Confidence Intervals Using Orthonormally Weighted Standardized Time Series

Robert D. Foley David Goldsman

December 24, 1998

Abstract

We extend the standardized time series area method for constructing confidence intervals for the mean of a stationary process. These intervals are based on orthonormally weighted standardized time series and are designed to have more degrees of freedom than their predecessors. The higher degrees of freedom result in smaller mean and variance of the length of the confidence intervals.

Authors' addresses: David Goldsman, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, sman@isye.gatech.edu; Robert D. Foley, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, rfoley@isye.gatech.edu.

Keywords: Simulation, Stationary Process, Variance Estimation, Standardized Time Series, Orthonormal Area Estimator.

1 Introduction

An important problem in computer simulation is that of finding confidence intervals for the mean μ of a weakly stationary, discrete-time stochastic process $Y_1, Y_2, \ldots$. For example, the Y_i's could represent successive customer waiting times from some simulated queueing process, and we could be interested in finding a confidence interval for the mean waiting time.

If $Y_1, Y_2, \ldots, Y_n$ were independent, identically distributed (i.i.d.) normal random variables, then

$$\bar{Y}_n \ \pm \ t_{n-1,\alpha/2} S_n/\sqrt{n} \tag{1}$$

would be a $100(1-\alpha)\%$ confidence interval for μ, where $\bar{Y}_n$ is the sample mean, $t_{n-1,\alpha/2}$ is the $1-\alpha/2$ quantile of the t distribution with $n-1$ degrees of freedom, and S_n^2 is the sample variance. However, when the random variables are dependent, the confidence intervals from (1) perform poorly. Typically, the Y_i's are positively correlated, S_n^2 underestimates the quantity $\sigma_n^2 \equiv n\mathsf{Var}(\bar{Y}_n)$, and the coverage is smaller than the desired level $1-\alpha$. A natural method for overcoming these problems is to group the data into b batches of equal size, and then to treat the batch means as i.i.d., normal data points. This *batch means* method gives confidence intervals of the form

$$\bar{Y}_n \ \pm \ t_{b-1,\alpha/2} S_{B,b}/\sqrt{b}, \tag{2}$$

where $S_{B,b}^2$ is the sample variance of the batch means. The degrees of freedom $b-1$ should be as large as possible *subject to obtaining the desired coverage*. Indeed, as b increases and the batch size becomes large, it can be shown that the expected value and variance of the half-length of the confidence intervals decrease—yielding better confidence intervals provided that the appropriate coverage is obtained (Schmeiser [19]).

Besides batch means, a variety of other methods have been proposed including spectral analysis, overlapping batch means, regeneration, autoregressive modeling, and standardized time series (STS). (See, e.g., [4] for a review of such techniques.) Our method is an extension of the STS area and weighted area estimators of Schruben [20] and Goldsman and Schruben [13].

The *standardized time series* of $Y_1, Y_2, \ldots, Y_n$ is

$$T_n(t) \ \equiv \ \frac{\lfloor nt \rfloor(\bar{Y}_n - \bar{Y}_{\lfloor nt \rfloor})}{\sigma\sqrt{n}} \ \text{ for } \ 0 \le t \le 1, \tag{3}$$

where $\bar{Y}_j \equiv \sum_{i=1}^{j} Y_i/j$ for $j = 1, 2, \ldots$, $\lfloor\cdot\rfloor$ is the greatest integer function, and $\sigma^2 = \lim_{n\to\infty} n\mathsf{Var}(\bar{Y}_n) > 0$. The quantity σ^2 is called the *variance parameter*; as well as being interesting in its own right, the accurate and precise estimation of σ^2 is key to the good performance of confidence intervals.

The sample mean $\bar{Y}_n$ and the random function $\sigma T_n(\cdot)$ are computed from the data $Y_1, Y_2, \ldots, Y_n$. As pointed out in [20], the entire time series $Y_1, Y_2, \ldots, Y_n$ can be reconstructed from $\bar{Y}_n$ and $\sigma T_n(\cdot)$; thus, all of the information in the original time series is retained. The (weighted) area method computes

$$A(n) \ \equiv \ \frac{1}{n}\sum_{k=1}^{n} w(\frac{k}{n})\sigma T_n(\frac{k}{n}), \tag{4}$$

where $w(t)$ is a certain weighting function (to be discussed in the sequel); this statistic is then used to form confidence intervals for μ based on the t distribution with only one degree of freedom. Obviously, there is still a great deal of unused information left in $\sigma T_n(\cdot)$. We develop a method in which more of the information can be extracted in a useful and meaningful way. Specifically, we will use d different weighting functions to obtain d statistics from $\sigma T_n(\cdot)$ that are asymptotically i.i.d., normal random variables with mean zero and variance σ^2. We can then construct confidence intervals based on the t distribution with d degrees of freedom; and it turns out (analogous to the method of batch means) that the higher d is the better. Even though d can be as large as desired, we shall see that as d grows, the amount of data needed to obtain coverage $1-\alpha$ also grows. But even obtaining as few as two or three degrees of freedom instead of one from $\sigma T_n(\cdot)$ results in confidence intervals with smaller expected lengths.

The STS procedure of [13, 20] is carried out on each of $b \geq 1$ batches of observations, thereby extracting a total of b degrees of freedom (and better confidence intervals). Although the discussion in our paper is limited to the case of one batch, we can easily generalize the results to more than one batch (see Song [21]). Our work bears similarities to that of Dzhaparidze [6, pp. 276–277], who uses $d \geq 1$ multiple weighting functions on the standardized periodogram of a single batch of observations; the purpose is to incorporate more degrees of freedom into goodness-of-fit tests on the spectral density of a time series. We will also use multiple weighting functions in this paper.

The current paper concentrates on the discrete-time setting, $Y_1, Y_2, \ldots$, though all of our results hold in the *continuous*-time setting, Y_t, $t \geq 0$ (and sometimes with simpler proofs). Furthermore, the continuous-time setting includes the discrete-time setting simply by defining $Y_t \equiv Y_{\lfloor t \rfloor}$. Nevertheless, we shall keep the exposition in the discrete-time setting for two reasons. First, it seems to be encountered more often in practice. Second (and more importantly), users supplied with $Y_1, Y_2, \ldots$ in the continuous-time setting would typically approximate the weighted area under the standardized time series with the Riemann sum $A(n)$ from (4); users would usually avoid the analogous weighted area expression using $Y_t = Y_{\lfloor t \rfloor}$, i.e., $A'(n) \equiv \int_0^1 w(t)\sigma T_n(t)\,dt$. Even if $A'(n)$ possesses desirable properties, there is no guarantee that these properties will hold for the Riemann approximation $A(n)$. Hence, we have gone to the discrete-time setting and used the Riemann sum as the definition of $A(n)$; this should ensure that our results hold both in theory and for what is actually computed in practice.

The remainder of the paper is organized as follows. §2 derives a class of estimators for σ^2 and confidence intervals for μ. §3 defines desirable properties and gives sufficient conditions for them to hold. §4 describes implementation aspects which substantially decrease the computation time and storage space requirements. §5 contains an extended analytical example for the first-order moving average process. Empirical aspects of the performance of the confidence intervals are studied in §6. We provide conclusions in §7. The reference [7] is an early version of this paper containing preliminary results.

2 Variance Estimators and Confidence Intervals

Similar to Glynn and Iglehart [8], we need a reasonable assumption about $Y_1, Y_2, \ldots$ in deriving our confidence intervals for μ.

Assumption FCLT There exist μ and positive σ such that as $n \to \infty$,

$$X_n \stackrel{\mathcal{D}}{\to} \sigma\mathcal{W},$$

where $\mathcal{W}$ is a standard Brownian motion, $\stackrel{\mathcal{D}}{\to}$ denotes convergence in distribution as $n \to \infty$, and

$$X_n(t) \equiv \frac{\lfloor nt \rfloor(\bar{Y}_{\lfloor nt \rfloor} - \mu)}{\sqrt{n}} \quad \text{for } 0 \leq t \leq 1.$$

Remark 1 Assumption FCLT appears as if it should be a result and not an assumption. Since [8] lists several different sets of sufficient conditions for Assumption FCLT to hold, the user can select whichever is appropriate.

Remark 2 The sample paths of X_n lie in $D[0,1]$, the space of functions on $[0,1]$ which are right-continuous and have left-hand limits; the sample paths of $\mathcal{W}$ lie in $C[0,1]$, the space of continuous functions on $[0,1]$. Even though the sample paths of X_n lie in $D[0,1]$, we can interpret $\stackrel{\mathcal{D}}{\to}$ in the uniform topology since the jump points occur at fixed time points; cf. Billingsley [3, p. 153].

Recall that a Brownian bridge $\mathcal{B}(\cdot)$ (or tied-down Brownian motion as it is sometimes called since it is conditioned to equal zero at both $t = 0$ and 1) can be obtained from a standard Brownian motion $\mathcal{W}(\cdot)$ by letting $\mathcal{B}(t) = t\mathcal{W}(1) - \mathcal{W}(t)$. Furthermore, $\mathcal{W}(1)$ and $\mathcal{B}(\cdot)$ are independent.

Theorem 1 Under Assumption FCLT,

$$(\sqrt{n}(\bar{Y}_n - \mu), \sigma T_n) \stackrel{\mathcal{D}}{\to} (\sigma\mathcal{W}(1), \sigma\mathcal{B}).$$

Proof For $x \in D[0,1]$, define $h(x) \equiv (x(1), y)$, where $y(t) = tx(1) - x(t)$ for $0 \leq t \leq 1$. So $h(\sigma\mathcal{W}) = (\sigma\mathcal{W}(1), \sigma\mathcal{B}(\cdot))$. Similarly, define $h_n(x) \equiv (x(1), z)$, where $z(t) = tx(1) - x(t) + (\lfloor nt \rfloor - nt)x(1)/n$. Note that $h_n(X_n) = (\sqrt{n}(\bar{Y}_n - \mu), \sigma T_n(\cdot))$. From Theorem 5.5 and the subsequent comments of [3], we need only show that h is continuous and h_n converges to h uniformly on compact sets. On $D[0,1]$, we use the uniform metric (i.e., sup norm); see [3, §18]. It is easy to verify that the above conditions hold. □

Remark 3 Three useful properties fall out of Theorem 1:

1. $\sqrt{n}(\bar{Y}_n - \mu)$ is asymptotically $\sigma\mathsf{Nor}(0,1)$,
2. σT_n is asymptotically σ times a Brownian bridge, and
3. $\sqrt{n}(\bar{Y}_n - \mu)$ and σT_n are asymptotically independent; thus, all information gleaned from σT_n will be asymptotically independent of $\sqrt{n}(\bar{Y}_n - \mu)$.

We can extract more information from $\sigma T_n(\cdot)$ than from Equation (4) by using more than one "weighting" function. In particular, let

$$A_i(n) \;\equiv\; \frac{1}{n}\sum_{k=1}^{n} w_i(\frac{k}{n})\sigma T_n(\frac{k}{n}) \tag{5}$$

for weighting functions $w_i(\cdot)$, $i = 1,\ldots,d$. The next theorem says that for large n, $(A_1(n),\ldots,A_d(n))$ behaves like the multivariate normal random vector $(A_1,\ldots,A_d)$, where

$$A_i \;\equiv\; \int_0^1 w_i(t)\sigma\mathcal{B}(t)\,dt.$$

Theorem 2 If Assumption FCLT holds and $w_i(\cdot)$ is continuous on $[0,1]$, $i = 1,\ldots,d$, then as $n \to \infty$, $(A_1(n),\ldots,A_d(n)) \xrightarrow{\mathcal{D}} (A_1,\ldots,A_d)$.

Proof This proof is similar to that of Theorem 1 though we need to redefine h and h_n. Let $h(x)$ be a d-dimensional vector whose ith component is

$$\int_0^1 w_i(t)x(t)\,dt$$

for $x \in D[0,1]$. So $h(\sigma\mathcal{B}(t))$ is $(A_1,\ldots,A_d)$. Similarly, define $h_n(x)$ as a d-dimensional vector whose ith component is

$$\int_0^1 w_{i,n}(t)x(t)\,dt$$

for $x \in D[0,1]$, where $w_{i,n}(t) \equiv w_i(\lfloor nt \rfloor/n)$, $0 \le t \le 1$. Thus, $h_n(\sigma T_n(t))$ is $(A_1(n),\ldots,A_d(n))$. Using the continuity of the w_i's, it is not hard to show that h is continuous and h_n converges to h uniformly on compact sets. $\square$

The continuous mapping theorem immediately gives us the next corollary.

Corollary 1 If Assumption FCLT holds and $w_i(\cdot)$ is continuous on $[0,1]$, $i = 1,\ldots,d$, then as $n \to \infty$,

$$V_W(n) \;\equiv\; \sum_{i=1}^{d} A_i^2(n)/d \;\xrightarrow{\mathcal{D}}\; \sum_{i=1}^{d} A_i^2/d.$$

The next theorem gives conditions that the weighting functions must satisfy so that $A_1, \ldots, A_d$ are i.i.d. $\mathsf{Nor}(0, \sigma^2)$ random variables. First, we need a definition.

Condition O The functions $w_1, \ldots, w_d$ are said to satisfy Condition O if they are orthonormal with respect to $r(s,t) \equiv (s \wedge t)[1 - (s \vee t)]$ over the unit square, where $\wedge$ denotes minimum and $\vee$ denotes maximum; i.e.,

$$\int_0^1 \int_0^1 w_i(s) w_j(t) r(s,t)\, ds\, dt \;=\; \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

Theorem 3 Suppose that Assumption FCLT holds and $w_1, \ldots, w_d$ are continuous on $[0,1]$. Then $A_1, \ldots, A_d$ are i.i.d. $\mathsf{Nor}(0, \sigma^2)$ random variables iff $w_1, \ldots, w_d$ satisfy Condition O.

Proof Clearly, $(A_1, \ldots, A_d)$ has a multivariate normal distribution with mean zero. We need to show that the covariance matrix is σ^2 times the identity matrix. Now,

$$\begin{aligned} \mathsf{Cov}(A_i, A_j) &= \mathsf{Cov}\left[\int_0^1 w_i(s)\sigma\mathcal{B}(s)\, ds, \int_0^1 w_j(t)\sigma\mathcal{B}(t)\, dt\right] \\ &= \sigma^2 \int_0^1 \int_0^1 w_i(s) w_j(t) \mathsf{Cov}(\mathcal{B}(s), \mathcal{B}(t))\, ds\, dt. \end{aligned}$$

Since $\mathsf{Cov}(\mathcal{B}(s), \mathcal{B}(t)) = r(s,t)$, the covariance matrix will be σ^2 times the identity matrix iff $w_1, \ldots, w_d$ are orthonormal with respect to $r(s,t)$. $\Box$

Theorem 3 only gives us the conditions the weighting functions must satisfy, not the weighting functions themselves. One method of obtaining orthonormal weighting functions is to take any set of linearly independent functions $v_1, \ldots, v_d$ and use the Gram-Schmidt procedure to orthonormalize them.

Example 1 Suppose we let $v_i(t) = t^{i-1}$, $i = 1, 2, 3, 4$. Note that the v_i's are linearly independent. Applying the Gram-Schmidt procedure with respect to $r(s,t)$ yields the orthonormal weighting functions

$$\begin{aligned} w_1(t) &= \sqrt{12} \\ w_2(t) &= \sqrt{720}(t - 1/2) \\ w_3(t) &= \sqrt{25200}(t^2 - t + 1/5) \\ w_4(t) &= 60(14t^3 - 21t^2 + 9t - 1). \end{aligned}$$

The following corollary is a consequence of Theorem 3 and Corollary 1.

Corollary 2 Suppose that Assumption FCLT holds and that $w_i(\cdot)$ is continuous on $[0,1]$, $i = 1, \ldots, d$. If $w_1, \ldots, w_d$ satisfy Condition O, then $\sum_{i=1}^{d} A_i^2/\sigma^2$ is a chi-squared random variable with d degrees of freedom, and $V_W(n) \xrightarrow{\mathcal{D}} \sigma^2\chi_d^2/d$.

Each $A_i^2(n)$, $i = 1, \ldots, d$, is an estimator for σ^2; we refer to $V_W(n)$ as the *orthonormally weighted area estimator* for σ^2. The area estimator from [20] corresponds to the case with $d = 1$ and $w_1(t) = \sqrt{12}$, while the weighted area estimators of [13] correspond to $d = 1$.

Once we have orthonormal weighting functions, we can compute the $A_i(n)$'s and then $V_W(n)$. Since $V_W(n)$ is asymptotically $\sigma^2\chi_d^2/d$, $\sqrt{n}(\bar{Y}_n - \mu)$ is asymptotically $\mathsf{Nor}(0, \sigma^2)$, and the two are asymptotically independent (cf. [8, 20]), we know that the pivot

$$\frac{\bar{Y}_n - \mu}{\sqrt{V_W(n)/n}} \xrightarrow{\mathcal{D}} t_d, \tag{6}$$

a t random variable with d degrees of freedom. Hence, an approximate $100(1-\alpha)\%$ confidence interval for μ is given by

$$\bar{Y}_n \pm t_{d,\alpha/2}\sqrt{\frac{V_W(n)}{n}}. \tag{7}$$

3 Properties

Suppose we have several orthonormal estimators $A_1^2(n), \ldots, A_d^2(n)$ computed from the same information. Since there is a virtually unlimited supply of orthonormal weighting functions, we should select those $A_i^2(n)$'s that have certain desirable properties. For example, although all of the $A_i^2(n)$'s are asymptotically unbiased for σ^2, there is usually bias for finite sample sizes. Therefore, one desirable property is that the bias of $A_i^2(n)$ as an estimator of σ^2 converges to zero quickly. In particular, we will call an estimator *first-order unbiased* for σ^2 if $\mathsf{E}[A_i^2(n)] = \sigma^2 + o(1/n)$. §3.2 gives results that will enable us to find the bias of $A_i^2(n)$ as an estimator of σ^2, and then to find first-order unbiased estimators for σ^2.

Asymptotically, the estimators $A_1^2(n), \ldots, A_d^2(n)$ are independent, but for finite sample size n, there is usually some correlation between the estimators. Thus, another desirable property for the estimators to have is that they are *first-order uncorrelated*, i.e., $\mathsf{Corr}(A_i^2(n), A_j^2(n)) = o(1/n)$ for $i \neq j$. This is the topic of §3.3. We would also prefer that the $A_i^2(n)$'s be approximately independent of the sample mean $\bar{Y}_n$. This property would help the distribution of the confidence interval pivot to be approximately t_d. We give relevant results in §3.4. Lastly, §3.5 presents asymptotic properties of our confidence interval estimators for μ.

3.1 Assumptions

Before we find first-order unbiased and uncorrelated estimators, we need to make some assumptions on the covariance function $R_k \equiv \mathsf{Cov}(Y_1, Y_{1+k})$ and on the weighting functions $w_1, \ldots, w_d$.

Assumptions FO

1. $Y_1, Y_2, \ldots$ is stationary,
2. Assumption FCLT holds,
3. $\sum_{k=1}^{\infty} k^2 |R_k| \; < \; \infty$,
4. $\sum_{k=-\infty}^{\infty} R_k \; = \; \lim_{n\to\infty} \sigma_n^2 \; = \; \sigma^2 \; > \; 0$,
5. w_i'' is continuous and bounded on $[0,1]$, $i = 1, \ldots, d$, and
6. $\int_0^1 \int_0^1 w_i(s) w_i(t) r(s,t)\, ds\, dt = 1$, $i = 1, \ldots, d$.

Assumption FO.3 is a technical condition on the underlying stochastic process that is satisfied by many stationary stochastic processes of interest; it is needed in our main proof. Assumption FO.4 should not be interpreted as a definition of σ^2 since that quantity was already implicitly defined as the appropriate constant needed in the functional central limit theorem. Sufficient conditions for σ^2 to be equal to $\sum_{k=-\infty}^{\infty} R_k$ appear in [8]. Assumption FO.5 is a mild technical condition on the weighting function that allows us to invoke the trapezoid rule in our main proof.

3.2 Bias and Variance of $A_i^2(n)$

In order to obtain an expression for the bias of $A_i^2(n)$ (and, implicitly, the bias of $V_W(n)$), we first define the quantities

$$\begin{aligned}
W_i(t) &\equiv \int_0^t w_i(s)\, ds, \quad 0 \le t \le 1; \quad i = 1, \ldots, d, \\
W_i &\equiv W_i(1), \quad i = 1, \ldots, d, \\
\bar{W}_i(t) &\equiv \int_0^t W_i(s)\, ds, \quad 0 \le t \le 1; \quad i = 1, \ldots, d, \\
\bar{W}_i &\equiv \bar{W}_i(1), \quad i = 1, \ldots, d, \\
c_{ij} &\equiv (W_i - \bar{W}_i)(W_j - \bar{W}_j) + \bar{W}_i \bar{W}_j, \quad i, j = 1, \ldots, d, \quad \text{and} \\
\gamma &\equiv -2 \sum_{k=1}^{\infty} k R_k.
\end{aligned}$$

Using Assumptions FO, we can derive the next theorem, a generalization of the discrete-time analog of Corollary 4.2 from Goldsman, Meketon, and Schruben [11].

Theorem 4 Under Assumptions FO,

$$\mathsf{Cov}(A_i(n), A_j(n)) \;=\; \sigma^2 \int_0^1 \int_0^1 w_i(s) w_j(t) r(s,t)\, ds\, dt + \frac{\gamma c_{ij}}{2n} + O(1/n^2).$$

Proof See the appendix. □

We now have an important corollary concerning the expected value of $A_i^2(n)$.

Corollary 3 Under Assumptions FO,

$$\mathsf{Var}(A_i(n)) \;=\; \mathsf{E}[A_i^2(n)] \;=\; \sigma^2 + \frac{\gamma c_{ii}}{2n} + O(1/n^2).$$

Corollary 3 gives the bias of $A_i^2(n)$ as an estimator of σ^2. The term $\gamma c_{ii}/2n$ constitutes the first-order bias, where γ is a constant that depends only on the stochastic process $Y_1, Y_2, \ldots$, and c_{ii} is a constant that depends only on w_i (cf. Song and Schmeiser [23]).

Example 2 For the w_i's from Example 1, we have $c_{11} = 3$, $c_{22} = 5$, $c_{33} = 7$, $c_{44} = 9$. So the weighted area estimator of σ^2 can be quite biased for stochastic processes with $|\gamma|$ large.

If we can select weighting functions with $c_{ii} = 0$, then $A_i^2(n)$ will be first-order unbiased for σ^2. Although a first-order unbiased estimator may not be unbiased, we would expect that such an estimator would have a smaller bias and, hence, yield "better" confidence intervals (see Sargent, Kang, and Goldsman [18]).

Example 3 If $w_1(t) = \sqrt{840}(3t^2 - 3t + 1/2)$, then $c_{11} = 0$, and so $A_1^2(n)$ is first-order unbiased for σ^2 (cf. [11]).

Here we define a condition on the w_i's that will be sufficient for an estimator to be first-order unbiased for σ^2.

Condition U The weighting function w_i is said to satisfy Condition U if $W_i = \bar{W}_i = 0$.

Corollary 4 Under Assumptions FO, the random variable $A_i^2(n)$ is first-order unbiased if w_i satisfies Condition U.

Proof Follows from the definition of c_{ii}. □

Remark 4 Although we have no control over γ (since it depends on the underlying stochastic process), $A_i^2(n)$ is first-order unbiased if it happens that $\gamma = 0$, e.g., when $Y_1, Y_2, \ldots$ are i.i.d.

Given a set of linearly independent weighting functions satisfying Condition U, we can use the Gram-Schmidt procedure to orthonormalize these weighting functions to satisfy Condition O. Fortunately, since Condition U is preserved under linear combinations, we see that it is preserved during the Gram-Schmidt procedure. Thus, we will be able to find weights that satisfy both Conditions O and U.

Example 4 Two polynomial weighting functions satisfying Conditions O and U are

$$\begin{aligned} w_1(t) &= \sqrt{\frac{63000}{19}}\left(\frac{1}{5} - \frac{9t}{10} + t^3\right) \quad \text{and} \\ w_2(t) &= \sqrt{574560}\left(\frac{-2}{57} + \frac{31t}{76} - t^2 + \frac{25t^3}{38}\right). \end{aligned}$$

It would be useful to have an infinite set of reasonably simple weighting functions that satisfy Conditions O and U. The next example gives such a set.

Example 5 An infinite sequence of weighting functions satisfying Conditions O and U is $w_i(t) = \sqrt{8}\pi i \cos(2\pi i t)$, $i = 1, 2, \ldots$.

Remark 5 The weightings from Example 5 are also suggested by Bartlett and Diananda [2] in the context of testing for autoregressive processes (also see Dzhaparidze [6]). In fact, the example is similar to a Karhunen-Loève expansion (cf. Ash and Gardner [1]) with the added wrinkle that the orthonormal functions must satisfy Condition U. We will use the weighting functions of Example 5 in our empirical studies.

3.3 Correlation between $A_i^2(n)$ and $A_j^2(n)$

As mentioned earlier, another desirable property for estimators to have is that they are first-order uncorrelated. Although we cannot always give conditions for $A_1^2(n), \ldots, A_d^2(n)$ to be first-order uncorrelated, we can do so for $A_1(n), \ldots, A_d(n)$. In fact, with no additional conditions beyond O and U, the next corollary shows that $\mathsf{Corr}(A_i(n), A_j(n)) = o(1/n)$ for $i \neq j$.

Corollary 5 If Assumptions FO hold, and if w_i and w_j, $i \neq j$, satisfy Conditions O and U, then $\mathsf{Corr}(A_i(n), A_j(n)) = O(1/n^2)$.

Proof Follows from Theorem 4 and Corollary 3. □

If we are willing make more assumptions, then we can make a stronger statement concerning $\mathsf{Corr}(A_i^2(n), A_j^2(n))$.

Corollary 6 Suppose Assumptions FO hold, $(A_i(n), A_j(n))$ is bivariate normal, and $\mathsf{Var}(A_\ell^2(n)) \to \mathsf{Var}(A_\ell^2)$, $\ell = i, j$. If w_i and w_j, $i \neq j$, satisfy Conditions O and U, then $\mathsf{Corr}(A_i^2(n), A_j^2(n)) = O(1/n^4)$.

Proof Since $(A_i(n), A_j(n))$ have respective means equal to zero, Patel and Read [16, p. 309] and Theorem 4 imply that

$$\mathsf{Cov}(A_i^2(n), A_j^2(n)) \;=\; 2\mathsf{Cov}^2(A_i(n), A_j(n)) \;=\; O(1/n^4).$$

The proof is completed by noting that $\mathsf{Var}(A_i^2(n))$ converges to a constant, viz., $2\sigma^4$. □

3.4 Correlation between $\bar{Y}_n$ and $A_i^2(n)$

We also stated earlier that another nice property would be for the sample mean $\bar{Y}_n$ to be approximately independent of the $A_i^2(n)$'s. We give the following related results, both of which show that the associated correlations disappear quickly as the sample size grows.

Theorem 5 If Assumptions FO hold, and if w_i satisfies Condition U, then $\mathsf{Corr}(\bar{Y}_n, A_i(n)) = o(1/n)$.

Proof See the appendix. □

Again, more assumptions will allow us to make a stronger statement concerning $\mathsf{Corr}(\bar{Y}_n, A_i^2(n))$.

Theorem 6 Suppose Assumptions FO hold and $(\bar{Y}_n, A_i(n))$ is bivariate normal. Then $\mathsf{Corr}(\bar{Y}_n, A_i^2(n)) = 0$.

Proof Without loss of generality, we can assume that $\mathsf{E}[\bar{Y}_n] = 0$. Then the result follows from Patel and Read [16, p. 309]. □

3.5 Confidence Intervals

The confidence intervals for μ given by Equation (7) are asymptotically valid, i.e., they achieve approximately the desired coverage as the run size $n \to \infty$ (with d fixed). Given that a particular confidence interval attains the nominal coverage $1-\alpha$, we might be interested in studying the *lengths* of the confidence intervals—the smaller the better if the desired coverage is obtained.

To this end, we define the half-length of the orthonormally weighted area confidence interval for μ as $H \equiv t_{d,\alpha/2}\sqrt{V_W(n)/n}$. Similar to Schmeiser [19] and Goldsman and Schruben [12], we find that as $n \to \infty$ (with d fixed),

$$\sqrt{n}\,H \;\xrightarrow{\mathcal{D}}\; \sigma t_{d,\alpha/2}\sqrt{\chi_d^2/d}.$$

Assuming uniform integrability, we can obtain

$$\sqrt{n}\,\mathsf{E}[H] \;\to\; \sigma t_{d,\alpha/2}\sqrt{\frac{2}{d}}\frac{\Gamma((d+1)/2)}{\Gamma(d/2)} \tag{8}$$

and

$$n\mathsf{Var}(H) \;\to\; \sigma^2 t_{d,\alpha/2}^2\left\{1-\frac{2}{d}\left[\frac{\Gamma((d+1)/2)}{\Gamma(d/2)}\right]^2\right\}, \tag{9}$$

where $\Gamma(\cdot)$ is the gamma function. The right-hand sides of Equations (8) and (9) both decrease in d.

Although Schmeiser [19] argues that there is not much to be gained in terms of $\mathsf{E}[H]$ and $\mathsf{Var}(H)$ by taking $d > 30$, it is interesting (at least pedagogically) to study the coverage of the confidence intervals for "large" d. In particular, we want to see how the coverage behaves as we push the value of d higher. (Also see Example 10 in §6.) If d is large and we assume that the $A_i^2(n)$'s, $i = 1,\ldots,d$, are approximately independent, then a central limit theorem (see Rohatgi [17]) implies that

$$V_W(n) \;=\; \frac{1}{d}\sum_{i=1}^{d} A_i^2(n) \;\approx\; \mathsf{Nor}(\mathsf{E}[V_W(n)],\mathsf{Var}(V_W(n))),$$

where the notation $\approx$ is taken to mean "is approximately distributed as". Thus,

$$\frac{\bar{Y}_n-\mu}{\sqrt{V_W(n)/n}} \;=\; \frac{(\bar{Y}_n-\mu)/\sqrt{\sigma^2/n}}{\sqrt{V_W(n)/\sigma^2}} \;\approx\; \widetilde{T} \;\equiv\; \frac{Z}{|X|^{1/2}},$$

where $Z \sim \mathsf{Nor}(0,1)$ and $X \sim \mathsf{Nor}(\mathsf{E}[V_W(n)],\mathsf{Var}(V_W(n)))/\sigma^2 \approx \mathsf{Nor}(1,2/d)$. Then the probability that the confidence interval (7) covers μ is approximately $\mathsf{Pr}\{|\widetilde{T}| \le t_{d,\alpha/2}\} =$

$2F_{\widetilde{T}}(t_{d,\alpha/2}) - 1$, where the cumulative distribution function (c.d.f.) of $\widetilde{T}$ is given by

$$\begin{aligned} F_{\widetilde{T}}(s) &\equiv \mathsf{Pr}\{\widetilde{T} \leq s\} = \mathsf{Pr}\{Z \leq s|X|^{1/2}\} \\ &= \int_{-\infty}^{\infty} \mathsf{Pr}\{Z \leq s|x|^{1/2}\} f_X(x)\,dx \doteq \int_0^{\infty} \Phi(sx^{1/2}) f_X(x)\,dx, \end{aligned} \tag{10}$$

and where $\Phi(\cdot)$ is the standard normal c.d.f., $f_X(\cdot)$ is the probability density function (p.d.f.) of X, and $\mathsf{Var}(X) \doteq 2/d$ is small enough to assume that $\mathsf{Pr}\{X < 0\}$ is negligible. (Note that the integral in (10) can be evaluated numerically.)

4 Implementation Aspects

We have presented the weighted area $A_i(n)$ as arising from a two-stage process: compute the standardized time series $\sigma T_n(t)$; then compute the weighted area $A_i(n)$ using $\sigma T_n(t)$ and the weighting function w_i. Intuitively, this is a good way to think of $A_i(n)$, but computationally, it is inefficient. Let us rewrite Equation (5) to see if we can find more efficient ways to compute $A_i(n)$ directly from $Y_1, Y_2, \ldots, Y_n$. (See Dzhaparidze [6] for similar results related to periodograms.) Equations (3) and (5) and some algebra yield

$$\begin{aligned} A_i(n) &= \frac{1}{n}\sum_{k=1}^{n} w_i(k/n)\frac{k(\bar{Y}_n - \bar{Y}_k)}{\sqrt{n}} \\ &= \frac{1}{n^{3/2}}\left[\left(\sum_{k=1}^{n}\frac{k}{n}w_i(\frac{k}{n})\right)\left(\sum_{j=1}^{n}Y_j\right) - \sum_{k=1}^{n}w_i(\frac{k}{n})\sum_{j=1}^{k}Y_j\right] \qquad (11) \\ &= \frac{1}{n^{3/2}}\sum_{j=1}^{n} g_i(j)Y_j, \qquad (12) \end{aligned}$$

where we define

$$g_i(j) \equiv \sum_{k=1}^{n}(k/n)w_i(k/n) - \sum_{k=j}^{n} w_i(k/n) \tag{13}$$

for a weighting function w_i.

Equation (12) provides an easy way to compute $A_i(n)$, $i = 1, \ldots, d$; it requires only that we use (13) to calculate (or store *a priori*) $g_i(j)$, $j = 1, \ldots, n$, before summing the linear combination of the Y_j's in one pass. This expression for $A_i(n)$ will allow us to derive its performance characteristics for the specific analytical examples to be presented in §5. Equation (11) gives an even more efficient way to calculate the $A_i(n)$'s, $i = 1, \ldots, d$, in one pass. This alternative requires only that we maintain and update the $2d+1$ quantities $\sum_{j=1}^{\ell} Y_j$, $\sum_{k=1}^{\ell}(k/n)w_i(k/n)$, and $\sum_{k=1}^{\ell} w_i(k/n)\sum_{j=1}^{k} Y_j$, $i = 1, \ldots, d$, the only nontrivial update being, for $\ell = 1, \ldots, n-1$,

$$\sum_{k=1}^{\ell+1} w_i(\frac{k}{n}) \sum_{j=1}^{k} Y_j \;=\; \sum_{k=1}^{\ell} w_i(\frac{k}{n}) \sum_{j=1}^{k} Y_j + w_i(\frac{\ell+1}{n}) \sum_{j=1}^{\ell+1} Y_j.$$

5 Analytical Results for the MA(1) Process

For some simple stochastic processes and low-order polynomial weighting functions, it is possible to carry out the algebra to obtain *exact* results on $V_W(n)$. Throughout this section, we find such exact results for a first-order moving average [MA(1)] process. The MA(1) is given by $Y_{j+1} = \theta\epsilon_j + \epsilon_{j+1}$, $j = 1, 2, \ldots$, where the ϵ_j's are i.i.d. $\mathsf{Nor}(0,1)$; so $R_0 = 1 + \theta^2$, $R_{\pm 1} = \theta$, and $R_k = 0$, otherwise. It is easy to see that for the MA(1), $\sigma^2 = (1+\theta)^2$ and $\gamma = -2\theta$.

§5.1 analyzes the bias, variance, and mean squared error (m.s.e.) of the weighted area and batch means estimators for σ^2. In §5.2, we study the correlation between $A_i^2(n)$ and $A_j^2(n)$. §5.3 is concerned with the correlation between the sample mean $\bar{Y}_n$ and $A_i^2(n)$; and §5.4 gives exact results for confidence interval coverage and expected half-length for some special cases.

5.1 Bias, Variance, and Mean Squared Error of $A_i^2(n)$

For an MA(1) process, Equation (12) gives

$$\begin{aligned} \mathsf{Cov}(A_1(n), A_2(n)) \;&=\; \frac{1}{n^3} \sum_{k=1}^{n} \sum_{\ell=1}^{n} g_1(k) g_2(\ell) \mathsf{Cov}(Y_k, Y_\ell) \\ =\; \frac{1}{n^3} \Bigg\{ (1+\theta^2) \sum_{j=1}^{n} g_1(j) g_2(j) &+ \theta \sum_{j=1}^{n-1} [g_1(j) g_2(j+1) + g_1(j+1) g_2(j)] \Bigg\}, \end{aligned} \tag{14}$$

whence

$$\mathsf{E}[A_1^2(n)] \;=\; \mathsf{Var}(A_1(n)) \;=\; \frac{1}{n^3} \left[(1+\theta^2) \sum_{j=1}^{n} g_1^2(j) + 2\theta \sum_{j=1}^{n-1} g_1(j) g_1(j+1) \right]. \tag{15}$$

We can now determine the bias of $A_1^2(n)$ as an estimator of σ^2.

Example 6 Consider the weighting function $w_1(t) = \sqrt{840}(3t^2 - 3t + 1/2)$ from Example 3. If we calculate $g_1(j)$ from Equation (13) and then plug into the above equation, we obtain

$$\mathsf{E}[A_1^2(n)] \;=\; \sigma^2 + \frac{7(\theta^2 - 10\theta + 1)}{2n^2} + \frac{105\theta}{n^3} + O(1/n^4).$$

This directly shows that $\mathsf{E}[A_1^2(n)] = \sigma^2 + O(1/n^2)$; hence, $A_1^2(n)$ is first-order unbiased (which makes sense in light of Example 3). Note that if θ happens to equal $(10 \pm \sqrt{96})/2$, then $A_1^2(n)$ is *second*-order unbiased for σ^2.

Since $A_1(n)$ is a linear combination of zero-mean, jointly normal random variables, we have $A_1(n) \sim \mathsf{Nor}(0, \mathsf{E}[A_1^2(n)])$. This implies that $A_1^2(n) \sim \mathsf{E}[A_1^2(n)]\chi_1^2$, and so $\mathsf{Var}(A_1^2(n)) = 2\mathsf{E}^2[A_1^2(n)]$. Thus, if we use the weighting function w_1 from Example 6, the m.s.e. of $A_1^2(n)$ as an estimator of σ^2 is $\mathsf{MSE}(A_1^2(n)) = \mathsf{Bias}^2[A_1^2(n)] + \mathsf{Var}(A_1^2(n)) = 2\sigma^4 + O(1/n^2)$.

Remark 6 (Comparison to Batch Means) Perhaps the most popular estimator for σ^2 is the *batch means* estimator mentioned in §1,

$$V_{B,b}(n) \;\equiv\; mS_{B,b}^2 \;=\; \frac{m}{b-1}\sum_{i=1}^{b}(\bar{Y}_{i,m} - \bar{Y}_n)^2 \;=\; \frac{m}{b-1}\left(\sum_{i=1}^{b}\bar{Y}_{i,m}^2 - b\bar{Y}_n^2\right),$$

where we divide the n observations into b adjacent, nonoverlapping batches, each of length m (assume $n = mb$), and where the ith batch mean is $\bar{Y}_{i,m} \equiv \sum_{k=1}^{m} Y_{(i-1)m+k}/m$ for $i = 1, \ldots, b$. From [18], we have $\mathsf{E}[V_{B,b}(n)] = \sigma^2 + \gamma(b+1)/n = \sigma^2 + O(1/n)$. So the weighted area estimator using a first-order unbiased weight is generally less biased for σ^2. From [5], we have $\mathsf{Var}(V_{B,b}(n)) = 2\sigma^4/(b-1) + O(1/n)$, and so the m.s.e. of $V_{B,2}(n)$, the batch means estimator for σ^2 based on $b = 2$ batches, is $2\sigma^4 + O(1/n)$—about the same as that of the weighted area estimator based on just $d = 1$ degree of freedom.

5.2 Correlation between $A_i^2(n)$ and $A_j^2(n)$

We illustrate the calculation of the $\mathsf{Corr}(A_1^2(n), A_2^2(n))$ for an MA(1) process.

Example 7 Consider the weighting functions w_1 and w_2 from Example 4 (which satisfy Conditions O and U). If we calculate $g_i(j)$ from Equation (13), $i = 1, 2$, and then substitute into Equation (14), we obtain

$$\mathsf{Cov}(A_1(n), A_2(n)) \;=\; \frac{5\sqrt{3}(\theta^2 - 10\theta + 1)}{19n^2} + O(1/n^3) \;=\; O(1/n^2).$$

Then [16, p. 309] yields

$$\mathsf{Cov}(A_1^2(n), A_2^2(n)) \;=\; 2\mathsf{Cov}^2(A_1(n), A_2(n)) \;=\; O(1/n^4).$$

Further, by Corollary 3, $\mathsf{Var}(A_i^2(n)) = 2\mathsf{Var}^2(A_i(n)) \to 2\sigma^4$ as $n \to \infty$ for $i = 1, 2$. Thus, $\mathsf{Corr}(A_1^2(n), A_2^2(n)) = O(1/n^4)$, a result that is in line with Corollary 6.

5.3 Correlation between $\bar{Y}_n$ and $A_i^2(n)$

We can also study the covariance between $A_i(n)$ and the sample mean $\bar{Y}_n$ for the MA(1). In particular,

$$\begin{aligned} n^{5/2}\mathsf{Cov}(\bar{Y}_n, A_i(n)) &= \sum_{k=1}^{n} g_i(k) \sum_{j=1}^{n} \mathsf{Cov}(Y_j, Y_k) \\ &= (1+\theta^2)\sum_{k=1}^{n} g_i(k) + 2\theta \sum_{k=1}^{n} g_i(k) - \theta(g_i(1)+g_i(n)) \;=\; -\theta(g_i(1)+g_i(n)), \end{aligned}$$

the last step a result of the fact that $\sum_{k=1}^{n} g_i(k) = 0$. Since $\bar{Y}_n$ and $A_i(n)$ are jointly normal, they will be *independent* if $\mathsf{Cov}(\bar{Y}_n, A_i(n)) = 0$, i.e., if $\theta = 0$ or $g_i(1) + g_i(n) = 0$. In addition, Theorem 6 immediately shows us that $\mathsf{Corr}(\bar{Y}_n, A_i^2(n)) = 0$.

Example 8 Let $w_1(t) = \sqrt{840}(3t^2 - 3t + 1/2)$ as in Example 3 (which satisfies Condition U). After some algebra,

$$g_1(j) = -\frac{\sqrt{840}}{4n^2}\left[4j^3 - 6j^2(n+1) + 2j(n^2+3n+1) - n(n+1)\right],$$

so that $g_1(1) + g_1(n) = 0$. Thus, for the MA(1) process with weighting function w_1, we have that $\bar{Y}_n$ and $A_1(n)$ are not only first-order uncorrelated, they are independent.

Unfortunately, it can be shown that this exact independence does *not* necessarily manifest itself for any w_i satisfying Condition U.

5.4 Properties of Confidence Intervals

We can give exact results for coverage and expected half-length for several special cases of the approximate $100(1-\alpha)\%$ confidence interval for μ given by Equation (7) with $d = 1$. For the remainder of this subsection, consider the first-order unbiased weighting function $w_1(t) = \sqrt{840}(3t^2 - 3t + 1/2)$ from Example 3. Since

- $A_1^2(n) \sim \mathsf{E}[A_1^2(n)]\,\chi_1^2$ (see §5.1),
- $\bar{Y}_n \sim \mathsf{Nor}(\mu, \sigma_n^2/n)$, and
- $\bar{Y}_n$ and $A_1^2(n)$ are independent (see Example 8),

the confidence interval pivot is exactly distributed as (cf. (6) with $d = 1$)

$$\widetilde{Y}_n \equiv \frac{\bar{Y}_n - \mu}{\sqrt{A_1^2(n)/n}} \sim t_1/u(n),$$

where $u(n) \equiv \sqrt{\mathsf{E}[A_1^2(n)]/\sigma_n^2}$.

The resulting two-sided weighted area confidence interval has coverage

$$\begin{aligned} \mathsf{CVG}_W &\equiv \mathsf{Pr}(-t_{1,\alpha/2} \le \widetilde{Y}_n \le t_{1,\alpha/2}) \\ &= \mathsf{Pr}(-t_{1,\alpha/2}u(n) \le t_1 \le t_{1,\alpha/2}u(n)) \\ &= \frac{2}{\pi}\arctan(t_{1,\alpha/2}u(n)). \end{aligned}$$

Since $A_1^2(n) \sim \mathsf{E}[A_1^2(n)]\chi_1^2$, the expected half-length is

$$\begin{aligned} \mathsf{E}[H_W] &\equiv t_{1,\alpha/2}\mathsf{E}[(A_1^2(n)/n)^{1/2}] \\ &= t_{1,\alpha/2}(\mathsf{E}[A_1^2(n)/n])^{1/2}\mathsf{E}[(\chi_1^2)^{1/2}] \\ &= t_{1,\alpha/2}\sqrt{2\mathsf{E}[A_1^2(n)]/n\pi} \end{aligned}$$

(cf. [18]). Similarly, one can also show that the batch means method with two batches yields coverage

$$\mathsf{CVG}_B \equiv \frac{2}{\pi}\arctan\Big(t_{1,\alpha/2}\sqrt{\mathsf{E}[V_{B,2}(n)]/\sigma_n^2}\,\Big)$$

and expected half-length

$$\mathsf{E}[H_B] \equiv t_{1,\alpha/2}\sqrt{2\mathsf{E}[V_{B,2}(n)]/n\pi}.$$

For this special example, the above results allow us to compare analytically the weighted area confidence interval based on $w_1(t)$ and the batch means confidence interval based on two batches—both of these confidence interval methods use one degree of freedom. The case of positive MA(1) parameter θ is not interesting since very small n yields coverages close to 90%. However, $\theta < 0$ is a different story; in this case, the MA(1) has negative serial correlation. Table 1 gives exact results when $\theta = -0.9$ and $1 - \alpha = 0.90$.

We see that the coverages for batch means slowly decrease to the nominal value 90%. The weighted area coverages dip below 90% but rebound to come within 0.01 of $1 - \alpha$ a bit *before* the batch means method. This peculiar "dip" is due to the weighted area variance estimator underestimating σ_n^2. The expected half-lengths for the weighted area confidence intervals are usually much smaller than those for the batch means method.

6 Empirical Performance Results

We now give empirical results comparing the orthonormally weighted area method with the batch means method in terms of confidence interval performance. We carried out a

n	CVG_B	CVG_W	$\mathsf{E}[H_B]$	$\mathsf{E}[H_W]$
4	0.942	0.956	2.937	3.914
8	0.941	0.932	1.474	1.273
16	0.940	0.907	0.742	0.472
32	0.939	0.878	0.377	0.188
64	0.936	0.859	0.193	0.087
128	0.932	0.861	0.102	0.049
256	0.926	0.874	0.056	0.032
512	0.919	0.885	0.032	0.022
1024	0.912	0.892	0.019	0.016
2048	0.907	0.896	0.013	0.011
4096	0.904	0.898	0.008	0.008
8192	0.902	0.899	0.006	0.006

Table 1: Exact Results for the MA(1) Process with $\theta = -0.9$ from §5.4

n	$C_{B,2}$	$C_{B,4}$	$C_{W,1}$	$C_{W,3}$
16	0.612	0.339	0.605	0.307
32	0.682	0.434	0.676	0.393
64	0.741	0.538	0.734	0.497
128	0.789	0.627	0.784	0.596
256	0.820	0.703	0.815	0.682
512	0.853	0.764	0.846	0.748
1024	0.867	0.804	0.859	0.791
2048	0.882	0.841	0.882	0.834
4096	0.888	0.865	0.887	0.860
8192	0.891	0.879	0.889	0.877

Table 2: Estimated Coverages for the M/M/1 Example 9

n	$C_{B,2}$	$C_{B,4}$	$C_{W,1}$	$C_{W,3}$
16	5.95	1.84	5.82	1.67
32	7.25	2.30	7.13	2.04
64	8.17	2.69	8.11	2.41
128	8.55	2.95	8.60	2.67
256	8.15	2.96	8.27	2.76
512	6.98	2.69	7.04	2.59
1024	5.62	2.25	5.63	2.22
2048	4.38	1.79	4.38	1.79
4096	3.23	1.36	3.24	1.37
8192	2.38	1.00	2.36	1.01

Table 3: Estimated Expected Half-Lengths for the M/M/1 Example 9

battery of Monte Carlo experiments involving simple ARMA, queueing, and inventory processes; the next empirical example can be viewed as representative, and shows that our new method is at least competitive with batch means when d is small.

Example 9 Consider the waiting time process of customers in a stationary M/M/1 queue with arrival rate 0.8 and service rate 1.0. We compare the performance of four confidence interval estimators: batch means with 2 and 4 batches from (2) ($C_{B,2}$ and $C_{B,4}$, respectively), and the orthonormally weighted area estimators from Example 5 using weighting functions $w_1, \ldots, w_d$ with $d = 1$ and 3 ($C_{W,1}$ and $C_{W,3}$, respectively). Note that $C_{B,2}$ and $C_{W,1}$ are based on 1 degree of freedom, while $C_{B,4}$ and $C_{W,3}$ are based on 3 degrees of freedom. Tables 2 and 3 give estimated coverages and expected half-lengths of the four confidence interval estimators. These results are based on 20000 independent replications of the appropriate simulation experiments. Coverage of 90% is desired; the standard error of the coverage estimates is about 0.002.

Table 2 shows that $C_{B,2}$ and $C_{W,1}$ have better coverage for small sample sizes, but all methods approach the desired coverage as n increases. Table 3 shows that the expected values of the half-lengths for $C_{B,4}$ and $C_{W,3}$ are much smaller than those from $C_{B,2}$ and $C_{W,1}$. Similar results hold for the variances of the half-lengths.

We see from Example 9 that for fixed n, coverage sometimes deteriorates as the degrees of freedom increases (whether we use batch means or the orthonormally weighted area method); cf. [18]. Clearly, the experimenter would like to maximize the degrees of freedom subject to proper coverage since the expected value and variance of the half-lengths tend to decrease with increasing degrees of freedom (see [19] and Table 3). The next example shows that it is possible for the orthonormally weighted confidence intervals to work with "large" degrees of freedom, even when only "limited" data is available.

Example 10 Suppose that we take $n = 128$ i.i.d. $\mathsf{Nor}(0,1)$ observations, and that we are interested in forming orthonormally weighted area 90% confidence intervals for μ based on $d = 30$ degrees of freedom. (Of course, batch means with one observation per batch and $n - 1 = 127$ degrees of freedom would have produced exactly 90% coverage for this example's scenario.) We shall use Example 5's weighting functions, which satisfy Conditions O and U. We conducted Monte Carlo simulation runs of the process to obtain 10000 independent realizations of the confidence intervals; the estimated coverage was 0.9099. Coverage inflated unacceptably for $d > 30$ at the expense of increased half-lengths.

How can we explain such behavior? By Equation (15) with $\theta = 0$, we have

$$\mathsf{E}[V_W(n)] \;=\; \frac{1}{d}\sum_{i=1}^{d}\mathsf{E}[A_i^2(n)] \;=\; \frac{1}{dn^3}\sum_{i=1}^{d}\sum_{k=1}^{n} g_i^2(k). \tag{16}$$

Similarly, (14) with $\theta = 0$ and Patel and Read [16, p. 309] imply

$$\begin{aligned}\mathsf{Var}(V_W(n)) \;&=\; \frac{1}{d^2}\sum_{i=1}^{d}\sum_{j=1}^{d}\mathsf{Cov}(A_i^2(n), A_j^2(n)) \\ &=\; \frac{2}{d^2}\sum_{i=1}^{d}\sum_{j=1}^{d}\mathsf{Cov}^2(A_i(n), A_j(n)) \;=\; \frac{2}{d^2n^6}\sum_{i=1}^{d}\sum_{j=1}^{d}\Big(\sum_{k=1}^{n} g_i(k)g_j(k)\Big)^2.\end{aligned} \tag{17}$$

For the i.i.d. $\mathsf{Nor}(0,1)$ example at hand, Equations (16) and (17) give $\mathsf{E}[V_W(n)] \doteq 1.0678$ and $\mathsf{Var}(V_W(n)) \doteq 0.07627$, respectively; so $V_W(n)$ is a bit biased for estimating the variance parameter $\sigma^2 = \lim_{n\to\infty} n\mathsf{Var}(\bar{Y}_n) = \mathsf{Var}(Y_i) = 1$, while $\mathsf{Var}(V_W(n))$ is a little bigger than the anticipated $\mathsf{Var}(\sigma^2\chi_d^2/d) = 2/d \doteq 0.06667$. Using our values for $\mathsf{E}[V_W(n)]$ and $\mathsf{Var}(V_W(n))$ into the p.d.f. $f_X(x)$ and integrating the right-hand side of (10) numerically via Mathematica (see [24]), we obtain an approximate coverage of $2F_{\widetilde{T}}(t_{30,0.05}) - 1 = 0.9094$ (which is very close to the Monte Carlo coverage result).

7 Conclusions

This article studied orthonormally weighted area variance estimators. These variance estimators are first-order unbiased, and thus compare favorably to the analogous batch means variance estimator. Our orthonormal estimators can also be used to construct asymptotically valid confidence intervals for the mean of a stationary stochastic process. The confidence intervals are based on more degrees of freedom than their competitors;

therefore, the new intervals tend to have smaller expected half-lengths and smaller half-length variances.

There are a number of ways to improve or otherwise use the orthonormal estimators.

1. In a sense, the orthonormally weighted area estimator is at the opposite end of the spectrum from the estimators in [13, 20]; those articles take one estimator ($d = 1$) with multiple batches, while we have one batch with several estimators ($d > 1$). The "best" procedure may lie between the two ends of the spectrum. If multiple batches are employed, the weighted area estimators are asymptotically independent of the classical batch means estimator (as the batch size grows); see Schruben [20] and our Theorem 1. Hence, the degrees of freedom from all of the estimators can be added—yielding better confidence intervals.

2. We can apply Meketon and Schmeiser's [15] methodology to our orthonormal area estimators. In [15], the authors divide n observations into $n - m + 1$ *overlapping* batches, each of size m. We would expect that the bias of the resulting overlapped orthonormal area estimators would be the same as that of the regular orthonormal area estimator, but that the overlapped estimators' variances would be somewhat *smaller*.

3. We would hope to apply our orthonormalization methodology to other variance estimators, perhaps the Cramér-von Mises estimators discussed in [9].

4. We can obtain confidence intervals for σ^2, a problem that is interesting in its own right. For example, starting with $V_W(n) \approx \sigma^2 \chi^2_d / d$, it is easy to show that an approximate $100(1 - \alpha)\%$ confidence interval for σ^2 is given by

$$\frac{d\, V_W(n)}{\chi^2_{d,\alpha/2}} \;<\; \sigma^2 \;<\; \frac{d\, V_W(n)}{\chi^2_{d,1-\alpha/2}},$$

   where $\chi^2_{d,1-\alpha/2}$ and $\chi^2_{d,\alpha/2}$ are appropriate quantiles from the χ^2_d distribution.

5. It is straightforward to use the methods of this article to derive first-order unbiased, orthonormally weighted area estimators for $\sigma_n^2 = n\mathsf{Var}(\bar{Y}_n)$ (instead of σ^2).

6. Song [21] and Song and Schmeiser [22] express several estimators (including the weighted area estimator) as quadratic forms and graph the coefficients of the quadratic forms in three dimensions (cf. Grenander and Rosenblatt [14]). These graphs yield considerable insight into the behavior of the orthonormally weighted area estimators.

8 Appendix

This appendix contains the proofs of Theorems 4 and 5. Throughout, n is the sample size, i and j are reserved for subscripts of the weighting functions, and k, ℓ, and m are integers. The following notation will be used in the appendix; the subscript D denotes the discrete analogs of previously defined quantities such as $W_i(t)$.

$$
\begin{aligned}
Z_m &\equiv \sum_{k=1}^{m} Y_k, \quad m = 1, \ldots, n, \\
V(m) &\equiv \mathsf{Var}(Z_m), \quad m = 1, \ldots, n, \\
W_{D,i}(\frac{m}{n}) &\equiv \frac{1}{n} \sum_{k=1}^{m} w_i(\frac{k}{n}), \quad m = 1, \ldots, n; \quad i = 1, \ldots, d, \\
W_{D,i} &\equiv W_{D,i}(1), \quad i = 1, \ldots, d, \\
\bar{W}_{D,i} &\equiv \frac{1}{n} \sum_{k=1}^{n-1} W_{D,i}(\frac{k}{n}), \quad i = 1, \ldots, d, \quad \text{and} \\
\beta &\equiv \max_{1 \le i \le d} \sup_{0 \le t \le 1} |w_i(t)|.
\end{aligned}
$$

The next four lemmas provide "big-Oh" results that will be used in the proof of Theorem 4.

Lemma 1 Under Assumptions FO,

$$
\begin{aligned}
\frac{V(n)}{n} &= \sigma^2 + \frac{\gamma}{n} + \frac{2}{n} \sum_{k=n}^{\infty} (k-n) R_k &\quad (18) \\
&= \sigma^2 + \frac{\gamma}{n} + O(\frac{1}{n^2}). &\quad (19)
\end{aligned}
$$

Proof Equation (18) follows from the definition of γ, Assumption FO.4, and the fact that $V(n) = nR_0 + 2\sum_{k=1}^{n-1}(n-k)R_k$; after noting that $n|\sum_{k=n}^{\infty}(k-n)R_k| \le \sum_{k=1}^{\infty} k^2 |R_k|$, Equation (19) follows from Assumption FO.3. □

Lemma 2 Under Assumptions FO, if $k \le \ell$, then

$$
\mathsf{Cov}(Z_\ell, Z_k) = k\sigma^2 + \frac{\gamma}{2} + \sum_{m=\ell}^{\infty}(m-\ell)R_m + \sum_{m=k}^{\infty}(m-k)R_m - \sum_{m=\ell-k}^{\infty}(m-(\ell-k))R_m.
$$

Proof Follows from the fact that $\mathsf{Cov}(Z_\ell, Z_k) = [V(\ell) + V(k) - V(\ell-k)]/2$ (cf. [10]) and Equation (18). □

Lemma 3 Under Assumptions FO,

$$\frac{1}{n^2}\sum_{k=1}^{n} w_i(\frac{k}{n})\mathsf{Cov}(Z_n, Z_k) \;=\; \sigma^2(W_{D,i} - \bar{W}_{D,i}) + \frac{\gamma}{2n}W_{D,i} + O(\frac{1}{n^2}).$$

Proof By Lemma 2 and summation by parts, we have

$$\begin{aligned}\frac{1}{n^2}\sum_{k=1}^{n} w_i(\frac{k}{n})\mathsf{Cov}(Z_n, Z_k) &= \frac{1}{n^2}\sum_{k=1}^{n} w_i(\frac{k}{n})\left(k\sigma^2 + \frac{\gamma}{2}\right) + \frac{P_i(n)}{n^2} \\ &= \sigma^2(W_{D,i} - \bar{W}_{D,i}) + \frac{\gamma}{2n}W_{D,i} + \frac{P_i(n)}{n^2},\end{aligned}$$

where

$$P_i(n) \;\equiv\; \sum_{k=1}^{n} w_i(\frac{k}{n})\left[\sum_{m=n}^{\infty}(m-n)R_m + \sum_{m=k}^{\infty}(m-k)R_m - \sum_{m=n-k}^{\infty}(m-(n-k))R_m\right].$$

To finish, we establish that $P_i(n)$ is of order $O(1)$.

$$\begin{aligned}|P_i(n)| &\le \beta\sum_{k=1}^{n}\left[\sum_{m=n}^{\infty} m|R_m| + \sum_{m=k}^{\infty} m|R_m| + \sum_{m=n-k}^{\infty} m|R_m|\right] \\ &= \beta\left[\sum_{m=n}^{\infty} nm|R_m| + \sum_{m=1}^{\infty} m(m\wedge n)|R_m| + \sum_{m=1}^{\infty} m((m+1)\wedge n)|R_m|\right] \\ &\le 4\beta\sum_{m=1}^{\infty} m^2|R_m| \;=\; O(1) \quad \text{(by Assumption FO.3).} \quad \square\end{aligned}$$

Lemma 4 Under Assumptions FO,

$$\begin{aligned}&\frac{1}{n^3}\sum_{k=1}^{n}\sum_{\ell=1}^{n} w_i(\frac{k}{n})w_j(\frac{\ell}{n})\mathsf{Cov}(Z_k, Z_\ell) \\ &\quad= \sigma^2[(W_{D,i} - \bar{W}_{D,i})W_{D,j} + W_{D,i}(W_{D,j} - \bar{W}_{D,j})] \\ &\qquad -\frac{\sigma^2}{n^3}\sum_{k=1}^{n}\sum_{\ell=1}^{n} w_i(\frac{k}{n})w_j(\frac{\ell}{n})(k\vee\ell) + \frac{\gamma}{2n}W_{D,i}W_{D,j} + O(\frac{1}{n^2}).\end{aligned}$$

Proof By Lemma 2 and summation by parts, we have

$$\frac{1}{n^3}\sum_{k=1}^{n}\sum_{\ell=1}^{n} w_i(\frac{k}{n})w_j(\frac{\ell}{n})\mathsf{Cov}(Z_k, Z_\ell)$$

$$
\begin{aligned}
&= \frac{1}{n^3}\sum_{k=1}^{n}\sum_{\ell=1}^{n} w_i(\frac{k}{n})w_j(\frac{\ell}{n})\left[(k\wedge\ell)\sigma^2+\frac{\gamma}{2}\right]+\frac{P_{ij}(n)}{n^3}\\
&= \frac{\sigma^2}{n^3}\sum_{k=1}^{n}\sum_{\ell=1}^{n} w_i(\frac{k}{n})w_j(\frac{\ell}{n})[k+\ell-(k\vee\ell)]+\frac{\gamma}{2n}W_{D,i}W_{D,j}+\frac{P_{ij}(n)}{n^3}\\
&= \sigma^2[(W_{D,i}-\bar{W}_{D,i})W_{D,j}+W_{D,i}(W_{D,j}-\bar{W}_{D,j})]\\
&\qquad -\frac{\sigma^2}{n^3}\sum_{k=1}^{n}\sum_{\ell=1}^{n} w_i(\frac{k}{n})w_j(\frac{\ell}{n})(k\vee\ell)+\frac{\gamma}{2n}W_{D,i}W_{D,j}+\frac{P_{ij}(n)}{n^3},
\end{aligned}
$$

where

$$
P_{ij}(n) \equiv \sum_{k=1}^{n}\sum_{\ell=1}^{n} w_i(\frac{k}{n})w_j(\frac{\ell}{n})\left[\sum_{m=\ell}^{\infty}(m-\ell)R_m+\sum_{m=k}^{\infty}(m-k)R_m-\sum_{m=|\ell-k|}^{\infty}(m-|\ell-k|)R_m\right].
$$

To finish, we establish that $P_{ij}(n)$ is of order $O(n)$.

$$
\begin{aligned}
|P_{ij}(n)| &\le \beta^2\sum_{k=1}^{n}\sum_{\ell=1}^{n}\left[\sum_{m=\ell}^{\infty} m|R_m|+\sum_{m=k}^{\infty} m|R_m|+\sum_{m=|\ell-k|}^{\infty} m|R_m|\right]\\
&= \beta^2\left[2n\sum_{\ell=1}^{n}\sum_{m=\ell}^{\infty} m|R_m|+\sum_{k=1}^{n}\sum_{\ell=1}^{n}\sum_{m=|\ell-k|}^{\infty} m|R_m|\right]\\
&= \beta^2\left[2n\sum_{m=1}^{\infty} m(m\wedge n)|R_m|+n\sum_{m=0}^{\infty} m|R_m|+2\sum_{\ell=1}^{n-1}(n-\ell)\sum_{m=\ell}^{\infty} m|R_m|\right]\\
&\le 5n\beta^2\sum_{m=1}^{\infty} m^2|R_m| \;=\; O(n) \quad \text{(by Assumption FO.3).} \quad \square
\end{aligned}
$$

We now have the machinery to prove the main result.

Proof of Theorem 4. By Equation (5),

$$
\begin{aligned}
\mathsf{Cov}(A_i(n),A_j(n)) &= \frac{\sigma^2}{n^2}\sum_{k=1}^{n}\sum_{\ell=1}^{n} w_i(\frac{k}{n})w_j(\frac{\ell}{n})\mathsf{Cov}(T_n(\frac{k}{n}),T_n(\frac{\ell}{n}))\\
&= \frac{1}{n^3}\sum_{k=1}^{n}\sum_{\ell=1}^{n} w_i(\frac{k}{n})w_j(\frac{\ell}{n})\left[\frac{k\ell}{n^2}V(n)-\frac{k}{n}\mathsf{Cov}(Z_n,Z_\ell)-\frac{\ell}{n}\mathsf{Cov}(Z_n,Z_k)+\mathsf{Cov}(Z_k,Z_\ell)\right]\\
&\qquad \text{(from Equation (3) and the definition of } Z_m)\\
&= \frac{1}{n}V(n)(W_{D,i}-\bar{W}_{D,i})(W_{D,j}-\bar{W}_{D,j})-\frac{1}{n^2}(W_{D,i}-\bar{W}_{D,i})\sum_{\ell=1}^{n} w_j(\frac{\ell}{n})\mathsf{Cov}(Z_n,Z_\ell)\\
&\qquad -\frac{1}{n^2}(W_{D,j}-\bar{W}_{D,j})\sum_{k=1}^{n} w_i(\frac{k}{n})\mathsf{Cov}(Z_n,Z_k)+\frac{1}{n^3}\sum_{k=1}^{n}\sum_{\ell=1}^{n} w_i(\frac{k}{n})w_j(\frac{\ell}{n})\mathsf{Cov}(Z_k,Z_\ell)
\end{aligned}
$$

(from summation by parts)

$$= \sigma^2 \left[W_{D,i}W_{D,j} - \bar{W}_{D,i}\bar{W}_{D,j} - \frac{1}{n^3}\sum_{k=1}^{n}\sum_{\ell=1}^{n} w_i(\frac{k}{n})w_j(\frac{\ell}{n})(k \vee \ell) \right]$$

$$+\frac{\gamma}{2n}\left[(W_{D,i} - \bar{W}_{D,i})(W_{D,j} - \bar{W}_{D,j}) + \bar{W}_{D,i}\bar{W}_{D,j} \right] + O(1/n^2) \quad (20)$$

(from Lemmas 1, 3, and 4).

At this point, we would like to replace the discrete approximations to integrals with the integrals themselves. To do this, we need three "big-Oh" expressions relating the approximations and the integrals. From the trapezoid rule for integration, we have

$$W_i(\frac{k}{n}) \;=\; W_{D,i}(\frac{k}{n}) + \frac{w_i(0) - w_i(\frac{k}{n})}{2n} + O(1/n^2), \quad (21)$$

which gives us the first of the three expressions needed:

$$W_{D,i} \;=\; W_i + \frac{w_i(1) - w_i(0)}{2n} + O(1/n^2). \quad (22)$$

To get a similar relationship for $\bar{W}_{D,i}$, start off with

$$\bar{W}_i \;=\; \frac{1}{n}\sum_{k=1}^{n} W_i(\frac{k}{n}) - \frac{W_i}{2n} + O(1/n^2),$$

replace $W_i(k/n)$ in the summation with the right-hand side of (21), and collect terms to obtain the second approximation needed:

$$\bar{W}_{D,i} \;=\; \bar{W}_i - \frac{w_i(0)}{2n} + O(1/n^2). \quad (23)$$

The third approximation relates the double summation in (20) with the integral

$$\int_0^1\int_0^1 w_i(s)w_j(t)(s \vee t)\,ds\,dt \;=\; \int_0^1\int_0^t w_i(s)w_j(t)t\,ds\,dt + \int_0^1\int_t^1 w_i(s)w_j(t)s\,ds\,dt. \quad (24)$$

The first term on the right-hand side of Equation (24) gives us

$$\int_0^1\int_0^t w_i(s)w_j(t)t\,ds\,dt \;=\; \int_0^1 tW_i(t)w_j(t)\,dt$$

$$= \frac{1}{n}\sum_{k=1}^{n}\frac{k}{n}W_i(\frac{k}{n})w_j(\frac{k}{n}) - \frac{W_i w_j(1)}{2n} + O(\frac{1}{n^2}) \quad \text{(from the trapezoid rule)}$$

$$= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{\ell=1}^{k}\frac{k}{n}w_i(\frac{\ell}{n})w_j(\frac{k}{n}) + \frac{1}{2n^2}\sum_{k=1}^{n}\frac{k}{n}w_j(\frac{k}{n})(w_i(0) - w_i(\frac{k}{n})) - \frac{W_i w_j(1)}{2n} + O(\frac{1}{n^2})$$

(by (21), the boundedness of w_j, and the definition of $W_{D,i}(k/n)$)

$$= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{\ell=1}^{k}\frac{k}{n}w_i(\frac{\ell}{n})w_j(\frac{k}{n}) + \frac{w_i(0)(W_j - \bar{W}_j) - \int_0^1 s w_i(s)w_j(s)\,ds - W_i w_j(1)}{2n} + O(\frac{1}{n^2})$$

(after applying integral approximations and then integration by parts). (25)

Similar machinations on the second term on the right-hand side of (24) yield

$$\int_0^1 \int_t^1 w_i(s)w_j(t)t \, ds \, dt$$
$$= \frac{1}{n^2} \sum_{k=1}^{n} \sum_{\ell=k+1}^{n} \frac{\ell}{n} w_i(\frac{\ell}{n}) w_j(\frac{k}{n}) + \frac{w_j(0)(W_i - \bar{W}_i) + \int_0^1 s w_i(s) w_j(s)\, ds - w_i(1)W_j}{2n} + O(\frac{1}{n^2}). \quad (26)$$

Substituting (25) and (26) into (24) and rewriting gives us the third "big-Oh" expression:

$$\frac{1}{n^2} \sum_{k=1}^{n} \sum_{\ell=1}^{n} \frac{(k \vee \ell)}{n} w_i(\frac{\ell}{n}) w_j(\frac{k}{n})$$
$$= \int_0^1 \int_0^1 w_i(s)w_j(t)(s \vee t)\, ds\, dt$$
$$- \frac{w_i(0)(W_j - \bar{W}_j) + w_j(0)(W_i - \bar{W}_i) - w_i(1)W_j - w_j(1)W_i}{2n} + O(\frac{1}{n^2}). \quad (27)$$

Now use the three expressions (22), (23), and (27) in (20) to obtain

$$\mathsf{Cov}(A_i(n), A_j(n)) = \sigma^2 \left[W_i W_j - \bar{W}_i \bar{W}_j - \int_0^1 \int_0^1 w_i(s) w_j(t)(s \vee t)\, ds\, dt \right]$$
$$+ \frac{\gamma}{2n} \left[(W_i - \bar{W}_i)(W_j - \bar{W}_j) + \bar{W}_i \bar{W}_j \right] + O(\frac{1}{n^2}). \quad (28)$$

To complete the proof, we use repeated integration by parts to simplify the coefficient of the σ^2 term in (28):

$$W_i W_j - \bar{W}_i \bar{W}_j - \int_0^1 \int_0^1 w_i(s) w_j(t)(s + t - s \wedge t)\, ds\, dt$$
$$= W_i W_j - \bar{W}_i \bar{W}_j - (W_i - \bar{W}_i) W_j - W_i (W_j - \bar{W}_j) + \int_0^1 \int_0^1 w_i(s) w_j(t)(s \wedge t)\, ds\, dt$$
$$= \int_0^1 \int_0^1 w_i(s) w_j(t)(s \wedge t)\, ds\, dt - (W_i - \bar{W}_i)(W_j - \bar{W}_j)$$
$$= \int_0^1 \int_0^1 w_i(s) w_j(t)(s \wedge t - st)\, ds\, dt. \quad \Box$$

Finally, we prove Theorem 5.

Proof of Theorem 5.

$$n^{5/2} \mathsf{Cov}(\bar{Y}_n, A_i(n)) = \sum_{k=1}^{n} w_i(\frac{k}{n}) \mathsf{Cov}(Z_n, \frac{k Z_n}{n} - Z_k)$$

$$
\begin{aligned}
&= V(n)\sum_{k=1}^{n}\frac{k}{n}w_i(\frac{k}{n}) - \sum_{k=1}^{n} w_i(\frac{k}{n})\mathsf{Cov}(Z_n, Z_k) \\
&= n(W_{D,i}/2 - \bar{W}_{D,i})\gamma + O(1) \\
&\quad \text{(by Lemmas 1 and 3 and summation by parts)} \\
&= o(n) \text{ (by (22), (23), and Condition U).}
\end{aligned}
$$

Thus, $\mathsf{Cov}(\bar{Y}_n, A_i(n)) = o(1/n^{3/2})$. Meanwhile, Lemma 1 implies that $n\mathsf{Var}(\bar{Y}_n) = V(n)/n \to \sigma^2$, and Corollary 3 implies that $\mathsf{Var}(A_i(n)) \to \sigma^2$. □

Acknowledgments. We thank Jeff Geronimo, Steve Hackman, Bruce Schmeiser, Lee Schruben, Whey-Ming Tina Song, and Jim Wilson for helpful comments. David Goldsman's work was supported by National Science Foundation Grants DDM-9012020 and DMI-9622269.

References

[1] R. B. Ash and M. F. Gardner (1975). *Topics in Stochastic Processes*, Academic Press, New York.

[2] M. S. Bartlett and P. H. Diananda (1950). "Extensions of Quenouille's Test for Autoregressive Schemes," *Journal of the Royal Statistical Society* **B12**, 108–115.

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

[4] P. Bratley, B. L. Fox, and L. E. Schrage (1987). *A Guide to Simulation*, 2nd Edition, Springer-Verlag, New York.

[5] C. Chien, D. Goldsman, and B. Melamed (1997). "Large-Sample Results for Batch Means," *Management Science* **43**, 1288–1295.

[6] K. Dzhaparidze (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis and Stationary Time Series*, Springer-Verlag, New York.

[7] R. D. Foley and D. Goldsman (1988). "Confidence Intervals with Orthonormally Weighted Standardized Time Series," *Proceedings of the 1988 Winter Simulation Conference* (ed. M. A. Abrams, P. L. Haigh, and J. C. Comfort), 422–424, The Institute of Electrical and Electronics Engineers, Piscataway, NJ.

[8] P. Glynn and D. L. Iglehart (1990). "Simulation Output Analysis Using Standardized Time Series," *Mathematics of Operations Research* **15**, 1–16.

[9] D. Goldsman, K. Kang, and A. F. Seila (1999). "Cramér-von Mises Variance Estimators for Simulations." To appear in *Operations Research*.

[10] D. Goldsman and M. S. Meketon (1986). "A Comparison of Several Variance Estimators," Technical Report J-85-12, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

[11] D. Goldsman, M. S. Meketon, and L. W. Schruben (1990). "Properties of Standardized Time Series Weighted Area Variance Estimators," *Management Science* **36**, 602–612.

[12] D. Goldsman and L. W. Schruben (1984). "Asymptotic Properties of Some Confidence Interval Estimators for Simulation Output," *Management Science* **30**, 1217–1225.

[13] D. Goldsman and L. W. Schruben (1990). "New Confidence Interval Estimators Using Standardized Time Series," *Management Science* **36**, 393–397.

[14] U. Grenander and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York.

[15] M. S. Meketon and B. W. Schmeiser (1984). "Overlapping Batch Means: Something for Nothing?" *Proceedings of the 1984 Winter Simulation Conference* (ed. S. Sheppard, U. W. Pooch, and C. D. Pegden), 227–230, The Institute of Electrical and Electronics Engineers, Piscataway, NJ.

[16] J. K. Patel and C. B. Read (1982). *Handbook of the Normal Distribution*, Marcel Dekker, New York.

[17] V. K. Rohatgi (1976). *An Introduction to Probability Theory and Mathematical Statistics*, John Wiley and Sons, New York.

[18] R. G. Sargent, K. Kang, and D. Goldsman (1992). "An Investigation of Finite-Sample Behavior of Confidence Interval Estimators," *Operations Research* **40**, 898–913.

[19] B. W. Schmeiser (1982). "Batch Size Effects in the Analysis of Simulation Output," *Operations Research* **30**, 556–568.

[20] L. W. Schruben (1983). "Confidence Interval Estimation Using Standardized Time Series," *Operations Research* **31**, 1090–1108.

[21] W.-M. T. Song (1988). *Estimators of the Variance of the Sample Mean: Quadratic Forms, Optimal Batch Sizes, and Linear Combinations*, Ph.D. Dissertation, School of Industrial Engineering, Purdue University, West Lafayette, Indiana.

[22] W.-M. T. Song and B. W. Schmeiser (1993). "Variance of the Sample Mean: Properties and Graphs of Quadratic-Form Estimators," *Operations Research* **41**, 501–517.

[23] W.-M. T. Song and B. W. Schmeiser (1995). "Optimal Mean-Squared-Error Batch Sizes," *Management Science* **41**, 110–123.

[24] S. Wolfram (1991). *Mathematica—A System for Doing Mathematics by Computer*, 2nd Edition, Addison-Wesley, Reading, Massachusetts.