In de Jong (2002), it is shown that for continuously differentiable $T(\cdot)$,

$$n^{-1/2}\sum_{t=2}^{n} u_t T(n^{-1/2}\sum_{j=1}^{t-1} u_j) \xrightarrow{d} \int_0^1 T(U(r))dU(r) + \Lambda \int_0^1 T'(U(r))dr \tag{5}$$

where Λ denotes a long-run variance parameter; this result suggests that the limit process for $G_n(\cdot)$ may not be nuisance-parameter free. An extra complication arises from the fact that we need to prove weak convergence, rather than pointwise convergence in our setting. Therefore, we propose to modify our statistic in a way that is reminiscent of the Phillips-Perron modification of the Dickey-Fuller test, and we suggest to consider $H_n^*(\cdot)$ (as defined below) instead of $G_n(\cdot)$. If u_t is i.i.d. and has a finite variance, it will be show below in Theorem 2 that $H_n^*(x) - G_n(x)$ converges to 0 uniformly in x. This means that the difference $H_n^*(x) - G_n(x)$ can be viewed as capturing the effect that the weak dependence of u_t has on the limit process for $G_n(\cdot)$.

The Phillips-Perron style modification of $\hat{t}_{H_0:\varphi_2=0}(c_1, c_2)$ is as follows. Let $r_j = j/k_n$ and $n_j = [jn/k_n]$ for $j \in [0, k_n]$, where k_n is an integer-valued bandwidth-type sequence that is nondecreasing in n and such that $k_n \to \infty$ and $k_n/n \to 0$. Define

$$U_n(r) = n^{-1/2} y_{[rn]} \tag{6}$$

and

$$H_n^*(x_2) = \sum_{j=1}^{k_n} (U_n(r_j) - U_n(r_{j-1}))(U_n(r_{j-1}) - x_2) I(U_n(r_{j-1}) > x_2) \tag{7}$$

where empty summations are defined as zero; this convention will ensure that $U_n(r)$ is well-defined for $r \in [0, n^{-1})$. Also, we assume $y_0 = 0$ in the proofs, which can be relaxed trivially to the assumption that y_0 is an arbitrary random variable with a finite variance. We assume that u_t is a linear process

$$u_t = \sum_{k=0}^{\infty} \phi_k \varepsilon_{t-k}, \tag{8}$$

and the variances of ε_t and u_t will be denoted by σ_ε^2 and σ^2 respectively. We will make the following assumptions regarding u_t:

Assumption 1

(a) $\sum_{k=0}^{\infty} k|\phi_k| < \infty$ *and* $E|\varepsilon_t|^p < \infty$ *for some* $p > 2$.

(b) *The distribution of* ε_t *is absolutely continuous with respect to the Lebesgue measure and has characteristic function* $\psi(s)$ *for which* $\lim_{|s|\to\infty} |s|^\eta \psi(|s|) = 0$ *for some* $\eta > 0$.

Under the above assumptions, $U_n(\cdot) \Rightarrow U(\cdot)$ and $U(r) = \lambda W(r)$, where $W(r)$ denotes Brownian motion and $\lambda > 0$.

For $H_n^*(\cdot)$, we have the following result.

Theorem 1 *Let* k_n *denote an integer-valued positive sequence that is nondecreasing in* n *and such that* $k_n \to \infty$ *and* $k_n n^{-1} \to 0$. *Then under Assumption 1 and* H_0, *for all* $K > 0$,

$$H_n^*(x_2) \Rightarrow \int_0^1 (U(r) - x_2) I(U(r) > x_2) dU(r) \tag{9}$$

on $[-K, K]$.

The result that $H_n^*(x_2) - G_n(x_2)$ is asymptotically negligible for i.i.d. u_t is stated formally below.

Theorem 2 *Let* k_n *denote an integer-valued positive sequence that is nondecreasing in* n *and such that* $k_n \to \infty$ *and* $k_n n^{-1} \to 0$. *Then under Assumption 1 and* H_0, *if* $\phi_0 = 1$ *and* $\phi_j = 0$ *for all* $j \geq 1$,

$$\sup_{x_2 \in \mathbb{R}} |H_n^*(x_2) - G_n(x_2)| = o_p(1). \tag{10}$$

3 Test statistic

Based on Theorem 1, we can now find the limit distribution of our modification of

$$\inf_{(c_1,c_2)\in C} \hat{t}_{H_0:\varphi_2=0}(c_1, c_2). \tag{11}$$

We will choose $C = \{(c_1, c_2) \in \mathbb{R}^2 : \underline{c}_{1n} \leq c_1 \leq c_2 \leq \overline{c}_{2n}\}$, and choose for $\underline{c}_{1n}$ and $\overline{c}_{2n}$ the β- and $(1-\beta)$-percentile of y_t respectively. This idea of a random upper bound to the parameter space seems to have been pioneered in Bec et al. (2004) and is also used in Bec et al. (2004). Under the assumptions made, $(n^{-1/2}\underline{c}_{1n}, n^{-1/2}\overline{c}_{2n})$ converges in distribution under the null, and the limit random variable will be denoted by $(\lambda \underline{d}_\beta, \lambda \overline{d}_\beta)$. Therefore, to determine the limit distribution of our statistic under the null hypothesis, in our proofs, we will consider $(x_1, x_2) = (n^{-1/2}c_1, n^{-1/2}c_2)$ everywhere and optimize the statistic over $(x_1, x_2) \in X$, where $X = \{(x_1, x_2) : n^{-1/2}\underline{c}_{1n} \leq x_1 \leq x_2 \leq n^{-1/2}\overline{c}_{2n}\}$.

Let

$$B_n(x_2) = n^{-1} \sum_{t=2}^{n} (n^{-1/2}y_{t-1} - x_2)^2 I(n^{-1/2}y_{t-1} > x_2) \tag{12}$$

and

$$s_n^2(x_1, x_2) = (n-2)^{-1} \sum_{t=2}^{n} \hat{u}_t(x_1, x_2)^2, \tag{13}$$

where $\hat{u}_t(x_1, x_2)$ denotes the residual from a linear regression of Δy_t on

$$(n^{-1/2}y_{t-1} - x_1)I(n^{-1/2}y_{t-1} < x_1) \quad \text{and} \quad (n^{-1/2}y_{t-1} - x_2)I(n^{-1/2}y_{t-1} > x_2).$$

Below, let

$$\hat{\sigma}_n^2 = n^{-1} \sum_{t=2}^{n} (\Delta y_t)^2 \tag{14}$$

and let $\hat{\lambda}_n^2$ denotes a consistent HAC estimator of the long-run variance of Δy_t, viz.

$$\hat{\lambda}_n^2 = n^{-1} \sum_{t=2}^{n} \sum_{s=2}^{n} \Delta y_t \Delta y_s k((t-s)/\gamma_n) \tag{15}$$

for a bandwidth sequence γ_n such that $\gamma_n n^{-1} + \gamma_n^{-1} \to 0$ and a kernel function $k(\cdot)$ such that $k(0) = 1$, $k(\cdot)$ is continuous at zero and $\sup_{x\geq 0} |k(x)| < \infty$, $\int_0^\infty \overline{k}(x)dx < \infty$ where $\overline{k}(x) = \sup_{y\geq x} |k(y)|$. It then follows from Jansson (2002) that under the null of a unit root in y_t and under Assumption 1, $|\hat{\lambda}_n - \lambda| = o_p(1)$.

Using these results, we can now establish the following theorem.

Theorem 3 *Let k_n denote an integer-valued positive sequence that is nondecreasing in n and such that $k_n \to \infty$ and $k_n n^{-1} \to 0$. Then under Assumption 1 and H_0,*

$$J_1 = \inf_{(x_1,x_2)\in X} (\hat{\sigma}_n/\hat{\lambda}_n) \frac{H_n^*(x_2)}{s_n(x_1,x_2)\sqrt{B_n(x_2)}}$$

$$\xrightarrow{d} \inf_{x_2\in[\underline{d}_\beta,\overline{d}_\beta]} \frac{\int_0^1 (W(r)-x_2)I(W(r)>x_2)dW(r)}{\sqrt{\int_0^1 (W(r)-x_2)^2 I(W(r)>x_2)dr}} = D_\beta. \tag{16}$$

The test statistic above is asymptotically pivotal. Table 1 gives critical values of D_β for $\beta = 0.10$ and $\beta = 0.05$. Similarly to the Dickey-Fuller and Phillips-Perron tests, we reject the null hypothesis for values of the test statistic smaller than the $(100\alpha)\%$-percentile for significance level α.

It is well-known that the threshold unit root model of Equation (1) will be stationary and strong mixing if the u_t are i.i.d. and $-2 < \varphi < 0$. This follows, for example, from the results in Doukhan (1994), p. 102, Theorem 7. However, as Seo (2006) points out, there appear to be no general results that establish weak dependence or stationarity properties of the threshold model if the errors u_t are weakly dependent themselves. The issue appears to be that results such as the ones of Doukhan use results from Markov chain theory that have no immediate equivalents if errors are not i.i.d. . Therefore, while we conjecture that under mixing and stationarity assumptions on u_t, y_t as generated in the threshold unit root model of Equation (1) will display weak dependence and/or stationarity properties too, we have no proof for this conjecture.

For studying the alternative that y_t is strictly stationary, we need to make the following assumption regarding y_t.

Assumption 2 *Either*

1. $y_t = \sum_{k=0}^{\infty} \phi_k \varepsilon_{t-k}$*, where ε_t satisfies the conditions of Assumption 1 and ε_t is strictly stationary; or*

2. *y_t is a strictly stationary sequence of strong mixing random variables and $Ey_t^4 < \infty$ and the strong mixing numbers $\alpha(\cdot)$ for y_t satisfy $\sum_{j=1}^{\infty} j^2 \alpha(j)^{(\nu-1)/\nu} < \infty$ for some $\nu > 1$.*

Using Assumption 2, we can obtain the following theorem that shows the consistency of our test against stationary alternatives. Let $\tilde{m}_n$ denote the sample median of y_t.

Theorem 4 *Assume that Assumption 2 hold and that for all $a \geq 1$,*

$$E\Delta y_t(y_{t-a} - \tilde{\mu})I(y_{t-a} > \tilde{\mu}) < 0 \tag{17}$$

and that $\tilde{m}_n \xrightarrow{p} \tilde{\mu}$ for some unique population median $\tilde{\mu}$. Also, assume that n/k_n is integer-valued. Then for some $L > 0$,

$$P(n^{1/2}k_n^{-1} \inf_{(x_1,x_2)\in X} (\hat{\sigma}_n/\hat{\lambda}_n)\frac{H_n^*(x_2)}{s_n(x_1,x_2)\sqrt{B_n(x_2)}} < -L) \to 1. \tag{18}$$

The median in the above theorem can be replaced by any γ-quantile for $\beta < \gamma < 1-\beta$. Theorem 4 establishes that the test statistic diverges to minus infinity at rate $n^{-1/2}k_n$ under the alternative. This implies that our test is consistent whenever $k_n = O(n^\eta)$ for $1/2 < \eta < 1$. Therefore, the property of robustness against correlation in the errors comes at the expense of a slower rate of divergence under a stationary alternative. A similar property can be observed in the so-called KPSS statistic of Kwiatkowski et al. (1992).

The condition of Equation (17) is needed to rule out possible unit-root type behavior in y_t. For the standard Dickey-Fuller test, we need $E\Delta y_t y_{t-1} < 0$ in order to derive a law of large numbers under the alternative, and our condition of Equation (17) is the analogue of that condition. In the AR(1) case $y_t = \rho y_{t-1} + u_t$ where $|\rho| < 1$ and u_t is i.i.d., it can be easily verified that the condition of Equation (17) holds. In that case,

$$E\Delta y_t(y_{t-a} - \tilde{\mu})I(y_{t-a} > \tilde{\mu}) = (\rho^a - 1)Ey_{t-a}(y_{t-a} - \tilde{\mu})I(y_{t-a} > \tilde{\mu}). \tag{19}$$

Now $E\Delta y_{t-a}(y_{t-a} - \tilde{\mu})I(y_{t-a} > \tilde{\mu}) > 0$ because if $\tilde{\mu} \geq 0$, $y_{t-a}(y_{t-a} - \tilde{\mu})I(y_{t-a} > \tilde{\mu}) \geq 0$, while if $\tilde{\mu} < 0$ we can write

$$Ey_t(y_t - \tilde{\mu})I(y_t > \tilde{\mu}) = Ey_{t-1}^2 I(y_{t-1} > \tilde{\mu}) - \tilde{\mu}Ey_tI(y_t > 0) - \tilde{\mu}Ey_tI(\tilde{\mu} \leq y_t \leq 0)$$

$$\geq Ey_{t-1}^2 I(y_{t-1} > \tilde{\mu}) - \tilde{\mu}Ey_tI(\tilde{\mu} \leq y_t \leq 0)$$

$$\geq Ey_{t-1}^2 I(y_{t-1} > \tilde{\mu}) - (1/2)\tilde{\mu}^2, \tag{20}$$

and the last expression is strictly positive.

Along the same lines it is also possible to test the null of a unit root against the alternative

$$\Delta y_t = \varphi(y_{t-1} - c_1)I(y_{t-1} < c_1) + \varphi(y_{t-1} - c_2)I(y_{t-1} > c_2) + u_t; \tag{21}$$

i.e. we set $\phi_1 = \phi_2 = \phi$ in the earlier model. We can now define, analogously to the previous statistics,

$$h(x_1, x_2, y) = (y - x_2)I(y > x_2) + (y - x_1)I(y \leq x_1), \tag{22}$$

$$\tilde{H}_n^*(x,x_2) = \sum_{j=1}^{k_n}(U_n(r_j) - U_n(r_{j-1}))h(x_1, x_2, U_n(r_{j-1})), \tag{23}$$

$$\tilde{B}_n(x_1, x_2) = n^{-1}\sum_{t=2}^{n} h(x_1, x_2, U_n(r_{j-1}))^2, \tag{24}$$

and

$$\tilde{s}_n^2(x_1, x_2) = (n-2)^{-1}\sum_{t=2}^{n} \tilde{u}_t(x_1, x_2)^2, \tag{25}$$

where the $\tilde{u}_t(x_1, x_2)$ denote the residuals from the regression of Δy_t on $h(x_1, x_2, n^{-1/2}y_{t-1})$. By copying the earlier reasoning, we can now prove a result analogous to that of Theorem 3 for a t-test that involves a regression using only one regressor. We will provide this result without proof.

Theorem 5 *Let k_n denote an integer-valued positive sequence that is nondecreasing in n and such that $k_n \to \infty$ and $k_n n^{-1} \to 0$. Then under Assumption 1 and H_0,*

$$J_2 = \inf_{(x_1,x_2)\in X}(\hat{\sigma}_n/\hat{\lambda}_n)\frac{\tilde{H}_n^*(x_2)}{\tilde{s}_n(x_1, x_2)\sqrt{\tilde{B}_n(x_2)}}$$

$$\xrightarrow{d} \inf_{x_2\in[\underline{d}_\beta,\overline{d}_\beta]}\frac{\int_0^1 h(x_1, x_2, W(r))dW(r)}{\sqrt{\int_0^1 h(x_1, x_2, W(r))^2 dr}} = D_\beta. \tag{26}$$

A consistency result of the type of Theorem 4 can also be obtained for this case; we will not state that result formally here.

4 Application: Purchasing Power Parity

In this section, we apply our newly developed J-test to the "Purchasing Power Parity" (PPP). PPP can be viewed as the international version of the "Law of One Price"; i.e. if two countries are engaging free trade, arbitrage should make purchasing powers of two countries' currencies to be equivalent. Therefore, it has been generally believed that PPP should hold at least in the long run. For example, Dornbusch and Krugman (1976) and Rogoff (1996) expressed their strong belief in PPP. (See Taylor and Taylor (2004)). Also, PPP has been one of the key building blocks of many international macroeconomic models. For the recent developments and debates in this area, see Taylor and Taylor (2004).

However, empirical evidence so far is at best mixed. Early studies that tested the random work hypothesis for real exchange rate, such as Roll (1979), Adler and Lehmann (1983) and Cumby and Obstfeld (1984), showed little supports for PPP. In the late 1980's and the early 1990's, even with the help of more sophisticated formal unit root tests, empirical studies generally failed to reject the unit root hypothesis. For example, see Taylor (1988) and Mark (1990). As Frankel (1986, 1990) pointed out, however, the power problem due to the small sample size of typically 15 years or so might be a blame for the failure to reject. See Froot and Rogoff (1995) and Lothian and Taylor (1996, 1997). However, even with the benefit of additional 10 or more years data typically available nowadays, the power of the conventional unit root tests, such as ADF and PP, did not improve significantly. See Sarno and Taylor (2002).

One of the many solutions for the power problem is to introduce nonlinear mean-reversion in real exchange rate. Nonlinearity can be well motivated by the transaction cost and the uncertainty that are especially important in the international trade. For some theoretical studies, see Benninga and Protopapadakis (1988), Dumas (1992) and Kilian and Taylor (2003). Also Taylor (2004) argues that nonlinearity may arise from the intervention operation of the central bank. In empirical studies, "Smooth Transition Autoregressive" (STAR) and "Exponential STAR" (ESTAR) have been typically used, and some positive results have been produced. See Michael et al. (1997), Taylor et al. (2001).

Following this line of research, we developed a new test that is based on the simple "Threshold Autoregressive" (TAR) model. We believe that TAR model is more straightforward to be interpreted than STAR and ESTAR,[1] but brings more serious theoretical challenges that is worth pursuing. Noticeable recent work similar to ours is Bec et al. (2004), which developed an unit root test that is also based on TAR model. However, along with the methodological differences, we found more supportive results for PPP.

PPP is defined as a cointegrating relationship among nominal exchange rate, home country's price and foreign country's price with a cointegrating vector of (1, 1, -1).

$$y_t = \ln(E_t) + \ln(P_t) - \ln(P_t^*) \tag{27}$$

where y_t is real exchange rate, E_t is nominal exchange rate, P_t is domestic price, and P_t^* is foreign price. If PPP holds, y_t must be stationary. Therefore, testing PPP is equivalent to the unit root test for y_t. We consider the following threshold autoregressive equation:

$$\Delta y_t = \rho\left[(y_{t-1} - c_1)I(y_{t-1} < c_1) + (y_{t-1} - c_2)I(y_{t-1} > c_2)\right] + u_t \tag{28}$$

Note that we impose a restriction that slope coefficients in both upper and lower regime are equal, however we allow the thresholds to be different between upper and lower regime. We used the J_2-statistic of Equation (26). We applied the J-test to PPP for 6 developed countries, such as Canada, UK, Germany, France, Italy and Spain. For the data set, we use monthly observations on consumer price index (CPI) and nominal exchange rate for those 6 countries against the US dollar. All data are extracted from Datastream. Test results are presented in Table 1 and 2. For all 6 countries, ADF and PP tests cannot reject the null hypothesis of no cointegration at any lag/bandwidth values at conventional significance level, meanwhile our J-test can reject the null hypothesis at many different values of n/k_n. These results support our conviction that the conventional unit root tests might suffer from the lack of power, and there might exist nonlinear mean-reversion in real exchange rate.

References

Adler, M. and B. Lehmann, "Deviations from Purchasing Power Parity in the Long Run," *Journal of Finance* 38 (1983), 1471–1487.

[1] Transaction cost and uncertainty create so called a "band of inaction" within which PPP does not hold. However, STAR and ESTAR cannot capture this phenomenon, meanwhile TAR model can do

Andrews, D. W., "Heteroscedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica* 59 (1991), 817–858.

Bec, F., A. Guay and E. Guerre, "Adaptive consistent unit root tests based on autoregressive threshold model," *Working Paper, CREST-ENSAE, France* (2006).

Bec, F., M. B. Salam and M. Carrasco, "Tests for unit-root versus threshold specification with an application to the purchasing power parity relationship," *Journal of Business and Economic Statistics* 22 (2004), 382–395.

Benninga, S. and A. A. Protopapadakis, "The Equilibrium Pricing of Exchange Rates and Assets When Trade Takes Times," *Journal of International Economics* 7 (1988), 129–149.

Berben, R. and D. van Dijk, "Unit root tests and asymmetric adjustment: a reassessment," *Research report EI-9902/A, Econometric Institute, Erasmus University* (1999).

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

Cumby, R. E. and M. Obstfeld, *International Interest Rate and Price Level Linkages under Flexible Exchange Rates: A Review of Recent Evidence*, Exchange Rate Theory and Practice (Chicago, University of Chicago Press, 1984).

de Jong, R. and J. Davidson, "The functional central limit theorem and weak convergence to stochastic integrals I: weakly dependent processes," *Econometric Theory* 16 (2000), 621–642.

de Jong, R. M., "Nonlinear estimators with integrated regressors but without exogeneity," *Mimeo, Ohio State University* (2002).

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

Dornbusch, R. and P. Krugman, "Flexible Exchange Rates in the Short Run," *Brookings Papers on Economic Activity* 3 (1976), 537–575.

Doukhan, P., *Mixing: properties and examples* (New York: Springer-Verlag, 1994).

Dumas, B., "Dynamic Equilibrium and the Real Exchange Rate in a Spatially Seperated Wrold," *Review of Financial Studies* 5 (1992), 153–180.

Frankel, J. A., *International Capital Mobility and Crowding-out in the U.S. Economy: Imperfect Integration of Financial Markets or of Goods Markets?* (in How Open is the U.S. Economy? Lexington, Mass. Lexington Books, 1986).

Frankel, J. A., *Zen and the Art of Modern Macroeconomics: A Commentary* (in Monetary Policy for a Volatile Global Economy. Washington, D.C., AEI Press, 1990).

Froot, K. A. and K. Rogoff, "Perspectives on PPP and Long-Run Real Exchange Rates," *Handbook of International Economics* (1995).

González, M. and J. Gonzalo, "Threshold unit root models," *Working Paper WP 97-50, Statistics and Econometrics, U. Carlos III de Madrid* (1997).

Hansen, B. E., "Stochastic equicontinuity for unbounded dependent heterogeneous arrays," *Econometric Theory* 12 (1996), 347–359.

Jansson, M., "Consistent covariance matrix estimation for linear processes," *Econometric Theory* 18 (2002), 1449–1459.

Kapetanios, G. and Y. Shin, "Unit root tests in three-regime SETAR models," *Econometrics Journal* 9 (2002), 252–278.

Kilian, L. and M. P. Taylor, "Why Is It So Difficult to Beat the Random Walk Forecast of Exchange Rate," *Journal of International Economics* 60 (2003), 85–107.

Kurtz, T. and P. Protter, "Weak limit theorems for stochastic integrals and stochastic differential equations," *Annals of Probability* 19 (1991), 1035–1070.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root," *Journal of Econometrics* 54 (1992), 159–178.

Lothian, J. R. and M. P. Taylor, "Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries," *The Journal of Political Economy* 104 (1996), 488–509.

Lothian, J. R. and M. P. Taylor, "Real Exchange Rate Behavior: The Problem of Power and Sample Size," *Journal of International Money and Finance* 16 (1997), 945–954.

Mark, N., "Real and Nominal Exchange Rates in the Long Run: An Empirical Investigation," *Journal of International Economics* 28 (1990), 115–136.

Michael, P., R. A. Nobay and D. A. Peel, "Transaction Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation," *Journal of Political Economy* 105 (1997), 862–879.

Phillips, P. and P. Perron, "Testing for a unit root in time series regression," *Biometrika* 75 (1988), 335–346.

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

Rogoff, K., "The Purchasing Power Parity Puzzle," *Journal of Economic Literature* 34 (1996), 647–668.

Roll, R., *Violations of Purchasing Power Parity and their Implications for Efficient International Commodity Markets*, volume 1 of *International Finance and Trade* (Cambridge. Mass., Ballinger, 1979).

Sarno, L. and M. P. Taylor, "Purchasing Power Parity and the Real Exchange Rate," *IMF Staff Papers* 49 (2002), 65–105.

Seo, M., "Unit root test in a threshold autoregression: asymptotic theory and residual-based block bootstrap," *Working paper, London School of Economics.* (2006).

Shintani, M. and Y. Park, "Testing for a unit root against transitional autoregressive models," *mimeo, Vanderbilt University* (2005).

Taylor, A. M. and M. P. Taylor, “The Purchasing Power Parity Debate,” *NBER Working Paper* (2004).

Taylor, M. P., “An Empirical Examination of Long-Run Purchasing Power Parity Using Cointegration Techniques,” *Applied Economics* 20 (1988), 1360–1381.

Taylor, M. P., “Is Official Exchange Rate Intervention Effective?,” *Economica* 71 (2004), 1–12.

Taylor, M. P., D. A. Peel and L. Sarno, “Nonlinear Mean-Reversion in Real Exchange Rates: Toward a Solution to the Purchasing Power Parity Puzzles,” *International Economic Review* 42 (2001), 1015–1042.

Tong, H., *Non-linear time series: a dynamical system approach* (Oxford: Clarendon Press, 1990).

Appendix 1: Critical values of the J_1 and J_2 statistic

Table 1: Critical values of the J_1 statistic

Threshold	1%	2.5%	5%	10%
$\beta = 0.05$	-3.476	-3.207	-2.975	-2.718
$\beta = 0.10$	-3.406	-3.128	-2.894	-2.638

Table 2: Critical values of the J_2 statistic

Threshold	1%	2.5%	5%	10%
$\beta = 0.05$	-4.079	-3.818	-3.580	-3.332
$\beta = 0.10$	-3.960	-3.688	-3.459	-3.202

Critical values were obtained using an Ox program; we used 100,000 replications and $n = 500$. The Ox program code is available upon request from the authors.

Appendix 2: Empirical results

Table 3: **Threshold Unit Root Test[1] : PPP**

Parameter[2]	ADF	PP	J_2 test
Canada (1971:1 ∼ 2004:6)			
1	-1.114	-0.891	-1.356
3	-1.122	-0.998	-1.725
5	-1.104	-1.033	-2.506
7	-1.248	-1.061	-2.160
9	-1.434	-1.098	-2.007
11	-1.827	-1.146	-2.772
13	-1.543	-1.196	-3.046
15	-1.536	-1.218	-4.468*
17	-1.631	-1.235	-6.452*
19	-1.786	-1.255	-5.387*
UK (1971:1 ∼ 2004:7)			
1	-2.713	-1.984	-1.853
3	-2.602	-2.351	-2.916
5	-2.491	-2.439	-4.214*
7	-2.391	-2.462	-5.973*
9	-2.500	-2.462	-2.895
11	-2.617	-2.472	-2.560
13	-2.611	-2.503	-7.522*
15	-2.272	-2.527	-4.330*
17	-2.307	-2.517	-7.812*
19	-2.724	-2.509	-4.498*
Germany (1971:1 ∼ 2001:12)			
1	-2.131	-1.576	-1.688
3	-2.145	-1.874	-2.474
5	-2.021	-1.956	-2.993
7	-2.016	-1.978	-2.712
9	-2.180	-2.005	-2.623
11	-2.270	-2.056	-2.624
13	-2.485	-2.108	-4.690*
15	-2.537	-2.159	-3.181
17	-2.293	-2.199	-6.723*
19	-2.256	-2.218	-5.360*

[1] * indicates rejection of the null hypothesis of no cointegration at 5% level of significance. (critical value = -3.629).

[2] lag length for ADF, HAC bandwidth parameter for PP, and n/k_n for J-test

Table 4: **Threshold Unit Root Test : PPP**

	ADF	PP	J_2 test
France (1971:1 ∼ 2001:12)			
1	-1.893	-1.321	-1.732
3	-2.052	-1.635	-2.369
5	-2.134	-1.760	-3.058
7	-2.007	-1.825	-3.060
9	-2.198	-1.862	-3.580
11	-2.120	-1.906	-2.014
13	-2.308	-1.939	-5.836*
15	-2.356	-1.981	-2.820
17	-2.148	-2.018	-7.800*
19	-2.057	-2.037	-9.015*
Italy (1971:1 ∼ 2001:12)			
1	-2.092	-1.315	-2.206
3	-1.961	-1.716	-2.991
5	-2.105	-1.803	-3.966*
7	-2.044	-1.853	-6.327*
9	-2.200	-1.883	-2.812
11	-2.230	-1.927	-2.659
13	-2.432	-1.969	-7.456*
15	-2.287	-2.008	-2.531
17	-2.044	-2.033	-6.785*
19	-1.878	-2.035	-6.969*
Spain (1973:1 ∼ 2001:12)			
1	-1.788	-1.412	-1.803
3	-1.737	-1.653	-2.262
5	-1.791	-1.735	-3.936*
7	-1.963	-1.799	-2.278
9	-2.226	-1.868	-3.762*
11	-2.102	-1.937	-2.306
13	-2.466	-1.983	-6.109*
15	-2.248	-2.024	-9.618*
17	-2.296	-2.057	-10.531*
19	-2.286	-2.084	-4.103*

[1] * indicates rejection of the null hypothesis of no cointegration at 5% level of significance. (critical value = -3.629).

[2] lag length for ADF, HAC bandwidth parameter for PP, and n/k_n for J-test

Appendix 2: Mathematical proofs

Lemma 1 *Let* (Y_{nj}, V_{nj}) *be an array of random variables such that* V_{nj} *is is i.i.d. for any* n, V_{nj} *is independent of* Y_{ni} *for all* $i < j$, *and* $EV_{nj} = EY_{nj} = 0$. *Also, assume that for some* $q > 1$,

$$\sup_{n\geq 1} \parallel V_{nj} \parallel_{2q} < \infty,$$

$$\sup_{n\geq 1} \parallel Y_{nj} \parallel_{2q} < \infty$$

and

$$\sup_{n\geq 1} \parallel \max_{1\leq j\leq k_n} |k_n^{-1/2} \sum_{i=1}^{j-1} Y_{ni}| \parallel_{2q} < \infty.$$

Then for any $K > 0$,

$$k_n^{-1/2} \sum_{j=1}^{k_n} V_{nj} g(k_n^{-1/2} \sum_{i=1}^{j-1} Y_{ni}, x) \Rightarrow \int_0^1 g(Y(r), x) dV(r) \tag{29}$$

on $[-K, K]$.

Proof of Lemma 1:

It follows from the FCLT (see e.g. de Jong and Davidson (2000)) that under the conditions of Lemma 1,

$$k_n^{-1/2} \sum_{i=1}^{[rk_n]} (V_{ni}, Y_{ni}) \Rightarrow (V(r), Y(r)).$$

Therefore, by Kurtz and Protter (1991), it follows that pointwise in x,

$$k_n^{-1/2} \sum_{j=1}^{k_n} V_{nj} g(k_n^{-1/2} \sum_{i=1}^{j-1} Y_{ni}, x) \xrightarrow{d} \int_0^1 g(Y(r), x) dV(r).$$

It remains to show stochastic equicontinuity, and in order to establish this, we verify the conditions of Theorem 4 of Hansen (1996). Obviously, because $|g(u,x)-g(u,x')| \leq |x-x'|$,

$$|V_{nj}g(k_n^{-1/2}\sum_{i=1}^{j-1} Y_{ni}, x) - V_{nj}g(k_n^{-1/2}\sum_{i=1}^{j-1} Y_{ni}, x')| \leq |V_{nj}||x-x'|,$$

implying that we can set $\lambda = 1$ for Hansen's theorem, that the condition of Hansen's Theorem 4 becomes that $q > 1$, and that Assumption 2 in Hansen (1996) becomes the conditions of

$$\limsup_{n\to\infty} n^{-1}\sum_{j=1}^{n} \| V_{nj}g(n^{-1/2}\sum_{i=1}^{j-1} Y_{ni}, x) \|_q^2 < \infty \tag{30}$$

and

$$\sup_{n\geq 1} \| V_{nj} \|_q < \infty.$$

The second condition is assumed, while the first condition holds because by Doob's inequality and the Burkholder inequality,

$$\| V_{nj}g(n^{-1/2}\sum_{i=1}^{j-1} Y_{ni}, x) \|_q$$

$$\leq \| V_{nj} \|_{2q} (K + \| \max_{1\leq j\leq n} |n^{-1/2}\sum_{i=1}^{j-1} Y_{ni}| \|_{2q})$$

implying that by assumption, the condition of Equation (30) holds. $\square$

Proof of Theorem 1:

Write

$$H_n^*(x) = \sum_{j=1}^{k_n}(U_n(r_j) - U_n(r_{j-1}))g(U_n(r_{j-1}), x)$$

where $g(u, x)$ is as before. Now by Phillips and Solo (1992), p. 976,

$$n^{1/2}(U_n(r_j) - U_n(r_{j-1})) = \sum_{t=n_{j-1}+1}^{n_j} u_t = (\sum_{k=0}^{\infty} \phi_k) \sum_{t=n_{j-1}+1}^{n_j} \varepsilon_t + \tilde{\varepsilon}_{n_{j-1}} - \tilde{\varepsilon}_{n_j},$$

implying that

$$H_n^*(x) = \sum_{j=1}^{k_n} (U_n(r_j) - U_n(r_{j-1}))g(U_n(r_{j-1}), x)$$

$$= k_n^{-1/2} \sum_{j=1}^{k_n} ((\sum_{k=0}^{\infty} \phi_k)(n/k_n)^{-1/2} \sum_{t=n_{j-1}+1}^{n_j} \varepsilon_t) g(U_n(r_{j-1}), x)$$

$$+ n^{-1/2} \sum_{j=1}^{k_n} (\tilde{\varepsilon}_{n_{j-1}} - \tilde{\varepsilon}_{n_j}) g(U_n(r_{j-1}), x). \tag{31}$$

We will show that the last term converges to 0 uniformly in $x \in [-K, K]$, and for the first term, we will prove weak convergence. To show that the last term in Equation (31) converges to 0, note that

$$n^{-1/2} \sum_{j=1}^{k_n} (\tilde{\varepsilon}_{n_{j-1}} - \tilde{\varepsilon}_{n_j}) g(U_n(r_{j-1}), x)$$

$$= n^{-1/2} \tilde{\varepsilon}_{n_0} g(U_n(r_0), x) - n^{-1/2} \tilde{\varepsilon}_{n_{k_n}} g(U_n(r_{k_n - 1}), x)$$

$$+ n^{-1/2} \sum_{j=1}^{k_n - 1} \tilde{\varepsilon}_{n_j} (g(U_n(r_j), x) - g(U_n(r_{j-1}), x)),$$

and note that, because $|g(u, x)| \leq |u| + |x|$,

$$\sup_{x \in [-K,K]} |n^{-1/2} \tilde{\varepsilon}_{n_0} g(U_n(r_0), x)| \leq (K + \sup_{r \in [0,1]} |U_n(r)|) n^{-1/2} \max_{t=1,\ldots,n} |\tilde{\varepsilon}_t| \xrightarrow{p} 0$$

because $U_n(\cdot) \Rightarrow U(\cdot)$ and $E|\tilde{\varepsilon}_t|^p < \infty$ for some $p > 2$ by Assumption 1. A similar argument holds for

$$n^{-1/2} \tilde{\varepsilon}_{n_{k_n}} g(U_n(r_{k_n - 1}), x).$$

Also, by the Cauchy-Schwartz inequality and because $|g(u,x) - g(u',x)| \leq |u - u'|$,

$$\sup_{x\in\mathbb{R}} |n^{-1/2} \sum_{j=1}^{k_n-1} \tilde{\varepsilon}_{n_j}(g(U_n(r_j),x) - g(U_n(r_{j-1}),x))|$$

$$\leq n^{-1/2} \sum_{j=1}^{k_n-1} |\tilde{\varepsilon}_{n_j}||g(U_n(r_j),x) - g(U_n(r_{j-1}),x)|$$

$$\leq n^{-1/2}(\sum_{j=1}^{k_n-1} \tilde{\varepsilon}_{n_j}^2)^{1/2}(\sum_{j=1}^{k_n-1} |g(U_n(r_j),x) - g(U_n(r_{j-1}),x)|^2)^{1/2}$$

$$\leq k_n^{1/2} n^{-1/2}(k_n^{-1} \sum_{j=1}^{k_n-1} \tilde{\varepsilon}_{n_j}^2)^{1/2}(\sum_{j=1}^{k_n-1} |U_n(r_j) - U_n(r_{j-1})|^2)^{1/2},$$

and because $E\tilde{\varepsilon}_{n_j}^2 < \infty$ and $\sup_{n\geq 1} \sum_{j=1}^{k_n-1} E|U_n(r_j) - U_n(r_{j-1})|^2 < \infty$, it follows that the last term is $O(k_n^{1/2} n^{-1/2}) = o(1)$.
To show that the first term of Equation (31) converges weakly on $[-K, K]$, we will apply Lemma 1. Define $Y_{nj} = k_n^{1/2}(U_n(r_j) - U_n(r_{j-1}))$ and $V_{nj} = (\sum_{k=0}^{\infty} \phi_k)(n/k_n)^{-1/2} \sum_{t=n_{j-1}+1}^{n_j} \varepsilon_t$. Then the first term of Equation (31) equals

$$k_n^{-1/2} \sum_{j=1}^{k_n} V_{nj} g(k_n^{-1/2} \sum_{i=1}^{j-1} Y_{ni}, x),$$

and therefore it suffices to verify the conditions of Lemma 1. Obviously Y_{nj} is i.i.d. for any n, independent of V_{ni} for $i < j$, and $EV_{nj} = 0$. Also, by the Burkholder and norm inequalities,

$$\sup_{n\geq 1} \| V_{nj} \|_{2q}$$

$$= (\sum_{k=0}^{\infty} \phi_k) \sup_{n\geq 1} \| (n/k_n)^{-1/2} \sum_{t=n_{j-1}+1}^{n_j} \varepsilon_t \|_{2q}$$

$$\leq (\sum_{k=0}^{\infty} \phi_k) \sup_{n\geq 1} \| (n/k_n)^{-1} \sum_{t=n_{j-1}+1}^{n_j} \varepsilon_t^2 \|_q^{1/2}$$

$$\leq (\sum_{k=0}^{\infty} \phi_k) \sup_{n \geq 1} ((n/k_n)^{-1} \sum_{t=n_{j-1}+1}^{n_j} \| \varepsilon_t^2 \|_q)^{1/2}$$

$$\leq (\sum_{k=0}^{\infty} \phi_k) \| \varepsilon_t \|_{2q} < \infty.$$

Furthermore,

$$\sup_{n \geq 1} \| \max_{1 \leq j \leq k_n} |k_n^{-1/2} \sum_{i=1}^{j-1} Y_{ni}| \|_{2q}$$

$$= \sup_{n \geq 1} \| \max_{1 \leq j \leq k_n} |U_n(r_{j-1})| \|_{2q}$$

$$\leq \sup_{n \geq 1} \| \max_{1 \leq t \leq n} |n^{-1/2} \sum_{j=1}^{t} u_j| \|_{2q}$$

$$\leq C \sup_{n \geq 1} \| n^{-1/2} \sum_{j=1}^{n} u_j \|_{2q}$$

$$\leq C \sup_{n \geq 1} \| (\sum_{k=0}^{\infty} \phi_k) n^{-1/2} \sum_{j=1}^{n} \varepsilon_j + n^{-1/2} \tilde{\varepsilon}_0 - n^{-1/2} \tilde{\varepsilon}_n \|_{2q}$$

$$\leq C' \| \varepsilon_j \|_{2q},$$

where the last inequality follows by the Burkholder inequality. The result now follows by setting $q = p/2$. $\square$

Proof of Theorem 2:

From Lemma 1, it follows that because u_t is i.i.d., $G_n(x)$ is stochastically equicontinuous under the conditions of the theorem. Also, from Theorem 1, it follows that $H_n^*(x)$ is stochastically equicontinuous, which implies that $H_n^*(x_2) - G_n(x_2)$ is stochastically equicontinuous. Therefore, noting that $\sup_{t=1,\ldots,n} |n^{-1/2} y_t| = O_p(1)$, it suffice to show that pointwise for each $x \in \mathbb{R}$, $H_n^*(x_2) - G_n(x_2) \xrightarrow{p} 0$. To show this, define

$$X_{ntj} = (n^{-1/2}\sum_{i=1}^{t-1}\varepsilon_i - x_2)I(n^{-1/2}\sum_{i=1}^{t-1}\varepsilon_i > x_2) - (n^{-1/2}\sum_{i=1}^{n_{j-1}}\varepsilon_i - x_2)I(n^{-1/2}\sum_{i=1}^{n_{j-1}}\varepsilon_i > x_2)$$

$$= g(n^{-1/2}\sum_{i=1}^{t-1}\varepsilon_i, x_2) - g(n^{-1/2}\sum_{i=1}^{n_{j-1}}\varepsilon_i, x_2),$$

and observe that

$$|g(n^{-1/2}\sum_{i=1}^{t-1}\varepsilon_i, x_2) - g(n^{-1/2}\sum_{i=1}^{n_{j-1}}\varepsilon_i, x_2)|$$

$$\leq |n^{-1/2}\sum_{i=n_{j-1}+1}^{t-1}\varepsilon_i|.$$

Next, note that for al $x_2 \in \mathbb{R}$,

$$E|H_n^*(x_2) - G_n(x_2)|^2$$

$$= E|n^{-1/2}\sum_{j=1}^{k_n}\sum_{t=n_{j-1}}^{n_j}\varepsilon_t X_{njt}|^2$$

$$= \sigma^2 n^{-1}\sum_{j=1}^{k_n}\sum_{t=n_{j-1}}^{n_j} EX_{njt}^2$$

$$\leq \sigma^2 k_n^{-1}\sum_{j=1}^{k_n} E\max_{n_{j-1}+1\leq t\leq n_j} X_{njt}^2$$

$$\leq C\sigma^2 k_n^{-1} \sum_{j=1}^{k_n} EX_{nj,n_j}^2$$

$$\leq C'\sigma^2 k_n^{-1} \sum_{j=1}^{k_n} n^{-1}(n/k_n) = O(k_n^{-1}) = o(1)$$

by assumption. □

For the proof of Theorem 3, we need several lemmas:

Lemma 2 *For all $K > 0$, under Assumption 1,*

$$\sup_{x_2 \in [-K,K]} |n^{-1} \sum_{t=2}^{n} u_t (n^{-1/2} y_{t-1} - x_2) I(n^{-1/2} y_{t-1} > x_2)| = o_p(1)$$

and

$$\sup_{x_1 \in [-K,K]} |n^{-1} \sum_{t=2}^{n} u_t (n^{-1/2} y_{t-1} - x_1) I(n^{-1/2} y_{t-1} < x_1)| = o_p(1).$$

Proof:

We will only show the first asserted result; the second result can be proven analogously. Uniform convergence follows from stochastic equicontinuity and pointwise convergence; stochastic equicontinuity follows from observing that

$$|n^{-1} \sum_{t=2}^{n} u_t g(n^{-1/2} y_{t-1}, x_2) - n^{-1} \sum_{t=2}^{n} u_t g(n^{-1/2} y_{t-1}, x_2')| \leq n^{-1} \sum_{t=1}^{n} |u_t||x_2 - x_2'|$$

and that $n^{-1} \sum_{t=1}^{n} E|u_t| = E|u_t| < \infty$ under Assumption 1. To show pointwise convergence, note that, for any sequence r_n such that $r_n \to \infty$ and $r_n n^{-1} \to 0$, defining $b_n = [n/r_n]$, for n large enough,

$$n^{-1} \sum_{t=2}^{n} u_t g(n^{-1/2} y_{t-1}, x_2)$$

$$= n^{-1} \sum_{i=1}^{r_n} \sum_{t=(i-1)b_n+1}^{ib_n} u_t g(n^{-1/2} y_{t-1}, x_2) + \sum_{i=r_n b_n+1}^{n} u_t g(n^{-1/2} y_{t-1}, x_2)$$

$$= n^{-1} \sum_{i=1}^{r_n} \sum_{t=(i-1)b_n+1}^{ib_n} u_t (g(n^{-1/2} y_{t-1}, x_2) - g(n^{-1/2} y_{ib_n}, x_2))$$

$$+ n^{-1} \sum_{i=1}^{r_n} \sum_{t=(i-1)b_n+1}^{ib_n} u_t g(n^{-1/2} y_{ib_n}, x_2) + n^{-1} \sum_{i=r_n b_n+1}^{n} u_t g(n^{-1/2} y_{t-1}, x_2).$$

Noting that the last term sums over less than b_n elements, it is easily seen to be $o_p(1)$. Also, because $|g(u,x)| \leq |u| + |x|$,

$$\| \sup_{x_2 \in \mathbb{R}} |n^{-1} \sum_{i=1}^{r_n} \sum_{t=(i-1)b_n+1}^{ib_n} u_t g(n^{-1/2} y_{ib_n}, x_2) \|_1$$

$$\leq n^{-1} \sum_{i=1}^{r_n} \| \sum_{t=(i-1)b_n+1}^{ib_n} u_t \|_2 (\| n^{-1/2} y_{ib_n} \|_2 + K),$$

and because

$$\| \sum_{t=(i-1)b_n+1}^{ib_n} u_t \|_2 \leq C b_n^{1/2},$$

this term is $O(n^{-1} r_n b_n^{1/2}) = O(n^{-1/2} r_n^{1/2}) = o(1)$. Also, because $|g(u,x) - g(u',x)| \leq |u - u'|$,

$$\sup_{x_2 \in \mathbb{R}} |n^{-1} \sum_{i=1}^{r_n} \sum_{t=(i-1)b_n+1}^{ib_n} u_t (g(n^{-1/2} y_{t-1}, x_2) - g(n^{-1/2} y_{ib_n}, x_2))|$$

$$\leq n^{-3/2} \sum_{i=1}^{r_n} \sum_{t=(i-1)b_n+1}^{ib_n} |u_t||y_{t-1} - y_{ib_n}|,$$

and by the Cauchy-Schwartz inequality,

$$\| n^{-3/2} \sum_{i=1}^{r_n} \sum_{t=(i-1)b_n+1}^{ib_n} |u_t||y_{t-1} - y_{ib_n}| \|_1$$

$$\leq \sigma n^{-3/2} \sum_{i=1}^{r_n} \sum_{t=(i-1)b_n+1}^{ib_n} \| y_{t-1} - y_{ib_n} \|_2$$

$$= O(n^{-3/2}(r_n b_n) b_n^{1/2}) = O(b_n^{1/2} n^{-1/2}) = o(1),$$

which completes the proof. $\square$

Lemma 3 *Under Assumption 1,*

$$\sup_{(x_1,x_2)\in X} |s_n^2(x_1,x_2) - \sigma^2| \xrightarrow{p} 0.$$

Proof of Lemma 3:

Note that by orthogonality of the regressors, for the OLS estimators $\hat{\phi}_1(x_1)$ and $\hat{\phi}_2(x_2)$ that originate from a regression of Δy_t on

$$(n^{-1/2}y_{t-1} - x_1)I(n^{-1/2}y_{t-1} < x_1) \quad \text{and} \quad (n^{-1/2}y_{t-1} - x_2)I(n^{-1/2}y_{t-1} > x_2),$$

we have

$$\hat{\phi}_1(x_1) = \frac{n^{-1}\sum_{t=2}^n u_t(n^{-1/2}y_{t-1} - x_1)I(n^{-1/2}y_{t-1} < x_1)}{n^{-1}\sum_{t=2}^n (n^{-1/2}y_{t-1} - x_1)^2 I(n^{-1/2}y_{t-1} < x_1)}$$

and

$$\hat{\phi}_2(x_2) = \frac{n^{-1}\sum_{t=2}^n u_t(n^{-1/2}y_{t-1} - x_2)I(n^{-1/2}y_{t-1} > x_2)}{n^{-1}\sum_{t=2}^n (n^{-1/2}y_{t-1} - x_2)^2 I(n^{-1/2}y_{t-1} > x_2)}.$$

We will concentrate on $\hat{\phi}_2(x_2)$; the argument for $\hat{\phi}_1(x_1)$ is identical. Note that because $\underline{c}_{1n}$ and $\overline{c}_{2n}$ are $O_p(1)$ and by Lemma 2, it follows that

$$\sup_{x_2\in[\underline{c}_{1n},\overline{c}_{2n}]} |n^{-1}\sum_{t=2}^n u_t(n^{-1/2}y_{t-1} - x_2)I(n^{-1/2}y_{t-1} > x_2)| = o_p(1)$$

and because the denominator is decreasing in x_2 and because $x_2 \in [\underline{c}_{1n}, \overline{c}_{2n}]$,

$$n^{-1} \sum_{t=2}^{n} (n^{-1/2} y_{t-1} - x_2)^2 I(n^{-1/2} y_{t-1} > x_2)$$

$$= n^{-1} \sum_{t=2}^{n} g(n^{-1/2} y_{t-1}, x_2)^2 \geq n^{-1} \sum_{t=2}^{n} g(n^{-1/2} y_{t-1}, n^{-1/2} \overline{c}_{2n})^2.$$

Noting that the last statistic converges in distribution to an a.s. positive random variable, it follows that $\sup_{x_2 \in [\underline{c}_{1n}, \overline{c}_{2n}]} |\hat{\phi}_2(x_2)| = o_p(1)$. Define $h(u, x) = (u - x)I(u < x)$. Then

$$|n^{-1} \sum_{t=2}^{n} (u_t - \hat{\phi}_2(x_2) g(n^{-1/2} y_{t-1}, x_2) - \hat{\phi}_1(x_1) h(n^{-1/2} y_{t-1}, x_1))^2 - n^{-1} \sum_{t=2}^{n} u_t^2|$$

$$\leq 2 \sup_{x_2 \in [\underline{c}_{1n}, \overline{c}_{2n}]} |n^{1/2} \hat{\phi}_2(x_2)| \sup_{x_2 \in [\underline{c}_{1n}, \overline{c}_{2n}]} |n^{-3/2} \sum_{t=2}^{n} u_t g(n^{-1/2} y_{t-1}, x_2)|$$

$$+ \sup_{x_2 \in [\underline{c}_{1n}, \overline{c}_{2n}]} |n^{1/2} \hat{\phi}_2(x_2)|^2 \sup_{x_2 \in [\underline{c}_{1n}, \overline{c}_{2n}]} |n^{-2} \sum_{t=2}^{n} g(n^{-1/2} y_{t-1}, x_2)^2|$$

$$+2 \sup_{x_1 \in [\underline{c}_{1n}, \overline{c}_{2n}]} |n^{1/2} \hat{\phi}_1(x_1)| \sup_{x_1 \in [\underline{c}_{1n}, \overline{c}_{2n}]} |n^{-3/2} \sum_{t=2}^{n} u_t g(n^{-1/2} y_{t-1}, x_1)|$$

$$+ \sup_{x_1 \in [\underline{c}_{1n}, \overline{c}_{2n}]} |n^{1/2} \hat{\phi}_1(x_1)|^2 \sup_{x_1 \in [\underline{c}_{1n}, \overline{c}_{2n}]} |n^{-2} \sum_{t=2}^{n} h(n^{-1/2} y_{t-1}, x_1)^2|$$

$$= o_p(1),$$

and noting that $|n^{-1} \sum_{t=1}^{n} (u_t^2 - \sigma^2)| = o_p(1)$, the result now follows from Lemma 2. $\square$

Proof of Theorem 3:

By the Skorokhod representation, for determining the limit distribution of our statistic, we can assume without loss of generality but with some abuse of notation that

$$\hat{\sigma}_n/\hat{\lambda}_n \xrightarrow{as} \sigma/\lambda,$$

$$\sup_{(x_1,x_2)\in X} |s_n^2(x_1,x_2) - \sigma^2| \xrightarrow{as} 0,$$

$$(\underline{c}_{n1}, \overline{c}_{n2}) \xrightarrow{as} (\lambda \underline{d}_\beta, \lambda \overline{d}_\beta),$$

$$\sup_{x_2\in[\underline{c}_{1n},\overline{c}_{2n}]} |B_n(x_2) - \int_0^1 U(r)^2 I(U(r) > x_2)dr| \xrightarrow{as} 0,$$

and

$$\sup_{x_2\in[\underline{c}_{1n},\overline{c}_{2n}]} |H_n^*(x_2) - \int_0^1 I(U(r) > x_2)U(r)dU(r)| \xrightarrow{as} 0.$$

Also, for $x_2 \in [\underline{c}_{n1}, \overline{c}_{n2}]$,

$$B_n(x_2) \geq \overline{c}_{n2}^2 n^{-1} \sum_{t=2}^{n} I(n^{-1/2} y_{t-1} > x_2) \geq \beta \overline{c}_{n2}^2.$$

Therefore, noting that

$$(\hat{\sigma}_n/\hat{\lambda}_n) \frac{H_n^*(x_2)}{s_n(x_1,x_2)\sqrt{B_n(x_2)}} = (\hat{\sigma}_n/\hat{\lambda}_n) \frac{H_n^*(x_2)}{s_n(x_1,x_2)\sqrt{B_n(x_2) - \beta\overline{c}_n^2 + \beta\overline{c}_n^2}},$$

it follows from the earlier uniform convergence observations that

$$\sup_{(x_1,x_2)\in X} |(\hat{\sigma}_n/\hat{\lambda}_n) \frac{H_n^*(x_2)}{s_n(x_1,x_2)\sqrt{B_n(x_2)}} - (\sigma/\lambda) \frac{\int_0^1 I(U(r) > x)U(r)dU(r)}{\sigma\sqrt{B(x)}}| \xrightarrow{as} 0.$$

Since the above limit statistic is continuous in $(\underline{c}_{1n}, \overline{c}_{2n})$, it now follows since $U(r) = \lambda W(r)$ that the limit distribution of the test statistic is

$$\inf_{x_2\in[\lambda\underline{d}_\beta,\lambda\overline{d}_\beta]} \frac{\int_0^1 I(\lambda W(r) > x)\lambda^2 W(r)dW(r)}{\sigma\sqrt{\int_0^1 I(\lambda W(r) > x)\lambda^2 W(r)^2 dr}} (\sigma/\lambda)$$

$$= \inf_{x_2 \in [\underline{d}_\beta, \overline{d}_\beta]} \frac{\int_0^1 I(W(r) > x) W(r) dW(r)}{\sqrt{\int_0^1 I(W(r) > x) W(r)^2 dr}}.$$

This completes the proof. □

Proof of Theorem 4:

First note that

$$\frac{H_n^*(x_2)}{s_n(x_1, x_2)\sqrt{B_n(x_2)}}$$

$$= \frac{\sum_{j=1}^{k_n} (U_n(r_j) - U_n(r_{j-1})(U_n(r_{j-1}) - x_2) I(U_n(r_{j-1}) > x_2)}{s(x_1, x_2)\sqrt{n^{-1}\sum_{t=2}^{n} (n^{-1/2} y_{t-1} - x_2)^2 I(n^{-1/2} y_{t-1} > x_2)}}$$

$$= (n^{-1/2} k_n) \frac{k_n^{-1} \sum_{j=1}^{k_n} (y_{n_j} - y_{n_{j-1}})(y_{n_{j-1}} - c_2) I(y_{n_{j-1}} > c_2)}{s(c_1, c_2)\sqrt{n^{-1}\sum_{t=2}^{n} (y_{t-1} - c_2)^2 I(y_{t-1} > c_2)}}.$$

Next, we will show that the three terms appearing in the above fraction converge in appropriate ways. Also, since we are optimizing over the random set X, the above statistic will always be less than or equal to the statistic evaluated at $(x_1, x_2) = (n^{-1/2}\tilde{m}_n, n^{-1/2}\tilde{m}_n)$, where $\tilde{m}_n$ denotes the median of y_t. Therefore, it suffices to show that for some constant $C > 0$,

$$\limsup_{n\to\infty} P(n^{1/2} k_n^{-1} (\hat{\sigma}_n / \hat{\lambda}_n) \frac{H_n^*(n^{-1/2}\tilde{m}_n)}{s_n(n^{-1/2}\tilde{m}_n, n^{-1/2}\tilde{m}_n)\sqrt{B_n(n^{-1/2}\tilde{m}_n)}} < -C) = 1.$$

Since $\tilde{m}_n \xrightarrow{p} \tilde{\mu}$ by assumption, we need to consider the uniform behavior of $s_n(\cdot,\cdot)$, $B_n(\cdot)$ and $H_n(\cdot)$ in a compact neighborhood $[\tilde{\mu} - \delta, \tilde{\mu} + \delta]$ of $\tilde{\mu}$. To show uniform convergence of $B_n(\cdot)$ in a compact neighborhood of $\tilde{\mu}$, note that pointwise in c_2, by the law of large numbers,

$$n^{-1}\sum_{t=2}^{n} (y_{t-1} - c_2)^2 I(y_{t-1} > c_2) \xrightarrow{p} E(y_{t-1} - c_2)^2 I(y_{t-1} > c_2)$$

under Assumption 2. Also,

$$|n^{-1}\sum_{t=2}^{n}(y_{t-1}-c_2)^2 I(y_{t-1}>c_2)-n^{-1}\sum_{t=2}^{n}(y_{t-1}-c_2')^2 I(y_{t-1}>c_2')|$$

$$\leq |c_2-c_2'|n^{-1}\sum_{t=2}^{n}(2|y_{t-1}|+|c_2|+|c_2'|),$$

implying that $B_n(\cdot)$ is stochastically equicontinuous, which together with pointwise convergence in probability implies uniform convergence in probability on $[\tilde{\mu}-\delta,\tilde{\mu}+\delta]$. Therefore,

$$B_n(n^{-1/2}\tilde{m}_n)\xrightarrow{p} E(y_{t-1}-\tilde{\mu})^2 I(y_{t-1}>\tilde{\mu}).$$

Similarly,

$$k_n^{-1}\sum_{j=1}^{k_n}(y_{n_j}-y_{n_{j-1}})(y_{n_{j-1}}-c_2)I(y_{n_{j-1}}>c_2)$$

also converges pointwise in probability and is stochastically equicontinuous on $[\tilde{\mu}-\delta,\tilde{\mu}+\delta]$, implying that

$$k_n^{-1}\sum_{j=1}^{k_n}(y_{n_j}-y_{n_{j-1}})(y_{n_{j-1}}-\tilde{\mu})I(y_{n_{j-1}}>\tilde{\mu})\xrightarrow{p} E(y_{n_j}-y_{n_{j-1}})(y_{n_{j-1}}-\tilde{\mu})I(y_{n_{j-1}}>\tilde{\mu}).$$

Under the assumptions of the theorem,

$$E(y_{n_j}-y_{n_{j-1}})(y_{n_{j-1}}-\tilde{\mu})I(y_{n_{j-1}}>\tilde{\mu})<0.$$

A similar argument can be used to show that $|s_n(n^{-1/2}\tilde{m}_n, n^{-1/2}\tilde{m}_n)^2-\sigma^2| = o_p(1)$. Also under Assumption 2, $\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2$ and $\hat{\lambda}_n^2 \xrightarrow{p} \lambda^2$ by the results of Andrews (1991). $\square$