

UC San Diego

Recent Work

Title

Unit Root Testing via the Continuous-Path Block Bootstrap

Permalink

<https://escholarship.org/uc/item/9qb4r775>

Authors

Paparoditis, Efstathios
Politis, Dimitris N

Publication Date

2001-03-01

2001-06

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

UNIT ROOT TESTING VIA THE CONTINUOUS-PATH BLOCK
BOOTSTRAP

BY

EFSTATHIOS PAPARODITIS

AND

DIMITRIS N. POLITIS

**DISCUSSION PAPER 2001-06
MARCH 2001**

Unit Root Testing via the Continuous-Path Block Bootstrap

Efstathios Paparoditis
Department of Mathematics and Statistics
University of Cyprus
P.O.Box 20537
CY 1678 Nicosia, Cyprus

Dimitris N. Politis
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-012
USA

Abstract

A new resampling procedure, the continuous-path block bootstrap, is proposed in the context of testing for integrated (unit root) time series. The continuous-path block bootstrap (CBB) is a nonparametric procedure that successfully generates unit root integrated pseudo time series retaining the important characteristics of the data, e.g., the dependence structure of the stationary process driving the random walk. As a consequence, the CBB can accurately capture the distribution of many unit root test statistics. Large sample theory for the new bootstrap methodology is developed and the asymptotic validity of CBB-based unit root testing is shown via a bootstrap functional limit theorem. Applications of the new procedure to least squares and Dickey-Fuller type test statistics of the unit root hypothesis are given. Finite-sample simulations confirm a good α -level accuracy and an increased power associated with CBB-based unit root testing.

Key words: Autocorrelation, Hypothesis Testing, Integrated Series, Non-stationary Series, Random Walk, Resampling.

1 Introduction

Consider time series data of the form X_1, X_2, \dots, X_n , where $\{X_t, t = 1, 2, \dots\}$ is a sequence of random variables with mean zero; for convenience we assume that $X_0 = 0$ although other common choices of the initial value can be also considered. Following the seminal work of Fuller (1996) and 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 decades. In particular, assume that the time series $\{X_t\}$ is either stationary, 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, where $Y_t := X_t - X_{t-1}$. The hypothesis test of interest can then be stated as:

$$\begin{aligned} H_0 & : \{X_t\} \text{ is } I(1) \\ & \text{versus} \\ H_1 & : \{X_t\} \text{ is stationary.} \end{aligned}$$

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 ρ parameter 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:

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

for $t = 1, 2, \dots$. 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. Finally, the quality of large sample approximations have been questioned in several simulation

studies where severe size distortions have been reported for some commonly used tests; see for instance Schwert (1989) or DeJong et al. (1992).

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—see Efron and Tibshirani (1993) and Shao and Tu (1995) for an overview of bootstrap methods.

Furthermore, in designing a nonparametric bootstrap procedure for testing purposes, an additional aspect must be taken into account which is important for good power performance. For such purposes, the bootstrap procedure should be able to reproduce the behavior of the test statistic under the null hypothesis (e.g., unit root integration) *whether the observed series obeys the null hypothesis or not*. In this connection, recent fruitful attempts to approach the unit root problem via bootstrap methods have been based on quite restrictive assumptions on the parametric structure of the model generating the data, assuming—more often than not—an AR(1) structure with i.i.d. residuals, i.e., equation (1.1) in connection to an i.i.d. sequence $\{U_t\}$; cf. Bertail (1994) and Ferretti and Romo (1996).

In the present paper, a new nonparametric bootstrap procedure is introduced, and—for reasons to be apparent shortly—is termed the *Continuous-Path Block Bootstrap testing procedure*; by its construction, the Continuous-Path Block Bootstrap testing procedure generates unit root time series while at the same time it manages to automatically (and nonparametrically) replicate the important weak dependence characteristics of the data, e.g., the dependence structure of the stationary process $\{U_t\}$. The Continuous-Path Block Bootstrap (CBB) testing procedure is a modification of a resampling algorithm introduced recently by Paparoditis and Politis (2000), and it is based on the block bootstrap of Künsch (1989) and Liu and Singh (1992).

To motivate the CBB, let us give an illustration demonstrating the failure of the block bootstrap (BB) under the presence of a unit root. Figure 1(a) shows a plot of (the natural logarithm of) the Dow Jones Utilities index series recorded daily from Aug. 28 to Dec. 18, 1972, while Figure 1(b) shows a realization of a BB pseudo replication of this series using block size 10. It is obvious visually that the bootstrap series is quite dissimilar to the original series, the most striking difference being the presence of strong discontinuities (of the ‘jump’ type) in the bootstrap series that—not surprisingly—occur every 10 time units, i.e., where the independent bootstrap blocks join.

Please insert Figure 1 about here

Figure 1(c) suggests a way to fix this problem by forcing the bootstrap sample path to be continuous. A simple way to do this is to *shift* each of the bootstrap blocks up or down with the goal of ensuring: (i) that the bootstrap series starts off at the same point as the original series,

and that (ii) the bootstrap sample path is continuous. Notably, the bootstrap blocks used in Figure 1(c) are the *exact same blocks* featuring in Figure 1(b).

At least as far as visual inspection of the plot can discern, the series in Figure 1(c) could just as well have been generated by the same probability mechanism that generated the original Dow Jones series. In other words, it is plausible that a bootstrap algorithm generating series such as the one in Figure 1(c) would be successful in mimicking important features of the original process; thus, the ‘‘Continuous-Path Block Bootstrap’’ of Figure 1(c) is expected to ‘work’ in this case.

Of course, the actual daily Dow Jones data are in discrete time, and talking about continuity is—strictly speaking—inappropriate. Nevertheless, an underlying continuous-time model may always be thought to exist, and the idea of continuity of sample paths is powerful and intuitive; hence the name ‘‘Continuous-Path Block Bootstrap’’ for our discrete-time methodology as well. Note that the above discussion describes just the rudimentary notion behind the CBB; we have here assumed that this series is indeed unit root integrated, and have postponed—until the next Section—the discussion of some necessary technical details, including the modification of the CBB that is required in the context of testing.

The paper is organized as follows. Section 2 describes in detail the CBB testing procedure and states its main characteristics. A bootstrap functional limit theorem for partial sum processes based on CBB pseudo-series is established in Section 3; consequently, the asymptotic validity of the CBB in approximating the distribution of some commonly used test statistics is shown in Section 4. Some extensions of the procedure are discussed in Section 5 while Section 6 examines the small sample performance of the CBB method. Section 7 summarizes our findings while all technical proofs are deferred to Section 8.

2 The Continuous-Path Block Bootstrap

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 $Y_t = X_t - X_{t-1}$. Note that under H_0 (which is equivalent to $\rho = 1$), it is obvious that the sequence $\{Y_t\}$ coincides with the sequence $\{U_t\}$ defined in equation (1.1). For technical reasons, we strengthen the above set-up by requiring that the weak dependence structure of $\{U_t\}$ satisfies one of two sets of conditions. The first one assumes that either $\{X_t\}$ is stationary and linear, or it is not stationary but its first difference series $\{Y_t\}$ is stationary and linear; as usual, linearity implies an MA(∞) representation with respect to some i.i.d. sequence $\{\varepsilon_t\}$. The second condition replaces linearity by a strong mixing assumption. We are now able to concisely state our assumptions in the following two conditions.

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

- (i) (Case $\rho = 1$). $X_t = X_{t-1} + U_t$ 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 generated by $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ where 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 $U_t = X_t - \rho X_{t-1}$ is always linear and stationary. For $\rho = 1$ this is so by assumption while for $\rho \neq 1$ we have

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

where $\Psi^+(L) = (1 - \rho L)\Psi(L)$, $\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$ and L is the shift operator defined by $L^k X_t := X_{t-k}$ for $k \in \mathbf{Z}$. 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 CBB 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 $\alpha(k) \rightarrow 0$ as $k \rightarrow \infty$ where the mixing coefficient $\alpha(k)$ is defined by

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

Here \mathcal{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: $E(U_t) = 0$, $E|U_t|^\beta < \infty$ for some $\beta > 2$, $f_U(0) > 0$, 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, $\sum_{k=1}^{\infty} \alpha(k)^{1-2/\beta} < \infty$, 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 $U_t = X_t - X_{t-1}$ is strong mixing. On the other hand, if $\{X_t\}$ is stationary ($\rho \neq 1$), then $\{X_t\}$ is a strong mixing process satisfying the conditions stated above. Since in this case $\{U_t\}$ is a moving average of $\{X_t\}$, i.e., $U_t = X_t - \rho X_{t-1}$, the process $\{U_t\}$ satisfies Condition B. Note that Condition B does not imply A, i.e., there are stationary processes satisfying Condition A which are not mixing; see Withers (1981) or Andrews (1984).

The Continuous-Path Block-Bootstrap (CBB) testing algorithm is now defined in the following six 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 $*$.

Continuous-Path Block Bootstrap (CBB) testing algorithm:

1. First calculate the centered residuals

$$\hat{U}_t = (X_t - \hat{\rho}_n X_{t-1}) - \frac{1}{n-1} \sum_{t=2}^n (X_t - \hat{\rho}_n X_{t-1}) \quad (2.1)$$

for $t = 2, 3, \dots, n$ where $\hat{\rho}_n = \hat{\rho}_n(X_1, X_2, \dots, X_n)$ is a consistent estimator of ρ based on the observed data $\{X_1, X_2, \dots, X_n\}$; see Remark 2.2 below.

Attention now focuses on the new variables \tilde{X}_t defined as follows:

$$\tilde{X}_t = \begin{cases} X_1 & \text{for } t = 1 \\ X_1 + \sum_{j=2}^t \hat{U}_j & \text{for } t = 2, 3, \dots, n. \end{cases}$$

2. Chose 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 CBB constructs a bootstrap pseudo-series X_1^*, \dots, X_l^* , where $l = kb + 1$, as follows.

3. Construction of the first bootstrap block of $b + 1$ observations. Set $X_1^* = X_1$ and

$$X_j^* := X_1 + [\tilde{X}_{i_0+j-1} - \tilde{X}_{i_0}]$$

for $j = 2, \dots, b + 1$. To elaborate:

$$\begin{aligned} X_2^* &:= X_1 + [\tilde{X}_{i_0+1} - \tilde{X}_{i_0}] \\ X_3^* &:= X_1 + [\tilde{X}_{i_0+2} - \tilde{X}_{i_0}] \\ &\vdots \\ X_{b+1}^* &:= X_1 + [\tilde{X}_{i_0+b} - \tilde{X}_{i_0}]. \end{aligned}$$

4. Construction of the $(m + 1)$ -th bootstrap block from the m -th block for $m = 1, 2, \dots, k - 1$.

Let

$$X_{mb+1+j}^* := X_{mb+1}^* + [\tilde{X}_{i_m+j} - \tilde{X}_{i_m}]$$

for $j = 1, \dots, b$. To elaborate:

$$\begin{aligned} X_{mb+2}^* &:= X_{mb+1}^* + [\tilde{X}_{i_m+1} - \tilde{X}_{i_m}] \\ X_{mb+3}^* &:= X_{mb+1}^* + [\tilde{X}_{i_m+2} - \tilde{X}_{i_m}] \\ &\vdots \\ X_{mb+1+b}^* &:= X_{mb+1}^* + [\tilde{X}_{i_m+b} - \tilde{X}_{i_m}]. \end{aligned}$$

5. Finally, compute the pseudo-statistic $\hat{\rho}^*$ which is nothing other than the statistic $\hat{\rho}_l$ based on the pseudo-data $\{X_1^*, \dots, X_l^*\}$.
6. Repeating steps 2–5 a great number of times (B times, say), we obtain the collection of pseudo-statistics $\hat{\rho}_1^*, \dots, \hat{\rho}_B^*$. As will be shown shortly, an empirical distribution (or just a histogram) based on the pseudo-statistics $\hat{\rho}_1^*, \dots, \hat{\rho}_B^*$ provides a consistent approximation of the distribution of $\hat{\rho}_n(X_1, \dots, X_n)$ 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 .

An intuitive way to understand the CBB construction is based on the discussion regarding Figure 1(c) in the Introduction and goes as follows: (i) construct a BB (block bootstrap) pseudo-series $\{X_t^*, t = 1, 2, \dots\}$ based on blocks of size equal to $b + 1$ taken from the series \tilde{X}_t ; (ii) shift the first block (of size $b + 1$) by an amount selected such that the bootstrap series starts off at the same point as the original series; (iii) shift the second BB block (of size $b + 1$) by another amount selected such that the first observation of this new bootstrap block matches exactly the last observation of the previous bootstrap block; (iv) join the two blocks but delete the last observation of the previous bootstrap block from the bootstrap series; (v) repeat parts (iii) and (iv) until all the generated BB blocks are used up. Note that a CBB series using block size b is associated to a BB construction with block size $b + 1$. This phenomenon is only due to the fact that we are dealing with discrete-time processes; it would not occur in a continuous-time setting. The reason for this is our step (iv) above: although we are effecting the matching of the first observation of a new bootstrap block to the last observation of the previous bootstrap block, it does not seem advisable to leave both occurrences of this common (matched) value to exist side-by-side; one of the two must be deleted as step (4) suggests.

Remark 2.1 Note that the CBB idea is not applied to the $\{X_t\}$ data; rather, it is applied to $\{\tilde{X}_t\}$ which is obtained by integrating the *centered* residuals \tilde{U}_t . The reason for this centering is that although the series $U_t = X_t - \rho X_{t-1}$ has a zero mean both under the null and under the alternative, the estimated innovations $\tilde{U}_t = X_t - \hat{\rho}_n X_{t-1}$ 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. Fortunately, recentering the innovations yields an easy fix-up; a similar necessity for residual centering has been recommended early on even in regular linear regression—see Freedman (1981).

Remark 2.2 The quantity $\hat{\rho}_n$ appearing in equation (2.1) is an appropriately chosen consistent estimator of the parameter ρ based on the data $\{X_1, X_2, \dots, X_n\}$. In particular, for the validity of the CBB we require that $\hat{\rho}_n$ satisfies

$$\hat{\rho}_n = \rho + O_P(n^{-(1+\delta(\rho))/2}) \quad (2.2)$$

where $\delta(\rho) = 1$ if $\rho = 1$, and $\delta(\rho) = 0$ if $\rho \neq 1$. Condition (2.2) is satisfied by many estimators; we elaborate with three specific examples:

Example 2.1 Let the parameter ρ have the meaning of the asymptotic lag-1 autocorrelation of series $\{X_t\}$, i.e., let

$$\rho = \lim_{t \rightarrow \infty} \frac{E X_t X_{t+1}}{E X_t^2}.$$

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

$$\frac{E X_t X_{t+1}}{E X_t^2} = 1 + O(1/t),$$

under the sole assumption that the series $\{U_t\}$ possesses a spectral density (which is guaranteed by either Condition A or B).

The above discussion showed that 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: $X_t = X_{t-1} + U_t$ where $\{U_t\}$ is stationary but with $Var(U_t) = 0$.

Let $\hat{\rho}_n = \hat{\rho}_{LS}$ be the (ordinary) least squares estimator obtained by regressing X_t on X_{t-1} given by

$$\hat{\rho}_{LS} = \frac{\sum_{t=2}^n X_t X_{t-1}}{\sum_{t=2}^n X_{t-1}^2}.$$

It is now well-known that the estimator $\hat{\rho}_{LS}$ above satisfies equation (2.2); see Brockwell and Davis (1991) for the stationary case, and Fuller (1996) or Phillips (1987) for the integrated case.

Example 2.2 We can similarly consider the parameter ρ signifying the asymptotic lag- k autocorrelation of the series $\{X_t\}$ for some fixed $k > 0$, i.e.,

$$\rho = \lim_{t \rightarrow \infty} \frac{E X_t X_{t+k}}{E X_t^2}.$$

Again under H_1 the series $\{X_t\}$ is stationary, and the limit is unnecessary. Similarly, under H_0 , we calculate that

$$\frac{E X_t X_{t+k}}{E X_t^2} = 1 + O(1/t),$$

using the fact that the series $\{U_t\}$ possesses a spectral density. Thus, it is apparent that H_0 implies $\rho = 1$; but is $\rho = 1$ equivalent to H_0 ? Technically speaking the answer is no, since a value of one for the lag- k autocorrelation of a series $\{X_t\}$ may mean either that $\{X_t\}$ is $I(1)$ (as in Example 2.1 above), or that $\{X_t\}$ is periodic with period k . If, however, the periodicity is ruled out, then $\rho = 1$ would effectively be equivalent to H_0 .

Now let $\hat{\rho}_n = \hat{\rho}_{LS,k}$ be the (ordinary) least squares estimator obtained by regressing X_t on X_{t-k} , i.e., let

$$\hat{\rho}_{LS,k} = \frac{\sum_{t=k+1}^n X_t X_{t-k}}{\sum_{t=k+1}^n X_{t-k}^2}.$$

It is now also true that equation (2.2) is satisfied; see Brockwell and Davis (1991) for the stationary case, and Hall (1989) for the unit root case where $\hat{\rho}_{LS,k}$ has been used as an instrumental variable estimator in the context of a specific model.

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, we have in this case that $U_t = X_t - X_{t-1}$ 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 in this case the integrated process $\{X_t\}$ can be expressed as $(1-L)(1+\sum_{j=1}^{\infty} \pi_j L^j)X_t = \varepsilon_t$, i.e., $\tilde{\pi}(z) = (1-z)\pi(z)$ has a unit root. If $\rho \neq 1$ then Condition A' implies that $X_t = -\sum_{j=1}^{\infty} \pi_j X_{t-j} + \varepsilon_t$ since in this case $\{X_t\}$ is by assumption linear and stationary, i.e., $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ and the power series $\Psi(z)$ has no zeros for $|z| \leq 1$. Therefore, Condition A' states that either is $\{X_t\}$ a stationary process possessing an infinite order autoregressive representation ($\rho \neq 1$) or $\{X_t\}$ is obtained by integrating such a process ($\rho = 1$).

If the process $\{X_t\}$ satisfies Condition A', then an alternative representation (identity) of X_t useful for testing purposes is given in the following lemma; cf. also Fuller (1996) for a similar representation in the case that $\{X_t\}$ is a finite order autoregressive process.

Lemma 2.1 *If $\{X_t\}$ satisfies Condition A' then*

$$X_t = \rho X_{t-1} + \sum_{j=1}^{\infty} a_j (X_{t-j} - X_{t-j-1}) + \varepsilon_t \quad (2.3)$$

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 $\rho = 1$ and $a_j = \pi_j$ for $j = 1, 2, \dots$, while if $\{X_t\}$ is stationary then

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

To see why the above statement is true note that in the unit root integrated case, representation (2.3) is obviously valid for $\rho = 1$ and $a_j = \pi_j$, since in this case $X_t = X_{t-1} + U_t$

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

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

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

$$\begin{aligned} \sum_{j=1}^{\infty} |a_j| &= \sum_{j=1}^{\infty} \left| \sum_{s=j+1}^{\infty} \pi_j \right| \\ &\leq \sum_{j=1}^{\infty} j|\pi_j| < \infty. \end{aligned}$$

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 (2.3) is given by $\rho = -\sum_{j=1}^{\infty} \pi_j$. Furthermore, this parameter always satisfies

$$\rho \leq 1. \tag{2.4}$$

To see this, note that the condition $\pi(z) \neq 0$ 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 $\lim_{z \rightarrow 1} \pi(z) = 1 + \sum_{j=1}^{\infty} \pi_j \geq 0$ which is just (2.4). Note that $\rho = 1$ implies that $\pi(z)$ has a unit root.

Example 2.3 Assume that the underlying process satisfies Condition A' and let $\hat{\rho}_n = \hat{\rho}_{DF}$ be the so-called augmented Dickey-Fuller estimator of ρ and which is obtained by fitting a truncated version of (2.3) to the observed series, i.e., by fitting the model

$$X_t = \rho X_{t-1} + \sum_{i=1}^p a_i (X_{t-i} - X_{t-i-1}) + \varepsilon_t. \tag{2.5}$$

To ensure consistency of $\hat{\rho}_{DF}$ the order $p = p(n)$ in the above equation is allowed to increase to infinity at some appropriate rate as the sample size n increases; see Said and Dickey (1984). Note that for fixed p , equation (2.5) is the set-up considered by Dickey and Fuller (1979) for finite order autoregressive processes of 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 $\hat{\rho}_{DF} = 1 + O(n^{-1})$. In Section 8 we extend their result by showing that if $\{X_t\}$ is a stationary process having the representation (2.3), then $\hat{\rho}_{DF} = \rho + O_P(n^{-1/2})$, where $\rho = -\sum_{j=1}^{\infty} \pi_j$. Therefore, the estimator $\hat{\rho}_{DF}$ considered in this example obeys the stochastic behavior (2.2). Recall that for the choice of ρ discussed here it follows from (2.4) that the value of ρ under the alternative is *not* necessarily in the interval $[-1, 1]$; whereas $\rho \leq 1$ always, it may be the case that $\rho < -1$. For instance, if the true process $\{X_t\}$ is the first order moving average process $X_t = \varepsilon_t - \theta\varepsilon_{t-1}$ 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 every $\theta \in (1/2, 1)$.

Note that while (2.3) is valid for all values of ρ considered, Condition A' does not necessarily imply that the stationary process $\{U_t\}$ has an AR(∞) representation with absolutely summable coefficients when 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 $X_t = -\sum_{j=1}^{\infty} \pi_j X_{t-j} + \varepsilon_t$. Now, if $U_t = \sum_{j=1}^{\infty} c_j U_{t-j} + \varepsilon_t$ exists then using the definition $U_t = X_t - \rho X_{t-1}$ and rearranging terms it follows that $c_0 = 1$ and $c_j = \pi_j + \rho c_{j-1}$ for $j = 1, 2, \dots$ which implies that $|c_j| \neq 0$ if $\rho \leq -1$.

What is apparent from the above three examples 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 all examples above, 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 2.3 Note that the estimator $\hat{\rho}_n$ used in calculating the residuals in step 1 of the CBB algorithm is the *same* estimator whose distribution is required in order to perform the test of $H_0 : \rho = 1$. In other words, the practitioner first selects a suitable statistic for the unit root test (making sure (2.2) is satisfied), and then uses the *same* statistic in constructing the CBB residuals. The reason this pairing is required can be better explained by the following simple argument.

Having decided on using a particular estimator $\hat{\rho}_n$ of the parameter ρ figuring in equation (1.1), recall that our (ideal) objective is to approximate (by bootstrap simulations) the distribution of $\hat{\rho}_n$ under H_0 ; in other words, we would like to be able to simulate the distribution of $\hat{\rho}_n$ in the context of equation (1.1) with $\rho = 1$. To do this, we would like to simulate $\{X_t\}$ data from the equation

$$X_t = X_{t-1} + U_t, \tag{2.6}$$

and then use the simulated data to compute the distribution of $\hat{\rho}_n$.

We now go back to the CBB testing construction to see how the CBB attempts to approximate the above simulation. First recall that step 1 of the CBB algorithm effectively sets

$$\tilde{X}_t = \tilde{X}_{t-1} + \hat{U}_t.$$

But if we momentarily neglect the residual re-centering of equation (2.1) we see that

$$\hat{U}_t \simeq X_t - \hat{\rho}_n X_{t-1}.$$

Combining the above two equations we get:

$$\tilde{X}_t \simeq \tilde{X}_{t-1} + X_t - \rho X_{t-1} - (\rho - \hat{\rho}_n) X_{t-1}$$

i.e.,

$$\tilde{X}_t \simeq \tilde{X}_{t-1} + U_t + (\hat{\rho}_n - \rho) X_{t-1}. \quad (2.7)$$

Comparing equation (2.7) to our “target” equation (2.6), it is apparent that the only way we will have our $\{\tilde{X}_t\}$ series mimicking correctly the $\{X_t\}$ series of equation (2.6), i.e., the $\{X_t\}$ series under the null hypothesis H_0 , is to have the term $(\hat{\rho}_n - \rho) X_{t-1}$ be negligible *under all possibilities for the true value of ρ* ; this is actually guaranteed by our equation (2.2). To drive this point further, note that 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 is what the CBB testing algorithm succeeds in doing.

Remark 2.4 The simple choice $\hat{\rho}_n \equiv 1$ in step 1 of the CBB algorithm was used in Paparoditis and Politis (2000) in the context of series that are *known* to have unit root nonstationarity. However, in the context of unit root *testing* discussed presently, such a choice of $\hat{\rho}_n$ is inappropriate. This is so because by (2.6) and (2.7), $\hat{\rho}_n$ must be a consistent estimator of ρ for *all* possible values of this parameter; in particular, $\hat{\rho}_n$ must satisfy eq. (2.2).

3 A Functional Limit Theorem for the CBB Partial Sum Process

The asymptotic properties of the CBB 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

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

and

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

where $U_1^* \equiv X_1$, $U_t^* = X_t^* - X_{t-1}^*$ for $t = 2, 3, \dots, l$ and $\sigma^{*2} = \text{Var}^*(l^{-1/2} \sum_{j=1}^l U_j^*)$. 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, \dots, X_n , the CBB 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 $T_n^* = T_n^*(X_1^*, X_2^*, \dots, X_n^*)$ is a random sequence based on bootstrap sample $X_1^*, X_2^*, \dots, X_n^*$ and G 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 $U_t = X_t - \rho X_{t-1}$ satisfies Condition A or Condition B and let $\hat{\rho}_n$ be an estimator of ρ such that equation (2.2) is satisfied. If $b \rightarrow \infty$ but $b/\sqrt{n} \rightarrow 0$ as $n \rightarrow \infty$, then*

$$S_1^* \Rightarrow W \quad \text{in probability.}$$

This basic result together with a bootstrap version of the continuous mapping theorem will enable us to apply the CBB-testing procedure 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 estimator $\hat{\rho}_n$. By the above theorem, we expect that the CBB testing procedure will succeed in approximating the distribution of any such statistic $\hat{\rho}_n$ under the null hypothesis of unit root integration 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 $\hat{\rho}_n$ of ρ satisfies equation (2.2). Such applications of the CBB testing procedure to two popular statistics are given in the next section.

4 Applications to Unit Root Testing

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

In the second application, validity of the CBB testing procedure in approximating the null distribution of the so-called augmented Dickey-Fuller test statistic based on the regression (2.5) is shown; cf. Fuller (1996), Dickey and Fuller (1991), Said and Dickey (1984) and our Example 2.3. This test is considered here because of its popularity and because of its good power properties reported in simulation studies; cf. Schwert (1989), DeJong et. al (1992)). See also Elliot et al. (1996) for a theoretical discussion.

Although large sample Gaussian approximations of their distribution under the null are well-known in the literature for both statistics considered, there are several reasons for expecting the bootstrap to perform better than such large sample approximations. First of all, we expect the bootstrap to be able to mimic more closely the finite sample behavior of the test statistics

considered under the null hypothesis. For instance, the bootstrap critical values are obtained here without imposing any distributional assumptions on the stationary process $\{U_t\}$ driving the random walk while the critical values commonly reported in the literature are based on Gaussian process assumptions; cf. Fuller (1996), Hamilton (1994). Furthermore, it is well known that the asymptotic distribution of the above statistics depends on nuisance (and difficult to estimate) parameters like the spectral density of the process at the origin. The CBB procedure manages to estimate such quantities automatically making an explicit nonparametric estimation superfluous.

Note that the asymptotic validity of the CBB testing procedure can be easily established also for studentized versions of the statistics considered in this section, i.e., for statistics based on the difference between estimator and hypothetical value of the parameter ρ divided by the estimated standard error of the estimator. Such an extension of our procedure is straightforward and we do not discuss it here in detail. Studentized versions of the statistics considered in this paper have attracted attention in the literature mainly because the asymptotic distribution of their ordinary (not studentized) versions depends on nuisance parameters.

4.1 Statistics Based on Ordinary Least Squares

Here we are concerned with the least squares estimator of the parameter ρ in the regression

$$X_t = \rho X_{t-1} + e_t. \quad (4.1)$$

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 are considered here, i.e., we show validity of the CBB procedure in approximating the distribution of the least squares estimator of the parameter ρ if the regression

$$X_t = \beta + \rho X_{t-1} + e_t \quad (4.2)$$

is fitted to the observed series. Let $\hat{\rho}_{LS}$ denote the least squares estimator of ρ in (4.1) and $\hat{\rho}_{LS,C}$ in (4.2). To approximate the distribution of $\hat{\rho}_{LS}$ we apply the CBB algorithm given in Section 2 by using the estimator $\hat{\rho}_{LS}$ in order to calculate the centered residuals \hat{U}_t in the first step. The pseudo-statistic $\hat{\rho}^*$ computed in step 5 is then given by the least squares estimator of the parameter of X_{t-1}^* obtained by regressing X_t^* on X_{t-1}^* . 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^* or X_{t-1}^* . The estimator of the coefficient of X_{t-1}^* in this regression is denoted in the following by $\hat{\rho}_{LS,C}^*$.

The following theorem shows the asymptotic validity of the CBB testing procedure in approximating the distribution of both least squares estimators considered.

Theorem 4.1 *Assume that the process $\{X_t\}$ satisfies Condition A or Condition B. If $b \rightarrow \infty$ but $b/\sqrt{n} \rightarrow 0$ as $n \rightarrow \infty$, then*

$$(i) \sup_{x \in \mathbb{R}} \left| P^* \left(l(\hat{\rho}_{LS}^* - 1) \leq x \mid X_1, X_2, \dots, X_n \right) - P_0 \left(n(\hat{\rho}_{LS} - 1) \leq x \right) \right| \rightarrow 0 \text{ in probability}$$

and

(ii) $\sup_{x \in \mathbb{R}} \left| P^* \left(l(\hat{\rho}_{LS,C}^* - 1) \leq x \mid X_1, X_2, \dots, X_n \right) - P_0 \left(n(\hat{\rho}_{LS,C} - 1) \leq x \right) \right| \rightarrow 0$ in probability,

where P_0 denotes the probability measure under the null hypothesis that $\{X_t\}$ is generated by integrating $\{U_t\}$.

Note that in the above theorem we do not assume that the observed series is a realization of a unit root integrated process, i.e., the CBB testing procedure approximates the distribution of the least squares estimators considered under the null hypothesis of unit root integration whether the observed process satisfies the null hypothesis or not.

Now, based on the above theorem, an (asymptotically) α -level CBB-based test of the null hypothesis of unit root integration can be obtained as follows: Reject the null hypothesis if $n(\hat{\rho}_{LS} - 1) < C_\alpha^*$ where C_α^* denotes the α -quantile of the bootstrap distribution of $l(\hat{\rho}_{LS}^* - 1)$, i.e., $C_\alpha^* = \inf\{u; P^*(l(\hat{\rho}_{LS}^* - 1) \leq u) \geq \alpha\}$. Similarly, a test based on the statistic $\hat{\rho}_{LS,C}$ can be constructed by rejecting the null hypothesis if $n(\hat{\rho}_{LS,C} - 1)$ is less than $C_{\alpha,\mu}^*$, where $C_{\alpha,\mu}^*$ denotes the α -quantile of the bootstrap distribution of $l(\hat{\rho}_{LS,C}^* - 1)$.

4.2 Dickey-Fuller Type Statistics

Consider now the problem of approximating the distribution of the estimator $\hat{\rho}_{DF}$ under the null hypothesis, where $\hat{\rho}_{DF}$ denotes the least squares estimator obtained by fitting the regression equation (2.5) to the observed series X_1, X_2, \dots, X_n . To do this, define the centered differences

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

$t = 2, 3, \dots, n$. To estimate the distribution of $\hat{\rho}_{DF}$ we apply the CBB algorithm as follows: We use the estimator $\hat{\rho}_{DF}$ to calculate the centered residuals \hat{U}_t in the first step of the algorithm and then generate the CBB series $X_1^*, X_2^*, \dots, X_l^*$ following steps 2 to 4. Additionally to the CBB series X_t^* , we also generate a pseudo-series of l centered differences denoted by $D_1^*, D_2^*, \dots, D_l^*$ as follows: For the first block of $b+1$ observations we set $D_1^* = 0$ and

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

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

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

where $j = 1, 2, \dots, b$. We then calculate the regression of X_t^* on X_{t-1}^* and on $D_{t-1}^*, D_{t-2}^*, \dots, D_{t-p}^*$. 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 motivate the above use of the CBB algorithm to approximate the distribution of $\hat{\rho}_{DF}$, recall our target regression (2.5) which relates X_t on X_{t-1} and on the lagged differences $X_{t-j} - X_{t-j-1}$, $j = 1, 2, \dots, p$. 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 $X_{t-j}^* - X_{t-j-1}^*$, $j = 1, 2, \dots, p$, will mimic the regression of X_t on X_{t-1} and on U_{t-j} , $j = 1, 2, \dots, p$. This, however, coincides with our target regression (2.5) only if $\rho = 1$, i.e., only if the observed series is indeed unit root integrated. Furthermore, as we have seen in Example 2.3, such an infinite order autoregressive representation for the process $\{U_t\}$ may not exist if $\rho \neq 1$.

Now, to understand where the definition of the new bootstrap variables D_t^* comes from, consider the bootstrap observations in the $(m+1)$ th block given by X_{mb+1+s}^* where $s \in \{1, 2, \dots, b\}$. Here we have $X_{mb+1+s}^* = X_{mb+s}^* + \hat{U}_{im+s+1}$ and \hat{U}_{im+s+1} behaves for large n like U_{im+s+1} which by (2.3) depends on the lagged differences $X_{im+s+1-j} - X_{im+s-j}$, $j = 1, 2, \dots$. Thus the bootstrap analogue of (2.5) will be to regress X_{mb+1+s}^* on X_{mb+s}^* and on $X_{im+s+1-j} - X_{im+s-j}$, $j = 1, 2, \dots, p$. Note that $D_{im+1+s-j}$ is just a centered version of $X_{im+s+1-j} - X_{im+s-j}$.

To approximate the distribution of $\hat{\rho}_{DF,\mu}$, i.e., of the least squares estimator of ρ in

$$X_t = \beta + \rho X_{t-1} + \sum_{i=1}^p a_i (X_{t-i} - X_{t-i-1}) + \varepsilon_t, \quad (4.3)$$

we include a constant term in the corresponding regression fitted to the pseudoseries $\{X_t^*, D_t^*, t = 1, 2, \dots, l\}$. The so obtained least squares estimator of the coefficient of X_{t-1}^* is denoted in the following by $\hat{\rho}_{DF,C}^*$.

Since in both cases discussed here the pseudoseries $X_1^*, X_2^*, \dots, X_n^*$ is generated in a non-parametric way using the CBB procedure and not the parametric form given in (2.5), we expect the bootstrap to be able to mimic correctly the ‘truncation effect’ on the distribution of $\hat{\rho}_{DF}$ and $\hat{\rho}_{DF,C}$ which is due to the fact that only a finite order version of (2.3) is fitted to the series. In other words, we expect the results based on the CBB testing procedure to be less sensitive with respect to the choice of the parameter p .

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

$$(i) \sup_{x \in \mathbb{R}} \left| P^* \left((l-p)(\hat{\rho}_{DF}^* - 1) \leq x \mid X_1, X_2, \dots, X_n \right) - P_0 \left((n-p)(\hat{\rho}_{DF} - 1) \leq x \right) \right| \rightarrow 0 \text{ in probability.}$$

and

$$(ii) \sup_{x \in \mathbb{R}} \left| P^* \left((l-p)(\hat{\rho}_{DF,C}^* - 1) \leq x \mid X_1, X_2, \dots, X_n \right) - P_0 \left((n-p)(\hat{\rho}_{DF,C} - 1) \leq x \right) \right| \rightarrow 0 \text{ in probability.}$$

Here P_0 denotes the probability measure under the null hypothesis that $\{X_t\}$ is generated by integrating $\{U_t\}$.

Some remarks on the power of the CBB testing procedure are in order. By equation (8.18) of Lemma 8.3 and the arguments used there we have that

$$\sqrt{n-p}(\hat{\rho}_{DF} - \rho) = \frac{1}{\sqrt{n-p}} \sum_{t=p+2}^n V_{t-1} \varepsilon_t + o_P(1) \quad (4.4)$$

where $V_{t-1} = e_1' V^{-1}(p) Y_{t-1}(p)$, $e_1 = (1, 0, \dots, 0)'$, $V(p) = E(Y_{t-1}(p) Y_{t-1}'(p))$ and $Y_{t-1}(p) = (X_{t-1}, X_{t-1} - X_{t-2}, \dots, X_{t-p} - X_{t-p-1})'$. By (4.4) and along the same lines as in the proof of Theorem 4 of Berk (1974) we get the following result.

Lemma 4.1 *Assume that the process $\{X_t\}$ satisfies Condition A' with $\rho \neq 1$ and that $p \rightarrow \infty$ as $n \rightarrow \infty$ such that $p^3/n \rightarrow 0$ and $\sqrt{n} \sum_{j=p+1}^{\infty} |a_j| \rightarrow 0$. Then*

$$\sqrt{n-p} \frac{\hat{\rho}_{DF} - \rho}{\sigma_{DF}} \Rightarrow N(0, 1), \quad (4.5)$$

where $\sigma_{DF}^2 = \sigma_\varepsilon^2 e_1' V^{-1}(p) e_1$.

For $\alpha \in (0, 1)$ let C_α and C_α^* be the α -quantiles of the distribution of $(n-p)(\hat{\rho}_{DF} - 1)$ and $(l-p)(\hat{\rho}_{DF}^* - 1)$ respectively under the null hypothesis of unit root integration. By Theorem 4.2 we have that $C_\alpha^* \rightarrow C_\alpha$ in probability and, therefore, that

$$\left| P\left((n-p)(\hat{\rho}_{DF} - 1) \leq C_\alpha^*\right) - P\left((n-p)(\hat{\rho}_{DF} - 1) \leq C_\alpha\right) \right| \rightarrow 0.$$

Now, since

$$\begin{aligned} P\left((n-p)(\hat{\rho}_{DF} - 1) \leq C_\alpha\right) &= P\left((n-p)(\hat{\rho}_{DF} - \rho) \leq C_\alpha - (n-p)(\rho - 1)\right) \\ &= P\left(\frac{\sqrt{n-p}(\hat{\rho}_{DF} - \rho)}{\sigma_{DF}} \leq \frac{C_\alpha}{\sqrt{n-p}\sigma_{DF}} - \frac{\sqrt{n-p}(\rho - 1)}{\sigma_{DF}}\right) \end{aligned}$$

we see –using (4.5)– that if $\rho \neq 1$ then

$$\lim_{n \rightarrow \infty} P\left((n-p)(\hat{\rho}_{DF} - 1) \leq C_\alpha^*\right) = \Phi(\infty) = 1, \quad (4.6)$$

since $\rho = -\sum_{j=1}^{\infty} \pi_j < 1$. Here Φ denotes the distribution function of the standard normal. Thus, under our assumptions, the power of the CBB based Dickey-Fuller type test approaches unity as the sample size increases.

In concluding this section we mention that the above consistency result of the CBB testing procedure can be also established for the case of the ordinary least squares estimator $\hat{\rho}_{LS}$ discussed in the previous section. Furthermore, it is not difficult to show that the tests considered have nondegenerate power against local alternatives converging to the null at the rate n^{-1} , i.e., if the true parameter is given by $\rho_n = 1 - c/n$ for some $c > 0$.

5 Extensions: The Case of Nonzero Mean

In this section we briefly indicate how the proposed CBB testing procedure can be extended to deal with the case where the observed process has a drift and the regression equation fitted includes a constant or a (linear) time trend component. To elaborate, consider the situation

where $\{X_t\}$ is a stochastic process which is either stationary around a (possibly nonzero) mean or its first difference series $\{Y_t\}$ with $Y_t = X_t - X_{t-1}$, is stationary with a (possibly nonzero) mean. Note that in the latter case (and if the first difference series is an i.i.d. sequence) we are simply in the case of a random walk with drift.

To elaborate on the appropriate modification of the CBB procedure to deal with this case, consider the process $\{U_t\}$ defined by

$$U_t := X_t - \beta - \rho X_{t-1} \quad (5.1)$$

where as before, ρ is a parameter such that $\rho = 1$ is equivalent to H_0 whereas $\rho \neq 1$ is equivalent to H_1 , and the constant β is defined such that $E(U_t) = 0$. The case $\beta = 0$ is the one considered in the previous sections. To estimate the distribution of an estimator $\hat{\rho}_n$ of ρ under the null hypothesis that $\rho = 1$, our procedure should be able to generate unit root integrated pseudo-series $X_1^*, X_2^*, \dots, X_l^*$ with a nonvanishing drift component. To do this, the following more general CBB testing algorithm can be applied:

1. First calculate the centered residuals

$$\hat{U}_t = (X_t - \hat{\beta} - \hat{\rho}_n X_{t-1}) - \frac{1}{n-1} \sum_{t=2}^n (X_t - \hat{\beta} - \hat{\rho}_n X_{t-1}) \quad (5.2)$$

for $t = 2, 3, \dots, n$ where $\hat{\beta}$ is \sqrt{n} -consistent estimator of β and $\hat{\rho}_n$ an estimator of ρ satisfying (2.2). Define now new variables \tilde{X}_t as follows:

$$\tilde{X}_t = \begin{cases} X_1 & \text{for } t = 1 \\ X_1 + \sum_{j=2}^t \hat{U}_j & \text{for } t = 2, 3, \dots, n. \end{cases}$$

2. Chose a positive integer $b (< n)$ and k as in step 2 of the CBB algorithm for the case $\beta = 0$.
3. For the first block of $b + 1$ observations set $X_1^* = X_1$ and

$$X_j^* := X_1 + (j-1)\hat{\beta} + [\tilde{X}_{i_0+j-1} - \tilde{X}_{i_0}]$$

for $j = 2, \dots, b + 1$.

4. To construct the $(m + 1)$ -th bootstrap block from the m -th block for $m = 1, 2, \dots, k - 1$ let

$$X_{mb+1+j}^* := X_{mb+1}^* + j\hat{\beta} + [\tilde{X}_{i_m+j} - \tilde{X}_{i_m}]$$

for $j = 1, \dots, b$.

5. Compute the pseudo-statistic $\hat{\rho}^*$ which is nothing other than the statistic $\hat{\rho}_l$ of interest based on the pseudo-data $\{X_1^*, \dots, X_l^*\}$.

The above version of the CBB algorithm can be used, for instance, to approximate the distribution of the least squares estimator of the coefficient of X_{t-1} under the null hypothesis that $\rho = 1$, if the model (4.2) or the model

$$X_t = \beta_0 + \beta_1 t + \rho X_{t-1} + e_t \quad (5.3)$$

is fitted to the observed series X_1, X_2, \dots, X_n . Similarly, the statistic of interest may be the least squares estimator of ρ in (4.3) or in

$$X_t = \beta_0 + \beta_1 t + \rho X_{t-1} + \sum_{i=1}^p a_i (X_{t-i} - X_{t-i-1}) + \varepsilon_t. \quad (5.4)$$

Note that in all cases above we allow for a nonzero mean β in (5.1). Thus the situation is different to that considered in the previous section since the distribution of the least squares estimators is affected; cf. Hamilton (1994) or Fuller (1996). Nevertheless, the theory developed can be easily extended to establish the asymptotic validity of the above generalized CBB proposal. For instance, denote by $\hat{\rho}_{LS,T}$ the least squares estimator of ρ in (5.3) and by $\hat{\rho}_{LS,T}^*$ the corresponding estimator using the bootstrap series $X_1^*, X_2^*, \dots, X_n^*$ generated according to the above CBB algorithm. The following result can then be established.

Theorem 5.1 *Assume that the process $\{X_t\}$ satisfies Condition B and assume that β in (5.1) is different from zero. If $b \rightarrow \infty$ but $b/\sqrt{n} \rightarrow 0$ as $n \rightarrow \infty$, then*

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

in probability, where P_0 denotes the probability measure under the null hypothesis that $\{X_t\}$ is generated by $X_t = \beta + X_{t-1} + U_t$.

6 Small Sample Performance

A small simulation study was conducted to evaluate the finite-sample performance of our CBB bootstrap testing procedure. 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 $\phi = 0.9$ and $\phi = 0.85$ corresponds to a stationary series $\{X_t\}$. Regarding the MA parameter θ , the values $-0.8, 0$, and 0.8 were chosen; $\theta = 0.8$ corresponds to a positive dependence, which—in combination with our ϕ choices—can yield data that could conceivably be mistaken for a unit root if the sample size is small enough; $\theta = 0$ corresponds to either a random walk with i.i.d. errors, or a stationary AR(1) model (according to whether $\phi = 1$ or $\phi \in \{0.85, 0.9\}$). Finally, the case where $\phi = 1$ (unit root) and $\theta = -0.8$ deserves special attention; here the MA polynomial has a root close to unity which for small

sample sizes yields series that can easily be mistaken for i.i.d., although in reality they are integrated.

The simulations were performed by generating a number of $M = 1000$ true $\{X_t\}$ series each of length $n + 200$ where the first 200 observations were discarded; we chose $n = 100$ and $n = 200$. From each generated data series the CBB was called to perform a α -level test of the unit root hypothesis H_0 . The test statistics that were used were based on the LS estimator $n(\hat{\rho}_{LS} - 1)$ of Example 2.1, and on Dickey-Fuller estimator $(n - p)(\hat{\rho}_{DF} - 1)$ of Example 2.3. The CBB was conducted by generating $B = 1000$ bootstrap series (for each true series) in order to perform the required Monte Carlo approximations.

In Tables 1 we report the empirical rejection probabilities of the CBB unit root test based on the least squares estimator $\hat{\rho}_{LS}$ with nominal level $\alpha = 0.05$ under different settings of the ARMA parameters ϕ and θ , different sample sizes, and different choices of the blocksize b . In Table 2 similar results are reported for the statistic $\hat{\rho}_{DF}$ where we now vary the order p of the lagged differences and—for computational reasons—fix the blocksize at $b = 8$ for the sample size $n = 100$ and $b = 10$ for $n = 200$.

The results of our small simulation are encouraging. Note that the case $\phi = 1$ in both Tables 1 and 2 correspond to rejections when H_0 is true; thus, we expect the entries there to be close to the nominal level $\alpha = 0.05$ of the test. This is indeed what appears to be happening with the exception of the case where $\theta = -0.8$. However, even in this case the size distortion for the ordinary least squares estimator is less severe than the one based on the asymptotic distribution; cf. Phillips and Perron (1988). Nevertheless, the case $\theta = -0.8$ is a well-known problematic situation in which—as discussed above—a practical ‘cancellation’ of the AR unit root with the MA ‘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 ‘cancellation’ 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).

Returning to our simulation results, note that the cases $\phi = 0.9$ and $\phi = 0.85$ correspond to rejections when H_0 is not true and should be rejected; thus, the corresponding entries measure the *power* of our test against specific alternatives, and we expect/hope the entries there to be close to one. This is indeed what is happening: the test CBB has power which increases with sample size n (as expected) when the alternative remains fixed. Quite remarkably, in the $n = 200$ case the power of our test is very high and quite close to one under *all* alternatives considered, even in the $\theta = 0.8$ case which is a case with appreciable positive correlation. Comparing the power results in both tables with those reported in other simulation studies for the same model and the same parameters, see for instance Philips and Perron (1988), Schwert (1989), Ng and Perron (1995), it seems that for both statistics considered here the power of CBB based tests outperforms that of the same tests based on asymptotic critical values. Furthermore, compared with the results of the aforementioned published simulation studies, the power of the CBB based augmented Dickey-Fuller type statistic $(n - p)(\hat{\rho}_{DF} - 1)$ reported in Table 2 seems to be less sensitive with respect to the choice of the order p .

Please Insert Table 1 and Table 2 about here

7 Conclusions

In this paper we have proposed a new resampling procedure, the continuous path block bootstrap testing procedure, that generates unit root integrated pseudoseries which retain the weak dependence structure of the observed series. The procedure manages to accurately capture the distribution of several unit root test statistics used in econometrics. Although we have restricted our considerations to two such popular test statistics, the theory presented is general enough in that it allows the application of the new resampling methodology to other test statistics too. Our theoretical results, as well as the limited empirical evidence presented in this paper, largely support the hypothesis that the new CBB testing procedure is a useful alternative to large sample Gaussian approximations commonly used in the econometric analysis of nonstationary time series. Notably, the idea of generating continuous path pseudoseries underlying the CBB methodology presented in this paper is not limited to the block bootstrap but can also be applied to other resampling techniques designed for stationary weak depended processes. For instance, preliminary results indicate that a continuous-path version of the stationary bootstrap of Politis and Romano (1994) can also be successfully employed in the context of unit root testing.

8 Auxiliary Results and Proofs

Proof of Theorem 3.1 Recall the definition of U_t^* and verify that the CBB random variable X_t^* is given by

$$\begin{aligned} X_t^* &= U_1^* + U_2^* + U_3^* + \dots + U_t^* \\ &= X_1 + \sum_{m=0}^{\lfloor (t-2)/b \rfloor} \sum_{s=1}^{\min\{b, t-mb-1\}} \widehat{U}_{i_m+s}, \end{aligned} \quad (8.1)$$

where \widehat{U}_t is defined in (2.1). For $0 \leq r \leq 1$ and by the construction of the CBB series, we have

$$\begin{aligned} S_l^*(r) &= \frac{1}{\sqrt{l}} \sum_{j=1}^{\lfloor lr \rfloor} U_j^* / \sigma^* \\ &= \frac{1}{\sqrt{l}} X_1 / \sigma^* + \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^B \widehat{U}_{i_m+s} / \sigma^* \end{aligned}$$

where $M_r = \lfloor (\lfloor lr \rfloor - 2)/b \rfloor$ and $B = \min\{b, \lfloor lr \rfloor - mb - 1\}$. Recall that

$$\widehat{U}_t = X_t - \hat{\rho}_n X_{t-1} - \frac{1}{n-1} \sum_{\tau=2}^n (X_\tau - \hat{\rho}_n X_{\tau-1}). \quad (8.2)$$

By Lemma 8.1 we have that $\sigma^{*2} \rightarrow \sigma^2$ 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

$$S_l^*(r) = \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^b \widehat{U}_{i_m+s} / \sigma^* - \frac{1}{\sqrt{l}} \sum_{s=B+1}^b \widehat{U}_{i_{M_r}+s} / \sigma^* + O_P(l^{-1/2}) \quad (8.3)$$

and

$$\sup_{0 \leq r \leq 1} \left| \frac{1}{\sqrt{l}} \sum_{s=B+1}^b \widehat{U}_{i_{M_r+s}} / \sigma^* \right| = O_P(k^{-1/2}) \quad (8.4)$$

we consider in the following only the first term on the right hand side of equation (8.3). We first show that uniformly in r

$$\left| \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^b \widehat{U}_{i_m+s} - \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^b (U_{i_m+s} - E^* U_{i_m+s}) \right| \rightarrow 0 \quad (8.5)$$

in probability. To establish (8.5) verify that

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

Now, if $\rho \neq 1$ then $\hat{\rho} - \rho = O_P(n^{-1/2})$ and, therefore,

$$(\hat{\rho} - \rho) \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^b (X_{i_m+s-1} - \frac{1}{n-1} \sum_{\tau=2}^n X_{\tau-1}) = O_{P^*}(n^{-1/2}) \quad (8.6)$$

uniformly in r , since $\{X_t\}$ is stationary. Furthermore, if $\rho = 1$, i.e., if $\{X_t\}$ is unit root integrated, then $\hat{\rho} - \rho = O_P(n^{-1})$ and

$$(\hat{\rho} - \rho) \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^b (X_{i_m+s-1} - \frac{1}{n-1} \sum_{\tau=2}^n X_{\tau-1}) = O_{P^*}(b^{1/2} n^{-1/2}) \quad (8.7)$$

uniformly in r since in this case we have that

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

To see the above statement note that

$$\begin{aligned} &E^* \left[\sum_{s=1}^b (X_{i_m+s-1} - \frac{1}{n-1} \sum_{\tau=2}^n X_{\tau-1}) \right] \\ &= \sum_{s=1}^b \left(\frac{1}{n-b} \sum_{t=1}^{n-b} X_{t+s-1} - \frac{1}{n-1} \sum_{\tau=2}^n X_{\tau-1} \right) \\ &= \frac{1}{(n-b)(n-1)} \left[\sum_{s=1}^b \left((n-1) \sum_{t=1}^{s-1} X_t + \sum_{t=n-b+s}^{n-1} X_t \right) + b(b-1) \sum_{\tau=2}^n X_{\tau-1} \right] \\ &= O_P(b^2(n-b)^{-1/2}) + O_P(b^2 n^{1/2} (n-b)^{-1}) \end{aligned} \quad (8.8)$$

and that

$$E^* \left[\sum_{s=1}^b \left(X_{i_m+s-1} - \frac{1}{n-1} \sum_{\tau=2}^n X_{\tau-1} \right) \right]^2 = O_P(b^2(n-b)). \quad (8.9)$$

Therefore, and because of the independence of the bootstrap blocks, we get

$$\begin{aligned} E^*[T_n^*]^2 &= \frac{1}{l} \sum_{m=0}^{M_r} E^* \left(\sum_{s=1}^b \left(X_{i_m+s-1} - \frac{1}{n-1} \sum_{\tau=2}^n X_{\tau-1} \right) \right)^2 \\ &\quad + \frac{1}{l} \sum_{m_1=0}^{M_r} \sum_{m_2=0, m_1 \neq m_2}^{M_r} E^* \left(\sum_{s=1}^b \left(X_{i_{m_1}+s-1} - \frac{1}{n-1} \sum_{\tau=2}^n X_{\tau-1} \right) \right) \\ &\quad \times E^* \left(\sum_{s=1}^b \left(X_{i_{m_2}+s-1} - \frac{1}{n-1} \sum_{\tau=2}^n X_{\tau-1} \right) \right) \\ &= O_P(kb^2(n-b)l^{-1}) + O_P(k^2b^4l^{-1}(n-b)^{-1}) \\ &= O_P(b^2k) + O_P(kb^3(n-b)^{-1}). \end{aligned}$$

From (8.6) and (8.7) it follows that uniformly r ,

$$\left| \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^b \hat{U}_{i_m+s} - \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^b \left(U_{i_m+s} - \frac{1}{n-1} \sum_{\tau=2}^n U_{\tau} \right) \right| \rightarrow 0 \quad (8.10)$$

in probability. Now, by the same arguments as those leading to (8.8) and (8.9) we get

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

in probability, uniformly in r . This together with (8.10) establishes (8.5).

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

$$\frac{1}{\sigma^*} \frac{1}{\sqrt{l}} \sum_{m=0}^{M_r} \sum_{s=1}^b (U_{i_m+s} - E^* U_{i_m+s}), \quad (8.11)$$

to the Brownian motion W on $[0, 1]$. For this note first that $[\lfloor lr \rfloor / b] = [kr]$ and therefore we can consider instead of (8.11) the asymptotically equivalent statistic

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

Note that the above expression differs from (8.11) solely by the fact that the first sum is up to $\lfloor lr \rfloor / b$ instead of $(\lfloor lr \rfloor - 2) / b$ terms. The above statistic can be written in the form

$$\frac{1}{\sigma^*} \frac{1}{\sqrt{l} \sigma^*} \sum_{m=0}^{\lfloor kr \rfloor} \sum_{s=1}^b (U_{i_m+s} - E^* U_{i_m+s}) = \frac{1}{\sqrt{k}} \sum_{m=0}^{\lfloor kr \rfloor} V_m^*, \quad (8.12)$$

where the random variables

$$V_m^* = \frac{1}{\sqrt{b}} \sum_{s=1}^b (U_{i_m+s} - E^* U_{i_m+s})$$

are independent and have mean zero under the bootstrap distribution. Note that by the definition of σ^{*2} and because $Var^*(b^{-1/2} \sum_{s=1}^b U_{i_m+s})$ we have that

$$|Var^*(V_m^*) - \sigma^{*2}| \rightarrow 0 \quad (8.13)$$

in probability. Consider now the partial sum

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

where $\{\tilde{V}_m^*, m = 0, 1, 2, \dots, [kr]\}$ with

$$\tilde{V}_m^* = \frac{1}{\sqrt{[kr] + 1} \sqrt{Var^*(V_m^*)}} V_m^*,$$

forms an array of independent random variables. Since by definition $Var^*(\tilde{V}_m^*) = 1/([kr] + 1)$ and

$$\begin{aligned} \frac{\sum_{m=0}^{[kr]} E^* |\tilde{V}_m^*|^{2+\delta}}{\left\{Var^*\left(\sum_{m=0}^{[kr]} \tilde{V}_m^*\right)\right\}^{(2+\delta)/2}} &= \frac{1}{([kr] + 1)^{(2+\delta)/2}} \sum_{m=0}^{[kr]} E^* \left| \frac{V_m^*}{\sqrt{Var^*(V_m^*)}} \right|^{2+\delta} \\ &= \frac{[kr] + 1}{([kr] + 1)^{(2+\delta)/2}} \frac{1}{(Var^*(V_m^*))^{(2+\delta)/2}} \\ &\quad \times \frac{1}{n-b} \sum_{t=1}^{n-b} \left(\frac{1}{\sqrt{b}} \sum_{s=1}^b (U_{t+s} - E^* U_{i_m+s}) \right)^{2+\delta} \\ &= O_P\left(\left([kr] + 1\right)^{-\delta/2}\right) \rightarrow 0, \end{aligned}$$

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

$$\sum_{m=0}^{[kr]} \tilde{V}_m^* \Rightarrow N(0, 1) \quad (8.14)$$

in probability. (8.14) and (8.13) implies then since

$$\frac{1}{\sqrt{k}\sigma^*} \sum_{m=0}^{[kr]} V_m^* = \sqrt{\frac{Var^*(V_m^*)}{\sigma^{*2}}} \sqrt{\frac{[kr] + 1}{k}} V^*(r), \quad (8.15)$$

that

$$\frac{1}{\sqrt{k}\sigma^*} \sum_{m=0}^{[kr]} V_m^* \Rightarrow W(r) \quad (8.16)$$

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 $(V^*(r_1), V^*(r_2), \dots, V^*(r_m)) \Rightarrow (W(r_1), W(r_2), \dots, W(r_m))$ in probability, for any fixed set of points $r_1 < r_2 < \dots < r_m$ 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 $\sum_{m=0}^{[kr]} \tilde{V}_m^*$ is a sum of independent random variables with mean zero and

$$\max_{0 \leq m \leq [kr]} \text{Var}^*(\tilde{V}_m^*) = \frac{1}{[kr] + 1} \rightarrow 0$$

as $n \rightarrow \infty$. Thus $V^* \Rightarrow W$ in probability which by (8.5), (8.12) and (8.15), implies the assertion of the theorem. \square

Lemma 8.1 *Under the assumptions of Theorem 3.1 and if $n \rightarrow \infty$, then*

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

$$(ii) \quad \sigma^{*2} := \text{Var}^*[l^{-1/2} \sum_{j=1}^l U_j^*] \rightarrow \sigma^2 := 2\pi f_U(0)$$

and

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

in probability.

Proof: To prove (i) note that by (8.5) we have

$$\begin{aligned} l^{-1} \sum_{j=1}^l U_j^* &= l^{-1} X_1 + l^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^b \hat{U}_{i_m+s} \\ &= l^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^b U_{i_m+s} + o_P(1) \end{aligned}$$

and that the first term in the right hand side of the above expression is the sample mean of a block bootstrap series which converges to $E(U_t) = 0$ under the assumptions of the lemma.

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

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

and by (8.5)

$$\left| l^{-1/2} \sum_{j=1}^l U_j^* - l^{-1/2} \sum_{m=0}^{k-1} \sum_{s=1}^b (U_{i_m+s} - E^*(U_{i_m+s})) \right| \rightarrow 0$$

in probability. Therefore,

$$E^* \left(l^{-1/2} \sum_{j=1}^l U_j^* \right) \rightarrow 0 \quad \text{in probability.}$$

Furthermore,

$$\begin{aligned} E^* \left(l^{-1/2} \sum_{j=1}^l U_j^* \right)^2 &= E^* \left(l^{-1/2} \sum_{m=0}^{k-1} \sum_{s=1}^b U_{i_m+s} \right)^2 + o_P(1) \\ &\rightarrow \sigma^2 \end{aligned}$$

because the first term on the right hand side of the last equality above is nothing else than the variance of the bootstrap sample mean based on a block bootstrap sample $\{U_{i_m+s}, m = 0, 1, \dots, k-1 \text{ and } s = 1, 2, \dots, b\}$ which is known to converge to σ^2 ; cf. Künsch (1989). \square

Lemma 8.2 *Let the conditions of Theorem 3.1 be satisfied. If $n \rightarrow \infty$ then*

- (i) $l^{-2} \sum_{t=2}^l X_{t-1}^{*2} \Rightarrow \sigma^2 \int_0^1 W^2(r) dr,$
- (ii) $l^{-1} \sum_{t=2}^l X_{t-1}^* U_t^* \Rightarrow \frac{1}{2}(\sigma^2 W^2(1) - \sigma_U^2),$
- (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: To prove (i) verify that

$$\begin{aligned} l^{-2} \sum_{t=2}^l X_{t-1}^{*2} &= l^{-2} \sum_{t=2}^l \left(\sum_{j=1}^{t-1} U_j^* \right)^2 \\ &= \sigma^{*2} \frac{1}{l} \sum_{t=2}^l S_t^{*2} ((t-1)/l) \\ &= \sigma^{*2} \sum_{t=2}^l \int_{(t-1)/l}^{t/l} S_t^{*2}([lr]) dr \\ &= \sigma^{*2} \int_0^1 S_t^{*2}([lr]) dr \\ &\Rightarrow \sigma^2 \int_0^1 W^2(r) dr \end{aligned}$$

in probability, by Theorem 3.1 and Lemma 8.1.

To establish (ii) we write

$$\begin{aligned}
l^{-1} \sum_{t=2}^l X_{t-1}^* (X_t^* - X_{t-1}^*) &= \frac{1}{l} \sum_{t=2}^l \left(\sum_{j=1}^{t-1} U_j^* \right) U_t^* \\
&= \frac{1}{2l} \sum_{t=2}^l \left[\left(\sum_{j=1}^t U_j^* \right)^2 - \left(\sum_{j=1}^{t-1} U_j^* \right)^2 - U_t^{*2} \right] \\
&= \frac{\sigma^{*2}}{2} \sum_{t=2}^l \left[S_l^{*2}(t/l) - S_l^{*2}((t-1)/l) - \frac{1}{l\sigma^{*2}} U_t^{*2} \right] \\
&= \frac{\sigma^{*2}}{2} \left[S_l^{*2}(1) - S_l^{*2}(1/l) \right] - \frac{1}{2l} \sum_{t=2}^l U_t^{*2} \\
&\Rightarrow \frac{1}{2} \left(\sigma^2 W^2(1) - \sigma_U^2 \right)
\end{aligned}$$

in probability, by Theorem 3.1 and Lemma 8.1.

Assertion (iii) follows because

$$\begin{aligned}
l^{-3/2} \sum_{t=2}^l X_{t-1}^* &= l^{-3/2} \sum_{t=1}^{l-1} \left(\sum_{j=1}^t U_j^* \right) \\
&= \sigma^* \frac{1}{l} \sum_{t=2}^l S_l^*((t-1)/l)
\end{aligned}$$

and (iv) because

$$l^{-1/2} \sum_{t=1}^l U_t^* = \sigma^* S_l^*(1).$$

The joint weak convergence to the above limits is established using the Cramer-Wold device; cf. Serfling (1980). \square

Proof of Theorem 4.1: To prove the first assertion of the theorem recall that

$$l(\hat{\rho}_{LS}^* - 1) = \left(l^{-2} \sum_{t=2}^l X_{t-1}^{*2} \right)^{-1} l^{-1} \sum_{t=2}^l X_{t-1}^* (X_t^* - X_{t-1}^*)$$

and apply Lemma 8.2 (i) and (ii) as well as the δ -method; cf. Serfling (1980). The second part of the theorem follows because the least squares estimator $(\hat{\beta}^*, \hat{\rho}_{LS,C}^*)'$ obtained by regressing X_t^* on X_{t-1}^* and on a constant is given by

$$\begin{pmatrix} \sqrt{l} \hat{\beta}^* \\ l(\hat{\rho}_{LS,C}^* - 1) \end{pmatrix} = \begin{pmatrix} 1 & l^{-3/2} \sum_{t=2}^l X_{t-1}^* \\ l^{-3/2} \sum_{t=2}^l X_{t-1}^* & l^{-2} \sum_{t=2}^l X_{t-1}^{*2} \end{pmatrix}^{-1} \begin{pmatrix} l^{-1/2} \sum_{t=2}^l U_t^* \\ l^{-1} \sum_{t=2}^l X_{t-1}^* U_t^* \end{pmatrix}.$$

The desired result follows then by Lemma 8.2 and the same arguments as for the first part of the theorem. \square

Lemma 8.3 *Let the process $\{X_t\}$ satisfy Condition A' 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 $X_{t-i} - X_{t-i-1}$, $i = 1, 2, \dots, p$. If the choice of p is such that $p = p(n) \rightarrow \infty$, $p^3/n \rightarrow 0$ and $\sqrt{n} \sum_{j=p+1}^{\infty} |a_j| \rightarrow 0$ as $n \rightarrow \infty$, then*

$$(n-p)^{(1+\delta(\rho))/2}(\hat{\rho}_{DF} - \rho) = O_P(1)$$

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

Proof: Note that under Condition A' we have $\rho \leq 1$. Since for $\rho = 1$ the assertion has been proved by Said and Dickey (1984), we need to consider only the case $\rho < 1$. For this let $\hat{\theta} = (\hat{\rho}_{DF}, \hat{a}_{1,n}, \dots, \hat{a}_{p,n})'$ be the least squares estimator of $\theta = (\rho, a_1, \dots, a_p)'$ and recall that

$$\hat{\theta} - \theta = \left(\sum_{t=p+2}^n Y_{t-1}(p) Y'_{t-1}(p) \right)^{-1} \sum_{t=p+2}^n Y_{t-1}(p) e_t \quad (8.17)$$

where $Y_{t-1}(p) = (X_{t-1}, X_t - X_{t-1}, \dots, X_{t-p} - X_{t-p-1})'$ and $e_t = X_t - \rho X_{t-1} - \sum_{j=1}^p a_j (X_{t-j} - X_{t-j-1})$. Let $V_n(p) = (n-p)^{-1} \sum_{t=p+2}^n Y_{t-1}(p) Y'_{t-1}(p)$, $V(p) = E(Y_{t-1}(p) Y'_{t-1}(p))$ and $e_1 = (1, 0, \dots, 0)'$ we then have that

$$\begin{aligned} \sqrt{n-p}(\hat{\rho}_{DF} - \rho) &= e_1' \sqrt{n-p}(\hat{\theta} - \theta) \\ &= e_1' V^{-1}(p) \frac{1}{\sqrt{n-p}} \sum_{t=p+2}^n Y_{t-1}(p) \varepsilon_t \\ &\quad + e_1' V^{-1}(p) \frac{1}{\sqrt{n-p}} \sum_{t=p+2}^n Y_{t-1}(p) (e_t - \varepsilon_t) \\ &\quad + e_1' (V_n^{-1}(p) - V^{-1}(p)) \frac{1}{\sqrt{n-p}} \sum_{t=p+2}^n Y_{t-1}(p) e_t. \end{aligned} \quad (8.18)$$

Let $V_{t-1} = e_1' V^{-1}(p) Y_{t-1}(p)$. The desired result follows then because under the assumptions made

$$\frac{1}{\sqrt{n-p}} \sum_{t=p+2}^n V_{t-1} (e_t - \varepsilon_t) = o_P(1). \quad (8.19)$$

$$\frac{1}{\sqrt{n-p}} \sum_{t=p+2}^n V_{t-1} \varepsilon_t = O_P(1), \quad (8.20)$$

$$\sqrt{p+1} \|V_n^{-1}(p) - V^{-1}(p)\| = o_P(1), \quad (8.21)$$

and

$$E \left\| \frac{1}{\sqrt{n-p}} \sum_{t=p+2}^n Y_{t-1}(p) e_t \right\|^2 \leq Cp(n-p) \sum_{j=0}^{\infty} \pi_j^2. \quad (8.22)$$

Here, $\|x\|$ denotes the Euclidean norm of a vector x while for a matrix A , $\|A\| = \sup_{\|x\| \leq 1} \|Ax\|$. We proceed by showing that assertions (8.19) to (8.22) are true.

We first show that

$$E \left\| \frac{1}{\sqrt{n-p}} \sum_{t=p+2}^n V_{t-1}(e_t - \varepsilon_t) \right\|^2 \leq C(n-p) \sum_{j=p+1}^{\infty} a_j^2. \quad (8.23)$$

From this and the assumption that $\sqrt{n} \sum_{j=p+1}^{\infty} |a_j| \rightarrow 0$ the validity of (8.19) is established. To see (8.23) note that $\|V^{-1}(p)\|$ is uniformly bounded above by a constant C for all p , which is given by one over the smallest eigenvalue of $V(p)$; cf. Berk (1974), p. 493. Furthermore, for $v = e_1' V^{-1}(p) = (v_1, v_2, \dots, v_{p+1})$ we have

$$V_{t-1} = v' Y_{t-1}(p) = \sum_{k=1}^{p+1} c_k X_{t-k}$$

where $c_1 = v_1 + v_2$, $c_j = v_{j+1} - v_j$ for $j = 2, 3, \dots, p$ and $c_{p+1} = -v_{p+1}$. Using $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ we then get

$$V_{t-1} = \sum_{j=0}^{\infty} b_j \varepsilon_{t-1-j} \quad (8.24)$$

where

$$b_j = \sum_{i=1}^{\min\{j, p+1\}} c_i \psi_{j-i}.$$

We next show that $\text{Var}(V_{t-1}) \leq C$, i.e., $\sum_{j=0}^{\infty} b_j^2 \leq C$. To see this note that

$$\begin{aligned} \text{Var}(V_{t-1}) &= e_1' V^{-1}(p) e_1 \\ &= \left(\gamma_0 - c_{1,2} \Gamma_D^{-1} c_{2,1} \right)^{-1} \end{aligned} \quad (8.25)$$

where $\Gamma_D = E[\tilde{Y}_{t-1}(p) \tilde{Y}_{t-1}'(p)]$, $\tilde{Y}_{t-1}(p) = (X_{t-1} - X_{t-2}, X_{t-2} - X_{t-3}, \dots, X_{t-p} - X_{t-p-1})'$ and $c_{1,2} = c_{2,1}' = (\gamma_0 - \gamma_1, \gamma_1 - \gamma_2, \dots, \gamma_p - \gamma_{p-1})$. (8.25) follows using the partition

$$V(p) = \begin{pmatrix} \gamma_0 & c_{1,2} \\ c_{2,1} & \Gamma_D \end{pmatrix}$$

of the matrix $V(p)$, the fact that Γ_D is nonsingular and well-known results for partitioned matrices. The nonsingularity of Γ_D is a consequence of Lemma 5.1.1 of Brockwell and Davis (1991) since $\text{Var}(X_t - X_{t-1}) = 2(\gamma_0 - \gamma_1) > 0$ and $\text{Cov}(X_t - X_{t-1}, X_{t+h} - X_{t+h-1}) \rightarrow 0$ for $h \rightarrow \infty$. Note that $\gamma_0 - c_{1,2} \Gamma_D^{-1} c_{2,1} > 0$ because this quantity is the mean square error $E(\hat{X}_t - X_t)^2$ of the best linear predictor \hat{X}_t of X_t , i.e., \hat{X}_t is the projection of X_t on the closed span $\overline{\text{span}}\{X_{t-1} - X_{t-2}, X_{t-2} - X_{t-3}, \dots, X_{t-p} - X_{t-p-1}\}$. Lemma 2 of Berk (1974) and the established properties of V_{t-1} give then the desired result (8.23).

(8.20) follows since by (8.24) and straightforward calculations we get

$$\begin{aligned} E\left[\frac{1}{\sqrt{n-p}} \sum_{t=p+2}^n V_{t-1} \varepsilon_t\right]^2 &= \frac{1}{n-p} \sum_{t=p+2}^n \sum_{s=p+2}^n E(V_{t-1} V_{s-1} \varepsilon_t \varepsilon_s) \\ &= E(\varepsilon_t^4) \sum_{j=0}^{\infty} b_j^2. \end{aligned}$$

To prove (8.21) note first that by equations (2.10) and (2.11) of Berk (1974) we get

$$(n-p)E\left[\frac{1}{n-p} \sum_{t=p+1}^n X_{t-1}(X_{t-j} - X_{t-j-1}) - 2(\gamma_{j-1} - \gamma_{j-2})\right]^2 \leq C_1,$$

and

$$(n-p)E\left[\frac{1}{n-p} \sum_{t=p+1}^n (X_{t-i} - X_{t-i-1})(X_{t-j} - X_{t-j-1}) - (-\gamma_{i-j+1} + 2\gamma_{i-j} - \gamma_{i-j-1})\right]^2 \leq C_2.$$

Using the above bounds and the fact that the squared Euclidean norm of a matrix is bounded by the sum of its squared elements we have

$$E\|V_n(p) - V(p)\|^2 \leq \frac{Cp^2}{n-p}. \quad (8.26)$$

Therefore,

$$pE\|V_n(p) - V(p)\|^2 \rightarrow 0 \quad (8.27)$$

because by assumption $p^3/n \rightarrow 0$ as $n \rightarrow \infty$. Now,

$$\begin{aligned} \|V_n(p)^{-1} - V(p)^{-1}\| &= \|V_n(p)^{-1}(V(p) - V_n(p))V(p)^{-1}\| \\ &\leq \|V_n(p)^{-1}\| \|V_n(p) - V(p)\| \|V(p)^{-1}\| \\ &= g_n G_n g \end{aligned}$$

with an obvious notation for g_n , G_n and g . Recall that g is bounded, let $q_n = \|V_n(p)^{-1} - V(p)^{-1}\|$ and verify that $g_n \leq q_n + g$. Since $G_n \rightarrow 0$ in mean square and g is bounded, we can choose n large enough so that $gG_n < 1$ which leads to the inequality

$$\|V_n(p)^{-1} - V(p)^{-1}\| \leq \frac{g^2 \|V_n(p) - V(p)\|}{1 - \|V_n(p) - V(p)\|g}, \quad (8.28)$$

from which $\sqrt{p}\|V_n(p)^{-1} - V(p)^{-1}\| \rightarrow 0$ is concluded using (8.27).

To see (8.22) let $X_s(p) = (X_{s-1}, X_{s-2}, \dots, X_{s-p})'$ and write

$$\begin{aligned} E \left\| \frac{1}{\sqrt{n-p}} \sum_{t=p+2}^n Y_{t-1}(p) e_t \right\|^2 &\leq 2E \left\| \frac{1}{\sqrt{n-p}} \sum_{t=p+2}^n (X_{t-1}, X'_{t-1}(p))' e_t \right\|^2 \\ &\quad + 2E \left\| \frac{1}{\sqrt{n-p}} \sum_{t=p+2}^n (0, X'_{t-2}(p))' e_t \right\|^2 \end{aligned}$$

and argue as in equation (2.13) of Berk (1974). \square

To proceed with the proof of Theorem 4.2 we fix some notation. Let \tilde{D}_n be the $(p+1) \times (p+1)$ diagonal matrix $D_n = \langle (l-p), \sqrt{l-p}, \dots, \sqrt{l-p} \rangle$, and let $C(p)$ and $\tilde{C}(p)$ be the block matrices

$$C(p) = \begin{pmatrix} c_{11} & \mathbf{0}' \\ \mathbf{0} & \Gamma_D \end{pmatrix} \quad \text{and} \quad \tilde{C}(p) = \begin{pmatrix} \tilde{c}_{11} & \tilde{c}_{12} & \mathbf{0}' \\ \tilde{c}_{21} & \tilde{c}_{22} & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & \Gamma_D \end{pmatrix},$$

where $\mathbf{0}$ is a $p \times 1$ zero vector, $c_{11} = \tilde{c}_{22} = (l-p)^{-2} \sum_{t=p+2}^l X_{t-1}^{*2}$, $\tilde{c}_{21} = \tilde{c}_{12} = (l-p)^{-1} \sum_{t=p+2}^l X_{t-1}^*$ and $\tilde{c}_{11} = 1$. Furthermore, let $Y_{t-1}^*(p) = (X_{t-1}^*, D_{t,1}^*, \dots, D_{t,p}^*)'$ and $\tilde{Y}_{t-1}^*(p) = (1, X_{t-1}^*, D_{t,1}^*, \dots, D_{t,p}^*)'$. The following lemma is then established.

Lemma 8.4 *Let the assumptions of Theorem 4.2 be satisfied. If $n \rightarrow \infty$ then*

$$(i) \quad \sqrt{p+1} \left\| D_n \left(\sum_{t=p+2}^l Y_{t-1}^*(p) Y_{t-1}^{*'}(p) \right)^{-1} D_n - C(p)^{-1} \right\| \rightarrow 0,$$

$$(ii) \quad \sqrt{p+1} \left\| D_n \left(\sum_{t=p+2}^l \tilde{Y}_{t-1}^*(p) \tilde{Y}_{t-1}^{*'}(p) \right)^{-1} D_n - \tilde{C}(p)^{-1} \right\| \rightarrow 0,$$

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

and

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

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

Proof: Consider (i). Let $D_t^*(p) = (D_{t-1}^*, D_{t-2}^*, \dots, D_{t-p}^*)'$ and $C_n^*(p) = D_n^{-1} \sum_{t=p+2}^l Y_{t-1}^*(p) Y_{t-1}^{*'}(p) D_n^{-1}$ be the block diagonal matrix

$$C_n^*(p) = \begin{pmatrix} (l-p)^{-2} \sum_{t=p+2}^l X_{t-1}^{*2} & (l-p)^{-3/2} \sum_{t=p+2}^l X_{t-1}^* D_{t-1}^{*'}(p) \\ (l-p)^{-3/2} \sum_{t=p+2}^l X_{t-1}^* D_t^*(p) & (l-p)^{-1} \sum_{t=p+2}^l D_t^*(p) D_t^{*'}(p) \end{pmatrix}.$$

Verify first that for $i, j \in \{1, 2, \dots, p\}$,

$$(l-p)E^* \left((l-p)^{-1} \sum_{t=p+2}^l D_{t-i}^* D_{t-j}^* - (-\gamma_{i-j+1} + 2\gamma_{i-j} - \gamma_{i-j-1}) \right)^2 \leq C_1 \quad (8.29)$$

and

$$(l-p)E^* \left((l-p)^{-3/2} \sum_{t=p+2}^l X_{t-1}^* D_{t-j}^* \right)^2 \leq C_2 \quad (8.30)$$

in probability, where the constants C_1 and C_2 do not depend on p . To see (8.29) recall that $E(X_{t-i} - X_{t-i-1})(X_{t-j} - X_{t-j-1}) = -\gamma_{i-j+1} + 2\gamma_{i-j} - \gamma_{i-j-1}$ and that the D_t^* are block bootstrap replicates of centered differences of the original observations. Let $\tilde{\gamma}(i, j) = -\gamma_{i-j+1} + 2\gamma_{i-j} - \gamma_{i-j-1}$ and note that

$$\sum_{t=p+2}^l D_{t-i}^* D_{t-j}^* = \sum_{m=\lceil (p+1)/2 \rceil}^{k-1} \sum_{s=\max\{1, p+1-mb\}}^b D_{i_m+s-i} D_{i_m+s-j}.$$

For computational simplicity we consider in the following the asymptotically equivalent quantity

$$(l-p)E^* \left((l-p)^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^b D_{i_m+s-i} D_{i_m+s-j} - \tilde{\gamma}(i, j) \right)^2$$

which equals

$$\begin{aligned} & (l-p)E^* \left((k)^{-1} \sum_{m=0}^{k-1} \left(k(l-p)^{-1} \sum_{s=1}^b D_{i_m+s-i} D_{i_m+s-j} - \tilde{\gamma}(i, j) \right) \right)^2 \\ & l^2 k^{-1} (l-p)^{-1} \sum_{m=0}^{k-1} E^* \left(b^{-1} \sum_{s=1}^b D_{i_m+s-i} D_{i_m+s-j} - \tilde{\gamma}(i, j) \right)^2 \\ & l^2 k^{-1} (l-p)^{-1} \sum_{t=1}^{n-b} \left(b^{-1} \sum_{s=1}^b D_{t+s-i} D_{t+s-j} - \tilde{\gamma}(i, j) \right)^2. \end{aligned}$$

Now, $b^{-1} \sum_{s=1}^b D_{t+s-i} D_{t+s-j}$ is a \sqrt{b} -consistent estimator of $\tilde{\gamma}(i, j)$ and, as in Berk (1974), we have

$$bE \left(b^{-1} \sum_{s=1}^b D_{t+s-i} D_{t+s-j} - \tilde{\gamma}(i, j) \right)^2 \leq C$$

where the constant C does not depend on i and j . We therefore get

$$\begin{aligned} l^2 k^{-1} (l-p)^{-1} \sum_{t=1}^{n-b} \left(b^{-1} \sum_{s=1}^b D_{t+s-i} D_{t+s-j} - \tilde{\gamma}(i, j) \right)^2 &= O_P(l^2 k^{-1} b^{-1} (l-p)^{-1}) \\ &= O_P(1) \end{aligned}$$

uniformly in i and j .

To see (8.30) note that

$$\begin{aligned}
& (l-p)^{-2} E^* \left(\sum_{m=0}^{k-1} \sum_{s=1}^b X_{i_m+s-1} D_{i_m+s-j} \right)^2 \\
&= (l-p)^{-2} \sum_{m=0}^{k-1} E^* \left(\sum_{s=1}^b X_{i_m+s-1} D_{i_m+s-j} \right)^2 \\
&= k(l-p)^{-2} (n-b) \sum_{t=1}^{n-b} \left(\sum_{s=1}^b X_{t+s-1} D_{t+s-j} \right)^2.
\end{aligned}$$

The desired result follows then since as in Said and Dickey (1984), p. 601/602, we have

$$b^{-2} \sum_{s_1=1}^b \sum_{s_2=1}^b E \left(X_{t+s_1-1} D_{t+s_1-i} X_{t+s_2-1} D_{t+s_2-i} \right) \leq C$$

where the constant C does not depend on i and j .

By (8.29) and (8.30) we have that

$$E^* \|C_n^*(p) - C(p)\|^2 \leq C \frac{(p+1)^2}{l-p}$$

and, therefore,

$$\sqrt{p+1} \|C_n^*(p) - C(p)\| \rightarrow 0 \tag{8.31}$$

in probability by the assumption that $p^3/n \rightarrow 0$. To show that (i) is true we use

$$\begin{aligned}
\sqrt{p+1} \|C_n^{*-1}(p) - C^{-1}(p)\| &\leq \|C_n^{*-1}(p)\| \|C_n^*(p) - C(p)\| \|C_n^{-1}(p)\| \\
&\leq g_n^* \|C_n^*(p) - C(p)\| g^*
\end{aligned}$$

with an obvious notation for g_n^* and g^* . Verify that $g_n^* \leq \|C_n^{*-1}(p) - C^{-1}(p)\| + g^*$. Since $E^* \|C_n^*(p) - C(p)\|^2 \rightarrow 0$ and g^* is bounded in probability, we can choose n large enough so that $g^* \|C_n^*(p) - C(p)\| < 1$ in probability. Thus we have the inequality

$$\|C_n^{*-1}(p) - C^{-1}(p)\| \leq \frac{g^* \|C_n^*(p) - C(p)\|}{1 - \|C_n^*(p) - C(p)\| g^*}$$

from which assertion (i) follows by (8.31).

Since assertion (ii) can be proved using the same arguments, the details are omitted.

To prove (iii) we first show that

$$\left| \frac{1}{l-p} \sum_{t=p+2}^l X_{t-1}^* (U_t^* - \sum_{j=1}^p a_j D_{t-j}^*) - \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b X_{i_m+s-1}^+ (U_{i_m+s}^+ - \sum_{j=1}^p a_j D_{i_m+s-1-j}^+) \right| \rightarrow 0 \tag{8.32}$$

in probability, where $U_t^+ = (X_t - \rho X_{t-1}) - (n-1)^{-1} \sum_{\tau=2}^n (X_\tau - \rho X_{\tau-1})$ and X_t^+ is the series obtained by replacing \widehat{U}_t by U_t^+ in the first step of the CBB algorithm. To see (8.32) note that

$$\frac{1}{l-p} \sum_{t=p+2}^l X_{t-1}^* (U_t^* - \sum_{j=1}^p a_j D_{t-j}^*) = \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b X_{i_m+s-1}^+ (\widehat{U}_{i_m+s} - \sum_{j=1}^p a_j D_{i_m+s-1-j}) \quad (8.33)$$

and consider the term

$$\frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b X_{i_m+s-1} \widehat{U}_{i_m+s}. \quad (8.34)$$

For this term we have

$$\begin{aligned} \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b X_{i_m+s-1} \widehat{U}_{i_m+s} &= \frac{1}{l-p} X_1 \sum_{m=0}^{k-1} \sum_{s=1}^b \widehat{U}_{i_m+s} \\ &\quad + \frac{1}{l-p} \sum_{m_1=0}^{k-1} \sum_{s_1=1}^b \left(\sum_{m_2=0}^{m_1-1} \sum_{s_2=1}^b \widehat{U}_{i_{m_2}+s_2} + \sum_{s_2=1}^{s_1-1} \widehat{U}_{i_{m_1}+s_2} \right) \widehat{U}_{i_{m_1}+s_1} \\ &= \frac{1}{l-p} X_1 \sum_{m=0}^{k-1} \sum_{s=1}^b \widehat{U}_{i_m+s} \\ &\quad + \frac{1}{2} \left[\left(\frac{1}{\sqrt{l}} \sum_{m=0}^{k-1} \sum_{s=1}^b \widehat{U}_{i_m+s} \right)^2 - \frac{1}{l} \sum_{m=0}^{k-1} \sum_{s=1}^b \widehat{U}_{i_m+s}^2 \right]. \end{aligned}$$

Now, by lemma 8.3 and the definition of \widehat{U}_t we have that $\widehat{U}_{i_m+s} = U_{i_m+s}^+ + O_P(n^{-1/2})$ uniformly in m and s . Thus

$$\frac{1}{l-p} X_1 \sum_{m=0}^{k-1} \sum_{s=1}^b \widehat{U}_{i_m+s} = \frac{1}{l-p} X_1 \sum_{m=0}^{k-1} \sum_{s=1}^b U_{i_m+s}^+ + O_P(n^{-1/2}). \quad (8.35)$$

Similarly, we get

$$\frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b \widehat{U}_{i_m+s}^2 = \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b (U_{i_m+s}^+)^2 + O_P((l-p)^{-1/2}) O_P(n^{-1/2}). \quad (8.36)$$

Furthermore, using

$$\begin{aligned} \frac{1}{\sqrt{l-p}} \sum_{m=0}^{k-1} \sum_{s=1}^b \widehat{U}_{i_m+s} &= \frac{1}{\sqrt{l-p}} \sum_{m=0}^{k-1} \sum_{s=1}^b U_{i_m+s}^+ \\ &\quad - (\hat{\rho} - \rho) \frac{1}{\sqrt{l-p}} \sum_{m=0}^{k-1} \sum_{s=1}^b (X_{i_m+s_1} - \frac{1}{n-1} \sum_{\tau=2}^{-1} X_{\tau-1}) \end{aligned}$$

and (8.6) and (8.7) we get

$$\left| \frac{1}{\sqrt{l-p}} \sum_{m=0}^{k-1} \sum_{s=1}^b \widehat{U}_{i_m+s} - \frac{1}{\sqrt{l-p}} \sum_{m=0}^{k-1} \sum_{s=1}^b U_{i_m+s}^+ \right| \rightarrow 0 \quad (8.37)$$

in probability. From (8.35), (8.36) and (8.37) we obtain (8.32).

We next show that

$$\frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b X_{i_m+s-1}^+ (U_{i_m+s}^+ - \sum_{j=1}^p a_j D_{i_m+s-1-j}) \Rightarrow \frac{1}{2} \sigma_\varepsilon^2 C_\Psi (W^2(1) - 1). \quad (8.38)$$

For this note that by (2.3) and the definition of D_{t-j} we have

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

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 from

$$T_{1,n}^* \Rightarrow \frac{1}{2} \sigma_\varepsilon^2 C_\Psi (W^2(1) - 1), \quad (8.39)$$

$$T_{2,n}^* \rightarrow 0 \quad (8.40)$$

and

$$T_{3,n}^* \rightarrow 0, \quad (8.41)$$

in probability. We proceed by showing that (8.39) to (8.41) are true.

Consider $T_{1,n}^*$ and let $\varepsilon_t^+ = \varepsilon_t - (n-1)^{-1} \sum_{\tau=2}^n \varepsilon_\tau$. We have

$$\begin{aligned} T_{1,n}^* &= \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b X_{i_m+s-1}^+ \varepsilon_{i_m+s}^+ \\ &= \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b (X_1 + \sum_{r=0}^{m-1} \sum_{s=1}^b U_{i_r+j}^+ + \sum_{j=1}^s U_{i_m+j}^+) \varepsilon_{i_m+s}^+ \\ &= \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b (\sum_{r=0}^{m-1} \sum_{s=1}^b U_{i_r+j}^+ + \sum_{j=1}^s U_{i_m+j}^+) \varepsilon_{i_m+s}^+ + O_P(l^{1/2}(l-p)^{-1}). \quad (8.42) \end{aligned}$$

Using the polynomial decomposition given in Phillips and Solo (1992) we can write

$$U_t^+ = \Psi(1) \varepsilon_t^+ - (1-L) \tilde{\Psi}(L) \varepsilon_t^+ \quad (8.43)$$

where L is the shift operator, $\Psi(1) = C_\Psi = \sum_{j=0}^{\infty} \psi_j$, $\tilde{\Psi}(L) = \sum_{j=0}^{\infty} \tilde{\psi}_j L^j$ and $\tilde{\psi}_j = \sum_{i=j+1}^{\infty} \psi_i$. Note that $\sum_{j=0}^{\infty} j^2 \psi_j^2 < \infty$ implies that $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Substituting (8.43) in (8.42) we get

$$\begin{aligned}
& \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b \left(\sum_{r=0}^{m-1} \sum_{s=1}^b U_{i_r+j}^+ + \sum_{j=1}^s U_{i_m+j}^+ \right) \varepsilon_{i_m+s}^+ \\
&= \Psi(1) \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b \left(\sum_{r=0}^{m-1} \sum_{s=1}^b \varepsilon_{i_r+j}^+ + \sum_{j=1}^s \varepsilon_{i_m+j}^+ \right) \varepsilon_{i_m+s}^+ \\
&\quad - \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b \left(\sum_{r=0}^{m-1} \sum_{s=1}^b (1-L) \tilde{\Psi}(L) \varepsilon_{i_r+j}^+ + \sum_{j=1}^s (1-L) \tilde{\Psi}(L) \varepsilon_{i_m+j}^+ \right) \varepsilon_{i_m+s}^+ \\
&= L_{1,n}^* + L_{2,n}^*
\end{aligned}$$

with an obvious notation for $L_{1,n}^*$ and $L_{2,n}^*$.

Now,

$$\begin{aligned}
L_{1,n}^* &= \Psi(1) \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b \left(\sum_{r=0}^{m-1} \sum_{s=1}^b \varepsilon_{i_r+j}^+ + \sum_{j=1}^s \varepsilon_{i_m+j}^+ \right) \varepsilon_{i_m+s}^+ \\
&= \Psi(1) \frac{1}{2} \left[\left(\frac{1}{\sqrt{l-p}} \sum_{m=0}^{k-1} \sum_{s=1}^b \varepsilon_{i_m+s}^+ \right)^2 - \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b (\varepsilon_{i_m+s}^+)^2 \right]
\end{aligned}$$

and because

$$\frac{1}{\sqrt{l-p}} \sum_{m=0}^{k-1} \sum_{s=1}^b \varepsilon_{i_m+s}^+ \Rightarrow N(0, \sigma_\varepsilon^2)$$

and

$$\frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b (\varepsilon_{i_m+s}^+)^2 \rightarrow \sigma_\varepsilon^2$$

in probability, we get using $C_\Psi = \Psi(1)$ that

$$L_{1,n}^* \Rightarrow \frac{1}{2} \sigma_\varepsilon^2 C_\Psi (W^2(1) - 1).$$

We next show that $L_{2,n}^* \rightarrow 0$ in probability. For this let $V_t^+ = \tilde{\Psi}(L) \varepsilon_t^+$ and note that

$$\begin{aligned}
L_{2,n}^* &= \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b \left[\sum_{r=0}^{m-1} \sum_{j=1}^b (V_{i_r+j}^+ - V_{i_r+j-1}^+) \right] \varepsilon_{i_m+s}^+ \\
&\quad + \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b \left[\sum_{j=1}^{s-1} (V_{i_m+j}^+ - V_{i_m+j-1}^+) \right] \varepsilon_{i_m+s}^+ \\
&= \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b \sum_{r=0}^{m-1} (V_{i_r+b}^+ - V_{i_r}^+) \varepsilon_{i_m+s}^+ \\
&\quad + \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b (V_{i_m+s-1}^+ - V_{i_m}^+) \varepsilon_{i_m+s}^+.
\end{aligned}$$

Since $\sum_{r=0}^{k-1} (V_{i_r+b}^+ - V_{i_r}^+) = O_P(k^{1/2})$ uniformly in r and $\sum_{r=0}^{k-1} \sum_{s=1}^b \varepsilon_{i_m+s}^+ = O_P(l^{1/2})$ we get

$$\frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b \sum_{r=0}^{m-1} (V_{i_r+b}^+ - V_{i_r}^+) \varepsilon_{i_m+s}^+ = O_P(k^{1/2} l^{1/2} (l-p)^{-1})$$

and

$$\frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b (V_{i_m+s-1}^+ - V_{i_m}^+) \varepsilon_{i_m+s}^+ = O_P((l-p)^{1/2}),$$

which shows that $L_{2,n}^* \rightarrow 0$ in probability.

Consider the term $T_{2,n}^*$ and verify that

$$\frac{1}{n-1} \sum_{t=2}^n (X_t - X_{t-1}) - \frac{1}{n-p} \sum_{t=p+2}^n (X_{t-j} - X_{t-p-j}) = O_P(p^{1/2}(n-p))$$

uniformly in j . Since $l^{3/2} \sum_{m=0}^{k-1} \sum_{s=1}^b X_{i_m+s-1}^+ = O_P(1)$ we get

$$T_{2,n}^* = O_P\left(\frac{l^{3/2} p^{3/2}}{(l-p)(n-p)}\right) \rightarrow 0$$

for $p^3/n \rightarrow 0$.

Finally, to see why $T_{3,n}^* \rightarrow 0$ in probability, note that

$$\frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b E^*(X_{i_m+s-1}^+)^2 = O_P(n-b)$$

and

$$\begin{aligned} \frac{1}{l-p} \sum_{m=0}^{k-1} \sum_{s=1}^b E^* \left(\sum_{j=p+1}^{\infty} a_j \left[(X_{i_m+s-j} - X_{i_m+s-j-1}) - \frac{1}{n-1} \sum_{t=p+2}^n (X_{t-j} - X_{t-j-1}) \right] \right)^2 \\ = O_P\left(\sum_{j=p+1}^{\infty} a_j^2 \right). \end{aligned}$$

By the Cauchy-Schwarz inequality with then get

$$T_{3,n}^* = O_P(\sqrt{n-p} \sum_{j=p+1}^{\infty} |a_j|)$$

which goes to zero under the assumptions of the lemma. \square

Proof of Theorem 4.2: We give the proof of the first part of the theorem since the second part is proved along the same lines.

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

$$\begin{aligned} (l-p)(\hat{\rho}_{DF}^* - 1) &= e_1' D_n (\hat{\theta}^* - \theta) \\ &= e_1^* C_n^{*-1}(p) D_n^{-1} \sum_{t=p+2}^l Y_{t-1}^*(p) e_t^* \end{aligned}$$

where

$$\begin{aligned} e_t^* &= X_t^* - \theta^{*'} Y_{t-1}^*(p) \\ &= U_t^* - \sum_{j=1}^p a_j D_{t-j}^* \end{aligned}$$

and

$$C_n^{*-1}(p) = D_n \left(\sum_{t=p+2}^l Y_{t-1}^*(p) Y_{t-1}^{*'}(p) \right)^{-1} D_n.$$

Write

$$\begin{aligned} (l-p)(\hat{\rho}_{DF}^* - 1) &= e_1' C^{-1}(p) D_n^{-1} \sum_{t=p+2}^l Y_{t-1}^*(p) e_t^* \\ &\quad + e_1' \left(C_n^{*-1}(p) - C^{-1}(p) \right) D_n^{-1} \sum_{t=p+2}^l Y_{t-1}^*(p) e_t^* \end{aligned}$$

and verify by straightforward calculations that

$$\left\| D_n^{-1} \sum_{t=p+2}^l Y_{t-1}^*(p) e_t^* \right\| = O_{P^*}(\sqrt{p+1}). \quad (8.44)$$

This together with lemma 8.4 (i) implies that

$$\left| e_1' \left(C_n^{*-1}(p) - C^{-1}(p) \right) D_n^{-1} \sum_{t=p+2}^l Y_{t-1}^*(p) e_t^* \right| \rightarrow 0 \quad (8.45)$$

in probability.

Now, since $C^{-1}(p)$ is a block diagonal matrix with first row given by $(c_{11}^{-1}, 0, 0, \dots, 0)$ we get

$$e_1' C^{-1}(p) D_n^{-1} \sum_{t=p+2}^l Y_{t-1}^*(p) e_t^* = \left((l-p)^{-2} \sum_{t=p+2}^l X_{t-1}^{2*} \right)^{-1} \frac{1}{l-p} \sum_{t=p+2}^l X_{t-1}^* e_t^*.$$

Thus

$$\left| (l-p)(\hat{\rho}_{DF}^* - 1) - \left((l-p)^{-2} \sum_{t=p+2}^l X_{t-1}^{2*} \right)^{-1} \frac{1}{l-p} \sum_{t=p+2}^l X_{t-1}^* e_t^* \right| \rightarrow 0$$

in probability. Because of this and the fact that

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

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

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

in probability. The assertion of the theorem is then established by the observation that the right hand side of the above expression is the asymptotic distribution of $(n-p)(\hat{\rho}_{DF} - 1)$ under the hypothesis that $\{X_t\}$ is unit root integrated; cf. Said and Dickey (1984). \square

References

- [1] Andrews, D. W. K. (1984), Nonstrong Mixing Autoregressive Processes. *Journal of Applied Probability*, 21, 930-934.
- [2] Berk, K. N. (1974), Consistent Autoregressive Spectral Estimates. *Annals of Statistics*, 2, 489-502.
- [3] Bertail. P. (1994), Un Test Bootstrap dans un Modèle AR(1). *Annales d'Économie et des Statistique*, 36, 57-79.
- [4] Billingsley, P. (1999), *Convergence of Probability Measures*. John Wiley: New York.
- [5] Brockwell, P. J. and Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd ed., Springer: New York.
- [6] Campbell, J. Y. and P. Perron (1991), Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots, *NBER Macroeconomics Annual*, MIT Press, Cambridge (Mass.).
- [7] DeJong, D. N., Nankervis, J. C., Savin, N. E. and Whiteman, C. H. (1992), The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors. *Journal of Econometrics*, 53, 323-343.
- [8] Dickey, D. A. and W. A. Fuller (1979), Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association* 74, 427-431.
- [9] Efron, B. and Tibshirani, R. (1993), *An Introduction to the Bootstrap*, Chapman and hall: New York.
- [10] Elliot, G., Rothenberg, T. J. and Stock, J. (1996), Efficient Tests for an Autoregressive Unit Root. *Econometrica*, 64, 813-836.

- [11] Ferretti, N. and Romo, J. (1996), Unit Root Bootstrap Tests for AR(1) models, *Biometrika*, 83, 849-860.
- [12] Freedman, D. A. (1981), Bootstrapping Regression Models, *Annals of Statistics* 9, 1218-1228.
- [13] Fuller, W. (1996), *Introduction to Statistical Time Series*, 2nd ed., John Wiley, New York.
- [14] Hall, A. (1989), Testing for a Unit Root in the Presence of Moving Average Errors. *Biometrika*, 76, 49-56.
- [15] Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton: New Jersey.
- [16] Künsch, H. R. (1989), The Jackknife and the Bootstrap for General Stationary Observations, *Annals of Statistics*, 17, 1217-1241.
- [17] Liu, R. Y. and Singh, K. (1992), Moving Blocks Jackknife and Bootstrap Capture Weak Dependence. In *Exploring the Limits of the Bootstrap*, ed. by LePage and Billard. John Wiley: New York, pp. 225-248.
- [18] Ng, S. and Perron, R. (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.
- [19] Paparoditis, E. and Politis, D. N. (2000), The Continuous-Path Block Bootstrap. In *Asymptotics in Statistics and Probability. Festschrift for George Roussas*, Madan Puri, Ed. To appear.
- [20] Phillips, P. C. B. (1987), Time Series Regression with a Unit Root. *Econometrica*, 2, 277-301.
- [21] Phillips, P. C. B. and Perron, P. (1988), Testing for a Unit Root in Time Series Regression, *Biometrika*, 75, 335-346.
- [22] Phillips, P. C. B. and Solo, V. (1992), Asymptotics for Linear Processes. *Annals of Statistics*, 20, 971-1001.
- [23] Politis, D. N. and Romano, J. P. (1994), The Stationary Bootstrap. *Journal of the American Statistical Association*, 89, 1303-1313.
- [24] Rosenblatt, M. (1985), *Stationary Sequences and Random Fields*. Birkhäuser: Boston.
- [25] Said, S. E. and Dickey, D. A. (1984), Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika*, 71, 599-607.
- [26] R. Serfling (1980), *Approximation Theorems of Mathematical Statistics*. John Wiley: New York.

- [27] Schwert, G. W. (1989), Tests for Unit Roots: A Monte Carlo Investigation. *Journal of Business & Economic Statistics*, 7, 147-159.
- [28] Shao, J. and Tu, D. (1995), *The Jackknife and the Bootstrap*. Springer-Verlag: New York.
- [29] Stock, J. (1994), Unit Roots and Trend Breaks, in: R. Engle and D. McFadden, eds. *Handbook of Econometrics*, Vol. 4, North-Holland.
- [30] Whithers, C. S. (1981), Conditions for a Linear Process to be Strong-Mixing. *Probability Theory and Related Fields*, 57, 477-480.

CAPTIONS FOR FIGURES

Figure 1 (a) Plot of the natural logarithm of the Dow Jones index series recorded daily from Aug. 28 to Dec. 18, 1972; (b) plot of a BB pseudo-series with block size 10; (c) plot of a CBB realization with block size 10.

| ϕ | θ | n=100 Blocksize | | | n=200 Blocksize | | |
|--------|----------|--------------------|-------|-------|--------------------|-------|-------|
| | | b=4 | b=6 | b=8 | b=5 | b=7 | b=10 |
| 1.0 | 0.8 | 0.046 | 0.053 | 0.049 | 0.049 | 0.041 | 0.050 |
| | 0.0 | 0.045 | 0.058 | 0.061 | 0.056 | 0.047 | 0.056 |
| | -0.8 | 0.437 | 0.377 | 0.371 | 0.495 | 0.462 | 0.469 |
| 0.9 | 0.8 | 0.603 | 0.637 | 0.614 | 0.986 | 0.984 | 0.981 |
| | 0.0 | 0.753 | 0.773 | 0.773 | 0.998 | 0.997 | 0.997 |
| | -0.8 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 0.85 | 0.8 | 0.876 | 0.859 | 0.860 | 1.000 | 1.000 | 0.999 |
| | 0.0 | 0.960 | 0.961 | 0.951 | 1.000 | 1.000 | 1.000 |
| | -0.8 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Table 1: Empirical rejection probabilities of the CBB unit root test with nominal level $\alpha = 0.05$ under different settings of the ARMA parameters ϕ and θ . The test statistic used here was $n(\hat{\rho}_{LS} - 1)$ of Example 2.1.

| ϕ | θ | n=100 Blocksize=8 | | | n=200 Blocksize=10 | | |
|--------|----------|----------------------|-------|-------|-----------------------|-------|-------|
| | | p=4 | p=6 | p=8 | p=5 | p=7 | p=9 |
| 1.0 | 0.8 | 0.010 | 0.012 | 0.012 | 0.055 | 0.054 | 0.055 |
| | 0.0 | 0.032 | 0.033 | 0.026 | 0.058 | 0.061 | 0.063 |
| | -0.8 | 0.170 | 0.141 | 0.126 | 0.218 | 0.187 | 0.165 |
| 0.9 | 0.8 | 0.292 | 0.299 | 0.300 | 0.982 | 0.966 | 0.965 |
| | 0.0 | 0.583 | 0.569 | 0.546 | 0.995 | 0.983 | 0.979 |
| | -0.8 | 1.000 | 0.999 | 0.997 | 1.000 | 1.000 | 1.000 |
| 0.85 | 0.8 | 0.567 | 0.563 | 0.541 | 0.998 | 0.998 | 0.997 |
| | 0.0 | 0.844 | 0.814 | 0.792 | 1.000 | 0.999 | 0.999 |
| | -0.8 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Table 2: Empirical rejection probabilities of the CBB unit root test with nominal level $\alpha = 0.05$ under different settings of the ARMA parameters ϕ and θ . The test statistic used here was $(n - p)(\hat{\rho}_{DF} - 1)$ of Example 2.3; the chosen CBB block sizes were: 8 (case $n = 100$), and 10 (case $n = 200$).

Figure 1a: Dow Jones Series (logarithm)

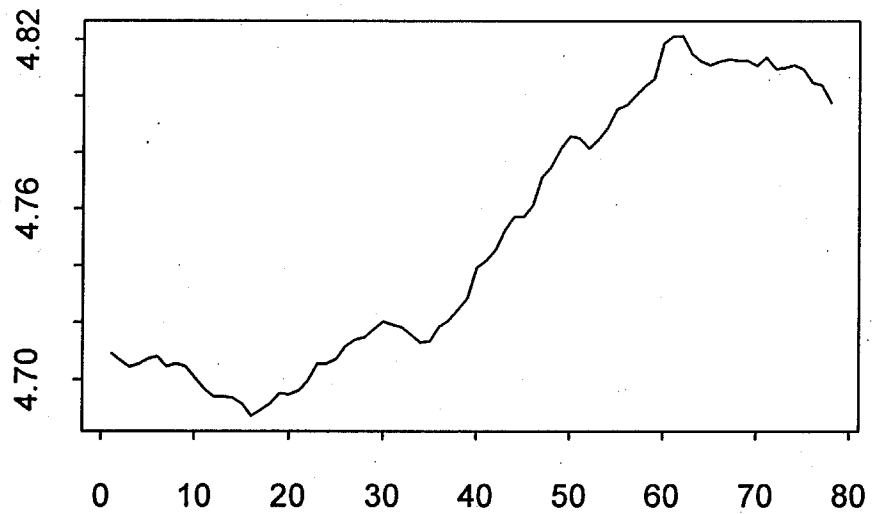


Figure 1b: Dow Jones BB pseudo-series

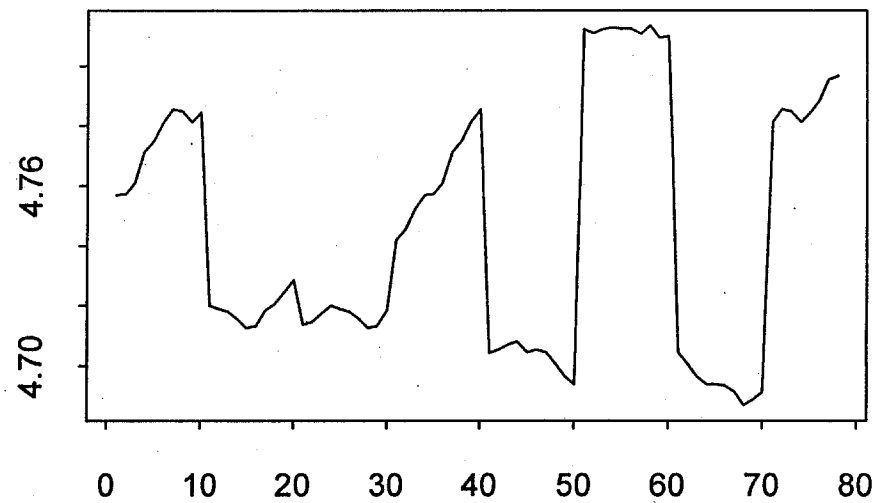


Figure 1c: Dow Jones CBB pseudo-series

