Consistency of the Stationary Bootstrap under Weak Moment Conditions

Sílvia Gonçalves
CIREQ, CIRANO and Département de sciences économiques, Université de Montréal

Robert de Jong
Michigan State University

May 29, 2003

Abstract

We prove the first order asymptotic validity of the stationary bootstrap of Politis and Romano (1994) under the existence of only slightly more than second moments. Our results improve upon previous results in the literature, which assumed finite sixth moments.

Keywords: Stationary bootstrap, second moments.

JEL codes: C15, C22.

1. Introduction

The consistency of the stationary bootstrap (SB) of Politis and Romano (1994) under general dependence conditions has recently been established by Gonçalves and White (GW) (2002). Nevertheless, current results require the restrictive assumption of finite sixth moments. In this note, we establish the first-order asymptotic validity of the SB under the existence of only slightly more than second moments. As in GW (2002), the data are assumed to satisfy a NED condition, which includes the more restrictive mixing assumption as a special case. We consider bootstrap variance estimation as well as bootstrap distribution estimation. Our results broaden considerably the scope for application of the SB in economics and finance, where the existence of higher order moments is often a concern.

Although we focus on the SB, similar results hold for the moving blocks bootstrap (MBB) of Künsch (1989) and Liu and Singh (1992). Lahiri (1999) shows that the MBB variance estimator has a smaller asymptotic mean squared error than the SB variance estimator, which may favour the use of the MBB in applications. However, as remarked by Politis and Romano (1994), the SB may be less sensitive to the choice of the block length than the MBB.

The consistency of the MBB variance estimator under weak moment conditions follows by an application of De Jong and Davidson's (2000) results for kernel variance estimators, given that the MBB variance estimator for the sample mean is equal to a Bartlett kernel variance estimator, up to terms of order $O_P\left(\ell^2/n\right)$, with ℓ the block length. These terms vanish in probability if $\frac{\ell^2}{n} \to 0$. In this case, the consistency of the MBB variance estimator holds even under minimal dependence conditions (cf. De Jong and Davidson, 2000). The consistency of the MBB distribution estimator follows as a corollary. To save space and because the SB is harder to analyse, here we only provide formal results for the SB.

2. Main Results

The SB variance estimator is given by

$$\hat{\sigma}_n^2 = n^{-1}\sum_{t=1}^{n}\left(X_{nt}-\bar{X}_n\right)^2 + 2\sum_{\tau=1}^{n-1} b_n(\tau)\, n^{-1}\sum_{t=1}^{n-\tau}\left(X_{nt}-\bar{X}_n\right)\left(X_{n,t+\tau}-\bar{X}_n\right),$$

where $b_n(\tau) = \left(1-\frac{\tau}{n}\right)(1-p_n)^{\tau} + \frac{\tau}{n}(1-p_n)^{n-\tau}$, with p_n such that $p_n \to 0$ and $np_n^2 \to \infty$.

We assume $\{X_{nt}, n, t = 1, 2, ...\}$ is L_2-near epoch dependent (L_2-NED) on a mixing process $\{V_t\}$, i.e. $\|X_{nt}\|_2 < \infty$ and $\nu_k \equiv \sup_{n,t}\left\|X_{nt} - E_{t-k}^{t+k}(X_{nt})\right\|_2 \to 0$ as $k \to \infty$. Here, $\|X\|_q = \left(E|X|^q\right)^{1/q}$, $q \geq 1$, and $E_{t-k}^{t+k}(\cdot) \equiv E\left(\cdot|\mathcal{F}_{t-k}^{t+k}\right)$, where $\mathcal{F}_{t-k}^{t+k} \equiv \sigma\left(V_{t-k}, \ldots, V_{t+k}\right)$ is the σ-field generated by $V_{t-k}, \ldots, V_{t+k}$. If $\nu_k = O\left(k^{-a-\varepsilon}\right)$ for some $\varepsilon > 0$ we say $\{X_{nt}\}$ is L_2-NED (on $\{V_t\}$) of size $-a$. $\{V_t\}$ is assumed to be strong mixing, with α-mixing coefficients α_k defined as usual such that $\alpha_k \to 0$ as $k \to \infty$. We make the following assumption:

Assumption 1. a) For some $r > 2$ and some $\delta > 0$, $\|X_{nt}\|_{r+\delta} \leq \Delta < \infty$ for all $n, t = 1, 2, \ldots$; b) $\{X_{nt}\}$ is L_2-NED on $\{V_t\}$ with NED coefficients ν_k of size -1; $\{V_t\}$ is an $\alpha-$mixing sequence with α_k of size $-\frac{r}{r-2}$; and c) $\{\mu_{nt} \equiv E(X_{nt})\}$ satisfies Assumption 2.2 of GW (2002).

We generalize GW's (2002) assumptions in two dimensions. First, we require slightly more than two finite moments whereas GW (2002) require slightly more than six moments. Second, we allow for more dependence as here $\nu_k = O\left(k^{-1-\varepsilon}\right)$ and $\alpha_k = O\left(k^{-\frac{r}{r-2}-\varepsilon}\right)$, as opposed to $\nu_k = O\left(k^{-\frac{2(r-1)}{r-2}}\right)$ and $\alpha_k = O\left(k^{-\frac{2r}{r-2}}\right)$ in GW (2002). While our size conditions on ν_k do not match the best possible dependence conditions for consistency of kernel variance estimators (cf. De Jong and Davidson (2000)), they are not too much stronger. Indeed, De Jong and Davidson (2000) require that ν_k be of size $-1/2$ as opposed to ν_k of size -1, with the same mixing size conditions as Assumption 1.b). Assumption 1.c) is satisfied if $\mu_{nt} = \mu$ for all t, n. Our main result is as follows:

Theorem 1. *Assume $\{X_{nt}\}$ satisfies Assumption 1. Then, if $p_n \to 0$ and $np_n^2 \to \infty$, $\hat{\sigma}_n^2 - \sigma_n^2 \xrightarrow{P} 0$, with $\sigma_n^2 \equiv Var\left(n^{-1/2}\sum_{t=1}^n X_{nt}\right)$.*

The consistency of the SB distribution of $\sqrt{n}\left(\bar{X}_n^* - \bar{X}_n\right)$ for the distribution of $\sqrt{n}\left(\bar{X}_n - \bar{\mu}_n\right)$ follows from Theorem 1 under a slightly stronger dependence condition than Assumption 1.b):

Assumption 1.b′). For $\delta > 0$ chosen as in Assumption 1.a), $\{X_{nt}\}$ is $L_{2+\delta}-$NED on $\{V_t\}$ with NED coefficients ν_k of size -1; $\{V_t\}$ is an $\alpha-$mixing sequence with α_k of size $-\frac{(2+\delta)(r+\delta)}{r-2}$.

Theorem 2. *Assume $\{X_{nt}\}$ satisfies Assumption 1 strengthened by Assumption 1.b′). Then, if $p_n \to 0$ and $np_n^2 \to \infty$, for any $\varepsilon > 0$, $P\left\{\sup_{x\in\mathbb{R}}\left|P^*\left[\sqrt{n}\left(\bar{X}_n^* - \bar{X}_n\right) \le x\right] - P\left[\sqrt{n}\left(\bar{X}_n - \bar{\mu}_n\right) \le x\right]\right| > \varepsilon\right\} \to 0$, where $\bar{\mu}_n \equiv n^{-1}\sum_{t=1}^n \mu_{nt}$ and P^* is the probability measure induced by the bootstrap.*

Theorem 2 justifies using the SB to build asymptotically valid confidence intervals for (or test hypotheses about) $\bar{\mu}_n$ under only slightly more than two finite moments. Possible applications of this result include White (2000) and Hansen (2003), where the SB is used to compute p-values.

A. Appendix

We present only abbreviated versions of the proofs of our results. For a version of this Appendix with more detailed proofs, see http://www.mapageweb.umontreal.ca/goncals.

Lemma A.1. *Let Z_{nt} be such that $E(Z_{nt}) = 0$ and $|Z_{nt}| \le C$, for all t, n, for some $C < \infty$. If Z_{nt} is L_2-NED on V_t, an $\alpha-$mixing process, then for fixed $\tau > 0$ and all $t < s \le t + \tau$, $|Cov(Z_{nt}Z_{n,t+\tau}, Z_{ns}Z_{n,s+\tau})| \le K_1\left(\alpha_{\left[\frac{s-t}{4}\right]} + \nu_{\left[\frac{s-t}{4}\right]}\right) + K_2\left(\alpha_{[\tau/4]} + \nu_{[\tau/4]}\right)^2$, for some finite constants K_1 and K_2 depending on C, but not on n, t, s or τ.*

Proof. See Lemma A.4 of GW (2002) and note that here the boundedness of Z_{nt} allows using Hölder's inequality with $q = \infty$ and $p = 1$ instead of $q = p = 2$.

Proof of Theorem 1. We show that $\tilde{\sigma}_n^2 - \sigma_n^2 \xrightarrow{P} 0$, where

$$\tilde{\sigma}_n^2 = n^{-1}\sum_{t=1}^n Z_{nt}^2 + 2\sum_{\tau=1}^{n-1} b_n(\tau)\, n^{-1}\sum_{t=1}^{n-\tau} Z_{nt}Z_{n,t+\tau} = n^{-1}\sum_{t=1}^n\sum_{s=1}^n Z_{nt}Z_{ns}b_n(|t-s|), \tag{A.1}$$

with $Z_{nt} \equiv X_{nt} - \mu_{nt}$. The result follows since $\hat{\sigma}_n^2 - \tilde{\sigma}_n^2 = o_P(1)$, if $np_n \to \infty$ (cf. GW, 2002, p. 1379). Let $h(a,x) = xI(|x| \le a) + aI(x > a) - aI(x < -a)$, $g(a,x) = (x-a)I(x > a) + (x+a)I(x < -a)$, and note $x = g(a,x) + h(a,x)$. For some K to be defined later, let $\widetilde{Z}_{nt} = g(K, Z_{nt}) - Eg(K, Z_{nt})$, $\overline{Z}_{nt} = h(K, Z_{nt}) - Eh(K, Z_{nt})$, and note $Z_{nt} = \widetilde{Z}_{nt} + \overline{Z}_{nt}$, since $EZ_{nt} = 0$. Thus, from (A.1), $\tilde{\sigma}_n^2 = A_{1n} + A_{2n} + A_{3n} + \widehat{\overline{\sigma}}_n^2$, where $A_{1n} \equiv n^{-1}\sum_{t=1}^n\sum_{s=1}^n \widetilde{Z}_{nt}\widetilde{Z}_{ns}b_n(|t-s|)$, $A_{2n} \equiv n^{-1}\sum_{t=1}^n\sum_{s=1}^n \widetilde{Z}_{nt}\overline{Z}_{ns}b_n(|t-s|)$, $A_{3n} \equiv n^{-1}\sum_{t=1}^n\sum_{s=1}^n \overline{Z}_{nt}\widetilde{Z}_{ns}b_n(|t-s|)$ and $\widehat{\overline{\sigma}}_n^2 \equiv n^{-1}\sum_{t=1}^n\sum_{s=1}^n \overline{Z}_{nt}\overline{Z}_{ns}b_n(|t-s|)$. The proof follows in three steps. In *Step 1*, we show $\lim_{K\to\infty}\limsup_{n\to\infty} E|A_{in}| = 0$, $i = 1, 2, 3$. In *Step 2*, we show $\widehat{\overline{\sigma}}_n^2 - \overline{\sigma}_n^2 = o_P(1)$, where $\overline{\sigma}_n^2 = Var\left(n^{-1/2}\sum_{t=1}^n \overline{Z}_{nt}\right)$. In *Step 3*, we show $\lim_{K\to\infty}\limsup_{n\to\infty}\left|\overline{\sigma}_n^2 - \sigma_n^2\right| = 0$.
Step 1. First, since we can write $b_n(x) = f_n\left(\frac{x}{n}\right) + f_n\left(1 - \frac{x}{n}\right)$, where $f_n(x) = (1 - |x|)I(|x| \le 1)\exp(-n|x|(-\log(1-p_n)))$, each A_{in} can be written as

$$n^{-1}\sum_{t=1}^n\sum_{s=1}^n \mathcal{X}_{nt}\mathcal{W}_{ns}f_n\left(\frac{|t-s|}{n}\right) + n^{-1}\sum_{t=1}^n\sum_{s=1}^n \mathcal{X}_{nt}\mathcal{W}_{ns}f_n\left(1 - \frac{|t-s|}{n}\right) \equiv a_{1n} + a_{2n},$$

where $\mathcal{X}_{nt} = \mathcal{W}_{nt} = \widetilde{Z}_{nt}$ for A_{1n}, while $\mathcal{X}_{nt} = \widetilde{Z}_{nt}$ and $\mathcal{W}_{nt} = \overline{Z}_{nt}$ for A_{2n} and A_{3n}. Second, note that

$$a_{1n} = (2\pi)^{-2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left[n^{-1/2} \sum_{t=1}^{n} \mathcal{X}_{nt} w_{1nt}(\xi_1, \xi_2) \right] \left[n^{-1/2} \sum_{s=1}^{n} \mathcal{W}_{ns} w_{2ns}(\xi_1, \xi_2) \right] \psi_1(\xi_1) \psi_2(\xi_2) \, d\xi_1 d\xi_2, \tag{A.2}$$

where $\psi_1(\xi_1) = 2\xi_1^{-2}(1 - \cos(\xi_1))$, $\psi_2(\xi_2) = 2\left(1 + \xi_2^2\right)^{-1}$, $w_{1nt}(\xi_1, \xi_2) = \exp\left(-it\left(n^{-1}\xi_1 + q_n\xi_2\right)\right)$ and $w_{2nt}(\xi_1, \xi_2) = \exp\left(it\left(n^{-1}\xi_1 + q_n\xi_2\right)\right)$, with $q_n = -\log(1 - p_n)$. Note $|w_{int}(\xi_1, \xi_2)| = 1$ for all n, t, ξ_1 and ξ_2, for $i = 1, 2$. A representation similar to (A.2) holds for a_{2n}, with a choice of different functions $\widetilde{w}_{1nt}(\cdot,\cdot)$ and $\widetilde{w}_{2nt}(\cdot,\cdot)$ having the same properties as $w_{1nt}(\cdot,\cdot)$ and $w_{2nt}(\cdot,\cdot)$. From (A.2), by Fatou's Lemma, Fubini's theorem and the Cauchy-Schwartz inequality,

$$E|a_{1n}| \leq (2\pi)^{-2} \iint \left\| n^{-1/2} \sum_{t=1}^{n} \mathcal{X}_{nt} w_{1nt}(\xi_1, \xi_2) \right\|_2 \left\| n^{-1/2} \sum_{s=1}^{n} \mathcal{W}_{ns} w_{2ns}(\xi_1, \xi_2) \right\|_2 |\psi_1(\xi_1)| \, |\psi_2(\xi_2)| \, d\xi_1 d\xi_2. \tag{A.3}$$

Under Assumption 1, when $\mathcal{X}_{nt} = \widetilde{Z}_{nt}$, we can show

$$\sup_{(\xi_1, \xi_2) \in \mathbb{R}^2} \left\| n^{-1/2} \sum_{t=1}^{n} \mathcal{X}_{nt} w_{1nt}(\xi_1, \xi_2) \right\|_2 \leq C f^{1/r}(K), \tag{A.4}$$

for $C < \infty$ and some function $f(K)$ such that $f(K) \to 0$ as $K \to \infty$. When $\mathcal{X}_{nt} = \overline{Z}_{nt}$,

$$\sup_{(\xi_1, \xi_2) \in \mathbb{R}^2} \left\| n^{-1/2} \sum_{t=1}^{n} \mathcal{X}_{nt} w_{1nt}(\xi_1, \xi_2) \right\|_2 \leq C. \tag{A.5}$$

Similar bounds apply to $\sup_{(\xi_1,\xi_2)\in\mathbb{R}^2} \left\| n^{-1/2} \sum_{s=1}^{n} \mathcal{W}_{ns} w_{2ns}(\xi_1, \xi_2) \right\|_2$ when $\mathcal{W}_{ns} = \widetilde{Z}_{ns}$ and $\mathcal{W}_{ns} = \overline{Z}_{ns}$, respectively. For A_{1n}, where $\mathcal{X}_{nt} = \mathcal{W}_{nt} = \widetilde{Z}_{nt}$, (A.3) and (A.4) imply $E|a_{1n}| \leq C f^{2/r}(K) (2\pi)^{-2} \int_{-\infty}^{+\infty} |\psi_1(\xi_1)| \, d\xi_1 \int_{-\infty}^{+\infty} |\psi_2(\xi_2)| \, d\xi_2 \leq C f^{2/r}(K)$, for $C < \infty$, given the absolutely integrability of ψ_1 and ψ_2. A similar result holds for A_{2n} and A_{3n}, implying that for each A_{in}, $\limsup_{n\to\infty} E|a_{1n}|$ can be made arbitrarily small by choosing K sufficiently large, since $f(K) \to 0$ as $K \to \infty$. By Markov's inequality, for this choice of K, $a_{1n} \overset{P}{\to} 0$. A similar proof applies to a_{2n} and thus $A_{in} \overset{P}{\to} 0$, for $i = 1, 2, 3$.

Next, we prove (A.4) and (A.5). For (A.4), define $\widetilde{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K) = \widetilde{Z}_{nt}(K) w_{1nt}(\xi_1, \xi_2)$, with the dependence of $\widetilde{Z}_{nt}$ on K now being explicit. For all K, $\widetilde{Z}_{nt}(K)$ is a Lipschitz function of Z_{nt}, and for all (ξ_1, ξ_2), $w_{1nt}(\xi_1, \xi_2)$ is a non random function bounded in absolute value by 1. Thus, for each (ξ_1, ξ_2, K), $\widetilde{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K)$ is mean-zero, L_2−NED on V_t with the same size as Z_{nt} (cf. Davidson, 1994, Theorem 17.12). But

$$\sup_{(\xi_1, \xi_2) \in \mathbb{R}^2} \left\| \widetilde{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K) \right\|_r = \sup_{(\xi_1, \xi_2) \in \mathbb{R}^2} |w_{1nt}(\xi_1, \xi_2)| \left\| \widetilde{Z}_{nt}(K) \right\|_r = \left\| \widetilde{Z}_{nt}(K) \right\|_r \leq 2 \left\| g(K, Z_{nt}) \right\|_r,$$

where the second equality holds by $|w_{1nt}(\xi_1, \xi_2)| = 1$ and the last inequality holds by the Minkowsky and the Jensen inequalities. By definition of $g(K, Z_{nt})$, we have $|g(K, Z_{nt})| \leq |Z_{nt} I(|Z_{nt}| > K)|$, implying

$$E|g(K, Z_{nt})|^r \leq E(|Z_{nt}|^r I(|Z_{nt}|^r > K^r)) \leq \sup_{t,n} E(|Z_{nt}|^r I(|Z_{nt}| > K)) \equiv f(K),$$

where $f(K) \to 0$ as $K \to \infty$. Thus, by Davidson's (1994) Corollary 17.6 (i), for each (ξ_1, ξ_2, K), $\widetilde{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K)$ is an L_2−mixingale of size $-\min\left\{1, \frac{r}{r-2}\left(\frac{1}{2} - \frac{1}{r}\right)\right\} = -\frac{1}{2}$, with mixingale constants $\widetilde{c}_{nt}(\xi_1, \xi_2, K)$ uniformly bounded by $\sup_{(\xi_1,\xi_2)\in\mathbb{R}^2} \left\| \widetilde{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K) \right\|_r \leq 2 f^{1/r}(K)$. By McLeish's (1975, Theorem 1.6) inequality, for all (ξ_1, ξ_2, K),

$$E\left(\sum_{t=1}^{n} \widetilde{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K)\right)^2 \leq E\left(\max_{j\leq n}\sum_{t=1}^{j} \widetilde{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K)\right)^2 \leq C\sum_{t=1}^{n} \widetilde{c}_{nt}^2(\xi_1, \xi_2, K) \leq Cnf^{2/r}(K),$$

proving (A.4). To prove (A.5), note that as before, $\overline{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K)$ is an L_2-mixingale of size $-1/2$ with constants $\overline{c}_{nt}(\xi_1, \xi_2, K)$ uniformly bounded by $\sup_{(\xi_1,\xi_2)\in\mathbb{R}^2} \left\|\overline{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K)\right\|_r$. But

$$\sup_{(\xi_1,\xi_2)\in\mathbb{R}^2} \left\|\overline{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K)\right\|_r = \sup_{(\xi_1,\xi_2)\in\mathbb{R}^2} |w_{1nt}(\xi_1, \xi_2)| \left\|\overline{Z}_{nt}(K)\right\|_r \leq 2\left\|h(K, Z_{nt})\right\|_r \leq 2\left\|Z_{nt}\right\|_r < 4\Delta^{1/r},$$

given Assumption 1.a) and $|h(K, Z_{nt})| \leq Z_{nt}$ for all t, n, thus implying (A.5) by McLeish's inequality. *Step 2.* For all $K > 0$, $\overline{Z}_{nt} = h(K, Z_{nt}) - Eh(K, Z_{nt})$ is a mean-zero, L_2-NED array on V_t of size -1, where V_t is $\alpha-$mixing of size $-\frac{r}{r-2}$, $r > 2$, hence of size -1. Because $\left|\overline{Z}_{nt}\right| \leq K$ for all t, n, we use Lemma A.1 to show that $Var\left(\widehat{\overline{\sigma}}_n^2\right) = O\left(\frac{1}{np_n^2}\right)$ and proceed as in GW (2002), proof of Theorem 2.1. *Step 3.* Since $\overline{Z}_{nt} = Z_{nt} - \widetilde{Z}_{nt}$, we have $\overline{\sigma}_n^2 = \sigma_n^2 + B_{1n} + B_{2n} + B_{3n}$, where $B_{1n} \equiv n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} E\left(Z_{nt}\widetilde{Z}_{ns}\right)$, $B_{2n} \equiv n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} E\left(\widetilde{Z}_{nt}Z_{ns}\right)$ and $B_{3n} \equiv n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} E\left(\widetilde{Z}_{nt}\widetilde{Z}_{ns}\right)$ can be made arbitrarily small by an argument similar to Step 1.

Proof of Theorem 2. Follow GW's (2002) proof of Theorem 2.2., verifying their condition (C2) as in Theorem 1.

References

[1] Davidson, J. (1994) *Stochastic limit theory.* Oxford: Oxford University Press.

[2] De Jong, R. and J. Davidson (2000) Consistency of Kernel Estimators of Heteroskedastic and Autocorrelated Covariance Matrices. *Econometrica*, 68, 407-423.

[3] Gonçalves, S. and H. White (2002) The Bootstrap of the Mean for Dependent Heterogeneous Arrays. *Econometric Theory,* 18, *1367-1384.*

[4] Hansen, P. (2003) Regression Analysis with Many Specifications: a Bootstrap Method to Robust Inference. *mimeo*, Brown University.

[5] Künsch, H.R. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217-1241.

[6] Lahiri, S.N. (1999) Theoretical comparisons of block bootstrap methods. *Annals of Statistics* 27, 386-404.

[7] Liu, R.Y. and K. Singh (1992) Moving blocks jackknife and bootstrap capture weak dependence In R. LePage and L. Billiard (eds), *Exploring the limits of the bootstrap*, pp. 224-248. New York: Wiley.

[8] McLeish, D.L. (1975) A maximal inequality and dependent strong laws. *Annals of Probability* 5, 616-621.

[9] Politis, D. and J. Romano (1994) The stationary bootstrap. *Journal of American Statistics Association* 89, 1303-1313.

[10] White, H. (2000) A Reality Check for Data Snooping. *Econometrica*, 68, 1097-1126.