The Bootstrap of the Mean for Dependent Heterogeneous Arrays*

Sílvia Gonçalves
C.R.D.E, CIRANO and Université de Montréal†

Halbert White
University of California, San Diego‡

First draft: October 2000
This version: April 2002

Abstract

Presently, conditions ensuring the validity of bootstrap methods for the sample mean of (possibly heterogeneous) near epoch dependent (NED) functions of mixing processes are unknown. Here we establish the validity of the bootstrap in this context, extending the applicability of bootstrap methods to a class of processes broadly relevant for applications in economics and finance. Our results apply to two block bootstrap methods: the moving blocks bootstrap of Künsch (1989) and Liu and Singh (1992) and the stationary bootstrap of Politis and Romano (1994). In particular, the consistency of the bootstrap variance estimator for the sample mean is shown to be robust against heteroskedasticity and dependence of unknown form. The first order asymptotic validity of the bootstrap approximation to the actual distribution of the sample mean is also established in this heterogeneous NED context.

Keywords: Block Bootstrap, near epoch dependence, sample mean

*We would like to thank a co-editor and three anonymous referees for many valuable suggestions.

†C.P.6128, succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada. Tel: (514) 343 6556. Email: silvia.goncalves@umontreal.ca. Gonçalves acknowledges financial support from Praxis XXI and a Sloan Dissertation fellowship.

‡9500 Gilman Drive, La Jolla, California, 92093-0508. USA . Tel: (858) 534-3502. Email: hwhite@weber.ucsd.edu. White's participation was supported by NSF grants SBR-9811562 and SES-0111238.

1. Introduction

Bootstrap methods have been most intensively studied for the case of independent identically distributed (i.i.d.) observations (e.g., Bickel and Freedman (1981), Singh (1981)). However, the failure of the i.i.d. resampling scheme to give a consistent approximation to the true limiting distribution of a statistic when observations are not independent (e.g. as remarked in Singh (1981)) has motivated several attempts in the literature to modify and extend Efron's (1979) idea to dependent data. Most of the extensions so far apply only to the stationary case. Bootstrap methods appropriate for stationary mixing processes have been proposed and studied by Künsch (1989) and Liu and Singh (1992) (the "moving blocks" bootstrap) and by Politis and Romano (1994a) (the "stationary bootstrap"), among others. As it turns out, the moving blocks bootstrap is robust to heterogeneity. Lahiri (1992) gives conditions ensuring the second order correctness of Künsch's bootstrap for the normalized sample mean of observations that are not necessarily stationary. More recently, Fitzenberger (1997) has shown that the moving blocks method can be validly applied to heterogeneous mixing processes in the context of linear regressions and quantile regressions. Similarly, Politis et al. (1997) have shown the validity of certain subsampling methods for heterogeneous mixing processes.

For applications in economics, mixing is too strong a dependence condition to be broadly applicable. Andrews (1984) gives an example of a simple AR(1) process that fails to be strong mixing. The need to accommodate such time series motivates the use of functions of mixing processes, the so-called processes near epoch dependent (NED) on an underlying mixing process (Billingsley, 1968; McLeish, 1975; Gallant and White, 1988). NED processes allow for considerable heterogeneity as well as dependence and include the mixing processes as a special case. An important example of the usefulness of near epoch dependence in economics concerns the standard ARCH (Engle, 1982) and GARCH (Bollerslev, 1986) processes widely used in economics and finance, for which the mixing properties are currently known only under certain restrictive assumptions (Carrasco and Chen, 2002). As Hansen (1991a) and Sin and White (1996) have shown, ARCH and GARCH processes are processes NED on an underlying mixing process, under mild regularity conditions. The NED concept thus makes possible a convenient theory of inference for these models that would otherwise be unavailable. The usefulness of NED is further illustrated by Davidson (2002), who establishes the NED property of a variety of nonlinear time series models, including the bilinear, GARCH and threshold autoregressive models.

Presently, conditions ensuring the validity of bootstrap methods for the sample mean of (possibly heterogeneous) NED functions of mixing processes are unknown. Our goal here is thus to establish the validity of the bootstrap in this context, extending the applicability of bootstrap methods to a class of processes broadly relevant for applications in economics and finance. As is usual in the bootstrap literature, establishing the validity of the bootstrap for the sample mean is an important step towards establishing its validity for more complicated statistics. In Gonçalves and White (2000) we build on the results given here to prove the validity of the bootstrap for general extremum estimators such as quasi-maximum likelihood and generalized method of moments estimators.

Our results apply to block bootstrap methods. Not only do they apply to the moving blocks bootstrap (MBB) scheme of Künsch (1989) and Liu and Singh (1992), which has not been studied with the degree of dependence considered here, but also to the stationary bootstrap (SB) of Politis and Romano (1994a), which has not yet been studied in a heterogeneous context. Our motivation in the latter case is to show that what is important about the stationary bootstrap is not its stationarity but that it is a bootstrap. In particular, we show that the consistency of moving block and stationary bootstrap variance estimators for the sample mean is robust against heteroskedasticity and dependence of unknown form. We also establish the first order asymptotic validity of the bootstrap approximation to the actual distribution of the sample mean in this heterogeneous, near epoch dependent context.

The main theoretical results are given in Section 2 and Section 3 concludes. An Appendix contains mathematical proofs.

2. Main Results

Suppose $\{X_{nt}, n, t = 1, 2, ...\}$ is a double array of not necessarily stationary (heterogeneous) random $d \times 1$ vectors defined on a given probability space $(\Omega, \mathcal{F}, P)$. By assuming that $\{X_{nt}\}$ is near epoch dependent on a mixing process, we permit a considerable degree of dependence and heterogeneity and include mixing processes as a special case. We define $\{X_{nt}\}$ to be NED on a mixing process $\{V_t\}$

provided $\|X_{nt}\|_2 < \infty$ and $v_k \equiv \sup_{n,t} \left\| X_{nt} - E_{t-k}^{t+k}(X_{nt}) \right\|_2$ tends to zero as $k \to \infty$ at an appropriate rate. Here and in what follows, $\|X\|_q = \left(\sum_i E|X_i|^q\right)^{1/q}$ for $q \geq 1$ denotes the L_q-norm of a random matrix 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 $v_k = O\left(k^{-a-\delta}\right)$ for some $\delta > 0$ we say $\{X_{nt}\}$ is NED (on $\{V_t\}$) of size $-a$. The sequence $\{V_t\}$ is assumed to be strong mixing although analogous results could be derived under the assumption of uniform mixing. 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.

Let $\mu_{nt} \equiv E(X_{nt})$ for $t = 1, 2, \ldots, n$, $n = 1, 2, \ldots$, and let $\bar{\mu}_n \equiv n^{-1} \sum_{t=1}^n \mu_{nt}$ be the vector of interest to be estimated by the multivariate sample mean $\bar{X}_n \equiv n^{-1} \sum_{t=1}^n X_{nt}$. Related studies such as Fitzenberger (1997) and Politis et al. (1997) have assumed common means across observations, $\mu_{nt} = \mu$, in which case $\bar{\mu}_n$ is just μ. Instead, we will assume that the population means μ_{nt} satisfy a less stringent homogeneity condition in order to establish our main consistency result.

Our goal is to conduct inference on $\bar{\mu}_n$ based on a realization of $\{X_{nt}\}$. Alternatively, we may be interested in constructing a confidence region for $\bar{\mu}_n$ or in computing an estimate of the covariance matrix of its estimator, the sample mean $\bar{X}_n$. The bootstrap can be used for these purposes.

We follow Lahiri (1999) in describing the block bootstrap methods of interest here. Let $\ell = \ell_n \in \mathbb{N}$ $(1 \leq \ell < n)$ denote the (expected) length of the blocks and let $B_{t,\ell} = \{X_{nt}, X_{n,t+1}, \ldots, X_{n,t+\ell-1}\}$ be the block of ℓ consecutive observations starting at X_{nt}; $\ell = 1$ corresponds to the standard bootstrap. Assume for simplicity that $n = k\ell$. The MBB resamples $k = n/\ell$ blocks randomly with replacement from the set of $n - \ell + 1$ overlapping blocks $\{B_{1,\ell}, \ldots, B_{n-\ell+1,\ell}\}$. Thus, if we let $I_{n1}, \ldots, I_{nk}$ be i.i.d. random variables uniformly distributed on $\{0, \ldots, n-\ell\}$, the MBB pseudo-time series $\left\{X_{nt}^{*(1)}, t = 1, \ldots, n\right\}$ is the result of arranging the elements of the k resampled blocks $B_{I_{n1}+1,\ell}, \ldots, B_{I_{nk}+1,\ell}$ in a sequence: $X_{n1}^{*(1)} = X_{n,I_{n1}+1}, X_{n2}^{*(1)} = X_{n,I_{n1}+2}, \ldots, X_{n\ell}^{*(1)} = X_{n,I_{n1}+\ell}, X_{n,\ell+1}^{*(1)} = X_{n,I_{n2}+1}, \ldots, X_{n,kl}^{*(1)} = X_{n,I_{nk}+\ell}$. Here and throughout, we use the superscript (1) in $X_{n,t}^{*(1)}$ to denote the bootstrap samples obtained by the MBB. Similarly, we will use the superscript (2) to denote bootstrap samples obtained by the SB resampling scheme.

Unlike the MBB, the stationary bootstrap resamples blocks of random size. Let $p = \ell^{-1}$ be a given number in $(0,1]$; $p = 1$ corresponds to the standard bootstrap. Let $L_{n1}, L_{n2}, \ldots$ be conditionally i.i.d. random variables having the geometric distribution with parameter p so that the probability of the event $\{L_{n1} = k\}$ is $(1-p)^{k-1} p$ for $k = 1, 2, \ldots$. Independent of $\{X_{nt}\}$ and of $\{L_{nt}\}$, let $I_{n1}, I_{n2}, \ldots$ be i.i.d. random variables having the uniform distribution on $\{1, \ldots, n\}$. The SB pseudo-time series $\left\{X_{nt}^{*(2)}\right\}$ can be obtained by joining the resampled blocks $B_{I_{n1},L_{n1}}, B_{I_{n2},L_{n2}}, \ldots, B_{I_{nK},L_{nK}}$, where $K = \inf\{k \geq 1 : L_{n1} + \ldots + L_{nk} \geq n\}$. Thus, the stationary bootstrap amounts to resampling blocks of observations of random length, where each block size has a geometric distribution with parameter p and expected length equal to $\frac{1}{p} = \ell$. We shall require $\ell = \ell_n$ to tend to infinity at an appropriate rate, which is equivalent to letting $p = p_n$ tend to zero. Hence, on average the lengths of the SB blocks tend to infinity with n as also holds for the (fixed) MBB block lengths.

In contrast to the MBB resampling method, the stationary bootstrap resample is a strictly stationary process (Politis and Romano, 1994a), conditional on the original data. As we show, this stationarity does not restrict its applicability solely to stationary processes.

Given the bootstrap resample $\{X_{n1}^{*(j)}, \ldots, X_{nn}^{*(j)}\}$, $j = 1, 2$, one can compute a resampled version of the statistic of interest, $\bar{X}_n^{*(j)} \equiv n^{-1} \sum_{t=1}^{n} X_{nt}^{*(j)}$. For stationary α-mixing processes, Künsch (1989) and Politis and Romano (1994a) show that their block bootstrap "works". As a consequence, by repeating this procedure a large number B of times, one can approximate the true distribution of $\sqrt{n}(\bar{X}_n - \bar{\mu}_n)$ by the approximate sampling distribution of $\sqrt{n}(\bar{X}_n^{*(j)} - \bar{X}_n)$, conditional on the original data, given by the empirical distribution of the B draws of $\sqrt{n}(\bar{X}_n^{*(j)} - \bar{X}_n)$. Likewise, an estimate of the covariance matrix of the scaled sample mean $\Sigma_n \equiv \text{var}\left(\sqrt{n}\bar{X}_n\right)$ is easily obtained by using the bootstrap covariance matrix $\hat{\Sigma}_{n,j} = \text{var}^*\left(\sqrt{n}\bar{X}_n^{*(j)}\right)$. (Here and in the following, a star appearing on E (var) denotes expectation (variance) with respect to $X_{n1}^{*(j)}, \ldots, X_{nn}^{*(j)}$ conditional on the data $X_{n1}, \ldots, X_{nn}$.) The goal of this section is to extend these results to the heterogeneous NED case.

Our first result establishes the consistency of the block bootstrap covariance matrix estimators for the sample mean when the observations are near epoch dependent on a mixing process. As is well known, neither of these bootstrap covariance estimators require resampling the observations. Indeed, following Künsch (1989, Theorems 3.1 and 3.4), the following formula for $\hat{\Sigma}_{n,1}$ is available:

$$\hat{\Sigma}_{n,1} = \tilde{R}_n(0) + \sum_{\tau=1}^{\ell-1} \left(1 - \frac{\tau}{\ell}\right) \left(\tilde{R}_n(\tau) + \tilde{R}_n'(\tau)\right), \tag{2.1}$$

where

$$\tilde{R}_n(\tau) = \sum_{t=1}^{n-\tau} \beta_{n,t,\tau} \left(X_{nt} - \bar{X}_{\gamma,n}\right)\left(X_{n,t+\tau} - \bar{X}_{\gamma,n}\right)',$$

and $\bar{X}_{\gamma,n} = \sum_{t=1}^{n} \gamma_{nt} X_{nt}$. The weights γ_{nt} and $\beta_{n,t,\tau}$ are given as follows (cf. Künsch, 1989, expressions (3.2) and (3.7)):

$$\gamma_{nt} = \begin{cases} \frac{t}{\ell(n-\ell+1)}, & \text{if } t \in \{1, \ldots, \ell-1\} \\ \frac{1}{n-\ell+1}, & \text{if } t \in \{\ell, \ldots, n-\ell+1\} \\ \frac{n-t+1}{\ell(n-\ell+1)}, & \text{if } t \in \{n-\ell+2, \ldots, n\}, \end{cases} \tag{2.2}$$

and

$$\beta_{n,t,\tau} = \begin{cases} \frac{t}{(\ell-|\tau|)(n-\ell+1)}, & \text{if } t \in \{1, \ldots, \ell-|\tau|-1\} \\ \frac{1}{n-\ell+1}, & \text{if } t \in \{\ell-|\tau|, \ldots, n-\ell+1\} \\ \frac{n-t-|\tau|+1}{(\ell-|\tau|)(n-\ell+1)}, & \text{if } t \in \{n-\ell+2, \ldots, n-|\tau|\}, \end{cases} \tag{2.3}$$

where $\sum_{t=1}^{n} \gamma_{nt} = 1$ and $\sum_{t=1}^{n-|\tau|} \beta_{n,t,\tau} = 1$.

Similarly, the SB covariance estimator can be calculated with the following formula (cf. Politis and Romano, 1994a, Lemma 1):

$$\hat{\Sigma}_{n,2} = \hat{R}_n(0) + \sum_{\tau=1}^{n-1} b_{n\tau} \left(\hat{R}_n(\tau) + \hat{R}_n'(\tau)\right), \tag{2.4}$$

where $\hat{R}_n(\tau)$ is the usual cross-covariance matrix estimator at lag τ, i.e. $\hat{R}_n(\tau) = n^{-1} \sum_{t=1}^{n-\tau} \left(X_{nt} - \bar{X}_n\right)\left(X_{n,t+\tau} - \bar{X}_n\right)'$ and

$$b_{n\tau} = \left(1 - \frac{\tau}{n}\right)(1-p)^{\tau} + \frac{\tau}{n}(1-p)^{n-\tau}$$

is the Politis and Romano (1994a) weight, with smoothing parameter $p_n = p \equiv \ell^{-1}$.

As is evident from (2.1) and (2.4), the MBB and the SB covariance matrix estimators for the sample mean are closely related to a lag window estimator of the spectral density matrix at frequency zero. In particular, as remarked in the univariate context by Fitzenberger (1997) and by Politis and Romano (1994a), the MBB variance estimator $\hat{\Sigma}_{n,1}$ is approximately equivalent to the Bartlett kernel variance estimator considered by Newey and West (1987). Politis and Romano (1994a) also discuss the relation between $\hat{\Sigma}_{n,1}$ and $\hat{\Sigma}_{n,2}$. They offer an interpretation for the SB variance estimator as a weighted average over ℓ of MBB variance estimators with fixed length ℓ, which suggests that $\hat{\Sigma}_{n,2}$ should be less sensitive to the choice of p than $\hat{\Sigma}_{n,1}$ is to the choice of ℓ. See Hall et al. (1995), Politis et al. (1997), Fitzenberger (1997), Horowitz (1999) and Politis and White (2001) for discussion of the important issue of blocksize choice in the context of mixing observations. We conjecture that the existing results on the optimal block choice for mixing data generalize straightforwardly to the NED context, although a proof is beyond the scope of this paper. In a recent theoretical study, Lahiri (1999) compares several block bootstrap variance estimators, including the MBB and the SB. He concludes that although (the univariate analogs of) $\hat{\Sigma}_{n,1}$ and $\hat{\Sigma}_{n,2}$ have the same asymptotic bias, the variance of $\hat{\Sigma}_{n,2}$ is larger than that of $\hat{\Sigma}_{n,1}$, suggesting that the SB method is asymptotically less efficient than the MBB. We note also that Horowitz (2001) discusses bootstrap methods for time series data that may perform even better than the MBB for certain types of processes (e.g. processes known to be finite order autoregressive processes).

Assumption 2.1 is used to establish our main consistency theorem.

Assumption 2.1

2.1.a) For some $r > 2$, $\|X_{nt}\|_{3r} \leq \Delta < \infty$ for all $n, t = 1, 2, \ldots$.

2.1.b) $\{X_{nt}\}$ is near epoch dependent (NED) on $\{V_t\}$ with NED coefficients v_k of size $-\frac{2(r-1)}{(r-2)}$; $\{V_t\}$ is an α-mixing sequence with α_k of size $-\frac{2r}{r-2}$.

As Theorem 2.1 below shows, under arbitrary heterogeneity in $\{X_{nt}\}$ the block bootstrap covariance estimators $\hat{\Sigma}_{n,j}$, $j = 1, 2$, are not consistent for Σ_n, but for $\Sigma_n + U_{n,j}$. The bias term $U_{n,j}$ is related to the heterogeneity in the means $\{\mu_{nt}\}$ and can be interpreted as the block bootstrap covariance matrix of $\sqrt{n}\bar{\mu}_n^{*(j)} = n^{-1/2} \sum_{t=1}^{n} \mu_{nt}^{*(j)}$ that would result if we could resample the vector time series $\{\mu_{nt}\}$. We call $\{\mu_{nt}^{*(j)}\}$ the "resampled version" of $\{\mu_{nt}\}$.

Theorem 2.1. *Assume $\{X_{nt}\}$ satisfies Assumption 2.1. Then, if $\ell_n \to \infty$ and $\ell_n = o\left(n^{1/2}\right)$, for $j = 1, 2$, $\hat{\Sigma}_{n,j} - (\Sigma_n + U_{n,j}) \xrightarrow{P} 0$, where $U_{n,j} \equiv \text{var}^* \left(n^{-1/2} \sum_{t=1}^{n} \mu_{nt}^{*(j)}\right)$ and $\mu_{nt}^{*(j)}$ is the resampled version of μ_{nt}.*

Theorem 2.1 makes clear that a necessary condition for the consistency of $\hat{\Sigma}_{n,j}$ for Σ_n is that $U_{n,j} \to 0$ as $n \to \infty$. A sufficient condition for $U_{n,j}$ to vanish is first order stationarity of $\{X_{nt}\}$: if $\mu_{nt} = \mu$ for all n, t , then $U_{n,j} = 0$. We have the following lemma.

Lemma 2.1. *If we let* $\bar{\mu}_{\gamma,n} = \sum_{t=1}^{n} \gamma_{nt}\mu_{nt}$*, then*

$$\begin{aligned} U_{n,1} &= \sum_{t=1}^{n} \beta_{n,t,0} \left(\mu_{nt} - \bar{\mu}_{\gamma,n}\right)\left(\mu_{nt} - \bar{\mu}_{\gamma,n}\right)' \\ &\quad + \sum_{\tau=1}^{\ell-1} \left(1 - \frac{\tau}{\ell}\right) \sum_{t=1}^{n-\tau} \beta_{n,t,\tau} \left[\left(\mu_{nt} - \bar{\mu}_{\gamma,n}\right)\left(\mu_{n,t+\tau} - \bar{\mu}_{\gamma,n}\right)' + \left(\mu_{n,t+\tau} - \bar{\mu}_{\gamma,n}\right)\left(\mu_{nt} - \bar{\mu}_{\gamma,n}\right)'\right], \\ \text{and } U_{n,2} &= n^{-1} \sum_{t=1}^{n} \left(\mu_{nt} - \bar{\mu}_{n}\right)\left(\mu_{nt} - \bar{\mu}_{n}\right)' \\ &\quad + \sum_{\tau=1}^{n-1} b_{n\tau} n^{-1} \sum_{t=1}^{n-\tau} \left[\left(\mu_{nt} - \bar{\mu}_{n}\right)\left(\mu_{n,t+\tau} - \bar{\mu}_{n}\right)' + \left(\mu_{n,t+\tau} - \bar{\mu}_{n}\right)\left(\mu_{nt} - \bar{\mu}_{n}\right)'\right]. \end{aligned}$$

Thus, the condition $\lim_{n\to\infty} U_{n,j} = 0$, $j = 1, 2$, can be interpreted as an homogeneity condition on the means. The following assumption ensures $\lim_{n\to\infty} U_{n,j} = 0$, $j = 1, 2$.

Assumption 2.2 $n^{-1} \sum_{t=1}^{n} \left(\mu_{nt} - \bar{\mu}_{n}\right)\left(\mu_{nt} - \bar{\mu}_{n}\right)' = o\left(\ell_n^{-1}\right)$, where $\ell_n = o\left(n^{1/2}\right)$.

Examples of the processes that satisfy Assumption 2.2 are those with a finite number of properly behaved breaks in mean, such as $X_t = \mu_1 1\left(t \le k_1\right) + \mu_2 1\left(k_1 < t \le k_2\right) + \mu_3 1\left(k_2 < t \le n\right) + Y_t$, where Y_t is a first order stationary process with mean zero, and $1\left(\cdot\right)$ denotes the indicator function. Assumption 2.2 entails that the break points k_1 and k_2 satisfy $k_1 = [n^{1/2}\tau_1]$ and $k_2 = [n^{1/2}\tau_2]$, where τ_1 and τ_2 are fixed constants in $(0, 1)$. Note that the split of observations between regimes is such that the last regime predominates, in that it has the largest proportion of the observations in the sample.

The following consistency result holds under Assumptions 2.1 and 2.2 and is an immediate consequence of the previous remark.

Corollary 2.1. *Assume* $\{X_{nt}\}$ *satisfies Assumptions 2.1 and 2.2. Then, if* $\ell_n \to \infty$ *and* $\ell_n = o\left(n^{1/2}\right)$*,* $\hat{\Sigma}_{n,j} - \Sigma_n \xrightarrow{P} 0$, $j = 1, 2$.

This result extends the previous consistency results by Künsch (1989) and Politis and Romano (1994a) to the case of dependent heterogeneous double arrays of random vectors, where the stationary mixing assumption is replaced by the more general assumption of a (possibly heterogeneous) double array near epoch dependent on a mixing process.

In particular, for $j = 2$, Corollary 2.1 contains a version of Politis and Romano's (1994a) Theorem 1 as a special case. Consider a strictly stationary α-mixing sequence $\{X_1, \ldots, X_n\}$ satisfying Assumption 2.1. Because a mixing process is trivially near epoch dependent on itself, the NED requirement is automatically satisfied. Corollary 2.1 achieves the same conclusion as Politis and Romano's (1994) Theorem 1 under the same moment conditions but weaker α-mixing size conditions ($\alpha_k = O\left(k^{-\lambda}\right)$ for some $\lambda > \frac{2r}{r-2}$ and $r > 2$ here in contrast to $\alpha_k = O\left(k^{-\lambda}\right)$ for some $\lambda > \frac{3(6+\varepsilon)}{\varepsilon}$ and $\varepsilon > 0$ there). We allow more dependence here, with the familiar trade-off between moment and memory conditions. Nevertheless, we require the stronger condition that $\ell_n = o\left(n^{1/2}\right)$, i.e. $n^{1/2}p_n \to \infty$ (with $p_n = \ell_n^{-1} \to 0$), while Politis and Romano (1994a) only require $\ell_n = o(n)$, i.e. $np_n \to \infty$. Imposing stationarity in our framework will ensure that $\Sigma_n \to \Sigma_\infty$ as $n \to \infty$, where $\Sigma_\infty = \text{var}(X_1) + 2\sum_{\tau=1}^{\infty} \text{var}(X_1, X_{1+\tau})$; hence, $\hat{\Sigma}_{n,2} \to \Sigma_\infty$ in probability, as Politis and Romano (1994a) conclude.

Similarly, for $j = 1$, our Corollary 2.1 specializes to Künsch's (1989) Corollary 3.1 when $\{X_t\}$ is a stationary α-mixing sequence, under the same moment conditions and weaker α-mixing conditions, but under the stronger requirement that $\ell_n = o\left(n^{1/2}\right)$ instead of $\ell_n = o(n)$. In particular, we show that the variance of $\hat{\Sigma}_{n,1}$ is of order $O\left(\frac{\ell^2}{n}\right)$, instead of Künsch's sharper result $O\left(\frac{\ell}{n}\right)$, which explains the loss of $\ell = o(n)$.

The next theorem establishes the first order asymptotic equivalence between the moving blocks and stationary bootstrap distributions and the limiting normal distribution for the multivariate sample mean. A slightly stronger dependence assumption is imposed to achieve this result. Specifically, we require $\{X_{nt}\}$ to be $L_{2+\delta}-$NED on a mixing process (see Andrews (1988)). We strengthen Assumption 2.1.b) slightly:

2.1.b′) For some small $\delta > 0$, $\{X_{nt}\}$ is $L_{2+\delta}-$NED on $\{V_t\}$ with NED coefficients v_k of size $-\frac{2(r-1)}{(r-2)}$; $\{V_t\}$ is an α-mixing sequence with α_k of size $-\frac{(2+\delta)r}{r-2}$.

Theorem 2.2. *Assume $\{X_{nt}\}$ satisfies Assumptions 2.1 and 2.2 strengthened by 2.1.b′). Let $\Sigma_n \equiv \text{var}\left(\sqrt{n}\bar{X}_n\right)$ be positive definite uniformly in n, i.e. Σ_n is positive semidefinite for all n and $\det \Sigma_n \geq \kappa > 0$ for all n sufficiently large. Then $\Sigma_n = O(1)$ and*

(i) $\Sigma_n^{-1/2}\sqrt{n}\left(\bar{X}_n - \bar{\mu}_n\right) \Rightarrow N\left(0, I_d\right)$ *under* P, *and*

$$\sup_{x\in\mathbb{R}^d}\left|P\left[\sqrt{n}\ \Sigma_n^{-1/2}\left(\bar{X}_n - \bar{\mu}_n\right) \leq x\right] - \Phi\left(x\right)\right| \to 0,$$

where Φ *is the standard multivariate normal cumulative distribution function, "* $\leq$ *" applies to each component of the relevant vector and "* $\Rightarrow$ *" denotes convergence in distribution.* *Moreover, if* $\ell_n \to \infty$ *and* $\ell_n = o\left(n^{1/2}\right)$, *then for any* $\varepsilon > 0$, *and for* $j = 1, 2$,

(ii) $\Sigma_n^{-1/2}\sqrt{n}\left(\bar{X}_n^{*(j)} - \bar{X}_n\right) \Rightarrow N\left(0, I_d\right)$ *under* P^* *with probability* P *approaching one, and*

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

where P^* *is the probability measure induced by the bootstrap, conditional on* $\{X_{nt}\}_{t=1}^n$.

For $j = 2$, this is an extension of Theorem 3 of Politis and Romano (1994a) for stationary mixing observations to the case of NED functions of a mixing process. For $j = 1$ and $d = 1$, Theorem 2.2 states a weaker conclusion than does Theorem 3.5 of Künsch (1989), since we prove convergence in probability, but not almost sure convergence. On the other hand, we permit heterogeneity and greater dependence.

In part (i) we state the usual asymptotic normality result for the multivariate sample mean. In part (ii) we prove the uniform convergence to zero (in probability) of the discrepancy between the actual distribution of $\sqrt{n}\left(\bar{X}_n - \bar{\mu}_n\right)$ and the block bootstrap approximation to it. This result follows from the fact that under our assumptions the distribution of $\Sigma_n^{-1}\sqrt{n}\left(\bar{X}_n^{*(j)} - \bar{X}_n\right)$, conditional on $X_{n1}, \ldots, X_{nn}$, converges weakly to the standard multivariate normal distribution for all double arrays $\{X_{nt}\}$ in a set with probability tending to one. In particular, we apply a central limit theorem for triangular arrays and use Assumption 2.1.b′) to ensure that Lyapounov's condition is satisfied under our heterogeneous NED context. Assumption 2.1.b) might well be sufficient to verify the weaker Lindeberg condition, as in Künsch (1989, Theorem 3.5), although we have not verified this.

Fitzenberger (1997) has recently proven the consistency of the moving blocks bootstrap approximation to the true sampling distribution of the sample mean for heterogeneous α-mixing processes. Our result extends his by allowing for near epoch dependence on mixing processes. However, in the purely strong mixing case, our moment and memory conditions are more stringent than his. In particular, his Theorem 3.1 only requires $E\left|X_t\right|^{2p+\delta} < C$, for small $\delta > 0$ and $p > 2$, and $\{X_t\}$ strong mixing of size $-\frac{p}{p-2}$. Politis et al. (1997, Theorem 3.1) have also established the robustness of the subsampling method for consistent sampling distribution estimation for heterogeneous and dependent data under mild moment conditions ($E\left|X_t\right|^{2+2\varepsilon} \leq \Delta < \infty$ for some $\varepsilon > 0$). Nevertheless, they also assume a strong mixing process, asymptotic covariance stationarity, and slightly stronger size requirements than ours on the mixing coefficients (α_k of size $-\frac{3(4+\varepsilon)}{\varepsilon}$).

A well known property of the MBB statistic $\sqrt{n}\left(\bar{X}_n^{*(1)} - \bar{X}_n\right)$ is that its (conditional) expected value is not zero. Indeed, under the MBB resampling scheme $E^*\left(\bar{X}_n^{*(1)}\right) = \sum_{t=1}^n \gamma_{nt} X_{nt}$, where the weights γ_{nt} are defined as in (2.2). If $\ell_n = o(n)$, one can show that $E^*\left(\bar{X}_n^{*(1)}\right) = \bar{X}_n + O_P\left(\frac{\ell_n}{n}\right)$ (see e.g. Lemma A.1 of Fitzenberger, 1997). Thus, the MBB distribution has a random bias $\sqrt{n}\left(E^*\left(\bar{X}_n^{*(1)}\right) - \bar{X}_n\right)$, which is of order $O_P\left(\frac{\ell_n}{n^{1/2}}\right)$. (For the SB no such problem exists since $E^*\left(\bar{X}_n^{*(2)}\right) = \bar{X}_n$.) As pointed out by Lahiri (1992), this random bias becomes predominant for second order analysis and prevents the MBB from providing second order improvements over the standard normal approximation. To correct for this bias, he suggests recentering the MBB distribution around the bootstrap mean, that is, to consider the bootstrap distribution of $\sqrt{n}\left(\bar{X}_n^{*(1)} - E^*\left(\bar{X}_n^{*(1)}\right)\right)$. The following result shows that under the assumptions of Theorem 2.2 recentering the MBB distribution around the MBB bootstrap mean is asymptotically valid (to first order) in this heterogeneous NED context.

Corollary 2.2. *Under the assumptions of Theorem 2.2, for all $\varepsilon > 0$,*

$$P\left\{\sup_{x \in \mathbb{R}^d}\left|P^*\left[\sqrt{n}\left(\bar{X}_n^{*(1)} - E^*\left(\bar{X}_n^{*(1)}\right)\right) \leq x\right] - P\left[\sqrt{n}\left(\bar{X}_n - \bar{\mu}_n\right) \leq x\right]\right| > \varepsilon\right\} \to 0,$$

where P^* is the probability measure induced by the MBB bootstrap, conditional on $\{X_{nt}\}_{t=1}^n$.

Theorem 2.2 and Corollary 2.2 justify the use of the MBB and SB distributions to obtain an asymptotically valid confidence interval for $\bar{\mu}_n$ instead of using a consistent estimate of the variance along with the normal approximation. For example, an equal tailed $(1-\alpha)\,100\%$ stationary bootstrap confidence interval for $\bar{\mu}_n$ when $d = 1$ would be $[\bar{X}_n - q_n^*\left(1 - \frac{\alpha}{2}\right), \bar{X}_n - q_n^*\left(\frac{\alpha}{2}\right)]$, where $q_n^*\left(\frac{\alpha}{2}\right)$ and $q_n^*\left(1 - \frac{\alpha}{2}\right)$ are the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles of the SB bootstrap distribution of $\bar{X}_n^{*(2)} - \bar{X}_n$.

3. Conclusion

In this paper we establish the first order asymptotic validity of block bootstrap methods for the sample mean of dependent heterogeneous data. Our results apply to the moving blocks bootstrap of Künsch (1989) and Liu and Singh (1992) as well as to the stationary bootstrap of Politis and Romano (1994a). In particular, we show that the MBB and the SB covariance estimators for the multivariate sample mean are consistent under a wide class of data generating processes, the processes near epoch dependent on a mixing process. We also prove the first order asymptotic equivalence between the block bootstrap distributions and the normal distribution in this heterogeneous near epoch dependent context.

Simulation results for the case of a multivariate linear regression (not reported here) reveal that the block bootstrap methods compare fairly well with standard kernel methods, in particular delivering better inferences for cases in which there is relatively high neglected serial correlation in the regression errors.

A. Appendix

To conserve space, this appendix contains abbreviated versions of the proofs of our results. A version of this appendix containing detailed proofs is available from the authors upon request.

In the proofs for simplicity we will consider $\{X_{nt}\}$ to be real-valued (i.e. $d = 1$). The results for the multivariate case follow by showing that the assumptions are satisfied for linear combinations $Y_{nt} \equiv \lambda' X_{nt}$ for any non-zero $\lambda \in \mathbb{R}^d$. Throughout the Appendix, K will denote a generic constant that may change from one usage to another. Furthermore, we shall use the notation $E_s^t(\cdot) = E\left(\cdot|\mathcal{F}_s^t\right)$ for $t \geq s$, where $\mathcal{F}_s^t = \sigma\left(V_t, \ldots, V_s\right)$, with t or s omitted to denote $+\infty$ and $-\infty$, respectively. $[x]$ will denote the integer part of x. Finally, the subscript n in ℓ_n and p_n will be implicitly understood throughout.

The mixingale property of zero mean NED processes on a mixing process is an important tool in obtaining our results. See, for example, Davidson (1994, Definition 16.5) for a definition of a mixingale. We will make use of the following lemmas in our proofs.

Lemma A.1. *(i) If $\{X_{nt}\}$ satisfies Assumption 2.1, $\left\{X_{nt} - \mu_{nt}, \mathcal{F}^t\right\}$ is an L_2-mixingale of size $-\frac{3r-2}{3(r-2)}$, $r > 2$, with uniformly bounded mixingale constants c_{nt}. (ii) If $\{X_{nt}\}$ satisfies Assumption 2.1 strengthened by Assumption 2.1.b$'$) with $\delta \leq 4$ and $r > 2$, then $\left\{X_{nt} - \mu_{nt}, \mathcal{F}^t\right\}$ is an $L_{2+\delta}$-mixingale of size*

$-\frac{3r-(2+\delta)}{3(r-2)}$, *with uniformly bounded mixingale constants* c_{nt}.

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

Lemma A.3. *Let* $\{Z_{nt}, \mathcal{F}^t\}$ *be an* L_p-*mixingale for some* $p \geq 2$ *with mixingale constants* c_{nt} *and mixingale coefficients* ψ_k *satisfying* $\sum_{k=1}^{\infty}\psi_k < \infty$. *Then,* $\left\|\max_{j\leq n}\left|\sum_{t=1}^{j} Z_{nt}\right|\right\|_p = O\left(\left(\sum_{t=1}^{n} c_{nt}^2\right)^{1/2}\right)$.

Lemma A.1 follows from Corollary 17.6.(i) of Davidson (1994), Lemma A.2 is Theorem 1.6 of McLeish (1975), and Lemma A.3 is a straightforward generalization of Hansen's (1991b) maximal inequality for L_p-mixingales, with[1] $p \geq 2$, to the double array setting. The following lemma generalizes Lemma 6.7 (a) in Gallant and White (1988, pp. 99-100).

Lemma A.4. *Assume* X_{nt} *is such that* $E(X_{nt}) = 0$ *and* $\|X_{nt}\|_{3r} \leq \Delta < \infty$ *for some* $r > 2$ *and for all* n, t. *If* $\{X_{nt}\}$ *is NED on* $\{V_t\}$ *and* $\{V_t\}$ *is* α*-mixing, then for fixed* $\tau > 0$ *and all* $t < s \leq t + \tau$,

$$|cov(X_{nt}X_{n,t+\tau}, X_{ns}X_{n,s+\tau})| \leq K_1\left(\alpha_{\left[\frac{s-t}{4}\right]}^{\frac{1}{2}-\frac{1}{r}} + v_{\left[\frac{s-t}{4}\right]}\right) + K_2 v_{\left[\frac{s-t}{4}\right]}^{\frac{r-2}{2(r-1)}} + K_3\left(\alpha_{\left[\frac{\tau}{4}\right]}^{\left(\frac{1}{2}-\frac{1}{r}\right)} + v_{\left[\frac{\tau}{4}\right]}\right)^2,$$

for some finite constants K_1, K_2 *and* K_3.

Proof. First, note that $|\text{cov}(X_{nt}X_{nt+\tau}, X_{ns}X_{n,s+\tau})| \leq |E(X_{nt}X_{n,t+\tau}X_{ns}X_{n,s+\tau})| + |E(X_{nt}X_{n,t+\tau})E(X_{ns}X_{n,s+\tau})|$. Next, note that $|E(X_{nt}X_{n,t+\tau})| \leq \Delta\left(5\Delta\alpha_{\left[\frac{\tau}{4}\right]}^{1/2-1/r} + 2v_{\left[\frac{\tau}{4}\right]}\right)$ (see Gallant and White, 1988, pp. 109-110), which implies the last bound on the covariance. To bound $|E(X_{nt}X_{n,t+\tau}X_{ns}X_{n,s+\tau})|$, define $m = s - t > 0$ and $\hat{Y}_{nmt\tau} \equiv E_{t-\tau-\left[\frac{m}{2}\right]}^{t+\tau+\left[\frac{m}{2}\right]}(X_{nt}X_{ns}X_{n,t+\tau})$, and note that

$$\begin{aligned}|E(X_{nt}X_{n,t+\tau}X_{ns}X_{n,s+\tau})| &\leq \left|E\left(\hat{Y}_{nmt\tau}X_{n,s+\tau}\right)\right| \\ &+ \left|E\left(X_{n,s+\tau}\left(X_{nt}X_{ns}X_{n,t+\tau} - \hat{Y}_{nmt\tau}\right)\right)\right|. \qquad \text{(A.1)}\end{aligned}$$

Since $\hat{Y}_{nmt\tau}$ is a measurable function of $\{V_{t-\tau-[m/2]}, \ldots, V_{t+\tau+[m/2]}\}$, it is measurable-$\mathcal{F}^{t+\tau+[m/2]}$. By an application of the law of iterated expectations and repeated applications of Hölder's inequality, $\left|E\left(\hat{Y}_{nmt\tau}X_{n,s+\tau}\right)\right| = \left|E\left(E^{t+\tau+[m/2]}\left(\hat{Y}_{nmt\tau}X_{n,s+\tau}\right)\right)\right| \leq \Delta^3\left\|E^{t+\tau+[m/2]}X_{n,s+\tau}\right\|_2$. To bound $\left\|E^{t+\tau+[m/2]}X_{n,s+\tau}\right\|_2$, note that $\mathcal{F}^{t+\tau+[m/2]} = \mathcal{F}^{s+\tau-k_1} \subseteq \mathcal{F}^{s+\tau-[m/2]}$ with $k_1 = m - [m/2] \geq [m/2]$, implying that $\left\|E^{t+\tau+[m/2]}X_{n,s+\tau}\right\|_2 \leq \left\|E^{s+\tau-[m/2]}X_{n,s+\tau}\right\|_2$ by Theorem 10.27 of Davidson (1994). By a similar argument as in McLeish (1975, Theorem 3.1), we have that $\left\|E^{s+\tau-[m/2]}X_{n,s+\tau}\right\|_2 \leq v_{[m/4]} + 5\Delta\alpha_{[m/4]}^{1/2-1/r}$. Thus, $\left|E\left(\hat{Y}_{nmt\tau}X_{n,s+\tau}\right)\right| \leq \Delta^3\left(5\Delta\alpha_{\left[\frac{m}{4}\right]}^{\frac{1}{2}-\frac{1}{r}} + v_{\left[\frac{m}{4}\right]}\right)$. To bound the second term

[1] Hansen (1992) gives the corrected version of this maximal inequality for $1 < p < 2$. Here, we will only use the result when $p \geq 2$, and we omit the case $1 < p < 2$.

in (A.1), by the Cauchy-Schwarz inequality, it suffices to show that $\left\|X_{nt}X_{ns}X_{n,t+\tau} - \hat{Y}_{nmt\tau}\right\|_2 \leq K v_{\left[\frac{m}{4}\right]}^{\frac{r-2}{2(r-1)}}$. Following an argument similar to that used by Davidson (1992, Lemma 3.5) and writing $E_J(\cdot)$ for $E_{t-\tau-J}^{t+\tau+J}(\cdot)$, X for X_{nt}, Y for X_{ns}, and Z for $X_{n,t+\tau}$, we obtain $\|XYZ - E_J(XYZ)\|_2 \leq \|XYZ - E_J(X)E_J(Y)E_J(Z)\|_2$ by Theorem 10.12 of Davidson (1994). Also, adding and subtracting appropriately, by the triangle inequality, $|XYZ - E_J(X)E_J(Y)E_J(Z)| \leq B\left(\mathfrak{X},\hat{\mathfrak{X}}\right)d\left(\mathfrak{X},\hat{\mathfrak{X}}\right)$, where $\mathfrak{X} = (X,Y,Z)'$; $\hat{\mathfrak{X}} = (E_J(X), E_J(Y), E_J(Z))'$; $B\left(\mathfrak{X},\hat{\mathfrak{X}}\right) = (|ZX| + |ZE_J(Y)| + |E_J(X)E_J(Y)|)$, and $d\left(\mathfrak{X},\hat{\mathfrak{X}}\right) = (|X - E_J(X)| + |Y - E_J(Y)| + |Z - E_J(Z)|)$. Now apply Lemma 4.1 of Gallant and White (1988, p. 47) with $b(\mathfrak{X}) = XYZ$, $b\left(\hat{\mathfrak{X}}\right) = E_J(X)E_J(Y)E_J(Z)$, and with $r > 2$, $q = \frac{3r}{2}$ and $p = \frac{q}{q-1} = \frac{3r}{3r-2}$. This yields $\|XYZ - E_J(XYZ)\|_2 \leq K\left(3\Delta^2\right)^{(r-2)/2(r-1)}(6v_J)^{(r-2)/2(r-1)}$. The result follows upon setting $J = [m/2] = \left[\frac{s-t}{2}\right]$, given that $v_{(\cdot)}$ is nonincreasing. ■

Proof of Theorem 2.1. The proof consists of two steps: (1) show $\tilde{\Sigma}_{n,j} - \Sigma_n \xrightarrow{P} 0$, and (2) show $\hat{\Sigma}_{n,j} - \left(\tilde{\Sigma}_{n,j} + U_{n,j}\right) \xrightarrow{P} 0$. In (1), we consider an infeasible estimator $\tilde{\Sigma}_{n,j}$ which is identical to $\hat{\Sigma}_{n,j}$ except that it replaces $\bar{X}_{\gamma,n}$ and $\bar{X}_n$ in (2.1) and (2.4) with μ_{nt} for $j = 1, 2$, respectively.

Proof of step 1. $(j = 1)$: Define $Z_{nt} \equiv X_{nt} - \mu_{nt}$ and $R_{nt}(\tau) = E(Z_{nt}Z_{n,t+\tau})$. By the triangle inequality,

$$\left|E\left(\tilde{\Sigma}_{n,1}\right) - \Sigma_n\right| \leq \sum_{t=1}^{n}\left|\beta_{n,t,0} - n^{-1}\right||R_{nt}(0)| + 2\sum_{\tau=1}^{\ell-1}\sum_{t=1}^{n-\tau}\left|\beta_{n,t,\tau} - n^{-1}\right||R_{nt}(\tau)| \tag{A.2}$$

$$+ 2\sum_{\tau=1}^{\ell-1}\frac{\tau}{\ell}\sum_{t=1}^{n-\tau}\left|\beta_{n,t,\tau}\right||R_{nt}(\tau)| + 2\sum_{\tau=\ell}^{n-1}n^{-1}\sum_{t=1}^{n-\tau}|R_{nt}(\tau)|. \tag{A.3}$$

The two terms in (A.2) are $O\left(\frac{\ell}{n-\ell+1}\right)$ and $O\left(\frac{\ell^2}{n-\ell+1}\right)$ respectively, and therefore they are $o(1)$. The two terms in (A.3) combined are bounded by $\frac{n}{n-\ell+1}\xi_n$, where $\frac{n}{n-\ell+1} \to 1$, and $\xi_n \equiv 2\sum_{\tau=1}^{n-1}\frac{\tau}{\ell}n^{-1}\sum_{t=1}^{n-\tau}|R_{nt}(\tau)|$ is $O\left(\ell^{-1}\right)$ given the assumed size conditions on α_k and v_k. Thus, $\lim_{n\to\infty}\left|E\left(\tilde{\Sigma}_{n,1}\right) - \Sigma_n\right| = 0$ given that $\ell_n \to \infty$ and $\ell = o\left(n^{1/2}\right)$.

To show that $\operatorname{var}\left(\tilde{\Sigma}_{n,1}\right) \to 0$, define $\tilde{R}_{n0}(\tau) = \sum_{t=1}^{n-|\tau|}\beta_{n,t,\tau}Z_{nt}Z_{n,t+|\tau|}$ and write $\operatorname{var}\left(\tilde{\Sigma}_{n,1}\right) = \sum_{\tau=-\ell+1}^{\ell-1}\sum_{\lambda=-\ell+1}^{\ell-1}\left(1 - \frac{|\tau|}{\ell}\right)\left(1 - \frac{|\lambda|}{\ell}\right)\operatorname{cov}\left(\tilde{R}_{n0}(\tau), \tilde{R}_{n0}(\lambda)\right)$. We show that $\operatorname{var}\left(\tilde{R}_{n0}(\tau)\right)$ is $O\left(\frac{n}{(n-\ell+1)^2}\right)$, which by the Cauchy-Schwarz inequality implies the result since $\sum_{\tau=-\ell+1}^{\ell-1}\sum_{\lambda=-\ell+1}^{\ell-1}\left(1 - \frac{|\tau|}{\ell}\right)\left(1 - \frac{|\lambda|}{\ell}\right) = \ell^2$ and $\ell = o\left(n^{1/2}\right)$. Following Politis and Romano (1994b), write

$$\operatorname{var}\left(\tilde{R}_{n0}(\tau)\right) = \sum_{t=1}^{n-|\tau|} \beta_{n,t,\tau}^2 \operatorname{var}\left(Z_{nt} Z_{n,t+|\tau|}\right) + 2\sum_{t=1}^{n-|\tau|} \sum_{s=t+1}^{n-|\tau|} \beta_{n,t,\tau}\beta_{n,s,\tau} \operatorname{cov}\left(Z_{nt}Z_{n,t+|\tau|}, Z_{ns}Z_{n,s+|\tau|}\right)$$

$$\leq \frac{1}{(n-\ell+1)^2} \sum_{t=1}^{n-|\tau|} \operatorname{var}\left(Z_{nt}Z_{n,t+|\tau|}\right) + \frac{2}{(n-\ell+1)^2} \sum_{t=1}^{n-|\tau|} \sum_{s=t+1}^{t+|\tau|} \left|\operatorname{cov}\left(Z_{nt}Z_{n,t+|\tau|}, Z_{ns}Z_{n,s+|\tau|}\right)\right|$$

$$+ \frac{2}{(n-\ell+1)^2} \sum_{t=1}^{n-|\tau|} \sum_{s=t+|\tau|+1}^{n-|\tau|} \left|\operatorname{cov}\left(Z_{nt}Z_{n,t+|\tau|}, Z_{ns}Z_{n,s+|\tau|}\right)\right|,$$

given that $\beta_{n,t,\tau} \leq \frac{1}{n-\ell+1}$ for all t and τ. For K sufficiently large, and given the mean zero property of $\{Z_{nt}\}$,

$$\begin{aligned}(n-\ell+1)^2 \operatorname{var}\left(\tilde{R}_{n0}(\tau)\right) \quad &\leq \quad Kn\left\{\Delta^2 + \sum_{k=1}^{\infty} \alpha_{\left[\frac{k}{4}\right]}^{\frac{1}{2}-\frac{1}{r}} + \sum_{k=1}^{\infty} v_{\left[\frac{k}{4}\right]} + \sum_{k=1}^{\infty} v_{\left[\frac{k}{4}\right]}^{\frac{r-2}{2(r-1)}}\right\} \\ &+Kn\left(|\tau|\, \alpha_{\left[\frac{|\tau|}{4}\right]}^{2\left(\frac{1}{2}-\frac{1}{r}\right)} + |\tau|\, v_{\left[\frac{|\tau|}{4}\right]}^2 + 2\,|\tau|\, \alpha_{\left[\frac{|\tau|}{4}\right]}^{\left(\frac{1}{2}-\frac{1}{r}\right)} v_{\left[\frac{|\tau|}{4}\right]}\right), \qquad \text{(A.4)}\end{aligned}$$

where we used Lemma A.4 to bound the covariances when $t < s \leq t + |\tau|$ and a result similar to Lemma 6.7 (a) in Gallant and White (1988, pp.99-100) when $s > t + |\tau|$. The sums in the curly brackets are finite, whereas the last term in (A.4) tends to 0 as $|\tau| \to \infty$ by the size assumptions on α_k and v_k, their monotonicity, and the fact that they tend to 0 as $k \to \infty$. Hence, $\operatorname{var}\left(\tilde{R}_{n0}(\tau)\right) \leq \frac{Kn}{(n-\ell+1)^2}$.

Proof of step 2. $(j = 1)$: Define $S_{n,1} = \sum_{\tau=-\ell+1}^{\ell-1}\left(1 - \frac{|\tau|}{\ell}\right)\sum_{t=1}^{n-|\tau|} \beta_{n,t,\tau} X_{nt} X_{n,t+|\tau|}$, and write

$$\hat{\Sigma}_{n,1} = S_{n,1} + \sum_{\tau=-\ell+1}^{\ell-1}\left(1-\frac{|\tau|}{\ell}\right)\sum_{t=1}^{n-|\tau|} \beta_{n,t,\tau}\left(-\bar{X}_{\gamma,n}X_{nt} - \bar{X}_{\gamma,n}X_{n,t+|\tau|} + \bar{X}_{\gamma,n}^2\right), \quad \text{and}$$

$$\tilde{\Sigma}_{n,1} = S_{n,1} + \sum_{\tau=-\ell+1}^{\ell-1}\left(1-\frac{|\tau|}{\ell}\right)\sum_{t=1}^{n-|\tau|} \beta_{n,t,\tau}\left(-\mu_{n,t+|\tau|}X_{nt} - \mu_{nt}X_{n,t+|\tau|} + \mu_{nt}\mu_{n,t+|\tau|}\right).$$

Then, $\hat{\Sigma}_{n,1} - \tilde{\Sigma}_{n,1} = A_{n1} + A_{n2} + A_{n3} + A_{n4}$, where

$$A_{n1} = -\left(\bar{X}_{\gamma,n} - \bar{\mu}_{\gamma,n}\right)\sum_{\tau=-\ell+1}^{\ell-1}\left(1-\frac{|\tau|}{\ell}\right)\sum_{t=1}^{n-|\tau|}\beta_{n,t,\tau}\left(Z_{nt} + Z_{n,t+|\tau|}\right),$$

$$A_{n2} = \sum_{\tau=-\ell+1}^{\ell-1}\left(1-\frac{|\tau|}{\ell}\right)\sum_{t=1}^{n-|\tau|}\beta_{n,t,\tau}\left(\mu_{nt} - \bar{\mu}_{\gamma,n}\right)Z_{n,t+|\tau|},$$

$$A_{n3} = \sum_{\tau=-\ell+1}^{\ell-1}\left(1-\frac{|\tau|}{\ell}\right)\sum_{t=1}^{n-|\tau|}\beta_{n,t,\tau}\left(\mu_{n,t+|\tau|} - \bar{\mu}_{\gamma,n}\right)Z_{nt},$$

$$A_{n4} = \sum_{\tau=-\ell+1}^{\ell-1} \left(1 - \frac{|\tau|}{\ell}\right) \sum_{t=1}^{n-|\tau|} \beta_{n,t,\tau} \left(\bar{X}_{\gamma,n}^2 - \left(\mu_{nt} + \mu_{n,t+|\tau|}\right) \bar{X}_{\gamma,n} + \mu_{nt}\mu_{n,t+|\tau|}\right),$$

with $\bar{\mu}_{\gamma,n} = \sum_{t=1}^{n} \gamma_{nt}\mu_{nt}$. If $\mu_{nt} = \mu$ for all t, $\bar{\mu}_{\gamma,n} = \mu$ since $\sum_{t=1}^{n} \gamma_{nt} = 1$, which implies $A_{n2} = A_{n3} = 0$ and $A_{n4} = \left(\bar{X}_{\alpha,n} - \bar{\mu}_{\alpha,n}\right)^2 \ell$, because $\sum_{t=1}^{n-|\tau|} \beta_{n,t,\tau} = 1$ for every τ. Below we show that $\bar{X}_{\gamma,n} - \bar{\mu}_{\gamma,n} = o_P\left(\ell^{-1}\right)$, which implies A_{n1} and A_{n4} are $o_P(1)$. Thus, $\hat{\Sigma}_{n,1} - \tilde{\Sigma}_{n,1} \xrightarrow{P} 0$. If μ_{nt} is not constrained to equal μ for every n, t, we obtain $A_{n4} = A'_{n4} + U_{n,1}$, where

$$\begin{aligned} A'_{n4} &= \left(\bar{X}_{\gamma,n} - \bar{\mu}_{\gamma,n}\right)^2 \ell + 2\left(\bar{X}_{\gamma,n} - \bar{\mu}_{\gamma,n}\right) \ell \bar{\mu}_{\gamma,n} \\ &\quad - \left(\bar{X}_{\gamma,n} - \bar{\mu}_{\gamma,n}\right) \sum_{\tau=-\ell+1}^{\ell-1} \left(1 - \frac{|\tau|}{\ell}\right) \sum_{t=1}^{n-|\tau|} \beta_{n,t,\tau} \left(\mu_{nt} + \mu_{n,t+|\tau|}\right), \quad \text{and} \end{aligned}$$

$$U_{n,1} = \sum_{\tau=-\ell+1}^{\ell-1} \left(1 - \frac{|\tau|}{\ell}\right) \sum_{t=1}^{n-|\tau|} \beta_{n,t,\tau} \left(\mu_{nt} - \bar{\mu}_{\gamma,n}\right) \left(\mu_{n,t+|\tau|} - \bar{\mu}_{\gamma,n}\right).$$

By Theorems 3.1 and 3.4 of Künsch (1989), $U_{n,1} = \text{var}^*\left(n^{-1/2} \sum_{t=1}^{n} \mu_{nt}^{*(1)}\right)$. Thus, it suffices to show that A_{n1}, A_{n2}, A_{n3} and A'_{n4} are $o_P(1)$.

We now show that $\bar{X}_{\gamma,n} - \bar{\mu}_{\gamma,n} = o_P\left(\ell^{-1}\right)$. Define $\phi_{nt}(x) = \omega_{nt}x$, where $\omega_{nt} \equiv \min\{t/\ell, 1, (n-t+1)/\ell\}$, and note that $\phi_{nt}(\cdot)$ is uniformly Lipschitz continuous. Next, write $\bar{X}_{\gamma,n} - \bar{\mu}_{\gamma,n} = (n-\ell+1)^{-1} \sum_{t=1}^{n} Y_{nt}$, where $Y_{nt} \equiv \phi_{nt}(Z_{nt})$ is a mean zero NED array on $\{V_t\}$ of the same size as Z_{nt} by Theorem 17.12 of Davidson (1994), satisfying the same moment conditions. Hence, by Lemma A.1 $\left\{Y_{nt}, \mathcal{F}^t\right\}$ is an L_2-mixingale of size $-\frac{3r-2}{3(r-2)}$, and thus of size $-1/2$, with uniformly bounded constants, and by Lemma A.2 $E\left(\max_{1\le j\le n}\left(\sum_{t=1}^{j} Y_{nt}\right)^2\right) = O(n)$. By Chebyshev's inequality, for $\varepsilon > 0$, $P\left[\ell\left(\bar{X}_{\gamma,n} - \bar{\mu}_{\gamma,n}\right) > \varepsilon\right] \le \frac{\ell^2}{\varepsilon^2(n-\ell+1)^2} E\left(\sum_{t=1}^{n} Y_{nt}\right)^2 = O\left(\frac{\ell^2 n}{(n-\ell+1)^2}\right) = o(1)$, if $\ell = o\left(n^{1/2}\right)$. This implies $A'_{n4} = o_P(1)$ and similarly $A_{n1} = o_P(1)$, given that $\sum_{\tau=-\ell+1}^{\ell-1}\left(1 - \frac{|\tau|}{\ell}\right) \sum_{t=1}^{n-|\tau|} \beta_{n,t,\tau}\left(Z_{nt} + Z_{n,t+|\tau|}\right) = O_P(\ell)$.

To prove that $A_{n3} = o_P(1)$, define $\mathcal{Y}_{nt\tau} \equiv \omega_{nt\tau}\left(\mu_{n,t+|\tau|} - \bar{\mu}_{\gamma,n}\right) Z_{nt} = \phi_{nt\tau}(Z_{nt})$, where $\omega_{nt\tau} = \min\left\{\frac{t}{\ell-|\tau|}, 1, \frac{n-t-|\tau|+1}{\ell-|\tau|}\right\}$, and $\phi_{nt\tau}(x) = \omega_{nt\tau}\left(\mu_{n,t+|\tau|} - \bar{\mu}_{\gamma,n}\right) x$ is uniformly Lipschitz continuous. Arguing as above, for each τ, $\left\{\mathcal{Y}_{nt\tau}, \mathcal{F}^t\right\}$ is an L_2-mixingale of size $-1/2$ by Lemma A.1, with mixingale constants $c_{nt\tau}^{\mathcal{Y}} \le K \max\{\|\mathcal{Y}_{nt\tau}\|_{3r}, 1\} \le K \max\{\|Z_{nt}\|_{3r}, 1\}$ which are bounded (uniformly in n, t and τ). Thus,

$$P\left[\left|\sum_{\tau=-\ell+1}^{\ell-1}\left(1 - \frac{|\tau|}{\ell}\right) \frac{1}{n-\ell+1} \sum_{t=1}^{n-|\tau|} \mathcal{Y}_{nt\tau}\right| \ge \varepsilon\right] \le \frac{1}{(n-\ell+1)\varepsilon}\left[\sum_{\tau=-\ell+1}^{\ell-1}\left(1 - \frac{|\tau|}{\ell}\right) E\left|\sum_{t=1}^{n-|\tau|} \mathcal{Y}_{nt\tau}\right|\right]$$

$$\le \frac{1}{(n-\ell+1)\varepsilon}\left[\sum_{\tau=-\ell+1}^{\ell-1}\left(1-\frac{|\tau|}{\ell}\right)\left(E\left(\sum_{t=1}^{n-|\tau|}\mathcal{Y}_{nt\tau}\right)^2\right)^{1/2}\right]$$

$$\le \frac{1}{(n-\ell+1)\varepsilon}\left[\sum_{\tau=-\ell+1}^{\ell-1}\left(1-\frac{|\tau|}{\ell}\right)\left(K\sum_{t=1}^{n-|\tau|}\left(c_{nt\tau}^{\mathcal{Y}}\right)^2\right)^{1/2}\right] \le K\frac{\ell n^{1/2}}{n-\ell+1} \to 0,$$

where the first inequality holds by Markov's inequality, the second inequality holds by Jensen's inequality, the third inequality holds by Lemma A.2 applied to $\{\mathcal{Y}_{nt\tau}\}$ for each τ, and the last inequality holds by the uniform boundedness of $c_{nt\tau}^{\mathcal{Y}}$. The proof of $A_{n2} = o_P(1)$ follows similarly.

The proof of the theorem for the SB follows closely that for the MBB, and we only present the relevant details. In step 1, let $\tilde{\Sigma}_{n,2} = \hat{R}_{n0}(0) + 2\sum_{\tau=1}^{n-1} b_{n\tau}\hat{R}_{n0}(\tau)$, where $\hat{R}_{n0}(\tau) = n^{-1}\sum_{t=1}^{n-\tau} Z_{nt}Z_{n,t+\tau}$, and follow Gallant and White (1988, p. 111; see also Newey and West, 1987) to show $E\left(\tilde{\Sigma}_{n,2}\right) - \Sigma_n \to 0$. Let $f_n(\tau) \equiv 1_{\{\tau\le n-1\}}|b_{n\tau}-1|\left(\alpha_{\left[\frac{\tau}{4}\right]}^{\frac{1}{2}-\frac{1}{r}} + v_{\left[\frac{\tau}{4}\right]}\right)$, and let ξ be the counting measure on the positive integers. By the dominated convergence theorem, $\lim_{n\to\infty}\left|E\left(\tilde{\Sigma}_{n,2}\right)-\Sigma_n\right|$ $\le \lim_{n\to\infty} 2K\sum_{\tau=1}^{n-1}|b_{n\tau}-1|\left(\alpha_{\left[\frac{\tau}{4}\right]}^{\frac{1}{2}-\frac{1}{r}} + v_{\left[\frac{\tau}{4}\right]}\right) = \lim_{n\to\infty}\int_0^\infty f_n(\tau)\,d\xi(\tau) = 0$, given that for each $\tau\in\mathbb{N}$, $\lim_{n\to\infty} b_n(\tau) = 1$, hence $\lim_{n\to\infty} f_n(\tau) = 0$, and given that $\sum_{\tau=1}^{\infty}\left(\alpha_{\left[\frac{\tau}{4}\right]}^{\frac{1}{2}-\frac{1}{r}} + v_{\left[\frac{\tau}{4}\right]}\right) < \infty$ by the size conditions on α_k and v_k. To prove that $\operatorname{var}\left(\tilde{\Sigma}_{n,2}\right) \to 0$, it then suffices that $\operatorname{var}\left(\hat{R}_{n0}(\tau)\right) = O\left(n^{-1}\right)$, which implies $\operatorname{var}\left(\tilde{\Sigma}_{n,2}\right) = O\left(\left(np_n^2\right)^{-1}\right) = o(1)$, given that $np_n^2 \to \infty$.

In step 2, define $S_{n,2} = n^{-1}\sum_{t=1}^{n} X_{nt}^2 + 2\sum_{\tau=1}^{n-1} b_{n\tau} n^{-1}\sum_{t=1}^{n-\tau} X_{nt}X_{n,t+\tau}$ and let $\hat{C}_n(\tau)$ $= n^{-1}\sum_{\tau=1}^{n} X_{nt}X_{n,t+\tau} - \bar{X}_n^2$ denote the *circular* autocovariance, which is based on the *extended* time series $\{X_{n1},\ldots,X_{nn},\, X_{n,n+1},\ldots,X_{n,2n},\ldots\}$, where $X_{n,i+jn} = X_{ni}$ for $0\le i\le n$ and $j = 1,2,\ldots$, with $X_{n0}\equiv X_{nn}$. As in Politis and Romano (1994a, Lemma 1), $\hat{C}_n(\tau) = \hat{R}_n(\tau) + \hat{R}_n(n-\tau)$, where $\hat{R}_n(\tau) = n^{-1}\sum_{t=1}^{n-\tau}\left(X_{nt}-\bar{X}_n\right)\left(X_{n,t+\tau}-\bar{X}_n\right)$. Using this property, $\hat{\Sigma}_{n,2} = S_{n,2} - \bar{X}_n^2 - 2\bar{X}_n^2\sum_{\tau=1}^{n-1}\left(1-\frac{\tau}{n}\right)(1-p)^\tau$, and

$$\begin{aligned}
\tilde{\Sigma}_{n,2} = {} & S_{n,2} - 2n^{-1}\sum_{t=1}^{n} X_{nt}\mu_{nt} + n^{-1}\sum_{t=1}^{n}\mu_{nt}^2 \\
& - 2\sum_{\tau=1}^{n-1}\left(1-\frac{\tau}{n}\right)(1-p)^\tau\left(n^{-1}\sum_{t=1}^{n} X_{nt}\mu_{n,t+\tau} + n^{-1}\sum_{t=1}^{n} X_{n,t+\tau}\mu_{nt}\right) \\
& + 2\sum_{\tau=1}^{n-1}\left(1-\frac{\tau}{n}\right)(1-p)^\tau\left(n^{-1}\sum_{t=1}^{n}\mu_{nt}\mu_{n,t+\tau}\right).
\end{aligned}$$

If $\mu_{nt} = \mu$ for all n, t then $\hat{\Sigma}_{n,2} - \tilde{\Sigma}_{n,2}$ simplifies to $-\left(\bar{X}_n-\mu\right)^2 - 2\left(\bar{X}_n-\mu\right)^2\sum_{\tau=1}^{n-\tau}\left(1-\frac{\tau}{n}\right)(1-p)^\tau$, which is $O_P\left(\frac{\ell}{n}\right)$, given that $\left(\bar{X}_n-\mu\right)^2 = O_P\left(n^{-1}\right)$ by a CLT for the sample mean, and given that

$\sum_{\tau=1}^{n-\tau}\left(1-\frac{\tau}{n}\right)(1-p)^{\tau}=O\left(p_n^{-1}\right)$. In the more general case, $\hat{\Sigma}_{n,2}-\tilde{\Sigma}_{n,2}=\zeta_{n1}+\zeta_{n2}+\zeta_{n3}+\zeta_{n4}+U_{n,2}$, where

$$\begin{aligned}
\zeta_{n1} &= -\left(\bar{X}_n-\bar{\mu}_n\right)^2; \;\; \zeta_{n2}=-2\left(\bar{X}_n-\bar{\mu}_n\right)^2\sum_{\tau=1}^{n-1}\left(1-\frac{\tau}{n}\right)(1-p)^{\tau};\\
\zeta_{n3} &= -2n^{-1}\sum_{t=1}^{n}\left(X_{nt}-\mu_{nt}\right)\left(\bar{\mu}_n-\mu_{nt}\right);\\
\zeta_{n4} &= -2\sum_{\tau=1}^{n-1}b_{n\tau}\left[n^{-1}\sum_{t=1}^{n-\tau}\left(X_{nt}-\mu_{nt}\right)\left(\bar{\mu}_n-\mu_{n,t+\tau}\right)+n^{-1}\sum_{t=1}^{n-\tau}\left(X_{n,t+\tau}-\mu_{n,t+\tau}\right)\left(\bar{\mu}_n-\mu_{nt}\right)\right];\\
U_{n,2} &= n^{-1}\sum_{t=1}^{n}\left(\mu_{nt}-\bar{\mu}_n\right)^2+2\sum_{\tau=1}^{n-1}b_{n\tau}n^{-1}\sum_{t=1}^{n-\tau}\left(\mu_{nt}-\bar{\mu}_n\right)\left(\mu_{n,t+\tau}-\bar{\mu}_n\right).
\end{aligned}$$

Now show that ζ_{n1} and ζ_{n2} are $O_P\left(n^{-1}\right)$ and $O_P\left(\frac{\ell_n}{n}\right)$ respectively, and that the remaining terms are $o_P(1)$ by an argument similar to the one used for the MBB to show that $A_{ni}=o_P(1)$. By Lemma 1 in Politis and Romano (1994a), $U_{n,2}\equiv \mathrm{var}^*\left(\sqrt{n}\bar{\mu}_n^{*(2)}\right)$. ■

Proof of Lemma 2.1. Immediate from the proof of Theorem 2.1. ■

Proof of Corollary 2.1. Immediate from Theorem 2.1 and the remark that follows it. ■

Proof of Theorem 2.2. (i) follows by Theorem 5.3 in Gallant and White (1988) and Polya's theorem (e.g. Serfling, 1980, p. 20). To prove (ii), first note that

$$\begin{aligned}
\Sigma_n^{-1/2}\sqrt{n}\left(\bar{X}_n^{*(j)}-\bar{X}_n\right) &= \Sigma_n^{-1/2}\sqrt{n}\left(\bar{Z}_n^{*(j)}-E^*\left(\bar{Z}_n^{*(j)}\right)\right)+\Sigma_n^{-1/2}\sqrt{n}\left(E^*\left(\bar{Z}_n^{*(j)}\right)-\bar{Z}_n\right)\\
&\quad+\Sigma_n^{-1/2}\sqrt{n}\left(\bar{\mu}_n^{*(j)}-\bar{\mu}_n\right)\\
&\equiv A_n^{(j)}+B_n^{(j)}+C_n^{(j)},
\end{aligned}$$

where $Z_{nt}\equiv X_{nt}-\mu_{nt}$ and $Z_{nt}^{*(j)}\equiv X_{nt}^{*(j)}-\mu_{nt}^{*(j)}$.

Proof for $j=1$. By Lemma A.1 of Fitzenberger (1997), $E^*\left(\bar{Z}_n^{*(1)}\right)=\bar{Z}_n+O_P\left(\frac{\ell}{n}\right)$. Thus, $B_n^{(1)}=O_P\left(\ell/n^{1/2}\right)=o_P(1)$, given that $\Sigma_n>\kappa>0$ and $\ell_n=o\left(n^{1/2}\right)$. Also, $E^*\left(C_n^{(1)}\right)=O(\ell/n)\to 0$ and $\mathrm{var}^*\left(C_n^{(1)}\right)=\Sigma_n^{-1/2}U_{n,1}^2\to 0$, which implies $C_n^{(1)}\overset{P^*}{\to}0$. Hence, it suffices to prove that $A_n^{(1)}\Rightarrow N(0,1)$, given $Z_{n1},\ldots,Z_{nn}$, with probability approaching one. Write $\bar{Z}_n^{*(1)}=k^{-1}\sum_{i=1}^{k}\mathcal{U}_{ni}$, where $\{\mathcal{U}_{ni}\}$ are i.i.d. with $P^*\left(\mathcal{U}_{ni}=\frac{Z_{n,j+1}+\ldots+Z_{n,j+\ell}}{\ell}\right)=\frac{1}{n-\ell+1}$, $j=0,\ldots,n-\ell$. Thus, $E^*\left(\bar{Z}_n^{*(1)}\right)=E^*\left(\mathcal{U}_{n1}\right)$ and $A_n^{(1)}=\Sigma_n^{-1/2}\sqrt{n}\left(k^{-1}\sum_{i=1}^{k}\left[\mathcal{U}_{ni}-E^*\left(\mathcal{U}_{n1}\right)\right]\right)\equiv\sum_{i=1}^{k}\tilde{Z}_{ni}$, where $\tilde{Z}_{ni}=\Sigma_n^{-1/2}n^{-1/2}\ell\left[\mathcal{U}_{ni}-E^*\left(\mathcal{U}_{n1}\right)\right]$, given $k=\frac{n}{\ell}$. In particular, $\left\{\tilde{Z}_{ni}\right\}$ are i.i.d. with $E^*\left(\tilde{Z}_{ni}\right)=0$ and $\mathrm{var}^*\left(\tilde{Z}_{n1}\right)=k^{-1}\frac{\hat{\Sigma}_{n,1}}{\Sigma_n}$. By Katz's (1963) Berry-Esseen Bound, for some small $\delta>0$,

$$\sup_{x\in\mathbb{R}}\left|P^*\left(\frac{\sum_{i=1}^k \tilde{Z}_{ni}}{\sqrt{\text{var}^*\left(\sum_{i=1}^k \tilde{Z}_{ni}\right)}}\leq x\right)-\Phi\left(x\right)\right|\leq K\left(\frac{\hat{\Sigma}_{n,1}}{\Sigma_n}\right)^{-1-\delta/2} kE^*\left|\tilde{Z}_{n1}\right|^{2+\delta}.$$

Since $\frac{\hat{\Sigma}_{n,1}}{\Sigma_n}\xrightarrow{P}1$ by Corollary 2.1, it suffices to show that $kE^*\left|\tilde{Z}_{n1}\right|^{2+\delta}\xrightarrow{P}0$. But

$$E\left|kE^*\left|\tilde{Z}_{n1}\right|^{2+\delta}\right|\leq\frac{n}{\ell\left(n-\ell+1\right)}\frac{1}{n^{1+\frac{\delta}{2}}\Sigma_n^{1+\delta/2}}\sum_{j=0}^{n-\ell}\left[\left\|\sum_{t=1}^{\ell}Z_{n,t+j}\right\|_{2+\delta}+\left\|\ell E^*\left(\mathcal{U}_{n,1}\right)\right\|_{2+\delta}\right]^{2+\delta}, \tag{A.5}$$

where the inequality follows by the Minkowski inequality. Under our assumptions,

$$\left\|\sum_{t=1}^{\ell}Z_{n,t+j}\right\|_{2+\delta}\leq\left\|\max_{1\leq i\leq\ell}\left|\sum_{t=j+1}^{j+i}Z_{nt}\right|\right\|_{2+\delta}\leq K\left(\sum_{t=j+1}^{j+\ell}c_{nt}^2\right)^{1/2}\leq K\ell^{1/2},$$

by Lemmas A.3 and A.4, given that the c_{nt} are uniformly bounded. Similarly, $\left\|\ell E^*\left(\mathcal{U}_{n,1}\right)\right\|_{2+\delta}=O\left(\ell^{1/2}\right)$, which from (A.5), implies $E\left|kE^*\left|\tilde{Z}_{n1}\right|^{2+\delta}\right|=O\left(\left(\frac{\ell}{n}\right)^{\delta/2}\right)\to 0$, given $\ell_n=o\left(n^{1/2}\right)$.

Proof for $j=2$. First, note that $B_n^{(2)}=0$ since for the SB we have $E^*\left(\bar{Z}_n^{*(1)}\right)=\bar{Z}_n$. Second, $C_n^{(2)}\xrightarrow{P^*}0$, given that $E^*\left(n^{-1/2}\sum_{t=1}^n\left(\mu_{nt}^{*(2)}-\mu_{nt}\right)\right)=0$, by the stationarity of the SB resampling scheme, and given that $\text{var}^*\left(n^{-1/2}\sum_{t=1}^n\left(\mu_{nt}^{*(2)}-\mu_{nt}\right)\right)\equiv U_{n,2}\to 0$ by Assumption 2.2. To prove that $A_n^{(2)}\Rightarrow N\left(0,1\right)$, we verify the conditions (C1), (C2) and (C3) in Politis and Romano's (1994) proof of their Theorem 2 and refer to their proof for more details. In our more general case, where Σ_n is not assumed to have a limit value Σ_∞^2, the appropriate version of these conditions is as follows:

(C1) $\frac{n\bar{Z}_n^2}{np}\xrightarrow{P}0$;

(C2) $\hat{C}_n\left(0\right)+2\sum_{\tau=1}^{\infty}\left(1-p\right)^{\tau}\hat{C}_n\left(\tau\right)-\Sigma_n\xrightarrow{P}0$;

(C3) $\frac{p}{n^{1+\frac{\delta}{2}}}\sum_{b=1}^{\infty}\sum_{\tau=1}^{n}\left|S_{\tau,b}-\frac{\bar{Z}_n}{p}\right|^{2+\delta}\left(1-p\right)^{b-1}p\xrightarrow{P}0$,

where in (C3) $S_{\tau,b}$ is defined as the sum of observations in block $B_{\tau,b}$.

Proof of (C1): This follows from $\bar{Z}_n=O_P\left(n^{-1/2}\right)$ (by the CLT) and $np_n\to\infty$.

Proof of (C2): First, define $\hat{\Sigma}_{n,\infty}=\hat{C}_n\left(0\right)+2\sum_{\tau=1}^{\infty}\left(1-p\right)^{\tau}\hat{C}_n\left(\tau\right)$ and $\hat{\Sigma}_n=\hat{C}_n\left(0\right)+2\sum_{\tau=1}^{n-1}\left(1-p\right)^{\tau}\hat{C}_n\left(\tau\right)$. Next, note that $\hat{C}_n\left(i\right)=\hat{C}_n\left(i+nj\right)$ for all $i=0,1,\ldots,n-1$ and $j=1,2,\ldots$, by the circularity properties of the extended time series. It follows that $\hat{\Sigma}_{n,\infty}=\hat{\Sigma}_n+2\sum_{j=1}^{\infty}\sum_{i=0}^{n-1}\left(1-p\right)^{nj+i}\hat{C}_n\left(nj+i\right)=\hat{\Sigma}_n+\left(\hat{\Sigma}_n+\hat{C}_n\left(0\right)\right)\sum_{j=1}^{\infty}\left(1-p\right)^{nj}$. By an argument similar to the proof of Theorem 2.1 for $j=2$, we can show that $\hat{\Sigma}_n-\Sigma_n\xrightarrow{P}0$. Hence, $\hat{\Sigma}_{n,\infty}-\Sigma_n\xrightarrow{P}0$, since $\hat{\Sigma}_n$ and $\hat{C}_n\left(0\right)$ are $O_P\left(1\right)$ and $\sum_{j=1}^{\infty}\left(1-p\right)^{nj}$ is $o_P\left(1\right)$, given $np_n\to\infty$ and $p_n\to 0$.

Proof of (C3): This follows provided $E\left|\sum_{t=\tau}^{\tau+b-1} Z_{nt}\right|^{2+\delta} \leq Kb^{1+\frac{\delta}{2}}$, where the constant K only depends on the mixingale coefficients of $\{Z_{nt}\}$. Apply Lemma A.4. ■

Proof of Corollary 2.2. Immediate from the proof of Theorem 2.2 for $j = 1$. ■

References

[1] Andrews, D.W.K. (1984) Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930-934.

[2] Andrews, D.W.K. (1988) Laws of large numbers for dependent nonidentically distributed random variables. *Econometric Theory* 4, 458-467.

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

[4] Billingsley, P. (1968) *Convergence of probability measures.* New York: Wiley.

[5] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.

[6] Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17-39.

[7] Davidson, J. (1992) A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes. *Econometric Theory* 8, 313-329.

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

[9] Davidson, J. (2002) Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics* 106, 243-269.

[10] Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7, 1-26.

[11] Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1006.

[12] Fitzenberger, B. (1997) The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics* 82, 235-287.

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

[14] Gonçalves, S. & H. White (2000) Maximum likelihood and the bootstrap for nonlinear dynamic models. UCSD Working Paper 2000-32.

[15] Hall, P, J. Horowitz, & B. Jing (1995) On blocking rules for the bootstrap with dependent data. *Biometrika* 82, 561-74.

[16] Hansen, B.E. (1991a) GARCH(1,1) processes are near epoch dependent. *Economics Letters* 36, 181-186.

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

[18] Hansen, B.E. (1992) Erratum. *Econometric Theory* 8, 421-422.

[19] Horowitz, J. (1999) The bootstrap. Forthcoming in *Handbook of econometrics*, Vol. 5.

[20] Horowitz, J. (2001) The bootstrap and hypothesis tests in econometrics. *Journal of Econometrics* 100, 37-40.

[21] Katz, M.L. (1963) Note on the Berry-Esseen theorem. *Annals of Mathematical Statistics* 34, 1107-1108.

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

[23] Lahiri, S.N. (1992) Edgeworth correction by 'moving block' bootstrap for stationary and nonstationary data. In R. LePage and L. Billiard (eds), *Exploring the limits of the bootstrap*, pp. 183-214. New York: Wiley.

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

[25] Liu, R.Y. & 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.

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

[27] Newey, W.K. & K.D. West (1987) A simple positive semi-definite, heteroskedastic and autocorrelation consistent covariance matrix. *Econometrica* 55, 703-708.

[28] Politis, D. & H. White (2001) Automatic block-length selection for the dependent bootstrap. UCSD manuscript.

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

[30] Politis, D. & J. Romano (1994b) Limit theorems for weakly dependent Hilbert space-valued random variables with applications to the stationary bootstrap. *Statistica Sinica* 4, 461-476.

[31] Politis, D., Romano, J. & M. Wolf (1997) Subsampling for heteroskedastic time series. *Journal of Econometrics* 81, 281-317.

[32] Serfling, R.J. (1980) *Approximation theorems of mathematical statistics.* New York: Wiley.

[33] Sin, C.-Y. & H. White (1996) Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics* 71, 207-225.

[34] Singh, K. (1981) On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics* 9, 1187-1195.