Unit Root Test in a Threshold Autoregression: Asymptotic Theory and Residual-based Block Bootstrap

by

Myunghwan Seo*
London School of Economics and Political Science

Contents:

Discussion Paper
No.EM/05/484
January 2005

The Suntory Centre
Suntory and Toyota International Centres for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020 7955 6679

* I am indebted to Bruce Hansen for his invaluable guidance and time. I also thank Peter Robinson, Oliver Linton, and David Hendry for helpful comments. Previous versions were presented at MEG 2003 and NASM 2004.

1 INTRODUCTION

Threshold autoregressive (TAR) models have been studied intensively last decade. The asymptotic properties of the least squares estimation are established by Chan (1993) and Chan and Tsay (1998). See also Hansen (1997) for a different asymptotic approach. Various testing methods have been developed for the presence of a threshold effect, see Chan and Tong (1990) and Hansen (1996) among others. On the other hand, testing the unit root null hypothesis in a TAR model is a rather recent area of investigation. For this purpose, the models are distinguished depending on whether the threshold variable is a level or a difference of the time series (we call them a level-based TAR model and a difference-based TAR model respectively). Unlike the differences, the level is nonstationary under the unit root hypothesis and the relevant distribution theory is completely different. Caner and Hansen (2001) provide a rigorous treatment of statistical tests for the difference-based model. In the absence of an appropriate distribution theory, however, the standard unit root test such as the ADF test has been commonly used for the level-based TAR models. The purpose of this paper is to develop such a distribution theory and a bootstrap for finite sample inferences.

The level-based TAR models have exhibited a great deal of empirical relevance. A leading example is the threshold cointegration model introduced by Balke and Fomby (1997), in which deviation from a long-run equilibrium follows a threshold autoregression, and the lagged level of the deviation is the threshold variable. See Lo and Zivot (2001) for a review of this model's wide application in the analysis of the purchasing power parity hypothesis or the law of one price hypothesis. Many economic and financial time series are also known to exhibit nonlinearities in mean, which may be well approximated by the level-based TAR model. For example, the estimated drift function of short-run interest rate by Ait-Sahalia (1996), appears to exhibit threshold effects based on the level of the rate.

This paper provides a formal distribution theory for unit root testing in a level-based TAR model, which allows for dependent heterogeneous innovations. The null model is a linear autoregression with a unit root. Since the threshold parameters are not identified under the null hypothesis, considered is the Wald test, which is the sup-Wald type test motivated by Davies (1987). We find that the Wald test has a non-standard asymptotic distribution; it is biased and depends on nuisance parameters in a complicated manner, and thus cannot be tabulated.

The weak convergence of stochastic integrals for dependent heterogeneous arrays over the parameter space is developed to derive the asymptotic distribution of the supremum type statistic. Since Park and Phillips (2001) have developed asymptotic theories for stochastic integrals for martingale arrays involving a broad range of nonlinear transformations, an active research, including this paper, has been undertaken to generalize the result to various situations. For a continuously differentiable function class, De Jong (2002) relaxes the martingale assumption in Park and Phillips (2001). On the other hands, Bec, Guay, and Guerre (2002) study the weak convergence for the case in which a nonstationary variable is appropriately normalized before the transformations.

We show that a residual-based block bootstrap (RBB) yields an asymptotically valid approximation to the sampling distribution of the Wald statistic. The nonparametric nature of a block bootstrap is appropriate to the weak assumption on the dependent structure of our model.[1] The RBB is close to that of Paparoditis and Politis (2003) in that the residuals are resampled by the block resampling of Künsch (1989) and then integrated to produce a bootstrap integrated process. Despite the similarity, the asymptotic developments are different in two aspects: First, the residuals are constructed from a TAR model. Second, and more importantly, the weak convergence of stochastic integrals over the parameter space should be established in the RBB context, unlike in the linear model in Paparoditis and Politis (2003).

The finite sample performances of the proposed bootstrap are examined and compared to the conventional augmented Dickey Fuller (ADF) test through Monte Carlo simulations. It is well documented in the literature that the ADF test loses power dramatically for some classes of threshold alternatives (see, for example, Pippenger and Goering (1993)). We find that the power gain from explicitly considering the threshold alternative is substantial for various data generating processes. The finding does not depend on different block length selections.

The newly developed testing strategy is illustrated through an investigation of the law of one price (LOP) hypothesis amongst used car markets in the US. An important feature of used car markets is the presence of uncertainty of quality, which has generated a large debate on the efficiency of these markets. An implication for the study of the LOP is that it (in addition to the transaction costs) will obstruct the arbitrage for at least a short period of time, and that the assumption of the linear adjustment in conventional cointegration models is not likely. In fact, we find that conventional cointegration tests,

[1] In fact, the first differences of the time series instead of the residuals can also be employed for the resampling, see Park (2002) and Chang and Park (2003). Although the difference-based bootstrap is known to yield an asymptotically valid approximation in the context of the linear autoregression, there is a concern about the power property of the bootstrap as demonstrated in Paparoditis and Politis (2003), and, therefore, this paper concentrates on the RBB.

such as the ADF test and Horvath and Watson (1995), fail to reject the null of no cointegration, while the Wald test of this paper and the supW test by Seo (2003) provide much stronger evidence for the LOP.

The remainder of this paper is organized as follows: Section 2 introduces the model and the Wald test, and develops the asymptotic theory for the test. The residual-based block bootstrap is introduced, and its asymptotic validity is established in Section 3. Section 4 presents simulation evidence for reasonable finite sample performance of the bootstrap. Section 5 illustrates the testing strategy through the study of the LOP in used car markets in the US. All proofs are collected in the appendix.

2 UNIT ROOT TESTING IN A TAR

Consider a threshold autoregressive model:

$$\Delta y_t = \alpha_1 y_{t-1} 1\{y_{t-1} \leq \gamma_1\} + \alpha_2 y_{t-1} 1\{y_{t-1} > \gamma_2\} + u_t, \tag{1}$$

$t = 1, \ldots, n$, where $1\{A\}$ is the indicator function that has value 1 if A is true, and value 0 otherwise. The threshold parameter $\gamma = (\gamma_1, \gamma_2)$, $\gamma_1 \leq \gamma_2$, is unknown and belongs to a compact set $\mathcal{S} \subset R^2$. The model (1), often called a band-TAR model, is a special case of a three-regime TAR model, and it becomes a two-regime TAR model when $\gamma_1 = \gamma_2$.

We want to test the null of a unit root process against the alternative of a stationary TAR process. Specifically, the null hypothesis of interest is :

$$H_0 : \alpha_1 = \alpha_2 = 0. \tag{2}$$

Note that the threshold effect also disappears under the null. Unfortunately, however, our understanding is not complete as to the stationarity conditions for general TAR processes. When the errors are independent, Chan, Petruccelli, Tong, and Woolford (1985) provide a necessary and sufficient condition of stationarity:

$$\alpha_1 < 0, \alpha_2 < 0, \text{ and } (\alpha_1 + 1)(\alpha_2 + 1) < 1, \tag{3}$$

which suggests that the natural alternative to H_0 should be

$$H_1 : \alpha_1 < 0 \text{ and } \alpha_2 < 0. \tag{4}$$

An important feature of the model (1) is that the threshold variable is not a lagged difference $y_{t-1} - y_{t-m}$ for some $m > 1$ but the lagged level y_{t-1}. The lagged difference is stationary, and Caner and Hansen (2001) develop a distribution theory for unit root testing in this case. Unlike the lagged difference, the lagged level y_{t-1} is nonstationary under the null of a unit root, so that the statistical inference for the testing relies on a totally different distribution theory from that of Caner and Hansen (2001).

2.1 TEST STATISTICS

The standard Wald statistic is employed to test the null (2). In particular, a constant and some lagged first difference terms are included in the estimation, as is common in the conventional unit root testing. That is, we compute the Wald statistic based on the following Dickey Fuller type regression:

$$\begin{aligned}\Delta y_t &= \hat{\alpha}_1(\gamma)\, y_{t-1} 1\{y_{t-1} \le \gamma_1\} + \hat{\alpha}_2(\gamma)\, y_{t-1} 1\{y_{t-1} > \gamma_2\} \\ &\quad + \hat{\mu}(\gamma) + \hat{\rho}_1(\gamma)\, \Delta y_{t-1} + \cdots + \hat{\rho}_p(\gamma)\, \Delta y_{t-p} + \hat{e}_t(\gamma), \end{aligned} \tag{5}$$

for each $\gamma \in \mathcal{S}$. We define $\hat{\sigma}^2(\gamma)$ as the residual variance from OLS estimation of (5), and $\hat{\sigma}_0^2$ as that of the null model. Then, we obtain the least squares (LS) estimators:

$$\hat{\gamma} = \arg\min_{\gamma \in \mathcal{S}} \hat{\sigma}^2(\gamma), \hat{\sigma}^2 = \hat{\sigma}^2(\hat{\gamma}), \hat{\alpha}_i = \hat{\alpha}_i(\hat{\gamma}), \text{ etc,} \tag{6}$$

and the Wald statistic:

$$W_n = n\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} - 1\right) = \sup_{\gamma \in \mathcal{S}} n\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2(\gamma)} - 1\right) = \sup_{\gamma \in \mathcal{S}} W_n(\gamma) \tag{7}$$

where $W_n(\gamma)$ is the standard Wald statistic with a fixed $\gamma \in \mathcal{S}$. The above equality follows, since $\hat{\sigma}_0^2$ is independent of γ, and $W_n(\gamma)$ is a decreasing function of $\hat{\sigma}^2(\gamma)$. Thus, the statistic W_n is the well-known "sup-Wald" statistic advocated by Davies (1987).

As will be shown shortly, the OLS estimator $\hat{\alpha}_i(\gamma)$ of the regression (5) is a consistent estimator for any p. In practice, we expect that $\hat{\alpha}_i(\gamma)$ is less biased than the OLS estimator from (1) by including those lagged difference terms. However, we do not attempt to increase the lag-order p as the sample size increases, as in Said and Dickey (1984), which is an approximation of an ARMA process by an infinite order AR process. Note that the approximation does not make sense in our case, since the error process u_t is not confined to an ARMA process.

2.2 ASYMPTOTIC DISTRIBUTION

We assume the following:

Assumption 1 *Let* $y_t = y_0 + \sum_{s=1}^{t} u_s$*, and* $\{u_t\}$ *be strictly stationary with mean zero and* $E|u_t|^{2+\delta} < \infty$*, for some* $\delta > 0$*, and strong mixing with mixing coefficients* a_m *satisfying* $\sum_{m=1}^{\infty} a_m^{1/2-1/(2+\delta)} < \infty$*. Furthermore,* $f_u(0) > 0$*, where* f_u *denotes the spectral density of* $\{u_t\}$*.*

Assumption 1 allows for conditional heteroskedasticity that is commonly observed in economic data. It differs from a finite order autoregressive process with iid innovations assumed in Bec, Guay, and Guerre (2002) or Kapetanios and Shin (2003).[2] The linear process need not be strong mixing, but can be shown to be strong mixing provided that the distribution of the innovation satisfies certain smoothness conditions (see section 14.4 of Davidson (1994)). Furthermore, the results developed in this paper including the bootstrap theory are readily modified to include the process.

The development of the asymptotic theory is based on the martingale approximation of Hansen (1992), which is based on the following representation of $\{u_t\}$. Define

$$\varepsilon_t = \sum_{s=0}^{\infty} \left(\mathrm{E}_t u_{t+s} - \mathrm{E}_{t-1} u_{t+s}\right), \ \zeta_t = \sum_{s=1}^{\infty} \mathrm{E}_t u_{t+s},$$

where $\mathrm{E}_t X = \mathrm{E}(X|\mathcal{F}_t)$ and $\mathcal{F}_t$ is the typical natural filtration. Then,

$$u_t = \varepsilon_t - \left(\zeta_t - \zeta_{t-1}\right), \tag{8}$$

and $\{\varepsilon_t\}$ is a martingale difference sequence, and $\{u_t\zeta_t - Eu_t\zeta_t\}$ is a uniformly integrable L^1-mixingale (see Hansen (1992), p. 499). The representation (8) yields a useful decomposition of a stochastic integral process:

$$\frac{1}{\kappa_n}\sum_{t=2}^{n} f\left(y_{t-1},\theta\right) u_t = \frac{1}{\kappa_n}\sum_{t=2}^{n} f\left(y_{t-1},\theta\right)\varepsilon_t + D_n, \tag{9}$$

where

$$D_n = \frac{1}{\kappa_n}\sum_{t=2}^{n}\left(f\left(y_t,\theta\right) - f\left(y_{t-1},\theta\right)\right)\zeta_t - \frac{1}{\kappa_n} f\left(y_{n-1},\theta\right)\zeta_n.$$

Of our interest is the weak convergence of the stochastic integral on a compact parameter space, say Θ, when $f(x,\theta) = x1\{x \le \theta\}$, $\theta \in \Theta$. Although the convergence of the first term in the right hand side of (9) and the convergence rate κ_n are developed by Park and Phillips (2001) for various classes of functions $f(\cdot,\theta)$, the development is made fixing θ at a give value. Hansen (1992) provides a limit of the bias D_n when $f(x) = x$ or x^2, and De Jong (2003) extends the result to one of the classes of

[2] I found these two working papers, independently developed of this paper, studying the simliar testing problem as the one in this paper. However, They do not develop any bootstrap theory.

functions considered by Park and Phillips (2001), which is different from the class of functions of our interest.

To state the asymptotic development, we introduce some notations. Let "$[x]$" denote the integer part of x, and "$\Rightarrow$" the weak convergence with respect to the uniform metric on the parameter space. Next, define the autocovariance function $r(k) = \mathrm{E} u_t u_{t+k}$ and let $\sigma^2 = r(0), \lambda = \sum_{s=1}^{\infty} r(s)$ and the long-run variance $\omega^2 = \sum_{s=-\infty}^{\infty} r(s) = \sigma^2 + 2\lambda$. Assume y_0 is zero for the sake of simplicity. It is well-known that, under Assumption 1, $\frac{1}{\sqrt{n}} y_{[nr]} = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t \Rightarrow B(r)$ where B is a Brownian motion with variance ω^2. We denote $\int_0^1 B(r)\,dr$ as $\int_0^1 B$, and similarly $\int_0^1 B(r)\,dB(r)$ as $\int_0^1 BdB$.

The following lemma is useful to develop the asymptotics of the bias term D_n, and stated more generally than necessary for our specific purpose. However, its generality may prove useful in other cases.

Lemma 1 *Suppose the sequence $\{w_t - \mu_w\}$ is a uniformly integrable L^1-mixingale and Assumption 1 holds. Then, for $k = 0, 1, 2, \ldots$*

$$\frac{1}{n^{1+k/2}} \sum_t y_t^k 1\{y_t \le \theta\} w_t \Rightarrow \mu_w \int_0^1 B^k 1\{B \le 0\}$$

on Θ.

Lemma 1 shows that $\left(y_t 1\{y_t \le \theta\}/\sqrt{n}\right)^k$ and w_t are asymptotically uncorrelated. Similar asymptotic uncorrelatedness between stationary process and nonstationary process is can be found in Theorem 3.3 of Hansen (1992) and Theorem 3 of Caner and Hansen (2001). The case of $k = 0$ is required for the following main theorem and of interest as the transformation of the limit Brownian motion is discontinuous. Unlike linear autoregression in which the bias is λ, we have the bias $\lambda \int_0^1 1\{B \le 0\}$ in the following theorem.

Theorem 2 *Under Assumption 1,*

$$\frac{1}{n} \sum_t y_{t-1} 1\{y_{t-1} \le \theta\} u_t \Rightarrow \int_0^1 B \cdot 1\{B \le 0\}\, dB + \lambda \int_0^1 1\{B \le 0\}$$

on Θ.

Now we turn to the convergence of the parameter estimators and the Wald statistic. To ease the exposition of the main theorem, we introduce some notations:

$$\begin{aligned}
\bar{B}_L &= B1\{B \le 0\} - \int_0^1 B1\{B \le 0\}, \ \bar{B}_U = B1\{B > 0\} - \int_0^1 B1\{B > 0\}, \\
G_p &= \begin{pmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r(1) & r(0) & \cdots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0) \end{pmatrix}, \ \tilde{r}_p = \begin{pmatrix} r(0) \\ r(0) + r(1) \\ \vdots \\ r(0) + \cdots + r(p-1) \end{pmatrix},
\end{aligned}$$

and $g_p = (r(1), \cdots, r(p))'$. Let ι_p be a p-dimensional vector of ones. The following theorem provides the limit distribution of $\hat{\alpha}_i, i = 1, 2$ and that of W_n.

Theorem 3 *Suppose that Assumption 1 holds. Then, as* $n \to \infty$,

$$(i) \begin{pmatrix} n\hat{\alpha}_1(\gamma) \\ n\hat{\alpha}_2(\gamma) \end{pmatrix} \Rightarrow \begin{pmatrix} \int_0^1 \bar{B}_L^2 & -\int_0^1 \bar{B}_L \bar{B}_U \\ -\int_0^1 \bar{B}_L \bar{B}_U & \int_0^1 \bar{B}_U^2 \end{pmatrix}^{-1} \begin{pmatrix} A_{p,L} \\ A_{p,U} \end{pmatrix},$$

$$(ii)\ W_n \Rightarrow \left(\sigma^2 - g_p' G_p^{-1} g_p\right)^{-1} \begin{pmatrix} A_{p,L} & A_{p,U} \end{pmatrix} \begin{pmatrix} \int_0^1 \bar{B}_L^2 & -\int_0^1 \bar{B}_L \bar{B}_U \\ -\int_0^1 \bar{B}_L \bar{B}_U & \int_0^1 \bar{B}_U^2 \end{pmatrix}^{-1} \begin{pmatrix} A_{p,L} \\ A_{p,U} \end{pmatrix},$$

on $\mathcal{S}$, *where*

$$\begin{aligned}
A_{p,L} &= \left(1 - g_p' G_p^{-1} \iota\right) \left(\int_0^1 \bar{B}_L dB + \lambda \int_0^1 1\{B \le 0\} \right) - g_p' G_p^{-1} \tilde{r}_p \int_0^1 1\{B \le 0\} \\
A_{p,U} &= \left(1 - g_p' G_p^{-1} \iota\right) \left(\int_0^1 \bar{B}_U dB + \lambda \int_0^1 1\{B > 0\} \right) - g_p' G_p^{-1} \tilde{r}_p \int_0^1 1\{B > 0\}.
\end{aligned}$$

Due to the recursion property of Brownian motion, the limit distributions above are well-defined, even though they are nonstandard and nonconventional. They depend on nuisance parameters, such as $\omega^2, \lambda, r(0), ..., r(p)$. The dependence on the data structure is quite complicated, and thus critical values cannot be tabulated. In the next section, we turn to a bootstrap procedure to approximate the sampling distribution of W_n.

3 RESIDUAL-BASED BLOCK BOOTSTRAP

We discuss a bootstrap approximation of the sampling distribution of the statistic (7). Under nonstationarity we cannot resample the data $\{y_t\}$ directly. Instead, we apply conventional bootstrap methods developed for dependent data to the first differences of y_t or to regression residuals using consistent estimators $\hat{\alpha}_i'$s, and then integrate the resampled ones. Specifically, we apply the overlapping block bootstrap of Künsch (1989) to the appropriately centered residuals and call it a residual-based block bootstrap (RBB).

In the context of conventional unit root testing, Paparoditis and Politis (2003) introduce the idea of RBB and establish its asymptotic validity. Although our RBB is similar to theirs, it should be noted that our procedure is distinguished in the manner of constructing centered residuals. Furthermore, the convergence of our bootstrapped Wald statistic is more complicated due to the involved nonlinearity.

Compared to a difference-based bootstrap, the RBB is believed to have a power advantage, because it does not impose the null hypothesis to obtain the residuals. Paparoditis and Politis (2003) show analytically that this is true in case of block bootstrap scheme in the context of conventional unit root testing. Expecting similar power advantage, we focus on the RBB. Following the convention in the literature, we denote the bootstrap quantities such as sample, probability measure, expectation, variance, etc, with an asterisk $*$.

3.1 BOOTSTRAP ALGORITHM

The RBB resampling is conducted through the following algorithm: First, we define the residuals using the estimators $\hat{\alpha}_i, \hat{\gamma}_i, i = 1, 2$ in (6) as

$$\hat{u}_t = \Delta y_t - \hat{\alpha}_1 y_{t-1} 1\{y_{t-1} \leq \hat{\gamma}_1\} - \hat{\alpha}_2 y_{t-1} 1\{y_{t-1} > \hat{\gamma}_2\}, \ t = 2, \ldots, n, \tag{10}$$

and then calculate centered residuals

$$\tilde{u}_t = \hat{u}_t - \frac{1}{n-b} \sum_{i=1}^{n-b} \frac{1}{b} \sum_{j=1}^{b} \hat{u}_{i+j}, \ t = 2, \ldots, n. \tag{11}$$

The centering here appears crucial to a valid bootstrap approximation, since $\hat{u}_t$ in (10) does not have mean zero while u_t has mean zero. However, the centering in (11) is different from that of Paparoditis and Politis (2003), who adopt a simple demeaning.

Second, we resample $\tilde{u}_t$ by the overlapping blocking scheme of Künsch (1989) and integrate the resampled $\tilde{u}_t'$s to get an integrated process. Specifically, for a positive integer $b\,(< n)\,, k = [(n-1)/b]$ and $l = kb + 1$, a bootstrap pseudo-series $y_1^*\,(= y_1)\,, \ldots, y_l^*$ is constructed as follows

$$y_t^* = y_{t-1}^* + \tilde{u}_{i_m + s}, \ t = 2, \ldots, l,$$

where $m = [(t-2)/b]\,, s = t - mb - 1$ and $i_0, i_1, \ldots, i_{k-1}$ are random variables drawn i.i.d. uniform on the set $\{1, 2, \ldots, n-b\}$.

Third, we compute the pseudo-statistic W_l^* using the generated bootstrap sample $y_1^*, \ldots, y_l^*$ as defined in (7). Repeating the second and third steps sufficiently many times, we can obtain the empirical quantiles of the bootstrap statistic W_l^*.

The centering (11) is commonly observed in the standard block bootstrap literature because the block bootstrapped series is not stationary, even in the first moment. It ensures that the expectation (based on the bootstrap distribution) of the sample mean of Δy_t^*, not of Δy_t^* itself, is zero. Hall, Horowitz, and Jing (1995) show that the centering (11) accelerates the convergence of the estimation error of the block bootstrap in case of a sample mean statistic with stationary data. This higher order property is, however, unclear in our case, in which an integration is involved.

3.2 CONSISTENCY OF THE RBB

In this subsection, we show that the bootstrap statistic W_l^* converges to the proper limit distribution defined in Theorem 3. Since the distribution of the statistic W_l^* depends on each realization of $\{y_t\}$, we define the following notation:

$$T_n^* \Rightarrow T \; in \; \mathrm{P},$$

meaning that the distance between the law of a statistic T_n^* based on the bootstrap sample and that of a random measure T tends to zero in probability for any distance metrizing weak convergence (see Paparoditis and Politis (2003)).

First, we establish an invariance principle for our bootstrap. Define a standardized partial sum process $\{S_l^*(r), 0 \le r \le 1\}$ by

$$\begin{aligned} S_l^*(r) &= \frac{1}{\sqrt{l}\omega^*}\sum_{t=1}^{j-1} u_t^* \text{ for } \frac{j-1}{l} \le r < \frac{j}{l} \;\; (j = 2, \ldots, l) \\ &= \frac{1}{\sqrt{l}\omega^*}\sum_{t=1}^{l} u_t^* \text{ for } r = 1, \end{aligned}$$

where $u_1^* = y_1^*, u_t^* = y_t^* - y_{t-1}^*$ for $t = 2, 3, \ldots, l$ and ${\omega^*}^2 = var^*\left(l^{-1/2}\sum_{j=2}^{l} u_j^*\right)$. Due to the fast rate of convergence of $\hat{\alpha}_i'$s, it can be shown that the partial sum process of the resampled centered residuals (11) is asymptotically equivalent to that of the resampled $\{u_t\}$. The invariance principle for the latter is derived in Paparoditis and Politis (2003). Let W denote a standard Brownian motion. Then we have the following invariance principle.

Theorem 4 *Suppose that Assumption 1 holds. If $b \to \infty$ such that $b/\sqrt{n} \to 0$ as $n \to \infty$, then*

$$S_l^* (r) \Rightarrow W(r) \text{ in } P,$$

on $[0,1]$.

Next, we establish the convergence of the stochastic integral for the RBB, i.e., a bootstrap version of Theorem 2. Without loss of generality, assume $u_t^* = 0$ if $t > l$. Let $u_t^* = \varepsilon_t^* - \left(\zeta_t^* - \zeta_{t-1}^*\right)$ where $\varepsilon_t^* = \sum_{j=0}^{\infty} \left(\mathrm{E}_t^* u_{t+j}^* - \mathrm{E}_{t-1}^* u_{t+j}^*\right)$ and $\zeta_t^* = \sum_{j=1}^{\infty} \mathrm{E}_t^* u_{t+j}^*$. Recalling that $u_t^* = \tilde{u}_{i_m+s}$, we have

$$E_t^* u_{t+j}^* = \left\{ \begin{array}{ll} u_{t+j}^*, & \text{if } j \le b-s \\ E^* u_{t+j}^*, & \text{o/w}, \end{array} \right. ,$$

so that

$$\varepsilon_t^* = \left\{ \begin{array}{ll} \sum_{j=0}^{b-1} u_{t+j}^*, & \text{if } s = 1, \\ 0, & \text{o/w}, \end{array} \right. , \text{ and } \zeta_t^* = \sum_{j=1}^{b-s} u_{t+j}^*,$$

since $\sum_{j=1}^{b} E^* u_{t+j}^* = 0$ for any $t \ge 2$. Note that ε_t^* is, in fact, an independent resampling of the sums of blocks. Based on this decomposition, we derive the following convergence of the stochastic integral for our RBB.

Theorem 5 *Suppose that Assumption 1 holds. If $b \to \infty$ such that $b/\sqrt{n} \to 0$ as $n \to \infty$, then*

$$\begin{array}{lll} (i) \ l^{-1-k/2} \sum_{t=2}^{l} {y_{t-1}^*}^k 1\left\{y_{t-1}^* \le \theta\right\} & \Rightarrow & \int_0^1 B^k 1\{B \le 0\} \text{ in } P, \\ (ii) \ l^{-1} \sum_{t=2}^{l} y_{t-1}^* 1\left\{y_{t-1}^* \le \theta\right\} u_t^* & \Rightarrow & \int_0^1 B 1\{B \le 0\}\, dB + \lambda \int_0^1 1\{B \le 0\} \text{ in } P. \end{array}$$

on Θ.

The consistency of the RBB of W_n follows.

Theorem 6 *Suppose that Assumption 1 holds. If $b \to \infty$ such that $b/\sqrt{n} \to 0$ as $n \to \infty$, then*

$$W_n^* \Rightarrow \left(\sigma^2 - g_p' G_p^{-1} g_p\right)^{-1} \left(\begin{array}{cc} A_{p,L} & A_{p,U} \end{array} \right) \left(\begin{array}{cc} \int_0^1 \bar{B}_L^2 & -\int_0^1 \bar{B}_L \bar{B}_U \\ -\int_0^1 \bar{B}_L \bar{B}_U & \int_0^1 \bar{B}_U^2 \end{array} \right)^{-1} \left(\begin{array}{c} A_{p,L} \\ A_{p,U} \end{array} \right) \text{ in } P,$$

on $\mathcal{S}$.

4 MONTE CARLO SIMULATION

This section examines the finite sample performance of the RBB of W_n, compared to that of the conventional ADF test. For the sake of fair comparison, we also apply RBB to the conventional ADF test as explained in Paparoditis and Politis (2003). Due to the heavy burden of computation, the number of simulation repetitions and of bootstrap iterations is set at 200 in the following computations.

Several details remain to be determined in order to implement the RBB in practice. One is the selection of the block length b and the lag order p. In this experiment we try several values of b and p to see how dependent the performance of the RBB is on those choices. We do not attempt a data-dependent method. Another is the matter of how to set the parameter space $\mathcal{S}$. Typically, we construct the set $\mathcal{S}$ in the form of an interval, such as $[-\bar{\gamma}, \bar{\gamma}]$, where $\bar{\gamma}$ is a sample quantile of $|y_t|$. Since $\bar{\gamma}$ should be bounded and, $\{y_t\}$ is an integrated process, we need to use lower and lower quantiles as the sample size increases. In practice, however, this argument does not answer the question of which quantile is optimal. In this regard, the bootstrap-based inference is expected to have some advantage over the asymptotics based inference, since the dependence on the particular choice of quantile is replicated by bootstrap.

In this experiment, we let $\bar{\gamma} = \max_{t \leq n} |y_{t-1}|$ due to the small sample size, and then compute a bootstrap statistic W_l^*, taking supremum of $W_l^*(\gamma)$ over the set

$$\left\{ \gamma_1, \gamma_2 \in [-\bar{\gamma}, \bar{\gamma}], \left| \sum_{t=2}^{n} 1\left\{ y_{t-1}^* \leq \gamma_1 \right\} \geq m \text{ and } \sum_{t=2}^{n} 1\left\{ y_{t-1}^* > \gamma_2 \right\} \geq m, \right. \right\}. \tag{12}$$

As mentioned above, $\bar{\gamma}$ could be other quantiles, and the choice of quantile does not matter much in RBB, as long as $\bar{\gamma}$ is not too low a quantile. The restriction in (12) is to guarantee that the regime-specific parameters α_i's are estimated properly; however, the constraint (12) is not binding in large samples because of the recursion property of the Brownian motion. The number m is set at 10 in our experiment.

We generate data from the following process:

$$\begin{aligned} \Delta y_t &= \alpha y_{t-1} 1\{|y_{t-1}| > \gamma\} + u_t \quad (13) \\ u_t &= \rho u_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \end{aligned}$$

where $\{\varepsilon_t\}$ follows iid standard normal distributions. As is common in the conventional unit root testing

		(ρ, θ)	$(0,0)$	$(-.5,0)$	$(.5,0)$	$(0,-.5)$	$(0,.5)$
ADF	$b=6$	$p=3$	.050	.065	.075	.085	.050
		$p=6$	.030	.065	.060	.045	.050
	$b=8$	$p=3$	.065	.065	.060	.080	.050
		$p=6$	.075	.075	.070	.060	.050
W_n	$b=6$	$p=3$	.040	.065	.070	.080	.075
		$p=6$	.055	.055	.090	.040	.060
	$b=8$	$p=3$	.080	.050	.060	.105	.100
		$p=6$	.085	.035	.070	.065	.095

Note: $n=100$. Nominal size 5%. RBB-based inference.

Table 1: Size of unit root tests

			(ρ, θ)	$(0,0)$	$(-.5,0)$	$(.5,0)$	$(0,-.5)$	$(0,.5)$
$n:100$	ADF	$b=6$	$\gamma=4$	.140	.115	.195	.125	.190
			$\gamma=8$	.110	.080	.145	.110	.105
	W_n	$b=6$	$\gamma=4$	.165	.140	.190	.115	.205
			$\gamma=8$	.195	.070	.195	.095	.180
$n:250$	ADF	$b=6$	$\gamma=4$	.380	.270	.755	.305	.745
			$\gamma=8$	.140	.105	.230	.140	.270
		$b=10$	$\gamma=4$	.415	.245	.730	.295	.700
			$\gamma=8$	.200	.125	.260	.185	.225
	W_n	$b=6$	$\gamma=4$	.715	.460	.735	.425	.765
			$\gamma=8$	.475	.250	.695	.145	.665
		$b=10$	$\gamma=4$	.725	.490	.645	.460	.740
			$\gamma=8$	.435	.245	.700	.250	.720

Note: Nominal size 5%. $p=3$. RBB-based inference.

Table 2: Power of unit root tests

literature, we consider the following combination of (ρ, θ) :

$$(0,0), (-0.5,0), (0.5,0), (0,-0.5), \text{ and } (0,0.5).$$

When α is not zero, we set the threshold parameter γ as 4 or 8. As the parameter γ increases, the no-adjustment region becomes larger, which may have an influence on the power of the tests.

We first study the size of nominal 5% tests. The data are simulated from (13) with $\alpha=0$ and with the sample size $n=100$. Note that there is no threshold effect if $\alpha=0$. We choose the block length $b=6,8$ and the lag order $p=3,6$. The rejection frequencies are reported in Table 1, which shows that both the ADF and W_n tests have reasonable size for most error types, even in this small sample size. One exception may be that W_n slightly over-rejects when the error u_t has a MA component, $b=8$ and $p=3$. It is difficult to find any dependence structure of rejection on the block length b or the lag order p.

Next, we examine the power properties of the tests. Let $\alpha = -0.1$ and let γ equal either 4 or 8. We also consider two different sample sizes of 100 and 250. The tests are the same as before except that the lag order p is fixed at 3. The block length b is 6 at the sample size 100, and both 6 and 10 at the sample size 250. Table 2 reports the simulation results. Across most parametrizations, W_n has better power than ADF, which can be seen more clearly as the sample size n increases from 100 to 250. Especially when $n = 250$ and $\gamma = 8$, the rejection frequencies of W_n are about two or three times higher than those of ADF, regardless of the block length. We can also find that the increase of the threshold parameter γ results in the decrease of power for both the ADF and W_n tests, more or less. This drop of power is natural in the sense that the higher γ means the broader no-adjustment region. Yet, this change in γ deteriorates ADF much more than it does W_n. For example, see the case with $(\rho, \theta) = (0, 0), (0.5, 0)$ and $(0, 0.5)$ when $n = 250$. Another feature of the simulation results is that both tests have relatively low power when the error u_t has a negative AR or MA component. This is due to the fact that the proportion of the no-adjustment region is higher for a given γ in those cases than in the other cases.

5 AN EMPIRICAL ILLUSTRATION

We investigate the law of one price (LOP) hypothesis amongst used car markets in the US. US Bureau of Labor Statistics Monthly Consumer Price Indexes of 29 different locations[3] are used for the period of December 1986-June 1996 (115 observations). This data set is the same as the one used in Lo and Zivot (2001) and in Seo (2003). Like these researchers, 28 bivariate systems of log prices are constructed with a benchmark city, New Orleans.

We perform several different cointegration tests using the cointegrating vector $(1, -1)'$ implied by the LOP. We compare the tests developed for threshold cointegration with those developed for conventional cointegration. In the univariate framework, we compare the RBB of the Wald statistic W_n developed in this paper to that of the conventional ADF. In the multivariate framework, the supW test of Seo (2003) is compared with the Wald test by Horvath and Watson (1995) (HW). In order to compute the p-values for the supW statistic, we follow the residual-based bootstrap developed in Seo (2003). All the bootstrap p-values are computed with 500 replications. For the lag order selection, the Schwarz

[3]Cities Abbv. : 1. Anchorage AN, 2. Atlanta AT, 3. Baltimore BT, 4. Boston BO, 5. Buffalo BU, 6. Chicago CH, 7. Cleveland CL, 8. Cincinnati CI, 9. Dallas DA, 10. Denver DN, 11. Detroit DT, 12. Honolulu HO, 13. Houston HS, 14. Kansas City KC, 15. Los Angeles LA, 16. Miami MA, 17. Milwaukee MI, 18. Minneapolis MS, 19. New York NY, 20. Philadelphia PH, 21. Pittsburgh PI, 22. Portland PO, 23. San Diego SD, 24. San Francisco SF, 25. Seattle SE, 26. St. Louis SL, 27. Tampa Bay TA, 28. Washington D.C. DC, Benchmark: New Orleans NO

Cities	supW(p-value)	HW(10%:8.3)	W_n(p-value)	ADF(p-value)
NY	0.009	1.214	0.96	0.9
PH	0.661	1.109	0.89	0.87
CH	0.22	0.285	0.85	0.96
LA	0.372	1.339	0.91	0.88
SF	0.7	1.549	0.62	0.89
Bo	0.293	0.7	0.93	1
Cl	0.432	5.371	0.32	0.56
Ci	0.71	2.017	0.03	1
DC	0.839	0.822	0.67	0.9
Ba	0.061	3.621	0.44	0.43
SL	0.087	0.03	0.43	0.97
MS	0.567	0.062	0.28	0.98
Ma	0.052	0.541	0.29	0.85
SD	0.299	3.789	0.41	0.37
Po	0.515	1.572	0.06	0.93
BU	0.038	0.707	0.77	0.98
DA	0.034	0.264	0.01	0.9
AT	0.064	7.714	0.27	0.5
AN	0.091	1.204	0.47	1
DN	0.044	2.838	0.17	0.1
DT	0.536	0.342	0.89	0.92
MI	0.677	2.86	0.08	0.88
KC	0.299	0.507	0.13	0.92
HS	0.506	0.511	0.85	0.88
HO	0.155	1.061	0.5	0.83
PI	0.888	0.575	0.01	0.95
TA	0.556	2.714	0.57	0.57
SE	0.854	0.416	0.23	0.9

Table 3: Tests for LOP in used car markets from 29 different locations (28 bivariate systems with a benchmark city, New Orleans)

Criterion (BIC) is used based on linear models for comparison.

Table 3 reports the results of these four tests. In the case of the conventional methods, HW and ADF, it is very hard to find evidence for the LOP. Only two systems are rejected for the LOP by those tests, on the basis of which Lo and Zivot (2001) argue that there is no LOP in the used car markets, since used cars are more heterogeneous than other categories that they consider, such as meats, fresh fruits and vegetables. Yet, it is unclear why used cars are more heterogeneous than meats or vegetables across locations. In contrast to those conventional tests, the RBB of W_n or the supW test provides much more evidence for the LOP in the used car markets. Thirteen systems are rejected for the LOP. These different testing results indicate that we need be more cautious about the used car markets.

An important feature of the used car markets is the presence of uncertainty of quality, which has

	Used car market	New car market
SupW	9	14
W_n	5	10
either SupW or W_n	13	17
HW	1	17
ADF	1	14
either HW or ADF	2	18

# of rejections of no cointegration out of 28 systems at 10% size

Table 4: LOP in used car and new car market

generated a large empirical debate as to whether the used car markets are efficient or not. An implication of the debate for the study of the LOP is that the uncertainty acts as a kind of transaction barrier in the used car markets. In other words, in the band type threshold cointegration models, a longer no-adjustment period is expected in used car markets than in, for example, new car markets. We make a simple comparison between the two markets by applying all four tests to the new car markets. The results are summarized in Table 4. In contrast to the testing results of the used car markets, that of the new car markets do not seem to depend on whether we use the threshold models or the linear models. This may imply that the nominal transaction costs are not significant enough to affect the power of the conventional tests, and that transaction barriers, such as the uncertainty, can be more important in analysis of the LOP in the used car markets. Arguably, a significant power loss of the conventional linear cointegration tests is likely, and so those tests can be quite misleading.

6 Conclusion

In this paper, we have developed the Wald statistic and the RBB to test the null hypothesis of a unit root in the threshold autoregression. From our simulation, the RBB of the Wald statistic outperforms the ADF test when the alternative is a stationary threshold autoregression, and vice versa when it is a stationary linear process. In practice, it will be prudent to apply both methods and to interpret the rejection by any of the two tests as the evidence for the rejection of the presence of unit root in the process.

An important avenue of future research will be the following two: First, the intercept also plays an important role in determining the stationarity of a threshold process. We did not pursue this issue in this paper, but it should be done to fully understand the stationarity of the process. Second, there is another hypothesis of interest in which the process is a nonstationary threshold process[4]. It is much more involved to develop a test to investigate this null hypothesis, since we do not have any asymptotic theory applicable to the process under this null.

A Proof of Theorems

We first prove the following lemma, which will be repeatedly used to prove main theorems. It is more general than required and holds for any bounded intervals.

Lemma 7 *Suppose $\{a_t, b_t\}$ is strictly stationary with $E|a_t| < \infty$ and $E|b_t| < \infty$, and $\{w_t\}$ is uniformly integrable. Then,*

$$E\left(\frac{1}{n}\sum_{t=1}^{n}\left|1\left\{a_t < y_t < b_t\right\}w_t\right|\right) \to 0.$$

Proof of Lemma Fix $\varepsilon > 0$.For any $c > 0$, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\mathrm{E}\left|1\left\{a_t < y_t < b_t\right\}w_t\right| \\
\leq\ & c\frac{1}{n}\sum_{t=1}^{n}\mathrm{E}1\left\{a_t < y_t < b_t\right\} + \sup_t \mathrm{E}\left[|w_t|\,1\left\{|w_t| > c\right\}\right]
\end{aligned}$$

First, due to the uniform integrability of $\{w_t\}$, there is a constant c s.t. $\sup_t \mathrm{E}[|w_t|\,1\{|w_t| > c\}] < \varepsilon/2$. Second, for any $M > 0$,

$$\mathrm{E}1\left\{a_t < y_t < b_t\right\} \leq \mathrm{E}1\left\{|y_t| < M\right\} + \Pr\left\{a_t < -M \text{ or } b_t > M\right\}, \tag{14}$$

and there exists $N(c)$ s.t., for all $n > N(c)$ and for small enough $\delta > 0$,

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\mathrm{E}1\left\{|y_t| < M\right\} &= \int_0^1 \mathrm{E}1\left\{\left|\frac{y_{[nr]}}{\sqrt{n}}\right| < \frac{M}{\sqrt{n}}\right\} \leq \int_0^1 \mathrm{E}1\left\{\left|\frac{y_{[nr]}}{\sqrt{n}}\right| < \delta\right\} \\
&\leq \int_0^1 \mathrm{E}1\left\{|B(r)| < \delta\right\} + \frac{\varepsilon}{8c} \leq \frac{\varepsilon}{4c},
\end{aligned} \tag{15}$$

by the weak convergence of $\frac{y_{[nr]}}{\sqrt{n}}$ and by the dominated convergence theorem. Furthermore, we can choose M large enough so that $\sup_t \Pr\left\{a_t < -M \text{ or } b_t > M\right\} < \frac{\varepsilon}{4c}$ by the strict stationarity of $\{a_t, b_t\}$.Thus, it follows from (14) and (15) that

$$c\frac{1}{n}\sum_{t=1}^{n}\mathrm{E}1\left\{a_t < y_t < b_t\right\} \leq \frac{\varepsilon}{2}$$

for all $n > N(c)$. Since ε is arbitrary, the proof is complete. ■

Remark 1 *The uniform integrability is quite general in that it includes any strictly stationary process with a finite first moment, any weakly stationary with a finite second moment.*

[4] That is, one of the $\alpha_i, i = 1, 2$ is nonzero, but the other is zero. Caner and Hansen (2001) call this a partial unit root process.

A.1 Proof of Theorem 1

For simplicity of exposition, assume $\theta \geq 0$ and let $\bar{\theta} = \max\{\theta \in \Theta\}$. Since $y_t^k 1\{0 < y_t \leq \theta\} \leq \bar{\theta}^k 1\{0 < y_t \leq \bar{\theta}\}$, we note that

$$\sup_{0\leq\theta\leq\bar{\theta}} \left| \frac{1}{n^{1+k/2}} \sum_{t=1}^{n} y_t^k 1\{0 < y_t \leq \theta\} w_t \right| \leq \frac{\bar{\theta}^k}{n^{k/2}} \frac{1}{n} \sum_{t=1}^{n} \left| 1\{0 < y_t \leq \bar{\theta}\} w_t \right|, \tag{16}$$

which is $o_p(1)$ due to Lemma 7. Therefore, it remains to show the convergence of

$$\frac{1}{n^{1+k/2}} \sum_{t=1}^{n} y_t^k 1\{y_t \leq 0\} w_t.$$

When $k > 0$, theorem 3.3 of Hansen (1992) applies directly to get the results. When $k = 0$, however, the transformation is not continuous and the theorem is not applicable.

Let $v_t = w_t - \mu_w$ and $1_\delta(y) = 1\{y \leq 0\} + \left(1 - \frac{1}{\delta} y\right) 1\{0 < y \leq \delta\}$ for some $\delta > 0$, and write that

$$\frac{1}{n} \sum_{t=1}^{n} 1\{y_t \leq 0\} v_t = \frac{1}{n} \sum_{t=1}^{n} 1_\delta(y_t) v_t - \frac{1}{n} \sum_{t=1}^{n} \left(1 - \frac{y_t}{\delta}\right) 1\{0 < y_t \leq \delta\} v_t.$$

The first term on the right hand side is $o_p(1)$ due to Theorem 3.3 of Hansen (1992). Noting that

$$\left| \left(1 - \frac{y_t}{\delta}\right) 1\{0 < y_t \leq \delta\} v_t \right| \leq \left| 1\{0 < y_t \leq \delta\} v_t \right|,$$

we conclude the second term is also $o_p(1)$ by Lemma 7. Finally, it follows from Theorem 3.1 of Park and Phillips (2001) that

$$\mu_w \frac{1}{n} \sum_{t=1}^{n} 1\{y_t \leq 0\} \Rightarrow \mu_w \int_0^1 1\{B \leq 0\}.$$

■

A.2 Proof of Theorem 2

Due to (16), we only have to develop the convergence of $\frac{1}{n}\sum_{t=2}^{n} y_{t-1} 1\{y_{t-1} \leq 0\} u_t$. Based on (8), write

$$\frac{1}{n} \sum_{t=2}^{n} y_{t-1} 1\{y_{t-1} \leq 0\} u_t = \frac{1}{n} \sum_{t=2}^{n} y_{t-1} 1\{y_{t-1} \leq 0\} \varepsilon_t + L_n + R_{1n} + R_{2n},$$

where

$$\begin{aligned} L_n &= \frac{1}{n} \sum_{t=2}^{n} u_t \zeta_t \cdot 1\{y_{t-1} \leq 0\} 1\{y_t \leq 0\} \\ R_{1n} &= \frac{1}{n} y_{n-1} 1\{y_{n-1} \leq 0\} \zeta_n \\ R_{2n} &= \frac{1}{n} \sum_{t=2}^{n} \left(y_t 1\{y_t \leq 0 < y_{t-1}\} + y_{t-1} 1\{y_{t-1} \leq 0 < y_t\}\right) \zeta_t. \end{aligned}$$

Since the transformation $s1\{s \le 0\}$ is continuous and $\{\varepsilon_t\}_{t=1}^n$ is a martingale difference sequence, it follows from Kurtz and Protter (1991) that

$$\frac{1}{n}\sum_{t=2}^{n} y_{t-1} 1\{y_{t-1} \le 0\} \varepsilon_t \Rightarrow \int_0^1 B \cdot 1\{B \le 0\} dB.$$

Next, since $y_t = y_{t-1} + u_t$ and $u_t\zeta_t$ is uniformly integrable, we note that, by Lemma 7,

$$\begin{aligned} & \frac{1}{n}\sum_{t=2}^{n} u_t\zeta_t \cdot 1\{y_{t-1} \le 0\} 1\{y_t \le 0\} - \frac{1}{n}\sum_{t=2}^{n} u_t\zeta_t \cdot 1\{y_t \le 0\} \\ = & \frac{1}{n}\sum_{t=2}^{n} u_t\zeta_t \cdot 1\{y_{t-1} + u_t \le 0 < y_{t-1}\} \\ = & o_p(1). \end{aligned}$$

Then, since $u_t\zeta_t - \lambda$ is a uniformly integrable L^1-mixingale, it follows from Theorem 1 that

$$L_n = \frac{1}{n}\sum_{t=2}^{n} u_t\zeta_t \cdot 1\{y_t \le 0\} + o_p(1) \Rightarrow \lambda \int_0^1 1\{B \le 0\}.$$

Finally, we show that R_{1n} and R_{2n} are $o_p(1)$. First, note that

$$\begin{aligned} \sup_{t \le n} \frac{1}{n} \left| y_t 1\{y_t \le 0\} \zeta_{t+1} \right| &\le \sup_{t \le n} \frac{1}{\sqrt{n}} |y_t| \sup_{t \le n} \frac{1}{\sqrt{n}} \left|\zeta_{t+1}\right| \\ &= O_p(1) o_p(1). \end{aligned}$$

Second, by replacing y_t by $y_{t-1} + u_t$, we have

$$\begin{aligned} & y_t 1\{y_t \le 0 < y_{t-1}\} + y_{t-1} 1\{y_{t-1} \le 0 < y_t\} \\ = & y_{t-1}\left(1\{0 < y_{t-1} \le -u_t\} + 1\{-u_t < y_{t-1} \le 0\}\right) + u_t 1\{0 < y_{t-1} \le -u_t\}, \end{aligned}$$

and, therefore,

$$\begin{aligned} |R_{2n}| &\le \frac{1}{n}\sum_{t=2}^{n} \left| y_{t-1}\zeta_t \left(1\{0 < y_{t-1} \le -u_t\} + 1\{-u_t < y_{t-1} \le 0\}\right)\right| \\ &\quad + \frac{1}{n}\sum_{t=2}^{n} \left| u_t\zeta_t 1\{0 < y_{t-1} \le -u_t\}\right| \\ &\le \frac{1}{n}\sum_{t=2}^{n} \left(|y_{t-1}\zeta_t| + |u_t\zeta_t|\right) 1\{|y_{t-1}| \le |u_t|\} \\ &\le \frac{1}{n}\sum_{t=2}^{n} 2\,|u_t\zeta_t|\, 1\{|y_{t-1}| \le |u_t|\}, \end{aligned} \tag{17}$$

which is $o_p(1)$ by Lemma 7. ∎

A.3 Proof of Theorem 3

We introduce some notations to ease our exposition. Let $\bar{u}_t = u_t - \frac{1}{n}\sum_{t=p+2}^{n} u_t$, and

$$\begin{aligned}
x_{1t} &= \begin{pmatrix} y_{t-1}1\{y_{t-1}\le\gamma_1\} - \frac{1}{n}\sum_{t=p+2}^{n} y_{t-1}1\{y_{t-1}\le\gamma_1\} \\ y_{t-1}1\{y_{t-1}>\gamma_2\} - \frac{1}{n}\sum_{t=p+2}^{n} y_{t-1}1\{y_{t-1}>\gamma_2\} \end{pmatrix}, \\
x_{2t} &= \left(u_{t-1} - \frac{1}{n}\sum_{t=p+1}^{n-1} u_t, \cdots, u_{t-p} - \frac{1}{n}\sum_{t=2}^{n-p} u_t\right)', \\
x_t &= \left(x_{1t}', x_{2t}'\right)'.
\end{aligned}$$

Here, the dependence of x_{1t} and x_t on γ is suppressed. We first derive limit distributions of $\hat{\alpha}_i's$. Under the null,

$$\begin{aligned}
\begin{pmatrix} n\hat{\alpha}_1(\gamma) \\ n\hat{\alpha}_2(\gamma) \end{pmatrix} &= \left(\frac{1}{n^2}\sum_{t=p+2}^{n} x_{1t}x_{1t}' - \frac{1}{n}\cdot\frac{1}{n}\sum_{t=p+2}^{n} x_{1t}x_{2t}'\left(\frac{1}{n}\sum_{t=p+2}^{n} x_{2t}x_{2t}'\right)^{-1}\frac{1}{n}\sum_{t=p+2}^{n} x_{2t}x_{1t}'\right)^{-1} \\
&\times\left(\frac{1}{n}\sum_{t=p+2}^{n} x_{1t}\bar{u}_t - \frac{1}{n}\sum_{t=p+2}^{n} x_{1t}x_{2t}'\left(\frac{1}{n}\sum_{t=p+2}^{n} x_{2t}x_{2t}'\right)^{-1}\frac{1}{n}\sum_{t=p+2}^{n} x_{2t}\bar{u}_t\right)
\end{aligned}$$

Similarly in (16), we have,

$$\sup_{\theta}\frac{1}{n}\sum_{t=p+2}^{n}\left|y_{t-1}1\{y_{t-1}\le\theta\} - y_{t-1}1\{y_{t-1}\le 0\}\right| = o_p(1),$$

where the supremum is taken over a compact set, and, by the continuous mapping theorem,

$$\frac{1}{n\sqrt{n}}\sum_{t=p+2}^{n} y_{t-1}1\{y_{t-1}\le 0\} \Rightarrow \int_0^1 B1\{B\le 0\}.$$

Therefore,

$$\frac{1}{n^2}\sum_{t=p+2}^{n} x_{1t}x_{1t}' \Rightarrow \begin{pmatrix} \int_0^1 \bar{B}_L^2 & -\int_0^1 \bar{B}_L\bar{B}_U \\ -\int_0^1 \bar{B}_L\bar{B}_U & \int_0^1 \bar{B}_U^2 \end{pmatrix}, \tag{18}$$

where $\bar{B}_L = B1\{B\le 0\} - \int_0^1 B1\{B\le 0\}$, $\bar{B}_U = B1\{B>0\} - \int_0^1 B1\{B>0\}$. Furthermore, by Theorem 2,

$$
\begin{aligned}
& \frac{1}{n}\sum_{t=p+2}^{n} x_{1t}\bar{u}_t \\
= & \begin{pmatrix} \frac{1}{n}\sum_{t=p+2}^{n} y_{t-1}1\{y_{t-1}\le\gamma_1\}u_t \\ \frac{1}{n}\sum_{t=p+2}^{n} y_{t-1}1\{y_{t-1}>\gamma_2\}u_t \end{pmatrix} - \begin{pmatrix} \frac{1}{n\sqrt{n}}\sum_{t=p+2}^{n} y_{t-1}1\{y_{t-1}\le\gamma_1\}\frac{1}{\sqrt{n}}\sum_{t=p+2}^{n} u_t \\ \frac{1}{n\sqrt{n}}\sum_{t=p+2}^{n} y_{t-1}1\{y_{t-1}>\gamma_2\}\frac{1}{\sqrt{n}}\sum_{t=p+2}^{n} u_t \end{pmatrix} \\
\Rightarrow & \begin{pmatrix} \int_0^1 B1\{B\le 0\}dB + \lambda\int_0^1 1\{B\le 0\} - B(1)\int_0^1 B1\{B\le 0\} \\ \int_0^1 B1\{B>0\}dB + \lambda\int_0^1 1\{B>0\} - B(1)\int_0^1 B1\{B>0\} \end{pmatrix} \\
= & \begin{pmatrix} \int_0^1 \bar{B}_L dB + \lambda\int_0^1 1\{B\le 0\} \\ \int_0^1 \bar{B}_U dB + \lambda\int_0^1 1\{B>0\} \end{pmatrix}, \qquad (19)
\end{aligned}
$$

on $\mathcal{S}$.

Next, we show that

$$
\frac{1}{n}\sum_{t=p+2}^{n} y_{t-1}1\{y_{t-1}\le\gamma_1\}u_{t-p} \Rightarrow \int_0^1 B1\{B\le 0\}dB + (\lambda+\bar{r}_p)\int_0^1 1\{B\le 0\}, \qquad (20)
$$

where $\bar{r}_p = r(0) + r(1)\cdots + r(p-1)$. Since

$$
\begin{aligned}
y_{t-1}1\{y_{t-1}\le\gamma_1\}u_{t-p} &= y_{t-p-1}1\{y_{t-p-1}\le\gamma_1\}u_{t-p} \\
& \quad + (y_{t-1}1\{y_{t-1}\le\gamma_1\} - y_{t-p-1}1\{y_{t-p-1}\le\gamma_1\})u_{t-p}
\end{aligned}
$$

and

$$
\frac{1}{n}\sum_{t=p+2}^{n} y_{t-p-1}1\{y_{t-p-1}\le\gamma_1\}u_{t-p} \Rightarrow \int_0^1 B1\{B\le 0\}dB + \lambda\int_0^1 1\{B\le 0\}
$$

from Theorem 2, it remains to show that

$$
\frac{1}{n}\sum_{t=p+2}^{n} (y_{t-1}1\{y_{t-1}\le\gamma_1\} - y_{t-p-1}1\{y_{t-p-1}\le\gamma_1\})u_{t-p} \Rightarrow \bar{r}_p\int_0^1 1\{B\le 0\}.
$$

To do so, note that

$$
\begin{aligned}
& y_{t-1}1\{y_{t-1}\le\gamma_1\} - y_{t-p-1}1\{y_{t-p-1}\le\gamma_1\} \\
= & u_{tp}1\{y_{t-1}\le\gamma_1\} + y_{t-p-1}(1\{\gamma_1 < y_{t-p-1}\le\gamma_1 - u_{tp}\} - 1\{\gamma_1 - u_{tp} < y_{t-p-1}\le\gamma_1\}),
\end{aligned}
$$

where $u_{tp} = u_{t-1} + \cdots + u_{t-p} = y_{t-1} - y_{t-p-1}$. Next, by Lemma 7, we have

$$
\begin{aligned}
& \frac{1}{n}\sum_{t=p+2}^{n} |y_{t-p-1}(1\{\gamma_1 < y_{t-p-1}\le\gamma_1 - u_{tp}\} - 1\{\gamma_1 - u_{tp} < y_{t-p-1}\le\gamma_1\})u_{t-p}| \\
\le & \frac{1}{n}\sum_{t=p+2}^{n} (|\bar{\gamma}| + |u_{tp}|)|u_{t-p}|1\{|y_{t-p-1}|\le|\bar{\gamma}| + |u_{tp}|\} \to_p 0, \qquad (21)
\end{aligned}
$$

and by Theorem 1,

$$\frac{1}{n}\sum_{t=p+2}^{n} u_{tp}u_{t-p}1\left\{y_{t-1}\leq\gamma_1\right\}\Rightarrow\bar{r}_p\int_0^1 1\left\{B\leq 0\right\}. \tag{22}$$

Then Theorem 2, (21), and (22) establish the convergence in (20), which in turn yields

$$\frac{1}{n}\sum_{t=p+2}^{n} x_{1t}x_{2t}'\Rightarrow\begin{pmatrix}\int_0^1\bar{B}_L dB+(\lambda+\bar{r}_1)\int_0^1 1\left\{B\leq 0\right\}, & \cdots & \int_0^1\bar{B}_L dB+(\lambda+\bar{r}_p)\int_0^1 1\left\{B\leq 0\right\}\\ \int_0^1\bar{B}_U dB+(\lambda+\bar{r}_1)\int_0^1 1\left\{B> 0\right\}, & \cdots & \int_0^1\bar{B}_U dB+(\lambda+\bar{r}_p)\int_0^1 1\left\{B> 0\right\}\end{pmatrix},$$

on $\mathcal{S}$. Finally, it follows from the law of large numbers that

$$\frac{1}{n}\sum_{t=p+2}^{n} x_{2t}x_{2t}'\Rightarrow G_p \text{ and } \frac{1}{n}\sum_{t=p+2}^{n} x_{2t}\bar{u}_t\Rightarrow g_p, \tag{23}$$

which completes the proof of part (i).

■

A.3.1 Limit distribution of W_n

We first derive the limit of $\hat{\sigma}^2(\gamma)$. Define κ_n as a $p+2$ dimensional diagnal matrix whose first two elements are n^{-1} and the others are $n^{-1/2}$. Then, the convergences $(18),(19),(20)$ and (23) yield

$$\begin{aligned}\hat{\sigma}^2(\gamma) &= \frac{1}{n}\sum_{t=p+2}^{n}\bar{u}_t^2-\left(n^{-1/2}\kappa_n\sum_{t=p+2}^{n}x_t\bar{u}_t\right)'\left(\kappa_n\sum_{t=p+2}^{n}x_t x_t'\kappa_n\right)^{-1}\left(n^{-1/2}\kappa_n\sum_{t=p+2}^{n}x_t\bar{u}_t\right)\\ &\Rightarrow \sigma^2-g_p'G_p^{-1}g_p. \end{aligned}\tag{24}$$

Finally, from (18), (**??**) and (24), we conclude (ii).

■

A.4 Proof of Theorem 4

Let $M_r=[([lr]-2)/b]$ and $B=\min\{b,[lr]-mb-1\}$. Then, $\frac{1}{\sqrt{l}}\sum_{t=1}^{[lr]}u_t^*$ can be written as $\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{j=1}^{B}\tilde{u}_{i_m+j}$, and it is sufficient to consider $\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{j=1}^{b}\tilde{u}_{i_m+j}$, as demonstrated in Theorem 3.1 (see (8.1) and the following sentence) of Paparoditis and Politis (2003) (hereafter PP).

Since $\frac{1}{n-b}\sum_{i=1}^{n-b}\sum_{j=1}^{b} u_{i+j} = \sum_{j=1}^{b} E^* u_{i_m+j}$, we can write

$$\begin{aligned}
\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{j=1}^{b}\tilde{u}_{i_m+j} &= \frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{j=1}^{b}\left(\hat{u}_{i_m+j} - \frac{1}{n-b}\sum_{i=1}^{n-b}\frac{1}{b}\sum_{s=1}^{b}\hat{u}_{i+s}\right) \\
&= \frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\left(\sum_{j=1}^{b} u_{i_m+j} - \sum_{j=1}^{b} E^* u_{i_m+j}\right) \\
&\quad -\hat{\alpha}_1\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\left(\sum_{j=1}^{b} y_{i_m+j-1}1\{y_{i_m+j-1}\le\gamma_1\} - \sum_{j=1}^{b} E^* y_{i_m+j}1\{y_{i_m+j}\le\gamma_1\}\right) \\
&\quad -\hat{\alpha}_2\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\left(\sum_{j=1}^{b} y_{i_m+j-1}1\{y_{i_m+j-1}>\gamma_2\} - \sum_{j=1}^{b} E^* y_{i_m+j}1\{y_{i_m+j}>\gamma_2\}\right),
\end{aligned}$$

the last three terms of which are denoted as $I_1, I_2,$ and I_3. Note that

$$\begin{aligned}
E^*\left[\sum_{j=1}^{b} y_{i_m+j-1}1\{y_{i_m+j-1}\le\gamma_1\}\right] &= \frac{1}{n-b}\sum_{t=1}^{n-b}\sum_{j=1}^{b} y_{t+j-1}1\{y_{t+j-1}\le\gamma_1\} \\
&\le b\sup_{t\le n}|y_t| = O_p\left(b\sqrt{n}\right), \qquad (25)
\end{aligned}$$

and

$$\begin{aligned}
E^*\left[\sum_{j=1}^{b} y_{i_m+j-1}1\{y_{i_m+j-1}\le\gamma_1\}\right]^2 &= \frac{1}{n-b}\sum_{t=1}^{n-b}\left[\sum_{j=1}^{b} y_{t+j-1}1\{y_{t+j-1}\le\gamma_1\}\right]^2 \\
&\le \left[b\sup_{t\le n}|y_t|\right]^2 = O_p\left(b^2 n\right), \qquad (26)
\end{aligned}$$

which in turn establish

$$\begin{aligned}
& E^*\left(\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{j=1}^{b}\left(y_{i_m+j-1}1\{y_{i_m+j-1}\le\gamma_1\} - \sum_{j=1}^{b} E^* y_{i_m+j}1\{y_{i_m+j}\le\gamma_1\}\right)\right)^2 \\
= \;& O_p(bn),
\end{aligned}$$

uniformly in $\gamma_1 \in \mathcal{S}$ and $r \in [0,1]$. Therefore, $E^* I_2^2 = O_p\left(bn^{-1}\right)$, and, similarly,$E^* I_3^2 = O_p\left(bn^{-1}\right)$. (This also proves Lemma 8 below).

Then, the convergence of $\frac{1}{\sqrt{l}}\sum_{t=1}^{[lr]} u_t^*$ is completely determined by I_1. However, I_1 is based on the resampling of $u_t's$ and its convergence to the Brownian motion is already developed in Theorem 3.1 of PP, and the convergence of ω^* is provided in the following lemma. ■

Similarly as in the above proof, the following lemma is straightforward from Lemma 8.1 of PP.

Lemma 8 *Under the assumptions of Theorem 4 and as $n \to \infty$, we have*

(i) $l^{-1}\sum_{j=1}^{l} u_j^* \to^p 0$,
(ii) ${\omega^*}^2 = var^*\left(l^{-1/2}\sum_{j=2}^{l} u_j^*\right) \to^p \omega^2$,
(iii) ${\sigma^*}^2 = l^{-1}\sum_{j=1}^{l} {u_j^*}^2 \to^p \sigma^2$.

A.5 Proof of Theorem 5

For part (i), note that, for any $\theta \in \Theta$, $\bar{\theta} = \max\{|\theta|; \theta \in \Theta\}$,

$$\begin{aligned} & l^{-2}\left|\sum_{t=2}^{l} y_{t-1}^{*^2} 1\left\{y_{t-1}^* \le \theta\right\} - \sum_{t=2}^{l} y_{t-1}^{*^2} 1\left\{y_{t-1}^* \le 0\right\}\right| \\ \le \;& l^{-2}\left|\sum_{t=2}^{l} y_{t-1}^{*^2} 1\left\{\left|y_{t-1}^*\right| \le \bar{\theta}\right\}\right| \le l^{-1}\bar{\theta} \to 0. \end{aligned}$$

Then, it follows from the continuous mapping theorem and Theorem 4 that

$$l^{-2}\sum_{t=2}^{l} y_{t-1}^{*^2} 1\left\{y_{t-1}^* \le 0\right\} \Rightarrow \int_0^1 B^2 1\{B \le 0\}.$$

For part (ii), without loss of generality, assume $u_t^* = 0$ if $t > l$. Let $u_t^* = \varepsilon_t^* - \left(\zeta_t^* - \zeta_{t-1}^*\right)$ where $\varepsilon_t^* = \sum_{j=0}^{\infty}\left(\mathrm{E}_t^* u_{t+j}^* - \mathrm{E}_{t-1}^* u_{t+j}^*\right)$ and $\zeta_t^* = \sum_{j=1}^{\infty}\mathrm{E}_t^* u_{t+j}^*$. Recalling that $u_t^* = \tilde{u}_{i_m+s}$ where $m = [(t-2)/b]$ and $s = t - mb - 1$, we have

$$E_t^* u_{t+j}^* = \begin{cases} u_{t+j}^*, & \text{if } j \le b-s \\ E^* u_{t+j}^*, & \text{o/w}, \end{cases}$$

and thus,

$$\begin{aligned} \varepsilon_t^* &= \begin{cases} \sum_{j=0}^{b-1} u_{t+j}^*, & \text{if } s = 1, \\ 0, & \text{if } s > 1, \end{cases} \\ \zeta_t^* &= \sum_{j=1}^{b-s} u_{t+j}^* + \sum_{j=b-s+1}^{l-t} E^* u_{t+j}^* = \sum_{j=1}^{b-s} u_{t+j}^*, \end{aligned}$$

since $\sum_{j=1}^{b} E^* u_{t+j}^* = 0$ for any $t \ge 2$.

Write

$$\begin{aligned} \frac{1}{l}\sum_{t=2}^{l} y_{t-1}^* 1\left\{y_{t-1}^* \le \theta\right\} u_t^* &= \frac{1}{l}\sum_{t=2}^{l} y_{t-1}^* 1\left\{y_{t-1}^* \le \theta\right\}\varepsilon_t^* + L_l^* + R_l^*, \qquad (27) \\ \text{where} \qquad L_l^* &= \frac{1}{l}\sum_{t=2}^{l} u_t^*\zeta_t^* \cdot 1\left\{y_{t-1}^* \le \theta\right\} 1\left\{y_t^* \le \theta\right\}, \\ R_l^* &= \frac{1}{l}\sum_{t=2}^{l}\left(y_t^* 1\left\{y_t^* \le \theta < y_{t-1}^*\right\} + y_{t-1}^* 1\left\{y_{t-1}^* \le \theta < y_t^*\right\}\right)\zeta_t^*. \end{aligned}$$

To show that $R_l^* = o_p(1)$ uniformly in $\theta \in \Theta$, write that

$$
\begin{aligned}
u_t^* \zeta_t^* &= \tilde{u}_{i_m+s} \sum_{j=s+1}^{b} \tilde{u}_{i_m+j} \\
&= \sum_{j=s+1}^{b} \left\{ \left(u_{i_m+s} - \frac{1}{n-b} \sum_{g=1}^{n-b} \frac{1}{b} \sum_{v=1}^{b} u_{v+g} \right) \left(u_{i_m+j} - \frac{1}{n-b} \sum_{g=1}^{n-b} \frac{1}{b} \sum_{v=1}^{b} u_{v+g} \right) \right\} \qquad (28) \\
&\quad + \hat{\alpha}_1^2 \sum_{j=s+1}^{b} \left\{ \begin{array}{c} \left(y_{i_m+s-1} 1\left\{ y_{i_m+s-1} \le \gamma_1 \right\} - \frac{1}{n-b} \sum_{g=1}^{n-b} \frac{1}{b} \sum_{v=1}^{b} y_{v+g} 1\left\{ y_{v+g} \le \gamma_1 \right\} \right) \\ \times \left(y_{i_m+j-1} 1\left\{ y_{i_m+j-1} \le \gamma_1 \right\} - \frac{1}{n-b} \sum_{g=1}^{n-b} \frac{1}{b} \sum_{v=1}^{b} y_{v+g} 1\left\{ y_{v+g} \le \gamma_1 \right\} \right) \end{array} \right\} \qquad (29) \\
&\quad + \hat{\alpha}_2^2 \sum_{j=s+1}^{b} \left\{ \begin{array}{c} \left(y_{i_m+s-1} 1\left\{ y_{i_m+s-1} > \gamma_2 \right\} - \frac{1}{n-b} \sum_{g=1}^{n-b} \frac{1}{b} \sum_{v=1}^{b} y_{v+g} 1\left\{ y_{v+g} > \gamma_2 \right\} \right) \\ \times \left(y_{i_m+j-1} 1\left\{ y_{i_m+j-1} > \gamma_2 \right\} - \frac{1}{n-b} \sum_{g=1}^{n-b} \frac{1}{b} \sum_{v=1}^{b} y_{v+g} 1\left\{ y_{v+g} > \gamma_2 \right\} \right) \end{array} \right\} \qquad (30)
\end{aligned}
$$

Note that (29) and (30) are $o_p(1)$ uniformly in t and γ as in the Proof of Theorem 4. Also note that $\frac{1}{n-b} \sum_{g=1}^{n-b} \frac{1}{b} \sum_{v=1}^{b} u_{v+g} = O_p\left(\frac{1}{\sqrt{n}} \right)$ (see Künsch (1989), p. 1227). Then, like (17), we have

$$
\begin{aligned}
\left| R_l^* \right| &\le \frac{1}{l} \sum_{t=2}^{l} \left(\left| \bar{\theta} \right| + 2 \left| u_t^* \zeta_t^* \right| \right) 1\left\{ \left| y_{t-1}^* \right| \le \bar{\theta} + \left| u_t^* \right| \right\} \\
&\le \frac{1}{l} \sum_{t=2}^{l} \left(\begin{array}{c} \left| \bar{\theta} \right| + 2 \left| \sum_{j=s+1}^{b} u_{i_m+s} u_{i_m+j} \right| + 2(b-s) \left| u_{i_m+s} \right| O_p\left(\frac{1}{\sqrt{n}} \right) \\ +2 \left| \sum_{j=s+1}^{b} u_{i_m+j} \right| O_p\left(\frac{1}{\sqrt{n}} \right) + o_p(1) \end{array} \right) \qquad (31) \\
&\quad \times 1\left\{ \left| y_{t-1}^* \right| \le \bar{\theta} + \left| u_t^* \right| \right\}.
\end{aligned}
$$

Since $\{u_t\}$ is independent of $\{i_m\}$, $\sum_{j=s+1}^{b} u_{i_m+s} u_{i_m+j}$ is uniformly integrable by Theorem 3.2 of Hansen (1992), not to mention $\sum_{j=s+1}^{b} u_{i_m+j}$ and u_{i_m+s}. And, write, for any $M_1 > 0$,

$$
\frac{1}{l} \sum_{t=2}^{l} E1\left\{ \left| y_{t-1}^* \right| \le \bar{\theta} + \left| u_t^* \right| \right\} \le \frac{1}{l} \sum_{t=2}^{l} E1\left\{ \left| y_{t-1}^* \right| \le \bar{\theta} + M_1 \right\} + \frac{1}{l} \sum_{t=2}^{l} E1\left\{ \left| u_t^* \right| > M_1 \right\}. \qquad (32)
$$

Then it follows from Theorem 4 and (15) in the proof of Lemma 7 that the first term in the right hand side of (32) is $o(1)$. Next, since

$$
\sup_x \left| P^*\left\{ u_t^* \le x \right\} - P^*\left\{ u_j^* \le x \right\} \right| \le 2b/n
$$

for any j and t, we observe that

$$
\frac{1}{l} \sum_{t=2}^{l} E^* 1\left\{ \left| u_t^* \right| > M_1 \right\} \le \sup_t E^* 1\left\{ \left| u_t^* \right| > M_1 \right\} \le P^*\left\{ u_2^* > M_1 \right\} + 2b/n.
$$

Furthermore, since $\{u_t\}$ is strictly stationary and independent of $\{i_m\}$, and $u_2^* = u_{i_1} + o_p(1)$, for any $\varepsilon > 0$, there is M_1 satisfying

$$
E\left[P^*\left\{ u_2^* > M_1 \right\} \right] \le P\left\{ u_{i_1} > M_1 - \varepsilon/2 \right\} + \varepsilon/2 < \varepsilon,
$$

which in turn yield that the second term in the right hand side of (32) is also $o(1)$. Finally, we conclude that $|R_l^*| = o_p(1)$ by the uniform integrability and the observation that (32) is negligible (as in the proof of Lemma 7).

Next, we have

$$E^* u_t^* \zeta_t^* = \sum_{j=1}^{b-s} E^* \tilde{u}_{i_m+s} \tilde{u}_{i_m+s+j} \to^p \lambda,$$

by a similar argument in Lemma 8. Then, by the uniform integrability above, Theorem 4, and the same argument as in the proof of Theorem 1 we can conclude that the limit of L_l^* is $\lambda \int_0^1 1\{B \leq 0\}$.

Finally, for the convergence of the first term in (27), write that

$$\frac{1}{l} \sum_{t=2}^{l} y_{t-1}^* 1\{y_{t-1}^* \leq \theta\} \varepsilon_t^* = \frac{1}{k} \sum_{m=1}^{k-1} Y_{m-1}^* 1\left\{Y_{m-1}^* \leq \theta/\sqrt{b}\right\} V_m^*,$$

where $V_m^* = \frac{1}{\sqrt{b}} \sum_{j=1}^{b} u_{mb+1+j}^* = \frac{1}{\sqrt{b}} \varepsilon_{mb+2}^*$ and $Y_m^* = \sum_{s=0}^{m} V_s^* = \frac{1}{\sqrt{b}} y_{mb+1}^*$. Note that $\{V_m^*\}$ is an independent identically distributed sequence under the bootstrap distribution, since $\{V_m^*\}$ is a normalized sum of each block that is resampled independently. And its mean is zero and its variance is $O_p(1)$ as shown in Lemma 8. Furthermore,

$$\begin{aligned} & E^* \sup_{0 \leq \gamma \leq \bar{\gamma}} \left| \frac{1}{k} \sum_{m=1}^{k-1} Y_{m-1}^* 1\left\{Y_{m-1}^* \leq \theta/\sqrt{b}\right\} V_m^* - \frac{1}{k} \sum_{m=1}^{k-1} Y_{m-1}^* 1\left\{Y_{m-1}^* \leq 0\right\} V_m^* \right| \\ \leq \;& \frac{\bar{\gamma}}{\sqrt{b}} \frac{1}{k} \sum_{m=1}^{k-1} Pr\left\{0 < Y_{m-1}^* \leq \bar{\theta}/\sqrt{b}\right\} E^* |V_m^*| = o_p(1), \end{aligned}$$

and the transformation $s1\{s \leq 0\}$ is continuous. Therefore, the convergence follows from the invariance principle in Theorem 4, the continuous mapping theorem, and the convergence to stochastic integral of Kurtz and Protter (1991). ■

A.6 Proof of Theorem 6

Since we already have the invariance principle for our RBB and the bootstrap version of Theorem 2 in hand, the proof of this Theorem is straightforward following the same line of argument of the proof of Theorem 3. ■

References

Ait-Sahalia, Y. (1996): "Testing continuous-time models of the spot interest rate," *Review of Financial Studies*, 9(2), 385–426.

Balke, N., and T. Fomby (1997): "Threshold cointegration," *International Economic Review*, 38, 627–645.

Bec, F., A. Guay, and E. Guerre (2002): "Adaptive Consistent Unit Root Tests Based on Autoregressive Threshold Model," Working Paper.

Caner, M., and B. E. Hansen (2001): "Threshold autoregression with a unit root," *Econometrica*, 69(6), 1555–1596.

Chan, K. S. (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, 21, 520–533.

Chan, K. S., J. D. Petruccelli, H. Tong, and S. W. Woolford (1985): "A Multiple-threshold AR(1) Model," *Journal of Applied Probability*, 22, 267–279.

Chan, K. S., and H. Tong (1990): "On Likelihood Ratio Tests for Threshold Autoregression," *Journal of the Royal Statistical Society, Series B, Methodological*, 52, 469–476.

Chan, K. S., and R. S. Tsay (1998): "Limiting properties of the least squares estimator of a continuous threshold autoregressive model," *Biometrika*, 85(2), 413–426.

Chang, Y., and J. Y. Park (2003): "A Sieve Bootstrap for the Test of a Unit Root," *Journal of Time Series Analysis*, 24(4), 379–400.

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford University Press.

Davies, R. (1987): "Hypothesis Testing when a Nuisance parameter is Present Only under the Alternative," *Biometrika*, 74, 33–43.

De Jong, R. (2003): "Nonlinear estimators with integrated regressors but without exogeneity," manuscript.

Hall, P., J. L. Horowitz, and B.-Y. Jing (1995): "On blocking rules for the bootstrap with dependent data," *Biometrika*, 82, 561–574.

Hansen, B. (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500.

——— (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64, 413–430.

——— (1997): "Approximate asymptotic p-values for structural change tests," *Journal of Business and Economic Statistics*, 15, 60–67.

Horvath, M., and M. Watson (1995): "Testing For Cointegration When Some of the Cointegrating Vectors are Prespecified," *Econometric Theory*, 11(5), 984–1014.

Kapetanios, G., and Y. Shin (2003): "Unit Root Tests in Three-Regime SETAR Models," Queen Mary, University of London.

Künsch, H. R. (1989): "The jackknife and the bootstrap for general stationary observations," *The Annals of Statistics*, 17, 1217–1241.

Kurtz, T., and P. Protter (1991): "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations," *Annals of Probability*, 19, 1035–1070.

Lo, M., and E. Zivot (2001): "Threshold Cointegration and Nonlinear Adjustment to the Law of One Price," *Macroeconomic Dynamics*, 5(4), 533–576.

Paparoditis, E., and D. N. Politis (2003): "Residual-based block bootstrap for unit root testing," *Econometrica*, 71(3), 813–855.

Park, J., and P. Phillips (2001): "Nonlinear Regressions with Integrated Time Series," *Econometrica*, 69, 117–161.

Park, J. Y. (2002): "An invariance principle for sieve bootstrap in time series," *Econometric Theory*, 18(2), 469–490.

Pippenger, M., and G. Goering (1993): "A note on the empirical power of unit root tests under threshold processes," *Oxford Bulletin of Economics and Statistics*, 55, 473–481.

Said, S. E., and D. A. Dickey (1984): "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, 71, 599–607.

Abstract

There is a growing literature on unit root testing in threshold autoregressive models. This paper makes two contributions to the literature. First, an asymptotic theory is developed for unit root testing in a threshold autoregression, in which the errors are allowed to be dependent and heterogeneous, and the lagged level of the dependent variable is employed as the threshold variable. The asymptotic distribution of the proposed Wald test is non-standard and depends on nuisance parameters. Second, the consistency of the proposed residual-based block bootstrap is established based on a newly developed asymptotic theory for this bootstrap. It is demonstrated by a set of Monte Carlo simulations that the Wald test exhibits considerable power gains over the ADF test that neglects threshold effects. The law of one price hypothesis is investigated among used car markets in the US.

Keywords: Threshold autoregression; unit root test; threshold cointegration; residual-based block bootstrap.

JEL No.: C12, C15, C22.

Contact details:
Myunghwan Seo
Department of Economics
London School of Economics
email: m.seo@lse.ac.uk