Econometric Theory, **18**, 2002, 469–490. Printed in the United States of America.
DOI: 10.1017.S0266466602182090

AN INVARIANCE PRINCIPLE FOR SIEVE BOOTSTRAP IN TIME SERIES

JOON Y. PARK
Seoul National University

This paper establishes an invariance principle applicable for the asymptotic analysis of sieve bootstrap in time series. The sieve bootstrap is based on the approximation of a linear process by a finite autoregressive process of order increasing with the sample size, and resampling from the approximated autoregression. In this context, we prove an invariance principle for the bootstrap samples obtained from the approximated autoregressive process. It is of the strong form and holds almost surely for all sample realizations. Our development relies upon the strong approximation and the Beveridge–Nelson representation of linear processes. For illustrative purposes, we apply our results and show the asymptotic validity of the sieve bootstrap for Dickey–Fuller unit root tests for the model driven by a general linear process with independent and identically distributed innovations. We thus provide a theoretical justification on the use of the bootstrap Dickey–Fuller tests for general unit root models, in place of the testing procedures by Said and Dickey and by Phillips.

1. INTRODUCTION

The bootstrap has become increasingly popular in econometrics and also in statistics and many other applied fields. This is in part due to the decreased computational cost but appears to be mostly because of its practical success and wide applicability. For a nice nontechnical introduction with an extensive survey of the bootstrap methodology, see Horowitz (1999). The method of bootstrap used to be applied mainly for the analysis of cross-sectional data, but recently it has also become popular in time series applications.

In this paper, we establish an invariance principle for bootstrap samples from linear processes with independent and identically distributed (i.i.d.) innovations. The underlying time series is approximated by a finite autoregressive process of order increasing with the sample size, and the bootstrap samples are drawn from the centered fitted residuals. Such a bootstrap is called sieve bootstrap by Bühlmann (1997), because it is based on an approximation of an infinite-

I thank a co-editor and two anonymous referees for useful comments and the Department of Economics at Rice University, where I am an adjunct professor, for its continuing hospitality and secretarial support. I am also very grateful to Don Andrews and Yoosoon Chang for helpful discussions and constructive suggestions, which have greatly improved the exposition of an earlier version of this paper. This research was supported by the Statistical Research Center for Complex Systems at Seoul National University. Address correspondence to: Joon Y. Park, School of Economics, Seoul National University, Seoul 151-742, Korea.

© 2002 Cambridge University Press 0266-4666/02 $9.50

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

dimensional and nonparametric model by a sequence of finite-dimensional parametric models. The sieve bootstrap has been studied earlier by Kreiss (1992), Bühlmann (1997, 1998), and Bickel and Bühlmann (1999). Along with the block bootstrap by Künsch (1989), it provides a standard tool for the bootstrap from dependent samples. Our work is most closely related in its aim to Bühlmann (1997) and Bickel and Bühlmann (1999). The former proves the bootstrap consistency for a class of nonlinear estimators, whereas the latter derives a bootstrap functional central limit theorem under a bracketing condition. They are of the weak form that holds only in probability.

The invariance principle we establish in the paper concerns the weak convergence of the bootstrap partial sum process to Brownian motion. None of the work cited previously deals with this type of invariance principle. It is of the strong form and holds almost surely for all sample realizations. With the continuous mapping theorem, it can be used to obtain asymptotic distributions of various bootstrapped statistics without making parametric assumptions on the underlying model. Our approach is built upon the strong approximation and the Beveridge–Nelson representation of linear processes. The invariance principles for the i.i.d. innovation and its bootstrap version are first developed using the strong approximation of the partial sum process by the standard Brownian motion, and subsequently the invariance principles for the general linear process with i.i.d. innovations and the corresponding bootstrapped process are established by their Beveridge–Nelson representations as in Phillips and Solo (1992).

For the purpose of illustration, we apply our results to analyze bootstrap asymptotics for Dickey–Fuller unit root tests. In particular, we prove the asymptotic validity of bootstrap Dickey–Fuller tests for models generated by linear processes with i.i.d. innovations. The Dickey–Fuller test statistics have the same limiting distributions as the corresponding bootstrap test statistics obtained by the sieve bootstrap procedure. One can therefore use Dickey–Fuller tests even when the underlying models are driven by innovations that are serially correlated, if the tests are based on the bootstrapped critical values. Thus bootstrap Dickey–Fuller tests can be an alternative to the tests by Said and Dickey and by Phillips, which were proposed to test for a unit root in general unit root models.

The rest of the paper is organized as follows. Section 2 presents some preliminary results on invariance principles for i.i.d. innovations and their bootstrap samples. In Section 3 we use the results from Section 2 to establish invariance principles for linear processes with i.i.d. innovations and for the samples obtained from the sieve bootstrap. Section 4 provides an application of our results to the test of a unit root. In particular, we develop the bootstrap asymptotic theory for Dickey–Fuller tests, and we show that they are asymptotically valid for unit root models driven by linear processes with i.i.d. innovations. Section 5 contains concluding remarks, and the mathematical proofs are given in Section 6. Finally, a word on notation. The standard notation used in probability and measure theory is used without reference, i.e., $\rightarrow_{a.s.}$, $\rightarrow_p$, $\rightarrow_{L^r}$, and

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

$\rightarrow_d$ imply, respectively, convergence almost surely, in probability, in L^r, and in distribution. A time series is denoted by (z_t), and whenever necessary the range of time index will be specified as in $(z_t)_{t=1}^n$ and $(z_t)_{t=1}^\infty$.

2. PRELIMINARY RESULTS

Throughout this section, we let (ε_t) be a sequence of i.i.d. random variables with mean zero. It is thought to be a sequence driving a model under investigation. Also, in the context of the bootstrap applications, it is supposed that the realization of (ε_t) itself or its estimates provides the values from which we resample. In this section, we first establish an invariance principle for (ε_t) and its bootstrap version. Primarily, this is a preliminary step toward an invariance principle for the sieve bootstrap in time series developed in the next section. Our results presented here, however, have a wider applicability and hence should be of independent interest. Though the existence of higher moments for (ε_t) will be required later for the main result of the paper, we simply let $\mathbf{E}\varepsilon_t^2 = \sigma^2 < \infty$ at the moment.

Let a partial sum process of (ε_t) be defined by

$$W_n(t) = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{[nt]} \varepsilon_k. \tag{1}$$

Here and elsewhere in the paper $[x]$ denotes the maximum integer that does not exceed x. Then by the classical Donsker's theorem (see, e.g., Billingsley, 1968) we have

$$W_n \rightarrow_d W \tag{2}$$

in the space $D[0,1]$ of cadlag functions, where W is the standard Brownian motion. The space $D[0,1]$ is usually equipped with the Skorohod topology. However, it is more convenient to consider the space endowed with the uniform norm $\|\cdot\|$ in the subsequent development of our theory.

We may actually have a stronger result than the weak convergence in (2). As a result of what is known as the Skorohod representation theorem (see, e.g., Pollard, 1984), we know that there exists a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ supporting a process W_n', say, and W such that W_n' has the same distribution as W_n and

$$W_n' \rightarrow_{a.s.} W. \tag{3}$$

Indeed, it follows from the result in Sakhanenko (1980) that we may choose W_n' satisfying the following condition.

LEMMA 2.1. *Let* $\mathbf{E}|\varepsilon_t|^r < \infty$ *for some* $r > 2$. *Then we have for any* $\delta > 0$

$$\mathbf{P}\{\|W_n' - W\| \geq \delta\} \leq n^{1-r/2} K_r \mathbf{E}|\varepsilon_t|^r,$$

where K_r *is an absolute constant depending only upon* r.

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

The result in Lemma 2.1, which is often called the strong approximation, is very useful in developing an invariance principle for bootstrap samples. This will be explained subsequently. In what follows, we will not distinguish W_n' from W_n, and we assume that W_n converges a.s. to W uniformly. This causes no loss in generality, because we are only concerned with distributional results.

For the bootstrap, we first obtain or estimate $(\varepsilon_t)_{t=1}^n$ from the sample of size n and get $(\hat{\varepsilon}_t)_{t=1}^n$, say. Then we resample from the empirical distribution of $(\hat{\varepsilon}_t)_{t=1}^n$, i.e., the distribution with point probability mass $1/n$ on their n observed values, to get the bootstrap sample $(\varepsilon_t^*)_{t=1}^n$. We may thus regard $(\varepsilon_t^*)_{t=1}^n$ as the i.i.d. samples from the empirical distribution of $(\hat{\varepsilon}_t)_{t=1}^n$. Both $(\hat{\varepsilon}_t)$ and (ε_t^*) are dependent upon the sample size n, and we may more precisely denote them as triangular arrays $(\hat{\varepsilon}_{nt})$ and (ε_{nt}^*). However, we will continue to write them as $(\hat{\varepsilon}_t)$ and (ε_t^*) in the subsequent discussions, following the usual convention.

Now we suppose that a realization of $(\hat{\varepsilon}_t)_{t=1}^\infty$ is given and consider the bootstrap probability space $(\Omega^*, \mathcal{F}^*, \mathbf{P}^*)$ conditional on the realization of $(\hat{\varepsilon}_t)_{t=1}^\infty$. For each realization of $(\hat{\varepsilon}_t)_{t=1}^\infty$, bootstrap samples $(\varepsilon_t^*)_{t=1}^n$ are regarded formally as random variables defined on this probability space. As mentioned earlier, the bootstrap samples $(\varepsilon_t^*)_{t=1}^n$ are obtained for each n from $(\hat{\varepsilon}_t)_{t=1}^n$. Under the probability $\mathbf{P}^*$, the bootstrap samples $(\varepsilon_t^*)_{t=1}^n$ become i.i.d. with the underlying distribution given by the empirical distribution of $(\hat{\varepsilon}_t)_{t=1}^n$. Naturally, the expectation with respect to $\mathbf{P}^*$ is signified by $\mathbf{E}^*$. We also use the conventional notations $\to_{p^*}$ and $\to_{d^*}$ to denote the convergences in probability and in distribution, respectively, for the functionals of bootstrap samples defined on $(\Omega^*, \mathcal{F}^*, \mathbf{P}^*)$. In particular, if the convergences occur a.s. for all realizations of $(\hat{\varepsilon}_t)_{t=1}^\infty$, then we write them, respectively, as $\to_{p^*}$ a.s. and $\to_{d^*}$ a.s. Of course, the former implies the latter. Whenever the sample mean $\bar{\varepsilon}_n$ of $(\hat{\varepsilon}_t)_{t=1}^n$ is nonzero, we assume that bootstrap samples are drawn from $(\hat{\varepsilon}_t - \bar{\varepsilon}_n)_{t=1}^n$ so that $\mathbf{E}^*\varepsilon_t^* = 0$ a.s.

For each realization of $(\hat{\varepsilon}_t)_{t=1}^\infty$, we consider

$$W_n^*(t) = \frac{1}{\hat{\sigma}_n\sqrt{n}} \sum_{k=1}^{[nt]} \varepsilon_k^*, \tag{4}$$

which corresponds to W_n defined in (1), where $\hat{\sigma}_n^2 = \mathbf{E}^*\varepsilon_t^{*2} = (1/n)\sum_{t=1}^n \hat{\varepsilon}_t^2$. As before, we let $W_n^{*\prime}$ be the distributionally equivalent copy of W_n^* in an expanded probability space rich enough to support a standard Brownian motion W^* such that

$$\mathbf{P}^*\{\|W_n^{*\prime} - W^*\| \geq \delta\} \leq n^{1-r/2} K_r \mathbf{E}^* |\varepsilon_t^*|^r \tag{5}$$

for any $\delta > 0$ and $r > 2$, where K_r is the absolute constant introduced in Lemma 2.1. This is possible, again due to the result by Sakhanenko (1980) as in Lemma 2.1. Following our earlier convention, we will not distinguish W_n^* in (4) from $W_n^{*\prime}$, and we simply assume that W_n^* and W^* are defined on the common probability space.

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

It is now obvious from the result in (5) that we can state the following theorem.

THEOREM 2.2. *If* $\mathbf{E}^*|\varepsilon_t^*|^r < \infty$ *a.s. and*

$$n^{1-r/2}\mathbf{E}^*|\varepsilon_t^*|^r \to_{a.s.} 0 \tag{6}$$

for some $r > 2$, *then*

$$W_n^* \to_{d^*} W \quad a.s. \tag{7}$$

as $n \to \infty$.

Thus we only need to show (6) to establish the invariance principle (7) for the bootstrap sample (ε_t^*). Note that (6) implies $W_n^* \to_{p^*} W^*$ a.s. as a result of (5) and therefore $W_n^* \to_{d^*} W^*$ a.s. Clearly, the latter convergence can be represented simply by $W_n^* \to_{d^*} W$ a.s. as in (7), because the limit process W^* is distributionally independent of the realizations of $(\hat{\varepsilon}_t)_{t=1}^{\infty}$.

In the simple case that (ε_t) are observable and the bootstrap samples are taken directly from the empirical distribution of $(\varepsilon_t - \bar{\varepsilon}_n)_{t=1}^n$, where $\bar{\varepsilon}_n$ is the sample mean of $(\varepsilon_t)_{t=1}^n$, we have

$$\mathbf{E}^*|\varepsilon_t^*|^r = \frac{1}{n}\sum_{t=1}^{n}|\varepsilon_t - \bar{\varepsilon}_n|^r.$$

Therefore, it suffices to require $\mathbf{E}|\varepsilon_t|^r < \infty$ for some $r > 2$ to derive the bootstrap invariance principle (7). Note that condition (6) is satisfied under such a moment condition, because $(1/n)\sum_{t=1}^n|\varepsilon_t - \bar{\varepsilon}_n|^r \to_{a.s.} \mathbf{E}|\varepsilon_t|^r$ by the strong law of large numbers.

For the simple case considered earlier, Basawa, Mallik, McCormick, Reeves, and Taylor (1991b) establish the bootstrap invariance principle (7) under a weaker assumption $\mathbf{E}\varepsilon_t^2 < \infty$.[1] They derive (7) directly by showing that the finite-dimensional distribution converges [in the Mallows metric (for details, see Bickel and Freedman, 1981)] and that the tightness condition holds. However, their approach does not readily extend to more complicated procedures such as the sieve bootstrap, which we will consider in the next section. We also refer to Ferretti and Romo (1996) for an invariance principle that is comparable to ours.

Let φ be a function on $D[0,1]$ continuous a.s. with respect to the Wiener measure. Once the invariance principle (2) is established, it follows immediately that

$$\varphi(W_n) \to_d \varphi(W) \tag{8}$$

as a result of the continuous mapping theorem. It is obvious from our development that a similar result holds for W_n^*; i.e., we may easily deduce from the bootstrap invariance principle that

$$\varphi(W_n^*) \to_{d^*} \varphi(W) \quad \text{a.s.} \tag{9}$$

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

In particular, if we apply (8) or (9) with $(\varphi(x))(t) = x(t) - tx(1)$ for $x \in D[0,1]$, then the limit process becomes the Brownian bridge $U(t) = W(t) - tW(1)$. Such limit process appears when we deal with the partial sum of $(\varepsilon_t - \bar{\varepsilon}_n)$ or $(\varepsilon_t^* - \bar{\varepsilon}_n^*)$, where $\bar{\varepsilon}_n$ and $\bar{\varepsilon}_n^*$ are, respectively, the sample means of (ε_t) and (ε_t^*). Also, we need to consider the detrended Brownian motion $W_\tau(t) = W(t) - (\int_0^1 W(s)\tau(s)'ds)(\int_0^1 \tau(s)\tau(s)'ds)^{-1}\tau(t)$ with $\tau \in D[0,1]^m$ for the asymptotic analysis of unit root models with deterministic trends. An invariance principle with limit process W_τ can be obtained from the application of (8) or (9) with $(\varphi(x))(t) = x(t) - (\int_0^1 x(s)\tau(s)'ds)(\int_0^1 \tau(s)\tau(s)'ds)^{-1}\tau(t)$ for $x \in D[0,1]$. As is well known, (8) or (9) yields the usual central limit theorem, because for $x \in D[0,1]$ $(\varphi(x))(t) = x(1)$ is a continuous functional.

3. INVARIANCE PRINCIPLE FOR SIEVE BOOTSTRAP

We consider a general linear process (u_t) given by

$$u_t = \pi(L)\varepsilon_t,$$

where (ε_t) is an i.i.d. random sequence and

$$\pi(z) = \sum_{k=0}^{\infty} \pi_k z^k.$$

More specifically, we let (ε_t) and $\pi(z)$ satisfy the following conditions.

Assumption 3.1. We assume that

(a) (ε_t) are i.i.d. random variables such that $\mathbf{E}\varepsilon_t = 0, \mathbf{E}\varepsilon_t^2 = \sigma^2$, and $\mathbf{E}|\varepsilon_t|^r < \infty$ for some $r > 4$,
(b) $\pi(z) \neq 0$ for all $|z| \leq 1$, and $\sum_{k=0}^{\infty} |k|^s |\pi_k| < \infty$ for some $s \geq 1$.

The conditions in Assumption 3.1 are not stringent and are routinely assumed in stationary time series analysis. Yet, they are sufficient to establish an invariance principle for the sieve bootstrap from the general linear process (u_t). The invariance principle that we will develop is of the strong form, which holds almost surely for all sample realizations. For a weak invariance principle that we only require to hold in probability, we may allow $r = 4$.[2]

Following Phillips and Solo (1992), we use the Beveridge–Nelson representation and write (u_t) as

$$u_t = \pi(1)\varepsilon_t + (\tilde{u}_{t-1} - \tilde{u}_t), \tag{10}$$

where

$$\tilde{u}_t = \sum_{k=0}^{\infty} \tilde{\pi}_k \varepsilon_{t-k}$$

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

with $\tilde{\pi}_k = \sum_{i=k+1}^{\infty} \pi_i$. Under the condition in Assumption 3.1(b), we have $\sum_{k=0}^{\infty} |\tilde{\pi}_k| < \infty$ as shown by Phillips and Solo (1992, p. 973). The time series $(\tilde{u}_t)$ is therefore well defined in both the a.s. and L^1 sense under Assumption 3.1(a).

If we let

$$V_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} u_k = (\sigma\pi(1))W_n(t) + \frac{1}{\sqrt{n}} (\tilde{u}_0 - \tilde{u}_{[nt]}), \tag{11}$$

then the invariance principle for (u_t),

$$V_n \to_d V = (\sigma\pi(1))W,$$

follows immediately from the continuous mapping theorem, because we have under Assumptions 3.1 that

$$\max_{1 \le k \le n} |n^{-1/2} \tilde{u}_k| \to_p 0$$

as shown in Phillips and Solo (1992, p. 978).

Under Assumption 3.1(b), we may write (u_t) as

$$\alpha(L) u_t = \varepsilon_t,$$

where

$$\alpha(z) = 1 - \sum_{k=1}^{\infty} \alpha_k z^k$$

with (α_k) satisfying $\sum |k|^s |\alpha_k| < \infty$, as shown in, e.g., Brillinger (1975). Therefore, it may be reasonable to approximate (u_t) by an autoregressive process of finite order, i.e.,

$$u_t = \alpha_1 u_{t-1} + \cdots + \alpha_p u_{t-p} + \varepsilon_{p,t}. \tag{12}$$

In the subsequent development of our theory, we let $p = p_n$ be a function of sample size n. More precisely, we can state the following assumption.

Assumption 3.2. p_n satisfies $p_n \to \infty$ and $p_n = o((n/\log(n))^{1/2})$.

We do not impose any lower limit to the divergence rate for p_n, and we may thus let it increase as slowly as we want. If we also assume $p_n/(\log n)^{1/4+\delta} \to \infty$ for some $\delta > 0$, all of our subsequent results hold also for $r = 4$ in Assumption 3.1(a). When there is some p_0 such that $\alpha_k = 0$ for all $k > p_0$ and (u_t) is generated by a finite autoregression, the conditions in Assumption 3.2 are of course not necessary. We only need to require $p_n \ge p_0$ for all large n. For the sake of notational brevity, we will denote by p as before instead of p_n the order of the approximated autoregression.

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

The parameters (α_k) and σ^2 can be consistently estimated from the regression (12). To show this, we let $(\hat{\alpha}_{p,k})$ be the ordinary least squares (OLS) estimators of (α_k) for $k = 1,\ldots,p$ and define

$$\hat{\alpha}_n(1) = 1 - \sum_{k=1}^{p} \hat{\alpha}_{p,k}.$$

Also, let $\hat{\sigma}_n^2$ be the usual error variance estimator.

LEMMA 3.1. *Let Assumptions 3.1 and 3.2 hold. Then it follows that*

$$\max_{1\le k\le p} |\hat{\alpha}_{p,k} - \alpha_k| = O((\log n/n)^{1/2}) + o(p^{-s}) \quad a.s.$$

for all large n. Moreover, we have

$$\hat{\sigma}_n^2 = \sigma^2 + O((\log n/n)^{1/2}) + o(p^{-s}) \quad a.s.$$

$$\hat{\alpha}_n(1) = \alpha(1) + O(p(\log n/n)^{1/2}) + o(p^{-s}) \quad a.s.$$

as $n \to \infty$.

The OLS estimators $(\hat{\alpha}_{p,k})$ are therefore consistent for (α_k), $k = 1,\ldots,p$. Moreover, as long as we let $p \to \infty$ as the sample size increases, the autoregressive coefficients (α_k) for $k > p$ become negligible in the limit, because $\sum_{k=p+1}^{\infty} \alpha_k = o(p^{-s})$ under our condition in Assumption 3.1(b).

For the sieve bootstrap, we first fit (u_t) with an approximated autoregression (12) of order p by OLS, i.e.,

$$u_t = \hat{\alpha}_{p,1} u_{t-1} + \cdots + \hat{\alpha}_{p,p} u_{t-p} + \hat{\varepsilon}_{p,t}$$

to get the fitted residuals $(\hat{\varepsilon}_{p,t})$. At this stage, we may use the usual order selection criterion such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). Under AIC and BIC, respectively, p is selected such that $\log \hat{\sigma}_n^2 + 2p/n$ and $\log \hat{\sigma}_n^2 + p \log n/n$ are minimized. If (u_t) is believed to be generated by a finite autoregression, BIC might be preferred because it yields a consistent estimator for p. See, e.g., An, Chen, and Hannan (1982). If not, AIC may be a better choice, because it leads to an asymptotically efficient choice for the optimal order of some projected infinite-order autoregressive process, as shown by Shibata (1980).

Our subsequent theory allows for such data-dependent selection rules. From Lemma 3.1, we have

$$\log \hat{\sigma}_n^2 = \log \sigma^2 + O((\log n/n)^{1/2}) + o(p^{-s}) \quad \text{a.s.}$$

Therefore, if we use AIC to select p, then p satisfies $p = o(n^{1/(1+s)})$ a.s., and the condition in Assumption 3.2 holds a.s. if $s > 1$. Likewise, BIC selects p such that $p = o((n/\log n)^{1/(1+s)})$ a.s., and the condition in Assumption 3.2 is

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

satisfied a.s. for $s \geq 1$. Note that $p \to \infty$ a.s. under both selection rules, unless (u_t) is generated by a finite autoregression.

The next step is to resample $(\varepsilon_t^*)_{t=1}^n$ from the empirical distribution of

$$\left(\hat{\varepsilon}_{p,t} - \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_{p,t}\right)_{t=1}^{n}$$

and generate recursively from (ε_t^*) an autoregressive process (u_t^*) of order p as

$$u_t^* = \hat{\alpha}_{p,1} u_{t-1}^* + \cdots + \hat{\alpha}_{p,p} u_{t-p}^* + \varepsilon_t^*.$$

For the actual generation of the bootstrap samples (u_t^*), their initial values must be set. The choice of the initial values of (u_t^*) may have an important effect on the actual performance of the bootstrap in finite samples. Also, to get stationary samples for (u_t^*), we must impose the appropriate stationarity conditions for the initial values, or we may repeat sampling until stationarity is achieved and throw the first generated values away. However, it is irrelevant in our subsequent asymptotic analysis and will not be discussed any further.

The bootstrapped process (u_t^*) may not be invertible in finite samples, though it should have small probability unless the sample size is very small. Given our results in Lemma 3.1, the problem of noninvertibility in (u_t^*) should vanish almost surely as the sample size increases. We may however prefer to use the Yule–Walker estimators (see, e.g., Brockwell and Davis, 1991, Secs. 8.1 and 8.2). It is well known that any finite autoregression fitted by the Yule–Walker estimators is invertible. All our subsequent results are also applicable for the bootstrap based on the Yule–Walker estimation of the autoregression (12).

Let

$$W_n^*(t) = \frac{1}{\hat{\sigma}_n \sqrt{n}} \sum_{k=1}^{[nt]} \varepsilon_k^*$$

be the bootstrap analogue of the process W_n introduced in the previous section. Then we have the following lemma.

LEMMA 3.2. *Let Assumptions 3.1 and 3.2 hold. Then condition (6) is satisfied and* $W_n^* \to_{d^*} W$ *a.s., as* $n \to \infty$.

The bootstrap invariance principle for (u_t^*) can be obtained from the Beveridge–Nelson representation, as in (10). If we let $\hat{\pi}_n(1) = 1/\hat{\alpha}_n(1)$ and

$$\tilde{u}_t^* = \frac{1}{\hat{\alpha}_n(1)} \sum_{k=1}^{p} \left(\sum_{i=k}^{p} \hat{\alpha}_{p,i}\right) u_{t-k+1}^*,$$

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

then we may write

$$V_n^*(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} u_k^* = (\hat{\sigma}_n \hat{\pi}_n(1)) W_n^*(t) + \frac{1}{\sqrt{n}} (\tilde{u}_0^* - \tilde{u}_{[nt]}^*). \tag{13}$$

Note that $\hat{\sigma}_n^2 \to_{a.s.} \sigma^2$ and $\hat{\pi}_n(1) \to_{a.s.} \pi(1)$ as $n \to \infty$, as shown in Lemma 3.1. Therefore, if we can show that

$$\mathbf{P}^* \left\{ \max_{1 \le t \le n} |n^{-1/2} \tilde{u}_t^*| > \delta \right\} \to_{a.s.} 0 \tag{14}$$

holds for any $\delta > 0$, the bootstrap invariance principle for (u_t^*)

$$V_n^* \to_{d^*} V = (\sigma \pi(1)) W \quad \text{a.s.}$$

would follow.

THEOREM 3.3. *Let Assumptions 3.1 and 3.2 hold. Then condition (14) holds and* $V_n^* \to_{d^*} V$ *a.s., as* $n \to \infty$.

The bootstrap invariance principle for the sieve bootstrap has now been established for general linear processes with i.i.d. innovations. It can be applied to obtain bootstrap asymptotics without making parametric assumptions on the underlying model. In particular, the limiting distributions of various bootstrap statistics from linear time series can be found simply by the application of the continuous mapping theorem.

4. SIEVE BOOTSTRAP FOR DICKEY–FULLER TESTS

We now show how the bootstrap invariance principle developed in the previous section can be applied to testing the unit root hypothesis. Let a time series (y_t) be generated as

$$y_t = D_t + x_t,$$

where (D_t) and (x_t) denote, respectively, the deterministic and stochastic components of (y_t). We specify (D_t) as

$$D_t = c_t' \beta, \tag{15}$$

where $c_t = (c_{1t}, \ldots, c_{mt})'$ is a vector of deterministic functions of time and $\beta = (\beta_1, \ldots, \beta_m)'$ is a parameter vector. Most commonly used such trends are $c_t = 1$ and $c_t = (1, t)'$. Furthermore, we let (x_t) be given by

$$x_t = \alpha x_{t-1} + u_t \tag{16}$$

with $\alpha = 1$ and let (u_t) be a linear process specified as in the previous section.[3] We may allow the initial value x_0 of (x_t) to be any random variable, and therefore we set $x_0 = 0$ in the subsequent development of our theory.

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

For expositional simplicity, we will mainly consider (y_t) with no deterministic trends, i.e., $(y_t) \equiv (x_t)$. It is easy to accommodate the presence of deterministic trends. For instance, our subsequent bootstrap methodology applies, without any modification, to (y_t) with deterministic trends if we detrend (y_t) by running an OLS regression

$$y_t = c_t' \hat{\beta}_n + \hat{x}_t$$

and use the fitted residuals $(\hat{x}_t)$ in place of (x_t).[4] Using $(\hat{x}_t)$ instead of (x_t) would, however, result in different asymptotic distributions for the test statistics. This will be explained in detail later.

To test the hypothesis $H_0: \alpha = 1$ in the model (16), we consider the statistics proposed and studied by Dickey and Fuller (1979, 1981). Their coefficient and t-statistics will be denoted by S_n and T_n, respectively. To present them more explicitly, we let $\hat{\alpha}_n$ be the OLS estimator of α, and we let $s(\hat{\alpha}_n)$ be the standard error for $\hat{\alpha}_n$ given by $\hat{\omega}_n(\sum_{t=1}^n x_{t-1}^2)^{-1/2}$, where $\hat{\omega}_n^2$ is the usual error variance estimator from regression (16). Then, we may write S_n and T_n as

$$S_n = n(\hat{\alpha}_n - 1) = \frac{(1/n)\left(x_n^2 - \sum_{t=1}^n u_t^2\right)}{(2/n^2)\sum_{t=1}^n x_{t-1}^2},$$

$$T_n = \frac{\hat{\alpha}_n - 1}{s(\hat{\alpha}_n)} = \frac{(1/n)\left(x_n^2 - \sum_{t=1}^n u_t^2\right)}{2\hat{\omega}_n\left((1/n^2)\sum_{t=1}^n x_{t-1}^2\right)^{1/2}}$$

under the unit root hypothesis.

We now have

$$S_n \approx \frac{V_n(1)^2 - \omega_n^2}{2\int_0^1 V_n(t)^2 dt} \to_d \frac{V(1)^2 - \omega^2}{2\int_0^1 V(t)^2 dt} \tag{17}$$

and similarly,

$$T_n \approx \frac{V_n(1)^2 - \omega_n^2}{2\hat{\omega}_n\left(\int_0^1 V_n(t)^2 dt\right)^{1/2}} \to_d \frac{V(1)^2 - \omega^2}{2\omega\left(\int_0^1 V(t)^2 dt\right)^{1/2}}, \tag{18}$$

where $\omega_n^2 = (1/n)\sum_{t=1}^n u_t^2$ and other notation is defined earlier. The convergence in distribution in (17) and (18) can be easily deduced from the laws of large numbers applied to ω_n^2 and $\hat{\omega}_n^2$ and the continuous mapping theorem ap-

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

plied to the weak convergence $V_n \to_d V$ because both $V(1)^2$ and $\int_0^1 V(t)^2 dt$ are continuous functionals of V in $D[0,1]$.

Let (u_t^*) be the bootstrap sample for (u_t), $u_t = x_t - x_{t-1}$, obtained as described in the previous section, and denote by (x_t^*) the bootstrap sample for (x_t) generated by

$$x_t^* = x_{t-1}^* + u_t^*$$

with initial value x_0, which is assumed to be zero to simplify the exposition. It is important to impose the unit root hypothesis when we generate the bootstrap samples (x_t^*). If we generate them by $x_t^* = \hat{\alpha}_n x_{t-1}^* + u_t^*$ using the estimated value $\hat{\alpha}_n$ of α in (16), then they would not behave as a unit root process. This is so, even though $\hat{\alpha}_n$ is superconsistent and converges to unity at a faster rate n. See Basawa et al. (1991a), Datta (1996), and Kreiss and Heimann (1996).

Let $\hat{\alpha}_n^*$ be the OLS coefficient estimator from the first-order autoregression of (x_t^*) on (x_{t-1}^*), and $s^*(\hat{\alpha}_n^*)$ be the bootstrap standard error for $\hat{\alpha}_n^*$, which is given by $\hat{\omega}_n(\sum_{t=1}^n x_{t-1}^{*2})^{1/2}$. Now we define

$$S_n^* = n(\hat{\alpha}_n^* - 1) = \frac{(1/n)\left(x_n^{*2} - \sum_{t=1}^n u_t^{*2}\right)}{(2/n^2)\sum_{t=1}^n x_{t-1}^{*2}},$$

$$T_n^* = \frac{\hat{\alpha}_n^* - 1}{s^*(\hat{\alpha}_n^*)} = \frac{(1/n)\left(x_n^{*2} - \sum_{t=1}^n u_t^{*2}\right)}{2\hat{\omega}_n\left((1/n^2)\sum_{t=1}^n x_{t-1}^{*2}\right)^{1/2}},$$

which are the bootstrap versions of the statistics S_n and T_n. Also, we define $\omega_n^{*2} = (1/n)\sum_{t=1}^n u_t^{*2}$.

LEMMA 4.1. *Let Assumptions 3.1 and 3.2 hold. Then we have, for any* $\delta > 0$, $\mathbf{P}^*\{|\omega_n^{*2} - \omega^2| \geq \delta\} \to_{a.s.} 0$ *as* $n \to \infty$.

THEOREM 4.2. *Let Assumptions 3.1 and 3.2 hold. Then we have*

$$S_n^* \to_{d^*} \frac{V(1)^2 - \omega^2}{2\int_0^1 V(t)^2 dt} \quad a.s.$$

$$T_n^* \to_{d^*} \frac{V(1)^2 - \omega^2}{2\omega\left(\int_0^1 V(t)^2 dt\right)^{1/2}} \quad a.s.$$

as $n \to \infty$.

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

Theorem 4.2 shows that the bootstrap statistics S_n^* and T_n^* have the same asymptotic distributions as the corresponding sample statistics S_n and T_n. Therefore, it establishes the asymptotic validity of the bootstrap Dickey–Fuller tests. It shows in particular that if based on the bootstrap critical values then the Dickey–Fuller tests can be used to test for the presence of a unit root in models driven by innovations that are correlated. The bootstrap Dickey–Fuller tests can therefore be an alternative to the testing procedures by Phillips (1987) and Said and Dickey (1984) (which are more often called augmented Dickey–Fuller tests).

The limiting distributions of the sample statistics S_n and T_n are not pivotal. They depend on the parameters σ^2, ω^2, and $\pi(1)$. Note that $V = (\sigma\pi(1))W$, where W is standard Brownian motion. It is thus not expected that the bootstrap statistics S_n^* and T_n^* provide asymptotic refinements. Therefore, the Dickey–Fuller tests based on the bootstrapped critical values are not necessarily better than the tests relying on the asymptotic critical values. However, for the model considered here, the Said–Dickey or Phillips statistics have limiting distributions that are free of nuisance parameters. Consequently, they may provide tests with more accurate finite sample sizes if based on the bootstrapped critical values.

The bootstrap consistency established previously assumes the absence of deterministic trends. However, it can be extended in a trivial manner to allow for the presence of deterministic trends. For the trend specification in (15), we just assume that there exists κ_i such that, if we define $\tau_{ni}(t) = n^{-\kappa_i}\sum_{k=1}^{n} c_{i,k-1}1\{(k-1)/n \le t < k/n\}$, then $\tau_{ni} \to_{L^2} \tau_i$ for some $\tau_i : [0,1] \to \mathbf{R}$ that is of bounded variation for $i = 1,\ldots,m$. Also, we let τ_i's be linearly independent in $L^2[0,1]$. Under these assumptions, we may show that the asymptotic distributions of S_n and T_n constructed from $(\hat{x}_t)$ are identical to those in (17) and (18), except that $V = (\sigma\pi(1))W$ is replaced by $V_\tau = (\sigma\pi(1))W_\tau$, where W_τ is the standard detrended Brownian motion introduced in Section 2 (see, e.g., Park, 1992). Moreover, it can also be shown without difficulty that the limiting distributions of their bootstrap counterparts S_n^* and T_n^* have the same limiting distributions.

5. CONCLUDING REMARKS

In this paper, we have developed an invariance principle for the sieve bootstrap from linear process with i.i.d. innovations. It can be used, in various contexts, to obtain the limiting distributions of bootstrap statistics without making parametric assumptions on the underlying model. As an illustration, we used the invariance principle to derive the limiting distributions of the Dickey–Fuller unit root tests and show their asymptotic validity for a model driven by a general linear process that is serially correlated. It appears that the bootstrap invariance principle established in the paper has wide applicability and can be used to analyze the asymptotic behavior of bootstrap samples in stationary and nonstationary time series models. Some of the applications are under way by the author and co-researchers.

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

6. MATHEMATICAL PROOFS

The proofs of Lemma 2.1 and Theorem 2.2 are omitted. Lemma 2.1 is due to Sakhanenko (1980), and the stated result in Theorem 2.2 follows immediately from (5) as mentioned in the following remarks.

Proof of Lemma 3.1. We write (u_t) as

$$u_t = \alpha_{p,1} u_{t-1} + \cdots + \alpha_{p,p} u_{t-p} + e_{p,t}, \tag{19}$$

where the coefficients $(\alpha_{p,k})$ are defined so that $(e_{p,t})$ are uncorrelated with (u_{t-k}) for $k = 1, \ldots, p$. We have

$$\max_{1 \le k \le p} |\hat{\alpha}_{p,k} - \alpha_{p,k}| = O((\log n/n)^{1/2}) \quad \text{a.s.}, \tag{20}$$

$$\sum_{k=1}^{p} |\alpha_{p,k} - \alpha_k| \le c \sum_{k=p+1}^{\infty} |\alpha_k| = o(p^{-s}), \tag{21}$$

where c is some constant. For the results in (20) and (21), see, e.g., Hannan and Kavalieris (1986, Theorem 2.1) and Bühlmann (1995, proof of theorem 3.1). The proof of the first part is immediate, because

$$|\hat{\alpha}_{p,k} - \alpha_k| \le |\hat{\alpha}_{p,k} - \alpha_{p,k}| + |\alpha_{p,k} - \alpha_k|.$$

Similarly, it follows that

$$\begin{aligned} |\hat{\alpha}_n(1) - \alpha(1)| &\le \sum_{k=1}^{p} |\hat{\alpha}_{p,k} - \alpha_{p,k}| + \sum_{k=1}^{p} |\alpha_{p,k} - \alpha_k| + \sum_{k=p+1}^{\infty} |\alpha_k| \\ &= O(p(\log n/n)^{1/2}) + o(p^{-s}) \end{aligned}$$

as was to be shown for the result on $\hat{\alpha}_n(1)$. For the result of $\hat{\sigma}_n^2$, see Bühlmann (1995, proof of theorem 3.2). He establishes the result for the Yule–Walker estimator, but it is easy to see that his result readily extends to the least squares estimator. ■

Proof of Lemma 3.2. We will show

$$n^{1-r/2} \mathbf{E}^* |\varepsilon_t^*|^r = n^{1-r/2} \left(\frac{1}{n} \sum_{t=1}^{n} \left| \hat{\varepsilon}_{p,t} - \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_{p,t} \right|^r \right) \to_{a.s.} 0 \tag{22}$$

as $n \to \infty$. Note that

$$\frac{1}{n} \sum_{t=1}^{n} \left| \hat{\varepsilon}_{p,t} - \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_{p,t} \right|^r \le c(A_n + B_n + C_n + D_n),$$

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

where c is some constant, and

$$A_n = \frac{1}{n}\sum_{t=1}^{n} |\varepsilon_t|^r,$$

$$B_n = \frac{1}{n}\sum_{t=1}^{n} |\varepsilon_{p,t} - \varepsilon_t|^r,$$

$$C_n = \frac{1}{n}\sum_{t=1}^{n} |\hat{\varepsilon}_{p,t} - \varepsilon_{p,t}|^r,$$

$$D_n = \left|\frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_{p,t}\right|^r.$$

Therefore, to deduce (22), we only need to show that $n^{1-r/2}A_n, n^{1-r/2}B_n$, $n^{1-r/2}C_n$ and $n^{1-r/2}D_n \to_{a.s.} 0$. This is what we now set out to do.

By the strong law of large numbers, $A_n \to_{a.s.} \mathbf{E}|\varepsilon_t|^r$, and therefore, $n^{1-r/2}A_n \to_{a.s.} 0$. Next, we show that

$$\mathbf{E}|\varepsilon_{p,t} - \varepsilon_t|^r = o(p^{-rs}) \tag{23}$$

holds uniformly in t, from which it follows that $n^{1-r/2}B_n \to_{a.s.} 0$. Note that if $p/(\log n)^{1/4+\delta} \to \infty$ for some $\delta > 0$ then $n^{1-r/2}B_n \to_{a.s.} 0$ also for $r = 4$, as we remarked following Assumption 3.2. To obtain (23), we write

$$\varepsilon_{p,t} = \varepsilon_t + \sum_{k=p+1}^{\infty} \alpha_k u_{t-k} \tag{24}$$

and apply Minkowski's inequality and use the stationarity of (u_t) to get

$$\mathbf{E}|\varepsilon_{p,t} - \varepsilon_t|^r \le \mathbf{E}|u_t|^r \left(\sum_{k=p+1}^{\infty} |\alpha_k|\right)^r = o(p^{-rs}).$$

Note that, due to the Marcinkiewicz–Zygmund inequality (see, e.g., Stout, 1974, Theorem 3.3.6),

$$\mathbf{E}|u_t|^r \le c\left(\sum_{k=0}^{\infty} \pi_k^2\right)^{r/2} \mathbf{E}|\varepsilon_t|^r < \infty, \tag{25}$$

where c is some constant, and therefore $\sum_{p+1}^{\infty} \alpha_k u_{t-k}$ is well defined in the L^r sense.

We now prove that $n^{1-r/2}C_n \to_{a.s.} 0$. Write

$$\begin{aligned}\hat{\varepsilon}_{p,t} &= u_t - \sum_{k=1}^{p} \hat{\alpha}_{p,k} u_{t-k} \\ &= \varepsilon_{p,t} - \sum_{k=1}^{p} (\hat{\alpha}_{p,k} - \alpha_{p,k}) u_{t-k} - \sum_{k=1}^{p} (\alpha_{p,k} - \alpha_k) u_{t-k},\end{aligned} \tag{26}$$

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

where $(\alpha_{p,k})$ are defined in (19). It follows that

$$|\hat{\varepsilon}_{p,t} - \varepsilon_{p,t}|^r \le c\left(\left|\sum_{k=1}^{p}(\hat{\alpha}_{p,k} - \alpha_{p,k})u_{t-k}\right|^r + \left|\sum_{k=1}^{p}(\alpha_{p,k} - \alpha_k)u_{t-k}\right|^r\right)$$

for $c = 2^{r-1}$. Therefore, if we define

$$C_{1n} = \frac{1}{n}\sum_{t=1}^{n}\left|\sum_{k=1}^{p}(\hat{\alpha}_{p,k} - \alpha_{p,k})u_{t-k}\right|^r,$$

$$C_{2n} = \frac{1}{n}\sum_{t=1}^{n}\left|\sum_{k=1}^{p}(\alpha_{p,k} - \alpha_k)u_{t-k}\right|^r,$$

then it suffices to show that $n^{1-r/2}C_{1n}, n^{1-r/2}C_{2n} \to_{a.s.} 0$.

Note that C_{1n} is majorized by

$$\left(\max_{1\le k\le p}|\hat{\alpha}_{p,k} - \alpha_{p,k}|^r\right)\frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{p}|u_{t-k}|^r$$

$$\le \left(\max_{1\le k\le p}|\hat{\alpha}_{p,k} - \alpha_{p,k}|^r\right)\frac{p}{n}\left(\sum_{t=0}^{n-1}|u_t|^r + \sum_{t=-1}^{1-p}|u_t|^r\right)$$

$$= O((\log n/n)^r)(p/n)O(n) = O(p(\log n/n)^r) \quad \text{a.s.}$$

by (20) and (25) and therefore, $C_{1n} \to_{a.s.} 0$. Moreover, we have

$$\mathbf{E}\left|\sum_{k=1}^{p}(\alpha_{p,k} - \alpha_k)u_{t-k}\right|^r \le \mathbf{E}|u_t|^r\left(\sum_{k=1}^{p}|\alpha_{p,k} - \alpha_k|\right)^r = o(p^{-rs})$$

by Minkowski's inequality and the stationarity of (u_t). Consequently, it follows that $n^{1-r/2}C_{2n} \to_{a.s.} 0$ for $r > 4$. If $p/(\log n)^{1/4+\delta} \to \infty$, $r = 4$ works also.

Finally, to deduce $n^{1-r/2}D_n \to_{a.s.} 0$, we show

$$\frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_{p,t} = \frac{1}{n}\sum_{t=1}^{n}\varepsilon_{p,t} + o(1) \quad \text{a.s.} = \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t + o(1) \quad \text{a.s.},$$

which hold if

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{k=p+1}^{\infty}\alpha_k u_{t-k} \to_{a.s.} 0, \tag{27}$$

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{p}(\alpha_{p,k} - \alpha_k)u_{t-k} \to_{a.s.} 0, \tag{28}$$

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{p}(\hat{\alpha}_{p,k} - \alpha_{p,k})u_{t-k} \to_{a.s.} 0 \tag{29}$$

as a result of (24) and (26).

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

To establish (27)–(29), we first define

$$S_n(i,j) = \sum_{t=1}^{n} \varepsilon_{t-i-j} \quad \text{and} \quad T_n(i) = \sum_{t=1}^{n} u_{t-i}.$$

Then it follows that

$$T_n(i) = \sum_{t=1}^{n} \sum_{j=0}^{\infty} \pi_j \varepsilon_{t-i-j} = \sum_{j=0}^{\infty} \pi_j S_n(i,j).$$

In what follows, we use c to denote a constant, which is not always the same. By successive applications of Doob's inequality and Burkholder's inequality (see, e.g., Hall and Heyde, 1980, Theorems 2.2 and 2.10), we have

$$\mathbf{E}\Big(\max_{1\le m\le n} |S_m(i,j)|^r \Big) \le cn^{r/2}$$

uniformly in i and j. Moreover, because

$$\Big[\mathbf{E}\Big(\max_{1\le m\le n} |T_m(i)|^r \Big)\Big]^{1/r} \le \sum_{j=0}^{\infty} |\pi_j| \Big[\mathbf{E}\Big(\max_{1\le m\le n} |S_m(i,j)|^r \Big)\Big]^{1/r}$$

we have

$$\mathbf{E}\Big(\max_{1\le m\le n} |T_m(i)|^r \Big) \le cn^{r/2}$$

uniformly in i.

We now let

$$L_n = \sum_{k=p+1}^{\infty} \alpha_k \sum_{t=1}^{n} u_{t-k}.$$

Then it follows that

$$\Big[\mathbf{E}\Big(\max_{1\le m\le n} |L_m|^r \Big)\Big]^{1/r} \le \sum_{k=p+1}^{\infty} |\alpha_k| \Big[\mathbf{E}\Big(\max_{1\le m\le n} |T_m(k)|^r \Big)\Big]^{1/r} \le cp^{-s} n^{1/2}$$

and consequently,

$$\mathbf{E}\Big(\max_{1\le m\le n} |L_m|^r \Big) \le cp^{-rs} n^{r/2}.$$

Therefore, we may show for any $\delta > 0$

$$L_n = o(p^{-s} n^{1/2} (\log n)^{1/r} (\log\log n)^{(1+\delta)/r}) = o(n) \quad \text{a.s.}$$

exactly as in Móricz (1976, pp. 309–310, below (4.8)). This proves (27).

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

Similarly, if we let

$$M_n = \sum_{k=1}^{p} (\alpha_{p,k} - \alpha_k) \sum_{t=1}^{n} u_{t-k}$$

then we have

$$\left[\mathbf{E}\Big(\max_{1\le m\le n} |M_m|^r \Big)\right]^{1/r} \le \sum_{k=1}^{p} |\alpha_{p,k} - \alpha_k| \left[\mathbf{E}\Big(\max_{1\le m\le n} |T_m(k)|^r \Big)\right]^{1/r} \le cp^{-s} n^{1/2}.$$

Therefore, we have

$$M_n = o(p^{-s} n^{1/2} (\log n)^{1/r} (\log\log n)^{(1+\delta)/r}) = o(n) \quad \text{a.s.},$$

which establishes (28). Finally, we consider

$$N_n = \sum_{k=1}^{p} (\hat{\alpha}_{p,k} - \alpha_{p,k}) \sum_{t=1}^{n} u_{t-k}.$$

If we let

$$Q_n = \sum_{k=1}^{p} \left| \sum_{t=1}^{n} u_{t-k} \right|$$

then N_n is dominated by

$$Q_n \max_{1\le k\le p} |\hat{\alpha}_{p,k} - \alpha_{p,k}|.$$

Moreover, we may deduce precisely as before that

$$\mathbf{E}\Big(\max_{1\le m\le n} |Q_m|^r \Big) \le cp^r n^{r/2}$$

and therefore for any $\delta > 0$

$$Q_n = o(pn^{1/2} (\log n)^{1/r} (\log\log n)^{(1+\delta)/r}) \quad \text{a.s.}$$

Consequently,

$$N_n = O((\log n/n)^{1/2}) Q_n = o(p(\log n)^{(r+2)/2r} (\log\log n)^{(1+\delta)/r}) = o(n) \quad \text{a.s.}$$

as was to be shown for (29). Therefore, the proof is complete. ■

Proof of Theorem 3.3. To show (14), we note that

$$\begin{aligned} \mathbf{P}^*\Big\{ \max_{1\le t\le n} |n^{-1/2} \tilde{u}_t^*| > \delta \Big\} &\le \sum_{t=1}^{n} \mathbf{P}^*\{|n^{-1/2} \tilde{u}_t^*| > \delta\} \\ &= n\mathbf{P}^*\{|n^{-1/2} \tilde{u}_t^*| > \delta\} \\ &\le (1/\delta^r) n^{1-r/2} \mathbf{E}^* |\tilde{u}_t^*|^r. \end{aligned}$$

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

The first inequality is obvious. The second equality follows from the stationarity of $(\tilde{u}_t^*)$, conditional on the realization of $(\hat{\varepsilon}_{p,t})$, and the third inequality is due to the Tchebyshev inequality.

To complete the proof, it now suffices to prove that

$$n^{1-r/2}\mathbf{E}^*|\tilde{u}_t^*|^r \to_{a.s.} 0,$$

which follows immediately if we show

$$n^{1-r/2}\mathbf{E}^*|u_t^*|^r \to_{a.s.} 0 \tag{30}$$

because

$$\mathbf{E}^*|\tilde{u}_t^*|^r \le \left(\frac{1}{|\hat{\alpha}_n(1)|}\sum_{k=1}^{p} k|\hat{\alpha}_{p,k}|\right)^r \mathbf{E}^*|u_t^*|^r$$

by Minkowski's inequality and the conditional stationarity of (u_t^*). Note that $\hat{\alpha}_n(1) \to_{a.s.} \alpha(1)$ as shown in Lemma 3.1 and that

$$\sum_{k=1}^{p} k|\hat{\alpha}_{p,k}| = \sum_{k=1}^{p} k|\alpha_{p,k}| + o(1) = \sum_{k=1}^{\infty} k|\alpha_k| + o(1) \quad \text{a.s.}$$

as shown in the proof of Lemma 3.1.

We may write for large n [for all n if we use the Yule–Walker method to estimate regression (12) (see, e.g., Brockwell and Davis, 1991, p. 240)]

$$u_t^* = \sum_{k=0}^{\infty} \hat{\pi}_{p,k}\varepsilon_{t-k}^* \tag{31}$$

with $(\hat{\pi}_{p,k})$ such that $\sum_{k=0}^{\infty}|\hat{\pi}_{p,k}| < \infty$ a.s. Therefore, due to the Marcinkiewicz–Zygmund inequality,

$$\mathbf{E}^*|u_t^*| \le c\left(\sum_{k=0}^{\infty}\hat{\pi}_{p,k}^2\right)^{r/2}\mathbf{E}^*|\varepsilon_t^*|^r$$

for some constant c. We therefore have (30) whenever (22) holds, which was shown earlier in the proof of Lemma 3.1. ■

Proof of Lemma 4.1. Let

$$\omega_*^2 = \mathbf{E}^*|u_t^*|^2$$

and note that

$$|\omega_n^{*2} - \omega^2| \le |\omega_n^{*2} - \omega_*^2| + |\omega_*^2 - \omega^2|$$

We first show that

$$\omega_*^2 \to_{a.s.} \omega^2 \tag{32}$$

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

From the Yule–Walker equations, we have

$$\omega_*^2 = (\sigma_*^2/\hat{\sigma}_n^2)\omega_n^2$$

where $\sigma_*^2 = \mathbf{E}^*|\varepsilon_t^*|^2$ and other notations were defined earlier. Because $\omega_n^2 \to_{a.s.} \omega^2$ by the strong law of large numbers, it suffices to show that

$$\sigma_*^2/\hat{\sigma}_n^2 \to_{a.s.} 1$$

to deduce (32). This, however, is immediate because $\hat{\sigma}_n^2 \to_{a.s.} \sigma^2$ as shown in Lemma 3.1 and

$$\sigma_*^2 = \hat{\sigma}_n^2 + \left(\frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_{p,t}\right)^2$$

the second term of which is of order $o(1)$ a.s., as we have shown in the proof of Lemma 3.2.

We now show

$$\mathbf{P}^*\{|\omega_n^{*2} - \omega_*^2| > \delta\} \le (1/\delta)^r \mathbf{E}^*|\omega_n^{*2} - \omega_*^2|^r \to_{a.s.} 0 \tag{33}$$

To deduce (33), we write (u_t^*) as in (31). Then we have

$$\omega_n^{*2} - \omega_*^2 = \frac{1}{n}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \hat{\pi}_{p,i}\hat{\pi}_{p,j}\sum_{t=1}^{n}(\varepsilon_{t-i}^*\varepsilon_{t-j}^* - \delta_{ij}\sigma_*^2)$$

where δ_{ij} is the Kronecker delta, and it follows that

$$\mathbf{E}^*|\omega_n^{*2} - \omega_*^2|^{r/2} \le n^{-r/2}\left(\sum_{k=0}^{\infty}|\hat{\pi}_{p,k}|\right)^r \mathbf{E}^*\left|\sum_{t=1}^{n}(\varepsilon_{t-i}^*\varepsilon_{t-j}^* - \delta_{ij}\sigma_*^2)\right|^{r/2}$$

by Minkowski's inequality.

By the successive applications of Burkholder's inequality and Minkowski's inequality, we have

$$\begin{aligned}\mathbf{E}^*\left|\sum_{t=1}^{n}(\varepsilon_{t-i}^*\varepsilon_{t-j}^* - \delta_{ij}\sigma_*^2)\right|^{r/2} &\le K\mathbf{E}^*\left|\sum_{t=1}^{n}(\varepsilon_{t-i}^*\varepsilon_{t-j}^* - \delta_{ij}\sigma_*^2)^2\right|^{r/4} \\ &\le Kn^{r/4}\mathbf{E}^*|\varepsilon_{t-i}^*\varepsilon_{t-j}^* - \delta_{ij}\sigma_*^2|^{r/2},\end{aligned}$$

where K is an absolute constant depending only upon r. Moreover,

$$\begin{aligned}&\mathbf{E}^*|\varepsilon_{t-i}^*\varepsilon_{t-j}^* - \delta_{ij}\sigma_*^2|^{r/2} \\ &\quad\le (r-2)/2(\mathbf{E}^*|\varepsilon_{t-i}^*\varepsilon_{t-j}^*|^{r/2} + |\sigma_*^2|^{r/2}) \le (r-2)\mathbf{E}^*|\varepsilon_t^*|^r\end{aligned}$$

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

as a result of the Cauchy–Schwarz inequality and Jensen's inequality. It follows that

$$\mathbf{E}^*|\omega_n^{*2} - \omega_*^2|^{r/2} \leq n^{-r/4} K \left(\sum_{k=0}^{\infty} |\hat{\pi}_{p,k}| \right)^r \mathbf{E}^*|\varepsilon_t^*|^r,$$

where K is an absolute constant depending upon r, and (33) would thus follow if

$$n^{-r/4}\mathbf{E}^*|\varepsilon_t^*|^r \to_{a.s.} 0.$$

This can be shown exactly as in the proof of Lemma 3.2. The stated result now follows from (32) and (33). ■

Proof of Theorem 4.2. The stated results follow immediately from Theorem 3.3 and Lemma 4.1.

NOTES

1. They assume that (ε_t) has a continuous distribution function. This condition, however, is not used in their proof for the bootstrap invariance principle.

2. Our assumptions here are comparable to those made in related works. Kreiss (1992) considers the model with (π_k) that decays exponentially. Bühlmann (1997) proves the weak form of bootstrap consistency for a class of nonlinear estimators under the condition $r \geq 4$ and $s \geq 1$. Bickel and Bühlmann (1999) establish a weak invariance principle for the bootstrap empirical distribution function for the case $r \geq 4$ and exponentially decaying (π_k).

3. Notice that we do not allow for unit root models generated by general martingale difference innovations.

4. Elliott, Rothenberg and Stock (1996) show that the power of the unit root test can be substantially improved by the generalized least squares detrending. Though we will not present the details here, the bootstrap asymptotics of their tests can also be analyzed using our method here.

REFERENCES

An, H.-Z., Z.-G. Chen, & E.J. Hannan (1982) Autocorrelation, autoregression and autoregressive approximation. *Annals of Statistics* 10, 926–936. Correction: 11, 1018.

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

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves, & R.L. Taylor (1991b) Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive processes. *Communications in Statistics—Theory and Methods* 20, 1015–1026.

Bickel, P.J. & P. Bühlmann (1999) A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* 5, 413–446.

Bickel, P.J. & D.A. Freedman (1981) Some asymptotic theory for the bootstrap. *Annals of Statistics* 9, 1196–1217.

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.

Brillinger, D.R. (1975) *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart and Winston.

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods*. New York: Springer-Verlag.

Bühlmann, P. (1995) Moving-average representation of autoregressive approximations. *Stochastic Processes and Their Applications* 60, 331–342.

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press

Bühlmann, P. (1997) Sieve bootstrap for time series. *Bernoulli* 3, 123–148.
Bühlmann, P. (1998) Sieve bootstrap for smoothing in nonstationary time series. *Annals of Statistics* 26, 48–83.
Datta, S. (1996) On asymptotic properties of bootstrap for AR(1) processes. *Journal of Statistical Planning and Inference* 53, 361–374.
Dickey, D.A. & W.A. Fuller (1979) Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.
Dickey, D.A. & W.A. Fuller (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49, 1057–1072.
Elliott, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.
Ferretti, N. & J. Romo (1996) Unit root bootstrap tests for AR(1) models. *Biometrika* 83, 849–860.
Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. New York: Academic Press.
Hannan, E.J. & L. Kavalieris (1986) Regression, autoregression models. *Journal of Time Series Analysis* 7, 27–49.
Horowitz, J.L. (1999) The Bootstrap. Mimeo, Department of Economics, University of Iowa.
Kreiss, J.P. (1992) Bootstrap procedures for AR(∞)-processes. In K.H. Jöckel, G. Rothe, & W. Sender (eds.), *Bootstrapping and Related Techniques*, Lecture Notes in Economics and Mathematical Systems 376. Heidelberg: Springer.
Kreiss, J.P. & G. Heimann (1996) Bootstrapping general first order autoregression. *Statistics and Probability Letters* 30, 87–98.
Künsch, H.R. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.
Móricz, F. (1976) Moment inequalities and the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 35, 298–314.
Park, J.Y. (1992) Canonical cointegrating regressions. *Econometrica* 60, 119–143.
Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.
Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.
Pollard, D. (1984) *Convergence of Stochastic Processes*. New York: Springer-Verlag.
Said, S.E. & D.A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–608.
Sakhanenko, A.I. (1980) On unimprovable estimates of the rate of convergence in invariance principle. In *Nonparametric Statistical Inference*, Colloquia Mathematica Societatis János Bolyai 32. pp. 779–783. Budapest, Hungary.
Shibata, R. (1980) Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics* 8, 147–164.
Stout, W.F. (1974) *Almost Sure Convergence*. New York: Academic Press.

https://doi.org/10.1017/S0266466602182090 Published online by Cambridge University Press