SUBSAMPLING FOR NONSTATIONARY TIME SERIES

by

Dimitris N. Politis
Purdue University

Joseph P. Romano
Stanford University

and

Michael Wolf
Stanford University

Technical Report #95-41

Department of Statistics
Purdue University

September 1995

*

Subsampling for nonstationary time series

Dimitris N. Politis
Department of Statistics
Purdue University
W. Lafayette, IN 47907

Joseph P. Romano *
Department of Statistics
Stanford University
Stanford, CA 94305

Michael Wolf
Department of Statistics
Stanford University
Stanford, CA 94305

August 1995

Abstract

In this article, a general theory for the construction of confidence intervals or regions in the context of nonstationary 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 nonstationary 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 nonstationary errors.

KEY WORDS: Time series, nonstationarity, subsampling, moving blocks bootstrap.

JEL-CLASSIFICATION: C14, C15.

*Joseph P. Romano, Phone: (415) 723-6326, Fax: (415) 725-8977, E-mail: romano@playfair.stanford.edu.

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. Singh (1981) showed that Efron's bootstrap fails to capture the dependence structure even for the sample mean of m-dependent data. Following this observation, there have been several attempts to modify and extend Efron's idea to dependent data.

Most extensions in the literature so far only apply to the stationary case. They can roughly be divided into resampling and subsampling methods.

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 (1986), Bose (1988)), nonparametric regression and nonparametric kernel spectral estimation (Härdle and Bowman (1988), Franke and Härdle (1987)). 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 nonstationary 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 nonstationariy 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. It turns out that Künsch's bootstrap enjoys some robustness property to nonstationarity, 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. This procedure is based on resampling blocks of random lengths, where the length of each block has a geometric distribution.

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) and Bühlmann (1994)).

Finally, while the moving blocks method is valid for parameters of the m-dimensional distribution of the observations, with m fixed, the subsampling method is valid also for parameters of the whole (infinite-dimensional) joint distribution. An example is the problem of estimating the spectral density function. Alternatively, an extension of the moving blocks method that can handle such cases was proposed in Politis and Romano (1992).

In this paper we present conditions which ensure that the subsampling method is still asymptotically valid for nonstationary observations. The paper is organized as follows. In section 2 the method is described, and the main theorems are presented. In section 3 we give a central limit theorem for nonstationary dependent random variables. In section 4 some applications and examples are discussed. In addition, a result for the validity of the moving blocks method for nonstationary data is stated. We briefly talk about the problem

of choosing of the blocksize in section 5, although no satisfactory answer is given. The paper concludes with two simulation studies in section 6. Proofs of technical results and tables for the simulation studies can be found in the appendix.

2 The General Theorem

In this section we describe the subsampling method, and give sufficient conditions under which it will still work in the nonstationary case.

Suppose $\{X_1, X_2, \ldots\}$ 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 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. To make this condition precise on a mathematical basis we introduce the concept of strong mixing coefficients. The original definition due to Rosenblatt typically applies to stationary sequences, so a modification for arbitrary (possibly nonstationary) sequences is needed:

Given a random sequence $\{Y_i\}$, let $\mathcal{F}_n^m$ be the σ-algebra generated by $\{Y_i, n \leq i \leq m\}$, and define the corresponding mixing sequence by

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

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 $k \to \infty$.

We will first discuss the case of real valued θ. The theory can be generalized considerably to allow for the construction of confidence regions for multivariate parameters or confidence bands for functions.

2.1 The Univariate Case

Let $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ be an estimator of $\theta(P) \in \mathbb{R}$, the parameter of interest. The crux of the method is to compute the corresponding statistic on subsamples to get an approximation to the sampling distribution of $\hat{\theta}_n$. Thus, let $\hat{\theta}_{b,a} = \hat{\theta}_b(X_a, \ldots, X_{a+b-1})$, the estimator of θ based on the subsample $X_a, \ldots, X_{a+b-1}$. In this notation we then have $\hat{\theta}_{n,1} = \hat{\theta}_n$. 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. Also define the corresponding cumulative distribution function:

$$J_{b,a}(x,P) = Prob_P\{\tau_b(\hat{\theta}_{b,a} - \theta(P)) \leq x\}. \tag{1}$$

Essentially, the only assumption that we will need to construct asymptotically valid confidence intervals 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 continuity point x of $J(P)$, $\frac{1}{n-b+1}\sum_{a=1}^{n-b+1} J_{b,a}(x,P) \to J(x,P)$, 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 $\{a_b\}$, $J_{b,a_b}(x) \to J(x)$ for every continuity point x of $J(\cdot, P)$, as $b \to \infty$.

Here, condition (ii) requires that the distribution function of the normalized statistic evaluated over a susample 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 will be clearly satisfied for stationary processes, but also for processes that exhibit asymptotic stationarity. For example, one can consider a Markov chain with an equilibrium distribution.

It is easy to see that Assumption A follows from Assumption B. If condition (ii) in Assumption A did not hold, there would have to exist a subsequence $\{b_k\}$ such that $J_{b_k,a_{b_k}}(x)$ is bounded away from $J(x)$ for some continuity point x of $J(\cdot, P)$. This contradicts condition (ii) in Assumption B.

In order to describe our method, let $Y_{b,j}$ be the block of size b of the consecutive data $\{X_j, \ldots, X_{j+b-1}\}$. Only a very weak assumption on b will be required. Typically, $b/n \to 0$ and $b \to \infty$ as $n \to \infty$. The approximation to $J_{n,1}(x, P)$ we study is defined by

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

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 $J_n(P)$, at least for large n. 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 $m \to \infty$. Let x be a continuity point of $J(\cdot, P)$. Then*

(i) $L_n(x) \to J(x, P)$ in probability.

(ii) If $J(\cdot, P)$ is continuous, then $sup_x|L_n(x) - J(x, P)| \to 0$ in probability.

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

$$\begin{aligned} c_{n,L}(1-\alpha) &= \inf\{x : L_n(x) \geq 1-\alpha\}, \\ c_{n,U}(1-\alpha) &= \sup\{x : L_n(x) \leq 1-\alpha\}. \end{aligned}$$

Correspondingly, define

$$\begin{aligned} c_L(1-\alpha, P) &= \inf\{x : J(x, P) \geq 1-\alpha\}, \\ c_U(1-\alpha, P) &= \sup\{x : J(x, P) \leq 1-\alpha\}. \end{aligned}$$

Let $\{c_n(1-\alpha)\}$ be any sequence of random variables such that

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

In other words, $c_n(1-\alpha)$ serves as a $(1-\alpha)$ quantile of the subsampling distribution $L_n(x)$.

If $J(\cdot, P)$ is continuous at $c_L(1-\alpha, P)$, then

$$Prob_P\{\tau_n[\hat{\theta}_{n,1} - \theta(P)] \leq c_n(1-\alpha)\} \to 1-\alpha \quad as\ n \to \infty.$$

Thus, the asymptotic coverage probability under P of the interval $I_1 = [\hat{\theta}_{n,1} - \tau_n^{-1} c_n(1 - \alpha),\ \infty)$ is the nominal level $1 - \alpha$.

Remark 2.1 In regular cases, $\tau_n = n^{1/2}$, and the assumptions on b simplify to $b/n \to 0$.

Remark 2.2 In the stationary case, condition (ii) of assumptions A or B will be trivially fulfilled, and the theorem reduces to Corollary 3.2 of Politis and Romano (1994b).

The interval I_1 corresponds to a one-sided hybrid percentile interval in the bootstrap literature (e.g. Hall (1992)). A two-sided equal-tailed confidence interval can be obtained by forming the intersection of two one-sided intervals. The two-sided analogue of I_1 is

$$I_2 = [\hat{\theta}_{n,1} - \tau_n^{-1} c_n(1 - \alpha/2),\ \hat{\theta}_{n,1} - \tau_n^{-1} c_n(\alpha/2)].$$

I_2 is called equal-tailed because it has approximately equal probability in each tail:

$$Prob_P\{\theta < \hat{\theta}_{n,1} - \tau_n^{-1} c_n(1 - \alpha/2)\} \doteq Prob_P\{\theta > \hat{\theta}_{n,1} - \tau_n^{-1} c_n(\alpha/2)\} \doteq \alpha/2.$$

As an alternative approach, two-sided symmetric confidence intervals could be constructed. A two-sided symmetric confidence interval is given by $[\hat{\theta}_{n,1} - \hat{c},\ \hat{\theta}_{n,1} + \hat{c}]$, where $\hat{c}$ is chosen so that $Prob_P\{\left|\hat{\theta}_{n,1} - \theta\right| > \hat{c}\} \doteq \alpha$. Hall (1988) showed that symmetric confidence intervals enjoy enhanced coverage and, even in asymmetric circumstances, can be shorter than equal-tailed confidence intervals. To construct two-sided symmetric subsampling intervals in practice we follow the traditional approach and estimate the two-sided distribution function

$$J_{n,1,|\cdot|}(x, P) = Prob_P\{\tau_n \left|\hat{\theta}_{n,1} - \theta(P)\right| \leq x\}. \tag{3}$$

The subsampling approximation to $J_{n,1,|\cdot|}(x, P)$ is defined by

$$L_{n,|\cdot|}(x) = \frac{1}{n - b + 1} \sum_{a=1}^{n-b+1} 1\{\tau_b \left|\hat{\theta}_{b,a} - \hat{\theta}_{n,1}\right| \} \leq x\}. \tag{4}$$

From Theorem 2.1 we can immediately follow the asymptotic validity of two-sided symmetric subsampling intervals.

Corollary 2.1 *Make the same assumptions as in Theorem 2.1 and denote by $J_{|\cdot|}(\cdot, P)$, $c_{L,|\cdot|}$ and $c_{n,|\cdot|}$ the obvious. Let x be a continuity point of $J_{|\cdot|}(\cdot, P)$. Then*

(i) $L_{n,|\cdot|}(x) \to J_{|\cdot|}(x, P)$ in probability.

(ii) If $J_{|\cdot|}(\cdot, P)$ is continuous, then $sup_x|L_{n,|\cdot|}(x) - J_{|\cdot|}(x, P)| \to 0$ in probability.

(iii) If $J_{|\cdot|}(\cdot, P)$ is continuous at $c_{L,|\cdot|}(1-\alpha, P)$, then

$$Prob_P\{\tau_n \left|\hat{\theta}_{n,1} - \theta(P)\right| \leq c_{n,|\cdot|}(1-\alpha)\} \to 1-\alpha \quad as\ n \to \infty.$$

Thus, the asymptotic coverage probability under P of the interval
$I_{SYM} = [\hat{\theta}_{n,1} - \tau_n^{-1} c_{n,|\cdot|}(1-\ \alpha),\ \hat{\theta}_{n,1} + \tau_n^{-1} c_{n,|\cdot|}(1-\ \alpha)]$ is the nominal level $1-\alpha$.

Symmetric confidence intervals, however, are not necessarily a superior choice. Results about good performance are only asymptotic in character and need not hold for all sample sizes. Furthermore, it is not always desirable to constrain a confidence interval to be symmetric. The asymmetry of an equal-tailed confidence interval may contain some valuable information about the location of the true parameter. We will compare the finite sample performance of symmetric and equal-tailed intervals in our simulation studies in chapter 6.

2.2 The multivariate case

It is often desirable to construct confidence regions for multivariate parameters $\theta(P) \in \mathbb{R}^k$ with $k > 1$. For example, this problems arises when we want to draw simultaneous inference on the parameters in a multivariate regression problem. Fortunately, our method can be extended to higher dimensions without much difficulty.

Again, let $\hat{\theta}_{b,a} = \hat{\theta}_b(X_a, \ldots, X_{a+b-1})$ be an estimator of $\theta(P)$ based on the subsample $X_a, \ldots, X_{a+b-1}$. Define $J_{b,a}(P)$ to be the sampling distribution of $\tau_b(\hat{\theta}_{b,a} - \theta(P))$. Rather than working with multivariate distribution functions we will look at indicators of Borel sets. For any Borel set $A \in \mathcal{B}^k$ define:

$$J_{b,a}(A, P) = Prob_P\{\tau_b(\hat{\theta}_{b,a} - \theta(P)) \in A\}. \tag{5}$$

It is then obvious how we need to modify our Assumptions A and B.

Assumption A.1: 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 $b/n \to 0$.

Assumption B.1: 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 $b \to \infty$.

Similarly to the univariate case, Assumption B.1 implies Assumption A.1.

Our method is analogous to the univariate case: Let $Y_{b,j}$ be the block of size b of the consecutive data $\{X_j, \ldots, X_{j+b-1}\}$. 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

$$L_n(A) = \frac{1}{n-b+1} \sum_{a=1}^{n-b+1} 1\{\tau_b(\hat{\theta}_{b,a} - \hat{\theta}_{n,1}) \in A\}. \tag{6}$$

Theorem 2.2 *Assume Assumption A.1 or Assumption B.1, 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:*

$$\begin{aligned} L_{n,\|\cdot\|}(x) &= \frac{1}{n-b+1} \sum_{a=1}^{n-b+1} 1\{\left\|\tau_b(\hat{\theta}_{b,a} - \hat{\theta}_{n,1})\right\| \le x\}, \\ J_{\|\cdot\|}(x,P) &= Prob\{\|Y\| \le x\}. \end{aligned}$$

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

$$\begin{aligned} c_{n,L}(1-\alpha) &= \inf\{x : L_{n,\|\cdot\|}(x) \geq 1-\alpha\}, \\ c_{n,U}(1-\alpha) &= \sup\{x : L_{n,\|\cdot\|}(x) \leq 1-\alpha\}. \end{aligned}$$

Correspondingly, define

$$\begin{aligned} c_L(1-\alpha, P) &= \inf\{x : J_{\|\cdot\|}(x, P) \geq 1-\alpha\}, \\ c_U(1-\alpha, P) &= \sup\{x : J_{\|\cdot\|}(x, P) \leq 1-\alpha\}. \end{aligned}$$

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

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

In other words, $c_n(1-\alpha)$ is a "smallest" point beyond which the subsampling distribution $L_{n,\|\cdot\|}(x)$ assigns probability less than or equal to α.

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

$$Prob_P\{\left\|\tau_n(\hat{\theta}_{n,1} - \theta(P))\right\| \leq c_n(1-\alpha)\} \to 1-\alpha \quad as\ n \to \infty.$$

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

3 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 nonstationary random variables. In various applications it turns out to be the needed method, as demonstrated in section 4.

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 a moment bound for strong mixing nonstationary random variables. Since it is a valuable tool, we state it here as a lemma rather than in the

appendix. The result is implicitly contained in a theorem of Doukhan (1994). However, the form in which the theorem is presented is not convenient for our purposes. Also, we will give more specific bounds for some special cases. A related bound assuming stationarity was given in Yokoyama (1980).

Lemma 3.1 (Moment bound) *Let $\{X_i\}$ be a sequence of mean zero random variables. Denote the corresponding mixing sequence by $\alpha_X(\cdot)$. Define, for $\tau \geq 2$ and $\delta > 0$*

$$C(\tau,\delta) \equiv \sum_{k=0}^{\infty}(k+1)^{\tau-2}\alpha_X^{\frac{\delta}{\tau+\delta}}(k), \tag{7}$$

$$L(\tau,\delta,d) \equiv \sum_{i=1}^{d}\|X_i\|_{\tau+\delta}^{\tau}, \tag{8}$$

$$D(\tau,\delta,d) \equiv Max\{L(\tau,\delta,d),[L(2,\delta,d)]^{\frac{\tau}{2}}\}. \tag{9}$$

Then the following bound holds

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

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*

$$E\left|\sum_{i=1}^{d}X_i\right|^{c} \leq B(c,\delta)D(c,\delta,d), \tag{10}$$

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

$$\begin{aligned} B(1,\delta) &\leq 1, \\ B(2,\delta) &\leq 18Max\{1,C(2,\delta)\}, \\ B(3,\delta) &\leq 102Max\{1,C(3,\delta)\}, \\ B(4,\delta) &\leq 3024Max\{1,C^2(4,\delta)\}. \end{aligned}$$

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

$$B \leq \left[3024Max\{1,C^2(4,\delta)\}\right]2^{4(4(2-\delta)/\delta+1)}. \tag{11}$$

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. It is in a form most useful for our purposes.

Corollary 3.1 (Concise moment bound) *If we assume in addition to the conditions of Lemma 3.1 that*

$$\|X_i\|_{2+2\delta} \leq \Delta \quad \textit{for all } i, \tag{12}$$

then

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

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

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

where $C(4,\delta)$ is defined as in (7).

We now present the central limit theorem for triangular arrays.

Theorem 3.1 *Let $\{X_{n,i}, 1 \leq i \leq d_n\}$ be a triangular array of mean zero random variables. Denote the mixing sequence corresponding to the n-th row by $\alpha_n(\cdot)$. Define*

$$\begin{aligned} S_{n,k,a} &\equiv \sum_{i=a}^{a+k-1} X_{n,i}, \\ T_{n,k,a} &\equiv k^{-\frac{1}{2}} \sum_{i=a}^{a+k-1} X_{n,i}, \\ \sigma^2_{n,k,a} &\equiv Var(T_{n,k,a}). \end{aligned}$$

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

- $\|X_{n,i}\|_{2+2\delta} \leq \Delta \quad \textit{for all } n, i,$ (13)
- $\sigma^2_{n,k,a} \to \sigma^2 > 0 \quad \textit{uniformly in } a \ (*),$ (14)
- $C_n(4) \equiv \sum_{k=0}^{\infty} (k+1)^2 \alpha_n^{\frac{\delta}{4+\delta}}(k) \leq K \quad \textit{for all } n.$ (15)

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

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

Remark 3.1

(a) This result exhibits the familiar trade-off between moment and dependence restrictions, as expressed by conditions (13) and (15). The larger δ, which corresponds to the largest finite moment, is, the lower the minimum mixing rate can be. If all moments exist ($\delta = \infty$), the mixing condition becomes essentially $\alpha_n(k) = o(k^{-3})$. For δ close to zero, on the other hand, the processes must be nearly independent.

(b) Condition (14) expresses a kind of "asymptotic covariance stationarity". The variance of a standardized block sum has to be close to some (positive) limiting value as long as the block size is large enough, independent of row or block index. It turns out that this condition is not very restrictive for our purposes, since we will be using the theorem to verify the uniformity condition (ii) of Assumption A or Assumption A.1.

4 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 4.1 (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 of size b of the consecutive data $\{X_a, \ldots, X_{a+b-1}\}$. 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:

$$J_{b,a}(x,P) = Prob_P\{b^{\frac{1}{2}}(\overline{X}_{b,a} - \theta) \leq x\}. \tag{16}$$

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

$$L_n(x) = \frac{1}{n-b+1}\sum_{a=1}^{n-b+1} 1\{b^{\frac{1}{2}}(\overline{X}_{b,a} - \overline{X}_{n,1}) \leq x\}. \tag{17}$$

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 4.1 *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*

$$\begin{aligned} T_{k,a} &\equiv k^{-\frac{1}{2}}\sum_{i=a}^{a+k-1} X_i, \\ \sigma^2_{k,a} &\equiv Var(T_{k,a}). \end{aligned}$$

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

- $\|X_i\|_{2+2\delta} \leq \Delta$ *for all i,* (18)
- $\sigma^2_{k,a} \to \sigma^2 > 0$ *uniformly in a,* (19)
- $C(4) \equiv \sum_{k=1}^{\infty}(k+1)^2\alpha_X^{\frac{\delta}{4+\delta}}(k) \leq K.$ (20)

Furthermore assume that $b/n \to 0$ and $b \to \infty$ as $n \to \infty$, and let $J(P) = N(0,\sigma^2)$.

Then conclusions (i) - (iii) of Theorem 2.1 will be true.

Remark 4.1 Even for stationary data we need $T_{n,1}$ to converge weakly to some nondegenerate limiting distribution [Politis and Romano (1994b)]. A reasonable condition for that to happen is $\sigma^2_{n,1} \to \sigma^2$. Taking this into account, our condition(19) does not seem prohibitive. In fact, it should be difficult to imagine a reasonable situation where $\sigma^2_{n,1} \to \sigma^2$, but condition (19) is violated.

Example 4.2 (Moving blocks for the mean case) Consider again the situation of Example 4.1. 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 nonstationary 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 $l \equiv l_n \equiv \lfloor \frac{b}{n} \rfloor$. 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 $\frac{1}{q}$ at each of the $Y_{b,a}$. Define a pseudo time series $\{X_1^*, ..., X_{bl}^*\}$ in the following way: Let $Y_{b,1}^*, \ldots, Y_{b,l}^*$ i.i.d. $\sim P_n^*$ and join them together

ERROR: undefined
OFFENDING COMMAND: fi

STACK:

(to)