Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 133 (2006) 639–672
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

A bootstrap theory for weakly integrated processes

Joon Y. Park*

Department of Economics, Rice University and Sungkyunkwan University, 6100 Main Street, Houston, TX 77005 1892, USA

Available online 18 August 2005

Abstract

This paper develops a bootstrap theory for models including autoregressive time series with roots approaching to unity as the sample size increases. In particular, we consider the processes with roots converging to unity with rates slower than n^{-1}. We call such processes *weakly* integrated processes. It is established that the bootstrap relying on the estimated autoregressive model is generally consistent for the weakly integrated processes. Both the sample and bootstrap statistics of the weakly integrated processes are shown to yield the same normal asymptotics. Moreover, for the asymptotically pivotal statistics of the weakly integrated processes, the bootstrap is expected to provide an asymptotic refinement and give better approximations for the finite sample distributions than the first order asymptotic theory. For the weakly integrated processes, the magnitudes of potential refinements by the bootstrap are shown to be proportional to the rate at which the root of the underlying process converges to unity. The order of boostrap refinement can be as large as $o(n^{-1/2+\epsilon})$ for any $\epsilon > 0$. Our theory helps to explain the actual improvements observed by many practitioners, which are made by the use of the bootstrap in analyzing the models with roots close to unity.
© 2005 Elsevier B.V. All rights reserved.

JEL classification: C12; C15; C22

Keywords: Bootstrap; Unit root; Weakly integrated processes; Consistency; Asymptotic refinement; Linear processes; Sieve bootstrap

*Tel.: +713 348 2354; fax: +713 348 5278.
E-mail address: jpark@rice.edu.

0304-4076/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2005.06.009

1. Introduction

It is now widely understood that the standard bootstrap theory based on the independent sampling extends well to the dependent time series model, if an appropriate model is fit and the bootstrap samples are obtained by resampling independent constituents of the model and reconstructing the data using estimated parameters. For stationary AR processes, Bose (1988) shows that the bootstrap is consistent and does provide the asymptotic refinement exactly as for the simple independent and identically distributed model, once the correct model is fit and the bootstrap samples are generated from the repeated samples of the fitted errors using estimated AR coefficients. The idea of fitting a model and recovering bootstrap samples based on the fitted model works even for an infinite order AR, as demonstrated by Kreiss (1992), as long as we increase the order of the fitted AR with the sample size at an appropriate rate. The procedure, which is often referred to as the sieve bootstrap, was further investigated and developed by Bühlmann (1997).

It is, however, well known that such a method relying on the fitted model does not work for nonstationary integrated processes. Indeed, Basawa et al. (1991) show that for AR(1) model the bootstrap based on the fitted regression becomes inconsistent if the process has a unit root, i.e., the resulting bootstrap distribution is different from the sample distribution even asymptotically. Though the result itself is surprising, it is not at all difficult to rationalize. We have a sharp discontinuity in the asymptotics of AR regressions around the neighborhood of the unit root, and the estimated model is not close enough to yield the same asymptotics. This is precisely the reason that we have the bootstrap inconsistency for the unit root model. If the unit root is imposed to generate samples, the bootstrap does what it is expected to do also for the unit root models. As shown recently by Park (2003a), the bootstrap becomes consistent and gives an asymptotic refinement if the presence of the unit root is imposed in generating bootstrap samples.

For the model with a root in the neighborhood of unity, the validity/invalidity of the bootstrap depends on whether or not the model can be estimated within the boundary that permits the continuity of the asymptotics. For the unit root model, the asymptotics are continuous in the $o_p(n^{-1})$-neighborhood, while the estimated coefficient is only in the $O_p(n^{-1})$-neighborhood. This would naturally lead to the inconsistency of the bootstrap based on the estimated model. Though not mentioned explicitly in the literature, it is also clear that the bootstrap becomes inconsistent for the models with roots approaching to unity at the n^{-1}-rate, which have been referred to as models having roots local-to-unity, since the estimation error is of order $O_p(n^{-1})$. For such models, the estimation error is of the same order as the rate at which the root is approaching to unity. This, however, is not so for all models with roots approaching to unity. As Park (2003b) shows, the processes with roots approaching to unity at a rate slower than n^{-1}, so called *weakly* integrated processes, have characteristics very different from the *near*-integrated processes with roots converging to unity at n^{-1}-rate.

This paper develops a bootstrap theory for the weakly integrated processes. The error involved in estimating the weakly integrated process is of a smaller order of magnitude than the rate of its root approaching to unity, and the estimation error

therefore becomes negligible. It is thus well expected that we have the bootstrap consistency for the weakly integrated processes. The estimated model generates the bootstrap samples with roots approaching to unity at the same rate as the original samples. Moreover, the bootstrap provides a refinement for the limiting distribution. As we show in the paper, the bootstrap distribution more closely approximate the distributions of sample statistics in finite samples. For the AR(1) model, the refinement can be as large as $o(n^{-1/2+\varepsilon})$ for some $\varepsilon>0$. Unlike the bootstrap for stationary models, the primary refinement for weakly integrated processes comes from utilizing the information on the estimated AR coefficients. For the weakly integrated processes, the estimated AR coefficients contain useful information and it is this information that provides the primary asymptotic refinement.

The theory developed in the paper helps to explain the actual improvements observed by many practitioners, which come from the use of the bootstrap in analyzing the models with roots close to unity. In fact, Nankervis and Savin (1996) clearly demonstrates through an extensive simulation that there is a huge potential for improvements, which can be achieved by using bootstrap in models with weakly integrated processes. Their simulation evidence indeed makes it clear that the potential for bootstrap refinement becomes larger as the root of the model gets closer to unity. Nevertheless, none of the existing bootstrap theory is able to give an insight to this observational fact. Our results provide some, if not all, obvious reconciliations. They predict that the bootstrap for the weakly integrated processes yields the distributions for sample statistics closer than their first order normal asymptotics just as for the stationary processes, and that the magnitudes of refinements become larger for the processes with the roots approaching to unity faster as we observe in practice.

The rest of the paper is organized as follows. Section 2 introduces the model and the main issues for the bootstrap of weakly integrated processes. Technical preliminaries involving the probabilistic embeddings of the partial sum processes to the limit Ornstein–Uhlenbeck processes are also given there with a discussion on the distributional effects of approximation errors. The sample and bootstrap asymptotics and their expansions are developed in Section 3. Also discussed there are the bootstrap refinements. The extensions to the bootstrap for more general weakly integrated processes driven by linear processes are made in Section 4. A sieve bootstrap based on an approximated finite AR model is considered and the relevant asymptotics are derived. Moreover, bootstrap refinements for more general models are also discussed. Section 5 concludes the paper, and all the mathematical proofs are given in Appendix.

2. The model and preliminaries

2.1. The model and main issues

Consider the time series (x_t) generated as

$$x_t = \alpha x_{t-1} + \varepsilon_t, \tag{1}$$

where we assume

Assumption 2.1. Let

$$\alpha = 1 - \frac{m}{n},$$

where $n, m \to \infty$ and $m/n \to 0$.

Assumption 2.2. Let (ε_t) be iid random variables with $\mathbf{E}\varepsilon_t = 0$, $\mathbf{E}\varepsilon_t^2 = \sigma^2$ and $\mathbf{E}|\varepsilon_t|^p < \infty$ for some $p > 2$.

In our formulation of α in Assumption 2.1, m just controls the rate of convergence for α to unity. Whenever it becomes more convenient to properly interpret our subsequent results, we will set m explicitly as a function of n, i.e.,

$$m = v(n) \tag{2}$$

such as $v(n) = \log n$ or n^κ with $0 < \kappa < 1$. The condition for (ε_t) in Assumption 2.2 is not necessary and will be relaxed to a more general stationary process later in Section 4. In particular, the iid assumption is not required for the subsequent development of our theory. We may easily allow them to be a general martingale difference sequence. The iid assumption is made just to make more meaningful the bootstrap of the model, which we will discuss below.

Under Assumptions 2.1 and 2.2, the time series (x_t) represents a process that behaves asymtotically as a random walk. If we let n be the sample size, our model describes a time series that has a root approaching to unity as the number of samples increases. Such a time series has been modelled previously by various authors using the formulation similar to ours with m replaced by a constant $c \neq 0$, and referred to as a near-integrated process or a process with a near-unit root or a root local-to-unity. Our model is different in that we let m grow as the sample size n increases.

For the formulation of α in Assumption 2.1, it is important how we set m in relation to n. We consider three possibilities: $m = 0$, $m = c \neq 0$ and $m \to \infty$ such that $m/n \to 0$. The first specification with $m = 0$ yields an exact unit root or integrated process. Both the second and third cases generate processes with roots that are asymptotically unity. The only difference between them is that the root for the former converges faster than that for the latter. However, it turns out that the distinction between the last two cases is much more meaningful than the one between the first two cases. The specifications $m = 0$ and $m = c \neq 0$ do generate processes behaving differently in large samples. Nevertheless, the difference in their asymptotic behaviors is of no qualitative nature and of no importance from the practical point of view. The models with time series specified as above with $m = 0$ and $m = c \neq 0$ have asymptotic properties that are largely comparable: They have estimators and test statistics converging at the same rates and limiting distributions having similar statistical properties. This is well known. On the other hand, as shown by Park (2003b), the specifications $m = c \neq 0$ and $m \to \infty$ such that $m/n \to 0$ yield time series having properties that are drastically different.

For this reason, we follow Park (2003b) and refer to the time series with each of the specifications $m = 0$, $m = c \neq 0$ and $m \to \infty$ such that $m/n \to 0$ respectively as the *exact*, *near* and *weakly* integrated processes. Under this convention, the time

series (x_t) in our model is a process having a weak unit root, and compares itself with a process having a root in the n^{-1}-neighborhood of unity, i.e., a near-unit root that has frequently been considered in the literature.

Our main purpose is to show that the bootstrap of a weakly integrated process does provide an asymptotic refinement, i.e., the bootstrap yields a distribution that is closer to the finite sample distribution, relative to the first order asymptotic theory. To bootstrap (x_t), we first fit the model

$$x_t = \hat{\alpha} x_{t-1} + \hat{\varepsilon}_t \tag{3}$$

and obtain bootstrap samples $(\varepsilon_t^*)_{t=1}^n$ from the centered fitted residuals

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

We may then generate bootstrap samples $(x_t^*)_{t=1}^n$ recursively by

$$x_t^* = \hat{\alpha} x_{t-1}^* + \varepsilon_t^*$$

starting from $x_0^* = x_0$. Throughout the paper, we assume $x_0 = 0$ for expositional simplicity.

It should be emphasized here that the suggested bootstrap procedure uses the estimated coefficient $\hat{\alpha}$ to generate bootstrap samples (x_t^*). As will be shown in the next section, we have

$$\hat{\alpha} = 1 - \frac{m}{n} + O_p\left(\frac{\sqrt{m}}{n} \right) \tag{4}$$

and therefore the bootstraps for models involving weakly integrated processes become generally consistent. Note that, for the weakly integrated processes, the bootstrap samples generated using the fitted AR coefficient behave again like weakly integrated processes. As can clearly be seen from (4), the estimation error in $\hat{\alpha}$ is of order smaller than the distance of α from unity, and becomes negligible as the sample size increases.

This is not so for the exact or near-integrated processes. For the exact unit root model, the bootstrap samples roughly behave like near-integrated processes if the estimated AR coefficient is used. Note that the estimated AR coefficient has an error of order $O_p(1/n)$ in this case. As shown by Basawa et al. (1991), the use of estimated AR coefficient would thus lead to bootstrap inconsistency. Likewise, it is easy to see that such a problem also arises for the near-unit root model. Our result in (4) holds in this case with $m = c$, and the estimation error becomes nonnegligible and affects bootstrap samples persistently even for large samples. The bootstrap samples from the near-unit root model, if generated using the estimated AR coefficient, are near-integrated processes just like the original samples, since the estimation error is of order $O_p(1/n)$. They, however, have differing local parameters c, and the bootstrap becomes inconsistent just as for the exact unit root model.

For the weakly integrated process, the bootstrap refinement comes primarily from utilizing the information on α in the sample that is revealed through $\hat{\alpha}$. The primary source of refinement here is therefore somewhat different from that of the usual

bootstrap, which is the empirical distribution estimating underlying distribution nonparametrically and consistently. For the weakly integrated process, the estimated AR coefficient $\hat{\alpha}$ contains useful information on its weak unit root property, and it is by utilizing this information that the bootstrap provides most significant refinement over the first order asymptotics. Therefore, the way that the bootstrap samples (ε_t^*) are obtained has only secondary importance, though it may well have nonnegligible effects on the bootstrap distribution of (x_t^*). Our theories are applicable to any parametric, as well as the usual nonparametric, bootstrap. As long as the bootstrap samples (x_t^*) are generated using the estimated AR coefficient $\hat{\alpha}$, they will have the weak unit root property inherited from (x_t) and thus provide the asymptotic refinement that we recognize and establish in the paper. This point will be made clear in the next section.

2.2. Technical preliminaries

The sample and bootstrap asymptotics developed in the paper rely on the probabilistic embedding

$$V_{mn}(r) =_d n^{-1/2} x_{[nr]} \tag{5}$$

for $r \in [0, 1]$, where $[z]$ denotes the largest integer not exceeding z. The process V_{mn} in (5) is defined for each n and $m \geqslant 0$. Recall that the root of (x_t) depends on m, as well as on n. For any fixed $m > 0$, it is well known that as $n \to \infty$

$$V_{mn} \to_d V_m,$$

where V_m is an Ornstein–Uhlenbeck process. More explicitly, we let V_0 be Brownian motion with variance σ^2, and define V_m as

$$V_m(r) = \int_0^r \exp(-m(r-s))\, \mathrm{d}V_0(s) \tag{6}$$

for $r \in [0, 1]$. Note that, if we set $m = 0$ in (6), V_m indeed reduces to the Brownian motion V_0.

As shown in Park (2003b), we may construct the processes V_{mn} and V_m in the same probability space so that

Lemma 2.3. *Under Assumptions* 2.1 *and* 2.2, *we have*

$$V_{mn}(r) = V_m(r) + O_p(n^{-1/2+1/p}) + O_p(mn^{-1})$$

uniformly in $r \in [0, 1]$ *and* $m \in \mathbf{R}_+$.

Lemma 2.3 allows us to represent, up to negligible errors, the distributions of various statistics of time series (x_t) by the integrals of the corresponding functionals of continuous process V_m. Furthermore, the magnitudes of the errors incurred by the approximations can be given explicitly. For the special case of $m = 0$, the result in Lemma 2.3 is well known, and has been used extensively in the analysis of nonlinear models with integrated processes. See, e.g., Park and Phillips (2001).

Before we present the corresponding result for the bootstrap samples (x_t^*), it is necessary to introduce some notations that will be used in the paper for the bootstrap samples and statistics. Note that we use the superscript "$*$" to signify the bootstrap samples and statistics, following the usual convention. Likewise, $\mathbf{P}^*$ denotes the bootstrap probability conditional on the realization of (ε_t), and $\mathbf{E}^*$ is used for the expectation taken with respect to $\mathbf{P}^*$. As usual, $\to_{d^*}$ denotes the weak convergence of distributions, and $=_{d^*}$ denotes the distributional equality, conditional on the realization of the samples. Moreover, 'in $\mathbf{P}$' means 'with probability close to unity for all large n', implying that the probability of (ε_t) being realized for which the designated statement holds can be made arbitrarily close to unity by requiring n to be sufficiently large. Our asymptotics in the paper involve the bootstrap stochastic order symbols, which are introduced in Chang and Park (2003). In particular, we denote by $Z_n^* = O_p^*(1)$ in $\mathbf{P}$ if for any $\epsilon > 0$ there exists K such that

$$\mathbf{P}\{\mathbf{P}^*\{|Z_n^*| > K\} > \epsilon\} < \epsilon$$

for all sufficiently large n.

We now define

$$V_{mn}^*(r) =_{d^*} n^{-1/2} x_{[nr]}^* \tag{7}$$

for $r \in [0,1]$, correspondingly as V_{mn} introduced in (5). Then we have

Lemma 2.4. *Under Assumptions* 2.1 *and* 2.2, *we have*

$$V_{mn}^*(r) = V_m(r) + O_p^*(n^{-1/2+1/p}) + O_p^*(mn^{-1}) \quad \textit{in } \mathbf{P}$$

uniformly in $r \in [0,1]$ *and* $m \in \mathbf{R}_+$.

Lemma 2.4 implies that we may have the probabilistic embedding for the bootstrap samples similarly as for the original samples. The distributions of various statistics of the bootstrap samples (x_t^*) can also be approximated by the corresponding functionals of V_m. Upon comparing the results in Lemmas 2.3 and 2.4, it is now well expected that the bootstrap would provide the asymptotic refinement for models with weakly integrated processes. Note that V_{mn} and V_{mn}^* have the common leading term V_m. Therefore, various functionals of V_{mn} and V_{mn}^* are expected to be represented by the same functionals of V_m up to the approximation errors, which would become asymptotically negligible under appropriate regularity conditions. This will be shown in later sections.

Lemmas 2.3 and 2.4 only provide the stochastic orders of the error terms, and cannot be used directly to show the asymptotic refinement of the bootstrap tests. In particular, our results in Lemmas 2.3 and 2.4 do *not* necessarily imply that V_{mn} can be approximated by V_m with an error which is distributionally of order $O(n^{-1/2+1/p})$ or $O(mn^{-1})$. Therefore, we may not readily compare the rejection probabilities of the bootstrap tests with those of the asymptotic tests. To investigate the distributional orders of the approximation errors appearing in Lemmas 2.3 and 2.4, we need to establish that

Corollary 2.5. *Under Assumptions* 2.1 *and* 2.2, *we have*

$$\mathbf{P}\left\{m^{\kappa}\sup_{0\leqslant r\leqslant 1}|V_{mn}(r)-V_m(r)|>\delta_{mn}\right\}\leqslant K\delta_{mn}\mathbf{E}|\varepsilon_t|^p,$$

$$\mathbf{P}^*\left\{m^{\kappa}\sup_{0\leqslant r\leqslant 1}|V^*_{mn}(r)-V_m(r)|>\delta_{mn}\right\}\leqslant K\delta_{mn}\mathbf{E}^*|\varepsilon^*_t|^p \quad \textit{in } \mathbf{P}$$

for any (δ_{mn}) *such that*

$$\delta_{mn}\geqslant \max(m^{\kappa p/(1+p)}n^{(2-p)/2(1+p)}, m^{1+\kappa}n^{-1+\epsilon})$$

with arbitrarily small $\epsilon>0$, *where* K *is an absolute constant depending only upon* p.

We now let $D[0,1]$ be the set of cadlag functions defined on $[0,1]$, and let Γ_m be a functional defined on $D[0,1]$ such that

$$|\Gamma_m(V_{mn})-\Gamma_m(V_m)|\leqslant Km^{\kappa}\sup_{0\leqslant r\leqslant 1}|V_{mn}(r)-V_m(r)|,$$

where K is an absolute constant possibly depending only upon p, and such that $\Gamma_m(V_m)$ has a density bounded uniformly in m. Clearly, Γ_m may be viewed as a Lipschitz functional defined on $D[0,1]$ endowed with the supremum norm. As a direct consequence of Corollary 2.5, we have

$$\mathbf{P}\{\Gamma_m(V_{mn})\leqslant x\}=\mathbf{P}\{\Gamma_m(V_m)\leqslant x\}+R_{mn},$$

$$\mathbf{P}^*\{\Gamma_m(V_{mn})\leqslant x\}=\mathbf{P}\{\Gamma_m(V_m)\leqslant x\}+R_{mn} \quad \text{in } \mathbf{P} \tag{8}$$

uniformly in $x\in\mathbf{R}$, where

$$R_{mn}=o(m^{\kappa/p(1+p)}n^{(2-p)/2(1+p)+\epsilon})+o(m^{1+\kappa}n^{-1+\epsilon})$$

for any $\epsilon>0$. This can be shown using the result in, e.g., Lemma A4 of Park (2003a).

As an illustration, we consider the kth sample moments of (x_t) and (x^*_t). It follows directly from the embeddings in (5) and (7) that

$$\frac{m^{k/2}}{n^{1+k/2}}\sum_{t=0}^{n-1}x_t^k=_d m^{k/2}\int_0^1 V^k_{mn}(r)\,\mathrm{d}r, \quad \frac{m^{k/2}}{n^{1+k/2}}\sum_{t=0}^{n-1}x_t^{*k}=_{d^*}m^{k/2}\int_0^1 V^{*k}_{mn}(r)\,\mathrm{d}r$$

Moreover, if we let M_k be the kth moment of $\mathbf{N}(0,\sigma^2/2)$ distribution and given by $M_k=0$ if $k=2j-1$ and $M_k=(\sigma^2/2)^j\Pi_{i=1}^j(2i-1)$ if $k=2j$ for $j=1,2,\ldots$, then $m^{k/2}\int_0^1 V^k_m(r)\,\mathrm{d}r\to_{\text{a.s.}}M_k$ and $m^{1/2}(m^{k/2}\int_0^1 V^k_m(r)\,\mathrm{d}r-M_k)$ converges weakly to normal law, as $m\to\infty$. This is shown in Park (2003b). The kth sample moment of (x_t) or (x^*_t) may therefore be effectively analyzed if we consider

$$\Gamma_m(V)=m^{1/2}\left(m^{k/2}\int_0^1 V^k(r)\,\mathrm{d}r-M_k\right). \tag{9}$$

Clearly, the functional Γ_m is Lipschitz and $\Gamma_m(V_m)$ has a density bounded uniformly in m. Our results in (8) are thus applicable for the functional Γ_m defined in (9) with $\kappa=(k+1)/2$.

All the test statistics we will subsequently consider can be represented as simple functions of the sample moments of (x_t) and (x_t^*). We may therefore directly compare the rejection probabilities of the bootstrap tests with those of the asymptotic tests as above using the functional Γ_m introduced in (9), and show that the bootstrap provides the asymptotic refinement. This, however, will not be formally pursued in the paper. The exact orders of the discrepancies in the rejection probabilities depend upon m, which we do not observe. Therefore, the absolute magnitudes of the orders are not very useful. In the subsequent lemmas, theorems and corollaries, we simply provide the results as those in Lemmas 2.3 and 2.4, identifying only the leading terms in the expansions with the stochastic orders of magnitude for the approximation errors. However, it should be emphasized here that all the error terms in our subsequent results can be made as small as $o_p(n^{-1/2+\epsilon})$ stochastically and $o(n^{-1/2+\epsilon})$ distributionally, for any $\epsilon>0$, under stringent enough moment conditions for (ε_t) and sufficiently slow divergence rates for m.

3. Sample asymptotics and bootstrap refinements

3.1. Sample asymptotics

As before, we denote by $\hat{\alpha}$ the least squares estimator for the AR coefficient α in regression (3). Let S_{mn} be the normalized coefficient given by

$$S_{mn}=\frac{n}{\sqrt{m}}(\hat{\alpha}-\alpha)=\frac{n\sum_{t=1}^{n}x_{t-1}\varepsilon_t}{\sqrt{m}\sum_{t=1}^{n}x_{t-1}^2}$$

and T_{mn} be the t-ratio given by

$$T_{mn}=\frac{\hat{\alpha}-\alpha}{s(\hat{\alpha})}=\frac{\sum_{t=1}^{n}x_{t-1}\varepsilon_t}{\hat{\sigma}\left(\sum_{t=1}^{n}x_{t-1}^2\right)^{1/2}},$$

where $\hat{\sigma}^2$ is the usual error variance estimator and $s(\hat{\alpha})$ is the standard error for $\hat{\alpha}$.

Lemma 3.1. *Under Assumptions* 2.1 *and* 2.2, *we have*

$$\frac{m}{n^2}\sum_{t=1}^{n}x_{t-1}^2=_d m\int_0^1 V_m^2(r)\,\mathrm{d}r+O_p(m^{1/2}n^{-1/2+1/p})+O_p(m^{3/2}n^{-1}),$$

$$\frac{\sqrt{m}}{n}\sum_{t=1}^{n}x_{t-1}\varepsilon_t=_d\sqrt{m}\int_0^1 V_m(r)\,\mathrm{d}V_0(r)+O_p(m^{1/2}n^{-1/2+1/p})+O_p(m^{3/2}n^{-1}),$$

as $n\to\infty$, *uniformly in* $m\in\mathbf{R}_+$.

Theorem 3.2. *Under Assumptions* 2.1 *and* 2.2, *we have*

$$S_{mn}=_d S_m+O_p(m^{1/2}n^{-1/2+1/p})+O_p(m^{3/2}n^{-1}),$$

$$T_{mn} =_d T_m + O_p(m^{1/2}n^{-1/2+1/p}) + O_p(m^{3/2}n^{-1}),$$

where

$$S_m = \frac{\sqrt{m}\int_0^1 V_m(r)\,\mathrm{d}V_0(r)}{m\int_0^1 V_m^2(r)\,\mathrm{d}r}, \quad T_m = \frac{\sqrt{m}\int_0^1 V_m(r)\,\mathrm{d}V_0(r)}{\sigma(m\int_0^1 V_m^2(r)\,\mathrm{d}r)^{1/2}},$$

as $n \to \infty$, *uniformly in* $m \in \mathbf{R}_+$, *and*

$$\frac{S_m}{\sqrt{2}},\ T_m \to_d \mathbf{N}(0,1),$$

as $m \to \infty$.

The role of m in our asymptotics should first be clarified to properly interpret the results in Theorem 3.2 and the subsequent results. To do so, we let (S, T) represent the limit random variables of (S_m, T_m), and assume that (S_{mn}, T_{mn}) is defined on the same probability space as (S_m, T_m) and (S, T). Then we write

$$S_{mn} = S + (S_m - S) + (S_{mn} - S_m),$$

$$T_{mn} = T + (T_m - T) + (T_{mn} - T_m).$$

It now follows from Theorem 3.2 that

$$A_m = S_m - S \quad \text{or } T_m - T$$

become small as $m \to \infty$, and that

$$B_{mn} = S_{mn} - S_m \quad \text{or } T_{mn} - T_m$$

can be made negligible uniformly in $m \in \mathbf{R}_+$ as $n \to \infty$.

Unless m increases too fastly relative to n, we may well expect

$$|A_m| > |B_{mn}| = O_p(m^{1/2}n^{-1/2+1/p}) + O_p(m^{3/2}n^{-1}). \tag{10}$$

Indeed, this is what is likely to happen when

$$m = o(\min(n^{2/3-4/3p}, n^{4/7})),$$

since A_m is of order at most $O_p(m^{-1/4})$, as one may see from proof of Theorem 3.2. Here we should notice that Theorem 3.2 does provide *asymptotic expansions* for (S_{mn}, T_{mn}), if A_m and B_{mn} satisfy the condition given in (10). In such cases, $S_m = S + (S_m - S)$ and $T_m = T + (T_m - T)$ represent the two leading terms of the expansions for S_{mn} and T_{mn}, respectively.

The precision and usefulness of the expansions, of course, depend on how slowly m increases relatively to n. Our expansions, in particular, are not valid when m increases as fast as n, in which case (x_t) becomes a stationary AR process. Recall that we assume $m/n \to 0$. If m increases at a slower rate relative to n, the approximation errors have reduced orders, and the leading terms in the expansions become more dominant. As m increases slowly (fastly) relative to n, m/n diminishes fastly (slowly) to zero, which in turn implies that $\alpha = 1 - m/n$ approaches fastly (slowly) to unity. Therefore, our expansions are generally more useful for the time series with roots

closer to unity. If, for instance, $m = \log n$ and $\alpha = 1 - m/n = 1 - \log n/n$, the expansions have the error terms of order $o_p(n^{-1/2+1/p} \log n)$. The order would further be reduced to $o_p(n^{-1/2+\epsilon})$ for any $\epsilon > 0$, if all moments of (ε_t) are finite. The leading terms in this case would thus provide reasonably good approximations for the finite sample distributions of S_{mn} and T_{mn}. Note that (S_{mn}, T_{mn}) are asymptotically pivotal, and the distributions of (S_m, T_m) and (S, T) do not depend upon any nuisance parameter.

Under stringent enough moment conditions for (ε_t), it readily follows from the results in Theorem 3.2 that

$$\mathbf{P}\{S_{mn} \leqslant x\} = \mathbf{P}\{S_m \leqslant x\} + o(m^{1/2} n^{-1/2+1/p+\epsilon}) + o(m^{3/2} n^{-1+\epsilon}), \tag{11}$$

$$\mathbf{P}\{T_{mn} \leqslant x\} = \mathbf{P}\{T_m \leqslant x\} + o(m^{1/2} n^{-1/2+1/p+\epsilon}) + o(m^{3/2} n^{-1+\epsilon}), \tag{12}$$

which hold uniformly in $x \in \mathbf{R}$ for any $\epsilon > 0$. The asymptotic expansions in (11) and (12) are more comparable to the Edgeworth type expansions for the usual stationary models, and also more directly applicable to investigate the asymptotic refinement of the bootstrap as we will see later. Given Corollary 2.5, the results in (11) and (12) can easily be derived exactly as in the proof of Corollary 3.8 in Park (2003a).

The usual asymptotics can also be easily derived from Theorem 3.2. Indeed, it can be easily deduced from Theorem 3.2 that

$$\frac{S_{mn}}{\sqrt{2}}, \ T_{mn} \to_d \mathbf{N}(0, 1)$$

as long as $n, m \to \infty$ such that $m/n \to 0$. Alternatively, we may set m explicitly as a function of n as in (2) with $v(n)$ such that $v(n)/n \to 0$ as $n \to \infty$. Then it follows from Theorem 3.2 that

$$\frac{S_{v(n)n}}{\sqrt{2}}, \ T_{v(n)n} \to_d \mathbf{N}(0, 1)$$

as $n \to \infty$, more conformably with the usual asymptotics. The models with weakly integrated processes, if correctly specified, have normal asymptotics in sharp constrast with the unit root and cointegrated models. The reader is referred to Park (2003b) for more details on the asymptotics of models with weakly integrated time series.

We also consider the estimators for α in the regressions with constant and linear time trend as given by

$$y_t = \hat{\mu} + \hat{\alpha}^{\mu} x_{t-1} + \hat{\varepsilon}_t, \tag{13}$$

$$y_t = \hat{\mu} + \hat{\tau} t + \hat{\alpha}^{\tau} x_{t-1} + \hat{\varepsilon}_t, \tag{14}$$

in addition to regression (3). In what follows, we will only consider the least squares estimators $\hat{\alpha}^{\mu}$ and $\hat{\alpha}^{\tau}$ of α in regressions (13) and (14). The least squares estimators of other parameters can be analyzed similarly.

Lemma 3.3. *We have*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t =_d \int_0^1 \mathrm{d}V_0(r) + O_p(n^{-1/2+1/p}),$$

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n} t\varepsilon_t =_d \int_0^1 r\,\mathrm{d}V_0(r) + O_p(n^{-1/2+1/p}),$$

$$\frac{\sqrt{m}}{n^{3/2}}\sum_{t=1}^{n} x_{t-1} =_d \sqrt{m}\int_0^1 V_m(r)\,\mathrm{d}r + O_p(m^{1/2}n^{-1/2+1/p}) + O_p(m^{3/2}n^{-1}),$$

$$\frac{\sqrt{m}}{n^{5/2}}\sum_{t=1}^{n} t x_{t-1} =_d \sqrt{m}\int_0^1 rV_m(r)\,\mathrm{d}r + O_p(m^{1/2}n^{-1/2+1/p}) + O_p(m^{3/2}n^{-1}),$$

as $n \to \infty$, *uniformly in* $m \in \mathbf{R}_+$.

Define $(S^{\mu}_{mn}, T^{\mu}_{mn})$ and $(S^{\tau}_{mn}, T^{\tau}_{mn})$, respectively for regressions (13) and (14), correspondingly as (S_{mn}, T_{mn}) defined for regression (3). Moreover, denote by $(S^{\mu}_{m}, T^{\mu}_{m})$ and $(S^{\tau}_{m}, T^{\tau}_{m})$ the leading terms in the asymptotic expansions of $(S^{\mu}_{mn}, T^{\mu}_{mn})$ and $(S^{\tau}_{mn}, T^{\tau}_{mn})$, respectively, analogously as our notations (S_m, T_m) for (S_{mn}, T_{mn}). Then we have

Corollary 3.4. *Let* $(S^{\mu}_{m}, T^{\mu}_{m})$ *and* $(S^{\tau}_{m}, T^{\tau}_{m})$ *be defined analogously as* (S_m, T_m) *with* V_m *replaced respectively by* V^{μ}_{m} *and* V^{τ}_{m}, *where*

$$V^{\mu}_{m}(r) =_d V_m(r) - \int_0^1 V_m(s)\,\mathrm{d}s,$$

$$V^{\tau}_{m}(r) =_d V_m(r) + (6t-4)\int_0^1 V_m(s)\,\mathrm{d}s - (12t-6)\int_0^1 sV_m(s)\,\mathrm{d}s.$$

Then Theorem 3.2 *continues to hold for* $(S^{\mu}_{mn}, T^{\mu}_{mn})$ *and* $(S^{\tau}_{mn}, T^{\tau}_{mn})$ *with* (S_m, T_m) *replaced respectively by* $(S^{\mu}_{m}, T^{\mu}_{m})$ *and* $(S^{\tau}_{m}, T^{\tau}_{m})$.

In particular, Corollary 3.4 implies that $(S^{\mu}_{mn}, T^{\mu}_{mn})$ and $(S^{\tau}_{mn}, T^{\tau}_{mn})$ have the leading expansion terms $(S^{\mu}_{m}, T^{\mu}_{m})$ and $(S^{\tau}_{m}, T^{\tau}_{m})$. Our earlier remarks on the interpretations of the leading expansion terms (S_m, T_m) for (S_{mn}, T_{mn}) also apply to the leading expansion terms $(S^{\mu}_{m}, T^{\mu}_{m})$ and $(S^{\tau}_{m}, T^{\tau}_{m})$ respectively for $(S^{\mu}_{mn}, T^{\mu}_{mn})$ and $(S^{\tau}_{mn}, T^{\tau}_{mn})$. Furthermore, the formal asymptotic expansions similar to those in (11) and (12) can also be obtained for both $(S^{\mu}_{mn}, T^{\mu}_{mn})$ and $(S^{\tau}_{mn}, T^{\tau}_{mn})$ under appropriate moment conditions for (ε_t).

We now turn to the distributions of the leading terms in our asymptotic expansions. Note that they are dependent only on m, and converge in distribution to normal distributions as $m \to \infty$. Figs. 1 and 2 present the densities of $(S_m, S^{\mu}_{m}, S^{\tau}_{m})$ and $(T_m, T^{\mu}_{m}, T^{\tau}_{m})$, respectively, for the values of $m = 10, 100$ and 1000, and compare them with their limit normal densities. The distributions of (S_m, T_m) are pretty close to the limit normal distributions, even when m is fairly small. Even for $m = 10$, their

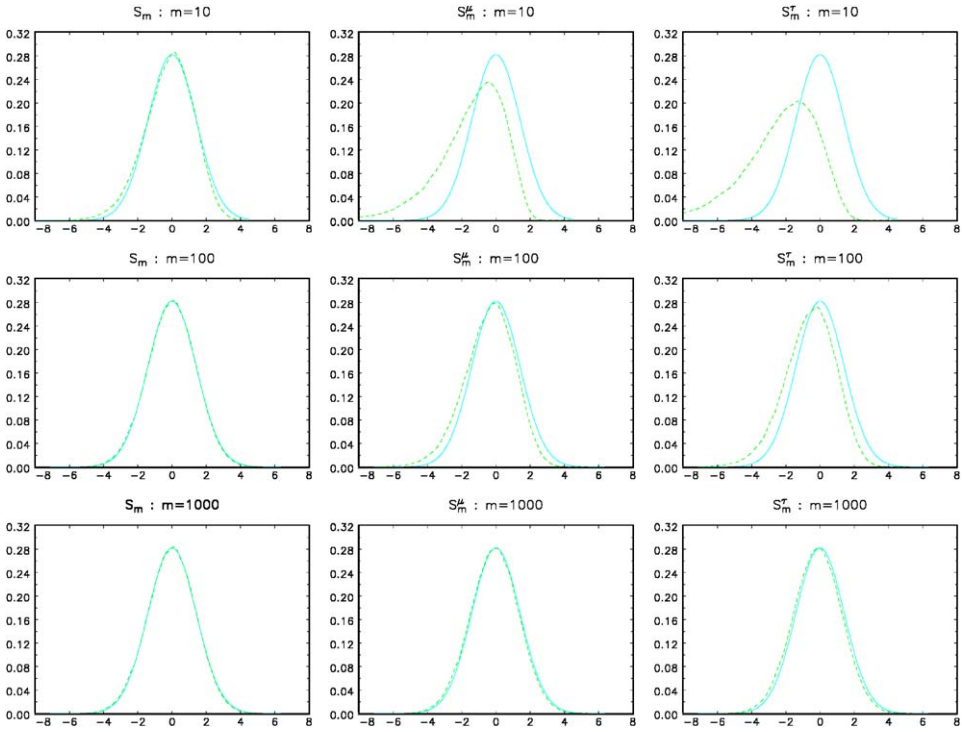


Fig. 1. Probability densities for S_m.

distributions are not significantly different from the limit normals. On the other hand, the distributions of (S_m^τ, T_m^τ) are very distinct from their limit distributions unless m becomes fairly large. When m is as small as 10, their discrepancies from the limit normals are quite substantial. It also appears that the distributions of (S_m^μ, T_m^μ) can be quite different from their normal limits, though the differences are not as large as (S_m^τ, T_m^τ). The distributions of S_m^μ and S_m^τ are noticeably skewed when m is small. This is not the case for T_m^μ and T_m^τ, whose distributions are mislocated but remain to be symmetric even for very small m.

3.2. Bootstrap refinements

Let $\hat{\alpha}^*$ be the least squares estimator for the AR coefficient α in regression (3) obtained using bootstrap samples (x_t^*). Define the bootstrap version of S_{mn} by

$$S_{mn}^* = \frac{n}{\sqrt{m}}(\hat{\alpha}^* - \hat{\alpha}) = \frac{n\sum_{t=1}^n x_{t-1}^* e_t^*}{\sqrt{m}\sum_{t=1}^n x_{t-1}^{*2}}$$

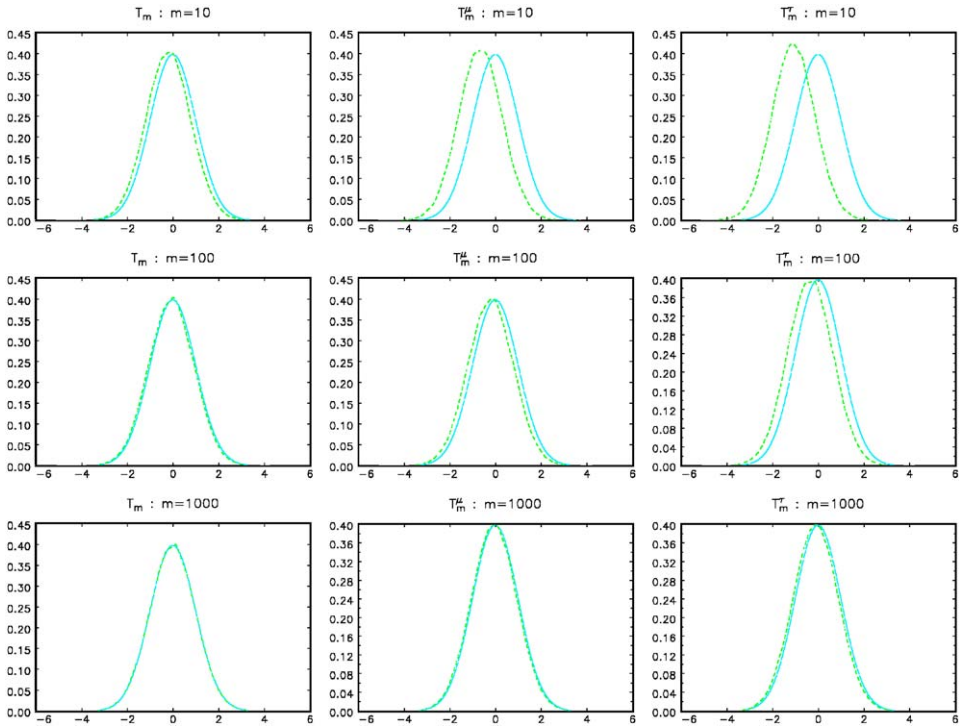


Fig. 2. Probability densities for T_m.

and the bootstrap version of T_{mn} by

$$T^*_{mn} = \frac{\hat{\alpha}^* - \hat{\alpha}}{s(\hat{\alpha}^*)} = \frac{\sum_{t=1}^n x^*_{t-1}\varepsilon^*_t}{\hat{\sigma}^*(\sum_{t=1}^n x^{*2}_{t-1})^{1/2}},$$

where $\hat{\sigma}^{*2}$ is the error variance estimator computed from bootstrap samples (x^*_t), and $s(\hat{\alpha}^*)$ is the standard error for $\hat{\alpha}^*$. Moreover, define $(S^{\mu*}_{mn}, T^{\mu*}_{mn})$ and $(S^{\tau*}_{mn}, T^{\tau*}_{mn})$ for the bootstrap versions of the regressions (13) and (14), respectively, analogously as (S^*_{mn}, T^*_{mn}) for the bootstrap version of regression (3).

Lemma 3.5. *Under Assumptions* 2.1 *and* 2.2, *we have*

$$\frac{m}{n^2}\sum_{t=1}^n x^{*2}_{t-1} =_{d^*} m\int_0^1 V^2_m(r)\,\mathrm{d}r + O^*_p(m^{1/2}n^{-1/2+1/p}) + O^*_p(m^{3/2}n^{-1}) \quad \textit{in } \mathbf{P}$$

$$\frac{\sqrt{m}}{n}\sum_{t=1}^n x^*_{t-1}\varepsilon^*_t =_{d^*} \sqrt{m}\int_0^1 V_m(r)\,\mathrm{d}V_0(r) + O^*_p(m^{1/2}n^{-1/2+1/p}) + O^*_p(m^{3/2}n^{-1}) \quad \textit{in } \mathbf{P}$$

as $n \to \infty$, *uniformly in* $m \in \mathbf{R}_+$.

Lemma 3.6. *Under Assumptions* 2.1 *and* 2.2, *we have*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t^* =_{d^*} \int_0^1 \mathrm{d}V_0(r) + O_p^*(n^{-1/2+1/p}) \quad \textit{in } \mathbf{P},$$

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n} t\varepsilon_t^* =_{d^*} \int_0^1 r\,\mathrm{d}V_0(r) + O_p^*(n^{-1/2+1/p}) \quad \textit{in } \mathbf{P},$$

$$\frac{\sqrt{m}}{n^{3/2}}\sum_{t=1}^{n} x_{t-1}^* =_{d^*} \sqrt{m}\int_0^1 V_m(r)\,\mathrm{d}r + O_p^*(m^{1/2}n^{-1/2+1/p}) + O_p^*(m^{3/2}n^{-1}) \quad \textit{in } \mathbf{P},$$

$$\frac{\sqrt{m}}{n^{5/2}}\sum_{t=1}^{n} tx_{t-1}^* =_{d^*} \sqrt{m}\int_0^1 rV_m(r)\,\mathrm{d}r + O_p^*(m^{1/2}n^{-1/2+1/p}) + O_p^*(m^{3/2}n^{-1}) \quad \textit{in } \mathbf{P},$$

as $n \to \infty$, *uniformly in* $m \in \mathbf{R}_+$.

Theorem 3.7. *Under Assumptions* 2.1 *and* 2.2, *we have*

$$S_{mn}^* =_{d^*} S_m + O_p^*(m^{1/2}n^{-1/2+1/p}) + O_p^*(m^{3/2}n^{-1}) \quad \textit{in } \mathbf{P},$$

$$T_{mn}^* =_{d^*} T_m + O_p^*(m^{1/2}n^{-1/2+1/p}) + O_p^*(m^{3/2}n^{-1}) \quad \textit{in } \mathbf{P},$$

where S_m *and* T_m *are defined in Theorem* 3.2.

Corollary 3.8. *Theorem* 3.7 *continues to hold for* $(S_{mn}^{\mu*}, T_{mn}^{\mu*})$ *and* $(S_{mn}^{\tau*}, T_{mn}^{\tau*})$ *with* (S_m, T_m) *replaced respectively by* (S_m^{μ}, T_m^{μ}) and (S_m^{τ}, T_m^{τ}).

Lemma 3.5 and Theorem 3.7 are completely analogous to Lemma 3.1 and Theorem 3.2, and provide the asymptotics for regression (3). Lemma 3.6 and Corollary 3.8 correspond to Lemma 3.3 and Corollary 3.4, and are for regressions (13) and (14) with intercept and linear time trend.

Theorem 3.7 shows that the bootstrap distributions of (S_{mn}^*, T_{mn}^*) are asymptotically identical to those of (S_{mn}, T_{mn}) as $n, m \to \infty$ such that $m/n \to 0$. Both have limit normal distributions. The bootstrap consistency is thus established. More importantly, however, Theorem 3.7 shows that the bootstrap provides the asymptotic refinements for (S_{mn}, T_{mn}). The asymptotic expansions for (S_{mn}, T_{mn}) and (S_{mn}^*, T_{mn}^*) have the same leading terms (S_m, T_m), which approximate the finite sample distributions of (S_{mn}, T_{mn}) and (S_{mn}^*, T_{mn}^*) up to the errors that become negligible if m increases slowly enough. Note that the leading terms in the asymptotic expansions of (S_{mn}^*, T_{mn}^*) have distributions not depending upon the sample realizations. Due to Corollary 3.8, all these discussions on the asymptotic refinement of the bootstrap for (S_{mn}, T_{mn}) extend to $(S_{mn}^{\mu}, T_{mn}^{\mu})$ and $(S_{mn}^{\tau}, T_{mn}^{\tau})$ for the regressions with intercept and linear time trend.

Analogously as (11) and (12), we may deduce from the results in Theorem 3.7 that

$$\mathbf{P}^*\{S^*_{mn} \leqslant x\} = \mathbf{P}\{S_m \leqslant x\} + o_p(m^{1/2}n^{-1/2+1/p+\epsilon}) + o_p(m^{3/2}n^{-1+\epsilon}), \tag{15}$$

$$\mathbf{P}^*\{T^*_{mn} \leqslant x\} = \mathbf{P}\{T_m \leqslant x\} + o_p(m^{1/2}n^{-1/2+1/p+\epsilon}) + o_p(m^{3/2}n^{-1+\epsilon}), \tag{16}$$

uniformly in $x \in \mathbf{R}$ for any $\epsilon > 0$. Similarly as (11) and (12), the bootstrap asymptotic expansions in (15) and (16) can be derived in a straightforward manner from Corollary 2.5 as in the proof of Corollary 3.11 in Park (2003a), under stringent enough moment conditions for (ε_t). The bootstrap asymptotic expansions for $(S^{\mu*}_{mn}, T^{\mu*}_{mn})$ and $(S^{\tau*}_{mn}, T^{\tau*}_{mn})$ corresponding to those in (15) and (16) can be obtained similarly.

Let a^*_λ and b^*_λ denote, respectively, the bootstrap critical values for the size λ tests based on the statistics S_{mn} and T_{mn}, which are given by

$$\mathbf{P}^*\{S^*_{mn} \leqslant a^*_\lambda\},\ \mathbf{P}^*\{T^*_{mn} \leqslant b^*_\lambda\} = \lambda.$$

Then we may easily deduce, by comparing (11) with (15) and (12) with (16), that

$$\mathbf{P}\{S_{mn} \leqslant a^*_\lambda\},\ \mathbf{P}\{T_{mn} \leqslant b^*_\lambda\} = \lambda + o_p(m^{1/2}n^{-1/2+1/p+\epsilon}) + o_p(m^{3/2}n^{-1+\epsilon}) \tag{17}$$

for any $\epsilon > 0$. The results in (17) show that the bootstrap provides the asymptotic refinements for S_{mn} and T_{mn} as long as (10) holds. In this case, the tests relying on the bootstrap critical values a^*_λ and b^*_λ, in place of their asymptotic values, have the actual rejection probabilities that are closer to their nominal values. Of course, similar results hold for $(S^{\mu}_{mn}, T^{\mu}_{mn})$ and $(S^{\tau}_{mn}, T^{\tau}_{mn})$ for the regressions with intercept and linear time trend.

The orders of the bootstrap refinements here depend on how fast $\alpha = 1 - m/n$ approaches to unity. If m increases slowly relative to n and $\alpha = 1 - m/n$ coverges to unity fastly, the magnitudes of the error terms become smaller and the common leading terms (S_m, T_m) in the asymptotic expansions of (S_{mn}, T_{mn}) and (S^*_{mn}, T^*_{mn}) more precisely represent their finite sample distributions. If we set $m = \log n$ so that $\alpha = 1 - m/n$ converges to the unity nearly as fast as in the case for the quasi-integrated process, then the orders of the bootstrap refinements can be made as large as $o^*_p(n^{-1/2+\epsilon})$ for any $\epsilon > 0$. This rate can actually be attained if the innovation has moments finite at all orders. The magnitudes of bootstrap refinements also depend upon how far away are the distributions of the leading terms from the limit normal distributions. Of three regressions (3), (13) and (14), we may expect most substantial bootstrap refinements for the regression with linear time trend, since as shown in Figs. 1 and 2 the distributions of the leading terms are most distinct from the limit normal distributions. This is indeed exactly what was found by Nankervis and Savin (1996) through an extensive simulation.

4. Extensions

4.1. A sieve bootstrap

Now we consider more general model

$$x_t = \alpha x_{t-1} + v_t, \tag{18}$$

where α is specified as in Assumption 2.1, and (v_t) is a general linear process given by

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

with $\pi(z) = \sum_{i=0}^{\infty} \pi_i z^i$.

Assumption 4.1. We assume that

(a) (ε_t) are iid random variables such that $\mathbf{E}\varepsilon_t = 0$, $\mathbf{E}\varepsilon_t^2 = \sigma^2$ and $\mathbf{E}|\varepsilon_t|^p < \infty$ for some $p \geqslant 4$,
(b) $\pi(z) \neq 0$ for all $|z| \leqslant 1$, and $\sum_{i=0}^{\infty} |i|^q |\pi_i| < \infty$ for some $q \geqslant 1$.

Let

$$V_{mn}(r) =_d \frac{x_{[nr]}}{\pi(1)\sqrt{n}}$$

and V_m be defined as in Section 2. Then we have

Lemma 4.2. *Under Assumptions* 2.1 *and* 4.1, *we have*

$$V_{mn}(r) = V_m(r) + O_p(n^{-1/2+1/p}) + O_p(mn^{-1})$$

uniformly in $r \in [0,1]$ *and* $m \in \mathbf{R}_+$.

Under Assumption 4.1, we may write (v_t) as an infinite order AR. In what follows, we will write

$$\beta(L)v_t = \varepsilon_t \tag{19}$$

with $\beta(z) = 1 - \sum_{i=1}^{\infty} \beta_i z^i$. Consequently, (x_t) is generated as

$$(1 - \alpha L)\beta(L)x_t = \varepsilon_t,$$

i.e., an infinite order AR process with a weak unit root. If we define $\beta_\kappa(z) = 1 - \sum_{i=1}^{\kappa-1} \beta_i z^i$, then we may write

$$(1 - \alpha L)\beta_\kappa(L)x_t = \varepsilon_{\kappa,t}.$$

We assume that κ satisfies

Assumption 4.3. Let

$$\max(m^{1/2q}, \min(n^{1/2q-1/pq}, m^{-1/q}n^{1/q})) \leqslant \kappa \leqslant n^{1/p}$$

as $n, m \to \infty$ such that $m/n \to 0$.

The time series (x_t) is now approximated by a κth order autoregression. The approximation order κ is assumed to increase with the sample size n at a controlled rate. Note that we may allow κ to increase slowly for larger values of q. If (v_t) is generated as an invertible ARMA process, then $q = \infty$. In this case, we may permit κ to grow at a logarithmic rate.

Bootstrap samples for the weakly integrated process (x_t) can be obtained similarly as before by first fitting (x_t) using an autoregression of order increasing with the sample size. For the fitted autoregression, we may use the linear specification

$$x_t = \hat{\alpha}_1 x_{t-1} + \cdots + \hat{\alpha}_\kappa x_{t-\kappa} + \hat{\varepsilon}_{\kappa,t} \tag{20}$$

or the nonlinear specification

$$x_t = \hat{\alpha} x_{t-1} + \sum_{i=1}^{\kappa-1} \hat{\beta}_i (x_{t-i} - \hat{\alpha} x_{t-i-1}) + \hat{\varepsilon}_{\kappa,t} \tag{21}$$

with order κ set to increase as the sample size. The fitted parameters in (20) and (21) are related each other by

$$\begin{aligned} &\hat{\alpha}_1 = \hat{\alpha} + \hat{\beta}_1, \\ &\hat{\alpha}_i = \hat{\beta}_i - \hat{\alpha}\hat{\beta}_{i-1}, \ i = 2, \ldots, \kappa - 1, \\ &\hat{\alpha}_\kappa = -\hat{\alpha}\hat{\beta}_{\kappa-1} \end{aligned} \tag{22}$$

and the fitted residuals are identical.

We may then obtain bootstrap samples (ε_t^*) from the centered fitted residuals

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

and generate the bootstrap samples (x_t^*) recursively using

$$x_t^* = \hat{\alpha}_1 x_{t-1}^* + \cdots + \hat{\alpha}_\kappa x_{t-\kappa}^* + \varepsilon_t^* \tag{23}$$

or

$$x_t^* = \hat{\alpha} x_{t-1}^* + \sum_{i=1}^{\kappa-1} \hat{\beta}_i (x_{t-i}^* - \hat{\alpha} x_{t-i-1}^*) + \varepsilon_t^* \tag{24}$$

correspondingly as the fitted models (20) and (21), given the intinial values $x_t^* = x_t$ for $t = 0, \ldots, -\kappa + 1$. The bootstrap samples (x_t^*) generated using the fitted models (20) and (21) are identical. The simple linear regression (20) may therefore be preferred to use in practise.

For the asymptotic analysis of the bootstrap sample (x_t^*) based on (20) or (21), it will be convenient to look at the fitted regressions

$$\left(x_t - \sum_{i=1}^{\kappa-1} \beta_i x_{t-i} \right) = \tilde{\alpha} \left(x_{t-1} - \sum_{i=1}^{\kappa-1} \beta_i x_{t-i-1} \right) + \tilde{\varepsilon}_{\kappa,t}, \tag{25}$$

$$(x_t - \alpha x_{t-1}) = \sum_{i=1}^{\kappa-1} \tilde{\beta}_i (x_{t-i} - \alpha x_{t-i-1}) + \tilde{\varepsilon}_{\kappa,t}. \tag{26}$$

Of course, these regressions are not feasible since α and (β_i) are unknown. They are introduced here simply to analyze the fitted regressions (20) and (21). From now on, we will concentrate on the estimation of the parameters in regression (21). The estimates of the parameters in regression (20) can be obtained using the relationships in (22). The next lemma shows that $\hat{\alpha}$ and $(\hat{\beta}_i)$ in regression (21) are asymptotically equivalent to $\tilde{\alpha}$ and $(\tilde{\beta}_i)$ in regressions (25) and (26).

Lemma 4.4. *Under Assumptions* 2.1, 4.1 *and* 4.3, *we have*

$$\frac{n}{\sqrt{m}}(\hat{\alpha} - \alpha) = \frac{n}{\sqrt{m}}(\tilde{\alpha} - \alpha) + O_p((m/n)^{1/2})$$

and for any κ *fixed*

$$\sqrt{n}(\hat{\beta}_i - \beta_i) = \sqrt{n}(\tilde{\beta}_i - \beta_i) + O_p((m/n)^{1/2})$$

uniformly in $i = 1, \ldots, \kappa - 1$ *and* $m \in \mathbf{R}_+$.

Lemma 4.5. *Under Assumptions* 2.1, 4.1 *and* 4.3, *we have*

$$\frac{n}{\sqrt{m}}(\tilde{\alpha} - \alpha) =_d \frac{\sqrt{m}\int_0^1 V_m(r)\,\mathrm{d}V_0(r)}{m\int_0^1 V_m^2(r)\,\mathrm{d}r} + O_p(n^{-1/2+1/p}) + O_p(mn^{-1})$$

as $n \to \infty$, *uniformly in* $m \in \mathbf{R}_+$.

The asymptotic properties of $(\tilde{\beta}_i)$ are well known. In particular, if we define $\tilde{\beta}_\kappa(z) = 1 - \sum_{i=1}^{\kappa-1} \tilde{\beta}_i z^i$, then we have

$$\begin{aligned}\tilde{\beta}_\kappa(1) &= \beta_\kappa(1) + O_p(\kappa n^{-1/2}) + o(\kappa^{-q}) \\ &= \beta(1) + O_p(\kappa n^{-1/2}) + o(\kappa^{-q})\end{aligned}$$

as shown in the proof of Lemma 3.1 in Park (2002). Due to the condition on κ in Assumption 4.3, and the result in Lemma 4.4, it therefore follows that

$$\hat{\beta}_\kappa(1) = \beta(1) + O_p(n^{-1/2+1/p}) + O_p(mn^{-1}) \tag{27}$$

where $\hat{\beta}_\kappa(z) = 1 - \sum_{i=1}^{\kappa-1} \hat{\beta}_i z^i$. The approximation error incurred by using $\hat{\beta}_\kappa(1)$ for $\beta(1)$ would thus become negligible within our error bound.

We now let

$$\hat{\pi}(1) = \frac{1}{\hat{\beta}_\kappa(1)}$$

and define

$$V_{mn}^*(r) =_{d^*} \frac{x_{[nr]}^*}{\hat{\pi}(1)\sqrt{n}}.$$

Then we have analogously as in Lemma 4.2 that

Theorem 4.6. *Under Assumptions* 2.1, 4.1 *and* 4.3, *we have*

$$V^*_{mn}(r) = V_m(r) + O^*_p(n^{-1/2+1/p}) + O^*_p(mn^{-1}) \quad \textit{in } \mathbf{P}$$

uniformly in $r \in [0,1]$ *and* $m \in \mathbf{R}_+$.

Our results in Lemma 4.2 and Theorem 4.6 make it clear that we may expect, also for the weakly integrated time series generated by general linear processes, the asymptotic refinements similar to those in the previous section established for the simple weakly integrated processes. We are only required to fit an approximated autoregression of order increasing with the sample size, and to obtain the bootstrap samples based on the fitted regression. The error bounds are exactly the same as for the first order autoregressive processes with weak unit roots.

4.2. Bootstrap refinements for general models

It follows rather straightforwardly from Lemma 4.2 and Theorem 4.6 that

Corollary 4.7. *Under Assumptions* 2.1, 4.1 *and* 4.3, *we have*

$$\frac{m}{n^2}\sum_{t=1}^{n} x_t^2 =_d m\int_0^1 V_m^2(r)\,\mathrm{d}r + O_p(m^{1/2}n^{-1/2+1/p}) + O_p(m^{3/2}n^{-1}),$$

$$\frac{m}{n^2}\sum_{t=1}^{n} x_t^{*2} =_{d^*} m\int_0^1 V_m^2(r)\,\mathrm{d}r + O^*_p(m^{1/2}n^{-1/2+1/p}) + O^*_p(m^{3/2}n^{-1}) \quad \textit{in } \mathbf{P}$$

as $n \to \infty$, *uniformly in* $m \in \mathbf{R}_+$, *which extends the results in Lemmas* 3.1 *and* 3.3 *to general weakly integrated time series driven by linear processes.*

One of the immediate implications of the results in Corollary 4.7 is the bootstrap refinement for the model

$$y_t = \beta x_t + u_t. \tag{28}$$

Let $\hat{\beta}$ be the least squares estimator for β. If the errors (u_t) are iid and independent of (x_t), we have

$$\begin{aligned}\frac{n}{\sqrt{m}}(\hat{\beta}-\beta) &= \left(\frac{m}{n^2}\sum_{t=1}^{n} x_t^2\right)^{-1}\frac{\sqrt{m}}{n}\sum_{t=1}^{n} x_t u_t,\\ &=_d \frac{\sqrt{m}\int_0^1 V_m(r)\,\mathrm{d}U(r)}{m\int_0^1 V_m^2(r)\,\mathrm{d}r} + O_p(m^{1/2}n^{-1/2+1/p}) + O_p(m^{3/2}n^{-1}),\end{aligned} \tag{29}$$

where U is Brownian motion independent of V_m. Similarly, we have in this case

$$\frac{n}{\sqrt{m}}(\hat{\beta}^* - \hat{\beta}) = \left(\frac{m}{n^2}\sum_{t=1}^{n} x_t^{*2}\right)^{-1} \frac{\sqrt{m}}{n}\sum_{t=1}^{n} x_t^* u_t^*$$
$$=_{d^*} \frac{\sqrt{m}\int_0^1 V_m(r)\,\mathrm{d}U(r)}{m\int_0^1 V_m^2(r)\,\mathrm{d}r} + O_p^*(m^{1/2}n^{-1/2+1/p}) + O_p^*(m^{3/2}n^{-1}) \quad \text{in } \mathbf{P}, \tag{30}$$

where (x_t^*) are bootstrap samples of (x_t) obtained as described in the previous subsection, and (u_t^*) are resamples of the centered fitted residuals $(\hat{u}_t - (1/n)\sum_{t=1}^{n}\hat{u}_t)$. Therefore, the bootstrap for regression (28) provides the refinement up to the order given by the maximum of $o(m^{1/2}n^{-1/2+1/p+\epsilon})$ and $o(m^{3/2}n^{-1+\epsilon})$ for arbitrarily small $\epsilon>0$, just as in the case of simple autoregression. The leading terms of the standardized $\hat{\beta}$ and $\hat{\beta}^*$ all have normal limit distributions, as shown in Park (2003b).

The model given above in (28) may represent a quite general *weak* cointegrating regression, if we specify (u_t) as a linear process jointly with (v_t) generating (x_t). To consider such a general model, we define

$$w_t = (u_t, v_t)'$$

and let (w_t) be a linear process given by

$$w_t = \Pi(L)\varepsilon_t,$$

where (ε_t) are now iid random vectors and $\Pi(z) = \sum_{i=0}^{\infty}\Pi_i z^i$. We impose conditions on (ε_t) and (Π_i) comparable to those in Assumption 4.1. Under this specification, the usual least squares method is not efficient. An efficient way of estimating β is to run the regression

$$y_t = \beta x_t + \sum_{|i|\leqslant\kappa}\beta_i v_{t-i} + u_{\kappa,t}. \tag{31}$$

Of course, (v_t) are not observed and should be replaced for the practical implementation by the fitted residuals $(\hat{v}_t)$ from the regression

$$x_t = \hat{\alpha}x_{t-1} + \hat{v}_t$$

and the number κ of leads and lags is assumed to satisfy Assumption 4.3.

To bootstrap the general weak cointegrating regression, we first let

$$\hat{z}_t = (\hat{u}_t, x_t)',$$

where $(\hat{u}_t)$ are the fitted residuals from regression (28), and fit

$$\hat{z}_t = \hat{A}_1\hat{z}_{t-1} + \cdots + \hat{A}_\kappa\hat{z}_{t-\kappa} + \hat{\varepsilon}_{\kappa,t}. \tag{32}$$

We may now obtain resamples (ε_t^*) from the centered residuals $(\hat{\varepsilon}_{\kappa,t})$ and construct the bootstrap samples (z_t^*) using the fitted vector autoregression in (32). The order κ needs not be the same for regressions (31) and (32). We just use the same symbol, since they are required to satisfy the same conditions.

Both the sample and bootstrap asymptotics for the general weak cointegrating regressions are quite similar to those for the simple weak cointegrating regressions driven by iid errors, if we fit the augmented regression (31) to estimate β. For the general weak cointegrating regressions, we indeed have exactly the same representations as in (29) and (30), only with reduced variance for the process U. We do not provide the details of the proofs here, because they are quite straightforward from our results in the previous section and the sample and bootstrap asymptotics developed in Chang et al. (2005) for the similar cointegrating regressions augmented with leads and lags of differenced regressors. The models with fitted mean and trend can be analyzed similarly. The finite order autoregressive model can also be considered as a special case within our framework. In particular, it is rather straightforward to show that the result by Inoue and Kilian (2002) continues to hold for weakly integrated processes. Undoubtedly, the bootstrap would provide refinements for more general models as well. It is indeed obvious from Park (2003b) that the bootstrap yields distributions closer to finite sample distributions compared to the first order asymptotics for nonlinear, as well as linear, regression models.

5. Conclusions

In this paper, we consider the bootstrap for weakly integrated processes with roots approaching to unity as the sample size increases at rates slower than n^{-1}. As shown in Park (2003b), models with such processes yield normal asymptotics, in sharp contrast to those with the exact unit roots or the roots converging to unity at rates equal to or faster than n^{-1}. For such models, the relevant asymptotic theories are generally nonstandard and nonnormal. We establish the bootstrap consistency and asymptotic refinement for models with weakly integrated processes. That is, it is shown that the usual bootstrap is not only first order equivalent to the asymptotics, but also yields the distributions that are closer to the finite sample distributions than the first order asymptotics if applied to the asymptotically pivotal statistics. It is well known that the bootstrap becomes inconsistent for models with exact unit roots, unless the unity of the root is imposed when we generate bootstrap samples.

We consider relatively simple models in the paper. This is, however, just for the concreteness of the arguments and by no means implies that the bootstrap works only for such simple models. The bootstrap theory presented here and the asymptotic theory developed in Park (2003b) indeed make it very clear that the bootstrap works for much more general models, including nonlinear and nonparametric models, with weakly integrated processes. The general conclusion drawn by our theory is also well expected to hold for more general models. Outside a certain proximity of the unit root, the bootstrap provides better approximations for finite sample distributions and therefore the bootstrap correction becomes more important, as the root approaches faster to unity. Inside an immediate neighborhood of the unit root, however, the bootstrap samples fail to mimic even the first order

asymptotics unless the exact information on the root is utilized when we generate the bootstrap samples.

Acknowledgements

I am grateful to Jean-Marie Dufour, Benoit Perron and two anonymous referees for their helpful comments. This paper was written for the presentation at the conference "Resampling Methods in Econometrics," held at University of Montreal, October 13–14, 2001. I wish to thank Jean Marie Dufour and Benoit Perron for their support and encouragement, and Yoosoon Chang for many constructive suggestions. This research was supported by Statistical Research Center for Complex Systems at Seoul National University.

Appendix A. Mathematical proofs

Let f be the density for $\mathbf{N}(0, 1/2)$, which is given by

$$f(x) = \frac{1}{\sqrt{\pi}} \mathrm{e}^{-x^2}.$$

As shown in Park (2003b), we have

Lemma A.1. *For $k \geqslant 0$, we have*

$$m^{k/2} \int_0^1 V_m^k(r)\,\mathrm{d}r \to_{\text{a.s.}} \sigma^k \int_{-\infty}^{\infty} x^k f(x)\,\mathrm{d}x,$$

$$m^{k/2} \int_0^1 r V_m^k(r)\,\mathrm{d}r \to_{\text{a.s.}} \frac{1}{2}\,\sigma^k \int_{-\infty}^{\infty} x^k f(x)\,\mathrm{d}x$$

as $m \to \infty$.

Proof of Lemma 2.3. The stated result follows as a special case of Lemma 2.3 in Park (2003b). The proof of Lemma 2.3 in Park (2003b), however, does not show how we may establish the corresponding result for the bootstrap samples. Here we give a more detailed proof to motivate the bootstrap version of the result given in Lemma 2.4 below.

It follows as in the proof of Lemma 2.3 in Park (2003b) that

$$V_{mn}(r) = V_{0n}(r) - m \int_0^r \exp(-m(r-s)) V_{0n}(s)\,\mathrm{d}s + R_{mn}(r),$$

where the remainder term R_{mn} is bounded by a constant multiple of

$$\frac{m}{n} \left| m \int_0^r \exp(-m(r-s)) V_{0n}(s)\,\mathrm{d}s \right| \tag{33}$$

uniformly in $r \in [0,1]$. However, due to Sakhanenko (1980), we may choose V_{0n} up to the distributional equivalence such that it is defined in the same probability space as the Brownian motion V_0 and

$$\mathbf{P}\left\{\sup_{0\leqslant r\leqslant 1}|V_{0n}(r)-V_0(r)|>c_n\right\}\leqslant Kc_n^{-p}n^{1-p/2}\mathbf{E}|\varepsilon_t|^p, \tag{34}$$

where K is an absolute constant depending only upon p. In particular, we have for $c_n = n^{-1/2+1/p}M$ with some large $M>0$

$$\mathbf{P}\left\{\sup_{0\leqslant r\leqslant 1}|V_{0n}(r)-V_0(r)|>n^{-1/2+1/p}M\right\}<\epsilon, \tag{35}$$

where $\epsilon>0$ is arbitrary. The stated result therefore follows directly from (33) and (35). □

Proof of Lemma 2.4. We first apply the result by Sakhanenko (1980) in (34) to V_{0n}^* to deduce

$$\mathbf{P}^*\left\{\sup_{0\leqslant r\leqslant 1}|V_{0n}^*(r)-V_0(r)|>c_n\right\}\leqslant Kc_n^{-p}n^{1-p/2}\mathbf{E}^*|\varepsilon_t^*|^p. \tag{36}$$

Then we show

$$\mathbf{E}^*|\varepsilon_t^*|^p=\frac{1}{n}\sum_{t=1}^n\left|\hat{\varepsilon}_t-\frac{1}{n}\sum_{t=1}^n\hat{\varepsilon}_t\right|^p=O_p(1).$$

The stated result would then follow immediately, analogously as in the proof of Lemma 2.3. Note that

$$\hat{\alpha}=\alpha+O_p(m^{1/2}n^{-1})=1+O_p(mn^{-1})$$

as we show in Theorem 3.2.

We write

$$\frac{1}{n}\sum_{t=1}^n\left|\hat{\varepsilon}_t-\frac{1}{n}\sum_{t=1}^n\hat{\varepsilon}_t\right|^p\leqslant c_p(A_n+B_n+C_n),$$

where c_p is a constant depending only on p, and

$$A_n=\frac{1}{n}\sum_{t=1}^n|\varepsilon_t|^p,\quad B_n=\frac{1}{n}\sum_{t=1}^n|\hat{\varepsilon}_t-\varepsilon_t|^p,\quad C_n=\left|\frac{1}{n}\sum_{t=1}^n\hat{\varepsilon}_t\right|^p.$$

It will be shown below that $A_n, B_n, C_n = O_p(1)$.

By the strong law of large numbers, $A_n\to_{\text{a.s.}}\mathbf{E}|\varepsilon_t|^p=O(1)$. To show that $B_n=O_p(1)$, we note that

$$\frac{1}{n}\sum_{t=1}^n|\hat{\varepsilon}_t-\varepsilon_t|^p=\left(\frac{m}{n}\right)^{p/2}\left(\frac{m}{n^2}\sum_{t=1}^n x_{t-1}^2\right)^{-p}\left|\frac{\sqrt{m}}{n}\sum_{t=1}^n x_{t-1}\varepsilon_t\right|^p\left(\frac{1}{n^{1+p/2}}\sum_{t=1}^n|x_{t-1}|^p\right).$$

As shown in Lemma 3.1 and Theorem 3.2, we have

$$\left(\frac{m}{n^2}\sum_{t=1}^{n} x_{t-1}^2\right)^{-1}, \frac{\sqrt{m}}{n}\sum_{t=1}^{n} x_{t-1}\varepsilon_t = O_p(1).$$

Moreover, we have

$$\frac{1}{n^{1+p/2}}\sum_{t=1}^{n} |x_{t-1}|^p =_d m^{p/2}\int_0^1 |V_{mn}(r)|^p \,\mathrm{d}r = O_p(1)$$

and it follows that $B_n = O_p((m/n)^{p/2})$. Finally,

$$\frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t = \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t - \frac{1}{\sqrt{n}}\left(\frac{m}{n^2}\sum_{t=1}^{n} x_{t-1}^2\right)^{-1}\left(\frac{\sqrt{m}}{n}\sum_{t=1}^{n} x_{t-1}\varepsilon_t\right)\left(\frac{\sqrt{m}}{n^{3/2}}\sum_{t=1}^{n} x_{t-1}\right)$$

from which we may easily deduce that $C_n = O_p(n^{-1/2})$. The proof is therefore complete. □

Proof of Corollary 2.5. The stated results can easily be deduced from (34) and (36) with

$$c_n = m^{-\kappa}\delta_{mn}.$$

Note that

$$(m^{-\kappa}\delta_{mn})^{-p} n^{1-p/2} = \delta_{mn}$$

yields

$$\delta_{mn} = m^{\kappa p/(1+p)} n^{(2-p)/2(1+p)}$$

as was to be shown. □

Proof of Lemma 3.1. Write

$$\frac{m}{n^2}\sum_{t=1}^{n} x_{t-1}^2 =_d m\int_0^1 V_{mn}^2(r)\,\mathrm{d}r.$$

We have

$$\begin{aligned}\int_0^1 V_{mn}^2(r)\,\mathrm{d}r - \int_0^1 V_m^2(r)\,\mathrm{d}r &= \int_0^1 (V_{mn}(r) - V_m(r))^2\,\mathrm{d}r + 2\int_0^1 V_m(r)(V_{mn}(r) - V_m(r))\,\mathrm{d}r\end{aligned}$$

and, by Cauchy–Schwarz inequality,

$$\left|\int_0^1 V_m(r)(V_{mn}(r) - V_m(r))\,\mathrm{d}r\right| \leqslant \left(\int_0^1 V_m^2(r)\,\mathrm{d}r\right)^{1/2}\left(\int_0^1 (V_{mn}(r) - V_m(r))^2\,\mathrm{d}r\right)^{1/2}.$$

The first result now follows directly from Lemma 2.3. Note that

$$\int_0^1 V_m^2(r)\,\mathrm{d}r = O(m^{-1}) \quad \text{a.s.}$$

due to Lemma A.1.

To derive the second result, we first construct V_{mn} through the Skorohod type embedding, i.e., the embedding of the partial sum of (ε_t) directly into a Brownian motion with properly chosen stopping times, as in Park (2003a). Then it follows that

$$\frac{\sqrt{m}}{n}\sum_{t=1}^{n} x_{t-1}\varepsilon_t =_d \sqrt{m}\int_0^1 V_{mn}(r)\,\mathrm{d}V_0(r) + O_p((m/n)^{1/2}).$$

Now we note that

$$\int_0^1 V_{mn}(r)\,\mathrm{d}V_0(r) - \int_0^1 V_m(r)\,\mathrm{d}V_0(r) = \int_0^1 (V_{mn}(r) - V_m(r))\,\mathrm{d}V_0(r)$$

has quadratic variation

$$\sigma^2\int_0^1 (V_{mn}(r) - V_m(r))^2\,\mathrm{d}r = O_p(n^{-1+2/p})$$

whose order is given by Lemma 2.3. The proof is therefore complete. □

Proof of Theorem 3.2. It follows from Lemma A.1 that

$$m\int_0^1 V_m^2(r)\,\mathrm{d}r \to_{\text{a.s.}} \sigma^2\int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \frac{\sigma^2}{2}.$$

Moreover, if we define a continuous martingale

$$M_m(r) = \sqrt{m}\int_0^r V_m(s)\,\mathrm{d}V_0(s)$$

it follows immediately that

$$[M_m](1) = m\int_0^1 V_m^2(r)\,\mathrm{d}r \to_{\text{a.s.}} \frac{\sigma^2}{2}$$

as $m \to \infty$, and that

$$[M_m, V_0](1) = \sqrt{m}\int_0^1 V_m(r)\,\mathrm{d}r \to_{\text{a.s.}} \sigma\int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = 0$$

as $m \to \infty$, due to Lemma A.1. We have thus shown that M_m is a continuous martingale such that it is asymptotically independent of V_0 and hence of V_m for all m, and it has the quadratic variation which converges to $1/2$. Therefore, we may now deduce that

$$M_m(1) \to_d \mathbf{N}\left(0, \frac{\sigma^2}{2}\right)$$

as $m \to \infty$. Finally, note that

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^2 - \frac{1}{n}\left(\frac{m}{n^2}\sum_{t=1}^{n} x_{t-1}^2\right)^{-1}\left(\frac{\sqrt{m}}{n}\sum_{t=1}^{n} x_{t-1}\varepsilon_t\right)^2 \\ &= \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^2 + O_p(n^{-1}) \\ &= \sigma^2 + O_p(n^{-1/2})\end{aligned}$$

which completes the proof. □

Proof of Lemma 3.3. The first two results are straightforward from (35). The third result can also be easily deduced from Lemma 2.3, since

$$\frac{\sqrt{m}}{n^{3/2}}\sum_{t=1}^{n} x_{t-1} =_d \sqrt{m}\int_0^1 V_{mn}(r)\,\mathrm{d}r.$$

Finally, we have

$$\frac{\sqrt{m}}{n^{5/2}}\sum_{t=1}^{n} t x_{t-1} =_d \sqrt{m}\int_0^1 r V_{mn}(r)\,\mathrm{d}r + R_{mn},$$

where the remainder term R_{mn} is bounded by

$$\sqrt{m}\int_0^1 \left|r - \frac{[nr]}{n}\right| |V_{mn}(r)|\,\mathrm{d}r \leqslant \frac{1}{n}\left(\sqrt{m}\int_0^1 |V_{mn}(r)|\,\mathrm{d}r\right) = O_p(n^{-1})$$

and the fourth result follows immediately from Lemma 2.3. □

Proof of Corollary 3.4. Given Lemma 3.3, the stated results follow exactly as in Theorem 3.2. The proof is therefore omitted. □

Proof of Lemma 3.5. Given Lemma 2.4, the proof is entirely analogous to that of Lemma 3.1. The proof is therefore omitted. □

Proof of Lemma 3.6. Given Lemma 2.4, the proof is entirely analogous to that of Lemma 3.3. The proof is therefore omitted. □

Proof of Theorem 3.7. Given Lemma 3.5, the stated results follow exactly as in Theorem 3.2. The proof is therefore omitted. □

Proof of Corollary 3.8. Given Lemma 3.6, the stated results follow exactly as in Corollary 3.4. The proof is therefore omitted. □

Proof of Lemma 4.2. The stated result follows exactly as in the proof of Lemma 2.3, only with the inequality (34) due to Sakhanenko (1980) replaced by the corresponding inequality extended by Akonom (1993) for linear processes. □

Proof of Lemma 4.4. For given κ, we consider the model

$$x_t = \alpha x_{t-1} + (x_{t-1} - \alpha x_{t-2})(\kappa)'\beta(\kappa) + \varepsilon_{\kappa,t},$$

where

$$\beta(\kappa) = (\beta_1, \ldots, \beta_{\kappa-1})',$$

$$(x_{t-1} - \alpha x_{t-2})(\kappa) = (x_{t-1} - \alpha x_{t-2}, \ldots, x_{t-\kappa+1} - \alpha x_{t-\kappa})'.$$

In the subsequent proof, we regard $\beta(\kappa)$ and $(x_{t-1} - \alpha x_{t-2})(\kappa)$ as scalars, and simply denote by β and $x_{t-1} - \alpha x_{t-2}$, respectively. This is purely for expositional brevity. The proof for the vector-valued $\beta(\kappa)$ and $(x_{t-1} - \alpha x_{t-2})(\kappa)$ are essentially identical, requiring only some obvious changes in notation.

Define

$$Q(\alpha, \beta) = \sum_{t=1}^{n} (x_t - (\alpha + \beta)x_{t-1} + \alpha\beta x_{t-2})^2$$

and denote by $\dot{Q}(\alpha, \beta)$ and $\ddot{Q}(\alpha, \beta)$ the first and second derivatives of $Q(\alpha, \beta)$. Also, we let

$$v_n = \operatorname{diag}\left(\frac{n}{\sqrt{m}}, \sqrt{n}\right)$$

and consider

$$0 = v_n^{-1}\dot{Q}(\alpha, \beta) + (v_n^{-1}\ddot{Q}(\bar{\alpha}, \bar{\beta})v_n^{-1})v_n(\hat{\alpha} - \alpha, \hat{\beta} - \beta)',$$

where $(\bar{\alpha}, \bar{\beta})$ lies on the line segment connecting $(\hat{\alpha}, \hat{\beta})$ and (α, β).

For the subsequent proof, we will use the fact

$$\sum_{t=1}^{n} x_{t-i}^2, \ \sum_{t=1}^{n} x_{t-i}x_{t-j} = O_p(n^2/m), \tag{37}$$

$$\sum_{t=1}^{n} x_{t-i}v_{t-j} = O_p(n), \tag{38}$$

$$\sum_{t=1}^{n} x_{t-i}\varepsilon_{\kappa,t} = O_p(m^{-1/2}n), \tag{39}$$

which hold uniformly for $i, j = 1, \ldots, \kappa - 1$. The result in (37) follows immediately from Lemma 4.2 as in the proof of Theorem 3.1. To deduce the result in (38), see the Proof of Lemma 4.1 in Park (2002) and the proof of Lemma 3.2 in Chang and Park (2002). To prove the result in (39), we first write

$$\sum_{t=1}^{n} x_{t-i}\varepsilon_{\kappa,t} = \sum_{t=1}^{n} x_{t-i}\varepsilon_t + \sum_{t=1}^{n} x_{t-i}(\varepsilon_{\kappa,t} - \varepsilon_t)$$

and write

$$x_{t-i} = \sum_{j=1}^{t-i} v_j - \frac{1-\alpha}{\alpha} \sum_{j=1}^{t-i-1} \alpha^{t-i-j}\left(\sum_{k=1}^{j} v_k\right).$$

Following the proof of Lemma 3.1 in Chang and Park (2002), we may deduce

$$\sum_{t=1}^{n}\left(\sum_{j=1}^{t-i} v_j\right)(\varepsilon_{\kappa,t}-\varepsilon_t)=o_p(\kappa^{-q}n),$$

$$\sum_{t=1}^{n}\left(\sum_{k=1}^{j} v_k\right)(\varepsilon_{\kappa,t}-\varepsilon_t)=o_p(\kappa^{-q}n),$$

uniformly in i and j. We therefore have

$$\sum_{t=1}^{n} x_{t-i}(\varepsilon_{\kappa,t}-\varepsilon_t)=o_p(\kappa^{-q}n),$$

uniformly in i. The result in (39) now follows immediately upon noticing

$$n\kappa^{-q}=O(m^{-1/2}n),$$

since $\kappa\geqslant m^{1/2q}$, and

$$\sum_{t=1}^{n} x_{t-i}\varepsilon_t=O_p(m^{-1/2}n),$$

which holds uniformly in $i\geqslant 1$.

It is easy to check that the conditions AD1–AD4 in Park and Phillips (2001) hold. Moreover, with the choice of $\mu_n=v_n^{1-\delta}$ for small $\delta>0$, the condition AD7 in Park and Phillips (2001) is satisfied. We now let

$$\bar{\alpha}=\alpha+\frac{\sqrt{m}}{n},\quad \bar{\beta}=\beta+\frac{1}{\sqrt{n}}$$

and define

$$\ddot{Q}(\bar{\alpha},\bar{\beta})-\ddot{Q}(\alpha,\beta)=\begin{pmatrix}\Delta_{11} & \Delta_{12}\\ \Delta_{21} & \Delta_{22}\end{pmatrix},$$

where

$$\Delta_{11}=\frac{1}{n}\sum_{t=1}^{n}x_{t-2}^2-\frac{2}{\sqrt{n}}\sum_{t=1}^{n}(x_{t-1}-\beta x_{t-2})x_{t-2}=O_p(n^{3/2}/m),$$

$$\Delta_{22}=\frac{m}{n^2}\sum_{t=1}^{n}x_{t-2}^2-\frac{2\sqrt{m}}{n}\sum_{t=1}^{n}(x_{t-1}-\alpha x_{t-2})x_{t-2}=O_p(m^{1/2}),$$

$$\begin{aligned}\Delta_{12}=&-\frac{\sqrt{m}}{n}x_{t-2}(x_{t-1}-\bar{\beta}x_{t-2})-\frac{1}{\sqrt{n}}(x_{t-1}-\alpha x_{t-2})x_{t-2}\\&-\frac{\sqrt{m}}{n}(x_{t-1}-\bar{\beta}x_{t-2})x_{t-2}-\frac{1}{\sqrt{n}}(x_{t-1}-\alpha x_{t-2})x_{t-2}\\=&\,O_p(n/m^{1/2}),\end{aligned}$$

which follow from (37) and (38). Consequently, we have

$$v_n^{-1}(\ddot{Q}(\bar{\alpha},\bar{\beta}) - \ddot{Q}(\alpha,\beta))v_n^{-1} = O_p(n^{-1/2}).$$

Moreover, if we denote by $\ddot{Q}_0(\alpha,\beta)$ the diagonal matrix whose diagonal entries are the same as those of $\ddot{Q}(\alpha,\beta)$, then

$$v_n^{-1}\ddot{Q}(\alpha,\beta)v_n^{-1} = v_n^{-1}\ddot{Q}_0(\alpha,\beta)v_n^{-1} + O_p((m/n)^{1/2}).$$

Note that the off-diagonal entry of $v_n^{-1}\ddot{Q}(\alpha,\beta)v_n^{-1}$ is given by

$$\begin{aligned}
&\frac{\sqrt{m}}{n}\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{n} x_{t-2}\varepsilon_{\kappa,t} + \sum_{t=1}^{n}(x_{t-1}-\alpha x_{t-2})(x_{t-1}-\beta x_{t-2})\right)\\
&\quad = \frac{\sqrt{m}}{n}\frac{1}{\sqrt{n}}\,O_p(m^{-1/2}n) + \frac{\sqrt{m}}{n}\frac{1}{\sqrt{n}}\,O_p(n)\\
&\quad = O_p((m/n)^{1/2})
\end{aligned}$$

as was to be shown.

We now have

$$v_n(\hat{\alpha}-\alpha,\hat{\beta}-\beta)' = -(v_n^{-1}\ddot{Q}_0(\alpha,\beta)v_n^{-1})^{-1}(v_n^{-1}\dot{Q}(\alpha,\beta)) + O_p((m/n)^{1/2}).$$

However, as can be easily shown,

$$v_n(\tilde{\alpha}-\alpha,\tilde{\beta}-\beta)' = -(v_n^{-1}\ddot{Q}_0(\alpha,\beta)v_n^{-1})^{-1}(v_n^{-1}\dot{Q}(\alpha,\beta))$$

and the proof is complete. □

Proof of Lemma 4.5. It follows from the proof of Lemma 3.2 in Park (2002) that

$$\sup_{0\leqslant r\leqslant 1}\left|\sum_{t=1}^{[nr]}(\varepsilon_{\kappa,t}-\varepsilon_t)\right| = O_p(\kappa^{-q}n^{1/2}).$$

We therefore have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_{\kappa,t} = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t + O_p(\kappa^{-q})$$

uniformly in $r\in[0,1]$. Consequently, if we define

$$y_t^{\kappa} = \beta_{\kappa}(L)x_t$$

and

$$V_{mn}^{\kappa}(r) = n^{-1/2}y_{[nr]}^{\kappa}$$

then it follows that

$$V_{mn}^{\kappa}(r) = V_m(r) + O_p(n^{-1/2+1/p}) + O_p(mn^{-1}) + O_p(\kappa^{-q}).$$

The stated result may now be derived easily. Note that

$$\kappa^{-q}\leqslant \max(n^{-1/2+1/p}, mn^{-1})$$

under the condition in Assumption 4.3, and the $O_p(\kappa^{-q})$ error term is of order smaller than either the $O_p(n^{-1/2+1/p})$ term or the $O_p(mn^{-1})$ term. □

Proof of Theorem 4.6. The fitted residual $(\hat{\varepsilon}_{\kappa,t})$ from regression (20) or (21) can also be obtained from the fitted regression

$$x_t = \bar{\alpha}x_{t-1} + \sum_{i=1}^{\kappa-1} \bar{\beta}_i(x_{t-i} - \alpha x_{t-i-1}) + \hat{\varepsilon}_{\kappa,t}, \tag{40}$$

which is just a reparametrized version of (20). Define

$$v_{\kappa,t} = (v_{t-1}, \ldots, v_{t-\kappa+1})'$$

and denote by $\bar{\beta} = (\bar{\beta}_1, \ldots, \bar{\beta}_{\kappa-1})'$ and $\tilde{\beta} = (\tilde{\beta}_i, \ldots, \tilde{\beta}_{\kappa-1})'$ the least squares estimates of $\beta = (\beta_1, \ldots, \beta_{\kappa-1})'$ from regressions (26) and (40), respectively. It follows that

$$\begin{aligned}\bar{\alpha} - \alpha = &\left(\sum_{t=1}^{n} x_{t-1}^2 - \left(\sum_{t=1}^{n} x_{t-1}v_{\kappa,t}'\right)\left(\sum_{t=1}^{n} v_{\kappa,t}v_{\kappa,t}'\right)^{-1}\left(\sum_{t=1}^{n} v_{\kappa,t}x_{t-1}\right)\right)^{-1} \\ &\times\left(\sum_{t=1}^{n} x_{t-1}\varepsilon_{\kappa,t} - \left(\sum_{t=1}^{n} x_{t-1}v_{\kappa,t}\right)(\tilde{\beta} - \beta)\right)\end{aligned}$$

and

$$\bar{\beta} = \tilde{\beta} - (\bar{\alpha} - \alpha)\left(\sum_{t=1}^{n} v_{\kappa,t}v_{\kappa,t}'\right)^{-1}\left(\sum_{t=1}^{n} v_{\kappa,t}x_{t-1}\right).$$

Moreover, we have

$$\begin{aligned}\hat{\varepsilon}_{\kappa,t} &= x_t - \bar{\alpha}x_{t-1} - v_{\kappa,t}'\bar{\beta} \\ &= ((x_t - \alpha x_{t-1}) - v_{\kappa,t}'\tilde{\beta}) - (\bar{\alpha} - \alpha)x_{t-1} \\ &\quad + (\bar{\alpha} - \alpha)\left(\sum_{t=1}^{n} x_{t-1}v_{\kappa,t}'\right)\left(\sum_{t=1}^{n} v_{\kappa,t}v_{\kappa,t}'\right)^{-1} v_{\kappa,t} \\ &= \tilde{\varepsilon}_{\kappa,t} - (\bar{\alpha} - \alpha)x_{t-1} + (\bar{\alpha} - \alpha)\left(\sum_{t=1}^{n} x_{t-1}v_{\kappa,t}'\right)\left(\sum_{t=1}^{n} v_{\kappa,t}v_{\kappa,t}'\right)^{-1} v_{\kappa,t},\end{aligned}$$

where $(\tilde{\varepsilon}_{\kappa,t})$ are the fitted residuals from regression (26).

Let $\|\cdot\|$ denote the usual Euclidean norm if applied to a vector, and the standard operator norm if applied to a matrix. We have

$$\left\|\left(\sum_{t=1}^{n} v_{\kappa,t}v_{\kappa,t}'\right)^{-1}\right\| = O_p(n^{-1}), \tag{41}$$

$$\left\|\sum_{t=1}^{n} v_{\kappa,t}x_{t-1}\right\| = O_p(\kappa^{1/2}n), \tag{42}$$

which follows from Lemma 3.2 of Chang and Park (2002) and our earlier result in (38). Moreover, we have

$$\|\tilde{\beta} - \beta\| = O_p(\kappa^{1/2} n^{-1/2}) + o(\kappa^{-q}), \tag{43}$$

which follows as in the proof of Lemma 3.5 in Chang and Park (2002a). We may have the corresponding 'in probability' result from e.g., Proposition 3.1 in Shibata (1980). As he himself noted, the normality assumption is not necessary there as long as sufficient moments exist.

It follows from (37), (39), (41)–(43) that

$$\begin{aligned}\tilde{\alpha} - \alpha &= O_p(m^{1/2} n^{-1}) + O_p(\kappa m n^{-3/2}) + o_p(\kappa^{1/2-q} m n^{-1}) \quad \text{if } \kappa \leqslant n/m \\ &= O_p(\kappa^{-1} m^{-1/2}) + O_p(n^{-1/2}) + o_p(\kappa^{-1/2-q}) \quad \text{if } \kappa \geqslant n/m.\end{aligned}$$

We may now easily deduce that

$$\tilde{\alpha} - \alpha = O_p(\min((m/n)^{1/2}, \kappa^{-1})) \tag{44}$$

For $\kappa \leqslant n/m$, we have

$$m^{1/2} n^{-1},\ \kappa m n^{-3/2},\ \kappa^{1/2-q} m n^{-1} = O(m n^{-1}),$$

since $m \to \infty, \kappa \leqslant \sqrt{n}$ and $\kappa^{1/2-q} \to 0$ with $q \geqslant 1$, respectively, and

$$m/n = O((m/n)^{1/2}), \quad O(\kappa^{-1})$$

in this case. On the other hand, if $\kappa \geqslant n/m$,

$$\kappa^{-1} m^{-1/2},\ n^{-1/2},\ \kappa^{-1/2-q} = O(\kappa^{-1}),$$

since $m \to \infty, \kappa \leqslant \sqrt{n}$ and $q \geqslant 1$, respectively, and

$$\kappa^{-1} m^{-1/2},\ n^{-1/2},\ \kappa^{-1/2-q} = O((m/n)^{1/2}),$$

since $\kappa^{-1} \leqslant m/n$, $m \to \infty$, $\kappa^{-1} \leqslant m/n$ and $q \geqslant 1$, respectively.

We have

$$\sum_{t=1}^{n} |x_{t-1}|^p = O_p(m^{-p/2} n^{1+p/2}) \tag{45}$$

for $p \geqslant 0$. Moreover, since

$$\left(\sum_{t=1}^{n} (u_{t-1}^2 + \cdots + u_{t-\kappa+1}^2)^{p/2} \right)^{2/p} \leqslant \sum_{i=1}^{\kappa-1} \left(\sum_{t=1}^{n} |u_{t-i}|^p \right)^{2/p} = O_p(\kappa n^{2/p})$$

by Minkowski's inequality, we also have

$$\sum_{t=1}^{n} \left\| v_{\kappa,t} - \frac{1}{n} \sum_{t=1}^{n} v_{\kappa,t} \right\|^p = O_p(\kappa^{p/2} n) \tag{46}$$

for $p \geqslant 2$.

Now write

$$\mathbf{E}^*|\varepsilon_t^*|^p = \frac{1}{n}\sum_{t=1}^{n}\left|\hat{\varepsilon}_{\kappa,t} - \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_{\kappa,t}\right|^p \leqslant c_p(A_n + B_n + C_n),$$

where c_p is a constant depending only upon p, and

$$A_n = \frac{1}{n}\sum_{t=1}^{n}\left|\tilde{\varepsilon}_{\kappa,t} - \frac{1}{n}\sum_{t=1}^{n}\tilde{\varepsilon}_{\kappa,t}\right|^p,$$

$$B_n = |\bar{\alpha}-\alpha|^p\left(\frac{1}{n}\sum_{t=1}^{n}|x_{t-1}|^p + \left|\frac{1}{n}\sum_{t=1}^{n}x_{t-1}\right|^p\right),$$

$$C_n = |\bar{\alpha}-\alpha|^p\left\|\left(\sum_{t=1}^{n}v_{\kappa,t}v_{\kappa,t}'\right)^{-1}\right\|^p\left\|\sum_{t=1}^{n}v_{\kappa,t}x_{t-1}\right\|^p\left(\frac{1}{n}\sum_{t=1}^{n}\left\|v_{\kappa,t} - \frac{1}{n}\sum_{t=1}^{n}v_{\kappa,t}\right\|^p\right).$$

It is shown by Park (2002) that $A_n = O_p(1)$ in the proof of Lemma 3.2. Also, we have from (44) and (45) that $B_n = O_p(1)$. Finally, it follows from (44) and (46), together with (41) and (42) that $C_n = O_p(1)$. We therefore have

$$\mathbf{E}^*|\varepsilon_t^*|^p = O_p(1)$$

and the stated result now follows as in the proof of Lemma 2.4. □

Proof of Corollary 4.7. The stated result follows immediately from Lemma 4.2 and Theorem 4.6, exactly as in the proofs of Lemmas 3.1 and 3.3. The details are therefore omitted. □

References

Akonom, J., 1993. Comportement asymptotique du temps d'occupation du processus des sommes partielles. Annales de l'Institut Henri Poincaré 29, 57–81.

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

Bose, A., 1988. Edgeworth correction by bootstrap in autoregression. Annals of Statistics 16, 1709–1722.

Bühlmann, P., 1997. Sieve bootstrap for time series. Bernoulli 3, 123–148.

Chang, Y., Park, J.Y., 2002. On the asymptotics of ADF tests for unit roots. Econometric Reviews 21, 431–448.

Chang, Y., Park, J.Y., 2003. A sieve bootstrap for the test of a unit root. Journal of Time Series Analysis 24, 379–400.

Chang, Y., Park, J.Y., Song, K., 2005. Bootstrapping cointegrating regressions. Journal of Econometrics, this issue, doi:10.1016/j.jeconom.2005.06.011.

Inoue, A., Kilian, L., 2002. Bootstrapping autoregressive processes with possible unit roots. Econometrica 70, 377–391.

Kreiss, J.P., 1992. Bootstrap procedures for AR(∞)-processes. In: Jockel, K.H., Rothe, G., Sender, W. (Eds.), Bootstrapping and Related Techniques, Lecture Notes in Economics and Mathematical Systems, vol. 376. Springer, Heidelberg.

Nankervis, J.C., Savin, N.E., 1996. The level and power of the bootstrap t test in the AR (1) model with trend. Journal of Business and Economic Statistics 14, 161–168.

Park, J.Y., 2002. An invariance principle for sieve bootstrap in time series. Econometric Theory 18, 469–490.

Park, J.Y., 2003a. Bootstrap unit root tests. Econometrica 71, 1845–1895.

Park, J.Y., 2003b. Weak unit roots, Mimeographed, Department of Economics, Rice University.

Park, J.Y., Phillips, P.C.B., 2001. Nonlinear regressions with integrated time series. Econometrica 69, 1452–1498.

Sakhanenko, A.I., 1980. On unimprovable estimates of the rate of convergence in invariance principle. In: Colloquia Mathematica Societatis János Bolyai, vol. 32, Nonparametric Statistical Inference, Budapest, Hungary, pp. 779–783.

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.