Econometric Theory, **24**, 2008, 1699–1716. Printed in the United States of America.
doi:10.1017/S0266466608080663

UNIT ROOT TEST IN A THRESHOLD AUTOREGRESSION: ASYMPTOTIC THEORY AND RESIDUAL-BASED BLOCK BOOTSTRAP

MYUNG HWAN SEO
London School of Economics

This paper develops a test of the unit root null hypothesis against a stationary threshold process. This testing problem is nonstandard and complicated because a parameter is unidentified and the process is nonstationary under the null hypothesis. We derive an asymptotic distribution for the test, which is not pivotal without simplifying assumptions. A residual-based block bootstrap is proposed to calculate the asymptotic p-values. The asymptotic validity of the bootstrap is established, and a set of Monte Carlo simulations demonstrates its finite-sample performance. In particular, the test exhibits considerable power gains over the augmented Dickey–Fuller (ADF) test, which neglects threshold effects.

1. INTRODUCTION

A time series with nonlinearity or breaks often generates a sample path, which is similar to that of a unit root process in a finite sample. A large literature has grown regarding testing the unit root null hypothesis against a stationary time series that exhibits nonlinearity or breaks. For example, Perron (1989) and Balke and Fomby (1997) considered break models and threshold models, respectively. In particular, the latter work has brought broad applied attention to this testing problem by introducing threshold cointegration, and Lo and Zivot (2001) and Bec and Rahbek (2004) review various applications. Although they concern a self-exciting threshold autoregression (SETAR) of Tong (1990), Enders and Granger (1998) proposed a momentum threshold autoregression (M-TAR) model whose thresholding is based on a difference of the series.

Although standard unit root tests such as the augmented Dickey–Fuller (ADF) test can be applied when the true process is a threshold-type model, the power of such tests can suffer significantly. A series of work has been done to develop

I am indebted to Bruce Hansen for his invaluable guidance and time. I thank Peter Robinson, Oliver Linton, David Hendry, and Howell Tong for helpful comments. I also thank the two referees and the co-editor. Address correspondence to Myung Hwan Seo, Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom; e-mail: m.seo@lse.ac.uk.

 0266-4666/08 $15.00

appropriate tests and econometric theory. Caner and Hansen (2001) developed an asymptotic theory for M-TAR and proposed a bootstrap. A cointegration test for the threshold cointegration model was developed by Seo (2006). This paper considers SETAR. Although all these models and tests appear similar, each demands distinct distribution theory depending on how nonlinearity and nonstationarity are combined. For instance, unlike M-TAR SETAR involves a nonlinear transformation of a nonstationary variable.

This paper develops a unit root test in a SETAR model:

$$\Delta y_t = \begin{cases} \alpha_1 y_{t-1} + u_t, & \text{if } y_{t-1} \le \gamma, \\ \alpha_2 y_{t-1} + u_t, & \text{if } y_{t-1} > \gamma, \end{cases} \tag{1}$$

where $t = 2, \ldots, n$. Hansen (1999) reviews its application to various economic time series including gross national product (GNP), industrial production, unemployment rate, stock volatilities, etc. Threshold cointegration entails SETAR to the error-correction term, and thus the unit root test developed in this paper can be applied. Discriminating nonstationarity from nonlinearity is important in this framework. Indeed, a series of papers has emerged to test the unit root hypothesis, $\alpha_1 = \alpha_2 = 0$, in the model (1). The case with fixed γ was studied by Enders and Granger (1998). The case with free γ has been considered by Kapetanios and Shin (2006), Bec, Guay, and Guerre (2008), and Park and Shintani (2005). These authors differ on the treatment of the parameter space for γ. Kapetanios and Shin (2006) consider a compact parameter space for γ, Bec et al. (2008) a parameter space that expands at the $\sqrt{n}$ rate, and Park and Shintani (2005) a random parameter space. Our development is based on the compact parameter space, which we demonstrate has some power advantage.

This paper differs from the aforementioned papers on the following issues. We allow for serial correlation and nonlinearity in the error u_t and develop a bootstrap theory. The presence of serial correlation is standard in the unit root testing, and that of nonlinearity is natural when the alternative model is nonlinear. We show that this generality introduces bias terms in a complicated way. Ng and Perron (1995) demonstrate that direct estimation of the bias exhibits poor finite-sample performance. The approach by Said and Dickey (1984) is not appropriate because of nonlinearity. Instead, we consider a residual-based block bootstrap (RBB) in the spirit of Paparoditis and Politis (2003) to allow for general dependence in the error u_t.

We establish the consistency of RBB, which generalizes that of the standard unit root tests in Paparoditis and Politis (2003). This is nontrivial because the testing problem is nonstandard as a result of the presence of an unidentified parameter under the null as in Davies (1987) and because the null model is an integrated process. The nonlinear transformation of an integrated process complicates the asymptotic and bootstrap theory in different manners because of the blocking resampling scheme. Although Caner and Hansen (2001) proposed a residual-based bootstrap for a unit root test in M-TAR, they did not provide theoretical justification.

The finite-sample performance of the proposed bootstrap test is examined through Monte Carlo simulations. The ADF test has some power over the threshold alternative and has been used in practice. We find that our bootstrap test exhibits reasonable finite-sample size property and substantial power gain over the ADF test, whereas a higher order property of the bootstrap is not studied. We also consider the random parameter space setup and find that the bootstrap associated with such construction does not improve upon the ADF test. These findings do not depend upon the block length selection.

The remainder of this paper is organized as follows. Section 2 introduces the model and the test statistic and develops the asymptotic theory for the test. The RBB is introduced and its asymptotic validity is established in Section 3. Section 4 presents simulation evidence for finite-sample performance of the bootstrap. Section 5 concludes. All proofs are collected in the Appendix.

2. UNIT ROOT TESTING IN SETAR

Rewrite the model (1) as follows:

$$\Delta y_t = \alpha_1 y_{t-1} 1\{y_{t-1} \leq \gamma\} + \alpha_2 y_{t-1} 1\{y_{t-1} > \gamma\} + u_t, \tag{2}$$

where $1\{\cdot\}$ is the indicator function and γ belongs to a compact set $\Gamma \subset \mathbf{R}$. The error process $\{u_t\}$ can be a serially dependent nonlinear process. The result in this paper naturally extends to three-regime SETARs, which are also employed widely in practice (see, e.g., Balke and Fomby, 1997).

This paper considers testing the unit root hypothesis

$$H_0 : \alpha_1 = \alpha_2 = 0 \tag{3}$$

against the alternative of a stationary SETAR process. Unfortunately, our understanding is not complete as to the stationarity conditions for general SETAR processes. When the errors are independent, Chan, Petruccelli, Tong, and Woolford (1985) showed that a necessary and sufficient condition of stationarity is $\alpha_1 < 0$, $\alpha_2 < 0$, and $(\alpha_1 + 1)(\alpha_2 + 1) < 1$, which suggests that the natural alternative to H_0 should be

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

As in the standard unit root testing, the least squares (LS) estimator of α_1 and α_2 is consistent under the null (3) despite the serial correlation in u_t. Different auxiliary regressions can be considered to get different but consistent estimators of α_1 and α_2 and therefore different statistics to test the null. Phillips and Perron (1988) considered a bias-corrected t-statistic based on the regression of Δy_t on the constant and y_{t-1}, and Said and Dickey (1984) proposed the ADF test based on the regression of Δy_t on the constant, y_{t-1}, and the lagged

Δy_t's. Although the Phillips–Perron test can be applied to more general error processes, it is often very difficult to estimate the bias term precisely for the sample size encountered in application. Although the ADF test often performs better than the Phillips–Perron test, it is justified only for the autoregressive moving average (ARMA) process. If $\{u_t\}$ is a nonlinear process, however, the ADF test is not bias-free but may reduce the bias as in prewhitening. We employ an ADF-type test statistic and approximate its sampling distribution by a RBB, which allows for nonlinear processes.

Specifically, we estimate the following auxiliary regression:

$$\Delta y_t = \hat{\alpha}_1(\gamma) y_{t-1} 1\{y_{t-1} \leq \gamma\} + \hat{\alpha}_2(\gamma) y_{t-1} 1\{y_{t-1} > \gamma\} + \hat{\mu}(\gamma) + \hat{\rho}_1(\gamma)\Delta y_{t-1} + \cdots + \hat{\rho}_p(\gamma)\Delta y_{t-p} + \hat{e}_t(\gamma), \tag{5}$$

where $\hat{\theta}(\gamma)$ and $\hat{e}(\gamma)$ are the LS estimate of a parameter $\theta = (\alpha_1, \alpha_2, \mu, \rho_1, \ldots, \rho_p)'$ and the regression residual for a fixed γ. Let $\hat{\sigma}^2(\gamma)$ denote the residual variance $n^{-1}\sum_{t=p+1}^{n} \hat{e}_t(\gamma)^2$, and $\hat{\sigma}_0^2$ that of the null model. Then, the LS estimators are

$$\hat{\gamma} = \arg\min_{\gamma \in \Gamma} \hat{\sigma}^2(\gamma), \qquad \hat{\sigma}^2 = \hat{\sigma}^2(\hat{\gamma}), \qquad \text{and } \hat{\theta} = \hat{\theta}(\hat{\gamma}). \tag{6}$$

The test is based on the Wald statistic

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

where $W_n(\gamma) = n(\hat{\sigma}_0^2/\hat{\sigma}^2(\gamma) - 1)$ is the Wald statistic for a fixed $\gamma \in \Gamma$. Thus, the statistic W_n is the well-known "sup-Wald" statistic advocated by Davies (1987). To obtain the asymptotic distribution of the statistic, make the following assumption.

Assumption 1. Let $y_t = y_0 + \sum_{s=1}^{t} u_s$ and $\{u_t\}$ be strictly stationary with mean zero and $\mathrm{E}|u_t|^{2+\delta} < \infty$, 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 $\{u_t\}$.

Next, introduce some notation. Let $[x]$ denote the integer part of x and $\Rightarrow$ denote the weak convergence of stochastic processes indexed either by $r \in [0,1]$ or by $\gamma \in \Gamma$ under the uniform metric. 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 $\omega^2 = \sum_{s=-\infty}^{\infty} r(s)$. Assume that y_0 is zero for simplicity. Wooldridge and White (1988) show that under Assumption 1 $(1/\sqrt{n}) y_{[nr]} \Rightarrow B(r)$, where B is a Brownian motion with variance ω^2. The following theorem serves as a building block for the derivation of the asymptotic distribution of the Wald statistic.

THEOREM 1. *Under Assumption 1, as $n \to \infty$,*

$$\frac{1}{n}\sum_t y_{t-1} 1\{y_{t-1} \le \gamma\} u_t \Rightarrow \int_0^1 B1\{B \le 0\} dB + \lambda \int_0^1 1\{B \le 0\}.$$

This result is derived using the martingale approximation of Hansen (1992) and extends that of Park and Phillips (2001), which considered martingale difference sequences for $\{u_t\}$. Serial correlation in $\{u_t\}$ introduces the bias term. Because of our assumption regarding compact parameter space, the parameter γ degenerates asymptotically. The remark following Theorem 5 in Section 3 discusses implications on our test in relation to the bootstrap. A consequence of this theorem is that the convergence rates of the slope estimators are the same as in the linear models, i.e., $\hat{\alpha}_i$'s are super-consistent, and the others are square root n consistent. We turn to the asymptotic distribution of the Wald statistic.

Some more notation is useful. Let G_p denote the covariance matrix of $(u_{t+1}, \ldots, u_{t+p})$ and $g_p, \tilde{r}_p$, and ι_p be p-dimensional vectors whose kth elements are $r(k), \sum_{i=1}^k r(i-1)$, and 1, respectively. Also, let

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

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

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

THEOREM 2. *Suppose that Assumption 1 holds. Then, as $n \to \infty$,*

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

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

3. RESIDUAL-BASED BLOCK BOOTSTRAP

This section describes our RBB. Because of the distributional discontinuity at the null (3), it is crucial to impose the null at a certain stage of resampling to achieve consistency of the bootstrap (see Basawa, Mallik, McCormick, Reeves, and Taylor, 1991). As our null model is the same as that of the standard unit root test, our bootstrap algorithm is similar to the ones previously proposed by Park (2002) and Paparoditis and Politis (2003) despite some differences in details. In particular, we closely follow Paparoditis and Politis (2003) as we adopt a blockwise resampling scheme rather than the sieve one to accommodate Assumption 1.

RBB proceeds as follows. First, consider the LS estimates $\hat{\alpha}_1, \hat{\alpha}_2$, and $\hat{\gamma}$ defined in (6) and construct a sequence of residuals

$$\hat{u}_t = \Delta y_t - \hat{\alpha}_1 y_{t-1} 1\{y_{t-1} \leq \hat{\gamma}\} - \hat{\alpha}_2 y_{t-1} 1\{y_{t-1} > \hat{\gamma}\}, \qquad t = 2, \ldots, n, \tag{8}$$

and that of centered ones as in Hall, Horowitz, and Jing (1995)

$$\tilde{u}_t = \hat{u}_t - \frac{1}{n-b} \sum_{i=1}^{n-b} \frac{1}{b} \sum_{j=1}^{b} \hat{u}_{i+j}, \qquad t = 2, \ldots, n. \tag{9}$$

Second, resample $\{\tilde{u}_t\}_{t=2}^n$ by the overlapping blocking scheme of Künsch (1989). For this, choose a positive integer $b < n - 1$ and let $k = [(n-1)/b]$ and $l = kb + 1$, where $[x]$ is the integer part of x. Given observations $\{\tilde{u}_t\}_{t=2}^n$, we construct $n - b$ blocks, for which the first block is $(\tilde{u}_2, \ldots, \tilde{u}_{b+1})$, the second is $(\tilde{u}_3, \ldots, \tilde{u}_{b+2})$, and so forth. Then, we draw k blocks independently with replacement from these $n - b$ blocks and connect them end-to-end, which are denoted by $u_2^*, \ldots, u_l^*$.

Third, construct a bootstrap sample $\{y_t^*\}_{t=1}^l$ by letting $y_1^* = y_1$, and $y_t^* = y_{t-1}^* + u_t^*$, $t = 2, \ldots, l$, and then a bootstrap statistic W_l^* as defined in (7) using this bootstrap sample.

Fourth, repeat the second and third steps sufficiently many times to obtain an empirical distribution of the bootstrap statistic W_l^*, which can be used to construct a bootstrap p-value of the statistic W_n.

We remark on this algorithm. Although this four-step approach is standard in the context of bootstrap unit root testing, there are various ways to modify the preceding procedure by changing ways to construct $\hat{u}_t, \tilde{u}_t$ and u_t^*. For instance, $\hat{u}_t$ may be replaced by Δy_t as in Park (2002). Paparoditis and Politis (2003) distinguished the preceding residual-based bootstrap from this difference-based bootstrap and showed that the former has better power than the latter. Although Paparoditis and Politis (2003) used the demeaning instead of (9) for simplicity, the centering in (9) facilitates the theoretical development that follows and exhibits some higher order advantage in the standard block bootstrap as shown in Hall et al. (1995). The bootstrap series $\{u_t^*\}$ may be generated from a different blocking scheme than Künsch's. The sieve resampling may also be employed if $\{u_t\}$ is a linear process.

The main difference of our bootstrap from the previous ones is that the statistic W_l^* is based on the threshold autoregression (5), which makes it more difficult to establish the consistency of our bootstrap. In particular, we need to establish Theorem 1 for the bootstrap sample $\{y_t^*\}$, which is not a trivial extension of bootstrap invariance principle or direct application of the proof of Theorem 1 to the bootstrap sample as the dependence structure of u_t^* is different from that of u_t.

The consistency of our bootstrap draws on the following invariance principle and convergence of stochastic integral. We introduce some notation. Define a standardized partial sum process $S_l^*(r) = (1/\sqrt{l}\omega^*)\sum_{t=0}^{[lr]} u_t^*$, $0 \le r \le 1$, where $u_0^* = 0, u_1^* = y_1^*, u_t^* = y_t^* - y_{t-1}^*$ for $t = 2,3,\ldots,l$ and ω^{*^2} is the variance of $l^{-1/2}\sum_{j=2}^{l} u_j^*$ conditional on a realization of $\{y_t\}$. Let W denote a standard Brownian motion and $T_n^* \Rightarrow T$ in P signify a convergence in which the distance between the law of a statistic T_n^* of a bootstrap sample and that of a random measure T tends to zero in probability for any distance metrizing weak convergence (refer to Paparoditis and Politis, 2003).

THEOREM 3. *Suppose that Assumption 1 holds. If $b \to \infty$ such that $b/\sqrt{n} \to 0$ as $n \to \infty$, then*

$$S_l^*(r) \Rightarrow W(r) \quad \textit{in } P.$$

Because of the fast rate of convergence of $\hat{\alpha}_i$'s, it can be shown that the partial sum process of the resampled centered residuals (9) is asymptotically equivalent to that of the resampled $\{u_t\}$. The invariance principle for the latter has been derived in Paparoditis and Politis (2003). Now, we turn to the convergence of stochastic integral. Although the following theorem makes use of the martingale approximation as in Theorem 1, the proofs are different because of the change in dependence structure, which is generated by the blocking resampling scheme.

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

$$\text{(i)} \quad l^{-1-k/2}\sum_{t=2}^{l} y_{t-1}^{*^k} 1\{y_{t-1}^* \le \gamma\} \Rightarrow \int_0^1 B^k 1\{B \le 0\} \textit{ in } P,$$

$$\text{(ii)} \quad l^{-1}\sum_{t=2}^{l} y_{t-1}^* 1\{y_{t-1}^* \le \gamma\} u_t^* \Rightarrow \int_0^1 B1\{B \le 0\}dB + \lambda\int_0^1 1\{B \le 0\} \textit{ in } P.$$

The asymptotic validity of RBB of W_n follows.

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

$$W_n^* \Rightarrow (\sigma^2 - g_p' G_p^{-1} g_p)^{-1} (A_{p,L} \quad A_{p,U}) \begin{pmatrix} \int_0^1 \bar{B}_L^2 & -\int_0^1 \bar{B}_L \bar{B}_U \\ -\int_0^1 \bar{B}_L \bar{B}_U & \int_0^1 \bar{B}_U^2 \end{pmatrix}^{-1} \times \begin{pmatrix} A_{p,L} \\ A_{p,U} \end{pmatrix} \quad in\ P.$$

We remark on two different ways of treating the parameter space Γ: fixed or random. Our asymptotic development with a fixed compact parameter space is limited in that the dependence of the sampling distribution of the statistic on a particular choice of Γ is not replicated in the asymptotic distribution. In the threshold literature, the set Γ is defined as the probability limit of an interval of the form $[q_n(\pi), q_n(1-\pi)]$, where $q_n(\pi)$ is the πth quantile of the threshold variable. In the testing problem in which we are interested, the threshold variable is $I(1)$ under the null, and the interval is random even asymptotically after the rescaling by the factor of the square root n. Park and Shintani (2005) study such an asymptotics with a random parameter space that is a function of π.

However, the bootstrap makes the discussion more involved. First, the bootstrap may improve upon the drawback of the fixed Γ because the bootstrap resampling replicates a particular choice of Γ. Second, the construction of Γ for the bootstrap statistic should be different for the two different approaches. In the case of the random Γ approach, we would set Γ^* as the bootstrap analogue of Γ, i.e., $[q_n^*(\pi), q_n^*(1-\pi)]$ where $q_n^*(\pi)$ is the πth quantile of the resampled threshold variable y_{t-1}^*. In the other case, Γ^* should be fixed at Γ. We may not expect much difference between the two in terms of the size of the test, as the distributional properties of y_t and y_t^* are similar under the null. In contrast, y_t is $I(0)$ under the alternative, whereas y_t^* is $I(1)$ because the bootstrap integrates the residuals. Then, we can easily expect that the interval $[q_n^*(\pi), q_n^*(1-\pi)]$ expands as the sample size increases, which will make the bootstrap critical value bigger than that of the fixed Γ. Therefore, the bootstrap with the fixed Γ should have a higher power than the one with the random Γ. See the simulation results in the next section. After all, Γ is a fixed compact set in the original sample under the alternative.

4. MONTE CARLO SIMULATION

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

Although the theory is developed for a two-regime model, the result readily extends to more general models. In practice, three-regime threshold autoregressions are commonly used. Because the stationarity of the model depends on the coefficients of the outside regimes, the testing problem is in principle the same as the two-regime model although some auxiliary assumptions are required under the null. A restricted model, which is also called band-type threshold autoregression, is considered in this simulation study:

$$\Delta y_t = \alpha_1 y_{t-1} 1\{y_{t-1} \leq \gamma_1\} + \alpha_2 y_{t-1} 1\{y_{t-1} > \gamma_2\} + u_t, \tag{10}$$

where $\gamma_1 < \gamma_2$. This model incorporates the no-adjustment periods in the middle regime and is commonly used in the threshold cointegration literature (see Lo and Zivot, 2001). The null hypothesis is the same as (3), and the test statistic can be constructed exactly the same as in (7) except that the supremum is now taken over both γ_1 and γ_2.

Several details remain to be determined to implement 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 the performance of RBB depends on those choices. We do not attempt data-dependent methods. The parameter space Γ is an interval $[-\bar{\gamma}, \bar{\gamma}]$, where $\bar{\gamma}$ is the maximum of $|y_t|$. As the sample size gets bigger, lower quantiles need to be employed to ensure boundedness. But the maximum seems to work fine in the sample size encountered in practice. We need to introduce a trimming parameter m, which is the minimum number of observations assigned to each regime when we estimate the model (5); however, this constraint 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. Specifically, a bootstrap statistic W_l^* is computed by taking the supremum of $W_l^*(\gamma)$ over the set

$$\left\{\gamma_1, \gamma_2 \in [-\bar{\gamma}, \bar{\gamma}], \middle| \sum_{t=2}^{n} 1\{y_{t-1}^* \leq \gamma_1\} \geq m \quad \text{and} \quad \sum_{t=2}^{n} 1\{y_{t-1}^* > \gamma_2\} \geq m\right\}. \tag{11}$$

Data are generated from (10) with restrictions $\alpha_1 = \alpha_2 = \alpha$ and $-\gamma_1 = \gamma_2 = \gamma$. Furthermore, let $u_t = \rho u_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, where $\{\varepsilon_t\}$ follows independent and identically distributed (i.i.d.) standard normal distributions. As is common in the conventional unit root testing literature, we consider the following combination of (ρ, θ): $(0, 0)$, $(-0.5, 0)$, $(0.5, 0)$, $(0, -0.5)$, and $(0, 0.5)$. The null hypothesis is $\alpha = 0$. When α is not zero, we set the threshold parameter γ at 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.

Table 1 summarizes the result. We only report the result with the block length $b = 6$ and the lag order $p = 3$ to save space. We tried different values of b and

Table 1. Size and power of unit root tests

	(ρ, θ)	$(0, 0)$	$(-0.5, 0)$	$(0.5, 0)$	$(0, -0.5)$	$(0, 0.5)$
$n = 100$						
$\alpha = 0$	ADF	0.050	0.065	0.075	0.085	0.050
	W_n	0.040	0.065	0.070	0.080	0.075
$\gamma = 4$	ADF	0.140	0.115	0.195	0.125	0.190
	W_n	0.165	0.140	0.190	0.115	0.205
$\gamma = 8$	ADF	0.110	0.080	0.145	0.110	0.105
	W_n	0.195	0.070	0.195	0.095	0.180
$n = 250$						
$\gamma = 4$	ADF	0.380	0.270	0.755	0.305	0.745
	W_n	0.715	0.460	0.735	0.425	0.765
$\gamma = 8$	ADF	0.140	0.105	0.230	0.140	0.270
	W_n	0.475	0.250	0.695	0.145	0.665

Note: 5% test; $p = 3$, $b = 6$, RBB.

p, and such cases are reported in Seo (2005). The results are similar to the one reported here. The rows with $\alpha = 0$ correspond to the empirical sizes of the tests. Both the ADF and W_n tests have reasonable size for most error types, even in this small sample size of 100.

The power of the tests is examined with $\alpha = -0.1$, and the rows with $\gamma =$ 4 and 8 correspond to this. We also consider two different sample sizes of 100 and 250. Across most parametrization, 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. We also find that the increase of the threshold parameter γ results in the decrease of power for both the ADF and W_n tests. 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 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 autoregressive (AR) or moving average (MA) component. This is due to the fact that the proportion of the no-adjustment region is bigger in those cases than in the other cases for a given γ.

Table 2 compares two different choices of Γ. The W_n^R indicates the bootstrap based on the random parameter space. Here the parameter space is set by trimming the lower and upper 10% of the threshold variables. As discussed in the preceding section, the power is significantly lower than that of W_n, which is based on the fixed parameter space. It is even lower than that of the ADF test.

Table 2. Power comparison of bootstrap unit root tests

	ADF	W_n^R	W_n
(ρ, θ)			
(0, 0)	0.455	0.420	0.690
(−0.5, 0)	0.210	0.185	0.415

Note: 5% test; $p = 3$, $b = 6$, $\gamma = 4$, $n = 250$.

5. CONCLUSION

This paper has developed the Wald test and RBB to test the null hypothesis of a unit root in the threshold autoregression. The simulation shows that our test 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 rejection by either of the two tests as evidence for the rejection of the presence of unit root in the process. Furthermore, the intercepts also play an important role in determining the stationarity of a threshold process. We leave this issue for future research.

REFERENCES

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

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves, & R.L. Taylor (1991) Bootstrapping unstable first-order autoregressive processes. *Annals of Statistics* 19, 1098–1101.

Bec, F., A. Guay, & E. Guerre (2008) Adaptive consistent unit root tests based on autoregressive threshold model. *Journal of Econometrics* 142, 94–133.

Bec, F. & A. Rahbek (2004) Vector equilibrium correction models with non-linear discontinuous adjustments. *Econometrics Journal* 7, 628–651.

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

Chan, K.S., J.D. Petruccelli, H. Tong, & S.W. Woolford (1985) A multiple-threshold AR(1) model. *Journal of Applied Probability* 22, 267–279.

Davies, R. (1987) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33–43.

Enders, W. & C. Granger (1998) Unit root tests and asymmetric adjustment with an example using the term structure of interest rates. *Journal of Business & Economic Statistics* 16, 304–311.

Hall, P., J.L. Horowitz, & 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.

Hansen, B. (1999) Testing for linearity. *Journal of Economic Surveys* 13, 551–576.

Kapetanios, G. & Y. Shin (2006) Unit root tests in three-regime SETAR models. *Econometrics Journal* 9, 252–278.

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

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

Lo, M. & E. Zivot (2001) Threshold cointegration and nonlinear adjustment to the law of one price. *Macroeconomic Dynamics* 5, 533–576.
Ng, S. & P. Perron (1995) Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90, 268–281.
Paparoditis, E. & D.N. Politis (2003) Residual-based block bootstrap for unit root testing. *Econometrica* 71, 813–855.
Park, J. (2002) On invariance principle for sieve bootstrap in time series. *Econometric Theory* 18, 469–490.
Park, J. & P. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69, 117–161.
Park, J. & M. Shintani (2005) Testing for a Unit Root against Transitional Autoregressive Models. Mimeo, Vanderbilt University.
Perron, P. (1989) The great crash, the oil price shock and the unit root hypothesis. *Econometrica* 57, 1361–1401.
Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.
Said, S.E. & D.A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–607.
Seo, M. (2005) Unit Root Test in a Threshold Autoregression: Asymptotic Theory and Residual-Based Block Bootstrap. Mimeo, London School of Economics, http://sticerd.lse.ac.uk/dps/em/em484.pdf.
Seo, M. (2006) Bootstrap testing for the null of no cointegration in a threshold vector error correction model. *Journal of Econometrics* 134, 129–150.
Tong, H. (1990) *Nonlinear Time Series*. Oxford University Press.
Wooldridge, J.M. & H. White (1988) Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* 4, 210–230.

APPENDIX

The following two lemmas will be repeatedly used to prove main theorems. For proofs, refer to the proofs of Lemmas 6 and 7 of Seo (2005).

LEMMA 6. *Suppose that Assumption 1 holds and* $\{a_t, b_t\}$ *is strictly stationary with* $\mathrm{E}|a_t| < \infty$ *and* $\mathrm{E}|b_t| < \infty$ *and* $\{w_t\}$ *is uniformly integrable. Then*

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

LEMMA 7. *Suppose that the sequence* $\{w_t - \mu_w\}$ *is a uniformly integrable* L^1-*mixingale and Assumption 1 holds. Then, for any integer k,*

$$\frac{1}{n^{1+k/2}}\sum_{t=1}^{n} y_t^k 1\{y_t \le \gamma\}w_t \Rightarrow \mu_w \int_0^1 B^k 1\{B \le 0\}.$$

Proof of Theorem 1. It is sufficient to derive the convergence of $(1/n)\sum_{t=2}^{n} y_{t-1}1\{y_{t-1} \le 0\}u_t$ for the same reason as in Lemma 7. Let $\zeta_t = \sum_{s=1}^{\infty}\mathrm{E}_t u_{t+s}$ and $\varepsilon_t = \sum_{s=0}^{\infty}(\mathrm{E}_t u_{t+s} - \mathrm{E}_{t-1}u_{t+s})$, where $\mathrm{E}_t X = \mathrm{E}(X|\mathcal{F}_t)$ and $\mathcal{F}_t$ is the natural filtration. Observe that $u_t = \varepsilon_t - (\zeta_t - \zeta_{t-1})$. Hansen (1992) shows that $\{\varepsilon_t\}$ is a martingale difference sequence and that $\{u_t\zeta_t - \mathrm{E}u_t\zeta_t\}$ is a uniformly integrable L^1-mixingale. Then,

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

where $R_{1n} = (1/n)y_{n-1}1\{y_{n-1} \le 0\}\zeta_n$, $R_{2n} = (1/n)\sum_{t=2}^{n}(y_t 1\{y_t \le 0 < y_{t-1}\} + y_{t-1}1\{y_{t-1} \le 0 < y_t\})\zeta_t$, and $L_n = (1/n)\sum_{t=2}^{n} u_t\zeta_t 1\{y_{t-1} \le 0\}1\{y_t \le 0\}$.

Because 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 $(1/n)\sum_{t=2}^{n} y_{t-1}1\{y_{t-1} \le 0\}\varepsilon_t \Rightarrow \int_0^1 B1\{B \le 0\}dB$. As $y_t = y_{t-1} + u_t$, we have

$$\frac{1}{n}\sum_{t=2}^{n} u_t\zeta_t[1\{y_{t-1}\le 0, y_t \le 0\} - 1\{y_t \le 0\}] = \frac{1}{n}\sum_{t=2}^{n} u_t\zeta_t 1\{y_{t-1} + u_t \le 0 < y_{t-1}\},$$

which is $o_p(1)$ by Lemma 6 because $u_t\zeta_t$ is uniformly integrable. This and Lemma 7 yield that

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

because $u_t\zeta_t - \lambda$ is a uniformly integrable L^1-mixingale.

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

$$\sup_{t\le n}\frac{1}{n}|y_t 1\{y_t \le 0\}\zeta_{t+1}| \le \sup_{t\le n}\frac{1}{\sqrt{n}}|y_t| \sup_{t\le n}\frac{1}{\sqrt{n}}|\zeta_{t+1}| = O_p(1)o_p(1).$$

Second, replace y_t by $y_{t-1} + u_t$ and note that

$$|R_{2n}| \le \frac{1}{n}\sum_{t=2}^{n}(|y_{t-1}\zeta_t| + |u_t\zeta_t|)1\{|y_{t-1}| \le |u_t|\} \le \frac{1}{n}\sum_{t=2}^{n} 2|u_t\zeta_t|1\{|y_{t-1}| \le |u_t|\}, \quad \textbf{(A.1)}$$

which is $o_p(1)$ by Lemma 6. ■

Proof of Theorem 2. Let $\bar{u}_t = u_t - (1/n)\sum_{t=p+2}^{n} u_t$ and $x_t = (x_{1t}', x_{2t}')'$, where

$$x_{1t} = \begin{pmatrix} y_{t-1}1\{y_{t-1} \le \gamma\} - \dfrac{1}{n}\displaystyle\sum_{t=p+2}^{n} y_{t-1}1\{y_{t-1} \le \gamma\} \\ y_{t-1}1\{y_{t-1} > \gamma\} - \dfrac{1}{n}\displaystyle\sum_{t=p+2}^{n} y_{t-1}1\{y_{t-1} > \gamma\} \end{pmatrix},$$

$$x_{2t} = \left(u_{t-1} - \frac{1}{n}\sum_{t=p+1}^{n-1} u_t, \ldots, u_{t-p} - \frac{1}{n}\sum_{t=2}^{n-p} u_t\right)'.$$

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

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

As in Lemma 7, $\sup_\gamma (1/n) \sum_{t=p+2}^{n} |y_{t-1} 1\{y_{t-1} \le \gamma\} - y_{t-1} 1\{y_{t-1} \le 0\}| = o_p(1)$, where the supremum is taken over a compact set. The continuous mapping theorem yields $1/(n\sqrt{n}) \sum_{t=p+2}^{n} y_{t-1} 1\{y_{t-1} \le 0\} \Rightarrow \int_0^1 B 1\{B \le 0\}$. Therefore,

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

Furthermore, by Theorem 1,

$$\frac{1}{n} \sum_{t=p+2}^{n} x_{1t} \bar{u}_t \Rightarrow \begin{pmatrix} \int_0^1 \bar{B}_L \, dB + \lambda \int_0^1 1\{B \le 0\} \\ \int_0^1 \bar{B}_U \, dB + \lambda \int_0^1 1\{B > 0\} \end{pmatrix}. \tag{A.3}$$

Next, we show that

$$\frac{1}{n} \sum_{t=p+2}^{n} y_{t-1} 1\{y_{t-1} \le \gamma\} u_{t-p} \Rightarrow \int_0^1 B 1\{B \le 0\} dB + \left(\lambda + \sum_{i=0}^{p-1} r(i) \right) \int_0^1 1\{B \le 0\}, \tag{A.4}$$

for which it is sufficient to show that

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

because of Theorem 1. To do so, note that

$$y_{t-1} 1\{y_{t-1} \le \gamma\} - y_{t-p-1} 1\{y_{t-p-1} \le \gamma\}$$
$$= u_{tp} 1\{y_{t-1} \le \gamma\} + y_{t-p-1} (1\{\gamma < y_{t-p-1} \le \gamma - u_{tp}\} - 1\{\gamma - u_{tp} < y_{t-p-1} \le \gamma\}),$$

where $u_{tp} = u_{t-1} + \cdots + u_{t-p} = y_{t-1} - y_{t-p-1}$. Next, it follows from Lemma 6 that

$$\frac{1}{n} \sum_{t=p+2}^{n} |y_{t-p-1} (1\{\gamma < y_{t-p-1} \le \gamma - u_{tp}\} - 1\{\gamma - u_{tp} < y_{t-p-1} \le \gamma\}) u_{t-p}|$$
$$\le \frac{1}{n} \sum_{t=p+2}^{n} (|\bar{\gamma}| + |u_{tp}|) |u_{t-p}| 1\{|y_{t-p-1}| \le |\bar{\gamma}| + |u_{tp}|\} \to_p 0 \tag{A.5}$$

and from Theorem 1 that

$$\frac{1}{n}\sum_{t=p+2}^{n} u_{tp}u_{t-p}1\{y_{t-1}\le\gamma\}\Rightarrow \bar{r}_p\int_0^1 1\{B\le 0\}. \qquad \textbf{(A.6)}$$

Then Theorem 2, (A.5), and (A.6) establish the convergence in (A.4), which in turn yields

$$\frac{1}{n}\sum_{t=p+2}^{n} x_{1t}x_{2t}'$$

$$\Rightarrow \begin{pmatrix} \int_0^1 \bar{B}_L\,dB + (\lambda+\bar{r}_1)\int_0^1 1\{B\le 0\}, & \ldots & \int_0^1 \bar{B}_L\,dB + (\lambda+\bar{r}_p)\int_0^1 1\{B\le 0\} \\ \int_0^1 \bar{B}_U\,dB + (\lambda+\bar{r}_1)\int_0^1 1\{B> 0\}, & \ldots & \int_0^1 \bar{B}_U\,dB + (\lambda+\bar{r}_p)\int_0^1 1\{B> 0\} \end{pmatrix}.$$

Next, it follows from the law of large numbers that

$$\frac{1}{n}\sum_{t=p+2}^{n} x_{2t}x_{2t}'\Rightarrow G_p \quad\text{and}\quad \frac{1}{n}\sum_{t=p+2}^{n} x_{2t}\bar{u}_t\Rightarrow g_p, \qquad \textbf{(A.7)}$$

which completes the convergence of $\hat{\alpha}_i$'s. It remains to derive the limit of $\hat{\sigma}^2(\gamma)$. Define κ_n as a $p+2$ dimensional diagonal matrix whose first two elements are n^{-1} and the others are $n^{-1/2}$. Then, the convergences (A.2)–(A.4) and (A.7) yield

$$\hat{\sigma}^2(\gamma) = \frac{1}{n}\sum_{t=p+2}^{n}\bar{u}_t^2 - \left(n^{-1/2}\kappa_n\sum_{t=p+2}^{n} x_t\bar{u}_t\right)'\left(\kappa_n\sum_{t=p+2}^{n} x_t x_t'\kappa_n\right)^{-1}\left(n^{-1/2}\kappa_n\sum_{t=p+2}^{n} x_t\bar{u}_t\right)$$

$$\Rightarrow \sigma^2 - g_p' G_p^{-1} g_p. \qquad \textbf{(A.8)}$$

■

The following notation is convenient for the development that follows. For a given t, define $m=[(t-2)/b]$ and $s=t-mb-1$. Let i_m be the $(m+1)$th random draw from the set $\{1,2,\ldots,n-b\}$ uniformly. As we connect the independently resampled blocks of length b end-to-end to construct $\{u_i^*\}_{i=1}^l$, u_t^* should be an observation in the block drawn $(m+1)$th and be the sth observation within that block. Thus, we can write $u_t^* = \tilde{u}_{i_m+s}$.

Proof of Theorem 3. Let $M_r = [([lr]-2)/b]$ and $B=\min\{b,[lr]-mb-1\}$. Then, $(1/\sqrt{l})\sum_{t=1}^{[lr]} u_t^* = (1/\sqrt{l})\sum_{m=0}^{M_r}\sum_{j=1}^{B}\tilde{u}_{i_m+j}$, and it is sufficient to consider $(1/\sqrt{l})\sum_{m=0}^{M_r}\sum_{j=1}^{b}\tilde{u}_{i_m+j}$, as demonstrated in Theorem 3.1 of Paparoditis and Politis (2003). See also (8.1) and the following discussion in Paparoditis and Politis (2003). Because $1/(n-b)\sum_{i=1}^{n-b}\sum_{j=1}^{b}u_{i+j} = \sum_{j=1}^{b}\mathrm{E}^*u_{i_m+j}$, let

$$\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{j=1}^{b}\tilde{u}_{i_m+j} = I_1+I_2+I_3,$$

where $I_1 = (1/\sqrt{l})\sum_{m=0}^{M_r}(\sum_{j=1}^{b}u_{i_m+j} - \sum_{j=1}^{b}\mathrm{E}^*u_{i_m+j})$, $I_2 = \hat{\alpha}_1(1/\sqrt{l})\sum_{m=0}^{M_r}(\sum_{j=1}^{b}y_{i_m+j-1}1\{y_{i_m+j-1}\le\gamma\} - \sum_{j=1}^{b}\mathrm{E}^*y_{i_m+j}1\{y_{i_m+j}\le\gamma\})$, and I_3 is similarly defined as I_2 with $\hat{\alpha}_1$ and the inequality $\leq$ in I_2 replaced by $\hat{\alpha}_2$ and $>$, respectively. Note first that

$$\mathrm{E}^*\left[\sum_{j=1}^{b} y_{i_m+j-1} 1\{y_{i_m+j-1} \leq \gamma\}\right] = \frac{1}{n-b} \sum_{t=1}^{n-b} \sum_{j=1}^{b} y_{t+j-1} 1\{y_{t+j-1} \leq \gamma\}$$

$$\leq b \sup_{t \leq n} |y_t| = O_p(b\sqrt{n}), \tag{A.9}$$

$$\mathrm{E}^*\left[\sum_{j=1}^{b} y_{i_m+j-1} 1\{y_{i_m+j-1} \leq \gamma\}\right]^2 \leq \left[b \sup_{t \leq n} |y_t|\right]^2 = O_p(b^2 n). \tag{A.10}$$

Because each block is drawn independently, (A.9) and (A.10) imply that

$$\mathrm{E}^*\left(\frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \left(\sum_{j=1}^{b} y_{i_m+j-1} 1\{y_{i_m+j-1} \leq \gamma\} - \sum_{j=1}^{b} \mathrm{E}^* y_{i_m+j} 1\{y_{i_m+j} \leq \gamma\}\right)\right)^2$$

$$= \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \mathrm{E}^*\left(\left(\sum_{j=1}^{b} y_{i_m+j-1} 1\{y_{i_m+j-1} \leq \gamma\} - \sum_{j=1}^{b} \mathrm{E}^* y_{i_m+j} 1\{y_{i_m+j} \leq \gamma\}\right)\right)^2,$$

which is $O_p(bn)$ uniformly in $\gamma \in \Gamma$ and $r \in [0,1]$. Therefore, $\mathrm{E}^* I_2^2 = O_p(bn^{-1})$, and, similarly, $\mathrm{E}^* I_3^2 = O_p(bn^{-1})$.

Then the convergence of $(1/\sqrt{l}) \sum_{t=1}^{[lr]} u_t^*$ 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 Paparoditis and Politis (2003), and the convergence of ω^* is provided in the following lemma. ■

Similarly as in the preceding proof, the following lemma is straightforward from Lemma 8.1 of Paparoditis and Politis (2003).

LEMMA 8. *Under the assumptions of Theorem 3 and as $n \to \infty$, we have (i) $l^{-1} \sum_{t=1}^{l} u_t^* \xrightarrow{p} 0$, (ii) $\omega^{*2} = \mathrm{var}^*(l^{-1/2} \sum_{t=2}^{l} u_t^*) \xrightarrow{p} \omega^2$, and (iii) $\sigma^{*2} = l^{-1} \sum_{t=1}^{l} u_t^{*2} \xrightarrow{p} \sigma^2$.*

Proof. (i) Note that $l^{-1} \sum_{t=1}^{l} u_t^* = l^{-1} \sum_{m=0}^{k-1} \sum_{j=1}^{b} u_{i_m+j} + o_p(1)$ as in the first paragraph of the proof of Theorem 3 and that $l^{-1} \sum_{m=0}^{k-1} \sum_{j=1}^{b} u_{i_m+j}$ is the sample mean of a block bootstrap series that converges to $\mathrm{E}u_t = 0$ (see Künsch, 1989). As (ii) and (iii) can be shown similarly, we show (ii). Using the same notation in the proof of Theorem 3, we write $l^{-1/2} \sum_{t=2}^{l} u_t^* = I_1 + I_2 + I_3$. It is shown in the same proof that $\mathrm{E}^* I_i \to 0$ and $\mathrm{E}^* I_i^2 \to 0$ in probability for $i = 2$ and 3. Thus, rewriting I_1 yields $\mathrm{var}^*(l^{-1/2} \sum_{t=2}^{l} u_t^*) = \mathrm{E}^*((1/\sqrt{l}) \sum_{m=0}^{M_r} \sum_{j=1}^{b} (u_{i_m+j} - \mathrm{E}^* u_{i_m+j}))^2 + o_p(1)$. The remaining steps are identical to the proof of (ii) of Lemma 8.1 of Paparoditis and Politis (2003). ■

Proof of Theorem 4. Part (i). Let $\bar{\gamma} = \max\{|\gamma|; \gamma \in \Gamma\}$. Then $y_{t-1}^{*2} |1\{y_{t-1}^* \leq \gamma\} - 1\{y_{t-1}^* \leq 0\}| \leq \bar{\gamma}^2$ for any $\gamma \in \Gamma$, and thus $l^{-2} |\sum_{t=2}^{l} y_{t-1}^{*2} 1\{y_{t-1}^* \leq \gamma\} - \sum_{t=2}^{l} y_{t-1}^{*2} 1\{y_{t-1}^* \leq 0\}| \to 0$. Then, it follows from the continuous mapping theorem and Theorem 3 that $l^{-2} \sum_{t=2}^{l} y_{t-1}^{*2} 1\{y_{t-1}^* \leq 0\} \Rightarrow \int_0^1 B^2 1\{B \leq 0\}$.

Part (ii). The proof is based on the martingale approximation for the bootstrapped sample and thus similar to the proof of Theorem 1. Let $\mathrm{E}_t^* X^* = \mathrm{E}^*(X^* | \mathcal{F}_t^*)$, where

$\mathcal{F}_t^*$ is the natural filtration associated with the bootstrapped sample. Without loss of generality, assume that $u_t^* = 0$ if $t > l$ and let $u_t^* = \varepsilon_t^* - (\zeta_t^* - \zeta_{t-1}^*)$ where $\varepsilon_t^* = \sum_{j=0}^{\infty}(\mathrm{E}_t^* u_{t+j}^* - \mathrm{E}_{t-1}^* u_{t+j}^*)$ and $\zeta_t^* = \sum_{j=1}^{\infty} \mathrm{E}_t^* u_{t+j}^*$. Because of the independent resampling of blocks of observation, $\mathrm{E}_t^* u_{t+j}^* = u_{t+j}^*$ if t and $t+j$ belong to the same block, and $\mathrm{E}_t^* u_{t+j}^* = \mathrm{E}^* u_{t+j}^*$ otherwise. Because of the centering of (9), $\sum_{j=1}^{b} \mathrm{E}^* u_{t+j}^* = 0$ for any $t = 2, \ldots, l-b$. Thus, $\varepsilon_t^* = 0$, if $s > 1$, and $\zeta_t^* = \sum_{j=1}^{b-s} u_{t+j}^*$. As a consequence, the integrated process of ε_t's up to time l corresponds to that of bootstrap samples (up to time k) that are independently resampled from the sums of blocks of observations. Therefore, defining R and L as in the proof of Theorem 1 and noting that $\zeta_l^* = 0$, we may write

$$\frac{1}{l}\sum_{t=2}^{l} y_{t-1}^* 1\{y_{t-1}^* \le \gamma\} u_t^* = \frac{1}{l}\sum_{t=2}^{l} y_{t-1}^* 1\{y_{t-1}^* \le \gamma\}\varepsilon_t^* + L_l^* + R_{2l}^*. \tag{A.11}$$

To show that $R_{2l}^* = o_p(1)$ uniformly in $\gamma \in \Gamma$, we may write $u_t^* \zeta_t^*$ as $\tilde{u}_{i_m+s} \sum_{j=s+1}^{b} \tilde{u}_{i_m+j}$, which can be rewritten as

$$\sum_{j=s+1}^{b}\left\{\left(u_{i_m+s} - \frac{1}{n-b}\sum_{g=1}^{n-b}\frac{1}{b}\sum_{v=1}^{b} u_{v+g}\right)\left(u_{i_m+j} - \frac{1}{n-b}\sum_{g=1}^{n-b}\frac{1}{b}\sum_{v=1}^{b} u_{v+g}\right)\right\} \tag{A.12}$$

$$+\, \hat{\alpha}_1^2 \sum_{j=s+1}^{b}\left\{\begin{aligned}&\left(y_{i_m+s-1} 1\{y_{i_m+s-1} \le \gamma\} - \frac{1}{n-b}\sum_{g=1}^{n-b}\frac{1}{b}\sum_{v=1}^{b} y_{v+g} 1\{y_{v+g} \le \gamma\}\right)\\ &\times\left(y_{i_m+j-1} 1\{y_{i_m+j-1} \le \gamma\} - \frac{1}{n-b}\sum_{g=1}^{n-b}\frac{1}{b}\sum_{v=1}^{b} y_{v+g} 1\{y_{v+g} \le \gamma\}\right)\end{aligned}\right\} \tag{A.13}$$

$$+\, \hat{\alpha}_2^2 \sum_{j=s+1}^{b}\left\{\begin{aligned}&\left(y_{i_m+s-1} 1\{y_{i_m+s-1} > \gamma\} - \frac{1}{n-b}\sum_{g=1}^{n-b}\frac{1}{b}\sum_{v=1}^{b} y_{v+g} 1\{y_{v+g} > \gamma\}\right)\\ &\times\left(y_{i_m+j-1} 1\{y_{i_m+j-1} > \gamma\} - \frac{1}{n-b}\sum_{g=1}^{n-b}\frac{1}{b}\sum_{v=1}^{b} y_{v+g} 1\{y_{v+g} > \gamma\}\right)\end{aligned}\right\}. \tag{A.14}$$

Note that (A.13) and (A.14) are $o_p(1)$ uniformly in t and γ because $\sup_{t,\gamma} y_t 1\{y_t \le \gamma\} = O_p(n^{1/2})$ as in the first paragraph of the proof of Theorem 3. Also note that $1/(n-b)\sum_{g=1}^{n-b}(1/b)\sum_{v=1}^{b} u_{v+g} = O_p(1/\sqrt{n})$ (see Künsch, 1989, p. 1227). Then, like (A.1), we have

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

Because $\{u_t\}$ 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 $M_1 > 0$,

$$\frac{1}{l}\sum_{t=2}^{l} \mathrm{E}1\{|y_{t-1}^*| \le \bar{\gamma} + |u_t^*|\} \le \frac{1}{l}\sum_{t=2}^{l} \mathrm{E}1\{|y_{t-1}^*| \le \bar{\gamma} + M_1\} + \frac{1}{l}\sum_{t=2}^{l} \mathrm{E}1\{|u_t^*| > M_1\}. \tag{A.15}$$

Then it follows from Theorem 3 and the proof of Lemma 6 in Seo (2005) that the first term on the right-hand side of (A.15) is $o(1)$. Next, $(1/l)\sum_{t=2}^{l} \mathrm{E}^* 1\{|u_t^*| > M_1\} \le \sup_{2\le t\le l} \mathrm{E}^* 1\{|u_t^*| > M_1\} \le P^*\{u_2^* > M_1\} + 2b/n$, because $\sup_{x\in R}|P^*\{u_t^* \le x\} - P^*\{u_j^* \le x\}| \le 2b/n$ for any j and t. Furthermore, because $\{u_t\}$ 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 $\mathrm{E}[P^*\{u_2^* > M_1\}] \le P\{u_{i_1} > M_1 - \varepsilon/2\} + \varepsilon/2 < \varepsilon$. This in turn yields that the second term on the right-hand side of (A.15) is also $o(1)$. Finally, we conclude that $|R_{2l}^*| = o_p(1)$ by uniform integrability and the fact that (A.15) is negligible (as in the proof of Lemma 6).

Next, a similar argument in Lemma 8 yields $\mathrm{E}^* u_t^* \zeta_t^* = \sum_{j=1}^{b-s} \mathrm{E}^* \tilde{u}_{i_m+s} \tilde{u}_{i_m+s+j} \xrightarrow{p} \lambda$. Then, it follows from uniform integrability, Theorem 3, and the same argument as in the proof of Lemma 7 in Seo (2005) that the limit of L_l^* is $\lambda \int_0^1 1\{B \le 0\}$.

Finally, for the convergence of the first term in (A.11), write that

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

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

$$\mathrm{E}^* \sup_{0\le\gamma\le\bar{\gamma}} \left| \frac{1}{k}\sum_{m=1}^{k-1} Y_{m-1}^* 1\{Y_{m-1}^* \le \gamma/\sqrt{b}\} V_m^* - \frac{1}{k}\sum_{m=1}^{k-1} Y_{m-1}^* 1\{Y_{m-1}^* \le 0\} V_m^* \right|$$

$$\le \frac{\bar{\gamma}}{\sqrt{b}} \frac{1}{k}\sum_{m=1}^{k-1} \Pr\{0 < Y_{m-1}^* \le \bar{\gamma}/\sqrt{b}\} \mathrm{E}^* |V_m^*| = o_p(1),$$

and the transformation $s1\{s \le 0\}$ is continuous. Therefore, the convergence follows from the invariance principle in Theorem 3, the continuous mapping theorem, and the convergence to stochastic integral of Kurtz and Protter (1991). ■

Proof of Theorem 5. Because we already have the invariance principle for RBB and the bootstrap version of Theorem 1 in hand, the proof of this theorem is straightforward following the same line of argument of the proof of Theorem 2. ■