ELSEVIER Journal of Econometrics 81 (1997) 281–317

JOURNAL OF Econometrics

Subsampling for heteroskedastic time series

D.N. Politis[a], Joseph P. Romano[b,*], Michael Wolf[b]

[a] *Department of Statistics, Purdue University, W. Lafayette, IN 47907, USA*
[b] *Department of Statistics, Stanford University, Stanford, CA 94305, USA*

(Received September 1995; received in revised form August 1996)

Abstract

In this article, a general theory for the construction of confidence intervals or regions in the context of heteroskedastic-dependent data is presented. The basic idea is to approximate the sampling distribution of a statistic based on the values of the statistic computed over smaller subsets of the data. This method was first proposed by Politis and Romano (1994b) for stationary observations. We extend their results to heteroskedastic observations, and prove a general asymptotic validity result under minimal conditions. In contrast, the usual bootstrap and moving blocks bootstrap are typically valid only for asymptotically linear statistics and their justification requires a case-by-case analysis. Our general asymptotic results are applied to a regression setting with dependent heteroskedastic errors. © 1997 Elsevier Science S.A.

Keywords: Time series; Heteroskedasticity; Subsampling; Moving blocks bootstrap
JEL classification: C14; C15

1. Introduction

It has been almost two decades since Efron (1979) introduced the bootstrap procedure for estimating sampling distributions of statistics based on independent and identically distributed (i.i.d.) observations. It is well known that, in the i.i.d. setup, the bootstrap often gives more accurate approximations than classical large sample approximations (e.g. Singh, 1981; Babu, 1986). However, when the observations are not necessarily independent the classical bootstrap no longer succeeds, as showed by Singh (1981). Most extensions in the literature so far only apply to the stationary case. They can roughly be divided into resampling and subsampling methods.

*Corresponding author.

PII S0304-4076(96)00016-X

There are, broadly speaking, two approaches to using resampling methods for strictly stationary dependent data. One is to apply Efron's bootstrap to an approximate i.i.d. setting by focusing on the residuals of some general regression model. Such examples include linear regression (Freedman, 1981; Freedman, 1984; Wu, 1986; Liu, 1988), autoregressive time series (Efron and Tibshirani, 1983; Bose, 1988), nonparametric regression and nonparametric kernel spectral estimation (Härdle and Bowman, 1988; Franke and Härdle, 1992). In all of the above situations the residuals are resampled, not the original observations. In addition to being restricted to relatively simple contexts where structural models are both plausible and tractable, little is known how this approach would perform for heteroskedastic observations. The fitted residuals will, in general, no longer behave like i.i.d. observations but exhibit some form of heteroskedasticity. However, it is known that Efron's bootstrap works reasonably well even when the data are independent but not identically distributed (Freedman, 1981; Liu, 1988; Liu and Singh, 1992), so one might hope for some robustness to heteroskedasticity as well. As a second approach, resampling methods for less restrictive contexts have been suggested more recently. They are based on "blocking" arguments, in which the data are divided into blocks and these blocks, rather than individual data values or estimated residuals, are resampled. Carlstein (1986) proposed non-overlapping blocks, whereas Künsch (1989) and Liu and Singh (1992) independently introduced the 'moving blocks' method which employs overlapping blocks. Subsequent research seems to have favored this scheme. Politis and Romano (1992) consider a blocks of blocks scheme to obtain valid inference of parameters of the infinite-dimensional joint distribution of the process, such as the spectrum. It turns out that Künsch's bootstrap enjoys some robustness property to heteroskedasticity, as was pointed out by Lahiri (1992) in the case of the sample mean. In both Carlstein's and Künsch's bootstrap blocks of fixed length are resampled, so that the newly generated pseudo-time series is no longer stationary. To fix this shortcoming, Politis and Romano (1994a) suggested the stationary bootstrap.

As an alternative to resampling methods, Politis and Romano (1994b) proposed the subsampling approach. Rather than resampling blocks from the original time series as ingredients to generating a new pseudo-time series, each individual subblock or subseries of observations is looked upon as a valid 'sub-time series' in its own right. The motivation is as follows. Each block, as a part of the original series, was generated by the true underlying probability mechanism. It then seems reasonable to hope that one can gain information about the sampling distribution of a statistic by evaluating it on all subseries, or 'subsamples'. On the other hand, building new pseudo-time series by joining randomly sampled, independent blocks together induces a different probability mechanism. Dependency will be reduced, and, for Carlstein's and Künsch's bootstrap, stationarity will be lost. However, in typical applications the underlying dependence is sufficiently weak. Therefore, the main contributions come from short lags which are well

approximated by the 'blocking' methods, ensuring that these methods nevertheless work.

Another attractive feature of the subsampling method is that it has been shown to be valid under very weak assumptions. Apart from regularity and dependency conditions, the only requirement, in the stationary setup, is that the sampling distribution of the properly normalized statistic of interest has a nondegenerate limiting distribution. The moving blocks method has essentially been shown to be valid for functions of linear statistics and smooth functionals only (see Künsch, 1989; Bühlmann, 1994).

In this paper we present conditions which ensure that the subsampling method is still asymptotically valid for heteroskedastic observations. The paper is organized as follows. In Section 2 the method is described, and the main theorems are presented. In Section 3 some applications and examples are discussed. In addition, a result for the validity of the moving blocks method for heteroskedastic data is stated. We talk about the problem of choosing of the blocksize in Section 4. Section 5 presents two simulation studies to address finite sample properties of the method. As a real-life example, we apply the subsampling method to variance ratio tests in Section 6. Section 7 gives a conclusion of the paper and some outlook on future research. Three appendices contain a central limit theorem for heteroskedastic-dependent random variables, the proofs of technical results, and tables for the outcomes of the simulation studies and the variance ratio tests.

2. The general theorem

Suppose $\{\ldots,X_{-1},X_0,X_1,\ldots\}$ is a sequence of random variables taking values in an arbitrary sample space S, and defined on a common probability space. Denote the joint probability law governing the infinite sequence by P. The goal is to construct a confidence region for some real or vector-valued parameter $\theta(P)$, on the basis of observing $\{X_1,\ldots,X_n\}$. The time series $\{X_i\}$ will be assumed to satisfy a certain weak-dependence condition. Specifically, given a random sequence $\{Y_i\}$, let $\mathcal{F}_n^m$ be the σ-algebra generated by $\{Y_i,\ n\leqslant i\leqslant m\}$, and define the corresponding mixing sequence by

$$\alpha_Y(k)=\sup_n \sup_{A,B} |P(A\cap B)-P(A)P(B)|,$$

where A and B vary over the σ-fields $\mathcal{F}_{-\infty}^n$ and $\mathcal{F}_{n+k}^{\infty}$, respectively. The sequence $\{Y_i\}$ is called α-mixing or strong mixing if $\alpha_Y(k)\to 0$ as $k\to\infty$.

Let $\hat{\theta}_n=\hat{\theta}_n(X_1,\ldots,X_n)$ be an estimator of $\theta(P)\in\mathbb{R}^k$. Let $\hat{\theta}_{b,a}=\hat{\theta}_b(X_a,\ldots,X_{a+b-1})$, the estimator of θ based on the subsample $X_a,\ldots,X_{a+b-1}$. Define $J_{b,a}(P)$ to be the sampling distribution of $\tau_b(\hat{\theta}_{b,a}-\theta(P))$, where τ_b is an appropriate normalizing constant. For any Borel set $A\in\mathcal{B}^k$ define:

$$J_{b,a}(A,P)=\mathrm{Prob}_P\{\tau_b(\hat{\theta}_{b,a}-\theta(P))\in A\}. \tag{1}$$

Essentially, the only assumption that we will need to construct asymptotically valid confidence regions for $\theta(P)$ is the following.

Assumption A. There exists a limiting law $J(P)$ such that
(i) $J_{n,1}(P)$ converges weakly to $J(P)$ as $n \to \infty$, and
(ii) for every Borel set A whose boundary has mass zero under $J(P)$,

$$\frac{1}{n-b+1} \sum_{a=1}^{n-b+1} J_{b,a}(A,P) \to J(A,P),$$

for any sequences n, b with $n, b \to \infty$ and $b/n \to 0$.

Condition (i) states that the estimator, properly normalized, has a limiting distribution. It is hard to conceive of any asymptotic theory free of such a requirement. Typically, much stronger assumptions are in force to ensure asymptotic normality. Condition (ii) states that the distribution functions of the normalized estimator based on the subsamples will be on average close to the distribution function of the normalized estimator based on the entire sample, for large n. A somewhat stronger condition is the following.

Assumption B. There exists a limiting law $J(P)$ such that
(i) $J_{n,1}(P)$ converges weakly to $J(P)$ as $n \to \infty$, and
(ii) for any index sequence $\{a_b\}$, $J_{b,a_b}(A,P) \to J(A)$, for every Borel set A whose boundary has mass zero under $J(P)$, as $b \to \infty$.

Here, condition (ii) requires that the distribution function of the normalized statistic evaluated over a subsample converges to the same limiting law as the distribution function of the normalized estimator based on the entire sample, *uniformly* in the starting point of the subsample. Assuming (i), the condition is satisfied for stationary processes, but also for processes that exhibit asymptotic stationarity. For example, one can consider a Markov chain with an equilibrium distribution. Assumption A follows from Assumption B.

In order to describe our method, let $Y_{b,a}$ be the block of size b of the consecutive data $\{X_a, \ldots, X_{a+b-1}\}$. Only a very weak assumption on b will be required. Typically, $b/n \to 0$ and $b \to \infty$ as $n \to \infty$. Now, let $\hat{\theta}_{b,a}$ be equal to the statistic $\hat{\theta}_b$ evaluated at the data set $Y_{b,a}$. The approximation to $J_{n,1}(A,P)$ we study is defined by

$$L_n(A) = \frac{1}{n-b+1} \sum_{a=1}^{n-b+1} 1\{\tau_b(\hat{\theta}_{b,a} - \hat{\theta}_{n,1}) \in A\}. \tag{2}$$

The motivation behind the method is the following. For any a, $Y_{b,a}$ is a 'true' subsample of size b. Hence, the *exact* distribution of $\tau_b(\hat{\theta}_{b,a} - \theta(P))$ is $J_{b,a}$. If condition (ii) of Assumption A is satisfied, then the empirical distribution of the $n-b+1$ values of $\tau_b(\hat{\theta}_{b,a} - \theta(P))$ should serve as good approximation to $J_n(P)$.

Replacing $\theta(P)$ by $\hat{\theta}_{n,1}$ is permissible because $\tau_b(\hat{\theta}_{n,1} - \theta)$ is of order τ_b/τ_n in probability and we will assume that $\tau_b/\tau_n \to 0$.

Theorem 2.1. Assume Assumption A or Assumption B, and that $\tau_b/\tau_n \to 0$, $b/n \to 0$ *and* $b \to \infty$ *as* $n \to \infty$. *Also, assume that* $\alpha_X(m) \to 0$ *as* $m \to \infty$. *Then*

(i) $L_n(A) \to J(A,P)$ *in probability, for each Borel set* A *whose boundary has mass zero under* $J(P)$.

(ii) $\rho_k(L_n, J(P)) \to 0$ *in probability for every metric* ρ_k *that metrizes weak convergence on* $\mathbb{R}^k$.

(iii) *Let* $\{Y_n\}$ *and* Y *be random vectors with* $\mathcal{L}(Y_n) = L_n$ *and* $\mathcal{L}(Y) = J(P)$. *Then, for any almost everywhere* $J(P)$ *continuous real function* f *and any metric* ρ_1 *which metrizes weak convergence on* $\mathbb{R}$, $\rho_1(\mathcal{L}(f(Y_n)), \mathcal{L}(f(Y))) \to 0$ *in probability. In particular, for a norm* $\|\cdot\|$ *on* $\mathbb{R}^k$ $\rho_1(\mathcal{L}(\|Y_n\|), \mathcal{L}(\|Y\|)) \to 0$ *in probability. This allows us to find confidence regions for* $\theta(P)$.

(iv) *Let* Y *be a random vector with* $\mathcal{L}(Y) = J(P)$. *For a norm* $\|\cdot\|$ *on* $\mathbb{R}^k$ *define univariate distributions* $L_{n,\|\cdot\|}$ *and* $J_{\|\cdot\|}(P)$ *in the following way*:

$$L_{n,\|\cdot\|}(x) = \frac{1}{n-b+1} \sum_{a=1}^{n-b+1} 1\{\|\tau_b(\hat{\theta}_{b,a} - \hat{\theta}_{n,1})\| \leqslant x\},$$

$$J_{\|\cdot\|}(x,P) = \text{Prob}\{\|Y\| \leqslant x\}.$$

For $\alpha \in (0,1)$, *let*

$$c_{n,L}(1-\alpha) = \inf\{x : L_{n,\|\cdot\|}(x) \geqslant 1-\alpha\},$$

$$c_{n,U}(1-\alpha) = \sup\{x : L_{n,\|\cdot\|}(x) \leqslant 1-\alpha\}.$$

Correspondingly, define

$$c_L(1-\alpha, P) = \inf\{x : J_{\|\cdot\|}(x,P) \geqslant 1-\alpha\},$$

$$c_U(1-\alpha, P) = \sup\{x : J_{\|\cdot\|}(x,P) \leqslant 1-\alpha\}.$$

Let $\{c_n(1-\alpha)\}$ *be a sequence such that*

$$c_{n,L}(1-\alpha) \leqslant c_n(1-\alpha) \leqslant c_{n,U}(1-\alpha).$$

If $J_{\|\cdot\|}(\cdot, P)$ *is continuous at* $c_L(1-\alpha)$, *then*

$$\text{Prob}_P\{\|\tau_n(\hat{\theta}_{n,1} - \theta(P))\| \leqslant c_n(1-)\} \to 1-\alpha \quad \text{as } n \to \infty.$$

Thus, the asymptotic coverage probability under P *of the region* $\{\theta : \|\tau_n(\theta - \hat{\theta}_{n,1})\| \leqslant c_n(1-\alpha)\}$ *is the nominal level* $1-\alpha$.

Consider the special case of univariate parameters, that is, $\theta(P) \in \mathbb{R}$. The standard choice for a norm on the real line is the absolute value function $|\cdot|$.

In this context, applying the subsampling method as outlined above boils down to estimating

$$J_{n,1,|\cdot|}(x,P)=\text{Prob}_P\{\tau_n\,|\,\hat{\theta}_{n,1}-\theta(P)|\leqslant x\}. \tag{3}$$

The resulting confidence regions are two-sided *symmetric* intervals $[\hat{\theta}_{n,1}-\hat{c},\hat{\theta}_{n,1}+\hat{c}]$, where $\hat{c}$ is chosen so that $\text{Prob}_P\{|\hat{\theta}_{n,1}-\theta|>\hat{c}\}\doteq\alpha$. However, in some scenarios we might be particularly interested in one-sided confidence intervals, upper or lower ones depending on context. In this case, we would have to estimate the 'standard' distribution function

$$J_{n,1}(x,P)=\text{Prob}_P\{\tau_b(\hat{\theta}_{n,1}-\theta(P))\leqslant x\}. \tag{4}$$

It is obvious that the subsampling method can be used here – take the Borel sets A of form $(-\infty,x]$. For a corresponding theorem see Politis et al. (1995). Following this route, we could construct two-sided *equal-tailed* intervals as the intersection of two one-sided intervals: $[\hat{\theta}_{n,1}-\hat{c}_1,\hat{\theta}_{n,1}+\hat{c}_2]$, where $\hat{c}_1$ and $\hat{c}_2$ are chosen so that $\text{Prob}_P\{\theta<\hat{\theta}_{n,1}-\hat{c}_1\}\doteq\alpha/2$ and $\text{Prob}_P\{\theta>\hat{\theta}_{n,1}+\hat{c}_2\}\doteq\alpha/2$. Hall (1988) showed that symmetric confidence intervals often enjoy enhanced coverage and, even in asymmetric circumstances, can be shorter than equal-tailed confidence intervals. We will compare the finite sample performance of symmetric and equal-tailed intervals in simulation studies in Section 5.

3. Applications

In this section we demonstrate the validity of the subsampling method in some specific situations: the univariate mean, smooth functions of the mean and multivariate linear regression. We also state a result concerning the moving blocks method for the mean case.

Example 3.1 (*The univariate mean*). Suppose $\{X_i\}$ is a sequence of random variables with common mean θ. Denote the joint probability law governing the sequence by P. The goal is to construct a confidence interval for θ, on the basis of observing $\{X_1,\ldots,X_n\}$. Let $\hat{\theta}_{b,a}=\hat{\theta}_b(X_a,\ldots,X_{a+b-1})\equiv b^{-1}\sum_{i=a}^{a+b-1}X_i\equiv\overline{X}_{b,a}$ be our estimator of θ based on the block $\{X_a,\ldots,X_{a+b-1}\}$. Define $J_{b,a}(P)$ to be the sampling distribution of $b^{\frac{1}{2}}(\overline{X}_{b,a}-\theta)$. Also define the corresponding cumulative distribution function:

$$J_{b,a}(x,P)=\text{Prob}_P\{b^{1/2}(\overline{X}_{b,a}-\theta)\leqslant x\}. \tag{5}$$

The approximation to $J_{n,1}(x,P)$ we study is defined by

$$L_n(x)=\frac{1}{n-b+1}\sum_{a=1}^{n-b+1}1\{b^{1/2}(\overline{X}_{b,a}-\overline{X}_{n,1})\leqslant x\}. \tag{6}$$

The following theorem gives sufficient moment and mixing conditions for which the subsampling technique will allow us to draw first order correct inference about θ.

Theorem 3.1. Let $\{X_i\}$ be a sequence of random variables defined on a common probability space. Denote the corresponding (generalized) mixing coefficients by $\alpha_X(\cdot)$. Define

$$T_{k,a} \equiv k^{-1/2} \sum_{i=a}^{a+k-1} X_i, \qquad \sigma_{k,a}^2 \equiv \mathrm{Var}(T_{k,a}).$$

Assume the following conditions. For some $\delta > 0$:

- $\|X_i\|_{2+2\delta} \leq \Delta$ *for all* i, (7)
- $\sigma_{k,a}^2 \to \sigma^2 > 0$ *uniformly in* a, (8)
- $C(4) \equiv \sum_{k=1}^{\infty} (k+1)^2 \alpha_X^{\delta/(4+\delta)}(k) \leq K$. (9)

Furthermore, assume that $b/n \to 0$ and $b \to \infty$ as $n \to \infty$, and let $J(P) = \mathrm{N}(0, \sigma^2)$.
Then the conclusions of Theorem 2.1 *will be true.*

Example 3.2 (Moving blocks for the mean case). Consider again the situation of Example 3.1. We will show that the moving blocks method, which was introduced by Künsch (1989) and Liu and Singh (1992) for the case of stationary time series, will still work in this heteroskedastic setting.

To describe the method (for the case of the mean), let $Y_{b,a}$ be the block of size b of the consecutive data $\{X_a, \ldots, X_{a+b-1}\}$ and let $l \equiv l_n \equiv \lfloor b/n \rfloor$. Conditional on the sample $\{X_1, \ldots, X_n\}$, denote the empirical distribution of $Y_{b,1}, \ldots, Y_{b,q}$ (where $q \equiv q_n \equiv n - b + 1$) by P_n^*, i.e., P_n^* puts mass $1/q$ on each of the $Y_{b,a}$. Define a pseudo-time series $\{X_1^*, \ldots, X_{bl}^*\}$ in the following way: Let $Y_{b,1}^*, \ldots, Y_{b,l}^*$ i.i.d. $\sim P_n^*$ and join them together to one big block: $\{X_1^*, \ldots, X_{bl}^*\} = \{Y_{b,1,1}^*, \ldots, Y_{b,1,b}^*, Y_{b,2,1}^*, \ldots, Y_{b,2,b}^*, \ldots, Y_{b,l,1}^*, \ldots, Y_{b,l,b}^*\}$. Here, of course, $Y_{b,j,i}^*$ denotes the ith element of the block $Y_{b,j}^*$.

The corresponding cumulative distribution function is given by

$$L_n^*(x) = \mathrm{Prob}_{P_n^*}\{(bl)^{1/2}(\overline{X^*}_{bl,1} - \overline{X}_{n,1}) \leq x\}. \tag{10}$$

The following theorem states that the moving blocks method is also asymptotically valid, provided we strengthen the moment and mixing conditions on the sequence $\{X_i\}$.

Theorem 3.2. Let $\{X_i\}$ be a sequence of random variables defined on a common probability space. Denote the corresponding (generalized) mixing coefficients by $\alpha_X(\cdot)$. Define

$$T_{k,a} \equiv k^{-1/2} \sum_{i=a}^{a+k-1} X_i, \qquad \sigma_{k,a}^2 \equiv \mathrm{Var}(T_{k,a}).$$

Assume the following conditions. For some $\delta > 0$:

- $\|X_i\|_{4+3\delta} \leqslant \Delta$ *for all* i, (11)
- $\sigma_{k,a}^2 \to \sigma^2 > 0$ *uniformly in* a, (12)
- $C(6) \equiv \sum_{k=1}^{\infty} (k+1)^4 \alpha_X^{\delta/(6+\delta)}(k) \leqslant K.$ (13)

Furthermore, assume that $b/n \to 0$ and $b \to \infty$ as $n \to \infty$, and let $J(P) = \mathrm{N}(0, \sigma^2)$.
Then the conclusions of Theorem 2.1 *will be true if we replace $L_n(\cdot)$ by $L_n^*(\cdot)$.*

Remark 3.1. Relative to the result for the subsampling method we need stronger moment and mixing conditions on the sequence $\{X_i\}$ here in order to show that the variance of the moving blocks distribution converges in probability to the proper limit. For details the reader is referred to the proof of the theorem in the appendix.

Remark 3.2. There have been previous results extending the moving blocks method for the sample mean to the heteroskedastic case. A result similar to Theorem 4.2 was obtained by Fitzenberger (1995) in his Theorem 3.1. Note, however, that he needs a stronger condition on the block size b, namely $b = o(n^{1/2})$. Furthermore, his proof is somewhat invalidated by some mistakes. Under stronger assumptions, Lahiri (1992) not only showed a result similar to Theorem 3.2, but also obtained second-order properties. However, he uses an even ore stringent requirement on the block size, namely $b = o(n^{1/4})$. As will be discussed in Section 4, this rate is too small in many situations.

Example 3.3 (Smooth functions of the multivariate mean). Suppose $\{X_i\}$ is a sequence of multivariate random variables with common mean $0 \in \mathbb{R}^k$, and $k > 1$. Denote the joint probability law governing the sequence by P. Assume that on the basis of observing $\{X_1, \ldots, X_n\}$ we are interested in finding a confidence region for $\zeta \equiv f(0)$, where $f(\cdot)$ is a smooth function from $\mathbb{R}^k$ to $\mathbb{R}^p$. Let $\hat{\zeta}_{b,a} = \hat{\zeta}_b(X_a, \ldots, X_{a+b-1}) \equiv f(b^{-1} \sum_{i=a}^{a+b-1} X_i) \equiv f(\overline{X}_{b,a})$ be our estimator of ζ based on the block $\{X_a, \ldots, X_{a+b-1}\}$. Define $J_{b,a}(P)$ to be the sampling distribution of $b^{1/2}(f(\overline{X}_{b,a}) - \zeta)$. For any Borel set $A \in \mathscr{B}^p$ let

$$J_{b,a}(A, P) = \mathrm{Prob}_P\{b^{1/2}(f(\overline{X}_{b,a}) - \zeta) \in A\}. \tag{14}$$

The approximation to $J_{n,1}(A,P)$ we study is defined by

$$L_n(A) = \frac{1}{n-b+1} \sum_{a=1}^{n-b+1} 1\{b^{1/2}(f(\overline{X}_{b,a}) - f(\overline{X}_{n,1})) \in A\}. \tag{15}$$

Theorem 3.3. Let $\{X_i\}$ be a sequence of random vectors with common mean $0 \in \mathbb{R}^k$. Denote the corresponding mixing coefficients by $\alpha(\cdot)$. Define $T_{k,a} \equiv k^{-\frac{1}{2}} \sum_{i=a}^{a+k-1} X_i$ and $\Sigma_{k,a} \equiv \mathrm{Cov}(T_{k,a})$. Assume the following conditions hold. For some $\delta > 0$:

- $\mathrm{E}|X_{i,j}|^{2+\delta} \leqslant \Delta$ *for all* i *and all* $1 \leqslant j \leqslant p$, (16)
- $\Sigma_{k,a} \to \Sigma > 0$ *uniformly in* a, (17)
- $C(4) \equiv \sum_{k=1}^{\infty} (k+1)^2 \alpha^{\delta/(4+\delta)}(k) \leqslant K$. (18)
- $f : \mathbb{R}^k \to \mathbb{R}^p$ *is continuously differentiable.* (19)

Furthermore assume that $b/n \to 0$ and $b \to \infty$ as $n \to \infty$, and let $J(P) = \mathrm{N}(0, J\Sigma J^{\mathrm{T}})$, where J is the Jacobian of f at 0. Then, the conclusions of Theorem 2.1 *will be true.*

Example 3.4 (Least-squares linear regression). Consider the linear model $y = X\beta + \varepsilon$, where y and ε are $(n \times 1)$ vectors, β is a $(p \times 1)$ vector and X a $(n \times p)$ matrix. Here X may be stochastic. The goal is to draw inference on β. In order to be able to apply the subsampling method we need to define subvectors and submatrices:

$$y_{b,a} \equiv (y_a, \ldots, y_{a+b-1})^{\mathrm{T}}, \qquad \varepsilon_{b,a} \equiv (\varepsilon_a, \ldots, \varepsilon_{a+b-1})^{\mathrm{T}}$$

and

$$X_{b,a} \equiv \begin{pmatrix} x_a^{\mathrm{T}} \\ \vdots \\ x_{a+b-1}^{\mathrm{T}} \end{pmatrix}, \quad \text{where } X = \begin{pmatrix} x_1^{\mathrm{T}} \\ \vdots \\ x_n^{\mathrm{T}} \end{pmatrix}.$$

Our estimator of β based on $X_{b,a}$ and $y_{b,a}$ then is defined as

$$\hat{\beta}_{b,a} \equiv (X_{b,a}^{\mathrm{T}} X_{b,a})^{-1} X_{b,a}^{\mathrm{T}} y_{b,a}. \tag{20}$$

In the classical regression model the ε_i's are assumed to be i.i.d. Here, we require only a weak dependence condition. Define $J_{b,a}(P)$ to be the sampling distribution of $b^{1/2}(\hat{\beta}_{b,a} - \beta)$. Also, for any Borel set $A \in \mathbb{R}^p$, define

$$J_{b,a}(A,P) = \mathrm{Prob}_P\{b^{1/2}(\hat{\beta}_{b,a} - \beta) \in A\}. \tag{21}$$

The approximation to $J_{n,1}(A,P)$ we study is defined by

$$L_n(A) = \frac{1}{n-b+1} \sum_{a=1}^{n-b+1} 1\{b^{1/2}(\hat{\beta}_{b,a} - \hat{\beta}_{n,1}) \in A\}. \tag{22}$$

Theorem 3.4. Let $\{(x_i, \varepsilon_i)\}$ be a sequence of random vectors defined on a common probability space. Denote the mixing coefficients for the $\{(x_i, \varepsilon_i)\}$ sequence by $\alpha(\cdot)$. Define

$$T_{k,a} \equiv k^{-\frac{1}{2}} \sum_{i=a}^{a+k-1} x_i \varepsilon_i,$$

$$V_{k,a} \equiv \mathrm{Cov}(T_{k,a}),$$

$$M_{k,a} \equiv \mathrm{E}(X_{k,a}^{\mathrm{T}} X_{k,a}/k).$$

Assume the following conditions hold. For some $\delta > 0$:

- $\mathrm{E}(x_{i,j}\varepsilon_i) = 0$ *for all* i *and all* $1 \leqslant j \leqslant p$, (23)
- $\mathrm{E}|x_{i,j}\varepsilon_i|^{2+2\delta} \leqslant \Delta_1$ *for all* i *and all* $1 \leqslant j \leqslant p$, (24)
- $\mathrm{E}|x_{i,j}|^{4+2\delta} \leqslant \Delta_2$ *for all* i *and all* $1 \leqslant j \leqslant p$, (25)
- $V_{k,a} \to V > 0$ *uniformly in* a, (26)
- $M_{k,a} \to M > 0$ *uniformly in* a, (27)
- $C(4) \equiv \sum_{k=1}^{\infty} (k+1)^2 \alpha^{\delta/(4+\delta)}(k) \leqslant K$. (28)

Furthermore assume that $b/n \to 0$ and $b \to \infty$ as $n \to \infty$, and let $J(P) = \mathrm{N}(0, M^{-1}VM^{-1})$.
Then the conclusions of Theorem 2.1 *will be true.*

Remark 3.3. Alternatively, the moving blocks method could be used for making inference on the regression parameter β. The procedure is analogous to the mean case, except the resampled blocks consist of regressors and the corresponding dependent variables. A related result concerning the validity of the moving blocks method for multivariate least-squares linear regression in the context of heteroskedastic data was presented in Fitzenberger (1995). In a small simulation study in Section 5 we compare finite sample performance of the subsampling and the moving blocks method in a linear regression set-up.

4. Choice of the blocksize

An apparent drawback of all blocking methods is that for applications a blocksize has to be chosen. For small to moderate sample sizes, performance can

depend rather critically on that choice, for example, as shown in our simulation studies in the next section. So far, contributions in the literature on how to pick the blocksize in practice have been restricted to the case of stationary data, and we shall give a brief overview here.

It is well known that for quite general statistics both the subsampling and the moving blocks variance estimators are asymptotically equivalent to a kernel smoothed (with Bartlett's kernel) estimator of the spectral density at the origin (cf. Künsch, 1989; Politis and Romano, 1994b), where the bandwidth of the kernel corresponds to the inverse of the blocksize. The relevant spectral density is the one of the process given by the influence function of the statistic of interest. Hence for variance estimation, choosing the blocksize is asymptotically equivalent to choosing the bandwidth for the Bartlett kernel. Bühlmann and Künsch (1994) exploit this equivalence.

A more important and ambitious goal than estimating a variance is the construction of confidence intervals. In Hall et al. (1995) it is shown that the optimal asymptotic rate of the block size for the moving blocks method depends significantly on context, being equal to $n^{1/3}$, $n^{1/4}$ and $n^{1/5}$ in the cases of variance or bias estimation, estimation of a one-sided distribution function and estimation of a two-sided distribution function, respectively. The latter two quantities are needed for construction of equal-tailed and symmetric confidence intervals, respectively. Therefore, it seems that the strategy of Bühlmann and Künsch (1994) is suboptimal for constructing confidence intervals. Hall et al. (1995) present a practical rule for selecting the block size empirically. It is based on the fact the asymptotic formula is $b \sim Cn^{1/k}$, where $k = 3$, 4 or 5 is known, and C is a constant that depends on the underlying process. The rule suggested provides a way for estimating the optimal block size for a time series of smaller length than the original, say $m < n$. Once this has been determined, at $\hat{b}_m$ say, the optimal blocks size $\hat{b}_n$ for the original series of length n may be estimated from the formula $\hat{b}_n = (n/m)^{1/k}\hat{b}_m$.

It is not difficult to modify this empirical rule for the subsampling method. Unfortunately, it did not work very well when we tested it via simulation studies. In particular, as will become clear from later simulation studies and from Table 1 in the appendix that the optimal blocksize for making inference on the univariate mean depends very much on the actual degree of dependence, at the very least for AR(1) processes. Loosely speaking, longer blocks are needed to capture greater dependence. Hall's method, however, turned out to be quite insensitive to changing the AR(1) parameter in our simulations.

We propose a new method that in a certain sense avoids having to find the "best" blocksize. One can think of the accuracy of an approximate or asymptotic confidence procedure in terms of its *calibration* (Loh, 1987). Suppose we use the procedure to construct a confidence region with nominal confidence level $1 - \lambda$. We can denote the actual confidence level by $1 - \alpha$. λ is known to us, α typically is not. If we knew the calibration function $h: 1 - \lambda \to 1 - \alpha$, we could

Table 1
Mean, AR(1) model, $n = 256$

Gaussian innovations								
Parameter	Interval	$b = 1$	$b = 4$	$b = 8$	$b = 16$	$b = 32$	CAL	Nominal level
$\rho = 0.2$	Moving blocks	0.90	0.95	0.95	0.93	0.90		95%
$\rho = 0.2$	Subsampling		0.96	0.95	0.93	0.91	0.95	95%
$\rho = 0.5$	Moving blocks	0.76	0.89	0.92	0.92	0.90		95%
$\rho = 0.5$	Subsampling		0.93	0.93	0.91	0.90	0.94	95%
$\rho = 0.8$	Moving blocks	0.47	0.72	0.83	0.87	0.86		95%
$\rho = 0.8$	Subsampling		0.74	0.82	0.85	0.85	0.92	95%
$\rho = -0.5$	Moving blocks	1.00	0.98	0.96	0.94	0.90		95%
$\rho = -0.5$	Subsampling		0.98	0.96	0.94	0.90	0.95	95%
Exponential innovations								
Parameter	Interval	$b = 1$	$b = 4$	$b = 8$	$b = 16$	$b = 32$	Cal.	Nominal level
$\rho = 0.2$	Moving blocks	0.90	0.94	0.95	0.93	0.90		95%
$\rho = 0.2$	Subsampling		0.96	0.95	0.92	0.90	0.94	95%
$\rho = 0.5$	Moving blocks	0.71	0.86	0.89	0.89	0.87		95%
$\rho = 0.5$	Subsampling		0.89	0.87	0.87	0.86	0.94	95%
$\rho = 0.8$	Moving blocks	0.46	0.73	0.81	0.86	0.85		95%
$\rho = 0.8$	Subsampling		0.72	0.80	0.83	0.83	0.92	95%
$\rho = -0.5$	Moving blocks	1.00	0.95	0.93	0.92	0.88		95%
$\rho = -0.5$	Subsampling		0.95	0.92	0.90	0.88	0.94	95%

construct a confidence region with exactly the desired coverage, by selecting the value of λ that satisfies $h(1 - \lambda) = 1 - \alpha$. Fortunately, the calibration function $h(\cdot)$ can be estimated by bootstrap methods. The basic bootstrap principle applies here as well. One generates pseudo-sequences from a known model P_n^*, then constructs confidence regions from each generated pseudo-sequence, and observes how frequently the parameter $\hat{\theta}_n$ is contained in those regions. In the context of dependent data, one needs to employ a more suitable bootstrap method, such as the moving blocks bootstrap or the stationary bootstrap, in order to generate appropriate bootstrap samples.

In the case we want to apply the calibration scheme to a blocking method, we can do it conditional on a *reasonable* blocksize. This means that we fix a sensible blocksize and calibrate the confidence procedure using that particular blocksize. The problem of finding the "best" blocksize is therefore reduced to the problem of finding a reasonable blocksize, which is a lot easier. In many scenarios we will have a pretty good idea what a reasonable blocksize will be, either from prior experience or related simulation studies. In case, we are in the dark with respect to how to pick a reasonable blocksize, see Remark 4.1 below.

To describe the general method more formally we can use the following algorithm.

4.1. Description of the calibration method

1. Generate K pseudo-sequences $X_1^{*k},\ldots,X_n^{*k}$, according to a suitable bootstrap distribution P_n^*.
   For each sequence, $k=1,\ldots,K$,
   1a. Compute an $1-\lambda$ level confidence region $\mathrm{CI}_{1-\lambda}^k$, for a grid of values of λ.
2. For each λ compute $\hat{h}(1-\lambda)=\#\{\hat{\theta}_n\in \mathrm{CI}_{1-\lambda}^k\}/K$.
3. Interpolate $\hat{h}(\cdot)$ between the grid values.
4. Find the value of λ satisfying $\hat{h}(1-\lambda)=1-\alpha$.
5. Construct a confidence interval with nominal level $1-\lambda$.

Remark 4.1. 1. A "suitable bootstrap method" in step 1 of the above requires a block size b_{MB} in the case of the moving blocks method and the equivalent counterpart p in the case of the stationary bootstrap. The choice of this block size will have a second-order effect and is, therefore, less important. To be on the safer side, a nested bootstrap could be employed to determine an appropriate block size. This means that we would use the same bootstrap method in steps 1 and 1a of the above algorithm with the *same* block size $b_{\mathrm{MB}}B$ (or the same p), limiting the grid of λ values to $\lambda=\alpha$. Repeating this algorithm for a number of b_{MB} (p) values, we then would select the value b_{MB} (p) that yields estimated coverage closest to $1-\alpha$.

2. If we use the calibration scheme to calibrate a blocking method, we need to start out with a reasonable block size. In situations, where we do not know what a reasonable block size is, we can use the following trick. In the same way as the actual confidence level can be regarded as function of the nominal confidence level (conditional on a fixed block size), it can be considered as a function of the block size (conditional on a fixed nominal level). Fixing the nominal level at the desired level, that is, choosing $\lambda=\alpha$, we can therefore estimate the block calibration function $g: b\to 1-\alpha$, using an analogous calibration algorithm:

   1* Generate K pseudo-sequences $X_1^{*k},\ldots,X_n^{*k}$, according to a suitable bootstrap distribution P_n^*.
   For each sequence, $k=1,\ldots,K$,
   1a^*. Compute an $1-\alpha$ level confidence region CI_b^k, for a selection of block sizes b.
   2* For each b compute $\hat{g}(b)=\#\{\hat{\theta}_n\in \mathrm{CI}_b^k\}/K$.
   A reasonable block size will then satisfy $\hat{g}(b)\simeq 1-\alpha$.

3. Two-sided equal-tailed intervals should always be computed as the intersection of two separately calibrated one-sided intervals. Particularly if the sampling distribution of $\hat{\theta}_n$ is asymmetric, the amount of calibration needed in the lower tail can be very different from the one needed in the upper tail.

It is obvious that calibrating an asymptotically correct procedure will again result in an asymptotically correct procedure. The hope is that the calibrated procedure will have better finite sample properties. In the context of i.i.d. observations, it has been shown that calibrating a procedure increases its asymptotic correctness. Two comprehensive references are Efron and Tibshirani (1993) and Hall (1992). Proving such a result in the context of dependent data is beyond the scope of this paper. We will shed some light on the performance of the calibrated subsampling method by means of simulation studies. Further research will be necessary to further explore theoretical and practical properties.

5. Simulation studies

In this section we will shed some light on the finite sample properties of the subsampling method and also the moving blocks method. We present two simulation studies, one for the case of the univariate sample mean and one for multivariate least-squares linear regression. Tables for both simulation studies are provided in Appendix C. One should note that simulation studies on blocking methods are computationally fairly expensive. This is even more true in the case of calibrated blocking methods. Since the subsampling method is a lot faster than the moving blocks bootstrap we restricted the calibration method to subsampling intervals for the sake of these simulation studies. Also, to keep the computations manageable we had to limit the number K of pseudo-sequences to estimate the calibration function $h(\cdot)$ to $K = 200$. For any real application, one should use $K = 1000$.

5.1. The univariate mean

In this simulation study we compare the finite sample performance of subsampling and moving blocks confidence intervals for the univariate mean. Performance is measured in terms of coverage probability of two-sided 95% level intervals. In order to generate the data we use heteroskedastic autoregressive AR(1) processes with different parameters. The innovations are independent but heteroskedastic. To be more specific, let $\tilde{v}_i$ be i.i.d. random variables. We then define weakly dependent, heteroskedastic time series by

$$\text{AR(1)}: \quad x_i = \rho x_{i-1} + v_i \quad \text{and} \quad v_i = a_i \tilde{v}_i.$$

Here $\{a_i\}$ denotes a sequence of real numbers that might be regarded as seasonal effects. Throughout, we chose $\{a_i\}$ to be the infinite repetition of the sequence $\{1, 1, 1, 2, 3, 1, 1, 1, 1, 2, 4, 6\}$. Without loss of generality, we fix the true mean at zero. This can be achieved by choosing a distribution with zero mean for the unscaled innovations, $\tilde{v}_i$. For our simulations we use two distributions, standard normal and centered exponential with scale parameter 1. As parameters for the AR(1) model $\rho = 0.2$, 0.5, 0.8 and -0.5 are considered.

We can quickly check that the assumptions of Theorem 3.1 are satisfied. The moment condition (7) and asymptotic covariance stationarity, condition (8), hold by our choice of distributions for the innovations and of the sequence of constants for the 'seasonal effects'. By Theorem 5 in Section 2.4 of Doukhan (1994), the strong mixing coefficients of the AR(1) and AR(2) processes $\{x_i\}$ are geometrically decreasing which implies (9).

Since we are drawing inference from dependent and heteroskedastic data, we cannot hope to do very well for small data sets. In the simulations we chose $n = 256$ as the sample size. 1000 random samples were generated for each scenario. The moving blocks intervals are based on 1000 resamples for each sample. The blocksizes range from $b = 1$ to $b = 32$ for the moving blocks method and from $b = 4$ to $b = 32$ for the subsampling method.

In Remark 2.2 we discussed equal-tailed and symmetric intervals. In all our simulations we found almost equal coverage probabilities. The difference between the two approaches was almost always less than 1% with no consistent winner. Since coverage was often off by several percent from the nominal level, we decided that this difference was not worth mentioning in comparison, and therefore only report the results for equal-tailed intervals.

Table 1 provides the results. We first discuss the findings for fixed blocksizes. As a first observation, it is striking how close the two methods are in terms of coverage probability. The difference is less than or equal to 1% in most cases, with no method dominating the other. Secondly, finite sample performance is far from perfect, and gets worse as the degree of dependence in the data increases. Thirdly, the distribution of the innovations certainly makes a difference, with somewhat worse results for exponential than for normal innovations.

In the AR(1) case we get satisfactory results only for $\rho = 0.2$. For blocksizes $b = 4$ and $b = 8$ coverage probabilities seem accurate, but for bigger blocksizes the confidence intervals definitely undercover. When increasing ρ we find that the optimal blocksize increases while overall performance decreases. The confidence intervals undercover for all block sizes, with relatively best performance at $b = 8$ and $b = 16$ for $\rho = 0.5$, and $b = 16$ and $b = 32$ for $\rho = 0.8$, respectively. As expected, Efron's bootstrap ($b = 1$) always does worst. For a negative autocorrelation parameter ($\rho = -0.5$) the story is different. We find overcoverage for small and undercoverage for large blocksizes, with exact coverage somewhere in between.

At this point two things have become clear. First, the choice of the blocksize definitely matters. Picking a "wrong" blocksize will yield bad finite sample performance. Secondly, in some scenarios even the "best" blocksize does not very well. For our simulations with $\rho = 0.8$ we found that the intervals undercover on the order of 10%, at best. We proposed the calibration method hoping that in such instances it will pick an appropriate nominal level, resulting in improved coverage. The results for calibrated symmetric subsampling intervals are also reported in Table 1. As starting points for the reasonable blocksize in our calibration algorithm we used $b = 10$ for $\rho = 0.2$, 0.5 and -0.5, and $b = 20$ for $\rho = 0.8$. Again, in any real application, a reasonable starting blocksize can be easily chosen by a block calibration method as discussed before. Except for $\rho = 0.8$, the calibrated subsampling intervals are right on target. For $\rho = 0.8$, they still underover by about 3%, but this is a considerable improvement over the "best" blocksize alone. Also, estimated coverage probabilities are basically identical for normal and exponential innovations, showing robustness of the calibration method towards skewed sampling distributions.

5.2. *Multivariate linear regression*

Our second simulation study is concerned with multivariate least squares linear regression. We assess finite sample performance of the subsampling and the moving blocks methods for inference on a (single) regression parameter. Performance is measured in terms of coverage probability of two-sided 95% confidence intervals. For comparison we also include some more standard kernel estimation approaches. Here the strategy is to find confidence intervals based on the t statistic constructed using the OLS regression coefficient estimator and a variance estimator obtained by a kernel technique. Andrews (1991) compared various kernels applied to covariance estimation in multivariate linear regression. He found that the so-called quadratic spectral (QS) kernel has certain asymptotic optimality properties, which were then substantiated in a Monte Carlo study. In a follow-up paper, Andrews and Monahan (1992) suggested prewhitened kernel estimators as an improvement over regular kernel estimators, at least when coverage probabilities are of main interest. Again, the prewhitened QS (QS-PW) kernel seems favored over other kernels. We were interested to see whether the subsampling or the moving blocks method could improve upon either of the QS kernel approaches. In addition, we decided to include another kernel in our simulation study, the Bartlett (BT) kernel, partly because it can be considered more traditional in the economics literature, and partly due to its close connection to the subsampling and moving blocks variance estimators. For all kernel methods, an automatic bandwidth selection, as proposed by Andrews (1991), is employed in our simulation study.

We use the same basic model as in Andrews (1991), given by

$$y_i = x_i^{\mathrm{T}} \beta + \varepsilon_i, \tag{29}$$

where $x_{i,1}=1$ and $x_i=(1,\tilde{x}_i)^T$ and β_0 are 5×1 vectors. Throughout we are concerned with construction confidence intervals for the regression parameter β_2. Without loss of generality, we set $\beta=\mathbf{0}$. As far as the data generating processes are concerned, we restrict ourselves to one of the processes considered by Andrews (1991), namely AR(1)-HET1, in which the errors and regressors are AR(1) processes with multiplicative heteroskedasticity overlaid on the errors. To be more specific, let $\tilde{\varepsilon}_i$ and the components of $\tilde{x}_i$ be independent random variables generated according to the same AR(1) model. Then we define

$$\text{AR(1)-HET1: } x_{i,j}=\rho x_{i-1,j}+v_{i,j},\ j=2,\ldots,5;$$

$$\tilde{\varepsilon}_i=\rho\tilde{\varepsilon}_{i-1}+v_i^{\varepsilon} \text{ and } \varepsilon_i=|x_{i,2}|\,\tilde{\varepsilon}_i.$$

In Andrews' notation HET stands for some kind of heteroskedasticity, although one should note that for this process the ε_i are only conditionally heteroskedastic. The same is true for all other HET models in Andrews (1991). Nevertheless, even conditional heteroskedasticity should make inference on β_2 more difficult. As far as the distribution of the innovations v_i^{ε} goes, Andrews (1991) did not venture beyond normality. In analogy to our simulation study for the univariate mean, however, we also examine exponential innovations. The values considered for the AR(1) parameter ρ are 0.2, 0.5, 0.8 and -0.5. Note also that in Andrews (1991) and Andrews and Mohanan (1992) both regressor and error variables were scaled so as to always have unconditional variance equal to one. We decided not to follow this strategy, but rather chose standard normal and centered exponential (with scale parameter 1) as distributions for the innovations. Thus, the variance of the regressors and the errors increases with the degree of dependence, which is more realistic. The sample size considered for our simulations is $n=128$, mainly since most results in the two papers by Andrews et al. ar for this size. 1000 random samples were generated for each scenario. The moving blocks intervals are based on 1000 resamples for each sample. The blocksizes considered range from $b=1$ to $b=32$ for the moving blocks method and from $b=5$ to $b=32$ for the subsampling method. As in the univariate mean case, we only compute calibrated subsampling intervals. We used $b=10$ as a starting blocksize for the calibration method.

We can quickly check that the conditions of Theorem 3.4 are satisfied.The moment conditions (23)–(25) are met by our choice for the distributions of the innovations. Asymptotic covariance stationarity [conditions (26) and(27)] follows trivially from the second order stationarity of the stochastic processes. This leaves to verify the mixing condition (28). By Theorem 5 in Section 2.4 of Doukhan (1994), the strong mixing coefficients of the (independent) AR(1) processes $x_{i,j}$, $j=2,\ldots,5$ and $\tilde{\varepsilon}_i$ are geometrically decreasing. By Theorem 3.49 in

Table 2
AR(1)-HET1 model, $n = 128$

Gaussian innovations										
Parameter	Interval	$b=1$	$b=5$	$b=6$	$b=7$	$b=8$	$b=12$	$b=16$	$b=32$	Nom. level
$\rho=0.2$	MB_{ET}	0.91	0.91	0.91	0.90	0.89	0.88	0.86	0.78	95%
$\rho=0.2$	SUB_{ET}		1.00	0.99	0.96	0.93	0.90	0.87	0.77	95%
$\rho=0.2$	MB_{SYM}	0.92	0.92	0.92	0.92	0.92	0.91	0.90	0.86	95%
$\rho=0.2$	SUB_{SYM}		1.00	0.99	0.97	0.95	0.92	0.89	0.81	95%
$\rho=0.5$	MB_{ET}	0.84	0.87	0.86	0.86	0.86	0.85	0.83	0.76	95%
$\rho=0.5$	SUB_{ET}		1.00	0.98	0.94	0.90	0.87	0.85	0.73	95%
$\rho=0.5$	MB_{SYM}	0.85	0.89	0.89	0.89	0.90	0.89	0.88	0.85	95%
$\rho=0.5$	SUB_{SYM}		1.00	0.99	0.96	0.94	0.90	0.88	0.80	95%
$\rho=0.8$	MB_{ET}	0.61	0.75	0.77	0.76	0.76	0.76	0.75	0.69	95%
$\rho=0.8$	SUB_{ET}		1.00	0.93	0.87	0.83	0.80	0.77	0.66	95%
$\rho=0.8$	MB_{SYM}	0.62	0.79	0.81	0.82	0.83	0.84	0.84	0.82	95%
$\rho=0.8$	SUB_{SYM}		1.00	0.97	0.91	0.88	0.85	0.82	0.79	95%
$\rho=-0.5$	MB_{ET}	0.85	0.88	0.88	0.87	0.87	0.86	0.85	0.79	95%
$\rho=-0.5$	SUB_{ET}		1.00	0.97	0.92	0.88	0.87	0.85	0.77	95%
$\rho=-0.5$	MB_{SYM}	0.85	0.89	0.90	0.90	0.90	0.90	0.89	0.85	95%
$\rho=-0.5$	SUB_{SYM}		1.00	0.98	0.95	0.93	0.90	0.89	0.84	95%
Exponential innovations										
Parameter	Interval	$b=1$	$b=5$	$b=6$	$b=7$	$b=8$	$b=12$	$b=16$	$b=32$	Nom. level
$\rho=0.2$	MB_{ET}	0.82	0.81	0.82	0.81	0.81	0.79	0.78	0.70	95%
$\rho=0.2$	SUB_{ET}		1.00	0.93	0.85	0.81	0.78	0.76	0.69	95%
$\rho=0.2$	MB_{SYM}	0.86	0.86	0.86	0.87	0.86	0.85	0.85	0.82	95%
$\rho=0.2$	SUB_{SYM}		1.00	0.98	0.93	0.89	0.85	0.83	0.79	95%
$\rho=0.5$	MB_{ET}	0.78	0.80	0.80	0.80	0.79	0.78	0.76	0.70	95%
$\rho=0.5$	SUB_{ET}		1.00	0.94	0.87	0.82	0.78	0.76	0.68	95%
$\rho=0.5$	MB_{SYM}	0.82	0.86	0.86	0.86	0.86	0.86	0.85	0.82	95%
$\rho=0.5$	SUB_{SYM}		1.00	0.98	0.93	0.90	0.85	0.84	0.80	95%
$\rho=0.8$	MB_{ET}	0.57	0.73	0.72	0.72	0.73	0.72	0.71	0.64	95%
$\rho=0.8$	SUB_{ET}		1.00	0.93	0.87	0.83	0.78	0.74	0.64	95%
$\rho=0.8$	MB_{SYM}	0.59	0.77	0.79	0.79	0.80	0.82	0.82	0.80	95%
$\rho=0.8$	SUB_{SYM}		1.00	0.97	0.91	0.88	0.83	0.81	0.75	95%
$\rho=-0.5$	MB_{ET}	0.83	0.84	0.84	0.83	0.83	0.82	0.80	0.74	95%
$\rho=-0.5$	SUB_{ET}		1.00	0.92	0.87	0.84	0.83	0.80	0.72	95%
$\rho=-0.5$	MB_{SYM}	0.85	0.89	0.89	0.89	0.89	0.88	0.87	0.83	95%
$\rho=-0.5$	SUB_{SYM}		1.00	0.96	0.92	0.89	0.88	0.86	0.82	95%

White (1984) the same type of decay holds for the sequence (x_i, ε_i). This, of course, implies condition (28).

The results of our study are provided in Tables 2 and 3. Again, we first briefly discuss the results for fixed blocksizes, given in Table 2. In some respects our

Table 3
AR(1)-HET1 model, $n = 128$

Gaussian innovations					
Parameter	BT	QS	QS-PW	SUB-CAL	Nominal level
$\rho = 0.2$	0.92	0.92	0.92	0.96	95%
$\rho = 0.5$	0.88	0.88	0.88	0.94	95%
$\rho = 0.8$	0.81	0.82	0.79	0.91	95%
$\rho = -0.5$	0.89	0.89	0.90	0.94	95%
Exponential innovations					
Parameter	BT	QS	QS-PW	SUB-CAL	Nominal level
$\rho = 0.2$	0.86	0.86	0.85	0.95	95%
$\rho = 0.5$	0.84	0.84	0.82	0.92	95%
$\rho = 0.8$	0.75	0.76	0.72	0.89	95%
$\rho = -0.5$	0.87	0.88	0.89	0.92	95%

findings are very different from those in the case of the univariate mean. Firstly, the subsampling and the moving blocks method no longer closely agree. The subsampling method is more sensitive to the choice of the blocksize, but gives better results for "the right choice". Fortunately, this choice is not affected very much by the degree of dependence. Secondly, symmetric intervals have significantly better coverage than two-sided intervals. Thirdly, the kind of dependence seems to affect coverage probabilities in a different manner. We do not see an inherently different effect of positive and negative AR(1) parameters. The moving blocks intervals undercover consistently, whereas subsampling intervals overcover for small blocksizes and undercover for large blocksizes, and this is true for *both* positive and negative AR(1) parameters.

The results of the kernel methods and the calibrated subsampling method are given in Table 2. The subsampling intervals clearly exhibit better coverage properties than the kernel intervals and are close to the target level, except for $\rho = 0.8$. The kernel intervals undercover consistently. Interestingly, the prewhitened kernel QS-PW seems not necessarily superior to the simple QS kernel. Especially if the degree of dependence is high, it appears to undercover even more. This is somewhat in contrast to the results in Andrews and Mohanan (1992). For exponential innovations, coverage probabilities are reduced for all intervals compared to normal intervals, typically in the range of 2 to 5%. However, the coverage reduction is largest for the kernel methods, indicating that they are somewhat less robust against the distribution of the innovations than subsampling or moving blocks methods. It seems that in using a normal approximation the kernel methods are paying a price for ignoring the potential skewness of the distribution of an estimator.

6. Variance ratio tests

A stylized version of the Efficient Market Hypothesis states that stock returns are serially independent and hence unpredictable. Combined with the additional assumption that log returns are identically distributed according to some normal distribution, this yields the historic log-random-walk-model, which has become the workhorse of the financial asset pricing theory. For example, it is the basis of the celebrated Black and Scholes (1973) model for pricing derivative securities. A weaker and maybe more interesting hypothesis is that log returns are serially uncorrelated, but possibly dependent. However, a number of recent papers have challenged these traditional views claiming that stock returns can be partially predicted, be it from past returns or other variables (e.g., the stock's dividend yield). A common way of testing predictability of returns from past returns has been the use of variance ratio tests. These tests use the simple fact that the variance of the sum of uncorrelated random variables is equal to the sum of the individual variances.

Assume that we observe the price of a stock at equal time intervals i. Define the one-period real total return as $R_{i+1} = (P_{i+1} + d_{i+1})/P_i$, where P_i is the end-of-period real stock price and d_i is the real dividends paid during period i. Here d_i could be zero in case the stock does not pay dividends. Assume that the log returns $r_i \equiv \log(R_i)$ are covariance stationary. We are interested in the null hypothesis of uncorrelated returns:

H_0 : $r_i \equiv \log(R_i)$ are serially uncorrelated with common mean μ and variance σ^2.

Suppose our sample consists of mq observations $r_1, \ldots, r_{mq}$. Unbiased estimators of μ and σ^2 are given by

$$\hat{\mu} = \frac{1}{mq} \sum_{i=1}^{mq} r_i, \tag{30}$$

$$\hat{\sigma}_1^2 = \frac{1}{mq - 1} \sum_{i=1}^{mq} (r_i - \hat{\mu})^2. \tag{31}$$

Under H_0 the following is also an unbiased estimator of σ^2:

$$\hat{\sigma}_q^2 = \frac{1}{l} \sum_{i=q}^{mq} \left(\sum_{k=1}^{q} r_{i+1-k} - q\hat{\mu} \right)^2, \tag{32}$$

where $l = q(mq - q + 1)(1 - q/(mq))$. However, if the log returns are serially correlated, this is no longer true. Under positive serial correlation $\hat{\sigma}_q^2$ will tend to be greater than σ^2, under negative serial correlation it will tend to be smaller than σ^2. A q-period variance ratio test statistic can therefore be defined by

$$M_r(q) = \frac{\hat{\sigma}_q^2}{\hat{\sigma}_1^2}. \tag{33}$$

It can easily seen to be a consistent (although biased) estimator of the general q-period variance ratio statistic

$$VR(q) = 1 + 2\sum_{k=1}^{q-1}(1 - k/q)\rho_k,$$

where ρ_k is the k-order autocorrelation coefficient of $\{r_i\}$.

The statistic $M_r(q)$ was introduced by Lo and MacKinlay (1988). It is obvious that under H_0 it should be close to unity. If the log returns r_i exhibit positive serial correlation, it will tend to be greater then unity, and vice versa for negative serial correlation. Under the null hypothesis, $M_r(q)$ will have a limiting normal distribution with mean one, which is even robust against some heteroskedasticity. However, the limiting variance very much depends on the dependency structure of the r_i and is nontrivial to estimate, see Lo and MacKinlay (1988). Applying this methodology for periods $q = 2, 4, 8$, and 16 to weekly data of the CRSP equal- and value-weighted indices from 12/1962 to 12/1992, Campbell et al. (1995) are able to reject the null hypothesis of uncorrelated r_i.

Malliaropulos (1996) employs a bootstrap approach to variance ratio tests in order to avoid having to rely on a limiting normal distribution. Using monthly observations of the FT-A All Share index from 01/1964 to 09/1993, he is not able to reject the stronger null hypothesis of i.i.d. r_i. Clearly, this is in contrast to the findings of Campbell et al. (1995). One should note however, that this bootstrap approach is only able to test the stronger i.i.d. hypothesis. In general, this is an undesirable property, since there is already a growing consensus among financial economists that (conditional) return volatilities change over time and second moments of log returns are correlated, which implies that returns are dependent. Moreover, for testing the stronger null hypothesis exact tests could be constructed.

Applying the subsampling method to variance ratio tests has a two-fold advantage. One does not have to estimate a complicated limiting distribution and it also works for dependent and heteroskedastic returns. We look at the CRSP equal- and value-weighted indices from 12/1947 to 12/1986 and the S&P 500 index from 12/1947 to 12/1994. All three data sets consist of monthly observations. Our strategy is to construct 95% calibrated subsampling intervals for the variance ratio statistic $VR(q)$ at various periods q and to check whether unity is contained in the intervals. We consider the periods $q = 2, 4$, and 8. We employed blocksizes of $b = 16, 32$, and 48, respectively, for the calibration method outlined in Section 4. The results are presented in Table 4 in Appendix C. While we find evidence for predictability for the CRSP equal-weighted index, we fail to do so for the CRSP value-weighted index and for the S&P 500 index. These results are somewhere in the middle between the findings of Campbell et al. (1995) and Malliaropulos (1996). However, one needs to take into account that we used different data sets – Campbell et al. analyzed *weekly* CRSP observations.

Table 4
Confidence Intervals for $VR(q)$

CRSP equal-weighted		
Period	$M_r(q)$	95% CI
$q=2$	1.153	[1.09, 1.22]
$q=4$	1.236	[1.07, 1.40]
$q=8$	1.367	[1.08, 1.65]
CRSP value-weighted		
Period	$M_r(q)$	95% CI
$q=2$	1.06	[1.01, 1.11]
$q=4$	1.109	[0.97, 1.24]
$q=8$	1.312	[0.99, 1.63]
S&P 500		
Period	$M_r(q)$	95% CI
$q=2$	1.023	[0.97, 1.08]
$q=4$	0.998	[0.90, 1.10]
$q=8$	1.066	[0.88, 1.26]

It is apparent that this brief discussion does not fully do justice to the problem of variance ratio tests. For one, the different methods should be applied to identical data sets in order to be really able to compare their results. Also, simulation studies seem in order to address finite sample performances. These issues are beyond the scope of this paper and will be explored in further research.

7. Conclusion

In this paper we have demonstrated that the subsampling method is a valid tool in heteroskedastic settings. In the stationary case, the asymptotic validity hinged on the simple assumption of a limit distribution for the normalized statistic based on the entire sample. Now, an additional sufficient condition is that the normalized statistic based on a subseries will be on average close to the same limiting distribution, at least for large samples. This allows for considerable local heteroskedasticity, and it applies to inference for multivariate parameters. To demonstrate the validity of the method for the mean, smooth functions of a multivariate mean and multivariate linear regression, we employed a central limit theorem for a triangular array of heteroskedastic weakly dependent random variables. We state this theorem together with a moment bound for heteroskedastic random variables in the appendix as results interesting in their own rights.

Preliminary simulation studies showed that the finite sample performance of blocking methods can strongly depend on the blocksize. Unfortunately, asymptotic theory provides little guidance as how to select a blocksize in practice. We proposed a calibration approach to deal with this problem. The idea is to start out with a reasonable blocksize and to then improve coverage accuracy by adjusting the nominal confidence level so as to better match the desired target confidence level of a confidence region.

As has been done in the context of i.i.d. data, it might be possible to derive second-order properties for calibrated confidence procedures in the context of dependent data. Due to the very involved mathematics of Edgeworth expansions for dependent data, such results basically would have to be restricted to the strictly stationary case. Future research will be devoted to this problem.

Two simulation studies were carried out for the cases of the univariate mean and multivariate least squares linear regression. They demonstrate that the calibration method yields good results, even when compared with the "best" blocksize. For the linear regression case we also compared the subsampling method with more standard kernel methods, as widely used in the econometrics literature. Our simulation studies suggest that calibrated subsampling intervals can result in considerable improvement for finite sample scenarios, in particular when the regression residuals exhibit strong serial correlation.

As an example, we applied the subsampling methodology to the problem of variance ratio tests, which are commonly used to check for serial correlations of random variables. Our findings using stock return data are somewhat different than results of more traditional methods relying on limiting normality and of a bootstrap method. We believe that this is one of many areas where subsampling can shed new light on issues that maybe have not been solved to complete satisfaction yet.

Appendix A. Central limit theorem for triangular arrays

Before applying our basic theorems for the construction of confidence regions, we will need a method to verify Assumptions A or B. In this section, we present a central limit theorem for a triangular array of weakly dependent heteroskedastic random variables. Central limit theorem for strong mixing random variables have been proved by Rosenblatt (1958), Ibragimov (1962), Oodaira and Yoshihara (1972), White and Domowitz (1984) and many others. A survey of the literature can be found in Doukhan (1994). Note that in many cases strict stationarity was assumed in addition to moment and mixing conditions. Our theorem is an extension of previous results, as it applies to triangular arrays. For the proof of the theorem we need the following moment bound for strong mixing heteroskedastic random variables. The result is implicitly contained in a theorem of Doukhan (1994).

Lemma A.1 (Moment bound, Doukhan, 1994). *Let $\{X_i\}$ be a sequence of mean zero random variables. Denote the corresponding mixing sequence by $\alpha_X(\cdot)$. Define, for $\tau \geqslant 2$ and $\delta > 0$*

$$C(\tau, \delta) \equiv \sum_{k=0}^{\infty} (k+1)^{\tau-2} \alpha_X^{\delta/(\tau+\delta)}(k), \tag{A.1}$$

$$L(\tau, \delta, d) \equiv \sum_{i=1}^{d} \|X_i\|_{\tau+\delta}^{\tau}, \tag{A.2}$$

$$D(\tau, \delta, d) \equiv \text{Max}\{L(\tau, \delta, d), [L(2, \delta, d)]^{\tau/2}\}. \tag{A.3}$$

Then the following bound holds:

$$\mathrm{E}\left|\sum_{i=1}^{d} X_i\right|^{\tau} \leqslant BD(\tau, \delta, d),$$

where B is a constant only depending on τ, δ and the mixing coefficients $\alpha_X(\cdot)$. We will be specific about the constant B for the two special cases τ is an even integer and $\tau = 2 + \delta$.

1. *If c is an even integer*

$$\mathrm{E}\left|\sum_{i=1}^{d} X_i\right|^{c} \leqslant B(c, \delta) D(c, \delta, d), \tag{A.4}$$

where bounds for the constants $B(c, \delta)$ can be computed recursively. For example, for c up to 4:

$$B(1, \delta) \leqslant 1,$$

$$B(2, \delta) \leqslant 18\,\text{Max}\{1, C(2, \delta)\},$$

$$B(3, \delta) \leqslant 102\,\text{Max}\{1, C(3, \delta)\},$$

$$B(4, \delta) \leqslant 3024\,\text{Max}\{1, C^2(4, \delta)\}.$$

2. *For $\tau = 2 + \delta$,*

$$B \leqslant [3024\,\text{Max}\{1, C^2(4, \delta)\}] 2^{4[4(2-\delta)/\delta+1)}. \tag{A.5}$$

In case we have an uniform bound on the $2 + 2\delta$ moments of the sequence $\{X_i\}$ we can obtain a less sharp but more concisely stated bound.

Corollary A.1 (Concise moment bound). *If we assume in addition to the conditions of Lemma* A.1 *that*

$$\|X_i\|_{2+2\delta} \leqslant \Delta \quad \textit{for all } i, \tag{A.6}$$

then

$$\mathrm{E}\left|\sum_{i=1}^{d} X_i\right|^{2+\delta} \leqslant \Gamma d^{1+\delta/2},$$

where Γ *is a constant that only depends on* Δ, δ *and the mixing coefficients* $\alpha_X(\cdot)$. *More explicitly,*

$$\Gamma = [3024 \operatorname{Max}\{1, C^2(4,\delta)\}] 2^{4(4(2-\delta)/\delta+1)} \Delta^{(2+\delta)(1+\delta/2)},$$

where $C(4,\delta)$ *is defined as in* (A.1).

We now present the central limit theorem for triangular arrays.

Theorem A.1. Let $\{X_{n,i}, 1 \leqslant i \leqslant d_n\}$ *be a triangular array of mean zero random variables. Denote the mixing sequence corresponding to the nth row by* $\alpha_n(\cdot)$. *Define*

$$S_{n,k,a} \equiv \sum_{i=a}^{a+k-1} X_{n,i},$$

$$T_{n,k,a} \equiv k^{-\frac{1}{2}} \sum_{i=a}^{a+k-1} X_{n,i},$$

$$\sigma_{n,k,a}^2 \equiv \operatorname{Var}(T_{n,k,a}).$$

Assume the following conditions hold: For some $\delta > 0$:

- $\|X_{n,i}\|_{2+2\delta} \leqslant \Delta$ *for all* n, i, (A.7)
- $\sigma_{n,k,a}^2 \to \sigma^2 > 0$ *uniformly in* a(*), (A.8)
- $C_n(4) \equiv \sum_{k=0}^{\infty} (k+1)^2 \alpha_n^{\delta/(4+\delta)}(k) \leqslant K$ *for all* n, (A.9)

where Δ *and* K *are finite constants independent of* n, k *or* a.

(*) *This means: For any sequence* $\{k_n\}$ *that tends to infinity with* n,

$\sup_a |\sigma_{n,k_n,a}^2 - \sigma^2| \to 0$ *as* $n \to \infty$.

Then $T_{n,d_n,1} \Rightarrow \mathrm{N}(0,\sigma^2)$, *i.e.* $d_n^{-1/2} \sum_{i=1}^{d_n} X_{n,i} \Rightarrow \mathrm{N}(0,\sigma^2)$.

Appendix B. Proofs of technical results

Proof of Theorem 2.1. For a set A denote its interior, closure and boundary by A^0, A^- and $\partial(A)$, respectively. For a set A and a positive constant ε define sets $A_{-\varepsilon} \subseteq A \subseteq A_{+\varepsilon}$ in the following way. Set $M_{A,\varepsilon} \equiv \bigcup_{x \in \partial(A)} B(x,\varepsilon)$, where $B(x,\varepsilon)$

denotes the (closed) ball with center x and radius ε. Then $A_{+\varepsilon} \equiv A \cup M_{A,\varepsilon}$ and $A_{-\varepsilon} \equiv A \backslash M_{A,\varepsilon}$.

We will now prove the theorem. To simplify the notation, let $q \equiv q_n \equiv n-b+1$. Let

$$U_n(A) = \frac{1}{q} \sum_{a=1}^{q} 1\{\tau_b(\hat{\theta}_{b,a} - \theta(P)) \in A\}.$$

To prove (i), it suffices to show that $U_n(A)$ converges in probability to $J(A,P)$ for every Borel set A whose boundary has measure zero under $J(P)$. This can be seen by noting that

$$L_n(A) = \frac{1}{q} \sum_{a=1}^{q} 1\{\tau_b(\hat{\theta}_{b,a} - \theta(P)) + \tau_b(\theta(P) - \hat{\theta}_{n,1}) \in A\},$$

so that for every $\varepsilon > 0$,

$$U_n(A_{-\varepsilon}) 1\{E_n\} \leqslant L_n(A) 1\{E_n\} \leqslant U_n(A_{+\varepsilon}),$$

where $1\{E_n\}$ is the indicator of the event $E_n \equiv \{\tau_b \| \theta(P) - \hat{\theta}_{n,1} \| \leqslant \varepsilon\}$. But, the event E_n has probability tending to one. So, with probability tending to one,

$$U_n(A_{-\varepsilon}) \leqslant L_n(A) \leqslant U_n(A_{+\varepsilon}).$$

Thus, if $A_{+\varepsilon}$ and $A_{-\varepsilon}$ are Borel sets whose boundaries have mass zero under $J(P)$, then $U_n(A_{\pm\varepsilon}) \to J(A_{\pm\varepsilon}, P)$ in probability implies

$$J(A_{-\varepsilon}, P) - \varepsilon \leqslant L_n(A) \leqslant J(A_{+\varepsilon}, P) + \varepsilon$$

with probability tending to one. Now, let $\varepsilon \to 0$ such that $A_{\pm\varepsilon}$ are Borel sets whose boundaries have mass zero under $J(P)$. Therefore, we may restrict our attention to $U_n(A)$.

Since $E(U_n(A)) = \frac{1}{q} \sum_{a=1}^{q} J_{b,a}(A,P)$, the proof of (i) reduces by Assumption A to showing that $\mathrm{Var}(U_n(A)) \to 0$ as $n \to \infty$. Define

$$I_{b,a} = 1\{\tau_b[\hat{\theta}_{b,a} - \theta(P)] \in A\}, \quad a = 1, \ldots, q,$$

$$s_{q,h} = \frac{1}{q} \sum_{a=1}^{q-h} \mathrm{Cov}(I_{b,a}, I_{b,a+h}).$$

Then

$$\begin{aligned} \mathrm{Var}(U_n(x)) &= \frac{1}{q} \left(s_{q,0} + 2 \sum_{h=1}^{q-1} s_{q,h} \right) \\ &= \frac{1}{q} \left(s_{q,0} + 2 \sum_{h=1}^{b-1} s_{q,h} + 2 \sum_{h=b}^{q-1} s_{q,h} \right) \\ &\equiv D^* + D, \end{aligned}$$

where $D^* = \frac{1}{q}(s_{q,0} + 2\sum_{h=1}^{b-1} s_{q,h})$ and $D = \frac{2}{q}\sum_{h=b}^{q-1} s_{q,h}$. It is readily seen that $|D^*| = O(b/q)$. To handle D we apply a well-known mixing inequality for bounded random variables (e.g., Theorem A.5 in Hall and Heyde, 1980). For $h \geq b$,

$$|\mathrm{Cov}(I_{b,a}, I_{b,a+h})| \leq 4\alpha_X(h - b + 1)$$

and therefore,

$$|D| \leq \frac{8}{q}\sum_{h=1}^{q-b} \alpha_X(h) \to 0.$$

Thus, both D and D^* converge to zero, which completes the proof of (i).

In order to prove (ii) we need the following result.

Lemma B.1 (Billingsley, 1968). *Let* $\{Q_n\}$ *and* Q *be probability measures on* $\mathbb{R}^k$. *Also, let* $\mathcal{U}$ *be a subclass of* $\mathcal{B}^k$ *such that*

(a) $\mathcal{U}$ *is closed under the formation of finite intersections and*

(b) *for every* x *in* $\mathbb{R}^k$ *and every positive* ε *there is an* A *in* $\mathcal{U}$ *with* $x \in A^0 \subseteq A \subseteq B^0(x, \varepsilon)$.

If $Q_n(A) \to Q(A)$ *for every* A *in* $\mathcal{U}$, *then* $Q_n \Rightarrow Q$, *or, equivalently,* $\rho_k(Q_n, Q) \to 0$ *for any metric that metrizes weak convergence on* $\mathbb{R}^k$.

Define a class of sets $\mathcal{U}_{J(P)}$ (the subscript indicates that the class may depend on $J(P)$) in the following way. Let D be a dense countable subset in the set of all points in $\mathbb{R}^k$ that have mass zero under $J(P)$. For each $x \in D$ let E_x be a dense countable subset containing positive real numbers ε for which $\partial(B^0(x, \varepsilon))$ has mass zero under $J(P)$. Now, set $\mathcal{V}_{J(P)} \equiv \bigcup_{x \in D, \varepsilon_x \in E_x} S(x, \varepsilon_x)$. Since $\mathcal{V}_{J(P)}$ is a countable union of countable sets, it is countable itself. Finally, define $\mathcal{U}_{J(P)}$ to contain all the finite intersections of elements of $\mathcal{V}_{J(P)}$. We see immediately that $\mathcal{U}_{J(P)}$ is countable again and meets the conditions (a) and (b) of Lemma B.1 Furthermore, each set $A \in \mathcal{U}_{J(P)}$ has a boundary of mass zero under $J(P)$.

Let $\{n_j\}$ be a subsequence of $\{n\}$. For each $A \in \mathcal{U}_{J(P)}$ we can then find a further subsequence $\{n_{j_l}\}$ such that $L_{n_{j_l}}(A) \to J(A, P)$ almost surely (by the fact that $L_n(A) \to J(A, P)$ in probability). Since $\mathcal{U}_{J(P)}$ is countable, there is a common subsequence $\{n_{j_m}\}$ such that, on a set of probability one, $L_{n_{j_m}}(A) \to J(A, P)$ for all $A \in \mathcal{U}_{J(P)}$. By Lemma B.1 then, $\rho_k(L_{n_{j_m}}, J(P)) \to 0$ almost surely and this shows that $\rho_k(L_n, J(P)) \to 0$ in probability.

The proof of (iii) is obvious once we have (ii). The proof of (iv) is very similar to the proof of Theorem 1 of Beran (1984) given our result (i).

Proof of Lemma A.1. This result is implicitly contained in Theorem 2 of Section 1.4 in Doukhan (1994). For a more detailed derivation see the proof of Lemma 3.1 in Politis et al. (1995).

Proof of Theorem A.1. In the proof of the theorem we will approximate a sum of weakly dependent random variables by a corresponding sum of independent random variables. The following lemma will help us to establish an upper bound of the sup difference of the corresponding characteristic functions. The proof is given in Ibragimov (1962).

Lemma B.2 (Ibragimov, 1962). *Let* $\{Z_t\}$ *denote a sequence of random vectors defined on a probability space, and let* $\mathcal{F}_a^b \equiv \sigma(Z_t; a \leqslant t \leqslant b)$. *Also denote the mixing sequence corresponding to the* Z_t *by* $\alpha_Z(\cdot)$. *Let* Y_1 *and* Y_2 *be random variables measurable with respect to* $\mathcal{F}_{-\infty}^n$ *and* $\mathcal{F}_{n+m}^{\infty}$, *respectively. In addition, let* Y_1' *and* Y_2' *be independent random variables having the same distribution as* Y_1 *and* Y_2, *respectively. Denote the characteristic functions of* $Y_1 + Y_2$ *and* $Y_1' + Y_2'$ *by* φ *and* φ', *respectively.*
Then $\sup_t |\varphi(t) - \varphi'(t)| \leqslant 16\alpha_Z(m)$.

We will now prove the theorem. The main idea of the proof is to split the sum $X_{n,1} + \cdots + X_{n,d_n}$ into alternate blocks of length b_n (the big blocks) and l_n (the small blocks). This is the traditional approach to proving central limit theorems for dependent random variables, and is commonly attributed to Markov or Bernstein (1927) ("Bernstein sums"). Define

$$U_{n,i} = X_{n,(i-1)(b_n+l_n)+1} + \cdots + X_{n,(i-1)(b_n+l_n)+b_n}, \quad 1 \leqslant i \leqslant r_n,$$

where r_n is the largest integer i for which $(i-1)(b_n+l_n)+b_n < d_n$. Further define

$$V_{n,i} = X_{n,(i-1)(b_n+l_n)+b_n+1} + \cdots + X_{n,i(b_n+l_n)+b_n}, \quad 1 \leqslant i \leqslant r_n,$$

$$V_{n,r_n} = X_{n,(r_n-1)(b_n+l_n)+b_n+1} + \cdots + X_{n,d_n}.$$

Then $S_{n,d_n,1} = \sum_{i=1}^{r_n} U_{n,i} + \sum_{i=1}^{r_n} V_{n,i}$, and the technique will be to choose the l_n small enough that $\sum_{i=1}^{r_n} V_{n,i}$ is small in comparison with $\sum_{i=1}^{r_n} U_{n,i}$ but large enough to ensure that the $U_{n,i}$ are nearly independent.

Let $b_n = \lfloor d_n^{3/4} \rfloor$ and $l_n = \lfloor d_n^{1/4} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part of a real number. Since r_n is the largest integer i such that $(i-1)(b_n + l_n) + b_n < d_n$,

$$b_n \sim d_n^{3/4}, \qquad l_n \sim d_n^{1/4}, \qquad r_n \sim d_n^{1/4}. \tag{B.1}$$

We will now proceed to show that $d_n^{-1/2} \sum_{i=1}^{r_n} V_{n,i}$ converges to zero in probability, as d_n tends to infinity. Since its expected value equals zero, it suffices to check that its variance tends to zero. First note that by Lemma A.1 and assumption (A.9), for all n, i

$$\mathrm{E}|V_{n,i}|^2 \leqslant 10K\Delta^2 l_n \equiv Bl_n. \tag{B.2}$$

Therefore,

$$\begin{aligned}
\mathrm{Var}^{1/2}\left(d_n^{-1/2}\sum_{i=1}^{r_n} V_{n,i}\right) &= \mathrm{E}^{1/2}\left(d_n^{-1/2}\sum_{i=1}^{r_n} V_{n,i}\right)^2 \\
&\leqslant \sum_{i=1}^{r_n} \mathrm{E}^{1/2}(d_n^{-1/2}V_{n,i})^2 \\
&\leqslant \left[\sum_{i=1}^{r_n-1}(Bl_n/d_n)^{1/2}\right]+[B(l_n+b_n)/d_n]^{1/2} \quad \text{(by (B.2))} \\
&\leqslant B^{1/2}\{r_n(l_n/d_n)^{1/2} + [(l_n+b_n)/d_n]^{1/2}\} \\
&\leqslant \mathrm{O}(d_n^{-1/8}) \quad \text{(by (B.1))} \\
&\to 0.
\end{aligned}$$

By Slutzky's Theorem it remains to prove that $d_n^{-1/2}\sum_{i=1}^{r_n} U_{n,i} \Rightarrow \mathrm{N}(0,\sigma^2)$.

Let $U'_{n,i}$, $1 \leqslant i \leqslant r_n$, be independent random variables having the same distributions as $U_{n,i}$, $1 \leqslant i \leqslant r_n$. By Lemma B.2 applied inductively, the characteristic functions of $d_n^{-1/2}\sum_{i=1}^{r_n} U_{n,i}$ and of $d_n^{-1/2}\sum_{i=1}^{r_n} U'_{n,i}$ differ by at most $16r_n\alpha_n(l_n)$. Note that by assumption (A.9) we may assume, without loss of generality, that $\alpha_n(k) \leqslant K/k^2$. Therefore,

$$\begin{aligned}
16r_n\alpha_n(l_n) &\leqslant 16r_nK/l_n^2 \\
&\leqslant \mathrm{O}(d_n^{-1/4}) \quad \text{(by (B.1))} \\
&\to 0.
\end{aligned}$$

Thus, the proof will be completed by showing that $d_n^{-1/2}\sum_{i=1}^{r_n} U'_{ni.} \Rightarrow \mathrm{N}(0,\sigma^2)$. This will be accomplished in two steps:

Step 1: $[\mathrm{Var}(\sum_{i=1}^{r_n} U'_{n,i})]^{2/(2+\delta)}\sum_{i=1}^{r_n}\mathrm{E}|U'_{n,i}|^{2+\delta} \to 0$.

Step 2: $\frac{1}{d_n}\mathrm{Var}(\sum_{i=1}^{r_n} U'_{n,i}) \to \sigma^2$.

The result then follows by Lyapounov's Central Limit Theorem and Slutsky's Theorem.

Proof of Step 1:

$$\begin{aligned}
\frac{1}{r_nb_n}\mathrm{Var}\left(\sum_{i=1}^{r_n} U'_{n,i}\right) &= \frac{1}{r_nb_n}\sum_{i=1}^{r_n}\mathrm{E}(U'_{n,i})^2 \\
&= \frac{1}{r_nb_n}\sum_{i=1}^{r_n}\mathrm{E}(U_{n,i})^2
\end{aligned}$$

$$= \frac{1}{r_n} \sum_{i=1}^{r_n} \mathrm{E}(b_n^{-1/2} U_{n,i})^2$$

$$\to \sigma^2 \quad \text{(by assumption (A.8))}. \tag{B.3}$$

Let us assume, without loss of generality, that $1 \leqslant K$. With Corollary A.1 then

$$\mathrm{E}|b_n^{-1/2} U'_{n,i}|^{2+\delta} \leqslant [3024 \,\mathrm{Max}\,\{1, C_n^2(4)\}] 2^{4(4(2-\delta)/\delta+1)} \Delta^{(2+\delta)(1+\delta/2)}$$

$$\leqslant [3024 K^2] 2^{4(4(2-\delta)/\delta+1)} \Delta^{(2+\delta)(1+\delta/2)}$$

$$\text{(by assumption (A.9))}$$

$$\equiv \Delta. \tag{B.4}$$

Finally,

$$\left[\mathrm{Var}\left(\sum_{i=1}^{r_n} U'_{n,i}\right)\right]^{2/(2+\delta)} \sum_{i=1}^{r_n} \mathrm{E}|U'_{n,i}|^{2+\delta}$$

$$= (1/r_n)^{(2+\delta)/2} \left[\frac{1}{r_n b_n} \mathrm{Var}\left(\sum_{i=1}^{r_n} U'_{n,i}\right)\right]^{2/(2+\delta)} \sum_{i=1}^{r_n} \mathrm{E}|b_n^{-1/2} U'_{n,i}|^{2+\delta}$$

$$\leqslant (1/r_n)^{(2+\delta)/2} \left[\frac{1}{r_n b_n} \mathrm{Var}\left(\sum_{i=1}^{r_n} U'_{n,i}\right)\right]^{2/(2+\delta)} r_n \Delta \quad \text{(by (B.4))}$$

$$\leqslant (1/r_n)^{(2+\delta)/2} \mathrm{O}(1)\,\mathrm{O}(r_n) \quad \text{(by (B.3))}$$

$$\leqslant \mathrm{O}((r_n)^{-\delta/2})$$

$$\to 0.$$

Proof of Step 2: We just showed that $(1/r_n b_n)\mathrm{Var}(\sum_{i=1}^{r_n} U'_{n,i}) \to \sigma^2$. But, $r_n b_n / d_n \to 1$ by (B.1).

Proof of Theorem 3.1. For the proof we will assume, without loss of generality, that true mean is equal to zero, i.e. $\theta = 0$. By Theorem 2.1 it is sufficient to verify our Assumption B, as the mixing condition $\alpha_X(k) \to 0$ is implied by assumption (9).

Condition (i) of Assumption B follows immediately from Theorem A.1. For condition (ii) let $\{i_b\}$ be an index sequence. Now apply Theorem A.1 to the triangular array, the nth row of which is the block of size b of the consecutive data $\{X_{i_b}, \ldots, X_{i_b+b-1}\}$ (recall that we consider b as a function of n).

Proof of Theorem 3.2. Again, we will assume without loss of generality that $\theta = 0$. To refresh the memory of the reader we will state a result which we are going to use in the proof. Its proof follows by application of Lindeberg's CLT.

Proposition B.1. Let F and $\{F_n\}$ be distribution functions with corresponding means and variances μ, $\{\mu_n\}$, σ^2 and $\{\sigma_n^2\}$, respectively. Assume

(1) $F_n \Rightarrow F$ *and*
(2) $\sigma_n^2 \to \sigma^2$.

Let $\{X_{n,i},\ 1 \leqslant i \leqslant d_n\}$ be a triangular array of random variables, the nth row of which satisfies: $X_{n,1}, \ldots, X_{n,d_n}$ i.i.d. $\sim F_n$ and define $\overline{X}_n = \frac{1}{d_n}\sum_{i=1}^{d_n} X_{n,i}$.
Then $d_n^{1/2}(\overline{X}_n - \mu_n) \Rightarrow \mathrm{N}(0, \sigma^2)$ as $d_n \to \infty$.

We will now prove the theorem. Note that

$$(bl)^{1/2}(\overline{X}^*_{bl,1} - \overline{X}_{n,1}) = l^{1/2}\left(\frac{1}{l}\sum_{i=1}^{l} b^{1/2}\{\overline{X^*}_{b,i} - \overline{X}_{n,1}\}\right) \equiv l^{1/2}\left(\frac{1}{l}\sum_{i=1}^{l} Z^*_{n,i}\right).$$

We see that the $Z^*_{n,i}$ are i.i.d. $\sim L_n$, where L_n is defined as in (6) from Example 3.1. Here, we have shown (under even weaker assumptions) that $L_n \Rightarrow \mathrm{N}(0, \sigma^2)$ in probability. The next step will be to check that $\mathrm{Var}(L_n) \to \sigma^2$ in probability also. An application of Proposition B.1 will then complete the proof.

The value of $\overline{X}_{n,1}$ has no effect on the variance of L_n. Therefore, by our definition $T_{b,a} \equiv b^{1/2}\overline{X}_{b,a}$,

$$\mathrm{Var}(L_n) = \frac{1}{q}\sum_{a=1}^{q} T_{b,a}^2 - \overline{T}_b^{\,2}, \quad \text{where } \overline{T}_b \equiv \frac{1}{q}\sum_{a=1}^{q} T_{b,a}.$$

Step 1: $\frac{1}{q}\sum_{a=1}^{q} T_{b,a}^2 \to \sigma^2$ in probability.

Proof. $\mathrm{E}[\frac{1}{q}\sum_{a=1}^{q} T_{b,a}^2] = \frac{1}{q}\sum_{a=1}^{q} \sigma_{b,a}^2 \to \sigma^2$ as $n \to \infty$ (by assumption (12).

Hence, it is sufficient to check that $\mathrm{Var}(\frac{1}{q}\sum_{a=1}^{q} T_{b,a}^2)$ tends to zero.
Define

$$s_{q,d} = \frac{1}{q}\sum_{a=1}^{q-d} \mathrm{Cov}(T_{b,a}^2, T_{b,a+d}^2).$$

Then

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{q}\sum_{a=1}^{q} T_{b,a}^2\right) &= \frac{1}{q}\left(s_{q,0} + 2\sum_{d=1}^{q-1} s_{q,d}\right) \\ &= \frac{1}{q}\left(s_{q,0} + 2\sum_{d=1}^{b-1} s_{q,d} + 2\sum_{d=b}^{q-1} s_{q,d}\right) \\ &\equiv A^* + A,\end{aligned}$$

where $A^* = \frac{1}{q}(s_{q,0} + 2\sum_{d=1}^{b-1} s_{q,d})$ and $A = \frac{2}{q}\sum_{d=b}^{q-1} s_{q,d}$.

By a well-known mixing inequality (e.g., Corollary A.2 in Hall and Heyde, 1980),

$$\begin{aligned}\mathrm{Cov}(T_{b,a}^2, T_{b,a+d}^2) &\leqslant 8\|T_{b,a}^2\|_{2+\delta}\,\|T_{b,a+d}^2\|_{2+\delta}\,\alpha_X^{\delta/(2+\delta)}(\mathrm{Max}\,\{0, d-b\}) \\ &\leqslant 8[\Delta^2 K_*]^2\,\alpha_X^{\delta/(2+\delta)}(\mathrm{Max}\,\{0, d-b\}),\end{aligned}$$

where K_* is a constant depending only upon K. This last inequality can be seen as follows

$$\begin{aligned}\|T_{b,a}^2\|_{2+\delta} &= \frac{1}{b}\left(\mathrm{E}\left|\sum_{i=a}^{a+b-1} X_i\right|^{4+2\delta}\right)^{1/(2+\delta)} \\ &\leqslant \frac{1}{b}(b^{2+\delta}\Delta^{4+2\delta}K_*)^{1/(2+\delta)} \\ &\qquad \text{(by Theorem A.1 and assumption (13))} \\ &\leqslant \Delta^2 K_*.\end{aligned}$$

Trivially now,

$$s_{q,d} = \frac{1}{q}\sum_{a=1}^{q-d} \mathrm{Cov}(T_{b,a}^2, T_{b,a+d}^2) \leqslant 8\Delta^4 K_*^2 \alpha_X^{\delta/(2+\delta)}(\mathrm{Max}\,\{0, d-b\}).$$

Hence,

$$A^* \leqslant 16\Delta^4 K_*^2 \frac{b}{q} \to 0,$$

and

$$A \leqslant 16\Delta^4 K_*^2 \frac{b}{q}\sum_{d=1}^{q-b} \alpha_X(d)^{\delta/(2+\delta)} \to 0 \quad \text{(by assumption (13))}.$$

Step 2: $\overline{T}_b \to 0$ in probability.

Proof. Analogous to the proof of Step 1.

From the proof of Theorem 3.1 we already know that $L_n \Rightarrow \mathrm{N}(0, \sigma^2)$. Thus, both $\mathrm{Var}(L_n) \to \sigma^2$ in probability and $\rho_1(L_n, \mathrm{N}(0, \sigma^2)) \to 0$ in probability, where ρ_1 is any metric that metrizes weak convergence on $\mathbb{R}$. Hence, for any subsequence n_l, one can extract a further subsequence n_{l_j} such that both $\rho_1(L_{n_{l_j}}, \mathrm{N}(0, \sigma^2)) \to 0$ and $\mathrm{Var}(L_{n_{l_j}}) \to \sigma^2$ almost surely. By Proposition B.1 it then follows that, on a set of probability one, $\rho_1(L_{n_{l_j}}^*, \mathrm{N}(0, \sigma^2)) \to 0$. This proves that $\rho_1(L_n^*, \mathrm{N}(0, \sigma^2))$ in probability.

Proof of Theorem 3.3. For the proof we will assume, without loss of generality, that the true mean is equal to zero, i.e. $\theta = 0$. The mixing condition in Theorem 2.1 is clearly implied by our assumption (18). Therefore, it will be sufficient to show that assumption B.1 is satisfied.

To prove part (ii) of assumption B.1, let a_b be an index sequence that may depend on b. We will accomplish the proof in two steps.

Step 1: $T_{b,a_b} \Rightarrow \mathrm{N}(0, \Sigma)$.

Proof. By the Cramér–Wold device it will suffice to look at linear combinations of the type $\lambda^{\mathrm{T}}\Sigma^{-1/2}b^{-1/2}X_{b,a_b}$, where λ is a vector in $\mathbb{R}^p$. Without loss of generality we may assume that λ has unit length. We observe that

- $\mathrm{E}(\lambda^{\mathrm{T}}\Sigma^{-1/2}X_{a_b+i}) = 0$ for $0 \leqslant i \leqslant b-1$,
- by assumption (16) and Minkowski, $\mathrm{E}|\lambda^{\mathrm{T}}\Sigma^{-1/2}X_{a_b+i}|^{2+2\delta} \leqslant \Delta^*$, where Δ^* is a constant that depends only on Δ and Σ (since $\|\lambda\| = 1$),
- $$\sigma^2_{b,a_b} \equiv \mathrm{Var}\left(\lambda^{\mathrm{T}}\Sigma^{-1/2}b^{-1/2}\sum_{i=0}^{b-1} X_{a_b+i}\right)$$
$$= \lambda^{\mathrm{T}}\Sigma^{-1/2}\mathrm{Cov}(T_{b,a_b})\Sigma^{-1/2}\lambda$$
$$= \lambda^{\mathrm{T}}\Sigma^{-1/2}\Sigma_{b,a_b}\Sigma^{-1/2}\lambda \to 1 \quad \text{by assumption (17).}$$

Apply Theorem A.1 to deduce that $\lambda^{\mathrm{T}}\Sigma^{-1/2}b^{-1/2}\sum_{i=0}^{b-1} X_{a_b+i} \Rightarrow \mathrm{N}(0,1)$ (to do so let $X_{n,i} = \lambda^{\mathrm{T}}\Sigma^{-1/2}X_{a_b+i-1}$ there). Since we did not specify λ, we may follow by the Cramér–Wold device that $\Sigma^{-1/2}T_{b,a_b} \Rightarrow \mathrm{N}(0, I)$ or, equivalently, that $T_{b,a_b} \Rightarrow \mathrm{N}(0, \Sigma)$.

Step 2: $b^{-1/2}(f(T_{b,a_b}) - f(\theta)) \Rightarrow \mathrm{N}(0, J\Sigma J^{\mathrm{T}})$.

Proof. This follows immediately from step 1, condition (19) and the multivariate δ-method.

Part (i) of assumption B.1 now follows trivially by letting $b = n$ and $a_b \equiv 1$ in (ii).

Proof of Theorem 3.4. The mixing condition in Theorem 2.1 is clearly implied by our assumption (28). Therefore, it will be sufficient to show that assumption B.1 is satisfied.

To prove part (ii) of assumption B.1, let a_b be an index sequence that may depend on b. We will accomplish the proof in two steps.

Step 1: $b^{-1/2}X_{b,a_b}^{\mathrm{T}}\varepsilon_{b,a_b} \Rightarrow \mathrm{N}(0, V)$.

Proof. By the Cramér–Wold device it will suffice to look at linear combinations of the type $\lambda^{\mathrm{T}}V^{-1/2}b^{-1/2}X_{b,a_b}^{\mathrm{T}}\varepsilon_{b,a_b}$, where λ is a vector in $\mathbb{R}^p$. Without loss of generality, we may assume that λ has unit length. We observe that

- $\mathrm{E}(\lambda^{\mathrm{T}}V^{-1/2}x_{a_b+i}\varepsilon_{a_b+i}) = 0$ for $0 \leqslant i \leqslant b-1$,
- by assumption (24) and Minkowski, $\mathrm{E}|\lambda^{\mathrm{T}}V^{-1/2}x_{a_b+i}\varepsilon_{a_b+i}|^{2+2\delta} \leqslant \Delta_1^*$, where Δ_1^* is a constant that depends only on Δ_1 and V (since $\|\lambda\| = 1$),
-

$$\sigma_{b,a_b}^2 \equiv \mathrm{Var}\left(\lambda^{\mathrm{T}}V^{-1/2}b^{-1/2}\sum_{i=0}^{b-1} x_{a_b+i}\varepsilon_{a_b+i}\right)$$

$$= \lambda^{\mathrm{T}}V^{-1/2}\mathrm{Cov}\left(b^{-1/2}\sum_{i=0}^{b-1} x_{a_b+i}\varepsilon_{a_b+i}\right)V^{-1/2}\lambda$$

$$= \lambda^{\mathrm{T}}V^{-1/2}V_{b,a_b}V^{-1/2}\lambda \to 1 \quad \text{by assumption (26).}$$

Apply Theorem A.1 to deduce that $\lambda^{\mathrm{T}}V^{-1/2}b^{-1/2}\sum_{i=0}^{b-1} x_{a_b+i}\varepsilon_{a_b+i} \Rightarrow \mathrm{N}(0,1)$ (to do so let $X_{n,i} = \lambda^{\mathrm{T}}V^{-1/2}x_{a_b+i-1}\varepsilon_{a_b+i-1}$ there). Since we did not specify λ, we may follow by the Cramér–Wold device that $V^{-1/2}b^{-1/2}X_{b,a_b}^{\mathrm{T}}\varepsilon_{b,a_b} \Rightarrow \mathrm{N}(0, I)$ or, equivalently, that $b^{-1/2}X_{b,a_b}^{\mathrm{T}}\varepsilon_{b,a_b} \Rightarrow \mathrm{N}(0, V)$.

Step 2: $b^{1/2}(\hat{\beta}_{b,a_b} - \beta) \Rightarrow \mathrm{N}(0, M^{-1}VM^{-1})$.

Proof. $b^{1/2}(\hat{\beta}_{b,a_b} - \beta) = (X_{b,a_b}^{\mathrm{T}}X_{b,a_b}/b)^{-1}b^{-1/2}X_{b,a_b}^{\mathrm{T}}\varepsilon_{b,a_b}$. By the imposed mixing and moment conditions on the sequence $\{x_i\}$ we have $(X_{b,a_b}^{\mathrm{T}}X_{b,a_b}/b) - M_{b,a_b} \to 0$ in probability. In fact, denote the (i,j)th entry of $[(X_{b,a_b}^{\mathrm{T}}X_{b,a_b}/b) - M_{b,a_b}]$ by $(D_b)_{i,j}$. Then

$$(D_b)_{i,j} = \frac{1}{b}\sum_{l=0}^{b-1} x_{a_b+l,i}\,x_{a_b+l,j} - \mathrm{E}(x_{a_b+l,i}\,x_{a_b+l,j}).$$

By definition, $\mathrm{E}((D_b)_{i,j}) = 0$. It will be sufficient to check that $\mathrm{Var}((D_b)_{i,j}) \to 0$. Define

$$s_{b,d} = \frac{1}{b}\sum_{a=a_b}^{a_b+b-(d+1)} \mathrm{Cov}(x_{a,i}\,x_{a,j}, x_{a+d,i}\,x_{a+d,j}).$$

Then

$$\mathrm{Var}((D)_{i,j}) = \frac{1}{b}\left(s_{b,0} + 2\sum_{d=1}^{b-1} s_{b,d}\right).$$

By the mixing inequality for bounded random variables again,

$$\begin{aligned}\mathrm{Cov}(x_{a,i}x_{a,j}, x_{a+d,i}x_{a+d,j}) &\leqslant 8\,\|x_{a,i}x_{a,j}\|_{2+\delta}\,\|x_{a+d,i}x_{a+d,j}\|_{2+\delta}\,\alpha(d)^{\delta/(2+\delta)} \\ &\leqslant 8\Delta_2^{2/(2+\delta)}\alpha(d)^{\delta/(2+\delta)} \quad \text{(by assumption (26))}.\end{aligned}$$

Therefore, $s_{b,d} \leqslant 8\Delta^{\frac{2}{2+\delta}}\alpha(d)^{\frac{\delta}{2+\delta}}$ and that implies

$$\mathrm{Var}((D)_{i,j}) \leqslant \frac{16}{b}\Delta^{2/(2+\delta)}\sum_{d=0}^{b-1}\alpha(d)^{\delta/(2+\delta)} \to 0.$$

Recalling that $M_{b,a_b} \to M$ by assumption (27), we thus have $(X_{b,a_b}^{\mathrm{T}}X_{b,a_b}/b) \to M$ in probability. By step 1, $b^{-1/2}X_{b,a_b}^{\mathrm{T}}\varepsilon_{b,a_b}$ is $O_P(1)$ and therefore

$$[(X_{b,a_b}^{\mathrm{T}}X_{b,a_b}/b)^{-1} - M^{-1}]b^{-1/2}X_{b,a_b}^{\mathrm{T}}\varepsilon_{b,a_b} \to 0 \quad \text{in probability.}$$

But, $M^{-1}b^{-1/2}X_{b,a_b}^{\mathrm{T}}\varepsilon_{b,a_b} \Rightarrow \mathrm{N}(O, M^{-1}VM^{-1})$ by step 1 again and this completes the proof of part (ii) of assumption B.1.

Part (i) of assumption B.1 now follows trivially by letting $b=n$ and $a_b \equiv 1$ in (ii).

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Andrews, D.W.K., Monahan, J.C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60, 953–966.

Babu, G.J., 1986. Bootstrapping statistics with linear combinations of Chi-squares as weak limrm. Sankhya Ser A. 56, 85–93.

Beran, R., 1984. Bootstrap methods in statistics. Jber. d. Dt. Math.-Vereins 86, 14–30.

Bernstein, S., 1927. Sur l'extension du théorème du calcul des probabilités aux sommes de quantités dependantes. Mathematische Annalen 97, 1–59.

Billingsley, P., 1968. Convergence of Probability Measures. John Wiley, New York.

Billingsley, P., 1986. Probability and Measure, 2nd ed. Wiley, New York.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Bose, A., 1988. Edgeworth correction by bootstrap in autoregressions. Annals of Statistics 14, 1709–1722.

Bühlmann, P., 1994. Blockwise bootstrapped empirical process for stationary sequences. Annals of Statistics 22, 995–1012.

Bühlmann, P., Künsch, H.R., 1994. Block length selection in the bootstrap for time series. Tech. Report Nr. 72, Seminar für Statistik, ETH Zürich.

Campbell, J.Y., Lo, A.W., MacKinlay, A.C., 1995. The econometrics of financial markets. Manuscript.

Carlstein, E., 1986. The use of subseries methods for estimating the variance of a general statistic from a stationary time series. Annals of Statistics 14, 1171–1179.

Davydov, Y.A., 1970. The invariance principle for stationary processes. Theory of Probability Application 14, 487–498.

Doukhan, P., 1994. Mixing: properties and examples. Lecture Notes in Statistics, Vol. 85, Springer, New York.

Efron, B., 1979. Bootstrap methods: another look at the jacknife. Annals of Statistics 7, 1–26.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

Fitzenberger, B., 1995. The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. Unpublished discussion paper, Universität Konstanz.

Frank, J., Härdle, W., 1992. On bootstrapping kernel spectral estimates. The Annals of Statistics, 20, 121–145.

Freedman, D.A., 1981. Bootstrapping regression models. Annals of Statistics 9, 1218–1228.

Freedman, D.A., 1984. On bootstrapping two-stage least-squares estimates in stationary linear models. Annals of Statistics 12, 827–842.

Hall, P., 1988. On symmetric bootstrap confidence intervals. Journal Royal Statistical Society B 50 (1) 35–45.

Hall, P., 1992. The Bootstrap and Edgeworth Expansion, Springer, New York.

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press, New York.

Hall, P., Horowitz J.L., Jing, B., 1995. On blocking rules for the bootstrap with dependent data. Biometrika, to appear.

Ibragimov, I.A., 1962. Some limit theorems for stationary processes, Theory of Probabilistic Applications 4, 349–382.

Künsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations. Annals of Statistics 17, 1217–1241.

Lahiri, S.N., 1992. Edgeworth correction by 'Moving block' bootstrap for stationary and nonstationary data. In Exploring the Limits of Bootstrap, pp. 183–214. John Wiley, New York.

Liu, R.Y., 1988. Bootstrap procedures under some non-iid models. Annals of Statistics 16, 1696–1708.

Liu, R.Y., Singh, K., 1992. Moving blocks jackknife and bootstrap capture weak dependence. In: LePage and Billard (Eds.), Exploring the limits of bootstrap Wiley, New York.

Lo, A.W., MacKinley, A.C., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. Review of Financial Studies 1, 41–66.

Loh, W.-Y., 1987. Calibrating confidence coefficients. Journal of American Statistical Association 82, 155–162.

Malliaropulos, D., 1996. Are long-horizon stock returns predictable? A bootstrap analysis. Journal of Business Finance and Accounting 23(1) 93–106.

Oodaira, H., Yoshihara, K., 1972. Functional central limit theorems for strictly stationary processes satisfying the strong mixing condition. Kodai Math. Seminar Report 24, 259–269.

Politis, D., Romano, J., 1992. A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. Annals of Statistics 20, 1985–2007.

Politis, D., Romano, J., 1994a. The stationary bootstrap. Journal of American Statistics Ass. 89, 1303–1313.

Politis, D., Romano, J., 1994b. Large sample confidence regions based on subsamples under minimal assumptions. Annals of Statistics 22, 2031–2050.

Politis, D., Romano, J., Wolf, M., 1995. Subsampling for nonstationary time series. Technical Report No. 483, Department of Statistics, Stanford University.

Rosenblatt, M., 1956. A central limit theorem and a strong mixing condition. Proceedings of National Academy of Science 42, 43–47.

Serfling, R.J., 1968. Contributions to central limit theory for dependent variables. Annals of Mathematics and Statistics 39, 1158–1175.

Singh, K., 1981. On the asymptotic accuracy of Efron's bootstrap. Annals of Statistics 9, 1187–1195.
Uteev, S., 1984. Inequalities and estimates of the convergence rate for the weakly dependent case. Advances in Probabilistics Theory 1985, Novosibirsk.
White, H., 1984. Asymptotic Theory for Econometricians. Academic Press, Orlando.
White, H., Domowitz I., 1984. Nonlinear regression with dependent observations. Econometrica 52, 143–161.
Wolkonoski, V.A., Rozanov, Y.A., 1959. Some limit theorems for random functions, Part I. Theory of Probabilistic Applications 4, 178–197.
Wu, C.F., 1986. Jacknife, bootstrap and other resampling methods in regression analysis. Annals of Statistics 14, 1261–1343.