Bootstrapping unit root tests

Anders Rygh Swensen
Department of Mathematics, University of Oslo,
P.O. Box 1053, N-0316 Blindern, Oslo , Norway

June 23, 2000

Abstract

In this paper we consider two bootstrap algorithms for testing unit roots under the condition that the differenced data are stationary. The first method consists of generating the resampled data after fitting an autoregressive model to the first differences of the observations. The second method consists of applying the stationary bootstrap to the first differences. Both procedures are shown to give methods that approach the correct asymptotic distribution under the null hypothesis of a unit root. We also present a Monte Carlo study comparing the two methods for some ARIMA models.

Keywords: Recursive bootstrap; Stationary bootstrap; Unit root.

1 Introduction

Assume that the observations $Y_1, \ldots, Y_T$ are generated according to the autoregressive scheme

$$Y_t = \alpha Y_{t-1} + U_t, (t = 1, \ldots, T), \alpha = 1 \tag{1}$$

where U_t is a strictly stationary process with mean zero.

The initial value in (1), Y_0, is assumed to satisfy

Assumption 1
Y_0 has a fixed distribution not depending on T.

Consider the estimator $\hat{\alpha}$ for α defined in Fuller (1996) as $\hat{\alpha} = \sum_{t=2}^{T} Y_t Y_{t-1} / \sum_{t=2}^{T} Y_{t-1}^2$. Phillips (1987) showed, under a mixing condition, that as $T \to \infty$, the distribution of $T(\hat{\alpha} - 1)$ converges towards the distribution of the random variable $\frac{1}{2}(W(1)^2 - \frac{\sigma_u^2}{\sigma^2}) / \int_0^1 W(s)^2 ds$ where $W(s)$ is a standard Wiener process on $C[0,1]$, the space of continuous functions defined on the interval $[0,1]$, $\sigma_u^2 = E(U_t^2)$ and $\sigma^2 = \lim_{T\to\infty} Var(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_t)$. For the t-statistic based on $\hat{\alpha}$, $t_\alpha = (\sum_{t=2}^{T} Y_{t-1}^2)^{1/2} (\hat{\alpha}_T - 1)/s_U$ where $s_U^2 = \sum_{t=2}^{T} (Y_t - \hat{\alpha}_T Y_{t-1})^2 / (T-2)$, a similar result exists.

The asymptotic distribution depends on the unknown parameters σ^2 and σ_u^2 even under the null hypothesis of a unit root, $H_0 : \alpha = 1$. There are several ways to deal with this. The most popular is the augmented Dickey-Fuller test proposed in Dickey and Fuller (1979) and Said and Dickey (1984) which consists of fitting an autoregressive process of appropriate order and base the test on the estimated coefficients. In Said and Dickey (1984) this was shown to yield a test which is asymptotically similar for general ARMA processes $\{U_t\}$ provided the order of the fitted autoregression increases as $o(T^{1/3})$.

Phillips (1987) suggested a nonparametric modification of the estimator $\hat{\alpha}$ which also has the asymptotic distribution $\frac{1}{2}(W(1)^2 - 1) / \int_0^1 W(s)^2 ds$. Using the modified statistics therefore yields a test which is asymptotically similar.

The alternative we shall consider is to estimate the distribution of $T(\hat{\alpha} - 1)$ by a bootstrap procedure, which will adapt automatically and approximate the distribution involving the unknown parameters. This is shown to be feasible in a recent paper by Ferretti and Romo (1996) in the case where the stationary process $\{U_t\}$ is an $AR(p)$ process of known order. The bootstrap sample is constructed based on the estimated errors of the autoregression using the recursive scheme defined by the model. They also demonstrated in a simulation study that the finite sample properties were satisfactory. This procedure is an example of what Li and Maddala (1997) calls the recursive bootstrap.

It is well known that bootstrapping in this situation is non-standard and present some special problems. Basawa et. al. (1991a) showed that in the case where $\{U_t\}$ are independent identically distributed variables and the bootstrap sample is constructed from the recursion $Y_t^* = \hat{\alpha} Y_{t-1}^* + U_t^*$, the asymptotic distribution is random. Following Basawa et. al. (1991b), Ferretti and Romo (1996) therefore used the recursion $Y_t^* = Y_{t-1}^* + U_t^*$. As usual U_t^* denotes the variables that are resampled. Ferretti and Romo (1996) used resampled variables based on centering the least squares residuals $Y_t - \hat{\alpha} Y_{t-1}$.

What makes bootstrapping of particular interest in testing unit roots, and in a multivariate setting cointegration analysis, is that often the asymptotic distribution of common statistics involves nonstandard distributions and nuisance parameters. Considerable effort have been spent to find estimators that are at least asymptotically similar. The implications of bootstrap based methods may therefore be quite far reaching, if it turns out that the approach to deal with nuisance parameters is satisfactory.

A problem using model-based resampling is that the structure of the data is unknown and must be identified from the observations. To ensure that the resampled data have the same structure as the original this identification must be correct. We shall therefore show that the conclusions from Ferretti and Romo (1976) are true for quite general processes $\{U_t\}$ provided the recursive residuals are constructed using autoregressive approximations where the order is $o(T^{1/3})$ as $T \to \infty$.

In addition to the recursive bootstrap there has been other suggestions how to exploit the bootstrap ideas for dependent data with the explicit purpose of avoiding the assumption that the generating model is known up to some unknown parameters. The block bootstrap methods consist of splitting the series in blocks of equal length and sample these with replacement. These resampled time series are not stationary conditional on the original data, which motivated Politis and Romano (1994) to introduce the stationary bootstrap, where the blocks are no longer of a fixed size but are determined according to a fixed, usually a geometric, distribution. A price for obtaining the conditional stationarity seems to have to be paid however. In estimating the unknown mean in a stationary time series Lahiri (1999) shows that the stationary bootstrap has larger mean square error than the other block bootstrap methods. As far as asymptotic distributions are concerned, we shall demonstrate that under quite general conditions using the stationary bootstrap yields the same limit as the recursive bootstrap.

The plan for the paper is as follows. In section 2 we treat the recursive bootstrap and provide results from some simulations to evaluate the performance of the method in finite samples. In section 3 the stationary bootstrap is similarly treated. Technical proofs can be found in section 4.

2 The recursive bootstrap

In this section we assume that the time series $\{Y_t\}_{t=1}^{\infty}$is generated by (1) and that $\{U_t\}$ can be written as a one sided moving average process of infinite order

$$U_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

where the random variables $\{\epsilon_t\}$ are independent, identically distributed and the weights satisfy some regularity conditions that will be specified below.

2.1 *Algorithm and asymptotic distribution*

The bootstrap algorithm that we shall consider is defined as follows. Let $Y_1, \ldots, Y_T$ be the observations. Compute the differences $\Delta Y_t = Y_t - Y_{t-1}, t = 2, \ldots, T$. Fit an autoregressive process of order k to the differences by the usual Yule -Walker procedure. Thus, let $\hat{\gamma}_k = (\hat{\gamma}(0), \ldots, \hat{\gamma}(k-1))'$ be the empirical covariances based on the differences and $\hat{\Gamma}_k$ be the estimated covariance matrix

$$\hat{\Gamma}_k = \begin{pmatrix} \hat{\gamma}(0) & \cdots & \hat{\gamma}(k-1) \\ \vdots & & \vdots \\ \hat{\gamma}(k-1) & \cdots & \hat{\gamma}(0) \end{pmatrix}.$$

The estimators $\hat{\boldsymbol{\phi}}_k = (\hat{\phi}_{1k}, \ldots, \hat{\phi}_{kk})'$ are then estimated by $\hat{\boldsymbol{\phi}}_k = \hat{\Gamma}_k^{-1}\hat{\gamma}_k$ and the variance by $\hat{\sigma}_k^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\phi}}_k'\hat{\gamma}_k$. The residuals are given by

$$\hat{\epsilon}_{t,T} = \Delta Y_t - \sum_{i=1}^{k} \hat{\phi}_{ik} \Delta Y_{t-i}, (t = k+2, \ldots, T)$$

where $\hat{\phi}_{0k} = 1$, and the centered residuals by

$$\tilde{\epsilon}_{t,T} = \hat{\epsilon}_{t,T} - \frac{1}{T-1-k} \sum_{t=k+2}^{T} \hat{\epsilon}_{t,T}, (t = k+2, \ldots, T).$$

Consider T observations from the stationary autoregressive process

$$U_t^* = \sum_{i=1}^{k} \hat{\phi}_{ik} U_{t-i}^* + \epsilon_t^*, t = 1, \ldots, T \tag{2}$$

where the errors $\{\epsilon_t^*\}, (t = 1, \ldots, T)$ are randomly sampled from the centered residuals with replacement. The use of stationary realizations follows a suggestion by Bühlmann (1997). As he points out, in practice one obtains the bootstrap sample $U_t^*, t = 1, \dots, T$ by running the recursion (2) for a while to eliminate the influence from the initial conditions. For notational purpose we do not indicate the dependence of $\{U_t^*\}$ and $\{\epsilon_t^*\}$ on T, but it is important to realize that the resampled residuals will now have a triangular array structure. The bootstrap sample $Y_1^*, \dots, Y_T^*$ is then defined by

$$Y_t^* = Y_{t-1}^* + U_t^*, t = 1, \dots, T$$

where Y_0^* is a random variable generated according a fixed distribution which should reflect the distribution of Assumption 1. The conditional distribution of $(Y_1^*, \dots, Y_T^*)$ given $(Y_1, \dots, Y_T)$ will, as is usual in the literature, be denoted by an $*$, so that P^*, E^* etc. refer to this distribution.

It may be worth mentioning that the scheme above is a modification of the procedure in Ferretti and Romo (1996) in three respects. First we use the differences ΔY_t where they employ the least squares residuals $Y_t - \hat{\alpha}_T Y_{t-1}$. In addition we use stationary realizations from the recursion (2). Finally, the coefficients of the fitted autoregressions are estimated by the Yule-Walker procedure, not by ordinary least squares.

We then compute

$$\hat{\alpha}_T^* = \frac{\sum_{t=2}^T Y_t^* Y_{t-1}^*}{\sum_{t=2}^T Y_{t-1}^{*2}} \tag{3}$$

and the corresponding t-statistic

$$t_{\alpha,T}^* = (\sum_{t=2}^T Y_{t-1}^{*2})^{1/2} (\hat{\alpha}_T^* - 1)/s_U \tag{4}$$

where $s_U^2 = \sum_{t=2}^T (Y_t - \hat{\alpha}_T Y_{t-1})^2/(T-2)$.

The bootstrap approximations to the distributions of $T(\hat{\alpha}_T - 1)$ and $t_{\alpha,T}$ are the distributions of $\hat{\alpha}_T^*$ and t_α^*, and these distributions can be approximated by the empirical distributions of the two statistics based on B bootstrap samples constructed as described above.

We shall now find the asymptotic distributions of α_T^* and $t_{\alpha,T}^*$ when $T \to \infty$. Before we state the necessary technical assumptions we recall that if $\{U_t\}$ also has an $AR(\infty)$ representation so that $\epsilon_t = \sum_{j=0}^\infty \phi_j U_{t-j}$ where $\phi_0 = 1$, the following two properties are equivalent, see Lemma 2.1 in Bühlmann (1997), (i)$\Phi(z) = \sum_{j=0}^\infty \phi_j z^j \neq 0$ for $|z| \leq 1$ and $\sum_{j=0}^\infty j|\phi_j| < \infty$ and (ii) $\Psi(z) = \sum_{j=0}^\infty \psi_j z^j \neq 0$ for $|z| \leq 1$ and $\sum_{j=0}^\infty j|\psi_j| < \infty$.

Assumption 2
(i) The random variables $\{\epsilon_t\}_{t=-\infty}^\infty$ are independent and identically distributed with $E(\epsilon_t) = 0, E(\epsilon_t^2) = \sigma_\epsilon^2$ and $E(\epsilon_t^4) < \infty$;

(ii) The power series $\Phi(z) = \sum_{j=0}^{\infty} \phi_j z^j$ is non-zero for $|z| \leq 1$ and $\sum_{j=0}^{\infty} j|\phi_j| < \infty$;
(iii) $k = k(T)$ satisfies $\frac{k^3}{T} \to 0$ as $T \to \infty$.

We also remind that the spectral density of $\{U_t\}$ is given by

$$f_U(\omega) = \frac{\sigma_\epsilon^2}{2\pi|\sum_j \phi_j e^{-ij\omega}|^2}, -\pi \leq \omega \leq \pi. \tag{5}$$

where $\sigma_\epsilon^2 = E(\epsilon_t^2)$.

Then the following theorem is valid

Theorem 1 *Suppose $Y_1, \ldots, Y_T$ is a time series generated according to (1) and satisfies Assumptions 1 and 2. Then for all x as $T \to \infty$,*

(i) $P^*(T(\hat{\alpha}_T^* - 1) \leq x) \to P(\frac{1}{2}(W(1)^2 - \frac{\sigma_u^2}{\sigma^2})/\int_0^1 W(s)^2 ds \leq x)$

(ii) $P^*(t_{\alpha,T}^* \leq x) \to P(\frac{\sigma}{2\sigma_u}(W(1)^2 - \frac{\sigma_u^2}{\sigma^2})/(\int_0^1 W(s)^2 ds)^{1/2} \leq x)$

in probability, where W is a standard Brownian motion.

A test having correct asymptotic level for $H_0 : \alpha = 1$ against $H_1 : \alpha < 1$ can now be constructed based on whether $\#\{b : \hat{\alpha}_T^* \leq \hat{\alpha}_T\}/B$ is less than the prescribed level. Using t_α an analogous test can be constructed. B and b refer to the number of bootstrap replications that are used.

Remark 1. Under reasonable conditions the bootstrap distribution can be used as an approximation to the sampling distribution of $\hat{\alpha}_T$ and $t_{\alpha,T}$. Thus the result due to Phillips (1988) referred to earlier will together with Theorem 1 imply that for all x

$$P^*(T(\hat{\alpha}_T^* - 1) \leq x) - P(T(\hat{\alpha}_T - 1) \leq x) \to 0 \text{ in probability as } T \to \infty$$

with a similar result for $t_{\alpha,T}$ and $t_{\alpha,T}^*$.

Remark 2. There are also other statistics than $\hat{\alpha}_T$ and $t_{\alpha,T}$ that are of interest. Often one will fit models of the type

$$Y_t = \mu + \alpha Y_{t-1} + U_t, (t = 1, \ldots, T),$$

where μ and α are estimated by ordinary least squares. When the distribution of $\{Y_t\}$ is governed by (1) there will be analogous results as those in Theorem 1 for this version of the estimator for α.

Remark 3. By estimating σ_u^2 and σ^2 one can modify $\hat{\alpha}_T^*$ and $t_{\alpha,T}^*$ so that the limit distributions do not depend the unknown parameters σ_u^2 and σ^2, i.e. the modifications are asymptotic pivots. This may be an advantage, since it is a conclusion from other situations that the approximations are improved by basing the bootstraps on pivotal- or approximate pivotal quantities. In the present case the situation may be more problematic since the modification involves estimating the variance σ^2 which can be difficult. For more discussion on this important issue see Li and Maddala (1997, p. 304).

Remark 4. In practice the order of the autoregressive approximation will often have to be data based. The methods that are used for the augmented Dickey-Fuller test, see e.g. Ng and Perron (1995), are natural candidates. A more informal procedure is to choose k so large that the estimated residuals are well behaved.

2.2 *Finite-sample simulations*

To get some idea of the finite sample performance of the methods of the previous section we have conducted some Monte Carlo experiments. Although the result from Theorem 1 only cover the distributions under the hypothesis we use the simulations to consider the performance of the test also under some reasonable alternatives. The observations were generated according to the recursion

$$Y_t = \alpha Y_{t-1} + U_t, t = 1, \ldots, T, Y_0 = 0.$$

for $\alpha = 0.7, 0.8, 0.9, 0.95, 0.99$ and 1.0. The series $\{U_t\}, t = -19, \ldots, T$ is an $ARMA(1,1)$ process $U_t = \phi U_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$ started at 0 and with the 20 first observations deleted to approximate the stationary initial distribution of U_1. The sequence $\{\epsilon_t\}$ consists of independent standard normal variates generated by the random number generator ran2 in Press et. al.(1992). Thus, in constructing the recursive bootstrap samples the recursion defined by (2) is started at 0.0 and the first 20 values are discarded in order to approximate the stationary distribution.

The tables are based on 10000 samples and 10000 bootstrap replications in each sample and show the fraction of samples for which the unit root hypothesis is rejected at significance level 0.05 using the $\hat{\alpha}^*$ and the t^*_α tests.

For all specifications of $\{U_t\}$that we consider the difference between the power of two bootstrap tests are small.

Table 1 approximately here

Table 1 shows the result where the stationary process $\{U_t\}$ is independent standard normal random variables for $T = 50$ and 100. For comparison also the power of the tests based on $\hat{\alpha}$ and t_α are included. The critical values are taken from the simulations in Tables 10.A.1 and 10.A.2 in Fuller (1996).

For both the $\hat{\alpha}^*$ and the t_α^* tests the power functions decrease with increasing k and using larger values of k than the correct one, $k = 0$, can lead too substantial loss of power.

In Table 2 $MA(1)$ processes are considered. For positive and moderately negative values of θ it is possible to find values of k such that the level is correct and at the same time the power is reasonable. However, the result for $\theta = -0.8$ indicates that it is necessary with k as large as 12 to obtain a reasonable nominal level.

Table 2 approximately here .

Table 3 confirms the impression from Table 2 for moderately large coefficients.

Table 3 approximately here

3 The stationary bootstrap

3.1 *Algorithm and asymptotic distribution*

The bootstrap algorithm we consider is defined as follows.

Compute the $T - 1$ differences

$$U_t = \Delta Y_t = Y_t - Y_{t-1}, (t = 2, \ldots, T)$$

where $Y_1, \ldots, Y_T$ are observations from a time series satisfying (1). Define the centered differences $X_t = X_{t,T} = U_t - \bar{U}_T$ where $\bar{U}_T = \sum_{j=2}^T U_j/(T-1)$. Let $B_{t,b} = B_{t,b}^T = \{X_t, \ldots, X_{t+b-1}\}$ be a block consisting of b succeeding centered differences starting with time t. To take care of the case where $s > T$ define $X_s = X_t$ when $s - t$ is a multiple of $T - 1$.

A bootstrap sample is generated in the following way. Let $I_1, I_2, \ldots$ be independent random variables uniformly distributed on $\{2, \ldots, T\}$ and independent of the independent geometrically distributed random variables $L_1, L_2, \ldots$ which have frequency distribution $P(L_i = m) = p(1-p)^{m-1}, (m = 1, 2, \ldots)$. Then form the blocks $B_{I_1,L_1}, B_{I_2,L_2}, \ldots$ until T elements are included. Denote the set by $\{X_1^*, \ldots, X_T^*\}$. Now the bootstrap sample, which we again denote by $Y_1^*, \ldots, Y_T^*$, is computed by the recursion

$$Y_t^* = Y_0^* + \sum_{s=1}^t X_s^*, (t = 1, \ldots, T)$$

where Y_0^* is generated according to a fixed distribution which should reflect the distribution in Assumption 1. The conditional distribution of $Y_1^*, \ldots, Y_T^*$ is also in this case denoted by P^*.

We can then compute $\hat{\alpha}_T^*$ and $t_{\alpha,T}^*$ from the formulas (3) and (4) and obtain a bootstrap approximation to the distribution of $T(\hat{\alpha}_T - 1)$ and $t_{\alpha,T}$ based on the stationary bootstrap.

We shall consider the asymptotic distribution of the stationary bootstrap under the following conditions, see Politis and Romano (1994),

Assumption 3
(i) The time series $\{U_t\}$ is strictly stationary with $E(U_t) = 0$ for all t;
(ii) If $\gamma(k) = E[U_t U_{t+k}]$, then $\gamma(0) + \sum_{r=0}^{\infty} |r\gamma(r)| < \infty$;
(iii) $\sum_{r,s,t} \kappa_4(r,s,t) = K < \infty$ where $\kappa_4(r,s,t)$ is the fourth joint cumulant of the distribution of $(U_j, U_{j+r}, U_{j+r+s}, U_{j+r+s+t})$.

The assumption (iii) above is the usual one to ensure that the variance of $\frac{1}{T}\sum_t U_t^2$ tends to zero, see e.g. Priestley (1981, p. 325), and implies that $\sigma^2 = \lim_{T\to\infty} Var(T^{-1/2}\sum_{t=1}^T U_t)$ can be consistently estimated.

The asymptotic properties of the bootstrap approximations are described in the following theorem.

Theorem 2 *Suppose $Y_1, \ldots, Y_T$ is a time series generated according to (1) and satisfies Assumptions 1 and 3. Then for all x, if $p = p_T \to 0$ and $Tp^3 \to \infty$ as $T \to \infty$*

(i) $P^*(T(\alpha_T^* - 1) \leq x) \to P(\frac{1}{2}(W(1)^2 - \frac{\sigma_u^2}{\sigma^2}) / \int_0^1 W(s)^2 ds \leq x)$

(ii) $P^*(t_{\alpha,T}^* \leq x) \to P(\frac{\sigma}{2\sigma_u}(W(1)^2 - \frac{\sigma_u^2}{\sigma^2})/(\int_0^1 W(s)^2 ds)^{1/2} \leq x)$

in probability, where W is a standard Brownian motion.

As for the recursive bootstrap tests having correct asymptotic level for $H_0 : \alpha = 1$ against $H_1 : \alpha < 1$ can now be constructed based on whether $\#\{b : \hat{\alpha}_T^* \leq \hat{\alpha}_T\}/B$ and $\#\{b : t_{\alpha,T}^* \leq \hat{t}_{\alpha,T}\}/B$ are less than the prescribed level.

Remark 1. The tuning parameter p plays a role which in many respect is similar to the length of the fitted autoregression in the recursive bootstrap. Politis and Romano (1994) contains some advice on how p may be chosen. It seems to be more problematic to determine than the lag length k.

3.2 *Finite-sample simulations*

We use the same setup as described for the recursive bootstrap. Again we remark the small differences between the power of the two bootstrap tests for all specifications of the stationary process in (1).

From Table 4 we see that when the time series $\{U_t\}$ is independent standard normal variables choosing small values of p will lead to loss of power and to empirical levels that are smaller than the nominal 0.05. In this respect p has the same property as the order of the fitted autoregressive process, k, in the recursive bootstrap.

Table 4 approximately here

In Table 5 $MA(1)$ processes are considered. Since the dependence is limited in this kind of processes one should expect the stationary bootstrap to perform well. The performance is satisfactory when the parameter θ is positive or moderately negative. For the problematic case with large negative values of θ the choice of p is critical. This is not so surprising since it is known that the unit root tests behave badly in this situation, see e.g. Schwert (1989). But even for $\theta = -0.4$, the mean block length $1/p$ has to be at least 10 to ensure that the size of the test is less than the nominal level, that is, much longer than the dependence may suggest.

Table 5 approximately here

Table 6 shows the performance for stationary processes that also has an autoregressive component. We see that in these cases the presence of autoregressive components can lead to substantial loss of power due to low empirical levels of the tests. This is regardless the choice of p. This confirms our statement in remark 4 after Theorem 1, that for the stationary bootstrap specification of p can be more difficult than the autoregressive lag, k, for the recursive bootstrap.

Table 6 approximately here

4 Proofs

The proofs of both theorems use the following well known expression for $\hat{\alpha}_T^*$, see e.g. Phillips (1987),

$$T(\hat{\alpha}_T^* - 1) = \frac{T\sum_{t=2}^T Y_{t-1}^* \Delta Y_t^*}{\sum_{t=2}^T Y_{t-1}^{*2}} = \frac{\frac{1}{2}\{(\frac{1}{\sqrt{T}}Y_T^*)^2 - \frac{Y_1^{*2}}{T} - \frac{\sum \Delta Y_t^{*2}}{T}\}}{\frac{\sum Y_{t-1}^{*2}}{T^2}}. \tag{6}$$

We shall consider weak convergence in $D[0,1]$, the space of functions defined on the interval $[0,1]$ being continuous from the right and having left limits. Pollard (1984) stressed that, when the limiting distribution is concentrated on the continuous functions, it is convenient to consider the uniform metric in $D[0,1]$. In our case the limit will be a Brownian motion and we therefore work with the uniform metric.

Proof of Theorem 1. By definition $U_t^* = Y_t^* - Y_{t-1}^*, (t = 2, \ldots, T)$ where $Y_t^*, (t = 1, \ldots, T)$ is the bootstrap sample. If $\mathbf{U}_t^* = (U_t^*, \ldots, U_{t-k+1}^*)'$ and $\boldsymbol{\epsilon}_t^* = (\epsilon_t^*, 0, \ldots, 0)'$, the recursion for constructing the bootstrap can be written

$$\mathbf{U}_t^* = \hat{A}_k \mathbf{U}_{t-1}^* + \boldsymbol{\epsilon}_t^*, (t = 1, \ldots, T)$$

where

$$\hat{A}_k = \begin{pmatrix} \hat{\phi}_{1k} & & \cdots & & \hat{\phi}_{kk} \\ 1 & 0 & \cdots & & 0 \\ \vdots & & & & \vdots \\ 0 & & \cdots & 1 & 0 \end{pmatrix}.$$

Solving the recursion backwards yields

$$\mathbf{U}_t^* = \sum_{i=0}^{t-1} \hat{A}_k^i \, \boldsymbol{\epsilon}_{t-i}^* + \hat{A}_k^t \mathbf{U}_0^* \tag{7}$$

so that

$$\sum_{j=1}^{t} \mathbf{U}_j^* = \sum_{j=0}^{t-1}\sum_{i=0}^{j} \hat{A}_k^i \, \boldsymbol{\epsilon}_{t-j}^* + \sum_{j=1}^{t} \hat{A}_k^j \mathbf{U}_0^*.$$

Here $\mathbf{U}_0^* = (U_0^*, \ldots, U_{-k+1}^*)'$ is a vector of k stationary random variables satisfying (2).

From the identity

$$I - \hat{A}_k^{j+1} = (I - \hat{A}_k)(I + \cdots + \hat{A}_k^j)$$

it follows that the sum may be written, when $I - \hat{A}_k$ is nonsingular,

$$(I-\hat{A}_k)^{-1}\sum_{j=0}^{t-1} \epsilon^*_{t-j} - (I-\hat{A}_k)^{-1}\hat{A}_k\sum_{j=0}^{t-1}\hat{A}_k^j\, \epsilon^*_{t-j} + (I-\hat{A}_k)^{-1}\hat{A}_k(I-\hat{A}_k^t)\mathbf{U}^*_0.$$

This can be simplified. Inserting (7) the expression becomes after some algebra, since $Y_t^* - Y_0^* = \sum_{j=1}^t \mathbf{e}_1'\mathbf{U}_j^*$ when $\mathbf{e}_1 = (1,0,\ldots,0)'$,

$$Y_t^* - Y_0^* = \mathbf{e}_1'(I-\hat{A}_k)^{-1}\sum_{j=0}^{t-1} \epsilon^*_{t-j} + \mathbf{e}_1'(\mathbf{U}^*_t - \mathbf{U}^*_0) - \mathbf{e}_1'(I-\hat{A}_k)^{-1}(\mathbf{U}^*_t - \mathbf{U}^*_0) \quad (8)$$

for $(t = 1,\ldots,T)$.

Now use that

$$\mathbf{e}_1'(I-\hat{A}_k)^{-1} = \frac{1}{1-\hat{\phi}_{1k}-\cdots-\hat{\phi}_{kk}}(1, \hat{\phi}_{2k}+\cdots+\hat{\phi}_{kk}, \ldots, \hat{\phi}_{kk})$$

so that

$$\mathbf{e}_1'(I-\hat{A}_k)^{-1}\sum_{j=1}^{t} \epsilon^*_j = \frac{1}{1-\hat{\phi}_{1k}-\cdots-\hat{\phi}_{kk}}\sum_{j=1}^{t}\epsilon^*_j.$$

Under the conditions in the theorem it is shown in Theorem 2 in Berk (1972) that

$$1/(1-\hat{\phi}_{1k}-\cdots-\hat{\phi}_{kk}) \to \sqrt{2\pi f_U(0)/\sigma^2_\epsilon} = \sigma/\sigma_\epsilon \text{ in probability}$$

where f_U is the spectral density of the time series $\{U_t\}$ defined in (5).

Now consider the random function $S_T^*(s) = \frac{1}{\sqrt{T}}\sum_{j=1}^{[Ts]}\epsilon^*_j, 0 \le s \le 1$ in $D[0,1]$. As usual $[\cdot]$ denotes the integer value. From Theorem V.19 in Pollard (1984) it follows that for all bounded continuous functions f defined on $D[0,1]$

$$E^*[f(S_T^*)] \to E[f(\sigma W)] \quad (9)$$

in probability. W denotes the standard Brownian motion.

To see that we first remark that $S_T^*(0) = 0$. Also S_T^* has independent increments given the observations $Y_1,\ldots,Y_T$, i.e. if $0 \le s < t < u \le 1$

$$P^*(S_T^*(u)-S_T^*(t) \le x_1, S_T^*(t)-S_T^*(s) \le x_2) = P^*(S_T^*(u)-S_T^*(t) \le x_1)P^*(S_T^*(t)-S_T^*(s) \le x_2)$$

for all real x_1 and x_2. This is a direct consequence of the definition of S_T^*. Furthermore, for all real r

$$\begin{aligned} E^*[exp\{ir(S_T^*(u) - S_T^*(t)\}] &= [E^*\{exp(ir\epsilon_1^*/\sqrt{T})\}]^{([Tu]-[Tt])} \\ = (1 - E^*(\epsilon_1^{*2})/2T + o_P(1))^{T(u-t)} &\to exp\{r(u-t)\sigma_\epsilon^2/2\} \end{aligned}$$

in probability since $E^*(\epsilon_t^{*j}) \to E(\epsilon_t^j), (j = 2, 4)$ in probability from Lemma 5.3 in Bühlmann (1997). Hence the increments $S_T^*(u) - S_T^*(t), (t < u)$ converge in P^* distribution to the distribution of $\sigma_\epsilon(W(u) - W(t))$. Finally, if $\delta > 0$

$$P^*(|S_T^*(u) - S_T^*(t)| \geq \delta/2) \leq 4E^*\{S_T^*(u) - S_T^*(t)\}^2/\delta^2 = 4([Tu] - [Tt])E^*(\epsilon_1^{*2})/T\delta^2$$

which tends to $4(u - t)\sigma_\epsilon^2/\delta^2$ in probability as $T \to \infty$. This concludes the verification of the conditions necessary to justify (9).

From Lemma 1 below it follows that $P(I - \hat{A}_k \text{ singular}) \to 0$.

Finally, we must consider the two last terms in (8). Define random elements in $D[0, 1]$ as described above, i.e. let $R_1(s) = \mathbf{e}_1'(\mathbf{U}_{[Ts]}^* - \mathbf{U}_0^*)/\sqrt{T}$ and $R_2(s) = \mathbf{e}_1'(I - \hat{A}_k)^{-1}(\mathbf{U}_{[Ts]}^* - \mathbf{U}_0^*)/\sqrt{T}$ when $0 \leq s \leq 1$. It follows from Lemma 2 and 3 that for all $\eta > 0$ $P^*(\sup_{0 \leq s \leq 1} |R_i(s)| > \eta), (i = 1, 2)$ tend to zero in probability as $T \to \infty$. This means, since we use the uniform norm, that the random functions tend to zero as elements in $D[0, 1]$, i.e. $E^*[f(R_i)] \to 0$ in probability for all bounded continuous functions f defined on $D[0, 1]$.

Hence, it follows from Assumption 1 and the continuous mapping theorem that in probability the conditional distribution of $(\frac{1}{T}Y_T^{*2}, \frac{1}{T^2}\sum_{t=1}^T Y_t^{*2})$ converges in distribution towards $(\sigma^2 W(1)^2, \sigma^2 \int_0^1 W(s)^2 ds)$.

We can then use the expression in the beginning of this section to complete the proof of the theorem by appealing to Slutsky's theorem, since from Theorem 3.1(i) in Bühlmann (1997) for all $\eta > 0$

$$P^*(|\frac{1}{T}\sum U_t^{*2} - \sigma_u^2| > \eta) \to 0$$

in probability. Also $s_U^2 \to \sigma_u^2$ in probability, where s_U^2 is defined after (4).

Lemma 1 *Under Assumption 2, $P(I - \hat{A}_k$ singular$) \to 0$ as $T \to \infty$*

Proof. Since $det|\lambda I - \hat{A}_k| = \lambda^k - \hat{\phi}_{1k}\lambda^{k-1} - \cdots - \hat{\phi}_{kk}$, $I - \hat{A}_k$ is nonsingular if and only if $1 - \hat{\phi}_{1k} - \cdots - \hat{\phi}_{kk} \neq 0$. But as proved in Theorem 2 in Berk (1972), $1 - \hat{\phi}_{1k} - \cdots - \hat{\phi}_{kk} \to (2\pi f_U(0)/\sigma_\epsilon^2)^{-1/2} \neq 0$ in probability where f_U is the spectral density of the time series $\{U_t\}$.

Let $\phi_{jk}, (j = 1, \ldots, k)$ be the coefficients minimizing $E(U_t - c_1 U_{t-1} - \cdots - c_k U_{t-k})^2$ with minimum σ_k^2. Then $\sigma_k^2 \to \sigma_\epsilon^2$ as $k \to \infty$, see Berk (1972). In the proof of the following lemmas we shall also make use of the fact that the fitted autoregressions are always causal, see e.g Brockwell and Davis (1991, p.240). The random variable U_t^* therefore has an $MA(\infty)$ representation $U_t^* = \sum_{j=0}^\infty \hat{\psi}_{j,T}\epsilon_{t-j}^*$. where $\sum |\hat{\psi}_{j,T}| < \infty$ and the coefficients are defined by the power series $\hat{\Psi}_T(z) = \sum \hat{\psi}_{j,T} z^j = 1/\sum_{j=0}^k \hat{\phi}_{j,k} z^j$. Finally we shall need a result proved in Theorem 3.1 in Bühlmann (1995) where it is shown that there exists a random variable T_0 so that $sup_{T > T_0} \sum_{j=0}^\infty |\hat{\psi}_{j,T}| < \infty$ almost surely.

Lemma 2 *Under Assumption 2, for all* $\eta > 0$, $P^*(\max_{-k+1 \leq t \leq T} |U_t^*/\sqrt{T}| > \eta) \to 0$ *in probability as* $T \to \infty$.

Proof. Since $k = o(T^{1/3})$, $(T+k)/T \to 1$. Furthermore, by the well known equality, see e.g. Hall and Heyde (1980, p. 53),

$$\begin{aligned} & P^*(\max_{-k+1 \leq t \leq T} |U_t^*/\sqrt{(T+k)}| > \eta) \\ = \; & P^*(\frac{1}{(T+k)} \sum_{t=-k+1}^{T} U_t^{*2} I[U_t^* > \eta\sqrt{(T+k)}] > \eta^2) \\ \leq \; & \frac{1}{\eta^2} E^*(U_1^{*2} I(U_1^* > \eta\sqrt{(T+k)}) \leq \frac{1}{\eta^4 (T+k)} E^*(U_1^{*4}). \end{aligned}$$

But $E^*(U_1^{*4}) \leq \sum_j \hat{\psi}_{j,T}^4 E^*(\epsilon_1^{*4}) + (\sum_j \hat{\psi}_{j,T}^2)^2 (E^*(\epsilon_1^{*2})^2$ and $\sum_j \hat{\psi}_{j,T}^k \leq \sup_j |\hat{\psi}_{j,T}|^{k-1} \sum_j |\hat{\psi}_{j,T}|$ $\leq (\sum_j |\hat{\psi}_{j,T}|)^k, (k = 2, 4)$. By using the result in Theorem 3.1 in Bühlmann (1995) referred to above the conclusion of the Lemma follows.

Lemma 3 *Under Assumption 2, for all* $\eta > 0$, $P^*(|max_{1 \leq t \leq T} \mathbf{e_1}'(I - \hat{A}_k)^{-1} \mathbf{U}_t^*/\sqrt{T}| > \eta) \to 0$ *in probability as* $T \to \infty$.

Proof. Using the expression for $\mathbf{e}'(I - \hat{A}_k)^{-1}$ we may write

$$\mathbf{e_1}'(I - \hat{A}_k)^{-1} \mathbf{U}_t^* = [U_t^* + (\hat{\phi}_{2k} + \cdots + \hat{\phi}_{kk}) U_{t-1}^* + \cdots + \hat{\phi}_{kk} U_{t-k}^*]/[1 - \hat{\phi}_{1k} - \cdots - \hat{\phi}_{kk}].$$

By the argument used in Lemma 1 the denominator converges in probability. Consider the numerator. Then

$$\begin{aligned} \sum_{j=2}^{k} |\sum_{l=j}^{k} \hat{\phi}_{lk}| \; & \leq \; \sum_{j=2}^{k} |\sum_{l=j}^{k} \phi_{lk}| + \sum_{j=2}^{k} \sum_{l=j}^{k} |\hat{\phi}_{lk} - \phi_{lk}| \\ \leq \sum_{j=2}^{k} j|\phi_{jk}| + \sum_{j=2}^{k} j|\hat{\phi}_{jk} - \phi_{jk}| \; & \leq \; \sum_{j=2}^{k} j|\phi_{jk}| + (k(k+1)/2) max_{j \leq k} |\hat{\phi}_{jk} - \phi_{jk}|. \end{aligned}$$

Now $\sum_{j=2}^{k} j|\phi_{jk}| < \infty$ where the bound is independent of k as k varies. This follows from Baxter's inequality, see e.g. Hannan and Deistler (1988, p.269) and the assumption that $\sum_{j=2}^{k} j|\phi_j| < \infty$. Also $k = o(T^{1/3})$, hence $k = o((T/log(T))^{1/2})$, and therefore $max_{j \leq k} |\hat{\phi}_{jk} - \phi_{jk}| = O((log(T)/T)^{1/2})$ almost surely by Theorem 7.4.5 in Hannan and Deistler (1988). Hence $\sum_{j=2}^{k} |\sum_{l=j+1}^{k} \hat{\phi}_{lk}|$ increases as

$$\sum_{j=2}^{k} |\sum_{l=j}^{k} \hat{\phi}_{lk}| = o((\log(T))^{1/2} T^{1/6})$$

almost surely. The proof is now concluded arguing as in Lemma 2 by noting that η may be allowed to depend on T and decrease at the rate $o(T^{-1/4})$.

In fact, a more careful inspection will reveal that $\sum_{j=2}^{k} |\sum_{l=j}^{k} \hat{\phi}_{lk}|$ increases at rate $o((\log(T)/T)^{1/2}k^2)$, and this has to be bounded by $T^{1/4}$, which is satisfied if $k^2 = o(T^{3/4-\delta})$ for some $\delta > 0$. Taking $\delta = 1/12$ so that $k = o(T^{1/3})$ does the job.

Proof of Theorem 2. The general outline of the proof follows Politis and Romano (1994). In particular Lemma 4 below is an invariance principle version of their central limit theorem.

The samples $(Y_1^*, \ldots, Y_T^*)$ and the distribution P^* in the present proof always refer to the stationary bootstrap defined in section 3.

We recall the following facts which can be found in Politis and Romano (1994). Let $S_{t,b} = S_{t,b}^T$ be the sum of the elements in block $B_{t,b}$. If the random variables $I = I_T$ and $L = L_T$ are independent, where I is uniformly distributed on $\{2, \ldots, T\}$ and L is geometrically distributed with mean $1/p_T$, the random variable $S_{I,L}$ has expectation $E^*[S_{I,L}] = E^*[E^*[S_{I,L}|L]$. But $E^*[S_{I,L}|L = l] = \frac{1}{T-1} \sum_{k=1}^{l} \sum_{j=2}^{T} X_{k+j} = \frac{l}{T-1} \sum_{j=2}^{T} X_j = 0$ from the definition of $X_s, s > T$ and the centering of the differences used for the construction of the bootstrap samples. Furthermore, $E^*[S_{I,L}^2|L = l] = l[\hat{C}_T(0) + 2\sum_{i=1}^{l-1}(1 - \frac{i}{l})\hat{C}_T(i)]$. Hence

$$E^*[S_{I,L}^2] = \frac{1}{p_T}[\hat{C}_T(0) + 2\sum_{i=1}^{\infty}(1 - p_T)^i \hat{C}_T(i)].$$

Here $\hat{C}_T(\cdot)$ is the so called circular covariance defined by

$$\hat{C}_T(i) = \frac{1}{T-1} \sum_{j=2}^{T} X_{j,T} X_{j+i,T} \text{ where } X_{j,T} = U_j - \bar{U}_T, (j = 2, \ldots, T)$$

and $X_{j,T} = X_{j-(T-1),T}$ when $j > T$.

In Politis and Romano (1994, p.1302) it is also shown that

$$\hat{C}_T(0) + 2\sum_{i=1}^{\infty}(1 - p_T)^i \hat{C}_T(i) \to \sigma^2 \tag{10}$$

in probability as $T \to \infty$.

Now let $L_1, L_2, \ldots$ be the independent geometrically distributed random variables used to define $Y_t^*, (t = 1, \ldots, T)$. Define $M_t = \inf\{k : L_1 + \cdots + L_k \geq t\}, (t = 1, \ldots, T)$ and let $N = N_t = \sum_{i=1}^{M_t} L_i$. Then define

$$Y_t^{**} = Y_N^*, (t = 1, \ldots, T)$$

as the sum of observations in N blocks of observations of lengths $L_1, \ldots, L_N$ and starting at observation $I_1, \ldots, I_N$. Also notice that the process $\{M_t - 1\}$

may alternatively be represented as indicating the number of successes in $t-1$ Bernoulli trials with probability of success p_T.

The processes Y_t^* and Y_t^{**} depend on T, but for notational reasons we do not index these processes on T in the sequel. Let $R = R_T = M_T - T$. Then R is geometrically distributed and $Z_T = \Delta Y_{T+1}^* + \cdots + \Delta Y_{T+R}^*$ has the same distribution as $S_{I,L}$.

Now, to prove Theorem 2 we write the right hand side of the expression (6) as

$$\frac{1}{2}\frac{\sum Y_{t-1}^{**2}}{\sum Y_{t-1}^{*2}}\frac{\frac{1}{T}(Y_T^{**} - Z_T)^2 - \frac{Y_1^{2*}}{T} - \frac{\sum \Delta Y_t^{*2}}{T}}{\frac{\sum Y_{t-1}^{**2}}{T^2}}.$$

From the continuous mapping theorem and Lemma 4 below it follows that in probability the conditional distribution of $(\frac{1}{T}Y_T^{**2}, \frac{1}{T^2}\sum Y_t^{**2})$ given $Y_1, \ldots, Y_T$ converges in distribution towards $(\sigma^2 W(1)^2, \sigma^2 \int_0^1 W(s)^2 ds)$.

Furthermore, from Lemma 5 it follows that $P^*(|\frac{\sum Y_{t-1}^{**2}}{\sum Y_{t-1}^{*2}} - 1| > \epsilon) \to 0$ for all $\epsilon > 0$ in probability when $p_T \to 0$ and $Tp_T^3 \to \infty$ and from Chebychev's inequality we get $P^*(|Z_T|/\sqrt{T} > \epsilon) \to 0$.

Slutsky's theorem now implies that in probability the conditional distribution of $T(\hat{\alpha}_T^* - 1)$ given $Y_1, \ldots, Y_T$ tends to the distribution of $\frac{1}{2}(W(1)^2 - \frac{\sigma_u^2}{\sigma^2})/\int_0^1 W(s)^2 ds$ since from Lemma 6 for all $\epsilon > 0$

$$P^*(|\frac{\sum \Delta Y_t^{*2}}{T} - \sigma_u^2| > \epsilon) \to 0$$

in probability as $T \to \infty$. A similar result also holds for $t_{\alpha,T}^*$.

It remains to verify Lemmas 4-6. Remark that in Lemmas 4 and 6 we only need $Tp_T \to \infty$. Define X_T^{**} by $X_T^{**}(s) = \frac{1}{\sqrt{T}}Y_{M_{[Ts]}}^{**}, 0 \le s \le 1$. Then X_T^{**} is an element in $D[0,1]$. The weak convergence result we need is contained in the following lemma.

Lemma 4 *Let W is a standard Brownian motion in $D[0,1]$. Under Assumptions 1 and 3*

$$E^*[f(X_T^{**})] \to E[f(\sigma W)]$$

in probability if $p_T \to 0$ and $Tp_T \to \infty$ for all bounded, continuous functions f on $D[0,1]$.

Proof. The process X_T^{**} has independent increments. To verify this property let $0 \le u < v < w \le 1$ and let $V_j, (j = 1, \ldots)$ be the sum of observations in the blocks of length L_j starting at observation I_j. Then for real numbers x_1 and x_2

$$P^*(X_T^{**}(v) - X_T^{**}(u) \le x_1, X_T^{**}(w) - X_T^{**}(v) \le x_2) =$$
$$\sum_{m,n} P^*(V_1 + \cdots + V_m \le x_1\sqrt{T}, V_1' + \cdots + V_n' \le x_2\sqrt{T} | M_{[Tv]} - M_{[Tu]} = m, M_{[Tw]} - M_{[Tv]} = n)$$
$$P(M_{[Tv]} - M_{[Tu]} = m, M_{[Tw]} - M_{[Tv]} = n)$$

where V_i and V_i' refer to independent pairs $(I_i, L_i), (I_i', L_i'), (i = 1, \ldots)$. Now, use the alternative representation of the process $M_t - 1$ as indicating the number of successes in $t-1$ Bernoulli trials with success probability p_T. Then clearly $M_{[Tv]} - M_{[Tu]}$ and $M_{[Tw]} - M_{[Tv]}$ are independent. Using this result in the expression above it is seen to reduce to $P^*(X_T^{**}(v) - X_T^{**}(u) \leq x_1) P^*(X_T^{**}(w) - X_T^{**}(v) \leq x_2)$.

From Pollard (1984), Theorem V.19, the result will then follow from the following two properties

(i) The P^* distribution of the increments $X_T^{**}(s) - X_T^{**}(r)$ for each pair $r < s$ converge in probability towards a $N(0, (s-r)\sigma^2)$ random variable for each pair $r < s$,

(ii) for all $\delta > 0$ the probabilities $P^*(|X_T^{**}(s) - X_T^{**}(r)| > \delta)$ are arbitrarily small when $s - r \to 0$, uniformly in r, s belonging to $[0,1]$.

The property (ii) follows from Markov's lemma since for all $\delta > 0$

$$P^*(|X_T^{**}(s) - X_T^{**}(r)| > \delta) \leq E^*[\{X_T^{**}(s) - X_T^{**}(r)\}^2]/\delta^2.$$

From the definition of X_T^{**} it follows that $X_T^{**}(s) - X_T^{**}(r) = \frac{1}{\sqrt{T}} \sum_{j=M_{[Tr]}+1}^{M_{[Ts]}} V_j$. The random variables $V_j, (j = 1, \ldots)$ are independent identically distributed as $S_{I,L}$. Since $E(S_{I,L}) = 0$, $\sum_{j=1}^{n} V_j$ is therefore a martingale with respect to the σ-algebras $\sigma(V_j, I_j, L_j, j \leq n)$. The random variables $M_{[Tr]}$ and $M_{[Ts]}$ are stopping times so that by reasoning as in the proof of Wald's equation in Neveu (1975, Proposition IV-4-21),

$$E^*[\{X_T^{**}(s) - X_T^{**}(r)\}^2] = \frac{1}{T} E^*[\sum_{j=M_{[Tr]}+1}^{M_{[Ts]}} V_j]^2 = \frac{1}{T} E(M_{[Ts]} - M_{[Tr]}) E^*[S_{I,L}^2].$$

By applying Wald's equation on the sum $L_1 + \cdots + L_n, (n = 1, \ldots)$ we get $|E(M_{[Ts]} - M_{[Tr]}) - T(s-r)p_T| \leq 1$. Thus by (10) property (ii) follows.

The property (i) can as explained in Politis and Romano (1994), be deduced from an extension of the central limit theorem for a sum with a random number of terms, see e.g Chung (1974), to a triangular array setting.

Lemma 5 *Under Assumptions 1 and 3 for any* $\epsilon > 0$

$$P^*(|\frac{\sum Y_{t-1}^{**2}}{\sum Y_{t-1}^{*2}} - 1| > \epsilon) \to 0$$

in probability if $p_T \to 0$ *and* $T p_T^3 \to \infty$ *as* $T \to \infty$.

Proof. We start from the equality

$$Y_t^{**2} - Y_t^{*2} = 2(Y_t^{**} - Y_t^*)Y_t^{**} - (Y_t^{**} - Y_t^*)^2$$

From Cauchy-Schwarz's inequality

$$\sum (Y_t^{**} - Y_t^*)Y_t^{**} \leq [\sum (Y_t^{**} - Y_t^*)^2]^{1/2}[\sum Y_t^{**2}]^{1/2}.$$

It follows from Lemma 4 that the distribution of the random variable $\frac{1}{T^2}\sum Y_t^{**2}$ converges in distribution.

Hence it is sufficient to show that

$$P^*(|\frac{1}{T^2}\sum (Y_t^{**} - Y_t^*)^2| > \epsilon) \to 0$$

in probability for any $\epsilon > 0$.

Let $W_{j,k}$ denote the k'th ($k = 1, \ldots$) observation in the block starting at observation I_j. Then, with $M = M_T$ and $I = I_M$,

$$\begin{aligned} \frac{1}{T^2}\sum (Y_t^{**} - Y_t^*)^2 &= \frac{1}{T^2}\sum_{j=1}^{M-1}(\sum_{k=0}^{L_j-1} kW_{j,k})^2 + \frac{1}{T^2}(\sum_{k=0}^{T-L_1+\cdots+L_{M-1}} kW_{I,k})^2 \\ &+ \frac{1}{T^2}(M-T)^2(\sum_{k=T-L_1+\cdots+L_{M-1}+1}^{L_1+\cdots+L_M} W_{I,k})^2 \qquad (11) \end{aligned}$$

which we write

$$\frac{1}{T^2}\sum_{j=1}^{M-1}(\sum_{k=0}^{L_j-1} kW_{l,k})^2 + Z_2 + Z_3 = \frac{1}{T^2}\sum_{j=1}^{M}(\sum_{k=0}^{L_j-1} kW_{j,k})^2 - Z_1 + Z_2 + Z_3. \qquad (12)$$

To bound the first term in (12) define, suppressing the index T,

$$S_{i,b}^k = \sum_{j=0}^{b-1} j^k X_{i+j}$$

so that $S_{i,b}^0 = S_{i,b}$ is the sum of the elements in the block $B_{i,b}^T$. Defining the corresponding random variables $V_j^k, (j = 1, \ldots)$ for the block defined by $(I_j, L_j), (j = 1, \ldots)$. Then the expectation of the first term in (12) can be written

$$\frac{1}{T^2}E^*[\sum_{j=1}^{M}(V_j^1)^2].$$

Now for fixed T the random variables $V_j^1, (i = 1, \ldots)$, are independent, so that

$$\sum_{j=1}^{n}(V_j^1)^2 - E^*[(V_j^1)^2]$$

is a sum of independent, identically distributed random variables with expectation zero. Then we can argue as in the proof of Lemma 4 using Wald's equation so that

$$\begin{aligned} E^*[\sum_{j=1}^{M}(V_j^1)^2] &= E^*[M]E^*[(S_{I,L}^1)^2] \\ &\leq (Tp_T+1)E^*[(S_{I,L}^1)^2]. \end{aligned}$$

Furthermore, from the Cauchy-Schwarz inequality

$$\begin{aligned} E^*[(S_{I,L}^1)^2] &= \frac{1}{T-1}\sum_{i=2}^{T}\sum_{l=0}^{\infty}(\sum_{j=1}^{l-1} jX_{i+j})^2 p_T(1-p_T)^{l-1} \\ &\leq \frac{1}{T-1}\sum_{i=2}^{T}\sum_{l=0}^{\infty}(\sum_{j=1}^{l-1} j^2)(\sum_{j=1}^{l-1} X_{i+j}^2)p_T(1-p_T)^{l-1} \\ &\leq const.\sum_{l=0}^{\infty} l^3\sum_{j=1}^{l-1}(\frac{1}{T-1}\sum_{i=2}^{T} X_{i+j}^2)p_T(1-p_T)^{l-1}. \end{aligned}$$

Due to the definition of $X_s, s > T$, the sum $\frac{1}{T-1}\sum_{i=2}^{T} X_{i+j}^2$ is independent of j and equals $\frac{1}{T-1}\sum_{i=2}^{T} X_i^2$. Hence,

$$E^*[(S_{I,L}^1)^2] \leq const.E[L^4]\frac{1}{T-1}\sum_{i=1}^{T} X_i^2$$

Since $E[L^4] = const.p_T^{-4}$, this means that the first term in (12) is bounded by $const.(\frac{1}{T-1}\sum_{i=2}^{T} X_i^2)(Tp_T+1)/T^2p_T^4$ which tends to zero in probability since $\frac{1}{T-1}\sum_{i=2}^{T} X_i^2 = \frac{1}{T-1}\sum_{i=2}^{T}(U_i-\bar{U}_T)^2 \to \sigma_u^2$ in probability.

Using the Cauchy-Schwarz inequality again it is seen that $E|Z_i|, (i = 1, 2, 3)$ also are bounded by $const.(\frac{1}{T-1}\sum_{i=2}^{T} X_i^2)E^*[L^4]/T^2$.

Lemma 6 *Under Assumptions 1 and 3 for any* $\epsilon > 0$

$$P^*(|\frac{\sum \Delta Y_t^{*2}}{T} - \sigma_u^2| > \epsilon) \to 0$$

in probability if $p_T \to 0$ *and* $Tp_T \to \infty$ *as* $T \to \infty$.

Proof. Denote M_T by M. Consider the random variable $Z = \sum_{j=0}^{L-1} X_{I+j}^2/L - \sigma_u^2$ defined on the block of $B_{I,L}$, and let $Z_j, (j = 1, \ldots)$ be similarly defined on the blocks defined by the pairs (I_j, L_j). Then $\sum \Delta Y_t^{*2}/T$ may be written

$$\sigma_u^2\frac{\sum_{j=1}^{M} L_j}{T} + \frac{\sum_{j=1}^{M} L_jZ_j}{T} - \frac{\sum_{t=T+1}^{M} \Delta Y_t^{*2}}{T}. \tag{13}$$

Since $T \leq \sum_{j=1}^{M} L_j \leq T + L_M$ and $E[L_M]/T = 1/Tp_T \to 0$, the first term in (13) converges in probability to σ_u^2.

By using arguments from the previous Lemmas $E^*[|\sum_{j=1}^{M} L_j Z_j|] \leq E^*[\sum_{i=1}^{M} L_j |Z_j|] = E[M]E^*[L|Z|] = (T+1)p_T E^*[L|Z|]$. Also $E^*[L|Z|] = E^*[|\sum_{j=0}^{L-1} X_{I+j}^2 - L\sigma_u^2|] = \sum_{l=1}^{\infty} |\sum_{j=2}^{T} X_{j+l}^2/(T-1) - \sigma_u^2| l p_T(1-p_T)^{l-1} = |\sum_{j=2}^{T} X_j^2/(T-1) - \sigma_u^2|/p_T$. Since $\sum_{j=2}^{T} X_j^2/(T-1) = \sum_{j=2}^{T}(U_j - \bar{U}_T)^2/(T-1) \to \sigma_u^2$ in probability, it then follows that the expectation of the absolute value of the second term in (13) tends to zero in probability.

The last term also clearly tends to 0 in probability, which concludes the proof of the Lemma.

References

[1] BASAWA, I.V., MALLIK, A.K., MCCORMICK,W.P.,REEVES,J.H. and TAYLOR, R.L.(1991a) Bootstrapping unstable first-order autoregressive processes. *Ann. Statist.* **19**, 1098-1101.

[2] BASAWA, I.V., MALLIK, A.K., MCCORMICK,W.P.,REEVES,J.H. and TAYLOR, R.L.(1991b) Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive processes. *Commun. Statist.* A **20**, 1015-1026.

[3] BERK, K.N. (1974) Consistent autoregressive spectral estimates. *Ann. Statist.* **2**, 489-502.

[4] BROCKWELL, P.J. and DAVIS, R.A. (1991). *Time Series: Theory and Methods* (2nd edn). New York: Springer.

[5] BÜHLMANN, P. (1995) Moving-average representation for autoregressive approximations. *Stochastic Process. Appl.* **60**, 331-342.

[6] BÜHLMANN, P. (1997) Sieve bootstrap for time series. *Bernoulli* **3**, 123-148.

[7] CHUNG, K.L. (1974) *A Course in Probability Theory* (2nd edn). New York: Academic Press.

[8] DICKEY, D.A. and FULLER, W.A. (1979). Distributions of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74** , 427-431.

[9] FERRETTI, N. and ROMO, J. (1996) Unit root bootstrap tests for AR(1) models. *Biometrika* **83**, 849-860.

[10] FULLER, W.A. (1996) *Introduction to Statistical Time Series* (2nd edn). New York: John Wiley.

[11] HALL, P. and HEYDE, C.C. (1980) *Martingale Limit Theory and Its Applications.* New York: Academic Press.

[12] HANNAN, E.J. and DEISTLER, M. (1988) *The Statistical Theory of Linear Systems.* New York: John Wiley.

[13] LAHIRI, S.N. (1999) Theoretical comparisons of block bootstrap methods. *Ann. Statist.* **27**, 386-404.

[14] LI, H. and MADDALA, G.S. (1997) Bootstrapping cointegrating regressions. *Journal of Econometrics* **80**, 297-318.

[15] NEVEU, J. (1975) *Discrete-parameter martingales.* Amsterdam: North Holland.

[16] NG, S. and PERRON, P. (1995) Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* **90**, 268-281.

[17] PHILLIPS, P.C.B. (1987) Time series regression with a unit root. *Econometrica* **55**, 277-301.

[18] POLITIS, D.N. and ROMANO, J.P. (1994) The stationary bootstrap. *Journal of the American Statistical Association* **89**, 1303-1313.

[19] POLLARD, D.(1984) *Convergence of Stochastic Processes.* New York: Springer.

[20] PRESS, W.H., TEUKOLSKY, S. A., VETTERLING, W.T. and FLANNERY, B.P. (1992). *Numerical Recipes in Fortran.* (2nd edn.) Cambridge: Cambridge University Press.

[21] PRIESTLEY, M. B.(1981) *Spectral Analysis and Time Series.* London: Academic Press.

[22] SAID, S.E. and DICKEY, D. A. (1984) Testing for unit root in autoregressive-moving average models of unknown order. *Biometrika* **71**, 599-607.

[23] SCHWERT, G. W. (1989) Testing for unit roots: A Monte Carlo investigation. *Journal of Business & Economic Statistics* **7**, 147-159.

T	k	Test	α					
			0.7	0.8	0.9	0.95	0.99	1.0
50		$\hat{\alpha}$	0.98	0.78	0.33	0.15	0.07	0.05
		$\hat{t}_\alpha$	0.97	0.78	0.34	0.15	0.07	0.05
	1	$\hat{\alpha}^*$	0.90	0.65	0.27	0.12	0.05	0.04
		t^*_α	0.89	0.65	0.27	0.12	0.06	0.04
	4	$\hat{\alpha}^*$	0.49	0.29	0.11	0.05	0.02	0.02
		t^*_α	0.50	0.30	0.11	0.05	0.02	0.02
	12	$\hat{\alpha}^*$	0.08	0.03	0.01	0.00	0.00	0.00
		t^*_α	0.07	0.03	0.01	0.00	0.00	0.00
100		$\hat{\alpha}$	1.00	1.00	0.77	0.31	0.07	0.05
		$\hat{t}_\alpha$	1.00	1.00	0.77	0.32	0.08	0.05
	1	$\hat{\alpha}^*$	1.00	0.99	0.69	0.29	0.07	0.05
		t^*_α	1.00	0.99	0.70	0.29	0.08	0.05
	4	$\hat{\alpha}^*$	0.98	0.88	0.47	0.19	0.05	0.03
		t^*_α	0.98	0.89	0.48	0.20	0.05	0.03
	12	$\hat{\alpha}^*$	0.59	0.39	0.14	0.05	0.01	0.01
		t^*_α	0.59	0.39	0.15	0.05	0.01	0.01

Table 1: Stationary process $\{U_t\}$ in (1); independent standard normal variables.

θ	k	Test	α					
			0.7	0.8	0.9	0.95	0.99	1.0
0.8	1	$\hat{\alpha}^*$	1.00	1.00	0.87	0.50	0.16	0.10
		t^*_α	1.00	0.99	0.85	0.49	0.15	0.10
	4	$\hat{\alpha}^*$	0.93	0.77	0.37	0.14	0.04	0.02
		t^*_α	0.93	0.77	0.37	0.15	0.04	0.02
	12	$\hat{\alpha}^*$	0.50	0.33	0.13	0.04	0.01	0.01
		t^*_α	0.51	0.34	0.14	0.05	0.01	0.01
-0.4	1	$\hat{\alpha}^*$	1.00	1.00	0.94	0.57	0.18	0.12
		t^*_α	1.00	1.00	0.95	0.58	0.18	0.12
	4	$\hat{\alpha}^*$	1.00	0.96	0.58	0.24	0.06	0.04
		t^*_α	1.00	0.96	0.58	0.24	0.06	0.04
	12	$\hat{\alpha}^*$	0.78	0.55	0.21	0.07	0.02	0.01
		t^*_α	0.78	0.56	0.21	0.07	0.02	0.01
-0.8	1	$\hat{\alpha}^*$	1.00	1.00	1.00	1.00	0.87	0.68
		t^*_α	1.00	1.00	1.00	1.00	0.87	0.68
	4	$\hat{\alpha}^*$	1.00	1.00	0.99	0.83	0.36	0.23
		t^*_α	1.00	1.00	0.99	0.84	0.36	0.23
	12	$\hat{\alpha}^*$	0.99	0.95	0.69	0.34	0.10	0.07
		t^*_α	0.99	0.96	0.70	0.35	0.10	0.07

Table 2: Stationary process $\{U_t\}$ in (1); $MA(1), T = 100$.

ϕ	θ	k	Test	α					
				0.7	0.8	0.9	0.95	0.99	1.0
0.4	0.4	1	$\hat{\alpha}^*$	1.00	0.98	0.79	0.43	0.14	0.09
			t^*_α	1.00	0.97	0.76	0.42	0.13	0.09
		4	$\hat{\alpha}^*$	0.92	0.77	0.40	0.17	0.04	0.03
			t^*_α	0.91	0.75	0.40	0.18	0.05	0.03
		12	$\hat{\alpha}^*$	0.44	0.30	0.12	0.05	0.01	0.01
			t^*_α	0.45	0.31	0.14	0.06	0.02	0.01
-0.4	-0.4	1	$\hat{\alpha}^*$	1.00	1.00	0.99	0.77	0.29	0.20
			t^*_α	1.00	1.00	0.99	0.78	0.29	0.20
		4	$\hat{\alpha}^*$	1.00	0.98	0.66	0.28	0.07	0.05
			t^*_α	1.00	0.99	0.67	0.28	0.08	0.05
		12	$\hat{\alpha}^*$	0.90	0.73	0.32	0.11	0.03	0.02
			t^*_α	0.90	0.73	0.32	0.12	0.03	0.02
0.4	0.0	1	$\hat{\alpha}^*$	1.00	0.97	0.63	0.27	0.07	0.04
			t^*_α	1.00	0.96	0.62	0.27	0.07	0.04
		4	$\hat{\alpha}^*$	0.95	0.81	0.43	0.18	0.05	0.03
			t^*_α	0.94	0.81	0.43	0.19	0.05	0.03
		12	$\hat{\alpha}^*$	0.47	0.31	0.12	0.04	0.01	0.01
			t^*_α	0.47	0.32	0.14	0.05	0.01	0.01

Table 3: Stationary process $\{U_t\}$ in (1); $ARMA(1,1), T = 100$.

T	p	Test	α					
			0.7	0.8	0.9	0.95	0.99	1.0
50		$\hat{\alpha}$	0.98	0.78	0.33	0.15	0.07	0.05
		t_α	0.97	0.78	0.34	0.15	0.07	0.05
	0.2	$\hat{\alpha}^*$	0.77	0.48	0.16	0.07	0.03	0.03
		t^*_α	0.77	0.49	0.17	0.07	0.03	0.03
	0.1	$\hat{\alpha}^*$	0.54	0.30	0.10	0.05	0.02	0.02
		t^*_α	0.54	0.30	0.11	0.05	0.02	0.02
	0.05	$\hat{\alpha}^*$	0.38	0.24	0.10	0.05	0.03	0.02
		t^*_α	0.37	0.23	0.10	0.05	0.02	0.02
100		$\hat{\alpha}$	1.00	1.00	0.77	0.31	0.07	0.05
		$\hat{t}_\alpha$	1.00	1.00	0.77	0.32	0.08	0.05
	0.2	$\hat{\alpha}^*$	1.00	0.97	0.56	0.21	0.05	0.03
		t^*_α	1.00	0.97	0.58	0.22	0.05	0.03
	0.1	$\hat{\alpha}^*$	0.99	0.87	0.40	0.14	0.03	0.02
		t^*_α	0.99	0.87	0.41	0.14	0.03	0.02
	0.05	$\hat{\alpha}^*$	0.86	0.63	0.25	0.08	0.02	0.01
		$t^*_{\hat{\alpha}}$	0.87	0.64	0.26	0.09	0.02	0.01

Table 4: Stationary process $\{U_t\}$ in (1); independent standard normal variables.

θ	p	Test	α					
			0.7	0.8	0.9	0.95	0.99	1.0
0.8	0.2	$\hat{\alpha}^*$	0.99	0.91	0.45	0.15	0.03	0.02
		t^*_α	0.99	0.90	0.45	0.15	0.04	0.02
	0.1	$\hat{\alpha}^*$	0.95	0.78	0.34	0.11	0.03	0.02
		t^*_α	0.94	0.77	0.35	0.12	0.03	0.02
	0.05	$\hat{\alpha}^*$	0.78	0.55	0.21	0.07	0.02	0.01
		t^*_α	0.77	0.55	0.22	0.08	0.02	0.01
-0.4	0.2	$\hat{\alpha}^*$	1.00	1.00	0.88	0.47	0.14	0.09
		t^*_α	1.00	1.00	0.89	0.48	0.14	0.09
	0.1	$\hat{\alpha}^*$	1.00	0.99	0.68	0.28	0.07	0.05
		t^*_α	1.00	0.99	0.69	0.29	0.07	0.05
	0.05	$\hat{\alpha}^*$	0.98	0.87	0.42	0.16	0.04	0.03
		t^*_α	0.99	0.87	0.43	0.16	0.04	0.03
-0.8	0.2	$\hat{\alpha}^*$	1.00	1.00	1.00	1.00	0.73	0.54
		t^*_α	1.00	1.00	1.00	1.00	0.74	0.54
	0.1	$\hat{\alpha}^*$	1.00	1.00	1.00	0.95	0.50	0.33
		t^*_α	1.00	1.00	1.00	0.95	0.50	0.33
	0.05	$\hat{\alpha}^*$	1.00	1.00	0.98	0.74	0.28	0.18
		t^*_α	1.00	1.00	0.98	0.75	0.28	0.19

Table 5: Stationary process $\{U_t\}$ in (1); $MA(1), T = 100$.

ϕ	θ	p	Test	α					
				0.7	0.8	0.9	0.95	0.99	1.0
0.4	0.4	0.2	$\hat{\alpha}^*$	0.96	0.79	0.32	0.10	0.02	0.01
			t^*_α	0.94	0.77	0.33	0.11	0.02	0.02
		0.1	$\hat{\alpha}^*$	0.88	0.66	0.26	0.09	0.02	0.01
			t^*_α	0.86	0.65	0.28	0.10	0.02	0.01
		0.05	$\hat{\alpha}^*$	0.67	0.46	0.17	0.06	0.01	0.01
			t^*_α	0.66	0.45	0.19	0.07	0.02	0.01
-0.4	-0.4	0.2	$\hat{\alpha}^*$	1.00	1.00	0.99	0.78	0.30	0.20
			t^*_α	1.00	1.00	0.99	0.79	0.30	0.20
		0.1	$\hat{\alpha}^*$	1.00	1.00	0.90	0.52	0.16	0.10
			t^*_α	1.00	1.00	0.91	0.52	0.16	0.10
		0.05	$\hat{\alpha}^*$	1.00	0.98	0.66	0.30	0.09	0.05
			t^*_α	1.00	0.98	0.67	0.30	0.08	0.05
0.4	0.0	0.2	$\hat{\alpha}^*$	0.98	0.84	0.36	0.11	0.03	0.02
			t^*_α	0.97	0.83	0.37	0.12	0.03	0.02
		0.1	$\hat{\alpha}^*$	0.91	0.70	0.28	0.09	0.02	0.01
			t^*_α	0.90	0.70	0.29	0.10	0.02	0.01
		0.05	$\hat{\alpha}^*$	0.71	0.48	0.18	0.06	0.01	0.01
			t^*_α	0.71	0.48	0.19	0.07	0.02	0.01

Table 6: Stationary process $\{U_t\}$ in (1); $ARMA(1,1), T = 100$.