Econometrica, Vol. 71, No. 3 (May, 2003), 813–855

RESIDUAL-BASED BLOCK BOOTSTRAP FOR UNIT ROOT TESTING

BY EFSTATHIOS PAPARODITIS AND DIMITRIS N. POLITIS[1]

A nonparametric, residual-based block bootstrap procedure is proposed in the context of testing for integrated (unit root) time series. The resampling procedure is based on weak assumptions on the dependence structure of the stationary process driving the random walk and successfully generates unit root integrated pseudo-series retaining the important characteristics of the data. It is more general than previous bootstrap approaches to the unit root problem in that it allows for a very wide class of weakly dependent processes and it is not based on any parametric assumption on the process generating the data. As a consequence the procedure can accurately capture the distribution of many unit root test statistics proposed in the literature. Large sample theory is developed and the asymptotic validity of the block bootstrap-based unit root testing is shown via a bootstrap functional limit theorem. Applications to some particular test statistics of the unit root hypothesis, i.e., least squares and Dickey-Fuller type statistics are given. The power properties of our procedure are investigated and compared to those of alternative bootstrap approaches to carry out the unit root test. Some simulations examine the finite sample performance of our procedure.

KEYWORDS: Autocorrelation, hypothesis testing, integrated series, nonstationary series, random walk, resampling.

1. INTRODUCTION

CONSIDER TIME SERIES DATA of the form $X_1, X_2, \ldots, X_n$, where $\{X_t, t = 1, 2, \ldots\}$ is a sequence of random variables. Following the seminal work of Dickey and Fuller (1979), statistical methods for detecting the possible presence of a unit root in the time series $\{X_t\}$ have attracted considerable attention over the last two decades. In particular, the assumption of interest is that the time series $\{X_t\}$ is either stationary around a (possibly nonzero) mean, or $I(1)$, i.e., integrated of order one; as usual, the $I(1)$ condition means that $\{X_t\}$ is not stationary, but its first difference series $\{Y_t\}$ is stationary (with a possibly nonzero mean), where $Y_t := X_t - X_{t-1}$. The hypothesis test setup can then be stated as:

$$H_0: \quad \{X_t\} \text{ is } I(1) \quad \text{versus}$$
$$H_1: \quad \{X_t\} \text{ is stationary.}$$

[1] This is a revised version of a paper circulated under the title "Unit Root Testing via the Continuous-path Block Bootstrap," which is now available as Discussion Paper 2001-06 from the Dept. of Economics, University of California, San Diego. We are very grateful to a co-editor for his detailed and insightful suggestions leading to an improved presentation of our results, as well as to the two referees for their most helpful comments. Many thanks are also due to Karim Abadir, Graham Elliot, Cameron Parker, Joe Romano, Hal White, and Mike Wolf for their helpful and encouraging remarks; special thanks are due to Stefano Fachin for posing some challenging questions early on, and to Mike Wolf for his generous help with some of the numerical work.

Throughout the paper we use the term 'stationary' as short-hand for 'strictly stationary.'

A first step in carrying out this hypothesis test is to choose a parameter ρ with the property that $\rho = 1$ is equivalent to H_0, whereas $\rho \neq 1$ is equivalent to H_1. A detailed discussion on different choices for the parameter ρ is given in the next section. After deciding on a particular choice for the ρ parameter, consider the new series $\{U_t\}$ defined by the equation:

$$U_t := X_t - \beta - \rho X_{t-1} \tag{1.1}$$

for $t = 1, 2, \ldots$ where the constant β is defined by $\beta = E(X_t - \rho X_{t-1})$, i.e., $E(U_t) = 0$. Equation (1.1) should be strictly considered as defining the new series $\{U_t\}$, and it is *not* to be thought of as the "model" generating the series $\{X_t\}$. In this paper, we do *not* assume a "model" for the $\{X_t\}$ series; the necessary technical assumptions placed on $\{X_t\}$ are stated in detail in Section 2. Nonetheless, definition (1.1) is very useful as the new series $\{U_t\}$ is easily seen to be stationary *always*: under H_0 and/or under H_1.

Numerous alternative procedures have been developed over the past three decades for testing the hypothesis that $\{X_t\}$ is integrated of order one (i.e., $\rho = 1$) against the alternative that it is integrated of order zero (i.e., $\rho \neq 1$); cf. Hamilton (1994) and Stock (1994) for an overview. The majority of these procedures employ certain estimators of the parameter ρ under different specifications of the estimated equation and use limiting distributions to obtain the rejection regions; cf. Fuller (1996) or Hamilton (1994). Nevertheless, the analysis is considerably complicated due to the stochastic behavior of the random quantities involved. For instance, it is well-known that the limiting distribution of the least squares (LS) estimator of the regression of X_t on X_{t-1} is nonstandard even in the simplest case of a random walk with i.i.d. residuals; this asymptotic distribution is shown to depend on the particular model fitted to the series, leading to different results for different specifications of a deterministic term. Moreover, allowing for serial correlation in the stationary process $\{U_t\}$ affects the limiting distribution by means of nuisance (and hard to estimate) parameters like the spectral density of the process at zero.

In situations like the above, where the limiting distribution of a statistic depends on difficult-to-estimate parameters, resampling methods have often in the past offered an alternative and potentially more powerful way to estimate the sampling behavior of a statistic of interest. However, none of the existing nonparametric bootstrap methods is directly applicable to the unit root nonstationary case considered here; this is true, for instance, for the block bootstrap (Künsch (1989), Liu and Singh (1992)) and the stationary bootstrap (Politis and Romano (1994)) since they are both designed for stationary weakly dependent processes.

In the paper at hand, a nonparametric block bootstrap testing procedure is introduced that is able to approximate the distribution under the null of various test statistics of the unit root hypothesis. By its construction, the testing procedure generates unit root time series by randomly selecting blocks of an appropriately defined residual process based on an empirical version of (1.1). It manages to automatically (and nonparametrically) replicate the important weak dependence characteristics of the data, e.g., the dependence structure of the stationary process $\{U_t\}$ and at the same time to mimic correctly the distribution of a particular test statistic under the null. This residual-based block bootstrap (RBB, for short) procedure is based on the block bootstrap of Künsch (1989) and Liu and Singh (1992); it constitutes a modification of the continuous-path block bootstrap algorithm introduced recently by Paparoditis and Politis (2001).

Different attempts to approach the unit root testing problem via bootstrap methods have been undertaken in the past where the theory developed has been based on restrictive assumptions on the parametric structure of the model generating the data; cf. Li and Maddala (1996) for an overview of some of the approaches proposed. Assuming a first order autoregressive process with i.i.d. errors, i.e., equation (1.1) in connection to an i.i.d. sequence $\{U_t\}$, Basawa et al. (1991), Bertail (1994), and Datta (1996) investigated the properties of a parametric autoregressive bootstrap based on i.i.d. resampling of model residuals. Ferretti and Romo (1996) extend this idea for the case where the error process is not i.i.d. but follows an autoregressive process the order of which is assumed to be finite and known. The case of an autoregressive process with finite (and known) order is also considered in a subsampling framework by Romano and Wolf (2001), and also in Park (2000) where second order size properties of an autoregressive parametric bootstrap test were investigated. Compared to these attempts our block bootstrap procedure is more general in that it is based on very weak assumptions on the dependence structure of the stationary process $\{U_t\}$, and it is not designed assuming any particular parametric structure of the process generating the data. Because of its generality, our RBB bootstrap approach can be applied to approximate the distribution of several unit root test statistics proposed in the literature; some important and popular examples are discussed in the sequel.

Note further that in designing a nonparametric bootstrap procedure for testing purposes, an additional aspect must be taken into account that is important for good power performance. For such purposes, the bootstrap procedure should be able to reproduce the sampling distribution of the test statistic under the null hypothesis (e.g., unit root integration) *whether the observed series obeys the null hypothesis or not*. For a successful unit root bootstrap test procedure it is *not* sufficient to be able to generate unit root pseudo-data, given unit root true data; the successful procedure must be able to generate unit root pseudo-data (with the correct dependence structure for the residuals) even if the true data happen to be stationary. This point has not been appropriately taken into account in the literature where bootstrap approaches are applied to the *differenced* observations and/or the theory of bootstrap validity is often derived *under the assumption* that the observed process *is* unit root integrated; cf. for example the bootstrap procedures proposed by Basawa et al. (1991), Park (2000), and Chang and Park (2001) or assumption $\beta = 1$ in Theorem 2.1 and 3.1 in Ferretti and Romo (1996). We show in this paper that the residual-based block bootstrap proposal has desirable global and local power properties. Furthermore, applying the block bootstrap to the differenced series fails if the null hypothesis is wrong, i.e., the corresponding bootstrap statistic diverges to minus infinity, leading to a loss of power.

The paper is organized as follows. Section 2 describes in detail the RBB testing procedure and states its main characteristics. A bootstrap functional limit theorem for partial sum processes based on randomly selected blocks of the residual process is established in Section 3; consequently, the asymptotic validity of the RBB bootstrap procedure in approximating the distribution of some commonly used test statistics is shown in Section 4. The global and local power properties of our procedure are investigated in Section 5 and a comparison with an alternative block bootstrap procedure based on differences is made. Section 6 discusses some practical implementation issues and examines the small sample performance of the RBB resampling method. Some comparisons to nonparametric as well as parametric bootstrap alternatives are given. Section 7 summarizes our findings while all technical proofs are deferred to Section 8.

2. RESIDUAL-BASED BLOCK BOOTSTRAP UNIT ROOT TESTING

As stated in the Introduction, we assume throughout the paper that the time series $\{X_t\}$ is either stationary (hypothesis H_1), or it is not stationary but its first difference series $\{Y_t\}$ is stationary (hypothesis H_0), where $Y_t = X_t - X_{t-1}$. For technical reasons, we strengthen the above set-up by requiring that the weak dependence structure of $\{U_t\}$ satisfy one of two sets of conditions. The first one assumes linearity, i.e., an MA(∞) representation with respect to some i.i.d. sequence $\{\varepsilon_t\}$ while the second condition replaces linearity by a strong mixing assumption.

CONDITION A: *$\{X_t\}$ satisfies one (and only one) of the following two conditions:*

(i) *(Case $\rho = 1$). $X_t = \beta + X_{t-1} + U_t$ where the process $\{U_t\}$ is generated by $U_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ with $\psi_0 = 1$, $\sum_{j=1}^{\infty} j|\psi_j| < \infty$, $C_\Psi = \sum_{j=0}^{\infty} \psi_j \neq 0$ and $\{\varepsilon_t\}$ a sequence of independent, identically distributed (i.i.d.) random variables with mean zero, positive variance σ_ε^2, and $E[\varepsilon_t^4] < \infty$.*

(ii) *(Case $\rho \neq 1$). $\{X_t\}$ is stationary and satisfies $X_t = a_0 + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ where $a_0 = \beta/(1-\rho)$ and the coefficients ψ_j and the sequence $\{\varepsilon_t\}$ satisfy the same conditions as above.*

Condition A simply states that the process $\{X_t\}$ is either a stationary linear process ($\rho \neq 1$) or it is generated by integrating such a linear process ($\rho = 1$). Note that in both cases the process $\{U_t\}$ defined by $U_t = X_t - \beta - \rho X_{t-1}$ is always linear and stationary. For $\rho = 1$ this is so by assumption while for $\rho \neq 1$ we have, since $\beta = E(X_t - \rho X_{t-1}) = (1-\rho)a_0$, that

$$U_t = -\beta + (1 - \rho L)X_t = \Psi^+(L)\varepsilon_t$$

where $\Psi^+(L) = (1-\rho L)\Psi(L) = \sum_{j=0}^{\infty} \psi_j^+ L^j$, $\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$ and L is the shift operator defined by $L^k X_t := X_{t-k}$ for $k \in \mathbb{Z}$. Clearly, $\sum_{j=0}^{\infty} j|\psi_j^+| < \infty$ and $\sum_{j=0}^{\infty} \psi_j^+ \neq 0$.

Apart from the linear class of stochastic processes, the RBB procedure can also be applied to approximate the distribution of interest in case the dependence structure of the stationary process driving the random walk is nonlinear but obeys a mixing condition. As usual, this is defined by means of the strong mixing coefficients; see e.g. Rosenblatt (1985). In particular, we say that the process $\{X_t\}$ is strong mixing if $\alpha(k) \to 0$ as $k \to \infty$ where the mixing coefficient $\alpha(k)$ is defined by

$$\alpha(k) = \sup_{A \in \mathscr{B}_{-\infty}^{0},\, B \in \mathscr{B}_{k}^{\infty}} |P(A \cap B) - P(A)P(B)|.$$

Here $\mathscr{B}_l^{l+m}$ denotes the σ-algebra generated by the set of random variables $\{X_l, X_{l+1}, \dots, X_{l+m}\}$. As an alternative to Condition A, we may impose the following condition on the process $\{X_t\}$.

CONDITION B: *For each value of ρ, the series $\{U_t\}$ is strong mixing and satisfies the following conditions*: $E(U_t) = 0, E|U_t|^{\kappa} < \infty$ *for some* $\kappa > 2$, $f_U(0) > 0$, *where f_U denotes the spectral density of $\{U_t\}$, i.e.*, $f_U(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_U(h) \exp\{i\lambda h\}$ *and* $\gamma_U(h) = E(U_t U_{t+h})$. *Furthermore*, $\sum_{k=1}^{\infty} \alpha(k)^{1-2/\kappa} < \infty$, *where $\alpha(\cdot)$ denotes the strong mixing coefficient of $\{U_t\}$.*

If $\{X_t\}$ is unit root integrated, then the above condition implies that the differenced process $X_t - X_{t-1}$ is strong mixing. On the other hand, if $\{X_t\}$ is stationary $(\rho \neq 1)$, then $\{X_t\}$ is itself a strong mixing process satisfying the conditions stated above. Note that Condition B does not imply A; see Withers (1981) or Andrews (1984).

The RBB testing algorithm is now defined in the following four steps below. As before, the algorithm is carried out conditionally on the original data $\{X_1, X_2, \dots, X_n\}$, and implicitly defines a bootstrap probability mechanism denoted by P^* that is capable of generating bootstrap pseudo-series of the type $\{X_t^*,\, t = 1, 2, \dots\}$. In the sequel, we denote quantities (expectation, variance, etc.) taken with respect to P^* with an asterisk *.

THE RBB TESTING ALGORITHM:

1. First calculate the centered residuals

$$\widehat{U}_t = (X_t - \tilde{\rho}_n X_{t-1}) - \frac{1}{n-1} \sum_{\tau=2}^{n} (X_\tau - \tilde{\rho}_n X_{\tau-1}) \tag{2.1}$$

for $t = 2, 3, \dots, n$ where $\tilde{\rho}_n = \tilde{\rho}_n(X_1, X_2, \dots, X_n)$ is a consistent estimator of ρ based on the observed data $\{X_1, X_2, \dots, X_n\}$; see Remarks 2.1 and 2.3 below.

2. Choose a positive integer $b(< n)$, and let $i_0, i_1, \dots, i_{k-1}$ be drawn i.i.d. with distribution uniform on the set $\{1, 2, \dots, n-b\}$; here we take $k = [(n-1)/b]$, where $[\cdot]$ denotes the integer part, although different choices for k are also possible. The procedure constructs a bootstrap pseudo-series $X_1^*, \dots, X_l^*$, where $l = kb + 1$, as follows:

$$X_t^* = \begin{cases} X_1 & \text{for } t = 1, \\ \hat{\beta} + X_{t-1}^* + \widehat{U}_{i_m + s} & \text{for } t = 2, 3, \dots, l, \end{cases}$$

where $m = [(t-2)/b]$, $s = t - mb - 1$, and $\hat{\beta}$ is a drift parameter that is either set equal to zero ($\hat{\beta} \equiv 0$), or $\hat{\beta} = \tilde{\beta}$ where $\tilde{\beta}$ is a $\sqrt{n}$-consistent estimator of β; see Remark 2.2 below.

3. Let $\hat{\rho}_n$ be the estimator used to perform the unit root test. Compute the pseudo-statistic $\hat{\rho}^*$, which is nothing other than the statistic $\hat{\rho}_l$ based on the pseudo-data $\{X_1^*, \ldots, X_l^*\}$.

4. Repeating steps 2–3 a great number of times (B times, say), we obtain the collection of pseudo-statistics $\hat{\rho}_1^*, \ldots, \hat{\rho}_B^*$. As will be shown shortly, an empirical distribution based on the pseudo-statistics $\hat{\rho}_1^*, \ldots, \hat{\rho}_B^*$ provides a consistent approximation of the distribution of $\hat{\rho}_n(X_1, \ldots, X_n)$ under the null hypothesis $H_0 : \rho = 1$. The α-quantile of the bootstrap distribution in turn yields a consistent approximation to the α-quantile of the true distribution (under H_0), which is required in order to perform an α-level test of H_0.

REMARK 2.1: The block bootstrap is a central part in the RBB procedure; note however, that the block bootstrap is not applied to the $\{X_t\}$ data directly, neither to its first differences; rather, the pseudo-series $X_1^*, X_2^*, \ldots, X_l^*$ is obtained by integrating randomly selected blocks of *centered* residuals $\widehat{U}_t$. The reason for this centering is that although the series $U_t = X_t - \beta - \rho X_{t-1}$ has a zero mean both under the null and under the alternative, the estimated innovations $\widetilde{U}_t = X_t - \tilde{\beta} - \tilde{\rho}_n X_{t-1}$ will likely have nonzero (sample) mean; this discrepancy has an important effect on the bootstrap distribution effectively leading to a random walk with *drift* in the bootstrap world. Note that $\widehat{U}_t$ defined in (2.1) is a centered version of $\widetilde{U}_t$ as defined in eq. (2.1), i.e., $\widehat{U}_t = \widetilde{U}_t - (1/(n-1)) \sum_{\tau=2}^{n} \widetilde{U}_\tau$, since the factor $\tilde{\beta}$ cancels out.

REMARK 2.2: $\hat{\beta}$ denotes the drift parameter in the RBB resampling scheme, which can be set equal to $\tilde{\beta}$, the latter being the estimator of the intercept term in the particular equation fitted to the series in order to obtain $\tilde{\rho}_n$; see the discussion below. Note that there are cases where it is appropriate to set $\hat{\beta} \equiv 0$ in the second step of the RBB algorithm. This is, for instance, true if we are interested in generating a unit root process without drift in order to approximate the distribution of the estimator $\hat{\rho}_n$ under this assumption. It is well known (see Hamilton (1994)) that the distribution of a particular estimator $\hat{\rho}_n$ used to test the unit root hypothesis is not only affected by the question whether the true series is generated by a model with an intercept term or not but also by the specification of the deterministic term in the model fitted to the observed series.

REMARK 2.3: The quantity $\tilde{\rho}_n$ appearing in equation (2.1) is an appropriately chosen consistent estimator of the parameter ρ based on the data $\{X_1, X_2, \ldots, X_n\}$. In particular, for the validity of the RBB testing procedure we require that $\tilde{\rho}_n$ satisfy the following conditions: If $\rho = 1$ then

$$(2.2) \qquad \tilde{\rho}_n = \rho + O_P\big(n^{-(2+\delta(\beta))/2}\big)$$

where $\delta(\beta) = 1$ if $\beta \neq 0$ and $\delta(\beta) = 0$ if $\beta = 0$. If $\rho \neq 1$, then

$$\tilde{\rho}_n = \rho + o_P(1) \tag{2.3}$$

for every $\beta \in \mathbb{R}$. Conditions (2.2) and (2.3) are satisfied by many estimators; we elaborate with two specific examples.

EXAMPLE 2.1: Assume for simplicity that $\beta = 0$ and let the parameter ρ have the meaning of the asymptotic lag-1 autocorrelation of series $\{X_t\}$, i.e., let

$$\rho = \lim_{t\to\infty} \frac{EX_t X_{t+1}}{EX_t^2}.$$

Note that under H_1 the series $\{X_t\}$ is stationary, and therefore the limit is unnecessary. Nevertheless, the limiting operation is required under H_0 (i.e., if the series $\{X_t\}$ is $I(1)$), in which case we can easily calculate that $EX_t X_{t+1}/EX_t^2 = 1 + O(1/t)$, under the sole assumption that the series $\{U_t\}$ possesses a spectral density (which is guaranteed by either Condition A or B).

Thus, if the series $\{X_t\}$ is $I(1)$, then $\rho = 1$. To show that H_0 is essentially equivalent to $\rho = 1$ in this case note that if $\rho = 1$, then either $\{X_t\}$ is $I(1)$, or it is the trivial stationary process with constant sample-paths (by the Cauchy-Schwarz inequality); but even this latter case can be put in the $I(1)$ framework: $X_t = \beta + X_{t-1} + U_t$ where $\{U_t\}$ is stationary but with $\mathrm{var}(U_t) = 0$.

Let $\tilde{\beta} = \hat{\beta}$ and $\tilde{\rho}_n = \hat{\rho}_{LS,C}$ be the (ordinary) LS estimators of β and ρ obtained by fitting

$$X_t = \beta + \rho X_{t-1} + e_t \tag{2.4}$$

to the observed series. It is well known that the LS estimator $\hat{\rho}_{LS,C}$ of ρ in the above regression equation satisfies conditions (2.2) and (2.3); see Brockwell and Davis (1991) for the stationary case, Phillips (1987a) for the integrated case with $\beta = 0$, and West (1988) for $\beta \neq 0$.

To introduce the next example we modify Condition A by restricting the class of linear processes considered to those possessing an infinite order autoregressive representation.

CONDITION A′: *The process $\{X_t\}$ satisfies Condition A and the power series $\Psi(z) = 1 + \sum_{j=1}^{\infty} \psi_j z^j$ is bounded, and bounded away from zero for $|z| \leq 1$.*

Condition $\Psi(z) \neq 0$ for $|z| \leq 1$ implies the existence of an infinite order autoregressive representation for $\{U_t\}$ if $\rho = 1$. In particular, in this case we have that $U_t = X_t - \beta - X_{t-1}$ has the representation $U_t = -\sum_{j=1}^{\infty} \pi_j U_{t-j} + \varepsilon_t$ where $\pi(z) = 1 + \sum_{j=1}^{\infty} \pi_j z^j = 1/\Psi(z)$. Note that, still in the $\rho = 1$ case, the integrated process $\{X_t\}$ can be also expressed as $(1-L)(1+\sum_{j=1}^{\infty} \pi_j L^j) X_t = \pi(1)\beta + \varepsilon_t$, i.e., the power series $\tilde{\pi}(z) = (1-z)\pi(z)$ has a unit root. If $\rho \neq 1$, then Condition A′

implies that $X_t = \pi(1)a_0 - \sum_{j=1}^{\infty} \pi_j X_{t-j} + \varepsilon_t$ since in this case $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ and the power series $\Psi(z)$ has no zeros for $|z| \le 1$. Therefore, Condition A′ states that either $\{X_t\}$ is a stationary process possessing an infinite order autoregressive representation ($\rho \ne 1$) or $\{X_t\}$ is obtained by integrating such a process ($\rho = 1$).

If the process $\{X_t\}$ satisfies Condition A′, then the following representation is very useful. It generalizes the one given in Fuller (1996) for finite order autoregressive processes.

LEMMA 2.1: *If $\{X_t\}$ satisfies Condition A′, then*

$$X_t = c + \rho X_{t-1} + \sum_{j=1}^{\infty} a_j (X_{t-j} - X_{t-j-1}) + \varepsilon_t \tag{2.5}$$

where $\sum_{j=1}^{\infty} |a_j| < \infty$ and the coefficients $\{a_j\}$ are defined as follows: If $\{X_t\}$ is unit root integrated, then $c = \pi(1)\beta$, $\rho = 1$, and $a_j = \pi_j$ for $j = 1, 2, \ldots$, while if $\{X_t\}$ is stationary, then $c = \pi(1)a_0$,

$$\rho = -\sum_{j=1}^{\infty} \pi_j, \quad \text{and} \quad a_j = \sum_{s=j+1}^{\infty} \pi_s \quad \text{for} \quad j = 1, 2, \ldots.$$

In the unit root integrated case, representation (2.5) is obviously valid for $\rho = 1$ and $a_j = \pi_j$, since in this case $X_t = \beta + X_{t-1} + U_t$ and $U_t = -\sum_{j=1}^{\infty} \pi_j U_{t-j} + \varepsilon_t$ by assumption. On the other hand, if $\{X_t\}$ is stationary, then the infinite order autoregressive representation of $\{X_t\}$ can be written as

$$\begin{aligned} X_t = \pi(1)a_0 - \sum_{j=1}^{\infty} \pi_j X_{t-1} + \sum_{j=2}^{\infty} \pi_j (X_{t-1} - X_{t-2}) \\ + \sum_{j=3}^{\infty} \pi_j (X_{t-2} - X_{t-3}) + \cdots + \varepsilon_t \end{aligned}$$

from which the postulated representation follows by the choice of ρ and a_j stated in the lemma. Since for $\rho \ne 1$ the process $\{X_t\}$ is stationary and $\sum_{j=1}^{\infty} j|\pi_j| < \infty$, we get by simple algebra that

$$\sum_{j=1}^{\infty} |a_j| = \sum_{j=1}^{\infty} \left| \sum_{s=j+1}^{\infty} \pi_j \right| \le \sum_{j=1}^{\infty} j|\pi_j| < \infty.$$

Thus if the process $\{X_t\}$ is generated by a stationary infinite order autoregressive process, then the parameter ρ appearing in definition (1.1) and equation (2.5) is given by $\rho = -\sum_{j=1}^{\infty} \pi_j$. Furthermore, this parameter always satisfies $\rho \le 1$. To see this, note that the condition $\pi(z) \ne 0$ for $|z| < 1$ implies (by the continuity of the power series $\pi(z)$ and the fact that $\pi(0) = 1$) that $\pi(z) > 0$ for $|z| < 1$. By continuity again we get $\lim_{z \to 1} \pi(z) = 1 + \sum_{j=1}^{\infty} \pi_j \ge 0$ which is just $\rho \le 1$. We are now ready to introduce our next example for the choice of the estimator $\tilde{\rho}_n$.

EXAMPLE 2.2: Assume that the underlying process satisfies Condition A′ and let $\tilde{\beta}=\hat{\beta}$ and $\tilde{\rho}_n=\hat{\rho}_{DF,C}$ be the so-called augmented Dickey-Fuller (DF) estimator of ρ obtained by fitting a truncated version of (2.5) to the observed series, i.e., by fitting the model

$$X_t=\beta+\rho X_{t-1}+\sum_{i=1}^{p}a_i(X_{t-i}-X_{t-i-1})+\varepsilon_t. \tag{2.6}$$

Consistency of $\hat{\rho}_{DF,C}$ requires that the order $p=p(n)$ in the above equation increase to infinity at some appropriate rate as the sample size n increases; see Said and Dickey (1984). For fixed p, equation (2.6) is the set-up considered by Dickey and Fuller (1979) for finite order autoregressive processes with known order.

From the discussion following Condition A′ it is clear that $\rho=1$ is equivalent to the null hypothesis of unit root integration. Under some assumptions on the rate with which p increases, Said and Dickey (1984) showed that in the unit root case and if $\beta=0$, $\hat{\rho}_{DF,C}=1+O_P(n^{-1})$. An extension of the arguments presented there shows that for $\beta\neq 0$, $\hat{\rho}_{DF,C}=1+O_P(n^{-3/2})$; cf. Hamilton (1994, p. 539–540) for the finite order autoregressive case. In Section 8 we consider the case where $\{X_t\}$ is a stationary process having the representation (2.5). We show that under some conditions on the rate at which p increases to infinity $\hat{\rho}_{DF,c}=\rho+O_P(n^{-1/2}c_n^{-1/2})$ where $\rho=-\sum_{j=1}^{\infty}\pi_j$ and where c_n is a sequence approaching zero as $n\to\infty$ but such that $n^{1/2}c_n^{1/2}\to\infty$; cf. Lemma 8.3. Therefore, the estimator $\hat{\rho}_{DF,C}$ considered in this example obeys the stochastic behavior (2.2) and (2.3).

In concluding the discussion of this example, note that for the choice of ρ discussed here the value of ρ under the alternative is *not* necessarily in the interval $[-1, 1]$; although $\rho\leq 1$ always, it may be the case that $\rho<-1$. For instance, if $X_t=\varepsilon_t-\theta\varepsilon_{t-1}$ with $|\theta|<1$, $\theta\neq 0$, then X_t has an autoregressive representation with $\pi_j=\theta^j$, i.e., $\rho=-\sum_{j=1}^{\infty}\pi_j=-\theta/(1-\theta)$, which is less than -1 for $\theta\in(1/2,1)$. Furthermore, Condition A′ does not necessarily imply that the stationary process $\{U_t\}$ has an AR(∞) representation with absolutely summable coefficients in the case that the original process $\{X_t\}$ is stationary and possesses such a representation. To see this recall that for $\rho\neq 1$, Condition A′ implies that $X_t=\pi(1)a_0-\sum_{j=1}^{\infty}\pi_jX_{t-j}+\varepsilon_t$. Now, if $U_t=\sum_{j=1}^{\infty}c_jU_{t-j}+\varepsilon_t$ holds true, then using the definition $U_t=X_t-\beta-\rho X_{t-1}$ and rearranging terms, it follows that $c_0=1$ and $c_j=\pi_j+\rho c_{j-1}$ for $j=1,2,\ldots$, which implies that $|c_j|\not\to 0$ if $\rho\leq -1$.

What is apparent from the two examples discussed so far, is that there are different possibilities for the meaning we attach to the parameter ρ figuring in equation (1.1). Whereas $\rho=1$ is equivalent to an integrated (unit root) series in both examples considered, the meaning of $\rho=c$ (where $c\neq 1$ is some constant) is different; this is a most important point in order to understand how our testing procedure behaves when the $\{X_t\}$ data are actually stationary.

REMARK 2.4: In the above examples, the parameter estimator $\tilde{\rho}_n$ used to calculate the residuals $\widehat{U}_t$ in (2.1) is obtained by fitting the corresponding regression equations *including* an intercept term β despite the fact that the true value of β may be equal to zero. In particular, if the practitioner wishes to apply the ordinary LS statistic to test the null hypothesis of unit root integration, then the estimator $\tilde{\rho}_n$ used should be the one obtained by fitting the LS regression (2.4) to the observed series. Similarly, if the practitioner wishes to apply the augmented DF type statistic for the same purpose, then the estimator $\tilde{\rho}_n$ should be the one obtained by fitting the regression (2.6). The reason why an intercept term is always included is that the estimator $\tilde{\rho}_n$ so obtained automatically satisfies conditions (2.2) and (2.3), i.e., the practitioner does not have to worry about the true value of the parameter β in (1.1). Note that the estimator $\tilde{\rho}_n$ used to perform the unit root test can be based on a different specification of the deterministic term than the one used in obtaining the estimator $\tilde{\rho}_n$; see Section 4.

3. A FUNCTIONAL LIMIT THEOREM FOR THE BOOTSTRAP PARTIAL SUM PROCESS

The asymptotic properties of the RBB testing procedure are largely based on the stochastic behavior of the standardized partial sum process $\{S_l^*(r), 0 \leq r \leq 1\}$ defined by

$$S_l^*(r) = \frac{1}{\sqrt{l}} \sum_{t=1}^{j-1} U_t^*/\sigma^* \quad \text{for} \quad \frac{(j-1)}{l} \leq r < \frac{j}{l} \qquad (j=2,\ldots,l) \tag{3.1}$$

and

$$S_l^*(1) = \frac{1}{\sqrt{l}} \sum_{t=1}^{l} U_t^*/\sigma^*, \tag{3.2}$$

where $U_1^* \equiv X_1$, $U_t^* = X_t^* - \hat{\beta} - X_{t-1}^*$ for $t = 2, 3, \ldots, l$, and $\sigma^{*^2} = \text{var}^*(l^{-1/2} \times \sum_{j=1}^{l} U_j^*)$. Note that $S_l^*(r)$ is a random element in the function space $D[0,1]$, i.e., the space of all real valued functions on the interval $[0,1]$ that are right continuous at each point and have finite left limits.

The following theorem shows that under a general set of assumptions on the process $\{X_t\}$, and conditionally on the observed series $X_1, X_2, \ldots, X_n$, the bootstrap partial sum process defined by (3.1) and (3.2) converges weakly to the standard Wiener process on $[0,1]$. This process is denoted in the following by W. To clarify some terminology used here and elsewhere in the paper, we note that if $T_n^* = T_n^*(X_1^*, X_2^*, \ldots, X_n^*)$ is a random sequence based on the bootstrap sample $X_1^*, X_2^*, \ldots, X_n^*$ and G is a random measure, then the notation $T_n^* \Rightarrow G$ *in probability* means that the distance between the law of T_n^* and the law of G tends to zero in probability for any distance metrizing weak convergence.

THEOREM 3.1: *Let $\{X_t\}$ be a stochastic process, assume that the process $\{U_t\}$ defined by $U_t = X_t - \beta - \rho X_{t-1}$ satisfies Condition A or Condition B, and let $\tilde{\rho}_n$ be* *an estimator of* ρ *such that equation* (2.2) *and* (2.3) *are satisfied. If* $b \to \infty$ *such that* $b/\sqrt{n} \to 0$ *as* $n \to \infty$, *then*

$$S_l^* \Rightarrow W \quad \textit{in probability}.$$

This basic result together with a bootstrap version of the continuous mapping theorem enables us to apply the block bootstrap proposal of this paper in order to approximate the null distribution of a variety of different test statistics proposed in the literature that correspond to different choices of the parameter ρ and the estimators $\tilde{\rho}_n$ and $\hat{\rho}_n$, provided the following two conditions are fulfilled: (a) The choice of the parameter ρ is such that $\rho = 1$ is equivalent to the null hypothesis of unit root integration, while $\rho \neq 1$ is equivalent to the alternative of a stationary process, and (b) the estimator $\tilde{\rho}_n$ of ρ used to calculate the residual series $\widehat{U}_t$ in (2.1) satisfies (2.2) and (2.3).

4. APPLICATIONS TO UNIT ROOT TESTING

In the first application we show consistency of the RBB procedure in approximating the distribution of the LS estimator obtained by regressing X_t on X_{t-1}, i.e., our Example 2.1. Interest in the corresponding test statistic that has been investigated by Dickey and Fuller (1979), Phillips (1987a), Phillips and Perron (1988), Abadir (1993), and Fuller (1996), occurs mainly because of its simplicity and the fact that it allows for testing the unit root integrated hypothesis without parameterizing the weak dependence structure of the process.

In the second application, validity of the RBB testing procedure in approximating the null distribution of the so-called augmented DF test statistic based on the regression (4.2) is shown; cf. Fuller (1996), Said and Dickey (1984), and our Example 2.2. For simplicity of exposition, we discuss in more detail the case where the generated process has no intercept, i.e., $\beta = 0$; the case where $\beta \neq 0$ with a possible linear trend is briefly addressed in Section 4.3.

4.1. *Statistics Based on the Lag-1 Autocorrelation*

Assume that $\beta = 0$ and consider the LS estimator of the parameter ρ in the regression

$$X_t = \rho X_{t-1} + e_t. \tag{4.1}$$

It is well known that, under the null hypothesis where $\{X_t\}$ is unit root integrated, the asymptotic distribution of this estimator is affected if a constant term is included in the regression (4.1) or not. Both cases can be handled by our bootstrap algorithm.

Let $\hat{\rho}_{LS}$ denote the LS estimator of ρ in (4.1) and $\hat{\rho}_{LS,C}$ that in (2.4); $\hat{\rho}_{LS,C}$ includes an intercept in the regression while $\hat{\rho}_{LS}$ does not. To approximate the distribution of $\hat{\rho}_{LS}$ we apply the RBB algorithm given in Section 2 by using the estimator $\tilde{\rho}_n = \hat{\rho}_{LS,C}$ in order to calculate the centered residuals $\widehat{U}_t$ in the first step. Furthermore, we set $\hat{\beta} \equiv 0$ in the second step while the pseudo-statistic $\hat{\rho}^*$ computed in step 4 is given by the LS estimator of the parameter of X^*_{t-1} obtained by regressing X^*_t on X^*_{t-1} without intercept, i.e., equation (4.1). We denote this estimator by $\hat{\rho}^*_{LS}$. Similarly, in order to approximate the distribution of $\hat{\rho}_{LS,C}$ we use the same bootstrap variables as above but we include a constant term in the regression of X^*_t on X^*_{t-1}. The estimator of the coefficient of X^*_{t-1} in this regression is denoted by $\hat{\rho}^*_{LS,C}$. Note that since we are interested in approximating the null distribution of $n(\hat{\rho}_{LS}-1)$ or of $n(\hat{\rho}_{LS,C}-1)$ under the assumption that $\beta = 0$, for both cases the generated pseudo-series $X^*_1, X^*_2, \ldots, X^*_l$ is unit root integrated *without* drift. The following theorem summarizes the behavior or our bootstrap proposal.

THEOREM 4.1: *Assume that the process $\{X_t\}$ satisfies Condition A or Condition B with $\beta = 0$. If $b \to \infty$ but $b/\sqrt{n} \to 0$ as $n \to \infty$, then*:

(i) $$\sup_{x\in\mathbb{R}} |P^*(l(\hat{\rho}^*_{LS}-1) \le x \mid X_1, X_2, \ldots, X_n) - P_0(n(\hat{\rho}_{LS}-1) \le x)| \to 0$$

in probability and

(ii) $$\sup_{x\in\mathbb{R}} |P^*(l(\hat{\rho}^*_{LS,C}-1) \le x \mid X_1, X_2, \ldots, X_n) - P_0(n(\hat{\rho}_{LS,C}-1) \le x)| \to 0$$

in probability, where P_0 denotes the probability measure corresponding to the case where the statistics $\hat{\rho}_{LS}$ and $\hat{\rho}_{LS,C}$ are computed from a stretch of size n from the unit root process obtained by integrating $\{U_t\}$.

4.2. *Dickey-Fuller Type Statistics*

Assume that $\{X_t\}$ satisfies Condition A′ with $\beta = 0$ and consider the problem of approximating the distribution of the augmented DF type estimator $\hat{\rho}_{DF}$ under the null hypothesis, where $\hat{\rho}_{DF}$ denotes here the LS estimator obtained by fitting the regression equation

$$X_t = \rho X_{t-1} + \sum_{i=1}^{p} a_i (X_{t-i} - X_{t-i-1}) + \varepsilon_t \tag{4.2}$$

to the observed series $X_1, X_2, \ldots, X_n$. To do this, define first the centered differences

$$D_t = X_t - X_{t-1} - \frac{1}{n-1}\sum_{t=2}^{n}(X_\tau - X_{\tau-1}),$$

$t = 2, 3, \ldots, n$. To estimate the distribution of $\hat{\rho}_{DF}$ we apply the RBB algorithm as follows: An estimator $\tilde{\rho}_n$ of ρ in (2.5) that satisfies (2.2) and (2.3) is used to calculate the centered residuals $\widehat{U}_t$ in the first step of the algorithm. Such an estimator is, for instance, $\hat{\rho}_{DF,C}$ obtained from (2.6). The pseudo-series $X_1^*, X_2^*, \ldots, X_l^*$ is then generated following steps 2 to 4 where $\hat{\beta} \equiv 0$ in the second step. Additionally to the bootstrap series X_t^*, we also generate a pseudo-series of l centered differences denoted by $D_1^*, D_2^*, \ldots, D_l^*$ as follows: For the first block of $b+1$ observations we set $D_1^* = 0$ and

$$D_j^* = D_{i_0+j-1}$$

for $j = 2, 3, \ldots, b+1$. For the $(m+1)$th block, $m = 1, \ldots, k-1$ we define

$$D_{mb+1+j}^* = D_{i_m+j},$$

where $j = 1, 2, \ldots, b$. We then calculate the regression of X_t^* on X_{t-1}^* and on $D_{t-1}^*, D_{t-2}^*, \ldots, D_{t-p}^*$. The least squares estimator of the coefficient of X_{t-1}^* in this regression, denoted by $\hat{\rho}_{DF}^*$, is used to approximate the distribution of the estimator $\hat{\rho}_{DF}$ under the null hypothesis.

To approximate the distribution of $\hat{\rho}_{DF,C}$, i.e., of the least squares estimator of ρ in (2.6), we just include a constant term in the corresponding regression fitted to the pseudo-series $\{X_t^*, D_t^*, t = 1, 2, \ldots, l\}$. The so obtained least squares estimator of the coefficient of X_{t-1}^* is denoted in the following by $\hat{\rho}_{DF,C}^*$.

To motivate the above use of the RBB algorithm to approximate the distribution of $\hat{\rho}_{DF}$ and $\hat{\rho}_{DF,C}$, consider for instance the estimator $\hat{\rho}_{DF}$ and recall our target regression (4.2), which relates X_t on X_{t-1} and on the lagged differences $X_{t-j} - X_{t-j-1}$, $j = 1, 2, \ldots, p$. Now, in the bootstrap world $X_{t-j}^* - X_{t-j-1}^* = U_{t-j}^*$, which for large n behaves like the random variable U_{t-j}, i.e., the bootstrap differences $X_{t-j}^* - X_{t-j-1}^*$ behave asymptotically like $X_{t-j} - \rho X_{t-j-1}$ and not like $X_{t-j} - X_{t-j-1}$. Thus regressing X_t^* on X_{t-1}^* and on the lagged differences $X_{t-j}^* - X_{t-j-1}^*$, $j = 1, 2, \ldots, p$, will mimic the regression of X_t on X_{t-1} and on U_{t-j}, $j = 1, 2, \ldots, p$. This, however, coincides with our target regression (4.2) only if $\rho = 1$, i.e., only if the observed series is indeed unit root integrated. Furthermore, as we have seen in Example 2.2, such an infinite order autoregressive representation for the process $\{U_t\}$ may not exist if $\rho \neq 1$.

Now, to understand from where the definition of the new bootstrap variables D_t^* comes, consider the bootstrap observations in the $(m+1)$th block given by X_{mb+1+s}^* where $s \in \{1, 2, \ldots, b\}$. Here we have $X_{mb+1+s}^* = X_{mb+s}^* + \widehat{U}_{i_m+s+1}$; note that, for large n, $\widehat{U}_{i_m+s+1}$ behaves like U_{i_m+s+1}, which by (2.5) depends on the lagged differences $X_{i_m+s+1-j} - X_{i_m+s-j}$, $j = 1, 2, \ldots$. Thus the bootstrap analogue of (4.2) will be to regress X_{mb+1+s}^* on X_{mb+s}^* and on $X_{i_m+s+1-j} - X_{i_m+s-j}$, $j = 1, 2, \ldots, p$. Note that $D_{i_m+1+s-j}$ is just a centered version of $X_{i_m+s+1-j} - X_{i_m+s-j}$.

THEOREM 4.2: *Assume that the process $\{X_t\}$ satisfies Condition A′ with $\beta = 0$. Assume further that $p \to \infty$ as $n \to \infty$ such that $p^3/\sqrt{n} \to 0$ and $\sqrt{n}\sum_{j=p+1}^{\infty} |a_j| \to 0$. If $b \to \infty$ such that $b/\sqrt{n} \to 0$ as $n \to \infty$, then*

$$\text{(i)} \qquad \sup_{x \in \mathbb{R}} |P^*((l-p)(\hat{\rho}_{DF}^* - 1) \leq x \mid X_1, X_2, \ldots, X_n) - P_0((n-p)(\hat{\rho}_{DF} - 1) \leq x)| \to 0$$

in probability and

(ii) $$\sup_{x\in\mathbb{R}} |P^*((l-p)(\hat{\rho}^*_{DF,C}-1) \leq x \mid X_1, X_2, \ldots, X_n) \\ - P_0((n-p)(\hat{\rho}_{DF,C}-1) \leq x)| \to 0$$

in probability.

Here P_0 denotes the probability measure corresponding to the case where the statistics $\hat{\rho}_{DF}$ and $\hat{\rho}_{DF,C}$ are computed from a stretch of size n from the unit root process obtained by integrating $\{U_t\}$.

We stress here the fact that in both cases discussed here the pseudo-series $X_1^*, X_2^*, \ldots, X_n^*$ is generated in a nonparametric way, i.e., using the RBB bootstrap procedure and not the parametric form given in (4.2). Because of this we expect the bootstrap to be able to mimic also correctly the 'truncation effect' on the distribution of $\hat{\rho}_{DF}$ and $\hat{\rho}_{DF,C}$. This truncation effect is due to the fact that a model with a finite lag p is fitted to the series at hand whereas an infinite order is ideally required because of the infinite order representation (2.5). In other words, we expect the results based on the RBB testing procedure to be less sensitive with respect to the choice of the parameter p. Section 6.2 presents numerical illustrations of this behavior.

4.3. *The Case of Nonzero Mean*

In order to cover the usual four cases of interest (see e.g. Chapter 17 of Hamilton (1994)), we now show that the RBB bootstrap algorithm can also be used to approximate the null distribution of our test statistics when the true process has an intercept term, i.e., that $\beta \neq 0$, and that the equation fitted to the observed series includes a constant or even a linear time trend component. We may, for instance, be interested in approximating the distribution of the LS estimator of the coefficient of X_{t-1} under the null hypothesis that $\rho = 1$, if either the model (2.4) or the model

$$X_t = \beta + \beta_1 t + \rho X_{t-1} + e_t \tag{4.3}$$

is fitted to the observed series $X_1, X_2, \ldots, X_n$. Similarly, the statistic of interest may be the LS estimator of ρ in (2.6) or in

$$X_t = \beta + \beta_1 t + \rho X_{t-1} + \sum_{i=1}^{p} a_i (X_{t-i} - X_{t-i-1}) + \varepsilon_t. \tag{4.4}$$

The theory developed in this paper can be easily applied to establish asymptotic validity of the RBB algorithm in this setting too. For instance, denote by $\hat{\rho}_{LS,T}$ the LS estimator of ρ in (4.3) and by $\hat{\rho}^*_{LS,T}$ the corresponding estimator using the bootstrap series $X_1^*, X_2^*, \ldots, X_n^*$. The bootstrap pseudo-series is generated here using the RBB bootstrap algorithm with $\tilde{\rho} = \hat{\rho}_{LS,C}$ in (2.1) and $\hat{\beta} = \tilde{\beta}$ in the second step, where $\tilde{\beta}$ denotes the estimator of β in (2.4). This is in contrast to the situation in Section 4.1 and 4.2 where we set $\hat{\beta} \equiv 0$ in the second step of the RBB algorithm. The following theorem summarizes the behavior of our bootstrap proposal in approximating the distribution of $n(\hat{\rho}_{LS,T} - 1)$ under the null hypothesis that $\rho = 1$ and $\beta_1 = 0$. An analogous result for approximating the distribution of the Dickey-Fuller statistic $(n-p)(\hat{\rho}_{DF,T} - 1)$ under the same null hypothesis is easily established.

THEOREM 4.3: *Assume that the process $\{X_t\}$ satisfies Condition B. If $b \to \infty$ but $b/\sqrt{n} \to 0$ as $n \to \infty$, then*

$$\sup_{x \in \mathbb{R}} \left| P^*\big(l\big(\hat{\rho}^*_{LS,T} - 1\big) \leq x \mid X_1, X_2, \ldots, X_n\big) - P_0\big(n\big(\hat{\rho}_{LS,T} - 1\big) \leq x\big) \right| \to 0$$

in probability, where P_0 denotes the probability measure corresponding to the case where the statistic $\hat{\rho}_{LS,T}$ is computed from a stretch of size n from the unit root process obtained by integrating the process $\{\beta + U_t\}$.

5. POWER CONSIDERATIONS

In this section the power properties of the RBB testing procedure are investigated and compared to those of an alternative block bootstrap scheme based on differences of the observed series. We restrict in the following our considerations to statistics based on the lag-1 autocorrelation and the corresponding ordinary least squares estimator $\hat{\rho}_{LS}$.

5.1. *Global and Local Power Properties of the RBB Procedure*

Consider first the case where the observed process is stationary and the parameter of interest ρ is the lag-1 autocorrelation with a (fixed) value in the interval $(-1, 1]$. For $\alpha \in (0, 1)$ let C^*_α be the α-quantile of the distribution of $l(\hat{\rho}^*_{LS} - 1)$, i.e., $C^*_\alpha = \inf\{C : P^*(l(\hat{\rho}^*_{LS} - 1) \leq C) \geq \alpha\}$; because the discontinuities of P^* vanish asymptotically, we may write $P^*(l(\hat{\rho}^*_{LS} - 1) \leq C^*_\alpha) \simeq \alpha$. Theorem 4.1 implies that as $n \to \infty$, $C^*_\alpha \to C_\alpha$ in probability, where C_α is the α-quantile of the asymptotic null distribution of $n(\hat{\rho}_{LS} - 1)$, i.e., the α-quantile of the distribution of the random variable

$$V_\infty := \big(W^2(1) - \sigma^2_U/\sigma^2\big) \Big/ \left(2 \int_0^1 W(r)^2\, dr\right), \tag{5.1}$$

where $\sigma^2_U = \operatorname{var}(U_t)$, $\sigma^2 = 2\pi f_U(0)$, and f_U denotes the spectral density of $\{U_t\}$. Recall that U_t is defined by $U_t = X_t - \rho X_{t-1}$ and that U_t is the differenced process only if the null hypothesis is true.

Let $\tau^2 = (c_{2,2} - 2\rho c_{1,2} + \rho^2 c_{1,1})/\gamma^2(0)$, where $c_{i+1,j+1} = \sum_{l=-\infty}^{\infty} \operatorname{cov}(X_0 X_i, X_l X_{l+j})$ and $\gamma(0) = \operatorname{var}(X_t)$. Let further $F_n(\cdot) = P(\sqrt{n}(\hat{\rho}_{LS} - \rho)/\tau \leq \cdot)$ and note that if the underlying process $\{X_t\}$ is stationary and satisfies Condition A or B, then $\sup_{x\in\mathbb{R}}|F_n(x)-\Phi(x)| \to 0$, where $\Phi(\cdot)$ denotes the distribution function of the standard normal; cf. Romano and Thombs (1996). Let $\beta_{RBB,n}(\rho;\alpha)$ be the power function of the α-level RBB test, i.e.,

$$\beta_{RBB,n}(\rho;\alpha) = P\big(n(\hat\rho_{LS}-1) \le C^*_\alpha \mid \rho \in (-1,1]\big).$$

We now have the following corollary, which shows consistency of the RBB based unit root test.

COROLLARY 5.1: *Under the assumptions of Theorem* 4.1, *we have*

$$\beta_{RBB,n}(\rho;\alpha) \xrightarrow{P} 1 \quad \textit{for all} \quad \rho \in (-1,1) \quad \textit{and}$$
$$\beta_{RRB,n}(1;\alpha) \xrightarrow{P} \alpha \quad \textit{as} \quad n\to\infty.$$

The proof of the corollary is immediate considering that

$$\beta_{RBB,n}(\rho;\alpha) = F_n\bigg(\frac{C^*_\alpha}{\sqrt{n}\tau} - \frac{\sqrt{n}(\rho-1)}{\tau}\bigg) = F_n\big(-\Omega_P\big(n^{-1/2}\big)+\Omega\big(n^{1/2}\big)\big) \xrightarrow{P} \Phi(+\infty)=1; \tag{5.2}$$

in the above, Ω is used to denote exact order of magnitude. In other words, for positive quantities A_n and B_n, $A_n=\Omega(B_n)$ if $A_n=O(B_n)$ and $B_n=O(A_n)$; similarly, we will write $A_n=\Omega_P(B_n)$ if $A_n=O_P(B_n)$ and $B_n=O_P(A_n)$.

Consider next the asymptotic power behavior of the RRB based unit root test for sequences of local alternatives converging to the null at the rate n^{-1}. To be more specific, assume that the underlying process satisfies the following condition.

CONDITION B′: *$X_t=\rho_n X_{t-1}+U_t$, $t=1,2,\ldots$, where $\rho_n=1+c/n$, $c<0$, and the stationary process $\{U_t\}$ satisfies: $E(U_t)=0$, $E|U_t|^\nu<\infty$ for some $\nu>2$, $f_U(0)>0$ and $\sum_{k=1}^\infty \alpha(k)^{1-2/\nu}<\infty$, where $\alpha(\cdot)$ denotes the strong mixing coefficient of $\{U_t\}$.*

Taking into account the asymptotic theory of the regression statistic $n(\hat\rho_{LS}-1)$ for near integrated processes (cf. Phillips (1987b)), the following theorem about the asymptotic local power behavior of the RBB based test can be established.

THEOREM 5.1: *Let $\{X_t\}$ satisfy Condition B′. If $b\to\infty$ as $n\to\infty$ such that $b/\sqrt{n}\to 0$, then*

$$\beta_{RBB,n}(\rho_n;\alpha) \to P(J\le C_\alpha - c),$$

in probability, where C_α is the α quantile of the distribution of $(W^2(1)-\sigma_U^2/\sigma^2)/(2\int_0^1 W^2(r)\,dr)$, J is a random variable having distribution $(\int_0^1 J_c(r)\,dW(r)+(1-\sigma_U^2/\sigma^2)/2)/(\int_0^1 J_c(r)^2\,dr)$ and $J_c(r)=\int_0^1 \exp\{(r-s)c\}\,dW(s)$ is the Ornstein-Uhlenbeck process generated by the stochastic differential equation $dJ_c(r)=cJ_c(r)\,dr+dW(r)$ with initial condition $J_c(0)=0$.

Thus, the above theorem confirms the good asymptotic performance of the RBB; even for alternatives contiguous to the null hypothesis, the RBB manages to achieve nontrivial power. Notice that like the case of fixed alternatives, also in the case of local alternatives considered here, a test based on $n(\hat{\rho}_{LS}-1)$ has an asymptotic null distribution that depends on nuisance parameters. A modification of this statistic leading to a distribution under the null that is free of nuisance parameters has been proposed by Phillips and Perron (1988). We stress here the fact that our bootstrap procedure can be successfully applied to approximate the null distribution of such a modified test statistic too.

5.2. *Comparison with a Block Bootstrap Based on Differences*

An alternative approach to implement the block bootstrap for unit root testing is to apply the block resampling scheme to the series of centered differences:

$$L_t = (X_t - X_{t-1}) - \frac{1}{n-1}\sum_{j=2}^{n}(X_j - X_{j-1}) \qquad (t = 2, 3, \ldots, n), \tag{5.3}$$

instead of applying it to the series of the centered residuals $\widehat{U}_t$ given in (2.1). The motivation for such an approach is to *impose* the null hypothesis of unit root integration to the observed series prior to applying the bootstrap. As mentioned in the Introduction, under more restrictive (e.g., parametric) assumptions on the underlying process, such a bootstrap approach based on differences has been widely used in the literature; cf. among others Nankervis and Savin (1996), Park (2000), Chang and Park (2001), and Psaradakis (2001).

A difference-based block bootstrap (DBB) procedure for unit root testing will generate pseudo series, say, $X_1^+, X_2^+, \ldots, X_l^+$, by following exactly the same steps as those of the RBB algorithm described in Section 2 with the only difference that the series of centered residuals $\widehat{U}_t$ used in the first step will be replaced by the series L_t of centered differences given in (5.3). Now let $\hat{\rho}_{LS}^+ = \sum_{t=2}^{l} X_t^+ X_{t-1}^+ / \sum_{t=2}^{l} X_{t-1}^{+2}$ be the least squares estimator of ρ; $\hat{\rho}_{LS}^+$ is the analogue of $\hat{\rho}_{LS}$ in the DBB world. The distribution of the DBB statistic $l(\hat{\rho}_{LS}^+ - 1)$ is then used to approximate the distribution of the statistic $n(\hat{\rho}_{LS}-1)$ under the null hypothesis of unit root integration.

To compare the above difference-based block-bootstrap procedure with our residual-based proposal, the following theorem is important. It deals with the asymptotic properties of the DBB estimator $\hat{\rho}_{LS}^+$.

THEOREM 5.2: *Assume that the process $\{X_t\}$ satisfies Condition A or B with $\beta = 0$. Let $b \to \infty$ as $n \to \infty$ such that $b/\sqrt{n} \to 0$.*

(i) *If $\rho = 1$, then*

$$\mathscr{L}\big(l(\hat{\rho}_{LS}^+ - 1)|X_1, X_2, \ldots, X_n\big) \Rightarrow \big(W^2(1) - \sigma_U^2/\sigma^2\big)\bigg/\bigg(2\int_0^1 W(r)^2\,dr\bigg),$$

in probability.

(ii) *If* $\rho \in (-1, 1)$ *then*

$$\mathscr{L}\big(k(\hat{\rho}_{LS}^{+} - 1) | X_1, X_2, \dots, X_n\big) \Rightarrow -\big(1 - \operatorname{corr}(X_t X_{t+1})\big) \Big/ \left(2 \int_0^1 W(r)^2 \, dr\right) < 0,$$

in probability.

Recall that the limiting distribution in part (i) of the above theorem is indeed the limiting distribution of $n(\hat{\rho}_{LS} - 1)$ if $\{X_t\}$ is unit root integrated. However, part (ii) of the same theorem shows that, under the alternative, the DBB estimator $\hat{\rho}_{LS}^{+}$ behaves totally differently as compared to the RBB estimator $\hat{\rho}_{LS}^{*}$. In particular, the DBB estimator converges to a different limit than the RBB estimator; in addition, this convergence occurs at the *slower* rate k as opposed to l for the RBB. As a careful reading of the proof of this theorem shows, the reason for the slow k-convergence of the DBB estimator $\hat{\rho}_{LS}^{+}$ under the alternative lies in the fact that, if the observed process is stationary, then integrating random blocks of differenced observations effects a cancelation of the differences found within each block. The result of this cancelation is that the partial sum process based on the increments $X_t^{+} - X_{t-1}^{+}$ of the DBB pseudoseries $X_1^{+}, X_2^{+}, \dots, X_l^{+}$ behaves essentially like a partial sum process of only k independent increments given by $X_{i_m+b} - X_{i_m}$, $m = 0, 1, \dots, k-1$; see Lemma 8.5.

There are some important implications of the above theorem for the behavior of the block bootstrap unit root testing based on differences. In particular, under the alternative, i.e., if the underlying process is stationary, then using the DBB bootstrap statistic $l(\hat{\rho}_{LS}^{+} - 1)$ to approximate the null distribution of $n(\hat{\rho}_{LS} - 1)$ fails. This is so because by part (ii) of Theorem 5.2 and the slower convergence rate of $\hat{\rho}_{LS}^{+}$, we get for every $\eta \in (0, 1)$ that

$$\lim_{n \to \infty} P\big(l(\hat{\rho}_{LS}^{+} - 1) \le -b^{\eta}\big) = 1,$$

since $lb^{-1}(\hat{\rho}_{LS}^{+} - 1)$ converges (in probability) to a negative random limit. Since $-b^{\eta} \to -\infty$, it follows that the DBB statistic $l(\hat{\rho}_{LS}^{+} - 1)$ diverges to $-\infty$ under the alternative!

Apart from this undesirable behavior, it might be interesting to investigate how the above failure of the DBB procedure affects the power properties of the corresponding unit root test. For this let $\beta_{DBB,n}(\rho; \alpha)$ be the power function of the DBB based test, i.e.,

$$\beta_{DBB,n}(\rho; \alpha) = P\big(n(\hat{\rho}_{LS} - 1) \le C_{\alpha}^{+} \mid \rho \in (-1, 1]\big),$$

where C_{α}^{+} denotes the α-quantile of the distribution of the DBB statistic $l(\hat{\rho}_{LS}^{+} - 1)$. Note that by part (ii) of Theorem 5.2 we have that C_{α}^{+} is negative and that $-C_{\alpha}^{+} = \Omega_P(b)$. Thus for $\rho \in (-1, 1)$,

$$\beta_{DBB,n}(\rho; \alpha) = F_n\left(\frac{C_{\alpha}^{+}}{\sqrt{n}\tau} - \frac{\sqrt{n}(\rho - 1)}{\tau}\right) = F_n(-\Omega_P(bn^{-1/2}) + \Omega(n^{1/2})). \tag{5.4}$$

Now, since $\Omega(n^{1/2})$ is the dominant term, it still follows that $\beta_{DBB,n}(\rho;\alpha) \xrightarrow{P} 1$ as $n \to \infty$ provided $b/n \to 0$ and $\rho \in (-1,1)$. However, a comparison of (5.4) with (5.2) shows that the difference between the power functions of the two block bootstrap tests (RBB and DBB) is due to the terms $C_\alpha^+/\sqrt{n}$ and $C_\alpha^*/\sqrt{n}$ respectively. Notice that both of those terms are negative while $-\sqrt{n}(\rho-1) > 0$ under the alternative. Furthermore, as $n \to \infty$, $-\sqrt{n}(\rho-1) \to +\infty$, $C_\alpha^+/\sqrt{n} \uparrow 0$ and $C_\alpha^*/\sqrt{n} \uparrow 0$. These facts imply that the slower the negative terms $C_\alpha^*/\sqrt{n}$ and $C_\alpha^+/\sqrt{n}$ approach 0, the lower the power of the corresponding test will be. Now, because $-C_\alpha^+/\sqrt{n} = \Omega_P(bn^{-1/2})$ while $-C_\alpha^*/\sqrt{n} = \Omega_P(n^{-1/2})$, the term $C_\alpha^*/\sqrt{n}$ converges to zero faster, confirming that the RBB test is asymptotically *more powerful* than the DBB test. The loss of power is essentially due to the slower convergence rate of the DBB statistic $\hat{\rho}_{LS}^+ - 1$. The previous discussion is summarized in the following corollary.

COROLLARY 5.2: *Under the assumptions of Theorem* 4.1, *and for all* $\rho \in (-1,1)$, *we have*

$$\beta_{DBB,n}(\rho;\alpha) \le \beta_{RBB,n}(\rho;\alpha)$$

with probability tending to one as $n \to \infty$.

We mention here that apart from the above differences in the convergence rate of the two test statistics considered, and as expressions (5.2) and (5.4) show, we expect the power of the DBB test to decrease when the block size b increases. This is because large values of b inflate the term C_α^+ which reduces the power of the DBB test.

In concluding this comparison we deal with the question of whether the DBB based test has power against sequences of $1/n$ local alternatives satisfying Condition B′. The next theorem states that regarding this class of local alternatives, the DBB based test has the same asymptotic behavior as the RBB based test. The reason for this is that in contrast to the case of fixed alternatives, $1/n$ local stationarity recovers $\sqrt{l}$-convergence of the bootstrap partial sum process based on differences; cf. Lemma 8.7.

THEOREM 5.3: *Let* $\{X_t\}$ *satisfy Condition B′. If* $b \to \infty$ *as* $n \to \infty$ *such that* $b/\sqrt{n} \to 0$, *then*

$$\beta_{DBB,n}(\rho_n;\alpha) \to P(J \le C_\alpha - c),$$

in probability, where C_α *and* J *are defined as in Theorem* 5.1.

The theoretical findings of this section are illustrated in Section 6.2 by means of some numerical examples.

6. IMPLEMENTATION ISSUES AND SMALL SAMPLE PERFORMANCE

6.1. *Choosing the Block Size in Practice*

Recall that our asymptotic results hold true for any block size b satisfying

$$b \to \infty \quad \text{but} \quad b/\sqrt{n} \to 0; \tag{6.1}$$

here of course the understanding is that b is a function of n, i.e., $b = b(n)$. Nevertheless, there are many choices of b that satisfy (6.1), and it is natural to ask whether there is an 'optimal' one. This is a familiar problem with all blocking methods, the practical implementation and performance of which is well known to be quite influenced by the actual block size used. The Residual-based Block Bootstrap is no exception, and it would be desirable to build a methodology towards 'optimal' block size choice.

To talk about an 'optimal' block size choice it is required to set a *criterion* that is to be optimized. In the usual application of block resampling or subsampling methods to stationary data, the criteria most often used are: (a) accuracy (e.g., mean squared error) in variance estimation; (b) accuracy in estimation of a distribution function; and (c) accuracy in achieving the nominal coverage of a confidence interval.

Typically, a higher-order expansion is developed involving one of the above accuracy measures. The expansion is then optimized with respect to b yielding an expression of the type

$$b_{opt} = Cn^{\gamma} \tag{6.2}$$

for the optimal block size. Ideally, γ is known but usually the constant C depends on unknown characteristics of the probability structure associated with the data series.

To fix ideas, consider the quite relevant example of performing a block bootstrap on the residuals $\{U_t\}$ with the objective of (accurate) estimation of the variance of the sample mean $\overline{U}_n = n^{-1}\sum_{t=1}^{n} U_t$. Recall that, under regularity conditions, $\mathrm{var}(\sqrt{n}\overline{U}_n) \to \sigma^2 := 2\pi f_U(0)$, where f_U is the spectral density of the series $\{U_t\}$. Thus, estimation of $\mathrm{var}(\sqrt{n}\overline{U}_n)$ is tantamount to estimation of $f_U(0)$. To achieve most accurate (in terms of mean squared error) estimation of $\mathrm{var}(\sqrt{n}\overline{U}_n)$ via the block bootstrap, it is well known that we must take $\gamma = 1/3$; see e.g. Künsch (1989). The constant C is however unknown as it depends—among other things—on the unknown function f and its smoothness near the origin.

In such a case when an expression of the type (6.2) is available with γ known but C unknown, there are two general approaches in the literature of block/bandwidth choice:

(i) *Plug-in methods*. Here the dependence of C on a few unknown features is exploited; the features (e.g., the spectral density function and its first one or two derivatives) are explicitly estimated, yielding an estimator $\widehat{C}$ of C to be used in (6.2). For the method to work well, accurate estimation of C is required. In the particular problem of variance estimation for the sample mean this approach was considered in Politis and White (2000) where estimation of C was based on fast-converging, infinite-order smoothing kernels; see Politis (2001) for more details.

(ii) *Cross-validation*. Although cross-validation is a general nonparametric technique, its particular application to choosing a block size was suggested by Hall, Horowitz, and Jing (1995); this is a very useful and easily implementable methodology that employs the notion of subsampling to effectively by-pass explicit estimation of C, while still yielding an estimator of b_{opt}.

In the problem at hand, a natural optimization criterion is to improve the speed of convergence of our RBB approximations. For example, part (i) of Theorem 4.1 can be restated as:

$$\text{(6.3)} \qquad \begin{aligned} &P^*\big(l(\hat{\rho}^*_{LS}-1)\le x|X_1,X_2,\dots,X_n\big)-P_0\big(n(\hat{\rho}_{LS}-1)\le x\big)\\ &\qquad = O_P(\delta_{b,n}), \quad \text{with} \quad \delta_{b,n}=o(1),\end{aligned}$$

as $b\to\infty$ but $b/\sqrt{n}\to 0$; here $\delta_{b,n}$ is a deterministic quantity that does not depend on x but does depend on b,n and the probabilistic structure of $\{X_t\}$. Although intractable at the moment, it is hoped that with further future work the quantity $\delta_{b,n}$ will be identified; minimization of $\delta_{b,n}$ with respect to the design parameter b would then yield an expression of the type (6.2), and consequently both aforementioned methods, plug-in and cross-validation, may be available for practical block size choice.

Until the quantity $\delta_{b,n}$ is pinpointed and a recommendation of the type (6.2) becomes available, the following heuristic ideas may be helpful. Recall that both distributions appearing in equation (6.3) converge to the distribution of the random variable V_∞ defined in (5.1). An immediate implication of part (i) of Theorem 4.1 is

$$\text{(6.4)} \qquad P^*(l(\hat{\rho}^*_{LS}-1)\le x|X_1,X_2,\dots,X_n)-P_W(V_\infty\le x)=o_P(1),$$

where P_W is the probability law associated with the Wiener process W. Consequently, we may try to choose b with the objective of minimizing the right-hand-side of (6.4) instead. To do this, recall that the distribution $P_W(V_\infty\le x)$ was seen to depend on two parameters, namely $\sigma_U^2=\mathrm{var}(U_t)$ and $\sigma^2=2\pi f_U(0)$. Therefore, for the approximation (6.4) to hold it must be true that the RBB procedure achieves an implicit estimaton of those two parameters. Recall, however, that estimating σ^2 is tantamount to estimating the variance of the sample mean $\overline{U}_n$. Thus, taking into account the RBB step of block bootstrapping the estimated residuals, it is intuitive that this implicit estimation of σ^2 by the RBB is achieved by a mechanism that is quite close to the block bootstrap estimator of the variance of $\overline{U}_n$; it follows that a block size given by equation (6.2) with $\gamma=1/3$ may be a good choice, in which case both aforementioned methods, plug-in and cross-validation, are directly applicable, as the problem has been effectively reduced to optimizing the block bootstrap variance estimator for the series $\{\widehat{U}_t\}$.

Nevertheless, the above recommendation is a heuristic one. In addition, our heuristic has the simpler objective of minimizing the right-hand side of (6.4) instead of the right-hand side of (6.3), which is more challenging. Last, but not least, note that there exist alternative philosophies for block size choice; important such examples are the *calibration method* and the *minimum volatility method*, both of which are described in detail in Politis, Romano, and Wolf (1999, Ch. 9.4) in connection with subsampling. In particular, the minimum volatility method seems to be easily applicable/extendible to the RBB set-up; it amounts to computing critical values for the RBB test using a range of different block sizes, and choosing b in a region where those critical values exhibit smallest volatility—see Politis, Romano, and Wolf (1999, pp. 201–202). More work is required in order to give analytical and/or empirical substantiation to the aforementioned preliminary block size choice ideas.

6.2. *Numerical Examples*

A small simulation study was conducted to evaluate the finite-sample performance of the RBB bootstrap testing procedure and to compare its performance with that of some other bootstrap procedures. For this purpose the simple ARMA(1,1) model,

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1},$$

was used to generate the observed series $\{X_t\}$ based on the i.i.d. Gaussian series $\{Z_t\} \sim N(0,1)$. The case $\phi = 1$ is the unit root case, whereas $\phi = 0.85$ corresponds to a stationary series $\{X_t\}$. Regarding the MA parameter θ, the values $-0.8, 0, 0.5$, and 0.8 were chosen; $\theta = 0.5$ and $\theta = 0.8$ correspond to a positive dependence; $\theta = 0$ corresponds to either a random walk with i.i.d. errors, or a stationary AR(1) model (according to whether $\phi = 1$ or $\phi = 0.85$). Finally, the case of negative correlation $\theta = -0.8$ has attracted some attention in the literature because the moving average polynomial has a root close to unity which, in combination with the unity autoregressive root, yields series that can easily be mistaken for i.i.d., especially when the sample size is not too big.

The simulations were performed by generating a number of $M = 2000$ true $\{X_t\}$ series each of length $n+100$ where the first 100 observations were discarded; we chose $n = 50$ and $n = 100$. From each generated data series the RBB bootstrap was called to perform an α-level test of the unit root hypothesis H_0. The test statistics used were based on the LS estimator in equation (2.4), i.e., the statistic $n(\hat{\rho}_{LS,C} - 1)$ and on the augmented DF estimator in equation (2.6), i.e., the statistic $(n-p)(\hat{\rho}_{DF,C} - 1)$. We denote the corresponding RBB-based tests by $Z^*(\hat{\rho}_{LS,C})$ and $ADF^*(\hat{\rho}_{DF,C})$ respectively. Note that the RBB bootstrap procedure was conducted by generating $B = 1000$ bootstrap series (for each true series) in order to perform the required Monte Carlo approximations.

The empirical performance of the RBB test—denoted by $Z^*(\hat{\rho}_{LS,C})$ in the tables—is compared with that of some other well-known tests. In particular, we consider the unit-root test proposed by Phillips and Perron (1988) that is obtained by correcting the statistic $n(\hat{\rho}_{LS,C} - 1)$ for nuisance parameters; the

Phillips-Perron test is denoted in the following by $Z(\hat{\rho}_{LS,C})$. Furthermore, we compare the performance of the RBB test with that of the DBB procedure based on differences, which was discussed in Section 5.2; we denote the DBB test by $Z^{+}(\hat{\rho}_{LS,C})$. Similarly, the $ADF^{*}(\hat{\rho}_{DF,C})$ test is compared with the Dickey-Fuller regression t test for a unit root in the autoregression (2.6); the latter is denoted in the following by $ADFt(\hat{\rho}_{DF,C})$. Note that Said and Dickey (1984) do not suggest a statistic based on the coefficient $\hat{\rho}_{DF,C}$ since the limit distribution of $(n-p)(\hat{\rho}_{DF,C}-1)$ depends on nuisance parameters. Thus, there is no analogue of our $ADF^{*}(\hat{\rho}_{DF,C})$ in Said and Dickey (1984). The $ADF^{*}(\hat{\rho}_{DF,C})$ test is also compared with a test based on a 'sieve bootstrap' approach (cf. Chang and Park (2001), Psaradakis (2001)) which is denoted by $ADF^{SIEVE}(\hat{\rho}_{DF,C})$.

Finally, we compare the RBB test with a test based on subsampling the Dickey-Fuller statistic; this test will be denoted by $ADF^{SUB}(\hat{\rho}_{DF,C})$. The validity of subsampling in this framework was recently shown in Romano and Wolf (2001) under an $AR(p)$ assumption; see also Chapter 12 in Politis, Romano, and Wolf (1999) where more details on this computer-intensive methodology may be found. Note that typically subsampling has the objective of forming confidence intervals for parameters of interest; nevertheless, the subsampling confidence intervals can be immediately inverted to yield tests of a point hypothesis such as our H_0.

Table I and Table II report the empirical rejection probabilities of the different unit root tests discussed above with nominal level $\alpha = 0.05$ and under different settings of the parameters ϕ and θ, different sample sizes n, different choices of the block size b, the autoregressive order p, and the subsampling block size B.

TABLE I

RESULTS OF MONTE CARLO EXPERIMENTS FOR THE LAG-1 AUTOCORRELATION[a]

			$\phi = 1.0$				$\phi = 0.85$		
			$\theta = 0.0$	$\theta = 0.5$	$\theta = 0.8$	$\theta = -0.8$	$\theta = 0.0$	$\theta = 0.5$	$\theta = 0.8$
$n = 50$	$Z(\hat{\rho}_{LS,C})$	$l = 3$	0.052	0.013	0.010	0.989	0.345	0.112	0.090
		$l = 5$	0.048	0.007	0.005	0.994	0.314	0.061	0.044
	$Z^{*}(\hat{\rho}_{LS,C})$	$b = 3$	0.056	0.022	0.019	0.982	0.355	0.144	0.120
		$b = 5$	0.065	0.018	0.018	0.989	0.356	0.145	0.124
	$Z^{+}(\hat{\rho}_{LS,C})$	$b = 3$	0.030	0.014	0.013	0.907	0.200	0.109	0.099
		$b = 5$	0.015	0.012	0.012	0.742	0.125	0.076	0.071
$n = 100$	$Z(\hat{\rho}_{LS,C})$	$l = 4$	0.049	0.035	0.032	0.991	0.803	0.497	0.457
		$l = 6$	0.048	0.023	0.021	0.994	0.813	0.442	0.391
	$Z^{*}(\hat{\rho}_{LS,C})$	$b = 4$	0.067	0.041	0.040	0.981	0.808	0.515	0.469
		$b = 6$	0.068	0.036	0.035	0.987	0.817	0.523	0.477
	$Z^{+}(\hat{\rho}_{LS,C})$	$b = 4$	0.047	0.036	0.035	0.893	0.613	0.442	0.418
		$b = 6$	0.032	0.026	0.029	0.784	0.532	0.402	0.378

[a] Empirical rejection probabilities of unit root tests with nominal level $\alpha = 0.05$ under different settings of the ARMA parameters ϕ and θ. The test statistic used here is $n(\hat{\rho}_{LS,C}-1)$ with asymptotic $(Z(\hat{\rho}_{LS,C}))$, RBB-based $(Z^{*}(\hat{\rho}_{LS,C}))$ and DBB-based $(Z^{+}(\hat{\rho}_{LS,C}))$ critical values; l denotes the truncation parameter used in the test statistic proposed by Phillips and Perron (1988) and b the bootstrap block size.

TABLE II

RESULTS OF MONTE CARLO EXPERIMENTS FOR THE DICKEY-FULLER TYPE STATISTIC[a]

				$\phi=1.0$				$\phi=0.85$		
				$\theta=0.0$	$\theta=0.5$	$\theta=0.8$	$\theta=-0.8$	$\theta=0.0$	$\theta=0.5$	$\theta=0.8$
$n=50$	$ADFt(\hat{\rho}_{DF,C})$		$p=2$	0.059	0.055	0.045	0.504	0.182	0.128	0.085
			$p=4$	0.072	0.077	0.064	0.228	0.154	0.137	0.096
	$ADF^*(\hat{\rho}_{DF,C})$	$b=4$,	$p=2$	0.068	0.038	0.026	0.824	0.343	0.207	0.158
			$p=4$	0.079	0.058	0.036	0.698	0.352	0.248	0.186
		$b=5$,	$p=2$	0.071	0.039	0.029	0.792	0.345	0.213	0.161
			$p=4$	0.078	0.051	0.037	0.651	0.358	0.245	0.185
	$ADF^{SIEVE}(\hat{\rho}_{DF,C})$		$p=2$	0.041	0.026	0.015	0.568	0.187	0.112	0.071
			$p=4$	0.030	0.028	0.019	0.234	0.142	0.113	0.077
	$ADF^{SUB}(\hat{\rho}_{DF,C})$	$B=4$,	$p=2$	0.117	0.076	0.077	0.938	0.435	0.357	0.373
		$B=5$,	$p=2$	0.152	0.113	0.110	0.954	0.526	0.446	0.468
		$B=10$,	$p=4$	0.078	0.056	0.059	0.528	0.297	0.235	0.217
$n=100$	$ADFt(\hat{\rho}_{DF,C})$		$p=2$	0.057	0.045	0.033	0.672	0.524	0.386	0.268
			$p=4$	0.061	0.048	0.042	0.331	0.416	0.374	0.292
			$p=6$	0.064	0.060	0.053	0.174	0.364	0.340	0.270
	$ADF^*(\hat{\rho}_{DF,C})$	$b=6$,	$p=2$	0.071	0.045	0.040	0.859	0.760	0.595	0.518
			$p=4$	0.073	0.052	0.036	0.704	0.731	0.611	0.505
			$p=6$	0.079	0.057	0.044	0.616	0.712	0.592	0.502
		$b=8$,	$p=2$	0.068	0.038	0.032	0.861	0.782	0.576	0.495
			$p=4$	0.069	0.053	0.038	0.714	0.740	0.584	0.486
			$p=6$	0.071	0.055	0.042	0.610	0.710	0.591	0.488
	$ADF^{SIEVE}(\hat{\rho}_{DF,C})$		$p=2$	0.035	0.019	0.010	0.724	0.611	0.458	0.336
			$p=4$	0.046	0.035	0.027	0.360	0.487	0.426	0.328
			$p=6$	0.039	0.036	0.030	0.197	0.364	0.331	0.275
	$ADF^{SUB}(\hat{\rho}_{DF,C})$	$B=6$,	$p=2$	0.128	0.096	0.101	0.961	0.893	0.841	0.843
			$p=4$	0.008	0.003	0.003	0.473	0.221	0.109	0.105
		$B=8$,	$p=2$	0.132	0.106	0.106	0.951	0.885	0.848	0.851
			$p=4$	0.037	0.022	0.017	0.601	0.499	0.337	0.339
		$B=10$,	$p=4$	0.057	0.036	0.029	0.606	0.574	0.484	0.461
		$B=20$,	$p=6$	0.073	0.062	0.062	0.345	0.565	0.479	0.488

[a] Empirical rejection probabilities of unit root tests with nominal level $\alpha=0.05$ under different settings of the ARMA parameters ϕ and θ. The test statistic used here is the augmented Dickey Fuller t-test with asymptotic critical values, the statistic $(n-p)(\hat{\rho}_{DF,C}-1)$ with RBB-based critical values ($ADF^*(\hat{\rho}_{DF,C})$) and with sieve bootstrap based critical values ($ADF^{SIEVE}(\hat{\rho}_{DF,C})$). $ADF^{SUB}(\hat{\rho}_{DF,C})$ denotes the test based on inverting the appropriate one-sided subsampling confidence interval. Finally, b denotes the block bootstrap size, B the subsampling block size, and p the order of the autoregression fitted.

Although the simulation study is limited, the results suggest several interesting conclusions. First the empirical sizes of the RBB-based tests are close to the nominal level of 5% with the only exception being the case where $\theta=-0.8$. For this particular case of strong negative correlation, it seems that using a nonparametric block bootstrap approach does not solve the poor size problems of the unit root tests considered. Note that similar problems occur more or less for all other alternative bootstrap methods considered here, with the sieve bootstrap procedure being somewhat better although far from satisfactory.

Regarding the DF-type test, an additional aspect to the above appears in the case $\theta = -0.8$. It is well known that the augmented DF t-test, i.e., the $ADFt(\hat{\rho}_{LS,C})$ statistic, also suffers size distortions for $\theta = -0.8$, which are attenuated as the lag length of the autoregression fitted increases; cf. Phillips and Perron (1988). For instance, the empirical rejection rate of the null hypothesis for $n = 50$, $\phi = 1.0$, and $\theta = -0.8$ equals 0.802 for $p = 1$, 0.504 for $p = 2$, 0.341 for $p = 3$, and 0.220 for $p = 4$, making evident that the size "correction" for $\theta = -0.8$ is achieved by increasing the lag length p. The price paid for this, however, is a drop in power; cf. also Ng and Perron (1995). Now, since the RBB-based $ADF^*(\hat{\rho}_{DF,C})$ test seems to be less sensitive to the choice of lag length p, the size distortions for $\theta = -0.8$ are only gradually corrected as p increases, leading to the high rejection rates reported in the table.

Nevertheless, the case $\theta = -0.8$ is a well-known problematic situation in which—as discussed above—a practical 'cancelation' of the autoregressive unit root with the moving average 'almost' unit root occurs, yielding series with sample paths closely resembling a white noise. As a matter of fact, many authors argue that in such a case the stationary model obtained after the 'cancelation' may provide a more parsimonious description of the data, and that consequently (false) rejections of H_0 are not necessarily a bad thing; see Campbell and Perron (1991), and Hamilton (1994) for a discussion.

A second conclusion of our simulation study is that using the RBB leads to improvements in terms of power that are in some cases substantial; we refer here to the results in Table I and Table II for $\phi = 0.85$ and $\theta \in \{0.0, 0.5, 0.8\}$. In particular, the power of the $Z^*(\hat{\rho}_{LS,C})$ based test is always bigger than that of the corresponding $Z(\hat{\rho}_{LS,C})$ asymptotic test with more clear improvements in the case of small n and positive error correlation ($\theta = 0.5, \theta = 0.8$). Furthermore, the numerical results confirm our theoretical analysis concerning the power behavior of the block bootstrap test based on differencing. The power of the DBB-based test $Z^+(\hat{\rho}_{LS,C})$ is much lower than that of the RBB-based test $Z^*(\hat{\rho}_{LS,C})$, and in most cases even lower than the power of the test based on asymptotic critical values. Note that the power of the $Z^+(\hat{\rho}_{LS,C})$ test is affected negatively by the block size b, i.e., increasing b leads to a loss of power; see Section 5.2 for an explanation of this behavior.

For the augmented DF test the gain in power from using our RBB procedure is considerable. The power of the $ADF^*(\hat{\rho}_{DF,C})$ test is not only always bigger than that of the $ADFt(\hat{\rho}_{DF,C})$ test, but for $\theta = 0.5$ and $\theta = 0.8$, the power of the RBB-based test is in some cases almost one and a half to two times that of the augmented DF t-test. Furthermore, as was intuitively expected, the power of the RBB-based augmented DF statistic $ADF^*(\hat{\rho}_{DF,C})$ seems to be less sensitive with respect to the choice of the lag length p. This interesting property is due to the particular way we implemented the block bootstrap procedure for the ADF test; see Section 4.2.

According to Table II, even in the case of linear ARMA alternatives, the RBB-based test $ADF^*(\hat{\rho}_{DF,C})$ seems to be more powerful than the sieve bootstrap based test $ADF^{SIEVE}(\hat{\rho}_{DF,C})$. This striking result is mainly due to two reasons: As for the DBB-based test $Z^{+}(\hat{\rho}_{LS,C})$, the $ADF^{SIEVE}(\hat{\rho}_{DF,C})$ test is based on *differencing* the observed series. If the alternative is true, such differencing leads to a considerable loss of power; the rationale is similar to the discussion of the DBB in Section 5.2. A second source of problems for the $ADF^{SIEVE}(\hat{\rho}_{DF,C})$ test is the sensitivity of its power on the value of the autoregressive order p. Since the sieve bootstrap series is generated by integrating a finite order autoregressive process, it shares the same problems regarding this sensitivity as the ordinary ADF test $ADFt(\hat{\rho}_{DF,C})$: Increasing p leads to a 'correction' of the size problems in the problematic case $\theta = -0.8$ but also to a drop of power.

The comparison of the RBB with the subsampling-based test is not so straightforward. From Table II it is apparent that the empirical results related to subsampling are quite sensitive to the choice of the subsampling size B, as well as to the order of the autoregression p. The test based on subsampling has been calculated for several values of B; however, for the sake of brevity, we present here only those values for which this method performs best, as well as those where the subsampling size B is comparable to the block size b of the RBB. The subsampling-based test seems to have difficulties in capturing the correct size of the test, making the need of a calibration quite apparent; see the discussion at the end of Section 6.1. Increasing p does not necessarily solve the problem, especially since p has to be chosen small with respect to the subsampling size B, whereas B must be chosen small with respect to n.

Nevertheless, we can compare the power of two tests only as long as the two tests have the same (or at least similar) size; a comparison of power between tests of different sizes is meaningless. Note that the high power values reported in Table II for $ADF^{SUB}(\hat{\rho}_{DF,C})$ in the cases where $n = 50$ and $B = 4$ or 5, or in the cases where $n = 100, B = 6$, and $p = 2$, or $B = 8$ and $p = 2$, are seriously inflated because of the size problems of the subsampling test in these particular cases. Thus, we must focus on the cases where the subsampling method achieves a reasonable size behavior, for instance, in the case where $n = 50, B = 10$, and $p = 4$, or in the cases where $n = 100, B = 8$ or 10, and $p = 4$, or $n = 100, B = 20$, and $p = 6$; in all those cases, the RBB-based test appears to be more powerful.

7. CONCLUSIONS

In the paper at hand we have proposed a block resampling procedure that generates unit root integrated pseudoseries that retain the weak dependence structure of the observed series. The procedure is based on very weak assumptions on the dependence structure of the stationary process driving the random walk and, as a consequence, it can be successfully applied to capture the distribution of many unit root test statistics commonly used in econometrics. Although we have restricted our consideration to two such popular test statistics, the theory developed is general enough in that it allows the application of the new resampling methodology to other test statistics too. In fact our procedure can be applied to any test statistic of the unit root hypothesis provided the parameter ρ and the estimator $\tilde{\rho}_n$ used to calculate the stationary residuals satisfy conditions (2.2) and (2.3).

Finite sample performance of the method was examined through some simulations and some comparisons to other bootstrap approaches have been given. The numerical findings are very encouraging. Although applying the RBB bootstrap method does not solve the size distortion problems of the unit root tests in the case of strong negative correlation, the numerical results show that a considerable improvement in terms of power can be achieved by using the RBB method. Such gains in power are quite substantial for the augmented DF statistic in particular.

Our theoretical results, as well as the limited empirical evidence presented in this paper, largely support the conclusion that the RBB testing procedure is a useful alternative to asymptotic distributions commonly used in the econometric analysis of nonstationary time series.

8. AUXILIARY RESULTS AND PROOFS

PROOF OF THEOREM 3.1: First note that for our asymptotic results we may, without loss of generality, assume that $X_0 = 0$. Now for $0 \leq r \leq 1$, and by construction of the RBB series, we have

$$S_l^*(r) = \frac{1}{\sqrt{l}} X_1/\sigma^* + \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^{B} \widehat{U}_{i_m+s}/\sigma^*$$

where $M_r = [([lr]-2)/b]$ and $B = \min\{b, [lr]-mb-1\}$. By Lemma 8.1 we have that $\sigma^{*^2} \to \sigma^2$ in probability where $\sigma^2 = 2\pi f_U(0)$ and f_U denotes the spectral density of $\{U_t\}$. Because of this, the fact that

$$S_l^*(r) = \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \widehat{U}_{i_m+s}/\sigma^* - \frac{1}{\sqrt{l}} \sum_{s=B+1}^{b} \widehat{U}_{i_{M_r}+s}/\sigma^* + O_P(l^{-1/2}) \tag{8.1}$$

and $\sup_{0\leq r\leq 1} |(1/\sqrt{l}) \sum_{s=B+1}^{b} \widehat{U}_{i_{M_r}+s}/\sigma^*| = O_P(k^{-1/2})$ we consider in the following only the first term on the right-hand side of (8.1). We first show that uniformly in r,

$$\left| \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \widehat{U}_{i_m+s} - \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \big(U_{i_m+s} - E^* U_{i_m+s}\big) \right| \to 0 \tag{8.2}$$

in probability. To establish (8.2) verify using the definitions of $\widehat{U}_t$ in (2.1) and of U_t in (1.1) that

$$\begin{aligned} \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \widehat{U}_{i_m+s} &= \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left(U_{i_m+s} - \frac{1}{n-1} \sum_{\tau=2}^{n} U_\tau \right) \\ &\quad - (\tilde{\rho} - \rho) \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left(X_{i_m+s-1} - \frac{1}{n-1} \sum_{\tau=2}^{n} X_{\tau-1} \right). \end{aligned}$$

Now if $\rho \neq 1$, then by condition (2.2) and (2.3) we have $\tilde{\rho} - \rho = o_P(1)$ and, therefore, by the stationarity of $\{X_t\}$ we get that, uniformly in r

$$(\tilde{\rho} - \rho)\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\left(X_{i_m+s-1} - \frac{1}{n-1}\sum_{\tau=2}^{n}X_{\tau-1}\right) \to 0 \tag{8.3}$$

in probability.

To deal with the case where $\rho = 1$ let $\widetilde{X}_t = \sum_{j=1}^{t} U_j$ and recall that by (1.1), $X_t = t\beta + \widetilde{X}_t$. We then have

$$\begin{aligned}
E^*&\left[\sum_{s=1}^{b}\left(X_{i_m+s-1} - \frac{1}{n-1}\sum_{\tau=2}^{n}X_{\tau-1}\right)\right] \\
&= \sum_{s=1}^{b}\left(\frac{1}{n-b}\sum_{t=1}^{n-b}X_{t+s-1} - \frac{1}{n-1}\sum_{\tau=2}^{n}X_{\tau-1}\right) \\
&= \beta\left[\frac{1}{n-b}\sum_{s=1}^{b}\sum_{t=1}^{n-b}(t+s-1) - \frac{b}{n-1}\sum_{\tau=1}^{n-1}\tau\right] \\
&\quad + \sum_{s=1}^{b}\left[\frac{1}{n-b}\sum_{t=1}^{n-b}\widetilde{X}_{t+s-1} - \frac{1}{n-1}\sum_{\tau=2}^{n}\widetilde{X}_{\tau-1}\right] \\
&= T_{1,n} + T_{2,n}
\end{aligned} \tag{8.4}$$

with an obvious notation for $T_{1,n}$ and $T_{2,n}$. It is easily seen by straightforward calculations that $T_{1,n} = O_P(b^2)$ and that

$$\begin{aligned}
T_{2,n} &= \frac{1}{(n-b)(n-1)} \\
&\quad \times \left[\sum_{s=1}^{b}(n-1)\left(\sum_{t=1}^{s-1}\widetilde{X}_t + \sum_{t=n-b+s}^{n-1}\widetilde{X}_t\right) + b(b-1)\sum_{\tau=2}^{n}\widetilde{X}_{\tau-1}\right] \\
&= O_P(b^2 n^{1/2}(n-b)^{-1}).
\end{aligned}$$

Similarly,

$$\begin{aligned}
E^*&\left(\sum_{s=1}^{b}\left(X_{i_m+s-1} - \frac{1}{n-1}\sum_{\tau=2}^{n}X_{\tau-1}\right)\right)^2 \\
&= \frac{1}{n-b}\sum_{t=1}^{n-b}\left[\sum_{s=1}^{b}\left(X_{t+s-1} - \frac{1}{n-1}\sum_{\tau=2}^{n-1}X_{\tau-1}\right)\right]^2 \\
&\leq \frac{2\beta^2}{n-b}\sum_{t=1}^{n-b}\left[\sum_{s=1}^{b}(t+s-1) - \frac{b}{n-1}\sum_{\tau=1}^{n-1}\tau\right]^2
\end{aligned}$$

$$+\frac{2}{n-b}\sum_{t=1}^{n-b}\left[\sum_{s=1}^{b}\left(\widetilde{X}_{t+s-1}-\frac{1}{n-1}\sum_{\tau=2}^{n}\widetilde{X}_{\tau-1}\right)\right]^2$$
$$=C_{1,n}+C_{2,n}.$$

Now, simple algebra shows that $C_{1,n}=O_P(b^2(n-b)^2)$ while as for the term $T_{2,n}$ we get $C_{2,n}=O_P(b^2(n-b))$. Let

$$T_n^*:=\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\left(X_{i_m+s-1}-\frac{1}{n-1}\sum_{\tau=2}^{n}X_{\tau-1}\right).$$

We have that if $\beta=0$, then $E^*(T_n^*)^2=O_P(b^2k)+O_p(kb^3(n-b)^{-1})$, while if $\beta\neq 0$, $E^*(T_n^*)^2=O_P(b(n-b)^2)$. Using (2.2) and (2.3) we conclude that for $\rho=1$

$$(\hat{\rho}-\rho)\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\left(X_{i_m+s-1}-\frac{1}{n-1}\sum_{\tau=2}^{n}X_{\tau-1}\right)=O_{P^*}(b^{1/2}n^{-1/2}). \tag{8.5}$$

From (8.3) and (8.5) it follows that uniformly in r,

$$\left|\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\widehat{U}_{i_m+s}-\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\left(U_{i_m+s}-\frac{1}{n-1}\sum_{\tau=2}^{n}U_\tau\right)\right|\to 0 \tag{8.6}$$

in probability. From (8.6) and because it is straightforward to show that, uniformly in r,

$$l^{-1/2}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\left(\frac{1}{n-1}\sum_{\tau=2}^{n}U_\tau-E^*U^*_{i_m+s}\right)\to 0$$

in probability, we get (8.2).

We next show the convergence of the centered bootstrap partial sum process

$$\frac{1}{\sigma^*}\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\left(U_{i_m+s}-E^*U_{i_m+s}\right), \tag{8.7}$$

to the Brownian motion W on $[0,1]$. For this note first that $[[l_r]/b]=[kr]$ and that, therefore, we can consider instead of (8.7) the asymptotically equivalent statistic

$$\frac{1}{\sigma^*}\frac{1}{\sqrt{l}}\sum_{m=0}^{[kr]}\sum_{s=1}^{b}\left(U_{i_m+s}-E^*U_{i_m+s}\right).$$

The above statistic can be written in the form

$$\frac{1}{\sigma^*}\frac{1}{\sqrt{l}\sigma^*}\sum_{m=0}^{[kr]}\sum_{s=1}^{b}\left(U_{i_m+s}-E^*U_{i_m+s}\right)=\frac{1}{\sqrt{k}}\sum_{m=0}^{[kr]}V_m^*, \tag{8.8}$$

where the random variables $V_m^* = b^{-1/2}\sum_{s=1}^{b}(U_{i_m+s} - E^* U_{i_m+s})$ are independent and have mean zero under the bootstrap distribution. By the definition of σ^{*^2} and because $\mathrm{var}^*(b^{-1/2}\sum_{s=1}^{b} U_{i_m+s}) \to 2\pi f_U(0)$ in probability (cf. Lemma 8.1(ii)), we have that

$$|\mathrm{var}^*(V_m^*) - \sigma^{*^2}| \to 0 \tag{8.9}$$

in probability. Consider now the partial sum

$$V^*(r) = \sum_{m=0}^{[kr]} \widetilde{V}_m^*,$$

where $\{\widetilde{V}_m^*,\ m = 0, 1, 2, \dots, [kr]\}$ forms an array of independent random variables and

$$\widetilde{V}_m^* = \frac{1}{\sqrt{[kr]+1}\sqrt{\mathrm{var}^*(V_m^*)}} V_m^*.$$

Since by definition $\mathrm{var}^*(\widetilde{V}_m^*) = 1/([kr]+1)$ and

$$\begin{aligned}
&\frac{\sum_{m=0}^{[kr]} E^* |\widetilde{V}_m^*|^{2+\kappa}}{\{\mathrm{var}^*(\sum_{m=0}^{[kr]} \widetilde{V}_m^*)\}^{(2+\kappa)/2}} \\
&\quad = \frac{1}{([kr]+1)^{(2+\kappa)/2}} \sum_{m=0}^{[kr]} E^* \left| \frac{V_m^*}{\sqrt{\mathrm{var}^*(V_m^*)}} \right|^{2+\kappa} \\
&\quad = \frac{[kr]+1}{([kr]+1)^{(2+\kappa)/2}} \frac{1}{(\mathrm{var}^*(V_m^*))^{(2+\kappa)/2}} \\
&\qquad \times \frac{1}{n-b} \sum_{t=1}^{n-b} \left(\frac{1}{\sqrt{b}} \sum_{s=1}^{b} \big(U_{t+s} - E^* U_{i_m+s}\big) \right)^{2+\kappa} \\
&\quad = O_P\big(([kr]+1)^{-\kappa/2}\big) \to 0,
\end{aligned}$$

we conclude by Liapunov's Theorem (cf. Serfling (1980)) that

$$\sum_{m=0}^{[kr]} \widetilde{V}_m^* \Rightarrow N(0,1) \tag{8.10}$$

in probability. Relations (8.10) and (8.9) imply then, since

$$\frac{1}{\sqrt{k}\sigma^*} \sum_{m=0}^{[kr]} V_m^* = \sqrt{\frac{\mathrm{var}^*(V_m^*)}{\sigma^{*^2}}} \sqrt{\frac{[kr]+1}{k}} V^*(r), \tag{8.11}$$

that

$$\frac{1}{\sqrt{k}\sigma^*} \sum_{m=0}^{[kr]} V_m^* \Rightarrow W(r) \tag{8.12}$$

in probability. Similarly, if $r_2 \geq r_1$ we get $(V^*(r_1), V^*(r_2) - V^*(r_1)) \Rightarrow (W(r_1), W(r_2) - W(r_1))$ in probability. This implies $(V^*(r_1), V^*(r_2)) \Rightarrow (W(r_1), W(r_2))$ in probability and an easy extension gives $(V^*(r_1), V^*(r_2), \ldots, V^*(r_m)) \Rightarrow (W(r_1), W(r_2), \ldots, W(r_m))$ in probability, for any fixed set of points $r_1 < r_2 < \cdots < r_m$ in [0, 1]. To conclude the proof of the theorem it remains to show tightness of $V^*(r)$. This, however, follows by a version of the functional limit theorem for partial sums of triangular arrays of independent random variables given in Billingsley (1999, p. 147), since $\sum_{m=0}^{[kr]} \widetilde{V}_m^*$ is a sum of independent random variables with mean zero and

$$\max_{0\leq m\leq [kr]} \operatorname{var}^*(\widetilde{V}_m^*) = \frac{1}{[kr]+1} \to 0$$

as $n \to \infty$. Thus $V^* \Rightarrow W$ in probability which, by (8.2), (8.8), and (8.11), implies the assertion of the theorem. *Q.E.D.*

LEMMA 8.1: *Under the assumptions of Theorem* 3.1 *and if* $n \to \infty$, *then*:

(i) $$l^{-1} \sum_{j=1}^{l} U_j^* \to 0,$$

(ii) $$\sigma^{*^2} := \operatorname{var}^*\left[l^{-1/2} \sum_{j=1}^{l} U_j^*\right] \to \sigma^2 := 2\pi f_U(0)$$

and

(iii) $$\sigma_U^{*^2} := l^{-1} \sum_{j=1}^{l} U_j^{*^2} \to \sigma_U^2 := E(U_t^2),$$

in probability.

PROOF: To prove (i) note that by (8.2) we have $l^{-1} \sum_{j=1}^{l} U_j^* = l^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} U_{i_m+s} + o_P(1)$ and that the first term in the right-hand side of the above expression is the sample mean of a block bootstrap series that converges to $E(U_t) = 0$.

Since the proof of (ii) and (iii) are very similar we show only (ii). For this recall that

$$\operatorname{var}^*\left[l^{-1/2} \sum_{j=1}^{l} U_j^*\right] = E^*\left[\left(l^{-1/2} \sum_{j=1}^{l} U_j^*\right)^2\right] - \left(E^*\left[l^{-1/2} \sum_{j=1}^{l} U_j^*\right]\right)^2.$$

Now,

$$\begin{aligned}\frac{1}{\sqrt{l}}\sum_{j=1}^{l}U_j^* = {} & \frac{1}{\sqrt{l}}\sum_{m=0}^{k-1}\sum_{s=1}^{b}\big(U_{i_m+s}-E^*U_{i_m+s}\big)\\ & +\frac{1}{\sqrt{l}}\sum_{m=0}^{k-1}\sum_{s=1}^{b}\left(E^*U_{i_m+s}-\frac{1}{n-1}\sum_{t=2}^{n-1}U_\tau\right)\\ & -(\tilde{\rho}-\rho)\frac{1}{\sqrt{l}}\sum_{m=0}^{k-1}\sum_{s=1}^{b}\left(X_{i_m+s-1}-\frac{1}{n-1}\sum_{t=2}^{n-1}X_{\tau-1}\right)+o(l^{-1/2})\\ = {} & \frac{1}{\sqrt{l}}\sum_{m=0}^{k-1}\sum_{s=1}^{b}\big(U_{i_m+s}-E^*U_{i_m+s}\big)+\widetilde{M}_n^*+\widetilde{T}_n^*+o(l^{-1/2}),\end{aligned}$$

with an obvious notation for $\widetilde{M}_n^*$ and $\widetilde{T}_n^*$. Now using arguments similar to those following the proof of Theorem 3.1, we get $E^*(\widetilde{M}_n^*)\to 0, E^*(\widetilde{M}_n^*)^2\to 0, E^*(\widetilde{T}_n^*)\to 0$, and $E^*(\widetilde{T}_n^*)^2\to 0$ in probability. Hence $E^*(l^{-1/2}\sum_{j=1}^{l}U_j^*)\to 0$ and $E^*(l^{-1/2}\sum_{j=1}^{l}U_j^*)^2=E^*(l^{-1/2}\sum_{m=0}^{k-1}\sum_{s=1}^{b}(U_{i_m+s}-E^*U_{i_m+s}))^2+o_P(1)\to\sigma^2$, in probability, because the first term on the right-hand side of the last equality is nothing other than the variance of $\sqrt{l}$ times the bootstrap sample mean based on a block bootstrap sample $\{U_{i_m+s},\ m=0,1,\dots,k-1 \text{ and } s=1,2,\dots,b\}$; cf. Künsch (1989). *Q.E.D.*

By Theorem 3.1, Lemma 8.1, and the continuous mapping theorem, the following lemma can be easily proved using standard arguments; see Paparoditis and Politis (2000) for details.

LEMMA 8.2: *Let the conditions of Theorem* 4.1 *be satisfied*. *If* $n\to\infty$, *then*:

(i) $$l^{-2}\sum_{t=2}^{l}X_{t-1}^{*^2}\Rightarrow\sigma^2\int_0^1 W^2(r)\,dr,$$

(ii) $$l^{-1}\sum_{t=2}^{l}X_{t-1}^*U_t^*\Rightarrow\frac{1}{2}(\sigma^2W^2(1)-\sigma_U^2),$$

(iii) $$l^{-3/2}\sum_{t=1}^{l}X_{t-1}^*\Rightarrow\sigma\int_0^1 W(r)\,dr,$$

(iv) $$l^{-1/2}\sum_{t=1}^{l}U_t^*\Rightarrow\sigma W(1),$$

in probability, *where joint weak convergence of the above limits also applies*.

PROOF OF THEOREM 4.1: To prove the first assertion of the theorem we use the expression for $l(\hat{\rho}^*_{LS}-1)$ in terms of the bootstrap variables X^*_t and apply Lemma 8.2(i) and (ii) as well as the δ-method; cf. Serfling (1980). The second part of the theorem follows using Lemma 8.2, the same arguments as for the first part and the fact that

$$\begin{pmatrix} \sqrt{l}\hat{\beta}^* \\ l(\hat{\rho}^*_{LS,C}-1) \end{pmatrix} = \begin{pmatrix} 1 & l^{-3/2}\sum_{t=2}^{l} X^*_{t-1} \\ l^{-3/2}\sum_{t=2}^{l} X^*_{t-1} & l^{-2}\sum_{t=2}^{l} X^{*2}_{t-1} \end{pmatrix}^{-1} \times \begin{pmatrix} l^{-1/2}\sum_{t=2}^{l} U^*_t \\ l^{-1}\sum_{t=2}^{l} X^*_{t-1}U^*_t \end{pmatrix}. \qquad Q.E.D.$$

To introduce the next lemma we first fix some notation. Let $\|x\|$ be the Euclidean length of the vector x and $\|A\| = \sup\{\|Ax\|, \|x\|=1\}$ for a matrix A. Let further $\delta(p) = (\delta_{1,p}, \delta_{2,p}, \ldots, \delta_{p,p})'$ be the coefficients of the orthogonal projection $\widehat{X}_{t-1}$ of X_{t-1} on the closed span $\mathcal{M}_{t-1,t-p} = \overline{sp}\{\Delta X_{t-i}, i = 1,2,\ldots,p\}$ and $\tilde{\delta}(p)$ be the $p+1$ dimensional vector $\tilde{\delta}(p) = ((1-\delta_{1,p}), (\delta_{1,p}-\delta_{2,p}), \ldots, (\delta_{p-1,p}-\delta_{p,p}), \delta_{p,p})'$. Notice that $\|\tilde{\delta}(p)\|^2 \le E(X_{t-1}-\widehat{X}_{t-1})^2$ and that $\lim_{p\to\infty} E(X_{t-1}-\widehat{X}_{t-1})^2 = 0$ since $X_{t-1} \in \mathcal{M}_{t-1,-\infty} = \overline{sp}\{\Delta X_{t-i}, i=1,2,\ldots\}$, i.e., $\|\tilde{\delta}(p)\| \to 0$ as $p \to \infty$.

LEMMA 8.3: *Let the process $\{X_t\}$ satisfy Condition A' with $\rho < 1$ and let $\hat{\rho}_{DF}$ be the least squares estimator of the coefficient of X_{t-1} in the regression of X_t on X_{t-1} and on $X_{t-i}-X_{t-i-1}$, $i=1,2,\ldots,p$. If the choice of p is such that $p=p(n)\to\infty$,*

$$\frac{p^{11/2}}{n^{1/2}\|\tilde{\delta}(p)\|^2} \to 0, \quad \textit{and} \quad \|\tilde{\delta}(p)\|^{-1}\sqrt{n}\sum_{j=p+1}^{\infty}|a_j| \to 0,$$

as $n\to\infty$, then

$$\hat{\rho}_{DF} - \rho = O_P\left(\frac{1}{\sqrt{n}\|\tilde{\delta}(p)\|}\right),$$

where $\rho = -\sum_{j=1}^{\infty}\pi_j$.

The proof of this lemma is given in Paparoditis and Politis (2002).

Notice that under the assumptions made $\sqrt{n}\|\tilde{\delta}(p)\| \to \infty$ as $n\to\infty$, i.e., under stationarity, $\hat{\rho}_{DF}$ is an asymptotically consistent estimator of ρ. Since $\|\tilde{\delta}(p)\|\to 0$ as $p\to\infty$, the first condition implies that under stationarity, p has to increase to infinity at a much lower rate than in the case where $\{X_t\}$ is unit root integrated.

To proceed with the proof of Theorem 4.2, let D_n be the diagonal matrix $D_n = \langle (l-p), \sqrt{l-p}, \ldots, \sqrt{l-p} \rangle$ of appropriate dimension, and let $C^*(p)$ and $\widetilde{C}^*(p)$ be the block diagonal matrices

$$C_D^*(p) = \begin{pmatrix} c_{1,1}^* & \mathbf{0}' \\ \mathbf{0} & \Gamma_p^* \end{pmatrix} \quad \text{and} \quad \widetilde{C}_D^*(p) \begin{pmatrix} \tilde{c}_{1,1}^* & \tilde{c}_{1,2}^* & \mathbf{0}' \\ \tilde{c}_{2,1}^* & \tilde{c}_{2,2}^* & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & \Gamma_p^* \end{pmatrix},$$

where $\mathbf{0}$ is a $p \times 1$ zero vector, $c_{1,1}^* = \tilde{c}_{2,2}^* = (l-p)^{-2} \sum_{t=p+2}^{l} X_{t-1}^{*^2}$, $\tilde{c}_{2,1}^* = \tilde{c}_{1,2}^* = (l-p)^{-1} \sum_{t=p+2}^{l} X_{t-1}^*$, $\tilde{c}_{1,1}^* = 1$, and $\Gamma_p^* = (l^{-1} \sum_{t=p+2}^{l} D_{t-i}^* D_{t-j}^*)_{i,j=1,2,\ldots,p}$. Furthermore, let $e_1' = (1, 0, 0, \ldots, 0)'$ of appropriate dimension,

$$\begin{aligned} Y_{t-1}^*(p) &= \left(X_{t-1}^*, D_{t-1}^*, \ldots, D_{t-p}^*\right)', \\ \widetilde{Y}_{t-1}^*(p) &= \left(1, X_{t-1}^*, D_{t-1}^*, \ldots, D_{t-p}^*\right)', \\ C^*(p) &= D_n^{-1} \sum_{t=p+2}^{l} Y_{t-1}^*(p) Y_{t-1}^{*'}(p) D_n^{-1}, \quad \text{and} \\ \widetilde{C}^*(p) &= D_n^{-1} \sum_{t=p+2}^{l} \widetilde{Y}_{t-1}^*(p) \widetilde{Y}_{t-1}^{*'}(p) D_n^{-1}. \end{aligned}$$

The following lemma can then be established.

LEMMA 8.4: *Let the assumptions of Theorem* 4.2 *be satisfied*. *If* $n \to \infty$, *then*

(i) $$\sqrt{p+1} \| e_1' (C^{*^{-1}}(p) - C_D^{*^{-1}}(p)) \| \to 0,$$

(ii) $$\sqrt{p+1} \| e_1' (\widetilde{C}^{*^{-1}}(p) - \widetilde{C}_D^{*^{-1}}(p)) \| \to 0,$$

(iii) $$(l-p)^{-1} \sum_{t=p+2}^{l} X_{t-1}^* \left(U_t^* - \sum_{j=1}^{p} a_j^* D_{t-j}^* \right) \Rightarrow \frac{1}{2} \sigma_\varepsilon^2 C_\Psi (W^2(1) - 1),$$

and

(iv) $$(l-p)^{-1/2} \sum_{t=p+2}^{l} \left(U_t^* - \sum_{j=1}^{p} a_j^* D_{t-j}^* \right) \Rightarrow \sigma_\varepsilon W(1),$$

in probability, *where joint convergence of the limits in* (iii) *and* (iv) *applies*.

PROOF: Consider (i). Verify using formulae for the inverse of partitioned matrices that

$$(8.13)\qquad e_1'\big(C^{*^{-1}}(p)-C_D^{*^{-1}}(p)\big)=\left(\frac{c_{1,2}^*\Gamma_p^{*^{-1}}c_{2,1}^*}{c_{1,1}^*\big(c_{1,1}^*-c_{1,2}^*\Gamma_p^{*^{-1}}c_{2,1}^*\big)},\frac{c_{1,2}^*\Gamma_p^{*^{-1}}}{c_{1,1}^*-c_{1,2}^*\Gamma_p^{*^{-1}}c_{2,1}^*}\right).$$

Note that under the assumptions made, $E^*((l-p)^{-3/2}\sum_{t=p+2}^{l}X_{t-1}^*D_{t-i}^*)^2\le(l-p)^{-1}C_2$ and therefore

$$(8.14)\qquad \|c_{1,2}^*\|=\left\|\left((l-p)^{-3/2}\sum_{t=p+2}^{l}X_{t-1}^*D_{t-i}^*,\ i=1,2,\ldots,p\right)'\right\|$$
$$=O_P(p^{1/2}(l-p)^{-1/2}).$$

Furthermore, since $E^*((l-p)^{-1/2}\sum_{t=p+1}^{l}D_{t-i}^*D_{t-j}^*-E(X_{t-i}-X_{t-i-1})(X_{t-j}-X_{t-j-1}))^2\le C$, we have $E^*\|\Gamma_p^*-\Gamma_p\|^2\le Cp^2(l-p)^{-1}$, where $\Gamma_p=(E\Delta X_{t-i}\Delta X_{t-j})_{i,j=1,2,\ldots,p}$. Since for every p the matrix Γ_p is positive definite, $\|\Gamma_p^{-1}\|$ is the reciprocal of the minimal eigenvalue of Γ_p. The spectral density of $\{\Delta X_t\}$ is given by $f_{\Delta X_t}(\lambda)=|1-e^{-i\lambda}|^2f_{X_t}(\lambda)$, where f_{X_t} denotes the spectral density of $\{X_t\}$. For the minimal eigenvalue of Γ_p we have

$$\inf_{\|x\|=1}\sum_{j=1}^{p}\sum_{k=1}^{p}x_j\operatorname{cov}(\Delta X_{t-j},\Delta X_{t-k})x_k$$
$$=\inf_{\|x\|=1}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{j=1}^{p}x_je^{ij\lambda}\right|^2f_{X_t}(\lambda)|1-e^{-i\lambda}|^2\,d\lambda$$
$$\ge\inf_{\lambda\in[0,\pi]}f_{X_t}(\lambda)\inf_{\|x\|=1}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{j=1}^{p}x_je^{ij\lambda}\right|^2|1-e^{i\lambda}|^2\,d\lambda$$
$$=\inf_{\lambda\in[0,\pi]}f_{X_t}(\lambda)\tilde\lambda_{\min},$$

where $\tilde\lambda_{\min}$ denotes the minimal eigenvalue of the $p\times p$ covariance matrix of the process with spectral density $(2\pi)^{-1}|1-e^{-i\lambda}|^2=\pi^{-1}(1-\cos(\lambda))$. Since this is a tridiagonal matrix with 2 on the main diagonal and -1 on the diagonal above and below the main diagonal, we have $\tilde\lambda_k=2(1-\cos((k\pi)/(p+1)))$, $k=1,2,\ldots,p$, and therefore

$$\left\|\Gamma_p^{-1}\right\|\le\frac{1}{K\big(1-\cos\big(\frac{\pi}{p+1}\big)\big)}.$$

Now, using $\|\Gamma_p^{*^{-1}}\|\le\|\Gamma_p^{-1}\|+\|\Gamma_p^{*^{-1}}-\Gamma_p^{-1}\|$ and standard arguments (cf. Berk (1974)), we get that

$$(8.15)\qquad \left\|\Gamma_p^{*^{-1}}\right\|\le O_P\left(\frac{1}{\big(1-\cos\big(\frac{\pi}{p+1}\big)\big)}\right)+O_P\left(\frac{p}{\sqrt{l-p}\big(1-\cos\big(\frac{\pi}{p+1}\big)\big)^2}\right).$$

Using (8.14) and (8.15) we get $c^*_{1,2}\Gamma_p^{*^{-1}}c^*_{2,1} \to 0$ in probability, i.e., $c^*_{1,1} - c^*_{1,2}\Gamma_p^{*^{-1}}c^*_{2,1} \to \sigma_\varepsilon^2 C_\Psi^2 \int_0^1 W^2(r)\,dr$. From (8.13), (8.14), and (8.15) it follows then that

$$\sqrt{p+1}\left\| e_1'\left(C^{*^{-1}}(p) - C_D^{*^{-1}}(p)\right)\right\| = O_P\left(\frac{p}{\sqrt{n}\left(1-\cos\left(\frac{\pi}{p+1}\right)\right)}\right),$$

which, taking into account that $1/(1-\cos(\pi/(p+1))) = O(p^2)$, concludes the proof of assertion (i). Since assertion (ii) can be proved using the same arguments, the details are omitted. To prove (iii) we use

$$\begin{aligned}(8.16)\qquad &\Bigg|\frac{1}{l-p}\sum_{t=p+2}^{l} X^*_{t-1}\left(U^*_t - \sum_{j=1}^{p} a_j D^*_{t-j}\right)\\ &\quad - \frac{1}{l-p}\sum_{m=0}^{k-1}\sum_{s=1}^{b} X^\star_{i_m+s-1}\left(U^\star_{i_m+s} - \sum_{j=1}^{p} a_j D_{i_m+s-1-j}\right)\Bigg| \to 0\end{aligned}$$

in probability, where $U^\star_t = (X_t - \rho X_{t-1}) - (n-1)^{-1}\sum_{\tau=2}^{n}(X_\tau - \rho X_{\tau-1})$ and $X^\star_t$ is the series obtained by replacing U_t by $U^\star_t$ in the first step of the RBB bootstrap algorithm.

We next show that

$$(8.17)\qquad \frac{1}{l-p}\sum_{m=0}^{k-1}\sum_{s=1}^{b} X^\star_{i_m+s-1}\left(U^\star_{i_m+s} - \sum_{j=1}^{p} a_j D_{i_m+s-1-j}\right) \Rightarrow \frac{1}{2}\sigma_\varepsilon^2 C_\Psi (W^2(1)-1).$$

For this note that by (2.5) and the definition of D_{t-j} we have

$$\begin{aligned}&\frac{1}{l-p}\sum_{m=0}^{k-1}\sum_{s=1}^{b} X^\star_{i_m+s-1}\left(U^\star_{i_m+s} - \sum_{j=1}^{p} a_j D_{i_m+s-1-j}\right)\\ &\quad = \frac{1}{l-p}\sum_{m=0}^{k-1}\sum_{s=1}^{b} X^\star_{i_m+s-1}\left(\varepsilon_{i_m+s} - \frac{1}{n-1}\sum_{\tau=2}^{n}\varepsilon_\tau\right)\\ &\qquad + \frac{1}{l-p}\sum_{m=0}^{k-1}\sum_{s=1}^{b}\sum_{j=1}^{p} a_j X^\star_{i_m+s-1}\\ &\qquad \times\left(\frac{1}{n-1}\sum_{\tau=2}^{n}(X_\tau - X_{\tau-1}) - \frac{1}{n-1}\sum_{\tau=p+2}^{n-p}(X_{\tau-j} - X_{\tau-j-1})\right)\\ &\qquad + \frac{1}{l-p}\sum_{m=0}^{k-1}\sum_{s=1}^{b}\sum_{j=p+1}^{\infty} a_j X^\star_{i_m+s-1}\\ &\qquad \times\left((X_{i_m+s-j} - X_{i_m+s-j-1}) - \frac{1}{n-1}\sum_{\tau=p+2}^{n-p}(X_{\tau-j} - X_{\tau-j-1})\right)\\ &\quad = T^*_{1,n} + T^*_{2,n} + T^*_{3,n}\end{aligned}$$

with an obvious notation for $T^*_{1,n}$, $T^*_{2,n}$, and $T^*_{3,n}$. The proof of assertion (iii) of the lemma is then concluded because $T^*_{1,n} \Rightarrow (1/2)\sigma^2_\varepsilon C_\Psi(W^2(1)-1)$, $T^*_{2,n} \to 0$ and $T^*_{3,n} \to 0$, in probability. Details are given in Paparoditis and Politis (2000). Assertion (iv) is proved along the same lines. *Q.E.D.*

PROOF OF THEOREM 4.2: We give the proof of the first part of the theorem since the proof of the second part is very similar.

Let $\theta = (1, a_1, \ldots, a_p)'$ and note that

$$(l-p)\big(\hat{\rho}^*_{DF}-1\big) = e_1' C^{*^{-1}}(p) D_n^{-1} \sum_{t=p+2}^{l} Y^*_{t-1}(p) e^*_t$$

where $e^*_t = X^*_t - \theta^{*'} Y^*_{t-1}(p) = U^*_t - \sum_{j=1}^{p} a_j D^*_{t-j}$. Write

$$(l-p)\big(\hat{\rho}^*_{DF}-1\big) = e_1' C_D^{*^{-1}}(p) D_n^{-1} \sum_{t=p+2}^{l} Y^*_{t-1}(p) e^*_t + e_1'\big(C^{*^{-1}}(p) - C_D^{*^{-1}}(p)\big) D_n^{-1} \sum_{t=p+2}^{l} Y^*_{t-1}(p) e^*_t$$

and verify by straightforward calculations that

$$\left\| D_n^{-1} \sum_{t=p+2}^{l} Y^*_{t-1}(p) e^*_t \right\| = O_{P^*}\big(\sqrt{p+1}\big). \tag{8.18}$$

This together with Lemma 8.4(i) implies that

$$\left| e_1'\big(C^{*^{-1}}(p) - C_D^{*^{-1}}(p)\big) D_n^{-1} \sum_{t=p+2}^{l} Y^*_{t-1}(p) e^*_t \right| \to 0 \tag{8.19}$$

in probability. Now, since $C_D^{*^{-1}}(p)$ is a block diagonal matrix, we get

$$e_1' C_D^{*^{-1}}(p) D_n^{-1} \sum_{t=p+2}^{l} Y^*_{t-1}(p) e^*_t = \left((l-p)^{-2} \sum_{t=p+2}^{l} X^{*^2}_{t-1} \right)^{-1} \times \frac{1}{l-p} \sum_{t=p+2}^{l} X^*_{t-1} e^*_t.$$

Thus

$$\left| (l-p)\big(\hat{\rho}^*_{DF}-1\big) - \left((l-p)^{-2} \sum_{t=p+2}^{l} X^{*^2}_{t-1} \right)^{-1} \frac{1}{l-p} \sum_{t=p+2}^{l} X^*_{t-1} e^*_t \right| \to 0$$

in probability. Because of this and the fact that

$$\left((l-p)^{-2}\sum_{t=p+2}^{l} X_{t-1}^{*^2}, (l-p)^{-1}\sum_{t=p+2}^{l} X_{t-1}^{*}e_t^{*}\right)$$
$$\Rightarrow \left(\sigma_\varepsilon^2 C_\Psi^2 \int W^2(r)\,dr, (1/2)\sigma_\varepsilon^2 C_\Psi (W^2(1)-1)\right),$$

in probability (cf. Lemma 8.2(i) and Lemma 8.4 (iii)), we conclude that

$$(l-p)\left(\hat{\rho}_{DF}^{*}-1\right) \Rightarrow \frac{1}{2C_\Psi}(W^2(1)-1)\Big/\int W^2(r)\,dr$$

in probability. *Q.E.D.*

PROOF OF THEOREM 5.1: Note first that for $\rho_n = 1 + c/n$ the LS estimator $\hat{\rho}_{LS}$ of ρ_n satisfies condition (2.2); cf. Theorem 1(a) of Phillips (1987b). As in the proof of Theorem 3.1 and under the assumptions made, we get using Lemma 1 of Phillips (1987b) that

$$\mathscr{L}\left(l\left(\hat{\rho}_{LS}^{*}-1\right)\big|X_1, X_2, \ldots, X_n\right) \Rightarrow \left(W^2(1)-\sigma_U^2/\sigma^2\right)\Big/\left(2\int_0^1 W^2(r)\,dr\right),$$

where uniform convergence also applies. Therefore, $C_\alpha^* \to C_\alpha$ in probability. Hence,

$$\begin{aligned}\beta_{RBB,\,n}(\rho_n;\alpha) &= P(n(\hat{\rho}_{LS}-1) \le C_\alpha^*)\\ &= P(n(\hat{\rho}_{LS}-\rho_n) \le C_\alpha^* - c) \to P(J \le C_\alpha - c).\end{aligned}$$ *Q.E.D.*

To prove Theorem 5.2 the following lemmas are needed.

LEMMA 8.5: *Let $\{X_t\}$ be a stationary process satisfying Condition A(ii) or B with $\beta = 0$. If $b \to \infty$ such that $b/\sqrt{n} \to 0$ as $n \to \infty$, then*

$$k^{-1/2}\sum_{j=1}^{[lr]} L_j^{+}\big/\sigma^{+} \Rightarrow W(r), \quad r \in [0,1],$$

in probability, where $L_1^{+} = X_1, L_t^{+} = X_t^{+} - X_{t-1}^{+}$ for $t = 2, 3, \ldots, l$, and $\sigma^{+^2} =$ $\mathrm{var}^{+}(k^{-1/2}\sum_{j=1}^{l} L_j^{+})$.

PROOF: As in the proof of Theorem 3.1, it suffices to consider $S_k^{+}(r) = k^{-1/2}$ $\sum_{m=0}^{M_r}\sum_{s=1}^{B} L_{i_m+s}^{+}/\sigma^{+}$, where $M_r = [([lr]-2)/b]$ and $B = \min\{b, [lr]-mb-1\}$.

Verify first by straightforward calculations that

$$\text{(8.20)}\qquad \sigma^{+^2} = \frac{1}{k}\sum_{m=0}^{k-1} E^+\left(\sum_{s=1}^{b} L_{i_m+s}\right)^2 + o_P(1)$$
$$= \frac{1}{n-b}\sum_{t=1}^{n-b}\left(X_{t+b}-X_t\right)^2 + o_P(1) \to 2\,\mathrm{var}(X_t).$$

Furthermore,

$$\text{(8.21)}\qquad \frac{1}{\sqrt{k}}\sum_{m=0}^{M_r}\sum_{s=1}^{B} L_{i_m+s} = \frac{1}{\sqrt{k}}\sum_{m=0}^{M_r} Y_{i_m} + o_P(1),$$

where the random variables $Y_{i_m} = (X_{i_m+b} - X_{i_m}) - E^+(X_{i_m+b} - X_{i_m})$ are independent, identically distributed, and the $o_P(1)$ term is mainly due to the fact that

$$\frac{1}{\sqrt{k}}\sum_{m=0}^{M_r}\left(E^+(X_{i_m+b} - X_{i_m}) - \frac{b}{n-1}\sum_{j=2}^{n}\left(X_j - X_{j-1}\right)\right) = O_P(k^{-1/2}).$$

Now, the assertion of the lemma follows using (8.20), (8.21), and the version of Donsker's theorem for partial sums of triangular arrays of independent random variables used in the proof of Theorem 3.1. *Q.E.D.*

LEMMA 8.6: *Under the assumptions of Lemma* 8.5 *and as* $n \to \infty$,

$$\text{(i)}\qquad l^{-1}\sum_{t=2}^{l} X_{t-1}^+(X_t^+ - X_{t-1}^+) \Rightarrow -(\mathrm{var}(X_t) - \mathrm{cov}(X_t, X_{t-1})),$$

$$\text{(ii)}\qquad (lk)^{-1}\sum_{t=2}^{l} X_{t-1}^{+^2} \Rightarrow 2\,\mathrm{var}(X_t)\int_0^1 W^2(r)\,dr,$$

in probability, where joint weak convergence of the above limits also applies.

PROOF: Let $S^+(r) = \sum_{t=1}^{j-1} L_t^+/\sigma^+$ for $(j-1)/l \le r \le j/l$, $j = 2, 3, \ldots, l$, and $S^+(1) = \sum_{t=1}^{l} L_t^+/\sigma^+$. Assertion (i) follows since

$$\begin{aligned} l^{-1}\sum_{t=2}^{l}\left(X_t^+ - X_{t-1}^+\right)X_{t-1}^+ &= l^{-1}\sum_{t=2}^{l} L_t^+\left(\sum_{j=1}^{t-1} L_j^+\right) \\ &= (2l)^{-1}\sum_{t=2}^{l}\left[\left(\sum_{j=1}^{t} L_j^+\right)^2 - \left(\sum_{j=1}^{t-1} L_j^+\right)^2\right] \\ &\quad - (2l)^{-1}\sum_{t=2}^{l} L_t^+ \end{aligned}$$

$$= (2b)^{-1}\sigma^{+^2}\Big[S_k^{+^2}(1) - S_k^{+^2}(1/l)\Big] - (2l)^{-1}\sum_{t=2}^{l} L_t^{+^2}$$
$$\to -\frac{1}{2}E(X_t - X_{t-1})^2$$
$$= -(\mathrm{var}(X_t) - \mathrm{cov}(X_t, X_{t+1})),$$

in probability. Assertion (ii) follows because

$$(lk)^{-1}\sum_{t=2}^{l} X_{t-1}^{+^2} = l^{-1}\sum_{t=2}^{l}\left(k^{-1/2}\sum_{j=1}^{t-1} L_j^+\right)^2$$
$$= \sigma^{+^2}\sum_{t=2}^{l}\int_{(t-1)/l}^{t/l}\big(k^{-1/2}S^+([lr])\big)^2\,dr$$
$$\Rightarrow 2\,\mathrm{var}(X_t)\int_0^1 W^2(r)\,dr,$$

in probability, by the continuous mapping theorem and using $\sigma^{+^2} \to 2\,\mathrm{var}(X_t)$, in probability. *Q.E.D.*

PROOF OF THEOREM 5.2: The case $\rho = 1$ follows essentially by the same arguments as in the proof of the validity of the RBB procedure with the main simplification that $\tilde{\rho} = 1$ in the proof of Theorem 3.1. The case $\rho \in (-1, 1)$ is proved using Lemma 8.6 and the same arguments as in the proof of the first assertion of Theorem 4.1. *Q.E.D.*

The following lemma is essential in establishing Theorem 5.3.

LEMMA 8.7: *Let $\{X_t\}$ satisfy Condition B'. If $b \to \infty$ such that $b/\sqrt{n} \to 0$ as $n \to \infty$, then*

$$l^{-1/2}\sum_{j=1}^{[lr]} L_j^+/\sigma^+ \Rightarrow W(r), \qquad r \in [0, 1],$$

in probability, where $L_1^+ = X_1$, $L_t^+ = X_t^+ - X_{t-1}^+$ for $t = 2, 3, \ldots, l$ and $\sigma^{+^2} =$ $\mathrm{var}^+(l^{-1/2}\sum_{j=1}^{l} L_j^+)$.

PROOF: Since the arguments are very similar to those used in the proof of Theorem 3.1 and Lemma 8.5, we stress only the essentials. Verify first by straightforward calculations that $\sigma^{+^2} \to 2\pi f_U(0)$. As in the proof of Theorem 3.1, consider $S_l^+(r) = l^{-1/2}\sum_{m=0}^{M_r}\sum_{s=1}^{b} L_{i_m+s}^+/\sigma^+$, where $M_r = [([lr]-2)/b]$. We then have

$$\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{s=1}^{b} L_{i_m+s}^+ = \frac{1}{\sqrt{k}}\sum_{m=0}^{M_r} Y_m^+ + R_l^+(r),$$

where $Y_m^+ = b^{-1/2}\sum_{s=1}^{b}\{(X_{i_m+s} - X_{i_m+s-1}) - E^+(X_{i_m+s} - X_{i_m+s-1})\}$ and

$$R_l^+(r) = -\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\left[\frac{1}{n-b}\sum_{t=1}^{n-b}(X_{t+s} - X_{t+s-1}) - \frac{1}{n-1}\sum_{t=2}^{n}(X_t - X_{t-1})\right].$$

Using $X_t - X_{t-1} = n^{-1}cX_{t-1} + U_t$, we get by straightforward calculations that uniformly in r, $R_l^+(r) = O_P(b/\sqrt{l}) \to 0$. Furthermore, because

$$Y_m^+ = \frac{1}{\sqrt{b}}\sum_{s=1}^{b}(U_{i_m+s} - E^+U_{i_m+s}) + \frac{c}{n\sqrt{b}}\sum_{s=1}^{b}(X_{i_m+s} - E^+X_{i_m+s}),$$

we get using Lemma 1 of Phillips (1987b) that

$$\mathrm{var}^+\left(\frac{c}{n\sqrt{b}}\sum_{s=1}^{b}(X_{i_m+s} - E^+X_{i_m+s})\right) = O_P(b/n) \to 0.$$

Therefore, we have uniformly in r,

$$\left|\frac{1}{\sqrt{l}}\sum_{m=0}^{M_r}\sum_{s=1}^{b} L_{i_m+s}^+ - \frac{1}{\sqrt{k}}\sum_{m=0}^{[kr]} V_m^+\right| \to 0,$$

in probability, where $V_m^+ = b^{-1/2}\sum_{s=1}^{b}(U_{i_m+s} - E^+U_{i_m+s})$. The remainder of the proof follows the proof of Theorem 3.1. *Q.E.D.*

PROOF OF THEOREM 5.3: Let $\sqrt{l}(\hat{\rho}_{LS}^+ - 1)$ be the least squares statistic based on the pseudoseries $X_1^+, X_2^+, \dots, X_l^+$. By Lemma 8.7 and along the same lines as in Lemma 8.6, we get

$$\mathscr{L}(l(\hat{\rho}_{LS}^+ - 1)|X_1, X_2, \dots, X_n) \Rightarrow (W^2(1) - \sigma_U^2/\sigma^2)/\left(2\int_0^1 W^2(r)\,dr\right)$$

in probability. The assertion of the theorem is then established using the same arguments as in the proof of Theorem 5.1. *Q.E.D.*

Department of Mathematics and Statistics, University of Cyprus, P.O.Box 20537, CY 1678 Nicosia, Cyprus; stathisp@ucy.ac.cy

and

Department of Mathematics, University of California, San Diego, La Jolla, CA 920930112, U.S.A.; politis@euclid.ucsd.edu

Manuscript received July, 2000; final revision received August, 2002.

REFERENCES

ABADIR, K. M. (1993): "The Limiting Distribution of the Autocorrelation Coefficient Under a Unit Root," *Annals of Statistics*, 21, 1058–1070.

ANDREWS, D. W. K. (1984): "Nonstrong Mixing Autoregressive Processes," *Journal of Applied Probability*, 21, 930–934.

BASAWA, I. V., A. K. MALLIK, W. P. MCCORMICK, J. H. REEVES, AND R. L. TAYLOR (1991): "Bootstrap Test of Significance and Sequential Bootstrap Estimation for Unstable First Order Autoregressive Processes," *Communications in Statistics-Theory and Methods*, 20, 1015–1026.

BERK, K. (1974): "Consistent Autoregressive Spectral Estimators," *Annals of Statistics*, 2, 489–502.

BERTAIL, P. (1994): "Un Test Bootstrap dans un Modéle AR(1)," *Annales d' Économie et des Statistique*, 36, 57–79.

BILLINGSLEY, P. (1999): *Convergence of Probability Measures*. New York: John Wiley.

BROCKWELL, P. K., AND R. A. DAVIS (1991): *Time Series: Theory and Methods*, 2nd ed. New York: Springer.

CAMBELL, J. Y., AND P. PERRON (1991): "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," *NBER Macroeconomics Annual*. Cambridge, Mass.: MIT Press.

CHANG, Y., AND J. Y. PARK (2001): "A Sieve Bootstrap for the Test of a Unit Root," *Journal of Time Series Analysis*, to appear.

DATTA, S. (1996): "On Asymptotic Properties of Bootstrap for AR(1) Processes," *Journal of Statistical Planning and Inference*, 53, 361–374.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

FERRETTI, N., AND J. ROMO (1996): "Unit Root Bootstrap Test for AR(1) Models," *Biometrika*, 83, 849–860.

FULLER, W. (1996): *Introduction to Statistical Time Series*, 2nd ed. New York: John Wiley.

HALL, P., J. L. HOROWITZ, AND B.-Y. JING (1995): "On Blocking Rules for the Bootstrap With Dependent Data," *Biometrika*, 82, 561–574.

HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton, New Jersey: Princeton University Press.

KÜNSCH, H. R. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241.

LI, H., AND G. S. MADDALA (1996): "Bootstrapping Time Series Models," *Econometric Reviews*, 15, 115–158.

LIU, R. Y., AND K. SINGH (1992): "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in *Exploring the Limits of the Bootstrap*, ed. by R. LePage and L. Billard. New York: John Wiley, 225–248.

NANKERVIS, J. C., AND N. E. SAVIN (1996): "The Level and Power of the Bootstrap *t* Test in the AR(I) Model with Trend," *Journal of Business and Economic Statistics*, 14, 161–168.

NG, S., AND R. PERRON (1995): "Unit Root Test 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., AND D. N. POLITIS (2000): "Unit Root Testing via the Continuous-Path Block Bootstrap," Discussion Paper 2001-06, Dept. of Economics, University of California, San Diego, USA.

——— (2001): "The Continuous-Path Block Bootstrap," in *Asymptotics in Statistics and Probability. Papers in Honor of George Roussas*, ed. by Madan Puri. Zeist: VSP Publications, pp. 305–320.

PAPARODITIS E., AND D. N. POLITIS (2002): "Bootstrapping Unit Root Tests for Autoregressive Processes," Preprint.

PARK, Y. J. (2000): "Bootstrap Unit Root Tests," Preprint.

PHILLIPS, P. C. B. (1987a): "Time Series Regression with a Unit Root," *Econometrica*, 2, 277–301.

——— (1987b): "Towards a Unified Asymptotic Theory for Autoregression, *Biometrika*, 74, 535–547.

PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

POLITIS, D. N. (2001): "Adaptive Bandwidth Choice," preprint.

POLITIS, D. N., AND J. P. ROMANO (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.
POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. New York: Springer.
POLITIS, D. N., AND H. WHITE (2000): "Automatic Block-length Selection for the Dependent Bootstrap," preprint.
PSARADAKIS, Z. (2001): "Sieve Bootstrap Tests for a Unit Root in Time Series Regressions," *Journal of Time Series Analysis*, 22, 577–594.
ROMANO, J. P., AND L. A. THOMBS (1996): "Inference for Autocorrelations under Weak Assumptions," *Journal of the American Statistical Association*, 91, 590–600.
ROMANO, J. P., AND M. WOLF (2001): "Subsampling Confidence Intervals for the Autoregressive Root," *Econometrica*, 69, 1283–1314.
ROSENBLATT, M. (1985): *Stationary Sequences and Random Fields*. Boston: Birkhäuser.
SAID, S. E., AND D. A. DICKEY (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599–607.
SERFLING, R. (1980): *Approximation Theorems of Mathematical Statistics*. New York: John Wiley.
STOCK, J. (1994): "Unit Roots and Trend Breaks," in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. Mc-Fadden. Amsterdam: North-Holland.
WEST, K. D. (1988): "Asymptotic Normality When Regressors Have a Unit Root," *Econometrica*, 56, 1397–1417.
WITHERS, C. S. (1981): "Conditions for a Linear Process to be Strong-Mixing," *Probability Theory and Related Fields*, 57, 477–480.