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, thus limiting the applicability of this bootstrap method.

The main purpose of this note is to 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. Nevertheless, as remarked by Politis and Romano (1994), the SB may be less sensitive to the choice of the (average) block length than the MBB is to the choice of the (fixed) block length.

For variance estimation, the consistency of the MBB under weak moment conditions follows straigthforwardly by an application of De Jong and Davidson's (2000a) consistency results for kernel variance estimators under minimal conditions. This is a consequence of the fact that the MBB variance estimator for the sample mean is equal to a Bartlett kernel variance estimator, up to terms of order $O_P\left(\ell^2/n\right)$, where ℓ is the block length. These terms vanish in probability under the condition that $\frac{\ell^2}{n} \to 0$. In this case, the consistency of the MBB variance estimator holds even under minimal dependence conditions (cf. De Jong and Davidson, 2000a). The consistency of the MBB distribution estimator follows as a corollary, under the same weak moment conditions but stronger size conditions on the NED coefficients. To conserve space and because the SB is harder to analyse, in this note we will 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 a positive sequence such that $p_n \to 0$ and $np_n^2 \to \infty$.

Our first goal is to show that $\hat{\sigma}_n^2 - \sigma_n^2 \xrightarrow{P} 0$, where $\sigma_n^2 \equiv Var\left(n^{-1/2}\sum_{t=1}^{n} X_{nt}\right)$, under weaker moment conditions than previously considered in the literature, but otherwise general dependence conditions. In particular, we assume that $\{X_{nt}, n, t = 1, 2, ...\}$ is L_2-near epoch dependent (L_2-NED) on a mixing process $\{V_t\}$, i.e. $\|X_{nt}\|_2 < \infty$ and $\nu_k \equiv \sup_{n,t}\left\|X_{nt} - E_{t-k}^{t+k}(X_{nt})\right\|_2$ tends to zero as $k \to \infty$ at an appropriate rate. Here and in what follows, $\|X\|_q = (E|X_i|^q)^{1/q}$ for $q \geq 1$ denotes the L_q-norm of a random variable X. Similarly, we let $E_{t-k}^{t+k}(\cdot) \equiv E\left(\cdot|\mathcal{F}_{t-k}^{t+k}\right)$, where $\mathcal{F}_{t-k}^{t+k} \equiv \sigma(V_{t-k}, \ldots, V_{t+k})$ is the σ-field generated by $V_{t-k}, \ldots, V_{t+k}$. In particular, 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 $-a$. The sequence $\{V_t\}$ is assumed to be strong mixing, where we define the strong or α-mixing coefficients as usual, i.e. $\alpha_k \equiv \sup_m \sup_{\{A\in\mathcal{F}_{-\infty}^m, B\in\mathcal{F}_{m+k}^{\infty}\}}|P(A\cap B) - P(A)P(B)|$, and we require $\alpha_k \to 0$ as $k \to \infty$ at an appropriate rate.

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 $-\frac{r}{r-2}$.

c) $\{\mu_{nt} \equiv E(X_{nt})\}$ satisfies Assumption 2.2 of GW (2002).

Assumption 1 generalizes Assumption 2.1 of GW (2002) in two dimensions. First, Assumption 1.a) requires slightly more than two finite moments whereas Assumption 2.1.a) of GW (2002) requires slightly more than six moments. Second, Assumption 1.b) allows for more dependence as it requires $\nu_k = O\left(k^{-1-\varepsilon}\right)$ 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 (2000a)), they are not too much stronger. Indeed, De Jong and Davidson (2000a) 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). As remarked by GW (2002), 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 $np_n^2 \to \infty$,*

$$\hat{\sigma}_n^2 - \sigma_n^2 \xrightarrow{P} 0.$$

We can use Theorem 1 to show the consistency of the stationary bootstrap distribution of $\sqrt{n}\left(\bar{X}_n^* - \bar{X}_n\right)$ for the distribution of $\sqrt{n}\left(\bar{X}_n - \bar{\mu}_n\right)$. As in GW (2002), we require $\{X_{nt}\}$ to satisfy a slightly stronger dependence condition than Assumption 1.b), namely we impose:

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 $\varepsilon > 0$,*

$$P\left\{\sup_{x\in\mathbb{R}}\left|P^*\left[\sqrt{n}\left(\bar{X}_n^* - \bar{X}_n\right) \leq x\right] - P\left[\sqrt{n}\left(\bar{X}_n - \bar{\mu}_n\right) \leq x\right]\right| > \varepsilon\right\} \to 0,$$

where $\bar{\mu}_n \equiv n^{-1}\sum_{t=1}^n \mu_{nt}$ and P^* is the probability measure induced by the bootstrap, conditional on $\{X_{nt}\}_{t=1}^n$.

Theorem 2 justifies using the SB distribution to build asymptotically valid confidence intervals for (or test hypotheses about) $\bar{\mu}_n$ under the existence of 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

For the proof of Theorem 1, we will use the following lemmas.

Lemma A.1. *Let* $\{Z_{nt}, \mathcal{F}^t\}$ *be an* L_2-*mixingale of size* $-1/2$ *with mixingale constants* c_{nt}. *Then,* $E\left(\max_{j\leq n}\left(\sum_{t=1}^{j} Z_{nt}\right)^2\right) = O\left(\sum_{t=1}^{n} c_{nt}^2\right)$.

Proof. See McLeish (1975), Theorem 1.6. ■

Lemma A.2. *Let* Z_{nt} *be such that* $E(Z_{nt}) = 0$ *and* $|Z_{nt}| \leq C$, *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* $t < s \leq t + \tau$,

$$|Cov(Z_{nt}Z_{n,t+\tau}, Z_{ns}Z_{n,s+\tau})| \leq K_1\left(\alpha_{\left[\frac{s-t}{4}\right]} + \nu_{\left[\frac{s-t}{4}\right]}\right) + K_2\left(\alpha_{[\tau/4]} + \nu_{[\tau/4]}\right)^2,$$

for some finite constants K_1 *and* K_2 *depending on* C, *but not on* n, t, s *or* τ.

Proof. The proof of this result follows closely that of Lemma A.4 of GW (2002). In particular, their proof relies on using the Cauchy-Schwartz inequality in several instances as a way to bound moments of NED arrays. Here, the boundedness of Z_{nt} allows using Hölder's inequality with $q = \infty$ and $p = 1$ instead of $q = p = 2$, which explains the improvements on the size conditions on $\alpha(\cdot)$ and $\nu(\cdot)$. Consider bounding $|E(Z_{nt}Z_{n,t+\tau})|$. As in Gallant and White (1988, pp. 109-110), we have

$$|E(Z_{nt}Z_{n,t+\tau})| \leq \left|E\left(Z_{n,t+\tau}E_{t-[\tau/2]}^{t+[\tau/2]}Z_{nt}\right)\right| + \left|E\left(Z_{n,t+\tau}\left(Z_{nt} - E_{t-[\tau/2]}^{t+[\tau/2]}Z_{nt}\right)\right)\right|.$$

Now, by Hölder's inequality with $q = \infty$ and $p = 1$, we have that

$$\left|E\left(Z_{n,t+\tau}\left(Z_{nt} - E_{t-[\tau/2]}^{t+[\tau/2]}Z_{nt}\right)\right)\right| \leq \|Z_{n,t+\tau}\|_\infty \left\|Z_{nt} - E_{t-[\tau/2]}^{t+[\tau/2]}Z_{nt}\right\|_1 \leq C\left\|Z_{nt} - E_{t-[\tau/2]}^{t+[\tau/2]}Z_{nt}\right\|_2 \leq C\nu_{[\tau/2]},$$

whereas we can show that

$$\begin{aligned}\left|E\left(Z_{n,t+\tau}E_{t-[\tau/2]}^{t+[\tau/2]}Z_{nt}\right)\right| &= \left|E\left(E_{t-[\tau/2]}^{t+[\tau/2]}Z_{nt}E\left(Z_{n,t+\tau}|\mathcal{F}^{t+[\tau/2]}\right)\right)\right| \leq \left\|E_{t-[\tau/2]}^{t+[\tau/2]}Z_{nt}\right\|_\infty \left\|E\left(Z_{n,t+\tau}|\mathcal{F}^{t+[\tau/2]}\right)\right\|_1 \\ &\leq C\left\|E\left(Z_{n,t+\tau}|\mathcal{F}^{t+[\tau/2]}\right)\right\|_1 \leq C\left(6\alpha_{[\tau/4]}\|Z_{n,t+\tau}\|_\infty + \nu_{[\tau/4]}\right).\end{aligned}$$

Thus, we have that $|E(Z_{nt}Z_{n,t+\tau})| \leq C\left(6\alpha_{[\tau/4]} + \nu_{[\tau/4]}\right)$. Similarly, to bound $|E(Z_{nt}Z_{n,t+\tau}Z_{ns}Z_{n,s+\tau})|$ when $t < s \leq t + \tau$, we follow GW's (2002) argument. For instance, consider bounding the second term in their expression (A1). By Hölder's inequality with $q = \infty$ and $p = 1$, we have that

$$\left|E\left(Z_{n,s+\tau}\left(Z_{nt}Z_{ns}Z_{n,t+\tau} - \hat{Y}_{nmt\tau}\right)\right)\right| \leq \|Z_{n,t+\tau}\|_\infty \left\|Z_{nt}Z_{ns}Z_{n,t+\tau} - \hat{Y}_{nmt\tau}\right\|_1 \leq C\left\|Z_{nt}Z_{ns}Z_{n,t+\tau} - \hat{Y}_{nmt\tau}\right\|_2.$$

To bound the last expression, we follow the same argument as in GW (2002) and use the fact that $|Z_{nt}| \leq C$ for all t, n to show that their function $B\left(\mathfrak{X}, \hat{\mathfrak{X}}\right)$ is bounded by $3C^2$. This implies that $\left\|Z_{nt}Z_{ns}Z_{n,t+\tau} - \hat{Y}_{nmt\tau}\right\|_2 \leq 3C^2\left(6\nu_{\left[\frac{s-t}{4}\right]}\right)$. ■

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

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

with $Z_{nt} = X_{nt} - \mu_{nt}$ a mean zero triangular array. This implies the result since, under Assumption 1, we can show that $\hat{\sigma}_n^2 - \tilde{\sigma}_n^2 = o_P(1)$, if $np_n \to \infty$ (see GW's (2002) proof of their Theorem 2.1. cf. step 2).

For the proof, the following expression for $\tilde{\sigma}_n^2$ is more convenient than (A.1); it follows by the properties of $b_n(\cdot)$:

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

where $b_n(x)$ can be written as

$$b_n(x) = f_n\left(\frac{x}{n}\right) + f_n\left(1 - \frac{x}{n}\right), \tag{A.3}$$

with

$$f_n(x) = (1 - |x|) I(|x| \leq 1) \exp(-n|x|(-\log(1-p_n))). \tag{A.4}$$

Following De Jong and Davidson (2000b), we introduce the following functions:

$$\begin{aligned} h(a,x) &= xI(|x| \leq a) + aI(x > a) - aI(x < -a), \\ g(a,x) &= (x-a) I(x > a) + (x+a) I(x < -a), \end{aligned}$$

and note that $x = g(a,x) + h(a,x)$. For some K to be defined later, we let

$$\tilde{Z}_{nt} = g(K, Z_{nt}) - Eg(K, Z_{nt}), \text{ and } \overline{Z}_{nt} = h(K, Z_{nt}) - Eh(K, Z_{nt}),$$

and note that $Z_{nt} = \tilde{Z}_{nt} + \overline{Z}_{nt}$, since $EZ_{nt} = 0$ by construction. Thus, from (A.2) it follows that

$$\begin{aligned} \tilde{\sigma}_n^2 &= n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \left(\tilde{Z}_{nt} + \overline{Z}_{nt}\right)\left(\tilde{Z}_{ns} + \overline{Z}_{ns}\right) b_n(|t-s|) \\ &= n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \tilde{Z}_{nt} \tilde{Z}_{ns} b_n(|t-s|) + n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \tilde{Z}_{nt} \overline{Z}_{ns} b_n(|t-s|) \\ &\quad + n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \overline{Z}_{nt} \tilde{Z}_{ns} b_n(|t-s|) + n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \overline{Z}_{nt} \overline{Z}_{ns} b_n(|t-s|) \\ &\equiv A_{1n} + A_{2n} + A_{3n} + \hat{\overline{\sigma}}_n^2. \end{aligned}$$

The proof will proceed in three steps.

In *Step 1*, we will show that $\lim_{K\to\infty} \limsup_{n\to\infty} E|A_{in}| = 0$, for $i = 1, 2, 3$.

In *Step 2*, we will show that $\hat{\overline{\sigma}}_n^2 - \overline{\sigma}_n^2 \to 0$ in probability, where $\overline{\sigma}_n^2 = Var\left(n^{-1/2} \sum_{t=1}^{n} \overline{Z}_{nt}\right)$.

In *Step 3*, we will show that $\lim_{K\to\infty} \lim \left|\overline{\sigma}_n^2 - \sigma_n^2\right| \to 0$.

Proof of Step 1. First, given (A.3) and (A.4), note that we can write each A_{in}, $i = 1, 2, 3$, 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 $\mathcal{X}_{nt}, \mathcal{W}_{nt}$ correspond to different choices of $\tilde{Z}_{nt}$ and/or $\overline{Z}_{nt}$ depending on the particular term A_{in}. In particular, $\mathcal{X}_{nt} = \mathcal{W}_{nt} = \tilde{Z}_{nt}$ for A_{1n}, while $\mathcal{X}_{nt} = \tilde{Z}_{nt}$ and $\mathcal{W}_{nt} = \overline{Z}_{nt}$ for A_{2n} and A_{3n}.

Second, notice that we can write

$$\begin{aligned} a_{1n} &\equiv n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \mathcal{X}_{nt} \mathcal{W}_{ns} f_n\left(\frac{|t-s|}{n}\right) \qquad \text{(A.5)} \\ &= (2\pi)^{-2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left[n^{-1/2} \sum_{t=1}^{n} \mathcal{X}_{nt} w_{1nt}(\xi_1, \xi_2)\right] \left[n^{-1/2} \sum_{s=1}^{n} \mathcal{W}_{ns} w_{2ns}(\xi_1, \xi_2)\right] \psi_1(\xi_1) \psi_2(\xi_2) d\xi_1 d\xi_2, \end{aligned}$$

where $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are real-valued, nonnegative absolutely integrable functions given by

$$\psi_1(\xi_1) = 2\xi_1^{-2}(1-\cos(\xi_1)), \ \psi_2(\xi_2) = 2\left(1+\xi_2^2\right)^{-1},$$

and $w_{1nt}(\cdot,\cdot)$ and $w_{2nt}(\cdot,\cdot)$ are complex-valued nonrandom functions defined as

$$w_{1nt}(\xi_1,\xi_2) = \exp\left(-it\left(n^{-1}\xi_1 + q_n\xi_2\right)\right) \text{ and } w_{2nt}(\xi_1,\xi_2) = \exp\left(it\left(n^{-1}\xi_1 + q_n\xi_2\right)\right),$$

with $q_n = -\log(1-p_n)$. Note that $|w_{int}(\xi_1,\xi_2)| = 1$ for all n, t, ξ_1 and ξ_2, for $i = 1, 2$.

A representation similar to (A.5) holds for a_{2n}, with a choice of different functions $\widetilde{w}_{1nt}(\cdot,\cdot)$ and $\widetilde{w}_{2nt}(\cdot,\cdot)$, which nonetheless share the same properties as $w_{1nt}(\cdot,\cdot)$ and $w_{2nt}(\cdot,\cdot)$.

To see why equation (A.5) holds, we note that

$$(1-|x|)\,I(|x|\leq 1) = (2\pi)^{-1}\int_{-\infty}^{+\infty}\exp(-ix\xi_1)\,2\xi_1^{-2}(1-\cos(\xi_1))\,d\xi_1 = (2\pi)^{-1}\int_{-\infty}^{+\infty}\exp(-ix\xi_1)\,\psi_1(\xi_1)\,d\xi_1$$

and

$$\exp(-|x|) = (2\pi)^{-1}\int_{-\infty}^{+\infty}\exp(-ix\xi_2)\,2\left(1+\xi_2^2\right)^{-1}d\xi_1 = (2\pi)^{-1}\int_{-\infty}^{+\infty}\exp(-ix\xi_2)\,\psi_2(\xi_2)\,d\xi_2,$$

which implies

$$f_n(x) = \left[(2\pi)^{-1}\int_{-\infty}^{+\infty}\exp(-ix\xi_1)\,\psi_1(\xi_1)\,d\xi_1\right]\left[(2\pi)^{-1}\int_{-\infty}^{+\infty}\exp\left(-ix\xi_2 n\left(-\log(1-p_n)\right)\right)\psi_2(\xi_2)\,d\xi_2\right].$$

From (A.5), noting that ψ_1 and ψ_2 are absolutely integrable functions, it follows by Fatou's Lemma, Fubini's theorem and the Cauchy-Schwartz inequality that

$$E|a_{1n}| \leq (2\pi)^{-2}\iint\left\|n^{-1/2}\sum_{t=1}^{n}\mathcal{X}_{nt}w_{1nt}(\xi_1,\xi_2)\right\|_2\left\|n^{-1/2}\sum_{s=1}^{n}\mathcal{W}_{ns}w_{2ns}(\xi_1,\xi_2)\right\|_2|\psi_1(\xi_1)|\,|\psi_2(\xi_2)|\,d\xi_1 d\xi_2. \tag{A.6}$$

As we will show next, under Assumption 1, when $\mathcal{X}_{nt} = \widetilde{Z}_{nt}$,

$$\sup_{(\xi_1,\xi_2)\in\mathbb{R}^2}\left\|n^{-1/2}\sum_{t=1}^{n}\mathcal{X}_{nt}w_{1nt}(\xi_1,\xi_2)\right\|_2 \leq Cf^{1/r}(K), \tag{A.7}$$

where C is some generic constant and $f(K)$ is a function defined below such that $f(K)\to 0$ as $K\to\infty$. When $\mathcal{X}_{nt} = \overline{Z}_{nt}$,

$$\sup_{(\xi_1,\xi_2)\in\mathbb{R}^2}\left\|n^{-1/2}\sum_{t=1}^{n}\mathcal{X}_{nt}w_{1nt}(\xi_1,\xi_2)\right\|_2 \leq C. \tag{A.8}$$

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.6) and (A.7) imply that

$$E|a_{1n}| \leq Cf^{2/r}(K)\,(2\pi)^{-2}\int_{-\infty}^{+\infty}|\psi_1(\xi_1)|\,d\xi_1\int_{-\infty}^{+\infty}|\psi_2(\xi_2)|\,d\xi_2 \leq Cf^{2/r}(K),$$

for some $C<\infty$. The second inequality holds by the absolutely integrability of the functions ψ_1 and ψ_2. Similarly, for A_{2n} and A_{3n} where $\mathcal{X}_{nt} = \widetilde{Z}_{nt}$ and $\mathcal{W}_{nt} = \overline{Z}_{nt}$, we will have that

$$E|a_{1n}| \leq Cf^{1/r}(K).$$

It follows that for each of the terms A_{1n}, A_{2n} and A_{3n}, $\limsup_{n\to\infty} E|a_{1n}|$ can be made arbitrarily small by choosing K sufficiently large, since $f(K) \to 0$ as $K \to \infty$. By Markov's inequality, it follows that, for this choice of K, $a_{1n} \xrightarrow{P} 0$. The term a_{2n} can be dealt with in a similar fashion, which completes the proof that $A_{in} \xrightarrow{P} 0$, for $i = 1, 2, 3$.

To complete the proof of step 1, we will prove (A.7) and (A.8).

We consider first the case in which $\mathcal{X}_{nt} = \widetilde{Z}_{nt}$ and define

$$\widetilde{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K) = \widetilde{Z}_{nt}(K)\, w_{1nt}(\xi_1, \xi_2),$$

where the dependence of $\widetilde{Z}_{nt}$ on K is now made 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, we can show that 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} (see Davidson, 1994, Theorem 17.12, p. 269).

Under Assumption 1, we can show that $\sup_{(\xi_1,\xi_2)\in\mathbb{R}^2} \left\| \widetilde{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K) \right\|_r \leq C f^{1/r}(K)$ for some function $f(K) \to 0$ for $K \to \infty$ (see below). By Corollary 17.6 (i) of Davidson (1994, p. 265), it follows that 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 respect to mixingale constants $\widetilde{c}_{nt}(\xi_1, \xi_2, K)$, which are uniformly bounded by $\sup_{(\xi_1,\xi_2)\in\mathbb{R}^2} \left\| \widetilde{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K) \right\|_r$. We have that

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

where the first equality holds by definition of $\widetilde{\mathcal{Z}}_{nt}(\cdot,\cdot,\cdot)$ and the non-randomness of $w_{1nt}(\xi_1, \xi_2)$, the second equality holds by the fact $|w_{1nt}(\xi_1, \xi_2)| = 1$, and the last inequality holds by an application of the Minkowsky and the Jensen inequalities, given that $\widetilde{Z}_{nt}(K) = g(K, Z_{nt}) - E(g(K, Z_{nt}))$. By the definition of $g(K, Z_{nt})$, we have that

$$|g(K, Z_{nt})| \leq |Z_{nt} I(|Z_{nt}| > K)|,$$

which implies that

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

where we can show that, under Assumption 1, $f(K)$ can be made arbitrarily small for K sufficiently large. The argument is the same as in Davidson (1994, p. 190):

$$\sup_{t,n} E|Z_{nt}|^{r+\delta} \geq \sup_{t,n} E\left(|Z_{nt}|^{r+\delta} I(|Z_{nt}| > K)\right) \geq K^\delta \sup_{t,n} E(|Z_{nt}|^r I(|Z_{nt}| > K)) \equiv K^\delta f(K),$$

and since under Assumption 1 $\sup_{t,n} E|Z_{nt}|^{r+\delta} \leq \Delta < \infty$ it must be the case that $f(K) \to 0$ (sufficiently fast) as $K \to \infty$ (otherwise $K^\delta f(K) \to \infty$). Thus, the mixingale constants of $\widetilde{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K)$ are such that

$$\widetilde{c}_{nt}(\xi_1, \xi_2, K) \leq \sup_{(\xi_1,\xi_2)\in\mathbb{R}^2} \left\| \widetilde{\mathcal{Z}}_{nt}(\xi_1, \xi_2, K) \right\|_r \leq 2 f^{1/r}(K).$$

Lemma A.1 now implies that for all (ξ_1, ξ_2, K),

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

which implies

$$\sup_{(\xi_1,\xi_2)\in\mathbb{R}^2}\left\|\left(n^{-1/2}\sum_{t=1}^{n}\widetilde{\mathcal{Z}}_{nt}(\xi_1,\xi_2,K)\right)\right\|_2 \le Cf^{1/r}(K),$$

thus proving (A.7).

To prove (A.8) when $\mathcal{X}_{nt}=\overline{Z}_{nt}$, note that by definition of h, for all K, $\overline{Z}_{nt}(K)$ is a Lipschitz function of Z_{nt}, and therefore we can show that for each (ξ_1,ξ_2,K), $\overline{\mathcal{Z}}_{nt}(\xi_1,\xi_2,K)\equiv\overline{Z}_{nt}(K)\,w_{1nt}(\xi_1,\xi_2)$ is a mean-zero, L_2-NED on V_t with the same size as Z_{nt}. As before, $\overline{\mathcal{Z}}_{nt}(\xi_1,\xi_2,K)$ is also an L_2-mixingale of size $-1/2$ with respect to mixingale constants $\overline{c}_{nt}(\xi_1,\xi_2,K)$ which are uniformly bounded by $\sup_{(\xi_1,\xi_2)\in\mathbb{R}^2}\left\|\overline{\mathcal{Z}}_{nt}(\xi_1,\xi_2,K)\right\|_r$. It follows that

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

where in particular the second-to-last inequality holds by the fact that $|h(K,Z_{nt})|\le Z_{nt}$ for all t,n, and the last inequality holds by Assumption 1 (i). Thus, by Lemma A.1,

$$E\left(\sum_{t=1}^{n}\overline{\mathcal{Z}}_{nt}(\xi_1,\xi_2,K)\right)^2\le C\left(\sum_{t=1}^{n}\overline{c}_{nt}^2(\xi_1,\xi_2,K)\right)\le C\left(16\Delta^{2/r}\right)n,$$

implying that

$$\sup_{(\xi_1,\xi_2)\in\mathbb{R}^2}\left\|\left(n^{-1/2}\sum_{t=1}^{n}\overline{\mathcal{Z}}_{nt}(\xi_1,\xi_2,K)\right)\right\|_2\le C,$$

for some $C<\infty$.

Proof of Step 2. To show that for all $K>0$, $\widehat{\overline{\sigma}}_n^2-\overline{\sigma}_n^2\xrightarrow{P}0$, note that $\overline{Z}_{nt}\equiv\overline{Z}_{nt}(K)=h(K,Z_{nt})-Eh(K,Z_{nt})$ 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 $\left|\overline{Z}_{nt}\right|\le K$ for all t,n, we can rely on Lemma A.2 to show that $Var\left(\widehat{\overline{\sigma}}_n^2\right)=O\left(\frac{1}{np_n^2}\right)$, following an argument similar to that of GW (2002, proof of Theorem 2.1). Similarly, we can show that $E\left(\widehat{\overline{\sigma}}_n^2\right)-\overline{\sigma}_n^2\to 0$ under Assumption 1.

Proof of Step 3. Given that $\overline{Z}_{nt}=Z_{nt}-\widetilde{Z}_{nt}$, we can write

$$\begin{aligned}\overline{\sigma}_n^2 &= E\left(n^{-1/2}\sum_{t=1}^{n}\left(Z_{nt}-\widetilde{Z}_{nt}\right)\right)^2=n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}E\left(\left(Z_{nt}-\widetilde{Z}_{nt}\right)\left(Z_{ns}-\widetilde{Z}_{ns}\right)\right)\\ &= n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}E(Z_{nt}Z_{ns})+n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}E\left(Z_{nt}\widetilde{Z}_{ns}\right)+n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}E\left(\widetilde{Z}_{nt}Z_{ns}\right)+n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}E\left(\widetilde{Z}_{nt}\widetilde{Z}_{ns}\right)\\ &= \sigma_n^2+B_{1n}+B_{2n}+B_{3n},\end{aligned}$$

where B_{1n}, B_{2n} and B_{3n} can all be made arbitrarily small by using an argument similar to that used in Step 1. Consider e.g. B_{1n}. We have that

$$\begin{aligned}B_{1n} &= n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}E\left(Z_{nt}\widetilde{Z}_{ns}\right)=E\left[\left(n^{-1/2}\sum_{t=1}^{n}Z_{nt}\right)\left(n^{-1/2}\sum_{s=1}^{n}\tilde{Z}_{ns}\right)\right]\\ &\le \left\|n^{-1/2}\sum_{t=1}^{n}Z_{nt}\right\|_2\left\|n^{-1/2}\sum_{s=1}^{n}\tilde{Z}_{ns}\right\|_2=O(1)\,O\left(f^{1/r}(K)\right),\end{aligned}$$

where the first inequality holds by an application of Cauchy-Schwartz and the last equality holds by application of mixingale inequalities to Z_{nt} and $\widetilde{Z}_{nt}$, respectively. In particular, the term $O\left(f^{1/r}(K)\right)$ follows by an argument similar to that used to show (A.7) in step 1. Choosing K sufficiently large will ensure that B_{1n} can be made arbitrarily small. The same argument applies to B_{2n} and B_{3n} since each of these contains $\widetilde{Z}_{nt}$ at least once. ■

Proof of Theorem 2. The proof follows closely that of GW's Theorem 2.2. (2002). A CLT for the sample mean of NED arrays holds under Assumption 1, verifying their condition (C1); (C2) follows by an argument similar to the proof of Theorem 1. For (C3), it suffices that $E\left|\sum_{t=\tau}^{\tau+b-1} Z_{nt}\right|^{2+\delta} \leq Cb^{1+\delta/2}$. In particular, Assumption 1.b$'$) ensures that Z_{nt} is an $L_{2+\delta}-$mixingale of size -1, the size requirement of Hansen's maximal inequality for L_p-mixingales, for $p \geq 2$ (cf. Hansen, 1991). ■

References

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

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

[3] De Jong, R. and J. Davidson (2000b) "The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I: Weakly Dependent Processes", *Econometric Theory*, 16, 621-642.

[4] Gallant, A.R. and H. White (1988) *A unified theory of estimation and inference for nonlinear dynamic models.* New York: Basil Blackwell.

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

[6] Hansen, B.E. (1991) Strong laws for dependent heterogeneous processes. *Econometric Theory* 7, 213-221.

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

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

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

[10] 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.

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

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

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