SUMMARY

THE USE OF SUBSERIES VALUES FOR ESTIMATING THE VARIANCE OF A GENERAL STATISTIC FROM A STATIONARY SEQUENCE

Edward G. Carlstein

University of North Carolina, Chapel Hill

Let $\{Z_i : -\infty < i < +\infty\}$ be a strictly stationary α-mixing sequence and let $t_n = t_n(Z_1, Z_2, \ldots, Z_n)$ be a statistic computed from the observed data $\vec{Z}_n = (Z_1, Z_2, \ldots, Z_n)$. Without specifying the dependence model giving rise to the sequence $\{Z_i\}$, and without specifying the marginal distribution of Z_i, we address the question of variance estimation for t_n. For estimating the variance of t_n from just the available data $\vec{Z}_n$, we propose computing subseries values $t_m(Z_{i+1}, Z_{i+2}, \ldots, Z_{i+m})$, $0 \leq i < i+m \leq n$. These subseries values are used as replicates in order to model the sampling variability of the statistic t. In particular we use adjacent non-overlapping subseries of length m_n, $m_n \to \infty$, $m_n/n \to 0$. Our basic variance estimator is just the usual sample variance computed amongst these subseries values (after appropriate standardization). This estimator is shown to be consistent under mild integrability conditions. A simulation study is conducted, leading to the introduction of overlapping subseries and improved performance of the variance estimator.

Running Heading: Using Subseries Values to Estimate Variance.

AMS 1980 subject classifications. Primary 62G05; Secondary 60G10.

Key Words and Phrases. Variance estimation, subseries values, general statistic, dependence, α-mixing, stationary sequence.

Supported by NSF Grant MCS-8102725.

THE USE OF SUBSERIES VALUES FOR ESTIMATING THE VARIANCE OF A GENERAL STATISTIC FROM A STATIONARY SEQUENCE

by

Edward G. Carlstein

1. Introduction.

Consider a strictly stationary sequence $\{Z_i : -\infty < i < +\infty\}$ from which we observe $\vec{Z}_n = (Z_1, Z_2, \ldots, Z_n)$, $n \geq 1$. A statistic $t_n = t_n(\vec{Z}_n)$ is computed from the observed series. In the absence of assumptions about the underlying dependence model in the sequence (e.g. autoregression), and in the absence of specific distributional assumptions about the Z_i's (e.g. joint normality), we would like to be able to estimate the variance of t_n from the available data $\vec{Z}_n$.

Most variance estimation techniques for general statistics have been aimed at iid samples, making heavy use of exchangeability in their schemes for generating replicates of t. This is true of the theory and intuition behind Tukey's (1958) "jackknife," Efron's (1979) "bootstrap," and Hartigan's (1969) "typical values." Recently, Freedman (1984) and Freedman and Peters (1984) have considered applying the bootstrap to a linear model with autoregressive component, but this still assumes additive iid perturbations. We propose computing the statistic t on subseries

$(Z_{i+1}, Z_{i+2}, \ldots, Z_{i+m})$, $0 \leq i < i+m \leq n$, within the sample $\vec{Z}_n$ as a way of obtaining replicates of t without disturbing the natural ordering in the data. Our basic variance estimator uses adjacent non-overlapping subseries of length m_n s.t. $m_n \to \infty$ and $m_n/n \to 0$ as $n \to \infty$.

Section 3 gives a detailed comparison of the motivating factors behind our variance estimator and those behind the standard variance estimators for iid data. In Section 4 we establish conditions under which our estimator will be consistent in the L_2 sense. Parallel theory is developed in Section 5 involving only $\mathbb{P}$-consistency of the variance estimator. These consistency results are combined with the asymptotic normality results of Carlstein (1984) to obtain asymptotic normality for general statistics from α-mixing sequences with the limiting distribution being free of the nuisance parameter σ^2. Then simulation studies are conducted in order to investigate the finite-sample performance of the variance estimator. The results of these studies (Section 6) give insight regarding the choice of subseries length (m_n); they also suggest a way to use longer overlapping subseries.

2. Notation and Definitions.

Let $\{Z_i(\omega) : -\infty < i < +\infty\}$ be a strictly stationary sequence of real-valued random variables (r.v.) defined on probability space $(\Omega, F, \mathbb{P})$.

Let $F_p^+ (F_q^-$ respectively) be the σ-field generated by

$\{Z_p(\omega),\ Z_{p+1}(\omega),\ldots\}$ ($\{\ldots,\ Z_{q-1}(\omega),\ Z_q(\omega)\}$ respectively).

For $N \geq 1$ denote: $\alpha(N) = \sup\{|\mathbb{P}\{A \cap B\} - \mathbb{P}\{A\}\mathbb{P}\{B\}| : A \in \mathcal{F}_N^+,\ B \in \mathcal{F}_0^-\}$, and define α-mixing to mean $\lim_{N\to\infty} \alpha(N) = 0$. This is a standard mixing condition which guarantees approximate independence between observations that are separated by a great distance (in time) (see Rosenblatt (1956)). α-mixing is known to be satisfied by normal, double-exponential, and Cauchy AR(1) sequences (Gastwirth and Rubin (1975)), as well as by Markov sequences with finite state space (Billingsley (1968), p. 167). In fact, Gastwirth and Rubin (1975) bound the mixing coefficient $\alpha(N)$ by $C\,|\rho|^N$ for the normal and double-exponential AR(1) sequences, and by $C\,N\,|\rho|^N$ for the Cauchy AR(1) sequence (where $-1 < \rho < 1$ is the AR parameter).

Let $t_n(z_1, z_2,\ldots, z_n)$ be a function from $\mathbb{R}^n \to \mathbb{R}^1$, defined for each $n \geq 1$ so that $t_n(Z_1(\omega), Z_2(\omega),\ldots,Z_n(\omega))$ is $\mathcal{F}$-measurable.

We will suppress the argument ω of $Z_i(\cdot)$ from here on.

Denote $\vec{Z}_n^i = (Z_{i+1}, Z_{i+2},\ldots,Z_{i+n})$ and $t_n^i = t_n(\vec{Z}_n^i)$; as a particular case: $\bar{Z}_n^i = \sum_{j=1}^{n} Z_{i+j}/n$.

For $B \geq 0$ denote: ${}_BX = X \cdot \mathbb{I}\{|X| < B\}$ and ${}^BX = X - {}_BX$.

<u>Definition</u>: Random variables $\{X_n\}$ will be said to be eventually uniformly integrable (e.u.i.) iff

$$\exists n_0 \text{ s.t. } \lim_{A\to\infty} \sup_{n \geq n_0} \mathbb{E}\{|{}^AX_n|\} = 0.$$

At times it will be convenient to use the equivalence:

$\{X_n\}$ are e.u.i. iff $\lim_{A\to\infty} \overline{\lim_{n\to\infty}}\ \mathbb{E}\{|{}^{A}X_n|\} = 0$.

3. The Variance Estimator.

It would be useful to have a procedure for estimating the variance σ^2 of a general statistic t_n^0 using only the available data $\vec{Z}_n^0$. In the same spirit as Carlstein (1984), we wish to avoid making assumptions about the specific marginal distribution F of Z_0 and about the dependence model M in $\{Z_i\}$. Moreover, calculation of the theoretical variance of t_n^0 in terms of the parameters of F and M --even if they were specified-- may be intractable. Hence our objective is a non-parametric variance estimator for general statistics from stationary α-mixing sequences.

In the special case where $\{Z_i\}$ is iid, non-parametric error estimation has been addressed within the broader context of sub-sampling and resampling. Hartigan's (1969) "typical values" can be used to obtain confidence intervals in a very general setting, without explicitly estimating σ^2. The "bootstrap" approach (see for example Efron (1982)) may be applied for estimating virtually any characteristic of the distribution of t_n^0 --including its variance. These techniques are based on the idea that by computing the statistic t on subsamples of the data, we can gain insight about the sampling distribution of t_n^0. This is the intuition behind the bootstrap: The empirical c.d.f. F_n from $\vec{Z}_n^0$ is "close" to the true c.d.f. F of Z_0, since $\vec{Z}_n^0$ is a random sample from F. We would like to observe many replications of t_n^0, each based on a new sample $\vec{Z}_n^0$ from F; as the number of such replications becomes large, the empirical distribution of t_n^0 would become close to the true sampling distribution of t_n^0. We are, however, stuck with but one sample $\vec{Z}_n^0$ from F. So instead of drawing many samples $\vec{Z}_n^0$ from F, we draw many "bootstrap" samples $\vec{Z}_n^{0*}$ from F_n (with replacement). Since F_n is close to F, the corresponding replications $t_n^{0*} = t_n(\vec{Z}_n^{0*})$ have an empirical distribution that satisfactorily approximates the true sampling distribution of t_n^0.

When non-trivial dependence is present in $\{Z_i\}$, there is in principle nothing wrong with using the empirical c.d.f. F_n to estimate the common marginal c.d.f. F. For example, ergodic theorems may be used to show that $F_n(t)$ converges to $F(t)$ with probability 1; and Gastwirth and Rubin (1975) have demonstrated that $(F_n(t) - F(t))n^{1/2}$ converges to a Gaussian process when $\{Z_i\}$

is α-mixing with $\alpha(k) = o(k^{-5/2})$. The problem is that the sampling distribution of t_n^0 depends not only on the marginal distribution F, but also on the dependence in $\vec{Z}_n^0$ (i.e. the joint distribution). Replications of t computed on bootstrap samples from F_n -- or even from F-- will not accurately reflect this dependence. The resultant "empirical distribution" of t_n^{0*} will be of little value in approximating the true sampling distribution of t_n^0. In fact, the dependence structure in $\{Z_i\}$ is preserved only by those subsamples of the form $\{\vec{Z}_k^j : 0 \leq j \leq n-k, \ n \geq k \geq 1\}$.

There are several competing considerations in designing a variance estimator based on $\{t_k^j : 0 \leq j \leq n-k, \ n \geq k \geq 1\}$. It is clear that the performance of such an estimator will depend upon how many representative subseries values t_k^j are used, how different the t_k^j's are from each other, and how accurately the t_k^j's model the behavior of t_n^0. For a particular value of k, one would not expect t_k^j and t_k^{j+1} to differ by much --especially in light of the dependence between $\vec{Z}_k^j$ and Z_{j+k+1}. Hence the collection of subseries values $\{t_k^j : 0 \leq j \leq n-k\}$ contains a great deal of redundancy that may not contribute information about t_n^0's sampling variability. The collection $\{t_k^{jk} : 0 \leq j \leq [n/k] - 1\}$, on the other hand, contains only non-overlapping subseries values. If k is growing, each t_k^{jk} will eventually behave as if it were independent of all but two of the other t_k^{jk}'s. Furthermore, if k remained fixed, a subseries value t_k^j would never be able to reflect the dependencies of lag k+1 or greater. These arguments suggest the use of $\{t_{k_n}^{jk_n} : 0 \leq j \leq [n/k_n] - 1\}$, with $k_n \to \infty$ as $n \to \infty$.

Within this framework it seems reasonable to consider $k_n = [\beta n]$ $(0 < \beta < 1)$, since the corresponding $t_{k_n}^{jk_n}$'s are based on subseries of the same order of magnitude as t_n^0 itself. Unfortunately, only about $1/\beta$ disjoint $t_{k_n}^{jk_n}$'s of this form will ever be available as representative subseries values. So an estimator based on such $t_{k_n}^{jk_n}$'s will never stabilize and home in on σ^2, even as $n \to \infty$. (Ironically, the bootstrap and typical-value methods use randomly selected subsets of the possible subsamples, since it is computationally impractical to use all the subsamples available.)

In light of these factors we propose the use of the subseries values $\{t_{m_n}^{jm_n} : 0 \leq j \leq [n/m_n] - 1\}$, where $\{m_n : n \geq 1\}$ are positive integers s.t. $m_n \to \infty$ and $m_n/n \to 0$ as $n \to \infty$. Thus we obtain an increasing number of subseries values (n/m_n), each of which is based on an ever-growing subseries $(\vec{Z}_{m_n}^{jm_n})$; and each $t_{m_n}^{jm_n}$ is becoming increasingly distant (m_n) from all but two of the other $t_{m_n}^{im_n}$'s.

From this point on we will assume the following set-up:

$s_n^i := s_n(\vec{Z}_n^i)$ is a statistic that is wholly computable from the data $\vec{Z}_n^i$, and does not involve any unknown parameters. $t_n^i := (s_n^i - \mathbb{E}\{s_n^0\})n^{\frac{1}{2}}$ is the correct theoretical standardization for s_n^i, in the sense that $\lim_{n\to\infty} \mathbb{E}\{(t_n^0)^2\} =: \sigma^2 \in (0,\infty)$. The proposed estimator for σ^2 is simply:

$$\hat{\sigma}_n^2 := (m_n^2/n) \sum_{i=0}^{[n/m_n]-1} (s_{m_n}^{im_n} - \bar{s}_{m_n})^2 \text{, where } \bar{s}_{m_n} := \sum_{i=0}^{[n/m_n]-1} s_{m_n}^{im_n}/[n/m_n].$$

This is nothing more than the usual sample variance amongst the standardized subseries values $\{m_n^{\frac{1}{2}} s_{m_n}^{jm_n} : 0 \le j \le [n/m_n] - 1\}$. In Section 6 we will investigate the choice of $\{m_n\}$, and we will introduce some modifications (involving longer overlapping subseries) which enhance the performance of $\hat{\sigma}_n^2$.

4. L_2-Consistency.

In this section we work out some theory for subseries values. The first main result is a law of large numbers for these entities. This result is used to obtain consistency of $\hat{\sigma}_n^2$. Finally we arrive at an asymptotic normality result for t_n^0 in which the limiting distribution is free of σ^2.

Let us begin with a useful truncation lemma:

<u>Lemma 1</u>: Let X be F_q^+-measurable and Y be F_p^--measurable, $q > p$. Suppose $\max\{\mathbb{E}\{X^2\}, \mathbb{E}\{Y^2\}\} \leq C < \infty$. Then for any $A > 0$:

$$|\mathbb{C}\{X,Y\}| \leq 4A^2\alpha(q-p) + 3C^{\frac{1}{2}}((\mathbb{E}\{({}^AX)^2\})^{\frac{1}{2}} + (\mathbb{E}\{({}^AY)^2\})^{\frac{1}{2}}) .$$

<u>Proof</u>: Using the representation $X = {}_AX + {}^AX$ we see that:

$$|\mathbb{C}\{X,Y\}| \leq |\mathbb{C}\{{}_AX, {}_AY\}| + |\mathbb{E}\{{}_AX \cdot {}^AY\}| + |\mathbb{E}\{{}^AX \cdot {}_AY\}| + |\mathbb{E}\{{}^AX \cdot {}^AY\}|$$

$$+ |\mathbb{E}\{{}_AX\}\mathbb{E}\{{}^AY\}| + |\mathbb{E}\{{}^AX\}\mathbb{E}\{{}_AY\}| + |\mathbb{E}\{{}^AX\}\mathbb{E}\{{}^AY\}| . \qquad (1)$$

The first term on the right hand side (r.h.s.) of (1) is bounded above by $4A^2\alpha(q-p)$, by Theorem 17.2.1, Ibragimov and Linnik (1971). The required bounds on the other terms follow from the Schwarz inequality. □

Applying this lemma we can establish the following law of large numbers for subseries values from an α-mixing process.

<u>Theorem 2</u>: Let $\{Z_i\}$ be α-mixing and let $f_n(\vec{Z}_n^i) = f_n^i$ be a statistic. Let $\{m_n : n \geq 1\}$ be s.t. $m_n \to \infty$ and $m_n/n \to 0$.

Define $$\bar{f}_{m_n} = \sum_{i=0}^{[n/m_n]-1} f_{m_n}^{im_n}/[n/m_n] .$$

If: $\lim_{n\to\infty} \mathbb{E}\{f_n^0\} = \phi \in \mathbb{R}^1$, and (2.a)

$(f_n^0)^2$ are e.u.i. ; (2.b)

then: $\bar{f}_{m_n} \xrightarrow{L_2} \phi$ as $n \to \infty$. (2.c)

Proof: By (2.a) it suffices to show $\lim_{n\to\infty} V\{\bar{f}_{m_n}\} = 0$.

$$\text{Now } \tfrac{1}{2}V\{\bar{f}_{m_n}\} \leq \sum_{0 \leq i \leq j \leq [n/m_n]-1} |\mathbb{C}\{f_{m_n}^{im_n}, f_{m_n}^{jm_n}\}| [n/m_n]^{-2} =$$

$$= \sum_{k=0}^{[n/m_n]-1} |\mathbb{C}\{f_{m_n}^0, f_{m_n}^{km_n}\}| ([n/m_n] - k)[n/m_n]^{-2}$$

$$\leq (4\mathbb{E}\{(f_{m_n}^0)^2\} + \sum_{k=2}^{[n/m_n]-1} |\mathbb{C}\{f_{m_n}^0, f_{m_n}^{km_n}\}|)[n/m_n]^{-1} .$$

The idea here is that the covariance between non-adjacent $f_{m_n}^{jm_n}$'s is dropping off as the separation (m_n) increases. So, although there are order n/m_n of these terms, their average becomes negligible as $n \to \infty$.

Formally, we note first that (by (2.b)) $\mathbb{E}\{(f_n^0)^2\}$ are bounded uniformly in $n \geq n_0$ by $C < \infty$. Assume now that n is sufficiently large so that $m_n \geq n_0$. Then for each $k \in \{2, 3, \ldots, [n/m_n]-1\}$ we have:

$|\mathbb{C}\{f^0_{m_n}, f^{km_n}_{m_n}\}| \leq 4A^2\alpha(m_n) + 6C^{\frac{1}{2}}(\mathbf{E}\{({}^A f^0_{m_n})^2\})^{\frac{1}{2}}$ for any $A > 0$, by Lemma 1.

Hence: $\frac{1}{2}V\{\bar{f}_{m_n}\} \leq 4[n/m_n]^{-1}C + 4A^2\alpha(m_n) + 6C^{\frac{1}{2}}(\mathbf{E}\{({}^A f^0_{m_n})^2\})^{\frac{1}{2}}$ for any $A > 0$.

Now take $\lim_{A\to\infty} \overline{\lim_{n\to\infty}} (\cdot)$ of this last expression. □

Now we are ready to prove the L_2-consistency of $\hat{\sigma}^2_n$. This result follows in part from Theorem 2, since $\hat{\sigma}^2_n$ is essentially a mean.

<u>Theorem 3</u>: Let $\{Z_i\}$ be α-mixing and let $\{m_n\}$ be s.t. $m_n \to \infty$ and $m_n/n \to 0$. Let s^i_n, t^i_n, σ^2, $\hat{\sigma}^2_n$ be as defined in Section 3.

If: $(t^0_n)^4$ are e.u.i. (3.a)

then: $\hat{\sigma}^2_n \xrightarrow{L_2} \sigma^2$ as $n \to \infty$. (3.b)

<u>Proof</u>: Write $\hat{\sigma}^2_n = \Sigma_n - (\bar{t}_{m_n})^2$, where $\bar{t}_{m_n} = \sum_{i=0}^{[n/m_n]-1} t^{im_n}_{m_n}/[n/m_n]$,

$\Sigma_n = [n/m_n]^{-1} \sum_{i=0}^{[n/m_n]-1} (t^{im_n}_{m_n})^2$. Clearly we only need to show

$\Sigma_n \xrightarrow{L_2} \sigma^2$ and $(\bar{t}_{m_n})^2 \xrightarrow{L_2} 0$. The former follows from Theorem 2.

In order to show $\bar{t}_{m_n} \xrightarrow{L_4} 0$, recall

<u>Lemma 4</u> (Chung (1974), p. 97): Let $r \in (0,\infty)$, and suppose that $\{X_n\}$ are s.t. $\mathbb{E}\{|X_n|^r\} < \infty$ $\forall n \geq n_0$, and $X_n \xrightarrow{\mathbb{P}} X$. Then:

$|X_n|^r$ are u.i. for $n \geq n_0$

iff

$X_n \xrightarrow{L_r} X$.

By (3.a), $\mathbb{E}\{(\bar{t}_{m_n})^4\} < \infty$ $\forall n$ s.t. $m_n \geq n_0$. And applying Theorem 2 we have $\bar{t}_{m_n} \xrightarrow{\mathbb{P}} 0$. Hence by Lemma 4 it will suffice to establish that $(\bar{t}_{m_n})^4$ are e.u.i.

Now $(\bar{t}_{m_n})^2 \leq \Sigma_n$, so that for $A > 0$: $\mathbb{E}\{(\bar{t}_{m_n})^4 \, \mathbb{I}\{(\bar{t}_{m_n})^4 \geq A\}\}$ $\leq \mathbb{E}\{(\Sigma_n)^2 \, \mathbb{I}\{(\Sigma_n)^2 \geq A\}\}$. Therefore we only need to show e.u.i. of $(\Sigma_n)^2$. But by (3.a) again we know that $\mathbb{E}\{(\Sigma_n)^2\} < \infty$ when $m_n \geq n_0$. And $\Sigma_n \to \sigma^2$ in $\mathbb{P}$ and in L_2 by Theorem 2, as discussed above. So Lemma 4 yields the required result. □

Notice that both Theorem 2 and Theorem 3 are logically independent of the question of convergence in distribution. These results give moment and integrability conditions that guarantee L_2-consistency of estimators based on the subseries values from an α-mixing sequence— regardless of whether the t_n^0's (or f_n^0's) are converging in distribution. Furthermore, we have not constrained the mixing coefficient α or the subseries length m_n in any way other than $\alpha(n) \to 0$, $m_n \to \infty$, $m_n/n \to 0$. In practice the L_2-consistency is desirable because it translates into shrinking variance and bias for the estimator.

We can now combine the variance estimation result (Theorem 3) with the distributional results of Carlstein (1984), and obtain:

<u>Theorem 5</u>: Let $\{Z_i\}$ be α-mixing and let s_n^i, $\{m_n\}$, $\hat{\sigma}_n^2$, t_n^i be as in Theorem 3.

If: $\exists\, \sigma^2 \in (0,\infty)$ s.t. (5.a)

$$\lim_{n\to\infty} (N_n/R_n)^{\frac{1}{2}}\, \mathbb{C}\{t_{N_n}^0,\ t_{R_n}^{M_n}\} = \sigma^2$$

whenever $\{N_n\}$, $\{M_n\}$, $\{R_n\}$ are s.t. $N_n \geq M_n + R_n \geq R_n \to \infty$; and

$$\overline{\lim_{n\to\infty}}\ \mathbb{E}\{(t_n^0)^4\} = 3\sigma^4; \tag{5.b}$$

then: $\hat{\sigma}_n^2 \xrightarrow{L_2} \sigma^2$ as $n \to \infty$; and (5.c)

$$(t_{N_n}^0/\hat{\sigma}_n,\ t_{R_n}^{M_n}/\hat{\sigma}_n) \xrightarrow{\mathcal{D}} N_2(0,\ 0,\ 1,\ 1,\ \rho) \text{ as } n \to \infty \tag{5.d}$$

whenever $\{N_n\}$, $\{M_n\}$, $\{R_n\}$ are s.t. $N_n \geq M_n + R_n \geq R_n \to \infty$ and $R_n/N_n \to \rho^2$.

Proof: We will begin by showing that $(t^0_{N_n}/\sigma,\ t^{M_n}_{R_n}/\sigma) \xrightarrow{\mathcal{D}} N_2(0, 0, 1, 1, \rho)$, via Theorem 4 of Carlstein (1984). Since $\mathbb{E}\{t^0_n\} \equiv 0$, it suffices to observe that (5.b) implies that $(t^0_n)^2$ are e.u.i.

Next we want to use Theorem 3 to conclude that (5.c) holds. In light of (5.a) with $M_n \equiv 0$ and $N_n = R_n = n$, it is enough to verify (3.a). But e.u.i. of $(t^0_n)^4$ follows directly from (5.b) together with $t^0_n \xrightarrow{\mathcal{D}} N(0,\sigma^2)$ (established above). □

(Condition (5.b) may of course be replaced by the less specific condition: $(t^0_n)^4$ are e.u.i.)

5. $\mathbb{P}$-Consistency.

In order to get the convergence in distribution (5.d) of Theorem 5, we really need just $\hat{\sigma}^2_n \xrightarrow{\mathbb{P}} \sigma^2$. It is possible to obtain results analogcus to Theorem 2 and Theorem 3, but which only require integrability conditions on the moments being estimated--not on the higher moments --and which yield only convergence in $\mathbb{P}$ for the subseries estimators. The trade-off, however, is that we must explicitly relate the subseries length m_n to the rate of decay in $\alpha(\cdot)$. Specifically, we use $[n/m_n]\ \alpha(m_n) \to 0$. This says essentially that if the dependence is strong (i.e. $\alpha(\cdot)$ decreases slowly), the subseries length should be large relative to n . This is reasonable since under strong dependence we need larger "gaps" separating non-adjacent subseries values if we want them to behave as if they were independent.

Proceeding in this spirit:

Theorem 6: Let $\{Z_i\}$ be α-mixing and let f_n^i, $\{m_n\}$, $\bar{f}_{m_n}$ be as in Theorem 2.

If: $\lim_{n\to\infty} \mathbb{E}\{f_n^0\} = \phi \in \mathbb{R}^1$, and (6.a)

f_n^0 are e.u.i., and (6.b)

$\lim_{n\to\infty} n\, \alpha(m_n)/m_n = 0$; (6.c)

then: $\bar{f}_{m_n} \xrightarrow{\mathbb{P}} \phi$ as $n \to \infty$. (6.d)

Proof: Denote $[n/m_n] = k_n$, $p_n = \sup\{j : j \text{ an even integer} \le k_n - 1\}$, $q_n = \sup\{j : j \text{ an odd integer} \le k_n - 1\}$. We write $\bar{f}_{m_n}$ as $\bar{f}_{m_n} = \bar{f}^1_{m_n} + \bar{f}^2_{m_n}$

where $\bar{f}^1_{m_n} = (f^0_{m_n} + f^{2m_n}_{m_n} + f^{4m_n}_{m_n} + \ldots + f^{p_n m_n}_{m_n})/k_n$, $\bar{f}^2_{m_n} = (f^{m_n}_{m_n} + f^{3m_n}_{m_n} + f^{5m_n}_{m_n} + \ldots + f^{q_n m_n}_{m_n})/k_n$. It will suffice to show that both $\bar{f}^1_{m_n} \xrightarrow{\mathbb{P}} \phi/2$ and $\bar{f}^2_{m_n} \xrightarrow{\mathbb{P}} \phi/2$. We consider $\bar{f}^1_{m_n}$ first.

Define r.v.'s $\{g^{jm_n}_{m_n} : j \in \{0, 2, 4, \ldots, p_n\}, n \ge 1\}$ having the same marginal distributions as $\{f^{jm_n}_{m_n} : j \in \{0, 2, 4, \ldots, p_n\}, n \ge 1\}$, but s.t. $\{g_{m_n}^{jm_n} : j \in \{0, 2, 4, \ldots, p_n\}\}$ are independent for fixed $n \geq 1$. Denote: $\psi_n(s) = \mathbb{E}\{\exp\{is\, \overline{f}^1_{m_n}\}\}$, $\tilde{\psi}_n(s) = \mathbb{E}\{\exp\{is\, \overline{g}_{m_n}\}\}$ $= (\theta_n(s))^{p_n/2+1}$, where $Y_{nj}(s) = \exp\{is\, f_{m_n}^{jm_n}/k_n\}$,

$\overline{g}_{m_n} = (g^0_{m_n} + g^{2m_n}_{m_n} + \ldots + g^{p_n m_n}_{m_n})/k_n$ and $\theta_n(s) = \mathbb{E}\{Y_{n0}(s)\}$.

Now,
$$|\psi_n(s) - \tilde{\psi}_n(s)| \leq |\psi_n(s) - \mathbb{E}\{\prod_{j=0,\,2,\,\ldots,\,p_n-2} Y_{nj}(s)\}\theta_n(s)| +$$
$$+ |\mathbb{E}\{\prod_{j=0,\,2,\,\ldots,\,p_n-2} Y_{nj}(s)\}$$
$$- \mathbb{E}\{\prod_{j=0,\,2,\,\ldots,\,p_n-4} Y_{nj}(s)\}\theta_n(s)| + \ldots$$
$$\ldots + |\mathbb{E}\{\prod_{j=0,\,2} Y_{nj}(s)\} - \theta^2_n(s)| \leq 16\, \alpha(m_n)p_n/2 ,$$

by Ibragimov and Linnik (1971), p. 307, because $|Y_{nj}(s)| = 1$ and $|\theta_n(s)| \leq 1$. Hence, by (6.c) it will suffice to show $\overline{g}_{m_n} \xrightarrow{\mathbb{P}} \phi/2$.

Put $r_n = p_n/2 + 1$; denote $X_{nj} = g_{m_n}^{2(j-1)m_n} - \phi$ for each $j \in \{1,2,\ldots,r_n\}$, $n \geq 1$. We will show $\sum_{j=1}^{r_n} X_{nj}/r_n \xrightarrow{\mathbb{P}} 0$. Note that for fixed n $\{X_{nj} : 1 \leq j \leq r_n\}$ are iid, with $\mathbb{E}\{X_{n1}\} \to 0$ as $n \to \infty$.

Also, $\{X_{n1}\}$ are e.u.i. by (6.b), which in turn implies that

$\lim_{n\to\infty} \mathbb{E}\{|{}^{r_n}X_{n1}|\} = 0$. Now truncating X_{nj} at r_n we obtain

$$r_n^{-1} \sum_{j=1}^{r_n} X_{nj} = r_n^{-1} \sum_{j=1}^{r_n} \left({}_{r_n}X_{nj} - \mathbb{E}\{{}_{r_n}X_{n1}\}\right) + \mathbb{E}\{{}_{r_n}X_{n1}\} +$$

$$+ r_n^{-1} \sum_{j=1}^{r_n} {}^{r_n}X_{nj} .$$

We will show that each of the 3 terms on the r.h.s. converges to zero in $\mathbb{P}$, using an argument similar to Chow and Teicher (1978), pp. 125-126.

$|\mathbb{E}\{{}_{r_n}X_{n1}\}| \le |\mathbb{E}\{X_{n1}\}| + |\mathbb{E}\{{}^{r_n}X_{n1}\}| \to 0$ as $n \to \infty$; also

$\mathbb{P}\{|r_n^{-1} \sum_{j=1}^{r_n} {}^{r_n}X_{nj}| > \varepsilon\} \le \mathbb{P}\{|X_{nj}| \ge r_n \text{ for some } 1 \le j \le r_n\} \le$

$r_n \mathbb{P}\{|X_{n1}| \ge r_n\} \le \mathbb{E}\{|{}^{r_n}X_{n1}|\} \to 0$ as $n \to \infty$; and lastly

$$r_n \mathbb{E}\{(r_n^{-1} \sum_{j=1}^{r_n} ({}_{r_n}X_{nj} - \mathbb{E}\{{}_{r_n}X_{n1}\}))^2\} = V\{{}_{r_n}X_{n1}\} \le \mathbb{E}\{({}_{r_n}X_{n1})^2\} =$$

$$= \sum_{j=0}^{r_n-1} \mathbb{E}\{(X_{n1})^2\, \mathbb{I}\{j \le |X_{n1}| < j+1\}\} \le \sum_{j=0}^{r_n-1} (j+1)^2 (\mathbb{P}\{|X_{n1}| \ge j\}$$

$$- \mathbb{P}\{|X_{n1}| \ge j+1\})$$

$$= \mathbb{P}\{|X_{n1}| \geq 0\} - r_n^2\, \mathbb{P}\{|X_{n1}| \geq r_n\} + \sum_{j=1}^{r_n - 1} (2j+1)\, \mathbb{P}\{|X_{n1}| \geq j\} \leq$$

$$\leq 1 + 3 \sum_{j=1}^{r_n} j\, \mathbb{P}\{|X_{n1}| \geq j\} \leq 3 \sum_{j=1}^{r_n} \mathbb{E}\{|{}^jX_{n1}|\} + 1 \,. \qquad (2)$$

Since $\{X_{n1}\}$ are e.u.i., $\exists\, C < \infty$ s.t.

$$\sum_{j=1}^{[A]} \mathbb{E}\{|{}^jX_{n1}|\} + \sum_{j=[A]+1}^{r_n} \mathbb{E}\{|{}^jX_{n1}|\} \leq A\, C + r_n\, \mathbb{E}\{|{}^AX_{n1}|\}\,, \text{ for any } A > 0\,,$$

for n sufficiently large. Substituting into (2), dividing through by r_n, and taking $\lim_{A\to\infty} \overline{\lim_{n\to\infty}} (\cdot)$ establishes the required convergence in $\mathbb{P}$ of $\overline{f}^1_{m_n}$. An exactly analogous argument may be used on $\overline{f}^2_{m_n}$. □

<u>Corollary 7</u>: Let $\{Z_i\}$ be α-mixing and let s_n^i, t_n^i, σ^2, $\hat{\sigma}_n^2$, $\{m_n\}$ be as in Theorem 3.

If: $(t_n^0)^2$ are e.u.i. , and (7.a)

$\lim_{n\to\infty} n\, \alpha(m_n)/m_n = 0$; (7.b)

then: $\hat{\sigma}_n^2 \xrightarrow{\mathbb{P}} \sigma^2$ as $n \to \infty$. (7.c)

Proof: Write $\hat{\sigma}_n^2 = \Sigma_n - (\bar{t}_{m_n})^2$ as in the proof of Theorem 3.

$\Sigma_n \xrightarrow{\mathbb{P}} \sigma^2$ follows directly from Theorem 6; so does $\bar{t}_{m_n} \xrightarrow{\mathbb{P}} 0$ since $\mathbb{E}\{t_n^0\} \equiv 0$. □

We can finally give a version of Theorem 4 of Carlstein (1984) whose conclusion is free of σ^2, and whose moment conditions are no stronger than those in that earlier result. Of course we pay by assuming more about the relationship between mixing rate and sub-series length.

Corollary 8: Let $\{Z_i\}$ be α-mixing and let s_n^i, $\{m_n\}$, $\hat{\sigma}_n^2$, t_n^i be as in Theorem 3.

If: $\exists\ \sigma^2 \in (0,\infty)$ s.t. (8.a)

$$\lim_{n\to\infty} (N_n/R_n)^{\frac{1}{2}}\, \mathbb{C}\{t_{N_n}^0\ ,\ t_{R_n}^{M_n}\} = \sigma^2$$

whenever $\{N_n\}$, $\{M_n\}$, $\{R_n\}$ are s.t.

$N_n \geq M_n + R_n \geq R_n \to \infty$; and

$(t_n^0)^2$ are e.u.i. ; and (8.b)

$\lim_{n\to\infty} n\ \alpha(m_n)/m_n = 0$; (8.c) then: $\hat{\sigma}_n^2 \xrightarrow{\mathbb{P}} \sigma^2$ as $n \to \infty$; and (8.d)

$$(t_{N_n}^0/\hat{\sigma}_n,\ t_{R_n}^{M_n}/\hat{\sigma}_n) \xrightarrow{\mathcal{D}} N_2(0,\ 0,\ 1,\ 1,\ \rho) \text{ as } n \to \infty \tag{8.e}$$

whenever $\{N_n\}$, $\{M_n\}$, $\{R_n\}$ are s.t.

$N_n \geq M_n + R_n \geq R_n \to \infty$ and $R_n/N_n \to \rho^2$.

<u>Proof</u>: This is an immediate consequence of Theorem 4 of Carlstein (1984) and Corollary 7 (above). □

Notice that if $\alpha(n) \leq C n \beta^n$, $0 < \beta < 1$, as in the normal, double-exponential and Cauchy autoregressive examples of Section 2, then choosing $m_n = n^\gamma$ $(0 < \gamma < 1)$ yields: $n\,\alpha(m_n)/m_n \leq C m_n^{1/\gamma} \beta^{m_n} \to 0$ as $n \to \infty$, as well as $m_n \to \infty$ and $m_n/n \to 0$.

The sample mean and sample fractile statistics are discussed as examples in Corollary 14 and Theorem 17 (respectively) of Carlstein (1984).

6. Simulation Study of $\hat{\sigma}_n^2$.

Section 3 gave intuitive motivation for the general form of $\hat{\sigma}_n^2$, and Sections 4 and 5 established certain reasonable asymptotic properties of this variance estimator. In the present section we consider the finite-sample behavior of $\hat{\sigma}_n^2$ and the choice of $\{m_n\}$, and we suggest some modifications of $\hat{\sigma}_n^2$ that yield superior performance. Here the method of investigation is large-scale simulation rather than theoretical calculation.

At this stage it is helpful to write $\hat{\sigma}_n^2$ as:

$$\hat{\sigma}_n^2 = \sum_{0 \leq i < j \leq [n/m_n] - 1} (s_{m_n}^{im_n} - s_{m_n}^{jm_n})^2 \, m_n / [n/m_n][n/m_n - 1] =$$

$$= \sum_{0 \leq i < j \leq [n/m_n] - 1} (t_{m_n}^{im_n} - t_{m_n}^{jm_n})^2 / [n/m_n][n/m_n - 1] .$$ There are $[n/m_n][n/m_n-1]/2$ squared paired differences, each contributing 2 replicates of $(t_{m_n}^{im_n})^2$ to our estimate of σ^2. As mentioned in Section 3, $\hat{\sigma}_n^2$ will be biased if m_n is not long enough to make $(t_{m_n}^{im_n})^2$ a good "representative." The cross-product terms $t_{m_n}^{im_n} t_{m_n}^{jm_n}$ will add to the bias if $jm_n - (i+1)m_n$ is not large enough to make $t_{m_n}^{im_n}$ and $t_{m_n}^{jm_n}$ approximately independent. And we need a fair number of $(t_{m_n}^{im_n})^2$ replicates if our estimator is to be stable. These considerations led us to define $\hat{\sigma}_n^2$ and $\{m_n\}$ with $m_n \to \infty$ and $n/m_n \to \infty$, and led us to impose α-mixing on the underlying sequence. The theoretical framework we arrived at was tractable, and yielded

encouraging results. But these same considerations also suggest modifications to improve the performance of $\hat{\sigma}_n^2$ for finite n .

For fixed n , we want our subseries to be as long as possible so that $(t_{m_n}^{im_n})^2$ reflects all of the "relevant" dependence in $\{Z_i\}$. We are restrained, however, by the fact that there are not enough non-overlapping long subseries. In practice, then, it is worthwhile to consider allowing the subseries to overlap so that quality need not be sacrificed for quantity. That is, we may use subseries starting at the same intervals $\{im_n : i = 0, 1, 2, \ldots\}$, but lasting for $k_n := \ell m_n$ terms rather than just m_n terms. (Here ℓ is a fixed positive integer.) The number of replicates available ($r_n := [n/m_n] - \ell + 1$) is virtually unchanged, but their approximate independence is undermined. On the other hand, since ℓ is fixed the asymptotic properties of the estimator will still hold: now each subseries is approximately independent of all but 2ℓ other subseries (rather than all but just 2 other subseries). In the finite-sample setting we would expect to reduce the bias of $\hat{\sigma}_n^2$ in so far as $(t_{k_n}^{im_n})^2$ is a better representative than $(t_{m_n}^{im_n})^2$. Yet the magnitude of the cross-product terms $t_{k_n}^{im_n} t_{k_n}^{jm_n}$ will probably be greater than that of $t_{m_n}^{im_n} t_{m_n}^{jm_n}$, especially for j-i small; this could offset the reductions in bias. And furthermore, although the number of $(t_{k_n}^{im_n})^2$ replicates is nearly the same as the number of $(t_{m_n}^{im_n})^2$ replicates, the covariances between the former are likely to be larger than the covariances between the latter. Hence the estimator $\hat{\sigma}_n^2$ based on the $(t_{k_n}^{im_n})^2$'s would have larger variance than the version based on $(t_{m_n}^{im_n})^2$'s. Our simulation study investigates these trade-offs.

Based on the above arguments, it would appear that the generalized variance estimator being proposed is:

$$\hat{\sigma}_n^2 = \sum_{0 \le i < j \le r_n - 1} (s_{k_n}^{im_n} - s_{k_n}^{jm_n})^2 \; k_n / r_n(r_n - 1),$$

which reduces to our old $\hat{\sigma}_n^2$ when $\ell = 1$. But reflecting on the case $s_n^i = \bar{Z}_n^i$, it is clear that the paired differences involving overlapping subseries require special treatment: for $1 \le j - i < \ell$, we have $\bar{Z}_{k_n}^{im_n} - \bar{Z}_{k_n}^{jm_n} = (\sum_{p=im_n+1}^{jm_n} Z_p - \sum_{p=im_n+k_n+1}^{jm_n+k_n} Z_p)/k_n$, which should be standardized by a factor of $k_n/((j-i)m_n)^{\frac{1}{2}}$ if it is to be used to model

$n^{\frac{1}{2}} s_n^0$. This suggests that the appropriate variance estimator (for mean-like statistics) is:

$$\hat{\sigma}_n^2 := \sum_{0 \le i < j \le r_n - 1} (s_{k_n}^{im_n} - s_{k_n}^{jm_n})^2 (\mathbb{I}\{j-i \ge \ell\} + \frac{\ell}{j-i} \mathbb{I}\{j-i < \ell\}) k_n / r_n (r_n - 1) ,$$

which again reduces to the old $\hat{\sigma}_n^2$ when $\ell = 1$.

To test the performance of this $\hat{\sigma}_n^2$, we began with a simple situation where theoretical checks can be made: $\{Z_i\}$ comes from an AR(1) sequence $Z_i = \phi Z_{i-1} + \varepsilon_i$, $|\phi| < 1$, $\varepsilon_i \sim$ iid $N(0,1)$, $s_n^i = \bar{Z}_n^i$. In this situation it is easy to show that $\sigma^2 = (1-\phi)^{-2}$. We considered weak, moderate, and strong positive dependences in $\{Z_i\}$ ($\phi = .1, .5, .9$) ; samples of realistic sizes for time-series analysis ($n = 100, 250, 500, 1000$) ; short, medium, and long "base-lengths" for the subseries ($m_n = [\ell n\ n]$, $[n^{\frac{1}{2}}]$, $[n^{3/4}]$) ; and subseries overlaps of 2/3, 1/2, and none (corresponding to $\ell = 3, 2, 1$) . For each combination of (ϕ, n, m_n, ℓ) , 1000 realizations of $\tilde{Z}_n^0$ (and hence $\hat{\sigma}_n^2$) were generated. The routine generating the ε_i's was adapted from the uniform random number generator of Wichmann and Hill (1982) and the inverse-normal approximation of Beasley and Springer (1977).

The criteria used to evaluate $\hat{\sigma}_n^2$ are: $\mathbb{E}\{\hat{\sigma}_n^2\}$, $V\{\hat{\sigma}_n^2\}$, and $MSE\{\hat{\sigma}_n^2\} := V\{\hat{\sigma}_n^2\} + (\mathbb{E}\{\hat{\sigma}_n^2\} - \sigma^2)^2$ (each of these being estimated from the 1000 realizations of $\hat{\sigma}_n^2$); the standard deviation of $\mathbb{E}\{\hat{\sigma}_n^2\}$ is estimated by $(V\{\hat{\sigma}_n^2\}/1000)^{\frac{1}{2}}$. Because the true values of σ^2 vary so dramatically ($\sigma^2 = 1.23$ for $\phi = .1$, $\sigma^2 = 100$ for $\phi = .9$), it aids comparisons across ϕ to consider $\mathrm{MSE}\{\hat{\sigma}_n^2\}/\sigma^4$ and also to consider $V\{\hat{\sigma}_n^2\}$ and $(\mathbb{E}\{\hat{\sigma}_n^2\} - \sigma^2)^2$ as proportions of $\mathrm{MSE}\{\hat{\sigma}_n^2\}$. The results are presented in Table 1.

As suggested by the theoretical results of Section 4, $\hat{\sigma}_n^2$ is converging to σ^2 in m.s.e. as n increases--for all values of ϕ and all choices of $\{m_n\}$ and ℓ.

Under weak dependence ($\phi = .1$), virtually all of the m.s.e. is due to variance because even the short subseries are long enough to represent the relevant dependence. Comparing across $m_n = \ell n\, n, n^{\frac{1}{2}}, n^{3/4}$ for fixed values of n and ℓ, we see the smallest variance for $m_n = \ell n\ n$ and the largest variance for $m_n = n^{3/4}$. This is due to the large number of subseries values (r_n) available when m_n is short (ℓn n), and the scarcity of subseries values when m_n is long ($n^{3/4}$). On the other hand, for fixed n and m_n we see a substantial increase in variance as ℓ increases. This cannot be attributed to the corresponding but relatively minor decrease in r_n (except perhaps in the case $m_n = n^{3/4}$). Rather, this effect must be from the larger covariances between the longer overlapping subseries values $(t_{k_n}^{im_n})^2$. Since variance is the name

TABLE 1. Simulation Study of $\hat{\sigma}_n^2$. $\{Z_i\}$ is an AR(1) sequence with coefficient ϕ. $\sigma^2 = \lim_{n\to\infty} (n\, V\{s_n^0\})$, $s_n^i = \bar{Z}_n^i$.

Based on 1000 realizations. n = sample size, m_n = "base" subseries length, ℓ = overlap factor, $k_n = \ell m_n$ = actual subseries length, r_n = # of subseries per sample.
*Best (or approximate best) in column for fixed n and ϕ. $\mathbb{E}$, V, MSE are simulated estimates of $\mathbb{E}\{\hat{\sigma}_n^2\}$, $V\{\hat{\sigma}_n^2\}$, $MSE\{\hat{\sigma}_n^2\}$ respectively.

						$\phi = .1$, $\sigma^2 = 1.23$					$\phi = .5$, $\sigma^2 = 4.0$					$\phi = .9$, $\sigma^2 = 100$				
	n	m_n	ℓ	k_n	r_n	$\mathbb{E}$	sd{$\mathbb{E}$}	V	V/MSE	MSE/σ^4	$\mathbb{E}$	sd{$\mathbb{E}$}	V	V/MSE	MSE/σ^4	$\mathbb{E}$	sd{$\mathbb{E}$}	V	V/MSE	MSE/σ^4
$m_n = \ell n\, n$	100	4	1	4	25	1.19	.011	.120*	.98	.080*	2.71	.026	.667*	.29	.146	15.3	.23	51.5*	.01	.713
	250	5	1	5	50	1.18	.008	.061*	.95	.042*	2.96	.019	.363*	.25	.090	20.7	.21	43.1*	.01	.634
	500	6	1	6	83	1.19	.006	.032*	.94	.023*	3.11	.015	.239*	.23	.065	25.2	.19	35.1*	.01	.564
	1000	6	1	6	166	1.19	.004	.016*	.89	.012*	3.11	.011	.127*	.14	.058	25.3	.13	16.5*	.00	.559
	100	4	2	8	24	1.21*	.014	.192	.99	.126	3.29	.040	1.57	.76	.129*	26.4	.45	200	.04	.562
	250	5	2	10	49	1.21*	.010	.090	.99	.060	3.42	.028	.81	.71	.071*	35.2	.41	169	.04	.437
	500	6	2	12	82	1.22*	.007	.055	1.00	.036	3.52	.023	.51	.69	.046*	41.8	.35	124	.04	.351
	1000	6	2	12	165	1.22*	.005	.027	1.00	.018	3.56	.016	.26	.57	.029*	42.6	.25	63	.02	.336
	100	4	3	12	23	1.18	.015	.237	.98	.158	3.38	.049	2.43	.86	.176	34.7	.65	423	.09	.469
	250	5	3	15	48	1.23*	.012	.137	1.00	.090	3.60	.035	1.20	.88	.085	46.4	.64	409	.12	.328
	500	6	3	18	81	1.21*	.009	.079	.99	.053	3.72	.028	.78	.91	.054	53.5	.52	275	.11	.244
	1000	6	3	18	164	1.22*	.006	.038	1.00	.025	3.66	.019	.37	.76	.031*	54.3	.34	118	.05	.220
$m_n = n^{1/2}$	100	10	1	10	10	1.23*	.018	.310	1.00	.203	3.42	.052	2.67	.89	.188	32.7	.61	367	.07	.489
	250	15	1	15	16	1.22*	.014	.189	1.00	.124	3.58	.040	1.61	.90	.112	47.2	.60	361	.11	.315
	500	22	1	22	22	1.24*	.012	.145	1.00	.095	3.73	.038	1.44	.95	.094	58.6	.62	381	.18	.209
	1000	31	1	31	32	1.23*	.010	.101	1.00	.066	3.78	.030	.93	.95	.061	69.8	.59	347	.28	.126
	100	10	2	20	9	1.18	.022	.489	.99	.323	3.58	.065	4.19	.96	.273	47.9	1.1	1157	.30	.388*
	250	15	2	30	15	1.24*	.018	.307	1.00	.202	3.85*	.056	3.14	.99	.198	61.5	1.0	1008	.40	.249*
	500	22	2	44	21	1.26*	.015	.232	.99	.153	3.91*	.047	2.17	.99	.136	76.7	1.0	1082	.67	.163*
	1000	31	2	62	31	1.22*	.012	.138	1.00	.091	3.93*	.039	1.53	1.00	.096	81.9	.9	796	.71	.112*
	100	10	3	30	8	1.26*	.027	.745	1.00	.489	3.77*	.084	6.98	.99	.440	55.1	1.5	2224	.52	.424
	250	15	3	45	14	1.23*	.022	.492	1.00	.323	3.76	.064	4.06	.99	.257	73.9	1.4	2049	.75	.273
	500	22	3	66	20	1.22*	.018	.314	1.00	.206	3.89*	.060	3.63	.99	.228	80.8	1.3	1687	.82	.206
	1000	31	3	93	30	1.24*	.015	.218	1.00	.143	4.01*	.049	2.42	1.00	.151	89.6	1.1	1218	.92	.133
$m_n = n^{3/4}$	100	31	1	31	3	1.19*	.037	1.39	1.00	.914	3.67*	.113	12.69	.99	.800	60.0*	1.9	3598	.69	.520
	250	62	1	62	4	1.24*	.031	.94	1.00	.617	3.83*	.097	9.35	1.00	.586	85.3*	2.2	4770	.96	.499
	500	105	1	105	4	1.18	.028	.79	1.00	.519	3.73	.098	9.56	.99	.602	89.0*	2.2	4976	.98	.510
	1000	177	1	177	5	1.23*	.028	.31	1.00	.532	3.85	.084	7.09	1.00	.444	91.9	2.0	4128	.98	.419
	100	31	2	62	2	1.13	.054	2.95	1.00	1.94	3.69*	.184	33.8	1.00	2.12	64.8*	3.1	9688	.89	1.09
	250	62	2	124	3	1.20*	.038	1.42	1.00	.93	4.20*	.133	17.6	1.00	1.10	89.9*	3.0	9082	.99	.92
	500	105	2	210	3	1.28*	.046	2.10	1.00	1.38	3.97*	.124	15.3	1.00	.96	94.4*	3.3	10721	1.00	1.08
	1000	177	2	354	4	1.25*	.034	1.14	1.00	.75	4.07*	.120	14.4	1.00	.90	90.0	2.5	6380	.98	.65
	100	31	3	93	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	250	62	3	186	2	1.31	.061	3.73	1.00	2.45	4.16*	.185	34.4	1.00	2.15	86.9*	4.0	16005	.99	1.62
	500	105	3	315	2	1.17*	.055	3.03	1.00	1.99	3.82*	.179	32.0	1.00	2.00	84.0	3.9	14889	.98	1.51
	1000	177	3	531	3	1.20*	.040	1.64	1.00	1.07	3.96*	.129	16.7	1.00	1.05	98.1*	3.3	10548	1.00	1.06

of the game here, the introduction of overlap doesn't pay off in a big way. But it is worth noting that in the cases where there was some significant bias (i.e. $m_n = \ell n\ n$, $\ell = 1$) , doubling the subseries length ($\ell = 2$) does eliminate it. In practice one is faced with a fixed n , a fixed (but unknown) ϕ , and a choice of estimators ($m_n = \ell n\ n$, $n^{1/2}$, $n^{3/4}$; $\ell = 1, 2, 3$) . So the "bottom line" of this analysis is to identify which of the 9 estimators is best for each criteria (bias, variance, m.s.e.), given n and ϕ . In Table 1 , an asterisk (*) indicates the best (or approximate best in the case of close races) estimator. Clearly, $m_n = \ell n\ n$, $\ell = 1$ is the big winner for all sample sizes when $\phi = .1$.

Moving on to the case of moderate dependence ($\phi = .5$) , we begin to see the biasing effect of insufficient subseries length. In the case of $m_n = \ell n\ n$, $\ell = 1$, where the bias is most substantial, there are pronounced gains for doubling and tripling the subseries length ($\ell = 2, 3$) . When $m_n = n^{1/2}$ and $\ell = 1$, there are again improvements for doubling, but less decidedly so for tripling. When $m_n = n^{3/4}$, the subseries are already so long that increasing ℓ is of little value. The pattern of variances is parallel to the $\phi = .1$ case: for fixed n , variance increases in response to fewer replications and increased overlap. Although this makes $m_n = \ell n\ n$, $\ell = 1$ the best choice for minimizing variance, the bias contribution is of enough consequence to push

$m_n = \ell n\ n$, $\ell = 2$ ahead in m.s.e. Note that $m_n = \ell n\ n$, $\ell = 3$ also beats $m_n = \ell n\ n$, $\ell = 1$ for $n \geq 250$ (in terms of m.s.e.). The variances for $m_n = n^{3/4}$ are so large--due to the excruciatingly small sizes of r_n --that it seems unwise to use such estimators, in spite of their relatively good performance in terms of bias. For $m_n = n^{\frac{1}{2}}$, $\ell = 2$ and 3, the biases are nearly as good as for $m_n = n^{3/4}$, but the variances and m.s.e. are much more reasonable (relative to σ^4). If one places higher priority on bias reduction, it would not be unreasonable to prefer $m_n = n^{\frac{1}{2}}$, $\ell = 2$: the estimator that minimizes variance and m.s.e. amongst those estimators that are "best" on the bias criteria. Thus there are several arguments supporting the use of overlapping subseries under moderate dependence.

When the dependence is strong ($\phi = .9$) we are embarrassed to find that, as n increases, the minimum variance estimator ($m_n = \ell n\ n$, $\ell = 1$) is zeroing-in on a value that is ¼ of the true σ^2 . Now it takes the mammoth subseries of $m_n = n^{3/4}$ to wipe out the bias portion of m.s.e. But again the variances of the $m_n = n^{3/4}$ estimators seem prohibitively large. The overall pattern of variances is as in the previous cases, but when bias and variance contributions are combined into m.s.e., the estimator with $m_n = n^{\frac{1}{2}}$, $\ell = 2$ is superior for all values of n .

On the whole, this simulation lends credence to the use of overlapping subseries, particularly $\ell = 2$ (since $\ell = 3$ seems to suffer from increased covariances more than it gains by reducing bias). And, for this range of sample sizes, $m_n = n^{3/4}$ yields too few replications for it to be a stable estimator.

Looking carefully at the case $\phi = .9$ (where the true value of σ^2 is 100), it appears that the gains for allowing overlap do not quite measure up to what would be expected. For example, when $k_n = 12$ with $\ell = 3$ the estimator has expectation 34.7, but when $k_n = 12$ with $\ell = 2$ the expectation is approximately 42. In principle the bias reduction should be constant for a fixed subseries length k_n, but here the estimator does worse when more overlap is involved. Similarly, the expectations for: $k_n = 31$, $\ell = 1$ $(m_n = n^{\frac{1}{2}})$; $k_n = 30$, $\ell = 2$; $k_n = 30$, $\ell = 3$ are respectively: 69.8; 61.5; 55.1. Likewise the expectations for: $k_n = 22$, $\ell = 1$; $k_n = 20$, $\ell = 2$ are: 58.6; 47.9. And those for: $k_n = 62$, $\ell = 1$; $k_n = 62$, $\ell = 2$ $(m_n = n^{3/4})$ are: 85.3; 64.8. In each case there is substantially more bias as ℓ increases. The explanation is that when $\ell = 1$ all of the $(t_{k_n}^{im_n} - t_{k_n}^{jm_n})^2$ terms contribute 2 replicates of $(t_{k_n}^{im_n})^2$, but when $\ell \geq 2$ there are pairs with $1 \leq j-i \leq \ell-1$ that instead contribute 2 replicates of $(t_{(j-i)m_n}^{im_n})^2$. Being based upon shorter subseries, these latter replicates do not have the de-biasing effect which was the motivation for introducing overlap. Our special standardization of these pairs by $k_n/((j-i)m_n)^{\frac{1}{2}}$ makes these terms the correct order of magnitude (if s_n^i is mean-like), and including them in our estimator gives us more paired differences and hence more stability. But in terms of bias we would be better off excluding them and defining:

$$\tilde{\sigma}_n^2 := \sum_{0 \leq i < j \leq r_n - 1} (s_{k_n}^{im_n} - s_{k_n}^{jm_n})^2 \, \mathbb{I}\{j-i \geq \ell\} k_n / (r_n - \ell)(r_n - \ell + 1)$$

as our variance estimator. Of course, if the original number of pairs $(r_n(r_n-1)/2)$ is small (e.g. $m_n = n^{3/4}$), the reduction to $(r_n-\ell)(r_n-\ell+1)/2$ non-overlapping pairs will be disastrous. And in general we would expect $\tilde{\sigma}_n^2$ to have larger variance but smaller bias than $\hat{\sigma}_n^2$. To investigate these effects simulations of $\tilde{\sigma}_n^2$ were conducted, but excluding $m_n = n^{3/4}$ (due to insufficient r_n), and excluding $\ell = 1$ (for obvious reasons) and $\ell = 3$ (because the m.s.e. of $\ell = 2$ was usually better). The results for $\phi = .1, .5, .9$ are in Table 2.

Under weak dependence ($\phi = .1$) the $\hat{\sigma}_n^2$ estimators were nearly unbiased, and so are the $\tilde{\sigma}_n^2$ estimators. (A bubble (∘) indicates that for fixed ϕ, n, m_n, ℓ the $\tilde{\sigma}_n^2$ estimator is superior to or approximately equivalent to the corresponding $\hat{\sigma}_n^2$ estimator. A

TABLE 2. Simulation Study of $\tilde{\sigma}_n^2$. $\{Z_i\}$ is an AR(1) sequence with coefficient ϕ.

$\sigma^2 = \lim_{n\to\infty}(n\,V\{s_n^0\})$, $s_n^i = \bar{Z}_n^i$. Based on 1000 realizations.

n = sample size, m_n = "base" subseries length, $\ell = 2$ = overlap factor,

$k_n = \ell m_n$ = actual subseries length, r_n = # of subseries per sample.

°Better than (or approximately equal to) corresponding value for $\hat{\sigma}_n^2$.

+Best (or approx. best) for criteria, for fixed n and ϕ, among all $\hat{\sigma}_n^2$ and $\tilde{\sigma}_n^2$ estimates.

$\mathbb{E}$, V, MSE are simulated estimates of $\mathbb{E}\{\tilde{\sigma}_n^2\}$, $V\{\tilde{\sigma}_n^2\}$, $MSE\{\tilde{\sigma}_n^2\}$ respectively.

		n	m_n	k_n	r_n	$\mathbb{E}$	sd{$\mathbb{E}$}	V	$\frac{V}{MSE}$	$\frac{MSE}{\sigma^4}$
$\phi = .1,\ \sigma^2 = 1.23$	$m_n = \ell n\, n$	100	4	8	24	1.21°+	.014	.190°	.99	.125°
		250	5	10	49	1.20°+	.009	.088°	.98	.059°
		500	6	12	82	1.21°+	.008	.056°	.98	.037°
		1000	6	12	165	1.22°+	.005	.027°	1.00	.018°
	$m_n = n^{1/2}$	100	10	20	9	1.17°	.022	.484°	.99	.321°
		250	15	30	15	1.22°+	.018	.340	1.00	.223
		500	22	44	21	1.23°+	.016	.247°	1.00	.162°
		1000	31	62	31	1.22°+	.013	.165	1.00	.108
$\phi = .5,\ \sigma^2 = 4.0$	$m_n = \ell n\, n$	100	4	8	24	3.22°	.039	1.50°	.71	.131°+
		250	5	10	49	3.41°	.029	.85°	.71	.075°+
		500	6	12	82	3.53°	.023	.52°	.70	.046°+
		1000	6	12	165	3.54°	.017	.27°	.56	.030°+
	$m_n = n^{1/2}$	100	10	20	9	3.72°+	.077	5.97	.99	.378
		250	15	30	15	3.74	.057	3.23°	.98	.206°
		500	22	44	21	3.89°+	.051	2.59	1.00	.163
		1000	31	62	31	3.93°+	.041	1.65°	.99	.104°
$\phi = .9,\ \sigma^2 = 100$	$m_n = \ell n\, n$	100	4	8	24	28.1°	.49	237	.04	.540°
		250	5	10	49	36.6°	.42	177°	.04	.419°
		500	6	12	82	42.6°	.37	140	.04	.344°
		1000	6	12	165	42.9°	.25	62°	.02	.333°
	$m_n = n^{1/2}$	100	10	20	9	48.5°	1.2	1532	.37	.418°+
		250	15	30	15	68.0°	1.2	1397	.58	.242°+
		500	22	44	21	77.9°	1.1	1240	.72	.173°+
		1000	31	62	31	84.2°	.9	851°	.77	.110°+

plus (+) indicates that for fixed ϕ and n the $\tilde{\sigma}_n^2$ estimator is the best or approximate best amongst all eleven $\hat{\sigma}_n^2$ and $\tilde{\sigma}_n^2$ estimators.) In terms of variance and m.s.e., $\tilde{\sigma}_n^2$ is almost as good as $\hat{\sigma}_n^2$. The variance of $\tilde{\sigma}_n^2$ is hurt more when r_n is small ($m_n = n^{1/2}$) because then the loss of $(2r_n-\ell)(\ell-1)/2$ overlapping pairs is relatively greater.

The story is similar for $\phi = .5$: $\tilde{\sigma}_n^2$ has about the same bias, variance and m.s.e. as $\hat{\sigma}_n^2$ did, but again $\tilde{\sigma}_n^2$ suffers in terms of variance when r_n is small. Notice that $\tilde{\sigma}_n^2$ is an optimal (+) estimator in terms of m.s.e. when $m_n = \ell n$ n , and is usually optimal for bias reduction when $m_n = n^{1/2}$.

Turning to the case of strong dependence ($\phi = .9$) , we now see significant gains in debiasing by using $\tilde{\sigma}_n^2$: every expectation shows an improvement relative to the corresponding entry for $\hat{\sigma}_n^2$, and five out of eight of these increments are in excess of 2 s.d. units. Once again the variance of $\tilde{\sigma}_n^2$ tends to be inferior, but its bias is so superior that $\tilde{\sigma}_n^2$ actually ends up with smaller m.s.e. than $\hat{\sigma}_n^2$ in six out of eight cases.

Overall it seems that $\tilde{\sigma}_n^2$ performs as well as $\hat{\sigma}_n^2$ in terms of bias and m.s.e., but somewhat worse in terms of variance. Moreover, whenever $\hat{\sigma}_n^2$ (ℓ=2) was optimal (*) for bias or m.s.e., $\tilde{\sigma}_n^2$ retained that optimality. And when the dependence is strong, $\tilde{\sigma}_n^2$ offers substantial gains over $\hat{\sigma}_n^2$ in terms of bias and m.s.e.

If one is more concerned with bias and m.s.e. than with variance, then nothing is to be lost by using $\tilde{\sigma}_n^2$. And if one would like "insurance" against strong dependence, then there is something to be gained in using $\tilde{\sigma}_n^2$.

An analogous simulation study was conducted in order to investigate the behavior of $\tilde{\sigma}_n^2$ when s_n^i is the ratio statistic $\sum_{t=i+1}^{i+n-1} Z_t Z_{t+1} / \sum_{t=i+1}^{i+n-1} Z_t^2$. The results here echo those for $s_n^i = \bar{Z}_n^i$: there are substantial gains in debiasing for using $\ell = 2$ rather than $\ell = 1$ (for fixed ϕ, n, m_n), and this effect is more pronounced under heavier dependence. The estimator using $m_n = \ell n\ n$, $\ell = 2$ minimizes m.s.e. when $\phi = .1$ and $\phi = .5$ (for all n); but when $\phi = .9$ the debiasing effect of long subseries is so important that $m_n = n^{\frac{1}{2}}$, $\ell = 2$ has the best m.s.e.

Thus there is further support for using longer overlapping subseries in $\tilde{\sigma}_n^2$. Throughout these simulations, $\ell = 2$ in particular has made noticeable improvements over $\ell = 1$, while $\ell = 3$ had unacceptably inflated variance. The choice of $\{m_n\}$ seems to hinge upon the strength of dependence in $\{Z_i\}$: when augmented by overlap ($\ell = 2$), $m_n = \ell n\ n$ was quite acceptable for $\phi = .1$ and $\phi = .5$; but for $\phi = .9$ the extra length of $m_n = n^{\frac{1}{2}}$ was really necessary to control the bias. (Recall that in Section 5 our theoretical work suggested the need to relate $\{m_n\}$ to $\alpha(\cdot)$

by $(n/m_n)\alpha(m_n) \to 0$. This again requires longer subseries under strong dependence.) Perhaps the "safest" and most intuitive estimator under unknown dependence would be $\tilde{\sigma}_n^2$ with $\ell = 2$, $m_n = n^{\frac{1}{2}}$. It gives equal priority to r_n (# of replicates) and m_n (base subseries length), but then beefs up the subseries length for debiasing ($k_n = \ell m_n = 2m_n$), and ignores the confounding overlapping pairs ($j-i < \ell$). And it minimized m.s.e. when $\phi = .9$, for all values of n, for both statistics s_n^i.

7. Acknowledgment.

I thank Professor John Hartigan, my thesis advisor, for his guidance on this research.

REFERENCES

Beasley, J.D. and Springer, S.G. (1977). The Percentage Points of the Normal Distribution. Appl. Stat., 26, 118-120.

Billingsley, P. (1968). Convergence of Probability Measures. John Wiley and Sons, New York.

Carlstein, E. (1984). Asymptotic Normality for a General Statistic from a Stationary Sequence. University of North Carolina Institute of Statistics, Mimeo Series #1561

Chow, Y.S. and Teicher, H. (1978). Probability Theory. Springer-Verlag, New York.

Chung, K.L. (1974). A Course in Probability Theory. Academic Press, New York.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. The Ann. of Stat., 7, 1-26.

Efron, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia.

Freedman, D. (1984). On Bootstrapping Two-stage Least-squares Estimates in Stationary Linear Models. The Ann. of Stat., 12, 827-842.

Freedman, D.A. and Peters, S.C. (1984). Bootstrapping a Regression Equation: Some Empirical Results. Jour. of the Am. Stat. Assoc., 79, 97-106.

Gastwirth, J.L. and Rubin, H. (1975). The Asymptotic Distribution Theory of the Empiric CDF for Mixing Stochastic Processes. The Ann. of Stat., 3, 809-824.

Hartigan, J.A. (1969). Using Subsample Values as Typical Values. Jour. of the Am. Stat. Assoc., 64, 1303-1317.

Ibragimov, I.A. and Linnik, Yu.V. (1971). Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff Publishing, Groningen, The Netherlands.

Rosenblatt, M. (1956). A Central Limit Theorem and a Strong Mixing Condition. Proc. of the Nat. Acad. of Sc., 42, 43-47.

Tukey, J.W. (1958). Bias and Confidence in Not-quite Large Samples. The Ann. of Math. Stat., 29, 614.

Wichmann, B.A. and Hill, I.D. (1982). An Efficient and Portable Pseudo-random Number Generator. Appl. Stat., 31, 188-190.

Ed Carlstein
Dept. of Statistics
Univ. of N.C.
Phillips Hall, 039A
Chapel Hill, N.C. 27514