2006

Combining Standardized Time Series Area and Cramér–von Mises Variance Estimators

Goldsman, David; Kang, Keebom; Kim, Seong-Hee; Seila, Andrew F.; Tokol, Gamze

Naval Research Logistics, Vol. 54 (2007)
http://hdl.handle.net/10945/46102

Combining Standardized Time Series Area and Cramér–von Mises Variance Estimators

David Goldsman,[1] Keebom Kang,[2] Seong-Hee Kim,[1] Andrew F. Seila,[3] Gamze Tokol[4]

[1] *H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia*

[2] *Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, California*

[3] *Terry College of Business, University of Georgia, Athens, Georgia*

[4] *Decision Analytics, Atlanta, Georgia*

Received 10 March 2006; revised 1 November 2006; accepted 4 November 2006
DOI 10.1002/nav.20214
Published online 22 January 2007 in Wiley InterScience (www.interscience.wiley.com).

Abstract: We propose three related estimators for the variance parameter arising from a steady-state simulation process. All are based on combinations of standardized-time-series area and Cramér–von Mises (CvM) estimators. The first is a straightforward linear combination of the area and CvM estimators; the second resembles a Durbin–Watson statistic; and the third is related to a jackknifed version of the first. The main derivations yield analytical expressions for the bias and variance of the new estimators. These results show that the new estimators often perform better than the pure area, pure CvM, and benchmark nonoverlapping and overlapping batch means estimators, especially in terms of variance and mean squared error. We also give exact and Monte Carlo examples illustrating our findings. © 2007 Wiley Periodicals, Inc. Naval Research Logistics 54: 384–396, 2007

Keywords: simulation; stationary process; variance estimation; standardized time series; area estimator; Cramér–von Mises estimator; Durbin–Watson estimator; batch means estimator

1. INTRODUCTION

Consider an experiment in which we want to estimate the mean μ of a steady-state simulation process, $Y_1, Y_2, \ldots, Y_n$, e.g., the mean time-in-system for parts produced on a continuously running manufacturing line. The usual estimator for μ is the sample mean $\bar{Y}_n$, and it is common to further provide an accompanying estimate of $\mathsf{Var}(\bar{Y}_n)$, or, almost equivalently, the *variance parameter*, $\sigma^2 \equiv \lim_{n\to\infty} n \,\mathsf{Var}(\bar{Y}_n)$.

There are a number of estimators for σ^2, as described in standard references such as [17]. One class of estimators is based on Schruben's standardized time series (STS) methodology [21]. Two specific examples of STS estimators with good properties are the weighted area and weighted Cramér–von Mises (CvM) estimators, introduced in [15] and [13], respectively. Although many estimators for σ^2 are available in the literature, including several "first-order unbiased" estimators, they can exhibit high variability as well as high mean squared error. Therefore, there have been efforts to develop less variable estimators, while still maintaining low bias. One way of doing this is by carefully combining multiple estimators, effectively re-using data to garner more information about σ^2. This article investigates new STS estimators to achieve these goals—estimators that combine the area and CvM approaches.

The paper is organized as follows. Section 2 presents necessary background material. Then in Section 3, we give three ways to combine the area and CvM estimators: the first is a simple linear combination; the second resembles a Durbin–Watson statistic; and the third is related to a jackknifed version of the first. We find that the new estimators sometimes have bias that is competitive with and variance that is often quite a bit better than the area and CvM estimators alone (among others). Section 4 gives exact and Monte Carlo examples illustrating these findings. Section 5 shows that batched versions of the new estimators compare favorably to the benchmark methods of nonoverlapping and overlapping batch means. Section 6 wraps up the discussion.

Correspondence to: D. Goldsman (sman@gatech.edu)
This article contains Supplementary Material available at http://www.interscience.wiley.com/jpages/0894-069X/suppmat

2. BACKGROUND

This section reviews results that will be needed in the remainder of the article. The subsections discuss, respectively, the standardized time series of a stochastic process, the STS weighted area estimator for σ^2, and the STS weighted CvM variance estimator.

2.1. Standardized Time Series

Schruben [21] defines the *standardized time series* of a stationary stochastic process $Y_1, Y_2, \ldots, Y_n$ (e.g., a simulation in steady state) as

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

where $\bar{Y}_j \equiv \sum_{k=1}^{j} Y_k / j$, $j = 1, 2, \ldots, n$, and $\lfloor \cdot \rfloor$ is the greatest integer function. We henceforth assume that $Y_1, Y_2, \ldots, Y_n$ satisfies the following mild "functional central limit theorem" condition:

Assumption FCLT. There exist μ and positive σ such that as $n \to \infty$, $X_n \Rightarrow \sigma \mathcal{W}$, where $\mathcal{W}$ is a standard Brownian motion process, " $\Rightarrow$ " denotes weak convergence as $n \to \infty$ (see, e.g., [5]), and

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

(See [10] for various sets of sufficient conditions for Assumption FCLT to hold.) Then, as in [9], [10], or [21], it can be shown that $T_n \Rightarrow \mathcal{B}$, where $\mathcal{B}$ is a standard Brownian bridge process on [0, 1]. It is well known that all finite-dimensional joint distributions of $\mathcal{B}$ are normal with $\mathsf{E}[\mathcal{B}(t)] = 0$ and $\mathsf{Cov}(\mathcal{B}(s), \mathcal{B}(t)) = \min(s, t) - st$, $0 < s, t < 1$. In addition, it is often convenient to express a Brownian bridge in terms of its underlying Brownian motion, $\mathcal{B}(t) = \mathcal{W}(t) - t\mathcal{W}(1)$.

2.2. The Weighted Area Estimator

As in [15], [16], and [21], we begin by defining random variables (r.v.'s) corresponding to the weighted area under the standardized time series and its limiting functional:

$$S(f; n) \equiv \frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k}{n}\right) \sigma T_n\left(\frac{k}{n}\right)$$

and

$$S(f) \equiv \int_0^1 f(t) \sigma \mathcal{B}(t)\, dt,$$

where the weighting function $f(t)$ has a continuous second derivative on [0, 1] and is chosen to satisfy $\mathsf{Var}(S(f)) = \sigma^2$, in which case $S(f) \sim \mathsf{Nor}(0, \sigma^2)$. In addition, let $A(f; n) \equiv S^2(f; n)$ and $A(f) \equiv S^2(f)$. Then under mild conditions, the continuous mapping theorem (CMT) ([5], Theorem 5.1) implies $A(f; n) \Rightarrow A(f) \sim \sigma^2 \chi_1^2$, and we call $A(f; n)$ the *weighted area* estimator for σ^2.

Theorem 1 gives expressions for the expected value and variance of $A(f; n)$. Before proceeding, we define the covariance function $R_k \equiv \mathsf{Cov}(Y_1, Y_{1+k})$, $k = 0, \pm 1, \pm 2, \ldots$, and the constant $\gamma \equiv -2 \sum_{k=1}^{\infty} k R_k$ (cf. [22]). In addition, let $F(t) \equiv \int_0^t f(s)\, ds$, $F \equiv F(1)$, $\bar{F}(t) \equiv \int_0^t F(s)\, ds$, and $\bar{F} \equiv \bar{F}(1)$. Finally, the notation $p(n) = o(q(n))$ means that $p(n)/q(n) \to 0$ as $n \to \infty$.

THEOREM 1 (see [9] and [15]): Suppose $Y_1, Y_2, \ldots, Y_n$ is a stationary process for which Assumption FCLT holds, $\sum_{k=1}^{\infty} k^2 |R_k| < \infty$, and $\sum_{k=-\infty}^{\infty} R_k = \sigma^2 > 0$. Then

$$\mathsf{E}[A(f; n)] = \sigma^2 + \frac{[(F - \bar{F})^2 + \bar{F}^2]\gamma}{2n} + o(1/n). \qquad (1)$$

If we also assume uniform integrability (see [5] for a definition and sufficient conditions) of $A^2(f; n)$, then as $n \to \infty$,

$$\mathsf{Var}(A(f; n)) \to \mathsf{Var}(A(f)) = 2\sigma^4. \qquad (2)$$

EXAMPLE 1: Application of Theorem 1 to Schruben's [21] original area estimator with constant weighting function $f_0(t) \equiv \sqrt{12}$, for all $t \in [0, 1]$, yields $\mathsf{E}[A(f_0; n)] = \sigma^2 + 3\gamma/n + o(1/n)$. It is possible to choose weights for which $F = \bar{F} = 0$; for such selections, the resulting estimator is *first-order unbiased* for σ^2, i.e., its bias is $o(1/n)$. Examples of weighting functions yielding first-order unbiased estimators are $f_2(t) \equiv \sqrt{840}(3t^2 - 3t + 1/2)$ and $f_{\cos,i}(t) = \sqrt{8}\pi i \cos(2\pi i t)$, $i = 1, 2, \ldots$ (see [9], [15], and [16]). □

2.3. The Weighted Cramér–von Mises Estimator

Now we define the area under the square of the STS and its limiting functional as

$$C(g; n) \equiv \frac{1}{n} \sum_{k=1}^{n} g\left(\frac{k}{n}\right) \sigma^2 T_n^2\left(\frac{k}{n}\right)$$

and

$$C(g) \equiv \int_0^1 g(t) \sigma^2 \mathcal{B}^2(t)\, dt,$$

respectively, where $g(t)$ is a weighting function normalized so that $\mathsf{E}[C(g)] = \sigma^2 \int_0^1 g(t) t(1 - t)\, dt = \sigma^2$ and $g''(t)$ is continuous on [0, 1]. Under mild assumptions, the

CMT implies that $C(g;n) \Rightarrow C(g)$, and we call $C(g;n)$ the *weighted Cramér–von Mises* estimator for σ^2.

Theorem 2 gives results on the expected value and variance of the weighted CvM estimator. Here, we define the notation $G \equiv \int_0^1 g(t)\,dt$.

THEOREM 2 (see [13]): Under the conditions of Theorem 1,

$$\mathsf{E}[C(g;n)] = \sigma^2 + \frac{\gamma}{n}(G-1) + o(1/n). \quad (3)$$

If we also assume uniform integrability of $C^2(g;n)$, then as $n \to \infty$,

$$\mathsf{Var}(C(g;n)) \to \mathsf{Var}(C(g)) = 4\sigma^4 \int_0^1 g(t)(1-t)^2 \int_0^t g(s)s^2\,ds\,dt. \quad (4)$$

EXAMPLE 2: Theorem 2 implies that the CvM estimator with constant weighting function $g_0(t) \equiv 6$ has $\mathsf{E}[C(g_0;n)] = \sigma^2 + 5\gamma/n + o(1/n)$. If one chooses weights having $G = 1$ (in addition to the other constraints on the weights), the theorem implies that $C(g;n)$ has bias $o(1/n)$; two examples from [13] are $g_2^\star(t) \equiv -24 + 150t - 150t^2$ and

$$g_4^\star(t) \equiv \frac{-1310}{21} + \frac{19270t}{21} - \frac{25230t^2}{7} + \frac{16120t^3}{3} - \frac{8060t^4}{3}.$$

□

EXAMPLE 3: Theorem 2 gives $\mathsf{Var}(C(g_0)) = 4\sigma^4/5$, $\mathsf{Var}(C(g_2^\star)) = 121\sigma^4/70$, and $\mathsf{Var}(C(g_4^\star)) = 1.042\sigma^4$. Although $\mathsf{Var}(C(g_2^\star))$ and $\mathsf{Var}(C(g_4^\star))$ are greater than $\mathsf{Var}(C(g_0))$, the estimators $C(g_2^\star;n)$ and $C(g_4^\star;n)$ are first-order unbiased for σ^2, while $C(g_0;n)$ is not. □

3. COMBINING THE AREA AND CvM ESTIMATORS

This section shows how one can combine the area and CvM estimators for σ^2 in such a way that the resulting estimator has reasonable bias and lower variance than either of its individual constituents. We shall take advantage of the fact that the covariance between the limiting area and CvM functionals, $\mathsf{Cov}(A(f), C(g))$, can be calculated exactly, a task that is carried out in Section 3.1. That result is used in Section 3.2 to propose three new variance estimators: a straightforward linear combination of the weighted area and CvM estimators, a Durbin–Watson-like estimator, and a jackknifed version of the first. Section 4 presents extended examples.

3.1. Covariance of $A(f)$ and $C(g)$

Our main covariance result, stated next, is proven in the Appendix.

LEMMA 1: If $f(t)$ and $g(t)$ are area and CvM weighting functions, respectively, then

$$\mathsf{Cov}(A(f), C(g)) = 2\sigma^4 \int_0^1 g(t)(t\bar{F} - \bar{F}(t))^2 dt.$$

EXAMPLE 4: Lemma 1 allows us to make the following calculations on the covariance and correlation of $A(f)$ and $C(g)$.

f, g	$\mathsf{Cov}(A(f), C(g))/\sigma^4$	$\mathsf{Corr}(A(f), C(g))$
f_0, g_0	$6/5$	0.949
$f_2, g_2^\star$	$37/22$	0.905
$f_{\cos,1}, g_2^\star$	$3(4\pi^2 + 375)/8\pi^4$	0.858

Of course, since the correlations are so close to unity, there may not be that much additional "information" to be gained by forming a simple linear combination of the particular area and CvM estimators under study; see Example 6. □

3.2. The New Estimators

In this subsection, we discuss some ways that we might be able to take advantage of the known correlation between $A(f)$ and $C(g)$. We present a motivational example and then two more-general estimators.

3.2.1. Motivational Example

EXAMPLE 5: Consider the linear combination $D_0(n) \equiv 2C(g_0;n) - A(f_0;n)$. By Examples 1 and 2, we have

$$\mathsf{E}[D_0(n)] = \sigma^2 + \frac{7\gamma}{n} + o\left(\frac{1}{n}\right).$$

Theorem 1 and Examples 3 and 4 imply

$$\mathsf{Var}(D_0(n)) \to 2\sigma^4/5.$$

Thus, $D_0(n)$ has comparatively low asymptotic variance but very high small-sample bias. □

3.2.2. Area + CvM Estimator

Example 5 suggests a more-general estimator,

$$D_\alpha(f,g;n) \equiv \alpha A(f;n) + (1-\alpha)C(g;n).$$

By the CMT,

$$D_\alpha(f,g;n) \Rightarrow D_\alpha(f,g) \equiv \alpha A(f) + (1-\alpha)C(g).$$

Theorems 1 and 2 immediately imply the following result.

THEOREM 3: Under the conditions of Theorem 1,

$$\mathsf{E}[D_\alpha(f,g;n)] = \sigma^2 + \frac{\gamma}{n}\left\{\frac{\alpha[(F-\bar{F})^2+\bar{F}^2]}{2} + (1-\alpha)(G-1)\right\} + o(1/n) \to \sigma^2.$$

Further, assuming uniform integrability of $D_\alpha^2(f,g;n)$, we have, as $n \to \infty$,

$$\begin{aligned}\mathsf{Var}(D_\alpha(f,g;n)) &\to \mathsf{Var}(D_\alpha(f,g)) \\ &= \alpha^2\mathsf{Var}(A(f)) + (1-\alpha)^2\mathsf{Var}(C(g)) \\ &\quad + 2\alpha(1-\alpha)\mathsf{Cov}(A(f),C(g)). \quad (5)\end{aligned}$$

In Eq. (5), note that $\mathsf{Var}(A(f)) = 2\sigma^4$, Theorem 2 gives $\mathsf{Var}(C(g))$, and Lemma 1 gives $\mathsf{Cov}(A(f),C(g))$. To find the value of α that minimizes $\mathsf{Var}(D_\alpha(f,g))$, we solve $\frac{d}{d\alpha}\mathsf{Var}(D_\alpha(f,g)) = 0$ for α, the result of which is

$$\hat{\alpha} = \frac{\mathsf{Var}(C(g)) - \mathsf{Cov}(A(f),C(g))}{\mathsf{Var}(A(f)) + \mathsf{Var}(C(g)) - 2\mathsf{Cov}(A(f),C(g))}. \quad (6)$$

Substituting $\hat{\alpha}$ from Eq. (6) into (5), we obtain the asymptotically optimal variance,

$$\mathsf{Var}(D_{\hat{\alpha}}(f,g)) = \frac{\mathsf{Var}(A(f))\cdot\mathsf{Var}(C(g)) - \mathsf{Cov}^2(A(f),C(g))}{\mathsf{Var}(A(f)) + \mathsf{Var}(C(g)) - 2\mathsf{Cov}(A(f),C(g))}, \quad (7)$$

where, for all of the examples under study, the denominator is >0.

EXAMPLE 6: Now we use the results of Example 4 and Eqs. (6) and (7) to obtain some variance-optimal linear combinations of $A(f)$ and $C(g)$.

f,g	$\hat{\alpha}$	$\mathsf{Var}(D_{\hat{\alpha}}(f,g))/\sigma^4$
f_0, g_0	-1	2/5
$f_2, g_2^\star$	0.1281	1.723
$f_{\cos,1}, g_2^\star$	0.2474	1.696

We see that the optimal choice of $\hat{\alpha} = -1$ for the linear combination of $A(f_0)$ and $C(g_0)$ is precisely that which we used for $D_0(n)$ in Example 5. Although $D_0(n)$'s asymptotic variance of $2\sigma^4/5$ is much lower than those of $D_{\hat{\alpha}}(f_2,g_2^\star;n)$ and $D_{\hat{\alpha}}(f_{\cos,1},g_2^\star;n)$, we recall from Example 5 that $D_0(n)$ has quite a bit of first-order bias, while Examples 1 and 2 imply that $D_{\hat{\alpha}}(f_2,g_2^\star;n)$ and $D_{\hat{\alpha}}(f_{\cos,1},g_2^\star;n)$ are first-order unbiased. On the other hand, suppose we compare $D_{\hat{\alpha}}(f_2,g_2^\star;n)$ and $D_{\hat{\alpha}}(f_{\cos,1},g_2^\star;n)$ to their component first-order unbiased CvM estimator $C(g_2^\star;n)$, which, by Example 3, has an asymptotic variance of about $1.729\sigma^4$. Then we notice that there is not really that much to be gained by using either choice of $D_{\hat{\alpha}}(f,g_2^\star;n)$ instead of $C(g_2^\star;n)$—a finding that makes sense in light of our remark at the end of Example 4 concerning the high correlation between $A(f)$ and $C(g)$ for f_2, $f_{\cos,1}$, and $g_2^\star$. □

3.2.3. *Durbin–Watson Estimator*

As established previously, we have $A(f_0;n) \Rightarrow A(f_0) = 12\sigma^2\bar{\mathcal{B}}^2$, where $\bar{\mathcal{B}} \equiv \int_0^1 \mathcal{B}(t)\,dt$ is the area under a Brownian bridge, and $C(g_0;n) \Rightarrow C(g_0) = 6\sigma^2\int_0^1 \mathcal{B}^2(t)\,dt$. Thus, we see from Example 5 that

$$\begin{aligned}D_0(n) \Rightarrow 2C(g_0) - A(f_0) &= 12\sigma^2\left(\int_0^1 \mathcal{B}^2(t)\,dt - \bar{\mathcal{B}}^2\right) \\ &= 12\sigma^2\int_0^1 (\mathcal{B}(t) - \bar{\mathcal{B}})^2\,dt,\end{aligned}$$

which bears a striking resemblance to a Durbin–Watson functional (cf. [8] and [24]).

In fact, we can generalize $D_0(n)$ by considering the estimator

$$D(h;n) \equiv \frac{1}{n}\sum_{k=1}^{n} h\left(\frac{k}{n}\right)\left(\sigma T_n\left(\frac{k}{n}\right) - \frac{S(f_0;n)}{\sqrt{12}}\right)^2$$

and the associated limiting functional

$$D(h) \equiv \sigma^2\int_0^1 h(t)(\mathcal{B}(t) - \bar{\mathcal{B}})^2\,dt,$$

where $h(t)$ is normalized so that $\mathsf{E}[D(h)] = \sigma^2$, and $h''(t)$ is continuous on [0, 1]. By the CMT, $D(h;n) \Rightarrow D(h)$, and we call $D(h;n)$ the *weighted Durbin–Watson* (DW) estimator for σ^2.

We henceforth use the notation $H(t) \equiv \int_0^t h(u)\,du$, $\bar{H}(t) \equiv \int_0^t H(u)\,du$, and $\hat{H}(t) \equiv \int_0^t \bar{H}(u)\,du$, along with $H \equiv H(1)$, $\bar{H} \equiv \bar{H}(1)$, and $\hat{H} \equiv \hat{H}(1)$. In order to normalize $h(t)$, we note that for $s \le t$,

$$\begin{aligned}q(s,t) &\equiv \mathsf{Cov}(\mathcal{B}(s) - \bar{\mathcal{B}}, \mathcal{B}(t) - \bar{\mathcal{B}}) \\ &= s(1-t) - \frac{s-s^2}{2} - \frac{t-t^2}{2} + \frac{1}{12}.\end{aligned}$$

Thus, $h(t)$ must satisfy

$$\mathsf{E}\left[\int_0^1 h(t)(\mathcal{B}(t) - \bar{\mathcal{B}})^2\,dt\right] = \int_0^1 h(t)q(t,t)\,dt = \frac{H}{12} = 1.$$

We now state the main theorem on the estimator $D(h;n)$, the proof of which is deferred to the Appendix.

THEOREM 4: Under the conditions of Theorem 1, we have

$$\mathsf{E}[D(h;n)] = \sigma^2\left(1 - \frac{h(0)-h(1)}{24n}\right) + \frac{\gamma}{n}(9 - \bar{H} + 2\hat{H}) + o(1/n). \quad (8)$$

(We can remove the $O(1/n)$ term in the coefficient for σ^2 in (8), simply by dividing the estimator by that known coefficient.) If we also assume uniform integrability of $D^2(h;n)$, then as $n \to \infty$,

$$\mathsf{Var}(D(h;n)) \to \mathsf{Var}(D(h)) = 4\sigma^4 \int_0^1 \int_0^t h(s)h(t)q^2(s,t)\, ds\, dt. \quad (9)$$

EXAMPLE 7: If we define $h_0(t) \equiv 12$ for all t, then $D(h_0;n) = D_0(n)$, from Example 5, where we found that $\mathsf{Var}(D(h_0;n)) \approx 2\sigma^4/5$ (which is very small) and $\mathsf{Bias}(D(h_0;n)) \approx 7\gamma/n$ (which is very high). In order to obtain a first-order unbiased DW estimator, Theorem 4 requires a weighting function such that $9 = \bar{H} - 2\hat{H}$, in addition to the normalizing constraint $H = 12$. The quadratic weighting function satisfying these constraints while yielding the minimum-variance estimator is $h_2(t) \equiv -198 + 1260t - 1260t^2$. Unfortunately, Eq. (9) implies that $\mathsf{Var}(D(h_2)) = 37\sigma^4/5$; so the price of first-order unbiasedness is an unacceptably high variance, at least for this example. In other work, details of which are not reported here, it can be shown that there is almost nothing to be gained by going up to cubic or quartic weighting functions. □

3.2.4. *Jackknife Estimator*

In light of the mixed reviews on the DW estimator—low variance but high bias or low bias but high variance—one might wonder whether there is any other way to take advantage of the excellent variance of $D(h_0;n)$. One possible alternative estimator looks like a jackknifed version of $D(h_0;n)$, namely,

$$D_{J,r}(h_0;n) \equiv \frac{D(h_0;n) - rD(h_0;rn)}{1-r} = \frac{2C(g_0;n) - A(f_0;n) - 2rC(g_0;rn) + rA(f_0;rn)}{1-r}, \quad (10)$$

where r is fixed in $(0,1)$. Bias and variance results for this estimator are summarized by the following theorem.

THEOREM 5: Under the assumptions of Theorem 1,

$$\mathsf{E}[D_{J,r}(h_0;n)] = \sigma^2 + o(1/n). \quad (11)$$

Further, assuming that $D^2_{J,r}(h_0;n)$ is uniformly integrable, then as $n \to \infty$,

$$\mathsf{Var}(D_{J,r}(h_0;n)) \to \frac{2(1 + r + 2r^2 - 2r^3)\sigma^4}{5(1-r)}. \quad (12)$$

Application of Example 5 or Eq. (8) to both terms on the right-hand side of Eq. (10) proves that $D_{J,r}(h_0;n)$ is first-order unbiased for σ^2. The second half of the theorem is proven in the Appendix.

REMARK 1: For finite n, the results of the theorem ought to hold, at least approximately, for any fixed r $(0 < r < 1)$ as long as the sample size n is large enough. As a matter of fact, as r approaches 0, it seems as if the variance of the first-order unbiased estimator $D_{J,r}(h_0;n)$ approaches the remarkable $2\sigma^4/5$. Of course, in real life, we do not have the luxury of arbitrarily large n; for small $r > 0$, this bodes poorly for the convergence of $D(h_0;rn)$, and hence $D_{J,r}(h_0;n)$, to their limiting Brownian bridge functionals (described in the Appendix). Indeed, Eq. (11)'s first-order unbiasedness result for $D_{J,r}(h_0;n)$ $(r > 0)$ breaks down badly when we take $r = 0$, as we know from Example 5. Thus, to be on the safe side, we recommend the choice $r = 1/2$, for which $\mathsf{Var}(D_{J,1/2}(h_0;n)) \to 7\sigma^4/5$, a very satisfying value, especially since $D_{J,1/2}(h_0;n)$ is first-order unbiased for σ^2. This variance estimator shall be addressed further in the examples of Section 4. □

4. EXAMPLES

To illustrate the performance of the new estimators, in this section we give exact calculations for a moving average process and a Monte Carlo example involving an autoregressive process. Specifically, we will find or estimate the expected value and variance of the area and CvM estimators from Sections 2.2 and 2.3, as well as the new estimators $D(h_0;n)$, $D_{\hat{\alpha}}(f_2, g_2^\star;n)$, and $D_{J,r}(h_0;n)$, from Sections 3.2.1, 3.2.2, and 3.2.4, respectively.

4.1. Exact Results: Moving Average Process

Suppose the underlying stationary process is a first-order moving average [MA(1)] process, given by $Y_i = \theta\epsilon_{i-1} + \epsilon_i$, $i \geq 1$, where the ϵ_i's are independent and identically distributed (i.i.d.) Nor(0, 1) r.v.'s. The MA(1) has covariance function $R_0 = 1 + \theta^2$, $R_{\pm 1} = \theta$, $R_k = 0$ elsewhere, whence we have $\sigma^2 = \sum_{j=-\infty}^{\infty} R_j = (1+\theta)^2$ and $\gamma = -2\sum_{j=1}^{\infty} jR_j = -2\theta$.

Table 1 displays exact expectations and variances for the variance estimators under study (some of which are also found in [13]). The tedious calculations required to obtain

Table 1. Exact expected values and variances of estimators for σ^2 for an MA(1) process.

Estimator	Expected value	Variance (to $O(n^{-2})$)
$A(f_0;n)$	$\sigma^2 + \frac{3\gamma}{n} - \frac{\sigma^2}{n^2} - \frac{3\gamma}{n^3}$	$2\sigma^4 + \frac{12\gamma\sigma^2}{n}$
$A(f_2;n)$	$\sigma^2 + \frac{7(\sigma^2+6\gamma)}{2n^2} + O(n^{-3})$	$2\sigma^4$
$C(g_0;n)$	$\sigma^2 + \frac{5\gamma}{n} - \frac{\sigma^2+6\gamma}{n^2} + \frac{\gamma}{n^3}$	$\frac{4\sigma^4}{5} + \frac{16\gamma\sigma^2}{5n}$
$C(g_2^\star;n)$	$\sigma^2 + \frac{4(\sigma^2+6\gamma)}{n^2} + O(n^{-3})$	$\frac{121\sigma^4}{70} + \frac{4.057\gamma\sigma^2}{n}$
$D(h_0;n)$	$\sigma^2 + \frac{7\gamma}{n} - \frac{\sigma^2+12\gamma}{n^2} + \frac{5\gamma}{n^3}$	$\frac{2\sigma^4}{5} + \frac{4\gamma\sigma^2}{5n}$
$D_\alpha(f_2,g_2^\star;n)$	$\sigma^2 + \frac{(\sigma^2+6\gamma)(4-\frac{\alpha}{2})}{n^2} + O(n^{-3})$	$1.723\sigma^4 + \frac{0.066\gamma\sigma^2}{n}$ $(\hat{\alpha} = 0.1281)$
$D_{J,r}(h_0;n)$	$\sigma^2 + \frac{\sigma^2+12\gamma}{rn^2} - \frac{5(1+r)\gamma}{r^2n^3}$	$\frac{2(1+r+2r^2-2r^3)\sigma^4}{5(1-r)} + \frac{4(1+r^2-r^3)\gamma\sigma^2}{5(1-r)^2n}$

these results are relegated to the online companion [12]. We see that all of the table entries match up with the corresponding theoretical results from Sections 2 and 3. Thus, all of the estimators under consideration in Table 1 are asymptotically unbiased for σ^2 as $n \to \infty$; but only $A(f_2;n)$, $C(g_2^\star;n)$, $D_\alpha(f_2, g_2^\star;n)$, and $D_{J,r}(h_0;n)$ are first-order unbiased. Further, $D(h_0;n)$ has the lowest variance (although it has the largest bias).

4.2. Monte Carlo Example: Autoregressive Process

Here we describe a Monte Carlo experiment based on the first-order autoregressive [AR(1)] process, which is defined as $Y_i = \phi Y_{i-1} + \epsilon_i$, $i \geq 1$, where the ϵ_i's are i.i.d. $\text{Nor}(0, 1-\phi^2)$ r.v.'s, and Y_0 is a $\text{Nor}(0,1)$ r.v. initialized independently of the others. The AR(1) has covariance function $R_k = \phi^{|k|}$ for all k, whence

$$\sigma^2 = \sum_{j=-\infty}^{\infty} R_j = \frac{1+\phi}{1-\phi}$$

$$\text{and} \quad \gamma = -2\sum_{j=1}^{\infty} jR_j = \frac{-2\phi}{(1-\phi)^2}. \quad (13)$$

4.2.1. Monte Carlo Setup

In the current example, we set $\phi = 0.9$, which corresponds to a fairly highly positive autocorrelation structure and variance parameter $\sigma^2 = 19$. We ran 100,000 independent replications of the process, each of which yielded AR(1) observations. For each of the 100,000 replications, we stored variance estimates for each of a variety of area, CvM, and combined estimators with sample sizes $n = 256$, 512, 1024, and 2048. Our main results are given in Table 2. The columns marked "$\widehat{\mathsf{E}}$" and "$\widehat{\mathsf{Var}}$" in Table 2 denote the sample means and variances calculated over the 100,000 replications for each selection of estimator and sample size. The right-most column of Table 2, denoted "True Var ($n \to \infty$)," is simply the list of limiting variances of the various estimators, obtained from Theorem 1, Example 3, and Theorem 5.

4.2.2. Discussion

We have a number of findings based on the results from Table 2.

- As the sample size n becomes large, the estimated expected values of the estimators all appear to approach $\sigma^2 = 19$, some faster than others. In fact, by the time $n = 512$ observations have been taken, the estimated expected values of the first-order unbiased estimators $A(f_2;n)$, $A(f_{\cos,1};n)$, $C(g_2^\star;n)$, $C(g_4^\star;n)$, and $D_{J,r}(h_0;n)$, $r \geq 0.4$, are all within 2% of $\sigma^2 = 19$. By the time $n = 2048$, the estimated expected values of all of the estimators except $A(f_0;n)$, $C(g_0;n)$, and $D(h_0;n)$—which are not first-order unbiased—are within 1% of σ^2.
- Note that the estimators' expected values are generally below σ^2. This makes sense since the AR(1)'s positive covariance function implies that the quantity $\gamma < 0$. Thus, e.g., if we appeal to Theorem 1's Eq. (1), we see that any first-order bias for the area estimator will often be negative as well. For instance, Example 1 and Eq. (13) imply

$$\mathsf{E}[A(f_0;256)] \doteq \sigma^2 + \frac{3\gamma}{256} = 19 - \frac{3(180)}{256} = 16.89,$$

  where $\doteq$ denotes "approximately equal to." This quantity is almost precisely equal to the corresponding Monte Carlo estimate of 16.99 from Table 2. (See [1] for more details.)
- As n becomes large, the variances of the estimators all appear to approach their respective theoretical limiting values. For example, we see that $\widehat{\mathsf{Var}}(C(g_2^\star;n)) \to 121\sigma^4/70 = 624.0$.

Table 2. Estimated expected values and variances of estimators for σ^2 for the AR(1) process with $\phi = 0.9$ ($\sigma^2 = 19$).

Estimator	$n = 256$		$n = 512$		$n = 1024$		$n = 2048$		True Var
	$\widehat{\mathsf{E}}$	$\widehat{\mathsf{Var}}$	$\widehat{\mathsf{E}}$	$\widehat{\mathsf{Var}}$	$\widehat{\mathsf{E}}$	$\widehat{\mathsf{Var}}$	$\widehat{\mathsf{E}}$	$\widehat{\mathsf{Var}}$	$(n \to \infty)$
$A(f_0;n)$	16.99	578	17.95	645	18.47	680	18.68	702	722.0
$A(f_2;n)$	18.14	656	18.73	699	18.98	716	18.89	708	722.0
$A(f_{\cos,1};n)$	18.10	653	18.73	701	19.00	718	18.88	706	722.0
$C(g_0;n)$	15.84	241	17.32	266	18.15	277	18.52	282	288.8
$C(g_2^\star;n)$	18.09	546	18.74	587	18.96	609	18.90	607	624.0
$C(g_4^\star;n)$	17.17	305	18.41	344	18.80	359	18.97	372	376.2
$D(h_0;n)$	14.64	119	16.70	134	17.81	141	18.43	142	144.4
$D_{J,0.1}(h_0;n)$	16.04	147	17.92	168	18.68	174	18.88	175	179.4
$D_{J,0.2}(h_0;n)$	16.92	185	18.35	212	18.84	221	18.92	222	228.2
$D_{J,0.3}(h_0;n)$	17.46	238	18.56	272	18.90	285	18.94	287	294.2
$D_{J,0.4}(h_0;n)$	17.79	308	18.68	355	18.96	370	18.93	373	383.1
$D_{J,0.5}(h_0;n)$	18.00	402	18.72	465	18.98	486	18.95	492	505.4
$D_{J,0.6}(h_0;n)$	18.17	538	18.79	621	19.01	653	18.91	662	681.6
$D_{J,0.7}(h_0;n)$	18.33	738	18.81	858	19.04	911	18.87	930	959.8
$D_{J,0.8}(h_0;n)$	18.60	1094	18.73	1293	18.99	1388	18.98	1428	1484
$D_{J,0.9}(h_0;n)$	18.71	1804	18.79	2344	18.96	2684	19.05	2785	2978

- As anticipated, some of the estimators have lower variance than others. Specifically, the pure CvM estimators, $D(h_0;n)$, and the jackknife estimators with $r \le 0.6$ all exhibit lower variance than do the pure area estimators. Of course, as pointed out previously, $D(h_0;n)$'s exceptionally low variance comes at the cost of high bias.

5. BATCHING

This section generalizes some of the previous estimators by incorporating the use of batching. Section 5.1 gives expected value and variance results for STS batched estimators, while Sections 5.2 and 5.3 review the benchmark methods of nonoverlapping and overlapping batch means. In Section 5.4, we compare all of these estimators in terms of their asymptotic bias, variance, and mean squared error (MSE). Batching typically increases estimator bias, but also reduces its variance; so it may be of interest to evaluate batching's overall effect on MSE. Finally, Section 5.5 presents a Monte Carlo example comparing the different variance estimators.

5.1. Batching the STS Estimators

All of our work so far has been for estimators for σ^2 based on one long batch of size n. It is well known that additional batching usually yields estimators having lower variance (and higher bias) than their one-batch predecessors. The method using nonoverlapping batches is especially simple.

- Divide the run into b contiguous, nonoverlapping batches, each of size m (assuming $n = mb$). Batch i consists of observations $Y_{(i-1)m+1}, Y_{(i-1)m+2}, \ldots, Y_{im}$, $i = 1, 2, \ldots, b$.
- Calculate estimators from each batch (instead of from the entire run). Using the obvious notation, we denote the area, CvM, and jackknife estimators from batch i as $A_i(f;m)$, $C_i(g;m)$, and $D_{J,r,i}(h_0;m)$, respectively.
- The batched area, CvM, and jackknife estimators for σ^2 are the sample means of the corresponding estimators from the individual nonoverlapping batches, i.e.,

$$\bar{\mathcal{A}}(f;b,m) \equiv \frac{1}{b}\sum_{i=1}^{b} A_i(f;m),$$

$$\bar{\mathcal{C}}(g;b,m) \equiv \frac{1}{b}\sum_{i=1}^{b} C_i(g;m),$$

and

$$\bar{\mathcal{D}}_{J,r}(h_0;b,m) \equiv \frac{1}{b}\sum_{i=1}^{b} D_{J,r,i}(h_0;m).$$

Since the batched estimators are simply linear combinations of estimators from each batch of size m, Eqs. (1), (3), and (11) immediately show that

$$\mathsf{E}[\bar{\mathcal{A}}(f;b,m)] = \mathsf{E}[A(f;m)] = \sigma^2 + \frac{[(F-\bar{F})^2 + \bar{F}^2]\gamma}{2m} + o(1/m), \quad (14)$$

$$\mathsf{E}[\bar{\mathcal{C}}(g;b,m)] = \mathsf{E}[C(g;m)] = \sigma^2 + \frac{\gamma}{m}(G-1) + o(1/m), \quad (15)$$

and

$$\mathsf{E}[\bar{\mathcal{D}}_{J,r}(h_0;b,m)] = \mathsf{E}[D_{J,r}(h_0;m)] = \sigma^2+o(1/m). \quad (16)$$

REMARK 2: We can fine-tune the expected values from Eqs. (14)–(16) for certain weighting functions. To do so, suppose we define $\delta \equiv -2\sum_{i=1}^{\infty} i^2 R_i$. Then Aktaran-Kalaycı et al. [1] find that under general conditions (for instance, if the covariance function decays exponentially)

$$\mathsf{E}[\bar{\mathcal{A}}(f_0;b,m)] = \sigma^2 + \frac{3\gamma}{m} - \frac{\sigma^2}{m^2} + o(1/m^2), \quad (17)$$

$$\mathsf{E}[\bar{\mathcal{A}}(f_2;b,m)] = \sigma^2 + \frac{7(\sigma^2+6\delta)}{2m^2} + o(1/m^2), \quad (18)$$

$$\mathsf{E}[\bar{\mathcal{C}}(g_0;b,m)] = \sigma^2 + \frac{5\gamma}{m} - \frac{\sigma^2+6\delta}{m^2} + o(1/m^2), \quad (19)$$

$$\mathsf{E}[\bar{\mathcal{C}}(g_2^\star;b,m)] = \sigma^2 + \frac{4(\sigma^2+6\delta)}{m^2} + o(1/m^2). \quad (20)$$

In addition, Eqs. (10), (17), and (19) imply that

$$\begin{aligned}\mathsf{E}[\bar{\mathcal{D}}_{J,r}(h_0;b,m)] &= \frac{2\mathsf{E}[C(g_0;m)] - \mathsf{E}[A(f_0;m)] - 2r\mathsf{E}[C(g_0;rm)] + r\mathsf{E}[A(f_0;rm)]}{1-r} \\ &= \sigma^2 + \frac{\sigma^2+12\delta}{rm^2} + o(1/m^2). \quad (21)\end{aligned}$$

We see that Eqs. (17)–(21) match up with the corresponding exact MA(1) results from Table 1 (using m in place of n and noting that $\gamma = \delta$ for the MA(1)). □

Before going on to the corresponding variance results, we note that the area, CvM, and jackknife estimators can be written as batched quadratic forms, i.e., quadratic forms in terms of the Y_i's within each batch. Then the main result of [4] allows us to make the intuitively pleasing assumption that estimators from two different batches are approximately independent—at least under suitable moment and mixing conditions and large enough batch size m. In this case, Eqs. (2), (4), and (12) imply

$$\mathsf{Var}(\bar{\mathcal{A}}(f;b,m)) \doteq \frac{\mathsf{Var}(A(f;m))}{b} \to \frac{2\sigma^4}{b}, \quad (22)$$

$$\mathsf{Var}(\bar{\mathcal{C}}(g;b,m)) \doteq \frac{\mathsf{Var}(C(g;m))}{b} \to \frac{\mathsf{Var}(C(g))}{b}, \quad (23)$$

and

$$\mathsf{Var}(\bar{\mathcal{D}}_{J,r}(h_0;b,m)) \doteq \frac{\mathsf{Var}(D_{J,r}(h_0;m))}{b} \to \frac{2(1+r+2r^2-2r^3)\sigma^4}{5(1-r)b}, \quad (24)$$

where the right-hand-side limits are taken as $m \to \infty$. So batching helps to decrease estimator variance (by a factor of b), although this is achieved at the cost of an increase in estimator bias (since m now appears instead of n in the expected value expressions).

5.2. Nonoverlapping Batch Means

We describe the method of nonoverlapping batch means (NBM), one of the most popular techniques in practice and one that we will use as a benchmark for comparison. The quantities $\bar{Y}_{i,m} \equiv \sum_{j=1}^{m} Y_{(i-1)m+j}/m$, $i = 1,2,\ldots,b$, are referred to as the *batch means* of the process $Y_1, Y_2, \ldots, Y_n$, and are often assumed to be i.i.d. normal random variables, at least for large enough batch size m. This i.i.d. assumption suggests the NBM estimator for σ^2,

$$\mathcal{N}(b,m) \equiv \frac{m}{b-1}\sum_{i=1}^{b}\left(\bar{Y}_{i,m} - \bar{Y}_n\right)^2$$

(see [20]). Under mild conditions, [11] and [23] show that $\mathcal{N}(b,m) \Rightarrow \sigma^2\chi^2_{b-1}/(b-1)$ as $m \to \infty$ with b fixed, where χ^2_ν denotes the chi-square distribution with ν degrees of freedom. Further, [6], [14], and [22] (among others) yield

$$\mathsf{E}[\mathcal{N}(b,m)] = \sigma^2 + \frac{\gamma(b+1)}{bm} + o(1/m) \quad (25)$$

and, as $m \to \infty$,

$$\mathsf{Var}(\mathcal{N}(b,m)) \to \frac{2\sigma^4}{b-1}. \quad (26)$$

We see from Eq. (25) that $\mathcal{N}(b,m)$ has mild first-order bias as an estimator for σ^2. On the other hand, Eq. (16) reveals that $\bar{\mathcal{D}}_{J,r}(h_0;b,m)$ is first-order unbiased for σ^2, and Eqs. (14) and (15) show the same for $\bar{\mathcal{A}}(f;b,m)$ and $\bar{\mathcal{C}}(g;b,m)$, respectively, as long as those estimators use the appropriate first-order unbiased weighting functions. Comparing Eq. (26) with (22)–(24), we conclude that the (asymptotic) variance of $\mathcal{N}(b,m)$ is at least as big as those of most of the other estimators.

5.3. Overlapping Batch Means

We also review the method of overlapping batch means (OBM), an alternative to NBM that seems to have better performance characteristics in many situations. The OBM variance estimator will serve as another benchmark for comparison.

First, define the ith overlapping batch mean as $\bar{Y}^{\mathrm{O}}_{i,m} \equiv \sum_{k=0}^{m-1} Y_{i+k}/m$, $i = 1,2,\ldots,n-m+1$. The OBM estimator for σ^2, due to [18] (with a slightly different scaling constant), is

$$\mathcal{O}(b,m) \equiv \frac{nm}{(n-m+1)(n-m)}\sum_{i=1}^{n-m+1}\left(\bar{Y}^{\mathrm{O}}_{i,m} - \bar{Y}_n\right)^2,$$

where we continue to define $b \equiv n/m$, although b can no longer be regarded as "the number of batches."

Under mild conditions, [3] and [7] show that, as the batch size m becomes large,

$$\mathsf{E}[\mathcal{O}(b,m)] = \sigma^2 + \frac{\gamma(b^2+1)}{mb(b-1)} + o(1/m) \doteq \sigma^2 + \frac{\gamma}{m} + o(1/m)$$

and

$$\mathsf{Var}\big[\mathcal{O}(b,m)\big] \to \frac{(4b^3-11b^2+4b+6)\sigma^4}{3(b-1)^4} \doteq \frac{4\sigma^4}{3b}$$

(also see [14], [18], and [22]), with the approximate results holding for large b. The expected value of the OBM estimator is almost the same as that for NBM, while OBM's asymptotic variance is superior to NBM's—and, indeed, competitive with the variances arising from the area, CvM, and jackknife estimators under study herein.

5.4. Asymptotic Bias, Variance, and Mean Squared Error

This subsection uses the bias and variance results from Sections 5.1–5.3 to obtain "optimal" MSE expressions for the variance estimators under consideration.

Consider an estimator $\widehat{V}$ for the variance parameter σ^2. As is the case with all of the variance estimators studied in the current paper, suppose that $\mathsf{Bias}(\widehat{V}) = \zeta_b/m^k + o(1/m^k)$ and $\mathsf{Var}(\widehat{V}) = \zeta_v/b + o(1/b)$ for some estimator-specific constants ζ_b and ζ_v and $k > 0$. Ignoring small-order noise terms, the MSE of $\widehat{V}$ is

$$\mathsf{MSE}(\widehat{V}) = \mathsf{Bias}^2(\widehat{V}) + \mathsf{Var}(\widehat{V}) = \frac{\zeta_b^2}{m^{2k}} + \frac{\zeta_v}{b}. \qquad (27)$$

To minimize this quantity, suppose that the number of batches is of the form $b = \omega n^{\epsilon}$ for some appropriately chosen $\omega > 0$ and $0 < \epsilon < 1$. Then Eq. (27) becomes

$$\mathsf{MSE}(\widehat{V}) = \frac{\omega^{2k}\zeta_b^2}{n^{(1-\epsilon)2k}} + \frac{\zeta_v}{\omega n^{\epsilon}}. \qquad (28)$$

The value of ϵ that minimizes this expression for MSE is determined by equating the exponents of n on the right-hand side of Eq. (28), so that $\epsilon = 2k/(1+2k)$. Thus,

$$\mathsf{MSE}(\widehat{V}) = n^{-2k/(1+2k)}\left(\omega^{2k}\zeta_b^2 + \frac{\zeta_v}{\omega}\right).$$

Table 3. Approximate large-sample bias, variance, and MSE of batched estimators for σ^2 for stationary processes.

Estimator	k	ζ_b	$b\,\zeta_v/\sigma^4$	$\mathsf{MSE}^\star$
$\bar{\mathcal{A}}(f_2;b,m)$	2	$3.5(\sigma^2+6\delta)$	2	$O(n^{-4/5})$
$\bar{\mathcal{C}}(g_2^\star;b,m)$	2	$4(\sigma^2+6\delta)$	121/70	$O(n^{-4/5})$
$\bar{\mathcal{D}}_{J,r}(h_0;b,m)$	2	$(\sigma^2+12\delta)/r$	$\frac{2(1+r+2r^2-2r^3)}{5(1-r)}$	$O(n^{-4/5})$
$\mathcal{N}(b,m)$	1	γ	2	$O(n^{-2/3})$
$\mathcal{O}(b,m)$	1	γ	4/3	$O(n^{-2/3})$

Minimizing this quantity with respect to ω, as detailed in [2], [14], and [22], we obtain the asymptotically optimal MSE,

$$\mathsf{MSE}^\star(\widehat{V}) = (1+2k)\left(\zeta_b\left(\frac{\zeta_v}{2nk}\right)^k\right)^{2/(1+2k)} = O\big(n^{-2k/(1+2k)}\big).$$

Table 3 gathers the bias, variance, and MSE results from this section together in one place (assuming large m and b). Among the estimators on this list, we see that the area (with weighting function $f_2(t)$), CvM (with weight $g_2^\star(t)$), and jackknife estimators all have the lowest order bias (highest k), as well as the lowest order "optimal" MSE. For fixed number of batches b, the jackknife (with, say, $r = 0.5$) and OBM estimators have the lowest variances.

5.5. Monte Carlo Example: Autoregressive Process

This subsection discusses an example involving a specific stochastic process in a non-asymptotic setting.

Consider the AR(1) process from Section 4.2, where we again take $\phi = 0.9$ (so that $\sigma^2 = 19$). We ran 100,000 independent replications of the process, each yielding AR(1) observations. For each of the replications, we stored variance estimates for each of the estimators $\bar{\mathcal{A}}(f_2;b,m)$, $\bar{\mathcal{C}}(g_2^\star;b,m)$, $\bar{\mathcal{D}}_{J,0.5}(h_0;b,m)$, $\mathcal{N}(b,m)$, and $\mathcal{O}(b,m)$ with sample sizes $n = 2048$, 8192, and $b = n/m = 4, 8, 16$. Table 4 gives the results. The columns marked "$\widehat{\mathsf{E}}$", "$\widehat{\mathsf{Var}}$", and "$\widehat{\mathsf{MSE}}$" in the table denote the sample mean, variance, and MSE calculated over the 100,000 replications for each selection of estimator and (b,m).

We make some points concerning the empirical results from Table 4.

- For fixed $b = n/m$, the estimated expected values of the estimators all appear to approach $\sigma^2 = 19$ as the batch size m increases (i.e., as the sample size n jumps from 2048 to 8192), in line with the theory. For very small batch sizes, the NBM and OBM estimators fare the best in terms of bias; but as m increases, the area, CvM, and jackknifed estimators always do at least a bit better than NBM and OBM—what is happening is that the first-order unbiasedness

Table 4. Estimated performance characteristics of batched area, CvM, jackknifed, NBM, and OBM estimators for σ^2 for the AR(1) process with $\phi = 0.9$, $n = 2048$, 8192, and $b = 4$, 8, 16 (for this process, $\sigma^2 = 19$).

		$b = 4$			$b = 8$			$b = 16$		
n	Estimator	$\widehat{\mathsf{E}}$	$\widehat{\mathsf{Var}}$	$\widehat{\mathsf{MSE}}$	$\widehat{\mathsf{E}}$	$\widehat{\mathsf{Var}}$	$\widehat{\mathsf{MSE}}$	$\widehat{\mathsf{E}}$	$\widehat{\mathsf{Var}}$	$\widehat{\mathsf{MSE}}$
2048	$\bar{\mathcal{A}}(f_2; b, m)$	18.75	176	176	18.12	82.2	83.0	16.12	32.4	40.7
	$\bar{\mathcal{C}}(g_2^\star; b, m)$	18.72	148	148	18.05	68.1	69.0	16.06	27.4	36.0
	$\bar{\mathcal{D}}_{J,0.5}(h_0; b, m)$	18.70	116	116	17.96	50.6	51.7	15.64	17.9	29.2
	$\mathcal{N}(b, m)$	18.54	230	230	18.19	94.8	95.5	17.50	41.3	43.6
	$\mathcal{O}(b, m)$	18.49	150	150	18.18	69.0	69.7	17.49	31.5	33.8
8192	$\bar{\mathcal{A}}(f_2; b, m)$	18.92	178	178	18.99	89.7	89.7	18.76	43.9	44.0
	$\bar{\mathcal{C}}(g_2^\star; b, m)$	18.94	154	154	18.99	76.3	76.3	18.73	36.7	36.8
	$\bar{\mathcal{D}}_{J,0.5}(h_0; b, m)$	18.97	125	125	18.96	60.8	60.8	18.70	28.9	29.0
	$\mathcal{N}(b, m)$	18.80	235	235	18.73	100	100	18.63	46.1	46.2
	$\mathcal{O}(b, m)$	18.82	151	151	18.76	69.2	69.3	18.61	32.3	32.4

properties of the area, CvM, and jackknifed estimators finally manifest themselves at moderate batch sizes, as predicted by Eqs. (18), (20), and (21). For example, $\widehat{\mathsf{E}}[\bar{\mathcal{A}}(f_2; 16, 128)] < \widehat{\mathsf{E}}[\mathcal{N}(16, 128)] \ll 19$, yet $\widehat{\mathsf{E}}[\mathcal{N}(16, 512)] < \widehat{\mathsf{E}}[\bar{\mathcal{A}}(f_2; 16, 512)] \doteq 19$. (See [1] for general results in this vein.)

- For fixed sample size n, the bias of each estimator decreases as the batch size m increases (i.e., as the ratio $b = n/m$ decreases), also as the theory predicts.
- As the batch size m becomes large for fixed b, the variances of the estimators all appear to approach their respective theoretical limiting values. For example, we see that $\widehat{\mathsf{Var}}(\bar{\mathcal{C}}(g_2^\star; 8, 1024)) = 76.3 \doteq 121\sigma^4/(70b) = 78.0$. We can also rank the variances. For instance,

$$\begin{aligned}\widehat{\mathsf{Var}}(\bar{\mathcal{D}}_{J,0.5}(h_0; 8, 1024)) &< \widehat{\mathsf{Var}}(\mathcal{O}(8, 1024))\\ &< \widehat{\mathsf{Var}}(\bar{\mathcal{C}}(g_2^\star; 8, 1024)) < \widehat{\mathsf{Var}}(\bar{\mathcal{A}}(f_2; 8, 1024))\\ &\doteq \widehat{\mathsf{Var}}(\mathcal{N}(8, 1024)),\end{aligned}$$

which is yet again consistent with the theory. In addition, the order of the respective MSEs is exactly the same, indicating that, in this example, the MSEs are dominated by their variance components. (We made no attempt to obtain the "optimal" asymptotic MSEs in this example, where our work in Section 5.4 would have mandated $b = \omega n^{2k/(1+2k)}$ and $m = \omega^{-1} n^{1/(1+2k)}$, which in turn would have produced unacceptably high bias.)

6. CONCLUSIONS

This article has studied various estimators for the variance parameter σ^2, all of which are based on combinations of standardized-time-series area and Cramér–von Mises estimators. Two estimators—one a simple linear combination of the area and CvM and one reminiscent of the Durbin–Watson statistic—can achieve first-order unbiasedness with appropriate choices of weighting functions, but are not especially impressive in terms of variance. A third estimator—related to a jackknifed version of the first—has more interesting performance characteristics. In addition to being first-order unbiased, the jackknifed estimator has variance that is lower than its constituent area estimator component and competitive with its constituent CvM component (both of which are biased).

The work above was initially applied to one long batch of observations; but it was obviously of interest to examine the effects of batching. In doing so, we found that the results of the one-batch case generalized naturally. Namely, if we keep the sample size fixed while increasing the number of batches (i.e., reducing the batch size), it turns out that estimator bias increases a bit, while estimator variance decreases substantially. When we applied batching, we saw that the jackknifed estimator fares well against the NBM, OBM, area, and CvM estimators in terms of bias, variance, and MSE. In fact, we recommend use of the jackknife estimator with $r = 0.5$ over the other estimators that we have studied in this paper.

A number of estimator augmentations are currently under investigation. For instance, one can perform an enhanced version of jackknifing, the result of which will be an additional savings in terms of estimator variance. We can also employ an overlapping estimator strategy to reduce variance as in the overlapping batch means technology of [18]. Finally, instead of concentrating on point estimator performance measures such as expected value, variance, and MSE, one might also be interested in formulating and evaluating confidence intervals for the mean μ of the underlying stationary stochastic process.

APPENDIX

Here we prove Lemma 1 and Theorems 4 and 5 from the main text. We first state a well-known fact that will be useful in the sequel.

LEMMA 2 (see [19]): If X_1 and X_2 are jointly normal with mean zero, then $\mathsf{Cov}(X_1^2, X_2^2) = 2\mathsf{Cov}^2(X_1, X_2)$.

PROOF OF LEMMA 1: By Lemma 2,

$$\begin{aligned}\mathsf{Cov}(A(f), C(g)) &= \sigma^4 \int_0^1 g(t)\mathsf{Cov}\left(\left(\int_0^1 f(s)\mathcal{B}(s)\,ds\right)^2, \mathcal{B}^2(t)\right) dt \\ &= 2\sigma^4 \int_0^1 g(t)\mathsf{Cov}^2\left(\int_0^1 f(s)\mathcal{B}(s)\,ds,\, \mathcal{B}(t)\right) dt \\ &= 2\sigma^4 \int_0^1 g(t)\left[\int_0^1 f(s)\mathsf{Cov}(\mathcal{B}(s), \mathcal{B}(t))\,ds\right]^2 dt \\ &= 2\sigma^4 \int_0^1 g(t)\left[\int_0^t f(s)s(1-t)\,ds + \int_t^1 f(s)t(1-s)\,ds\right]^2 dt,\end{aligned}$$

and the result follows after using integration by parts on $\int f(t)t\,dt$. □

PROOF OF THEOREM 4: After some algebra, we can rewrite the expression for $D(h;n)$ in terms of various area and CvM estimators.

$$\begin{aligned}D(h;n) &= C(h;n) - \frac{2}{\sqrt{12}}S(f_0;n)S(h;n) + \frac{A(f_0;n)}{12n}\sum_{k=1}^{n} h\left(\frac{k}{n}\right) \\ &= C(h;n) - \frac{(S(f_0;n)+S(h;n))^2 - S^2(f_0;n) - S^2(h;n)}{\sqrt{12}} \\ &\quad + \frac{A(f_0;n)}{12n}\sum_{k=1}^{n} h\left(\frac{k}{n}\right) \\ &= C(h;n) + \frac{A(h;n) - A(f_0+h;n)}{\sqrt{12}} \\ &\quad + \left(\frac{1}{\sqrt{12}} + \frac{1}{12n}\sum_{k=1}^{n} h\left(\frac{k}{n}\right)\right)A(f_0;n) \\ &\quad \text{(since } S(f_0+h;n) = S(f_0;n) + S(h;n) \text{ and } A(f;n) = S^2(f;n)) \\ &= C(h;n) + \frac{A(h;n) - A(f_0+h;n)}{\sqrt{12}} \\ &\quad + \left(\frac{1}{\sqrt{12}} + \frac{1}{12}\int_0^1 h(s)\,ds + \frac{h(1)-h(0)}{24n} + o(1/n)\right)A(f_0;n) \\ &\quad \text{(by the trapezoid rule)} \\ &= C(h;n) + \frac{A(h;n) - A(f_0+h;n)}{\sqrt{12}} \\ &\quad + \left(1 + \frac{1}{\sqrt{12}} + \frac{h(1)-h(0)}{24n} + o(1/n)\right)A(f_0;n) \\ &\quad \text{(since we require } H = 12). \end{aligned} \tag{29}$$

Before explicitly calculating $\mathsf{E}[D(h;n)]$, we calculate some intermediate quantities. First, we recall from Example 1 that $\mathsf{E}[A(f_0;n)] = \sigma^2 + 3\gamma/n + o(1/n)$. We need to get similar expressions for $\mathsf{E}[C(h;n)]$, $\mathsf{E}[A(h;n)]$, and $\mathsf{E}[A(f_0+h;n)]$, but care must be taken since the weighting functions $h(t)$ and $f_0(t)+h(t)$ used in $C(h;n)$, $A(h;n)$, and $A(f_0+h;n)$ are not necessarily normalized with respect to σ^2, i.e., $\mathsf{E}[C(h)]$, $\mathsf{E}[A(h)]$, and $\mathsf{E}[A(f_0+h)] \neq \sigma^2$.

In order to derive an expression for $\mathsf{E}[C(h;n)]$, first consider

$$\mathsf{E}[C(h)] = \kappa_h\sigma^2,$$

where

$$\kappa_h \equiv \int_0^1 h(t)\mathsf{E}[\mathcal{B}^2(t)]\,dt = \int_0^1 h(t)t(1-t)\,dt = \bar{H} - 2\hat{H}, \tag{30}$$

after integration by parts twice and some algebra. Then the weighting function $h(t)/\kappa_h$ normalizes the CvM estimator in the sense that $\mathsf{E}[C(h/\kappa_h)] = \sigma^2$. Thus, we have

$$\mathsf{E}[C(h/\kappa_h;n)] = \sigma^2 + \frac{((H/\kappa_h)-1)\gamma}{n} + o(1/n),$$

so that

$$\mathsf{E}[C(h;n)] = \kappa_h\sigma^2 + \frac{(H-\kappa_h)\gamma}{n} + o(1/n). \tag{31}$$

Similarly, we can derive an expression for $\mathsf{E}[A(h;n)]$. To this end, consider

$$\mathsf{E}[A(h)] = \mathsf{Var}(S(h)) = \eta_h\sigma^2,$$

where

$$\eta_\xi \equiv \int_0^1\int_0^1 \xi(s)\xi(t)\mathsf{Cov}(\mathcal{B}(s),\mathcal{B}(t))\,ds\,dt,$$

for an arbitrary weighting function $\xi(t)$. Then the weighting function $h(t)/\sqrt{\eta_h}$ normalizes the area estimator in the sense that $\mathsf{E}[A(h/\sqrt{\eta_h})] = \sigma^2$. Thus, we have

$$\mathsf{E}[A(h/\sqrt{\eta_h};n)] = \sigma^2 + \frac{[(H-\bar{H})^2 + \bar{H}^2]\gamma}{2n\eta_h} + o(1/n),$$

so that

$$\mathsf{E}[A(h;n)] = \eta_h\sigma^2 + \frac{[(H-\bar{H})^2 + \bar{H}^2]\gamma}{2n} + o(1/n). \tag{32}$$

By exactly the same reasoning, we obtain

$$\mathsf{E}[A(f_0+h;n)] = \eta_{f_0+h}\sigma^2 + \frac{[(F_0+H-\bar{F}_0-\bar{H})^2 + (\bar{F}_0+\bar{H})^2]\gamma}{2n} + o(1/n), \tag{33}$$

where, since $f_0(t) = \sqrt{12}\ \forall t$,

$$\begin{aligned}\eta_{f_0+h} &= \int_0^1\int_0^1 (f_0(s)+h(s))(f_0(t)+h(t))\mathsf{Cov}(\mathcal{B}(s),\mathcal{B}(t))\,ds\,dt \\ &= \eta_{f_0} + \eta_h + 2\sqrt{12}\int_0^1\int_0^1 h(s)\mathsf{Cov}(\mathcal{B}(s),\mathcal{B}(t))\,ds\,dt \\ &= 1 + \eta_h + 2\sqrt{12}\left[\int_0^1\int_0^t h(s)s(1-t)\,ds\,dt \right. \\ &\qquad \left. + \int_0^1\int_t^1 h(s)t(1-s)\,ds\,dt\right] \\ &= 1 + \eta_h + \sqrt{12}(\bar{H} - 2\hat{H}),\end{aligned} \tag{34}$$

after integration by parts.

Finally, we substitute (31), (32), and (33) into Eq. (29) to get

$$
\begin{aligned}
&\mathsf{E}[D(h;n)]\\
&= \mathsf{E}[C(h;n)] + \frac{\mathsf{E}[A(h;n)] - \mathsf{E}[A(f_0+h;n)]}{\sqrt{12}}\\
&\quad + \left(1 + \frac{1}{\sqrt{12}} + \frac{h(1)-h(0)}{24n} + o(1/n)\right)\mathsf{E}[A(f_0;n)]\\
&= \left[\kappa_h\sigma^2 + \frac{(H-\kappa_h)\gamma}{n}\right] + \frac{1}{\sqrt{12}}\left[\eta_h\sigma^2 + \frac{[(H-\bar{H})^2+\bar{H}^2]\gamma}{2n}\right]\\
&\quad - \frac{1}{\sqrt{12}}\left[\eta_{f_0+h}\sigma^2 + \frac{[(F_0+H-\bar{F}_0-\bar{H})^2+(\bar{F}_0+\bar{H})^2]\gamma}{2n}\right]\\
&\quad + \left(1 + \frac{1}{\sqrt{12}} + \frac{h(1)-h(0)}{24n}\right)\left[\sigma^2 + \frac{3\gamma}{n}\right] + o(1/n)\\
&= \sigma^2\left[\kappa_h + \frac{\eta_h}{\sqrt{12}} - \frac{\eta_{f_0+h}}{\sqrt{12}} + 1 + \frac{1}{\sqrt{12}} + \frac{h(1)-h(0)}{24n}\right]\\
&\quad + \frac{(9-\kappa_h)\gamma}{n} + o(1/n)\\
&\qquad \text{(after algebra, noting that } H = 12,\ F_0 = \sqrt{12}, \text{ and } \bar{F}_0 = \sqrt{3}). \qquad (35)
\end{aligned}
$$

The result for $\mathsf{E}[D(h;n)]$ follows by substituting (30) and (34) into Eq. (35).

To obtain the variance result, note that

$$
\mathsf{Var}(D(h)) = \sigma^4 \int_0^1\int_0^1 h(s)h(t)\mathsf{Cov}((\mathcal{B}(s)-\bar{\mathcal{B}})^2, (\mathcal{B}(t)-\bar{\mathcal{B}})^2)\,ds\,dt
$$

and apply Lemma 2 and symmetry. □

PROOF OF THEOREM 5: To begin, we define a Brownian bridge on $[0,r]$ as $\mathcal{B}([0,r];t) \equiv (\mathcal{W}(rt) - t\mathcal{W}(r))/\sqrt{r}$. Under Assumption FCLT, the standardized time series based on rn observations converges to this process as $n \to \infty$ (with r fixed, $0 < r < 1$), i.e., $T_{rn}(t) \Rightarrow \mathcal{B}([0,r];t)$. Using this notation, we see that

$$
\begin{aligned}
&A(f_0;rn)\\
&\quad \Rightarrow 12\sigma^2\left(\int_0^1 \mathcal{B}([0,r];t)\,dt\right)^2 = \frac{12\sigma^2}{r}\left(\int_0^1 \mathcal{W}(rt)\,dt - \frac{\mathcal{W}(r)}{2}\right)^2\\
&C(g_0;rn) \Rightarrow 6\sigma^2\int_0^1 \mathcal{B}^2([0,r];t)\,dt = \frac{6\sigma^2}{r}\int_0^1 (\mathcal{W}(rt) - t\mathcal{W}(r))^2\,dt.
\end{aligned}
$$

Before proving the result of the theorem, we state a few intermediate results that will be useful along the way. First, as $n \to \infty$, Lemma 2 gives

$$
\begin{aligned}
&\mathsf{Cov}(A(f_0;n), A(f_0;rn))\\
&\quad \to \frac{144\sigma^4}{r}\mathsf{Cov}\left(\left(\int_0^1 \mathcal{W}(t)\,dt - \frac{\mathcal{W}(1)}{2}\right)^2, \left(\int_0^1 \mathcal{W}(rs)\,ds - \frac{\mathcal{W}(r)}{2}\right)^2\right)\\
&\quad = \frac{288\sigma^4}{r}\mathsf{Cov}^2\left(\int_0^1 \mathcal{W}(t)\,dt - \frac{\mathcal{W}(1)}{2}, \int_0^1 \mathcal{W}(rs)\,ds - \frac{\mathcal{W}(r)}{2}\right)\\
&\quad = \frac{288\sigma^4}{r}\left[\int_0^1\int_0^1 \mathsf{Cov}(\mathcal{W}(t),\mathcal{W}(rs))\,dt\,ds\right.\\
&\qquad\qquad - \frac{1}{2}\int_0^1 \mathsf{Cov}(\mathcal{W}(t),\mathcal{W}(r))\,dt\\
&\qquad\qquad \left. - \frac{1}{2}\int_0^1 \mathsf{Cov}(\mathcal{W}(1),\mathcal{W}(rs))\,ds + \frac{r}{4}\right]^2\\
&\quad = \frac{288\sigma^4}{r}\left[\left\{\int_0^r\int_0^{t/r} + \int_0^r\int_{t/r}^1 + \int_r^1\int_0^1\right\}\mathsf{Cov}(\mathcal{W}(t),\mathcal{W}(rs))\,ds\,dt\right.\\
&\qquad\qquad - \frac{1}{2}\left\{\int_0^r + \int_r^1\right\}\mathsf{Cov}(\mathcal{W}(t),\mathcal{W}(r))\,dt\\
&\qquad\qquad \left. - \frac{1}{2}\int_0^1 \mathsf{Cov}(\mathcal{W}(1),\mathcal{W}(rs))\,ds + \frac{r}{4}\right]^2\\
&\qquad \text{(where we use some abuse of integral notation)}\\
&\quad = \frac{288\sigma^4}{r}\left[\int_0^r\int_0^{t/r} rs\,ds\,dt + \int_0^r\int_{t/r}^1 t\,ds\,dt + \int_r^1\int_0^1 rs\,ds\,dt\right.\\
&\qquad\qquad \left. - \frac{1}{2}\int_0^r t\,dt - \frac{1}{2}\int_r^1 r\,dt - \frac{1}{2}\int_0^1 rs\,ds + \frac{r}{4}\right]^2\\
&\quad = 2r^3\sigma^4. \qquad (36)
\end{aligned}
$$

Second, using similar machinations, we have

$$
\begin{aligned}
&\mathsf{Cov}(A(f_0;n), C(g_0;rn))\\
&\quad \to \frac{72\sigma^4}{r}\mathsf{Cov}\left(\left(\int_0^1 \mathcal{W}(t)\,dt - \frac{\mathcal{W}(1)}{2}\right)^2, \int_0^1 (\mathcal{W}(rs) - s\mathcal{W}(r))^2\,ds\right)\\
&\quad = \frac{144\sigma^4}{r}\int_0^1\left[\mathsf{Cov}\left(\int_0^1 \mathcal{W}(t)\,dt - \frac{\mathcal{W}(1)}{2}, \mathcal{W}(rs) - s\mathcal{W}(r)\right)\right]^2 ds\\
&\quad = 6r^3\sigma^4/5. \qquad (37)
\end{aligned}
$$

Third,

$$
\begin{aligned}
&\mathsf{Cov}(C(g_0;n), A(f_0;rn))\\
&\quad \to \frac{72\sigma^4}{r}\mathsf{Cov}\left(\int_0^1 (\mathcal{W}(t) - t\mathcal{W}(1))^2\,dt, \left(\int_0^1 \mathcal{W}(rs)\,ds - \frac{\mathcal{W}(r)}{2}\right)^2\right)\\
&\quad = \frac{144\sigma^4}{r}\int_0^1\left[\mathsf{Cov}\left(\mathcal{W}(t) - t\mathcal{W}(1), \int_0^1 \mathcal{W}(rs)\,ds - \frac{\mathcal{W}(r)}{2}\right)\right]^2 dt\\
&\quad = 6r^2\sigma^4/5. \qquad (38)
\end{aligned}
$$

Fourth,

$$
\begin{aligned}
&\mathsf{Cov}(C(g_0;n), C(g_0;rn))\\
&\quad \to \frac{36\sigma^4}{r}\mathsf{Cov}\left(\int_0^1 (\mathcal{W}(t) - t\mathcal{W}(1))^2\,dt, \int_0^1 (\mathcal{W}(rs) - s\mathcal{W}(r))^2\,ds\right)\\
&\quad = \frac{72\sigma^4}{r}\int_0^1\int_0^1\left[\mathsf{Cov}(\mathcal{W}(t) - t\mathcal{W}(1), \mathcal{W}(rs) - s\mathcal{W}(r))\right]^2 ds\,dt\\
&\quad = 4r^2\sigma^4/5. \qquad (39)
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&(1-r)^2\,\mathsf{Var}(D_{J,r}(h_0;n))\\
&\quad = \mathsf{Var}(D(h_0;n)) + r^2\,\mathsf{Var}(D(h_0;rn)) - 2r\,\mathsf{Cov}(D(h_0;n), D(h_0;rn))\\
&\quad = \mathsf{Var}(D_0(n)) + r^2\,\mathsf{Var}(D_0(rn))\\
&\qquad - 2r\,\mathsf{Cov}(2C(g_0;n) - A(f_0;n),\ 2C(g_0;rn) - A(f_0;rn))\\
&\quad = \mathsf{Var}(D_0(n)) + r^2\,\mathsf{Var}(D_0(rn)) - 8r\,\mathsf{Cov}(C(g_0;n), C(g_0;rn))\\
&\qquad + 4r\,\mathsf{Cov}(C(g_0;n), A(f_0;rn)) + 4r\,\mathsf{Cov}(A(f_0;n), C(g_0;rn))\\
&\qquad - 2r\,\mathsf{Cov}(A(f_0;n), A(f_0;rn)), \qquad (40)
\end{aligned}
$$

and Eq. (12) follows by Example 5 and Eqs. (36)–(39). □

ACKNOWLEDGMENTS

We thank the referees and Associate Editor for their substantive comments. This work was partially supported by National Science Foundation Grant DMI-0400260.

REFERENCES

[1] T. Aktaran-Kalaycı, C. Alexopoulos, N.T. Argon, D. Goldsman, and J.R. Wilson, Exact expected values of variance estimators in steady-state computer simulation, Nav Res Logist, forthcoming.

[2] T. Aktaran-Kalayci, D. Goldsman, and J.R. Wilson, Exact mean squared errors of variance estimators in steady-state computer simulation, Technical Report, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 2006.

[3] C. Alexopoulos, N.T. Argon, D. Goldsman, G. Tokol, and J.R. Wilson, Overlapping variance estimators for simulation, Oper Res, forthcoming.

[4] C. Alexopoulos, D. Goldsman, and G. Tokol, Properties of batched quadratic-form variance parameter estimators for simulations, INFORMS J Comput 13 (2001), 149–156.

[5] P. Billingsley, Convergence of probability measures, Wiley, New York, 1968.

[6] C. Chien, D. Goldsman, and B. Melamed, Large-sample results for batch means, Manage Sci 43 (1997), 1288–1295.

[7] H. Damerdji, Mean-square consistency of the variance estimator in steady-state simulation output analysis, Oper Res 43 (1995), 282–291.

[8] J. Durbin, Distribution theory for tests based on the sample distribution function, Society for Industrial and Applied Mathematics, Philadelphia, 1973.

[9] R.D. Foley and D. Goldsman, Confidence intervals using orthonormally weighted standardized time series, ACM Trans Model Comput Simul 9 (1999), 297–325.

[10] P.W. Glynn and D.L. Iglehart, Simulation output analysis using standardized time series, Math Oper Res 15 (1999), 1–16.

[11] P.W. Glynn and W. Whitt, Estimating the asymptotic variance with batch means, Oper Res Lett 10 (1991), 431–435.

[12] D. Goldsman, K. Kang, S.-H. Kim, A.F. Seila, and G. Tokol, Online companion to "Combining standardized time series area and Cramér–von Mises variance estimators," Technical Report, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA. Available online at www.interscience.wiley.com/jpages/0894-069X/suppmat.

[13] D. Goldsman, K. Kang, and A.F. Seila, Cramér–von Mises variance estimators for simulations, Oper Res 47 (1999), 299–309.

[14] D. Goldsman and M.S. Meketon, A comparison of several variance estimators, Technical Report J-85-12, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 1986.

[15] D. Goldsman, M.S. Meketon, and L.W. Schruben, Properties of standardized time series weighted area variance estimators, Manage Sci 36 (1990), 602–612.

[16] D. Goldsman and L.W. Schruben, New confidence interval estimators using standardized time series, Manage Sci 36 (1990), 393–397.

[17] A.M. Law and W.D. Kelton, Simulation modeling and analysis (3rd edition), McGraw–Hill, New York, 2000.

[18] M.S. Meketon and B.W. Schmeiser, Overlapping batch means: Something for nothing?, Proceedings of the 1984 Winter Simulation Conference, S. Sheppard, U.W. Pooch, and C.D. Pegden (Editors), Institute of Electrical and Electronics Engineers, Piscataway, NJ, 1984, pp. 227–230.

[19] J.K. Patel and C.B. Read, Handbook of the normal distribution (2nd edition), Dekker, New York, 1996.

[20] B.W. Schmeiser, Batch size effects in the analysis of simulation output, Oper Res 30 (1982), 556–568.

[21] L.W. Schruben, Confidence interval estimation using standardized time series, Oper Res 31 (1983), 1090–1108.

[22] W.-M.T. Song and B.W. Schmeiser, Optimal mean-squared-error batch sizes, Manage Sci 41 (1995), 110–123.

[23] N.M. Steiger and J.R. Wilson, An improved batch means procedure for simulation output analysis, Manage Sci 48 (2001), 1569–1586.

[24] G.S. Watson, Goodness-of-fit tests on a circle, Biometrika 48 (1961), 109–114.