An Improved Standardized Time Series Durbin–Watson Variance Estimator for Steady-State Simulation

Demet Batur[a,1], David Goldsman[b], Seong-Hee Kim[b]

[a] *Industrial and Management Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA*
[b] *H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA*

Abstract

We discuss an improved jackknifed Durbin–Watson estimator for the variance parameter from a steady-state simulation. The estimator is based on a combination of standardized time series area and Cramér–von Mises estimators. Various examples demonstrate its efficiency in terms of bias and variance compared to other estimators.

Keywords: Simulation; Stationary Process; Variance Estimation; Standardized Time Series; Durbin–Watson Estimator; Batched Estimators

1. Introduction

Consider a steady-state simulation output process, $\boldsymbol{Y} \equiv \{Y_1, Y_2, \ldots, Y_n\}$, where we typically estimate the unknown mean, μ, by $\bar{Y}_n$, the sample average of the first n observations. In order to give a measure of the precision of $\bar{Y}_n$ or to build a confidence interval for μ, we can also estimate the *variance parameter*, $\sigma^2 \equiv \lim_{n\to\infty} n\text{Var}(\bar{Y}_n)$. There are a number of different techniques in the literature devoted to the estimation of σ^2, e.g., the methods of nonoverlapping batch means (NBM) [16], overlapping batch means (OBM) [15], and standardized time series (STS) [17]. Among the estimators based on the STS methodology are the so-called area [14] and Cramér–von Mises (CvM) [12] estimators. Goldsman et al. [11] combine these two types of estimators to obtain new STS estimators—the Durbin–Watson (DW) and jackknifed DW (JDW)—with competitive bias and lower asymptotic variance than many of their competitors. In this paper, we improve the JDW estimator in such a way that the resulting modified jackknifed Durbin–Watson (MJDW) estimator has an asymptotic variance even smaller than that of the JDW estimator while maintaining almost the same bias value. The interesting story here is that one can "re-use" data to come up with more-efficient variance estimators. For instance, [3] and [15] use each individual observation over and over in the computation of various overlapping estimators; and [2] forms estimators by linearly combining estimators that differ only in the batch sizes used. In the current paper, we shall re-use data by linearly combining different types of estimators for σ^2—the area and CvM—along with different batch sizes.

The paper is organized as follows: §2 outlines background about the area, CvM, DW, and JDW estimators. §3 defines and gives properties of the improved estimator. §4 discusses batched versions of all the estimators, including NBM and OBM. §5 provides a Monte Carlo example illustrating estimator performance, along with a summary. The Online Companion [5] gives a proof of the paper's main result (Theorem 1) and additional performance examples.

[1] Corresponding author.
E-mail addresses: `dbatur@gmail.com` (Batur), `sman@gatech.edu` (Goldsman), `skim@isye.gatech.edu` (Kim).

2. Background

Henceforth, we assume that the output data $\boldsymbol{Y}$ are from a stationary stochastic process (e.g., a steady-state simulation) that satisfies the following.

Standing Assumptions 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 [6], and

$$X_n(t) \;\equiv\; \frac{\lfloor nt \rfloor (\bar{Y}_{\lfloor nt \rfloor} - \mu)}{\sqrt{n}} \quad \text{for} \quad t \in [0,1],$$

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 also let $R_k \equiv \mathrm{Cov}(Y_1, Y_{1+k})$, $k = 0, 1, 2, \ldots$, and assume that $|R_k| = O(\delta^k)$ for some $\delta \in (0,1)$, so that the constants $\gamma_j \equiv 2\sum_{k=1}^{\infty} k^j R_k$, $j = 0, 1, 2, \ldots$, are well defined [1, 18].

Before reviewing various estimators for σ^2, we first define the STS of $\boldsymbol{Y}$,

$$T_n(t) \;\equiv\; \frac{\lfloor nt \rfloor (\bar{Y}_n - \bar{Y}_{\lfloor nt \rfloor})}{\sigma\sqrt{n}} \quad \text{for} \quad t \in [0,1].$$

Under the Standing Assumptions, [10] and [17] show that $T_n \Rightarrow \mathcal{B}$, a Brownian bridge on $[0,1]$, i.e., a Gaussian process with $\mathrm{E}[\mathcal{B}(t)] = 0$ and $\mathrm{Cov}(\mathcal{B}(s), \mathcal{B}(t)) = \min(s,t) - st$. The cited references use this fact to show that all of the variance estimators considered herein converge to limiting random variables having expectation σ^2. We shall also assume uniform integrability [6] of the squares of the estimators, thereby establishing the limiting variance of each estimator.

As in [14], the STS *area estimator* for σ^2 and its limiting functional are given by

$$A(f;n) \;\equiv\; \left(\frac{1}{n}\sum_{k=1}^{n} f(\tfrac{k}{n})\sigma T_n(\tfrac{k}{n})\right)^2 \;\Rightarrow\; A(f) \;\equiv\; \left(\int_0^1 f(t)\sigma\mathcal{B}(t)\,\mathrm{d}t\right)^2 \quad \text{as } n \to \infty,$$

where the weight $f(t)$ is normalized so that $\mathrm{E}[A(f)] = \sigma^2$ and $\frac{\mathrm{d}^2}{\mathrm{d}t^2}f(t)$ is continuous on $[0,1]$.

Example 1. We consider area estimators with two weight functions from the literature [14, 17]: $f_0(t) \equiv \sqrt{12}$ and $f_2(t) \equiv \sqrt{840}(3t^2 - 3t + 1/2)$ for $t \in [0,1]$. From [1], we have

$$\mathrm{E}[A(f_0;n)] \;=\; \sigma^2 - \frac{3\gamma_1}{n} - \frac{\sigma^2}{n^2} + O(\frac{1}{n^3}), \tag{1}$$

$$\mathrm{E}[A(f_2;n)] \;=\; \sigma^2 + \frac{7(\sigma^2 - 6\gamma_2)}{2n^2} + O(\frac{1}{n^3}). \tag{2}$$

Note that the area estimator with weight $f_2(t)$ is first-order unbiased for σ^2. Further, it can be shown that all area estimators have asymptotic variance $2\sigma^4$ as $n \to \infty$.

As in [12], the STS *CvM estimator* for σ^2 and its limiting functional are

$$C(g;n) \;\equiv\; \frac{1}{n}\sum_{k=1}^{n} g(\tfrac{k}{n})\sigma^2 T_n^2(\tfrac{k}{n}) \;\Rightarrow\; C(g) \;\equiv\; \int_0^1 g(t)\sigma^2\mathcal{B}^2(t)\,\mathrm{d}t \quad \text{as } n \to \infty,$$

where the weight $g(t)$ is normalized so that $\mathrm{E}[C(g)] = \sigma^2$ and $\frac{\mathrm{d}^2}{\mathrm{d}t^2}g(t)$ is continuous on $[0,1]$.

Example 2. We consider three weights from [12]: $g_0(t) \equiv 6$, $g_2^\star(t) \equiv -24 + 150t - 150t^2$, and

$$g_4^\star(t) \;\equiv\; \frac{-1310}{21} + \frac{19270t}{21} - \frac{25230t^2}{7} + \frac{16120t^3}{3} - \frac{8060t^4}{3}, \quad \text{for } t \in [0,1],$$

the latter two of which are known to minimize asymptotic variance among all first-order unbiased CvM estimators having quadratic and quartic weights. From [1], we have

$$\mathrm{E}[C(g_0;n)] \;=\; \sigma^2 - \frac{5\gamma_1}{n} + \frac{6\gamma_2 - \sigma^2}{n^2} + O(\frac{1}{n^3}) \quad \text{and} \quad \mathrm{Var}(C(g_0)) \;=\; 0.8\sigma^4, \tag{3}$$

$$\mathrm{E}[C(g_2^\star;n)] \;=\; \sigma^2 + \frac{4(\sigma^2 - 6\gamma_2)}{n^2} + O(\frac{1}{n^3}) \quad \text{and} \quad \mathrm{Var}(C(g_2^\star)) = 1.729\sigma^4, \tag{4}$$

$$\mathrm{E}[C(g_4^\star;n)] \;=\; \sigma^2 + \frac{655(\sigma^2 - 6\gamma_2)}{63n^2} + O(\frac{1}{n^3}) \quad \text{and} \quad \mathrm{Var}(C(g_4^\star)) = 1.042\sigma^4. \tag{5}$$

As a special case of [11], the STS *DW estimator* for σ^2 and its limiting functional are

$$D(n) \;\equiv\; 2C(g_0;n) - A(f_0;n) \;\Rightarrow\; D \;\equiv\; 2C(g_0) - A(f_0) \quad \text{as } n \to \infty. \tag{6}$$

Using (1), (3), (6), and the work in [11], we have

$$\mathrm{E}[D(n)] = \sigma^2 - \frac{7\gamma_1}{n} + \frac{12\gamma_2 - \sigma^2}{n^2} + O(\frac{1}{n^3}) \quad \text{and} \quad \mathrm{Var}(D(n)) \to 2\sigma^4/5 \quad \text{as } n \to \infty. \tag{7}$$

The DW estimator $D(n)$ has low asymptotic variance but high bias. To reduce bias while maintaining small variance, a "jackknifed" (JDW) estimator is introduced in [11],

$$D_{\mathrm{J},r}(n) \;\equiv\; \frac{D(n)}{1-r} - \frac{rD(rn)}{1-r}, \tag{8}$$

where r is fixed in (0,1), and where we assume for convenience that rn is an integer.

Example 3. For the choice $r = 0.5$, Equations (7) and (8) and the work in [11] imply

$$\mathrm{E}[D_{\mathrm{J},0.5}(n)] = \sigma^2 + \frac{2(\sigma^2 - 12\gamma_2)}{n^2} + O(\frac{1}{n^3}) \quad \text{and} \quad \mathrm{Var}(D_{\mathrm{J},0.5}(n)) \to 1.4\sigma^4 \quad \text{as } n \to \infty. \tag{9}$$

3. The Modified Jackknifed Durbin–Watson Estimator

The main result of this paper is a new estimator that is first-order unbiased for σ^2 but has smaller asymptotic variance than the JDW estimator. Let $\widetilde{D}(s)$, $s = 1, 2, \ldots, n$, be the DW estimator for σ^2 based on the *last* s observations out of the original n. We can generalize the JDW estimator $D_{\mathrm{J},r}(n)$ from Equation (8) by including $\widetilde{D}((1-r)n)$ in the mix, thus yielding the *modified jackknifed Durbin–Watson* (MJDW) estimator for σ^2,

$$\widetilde{D}_{\mathrm{J},r}(n) \;\equiv\; \beta_1 D(n) + \beta_2 D(rn) + \beta_3 \widetilde{D}((1-r)n), \tag{10}$$

where

$$\beta_1 \;\equiv\; \frac{-1}{2(4r^4 - 8r^3 + 2r^2 + 2r - 1)}, \;\; \beta_2 \;\equiv\; \frac{r(4r^3 - 6r^2 - r + 2)}{2(4r^4 - 8r^3 + 2r^2 + 2r - 1)}, \tag{11}$$

$\beta_3 \equiv 1 - \beta_1 - \beta_2$, and r is fixed in $(0,1)$. The β's are determined so that the MJDW estimator is first-order unbiased. In fact, we have the following result, proven in our Online Companion.

Theorem 1. If $\boldsymbol{Y}$ is a stationary process for which the Standing Assumptions hold, then

$$\mathrm{E}[\widetilde{D}_{\mathrm{J},r}(n)] \;=\; \sigma^2 + \left(\frac{8r^4 - 16r^3 + 3r^2 + 5r - 2}{2r(1-r)(4r^4 - 8r^3 + 2r^2 + 2r - 1)}\right)\frac{(\sigma^2 - 12\gamma_2)}{n^2} + O(\frac{1}{n^3}). \quad (12)$$

Further, assuming that the family $\{\widetilde{D}^2_{\mathrm{J},r}(n) : n = 1, 2, \ldots\}$ is uniformly integrable, then

$$\mathrm{Var}(\widetilde{D}_{\mathrm{J},r}(n)) \;\to\; \left(\frac{4r^4 - 8r^3 + 3r^2 + r - 2}{4r^4 - 8r^3 + 2r^2 + 2r - 1}\right)\frac{\sigma^4}{5} \quad \text{as } n \;\to\; \infty. \quad (13)$$

Example 4. As in Example 3, we take $r = 0.5$, so that the estimator takes the intuitive form $\widetilde{D}_{\mathrm{J},0.5}(n) = 2D(n) - \frac{1}{2}\big(D(n/2) + \widetilde{D}(n/2)\big)$. In this case, (9) and (12) imply

$$\mathrm{E}[\widetilde{D}_{\mathrm{J},0.5}(n)] \;=\; \mathrm{E}[D_{\mathrm{J},0.5}(n)] \;=\; \sigma^2 + \frac{2(\sigma^2 - 12\gamma_2)}{n^2} + O(\frac{1}{n^3}). \quad (14)$$

The choice $r = 0.5$ leads to the minimum asymptotic variance for the MJDW estimator, $\mathrm{Var}(\widetilde{D}_{\mathrm{J},0.5}(n)) \to 1.2\sigma^4$, which compares favorably to $\mathrm{Var}(D_{\mathrm{J},0.5}(n)) \to 1.4\sigma^4$.

4. Batching

Batching is a way to reduce the variance of a variance estimator—although at the cost of a possible increase in its bias. This section discusses various estimators that employ batching. To begin, we divide a long run of n stationary observations into b nonoverlapping batches, each of size m (assuming $n = bm$), so that batch i consists of observations $Y_{(i-1)m+1}, \ldots, Y_{im}$, $i = 1, 2, \ldots, b$. Then we form an estimator from each batch and take the sample average of the estimators to obtain an overall "batched" STS estimator for the variance parameter.

To put this on firmer footing, let $\mathcal{V}_i(b, m)$ denote a generic STS estimator computed from the ith batch of size m, $i = 1, 2, \ldots, b$, and suppose the resulting batched estimator is of the form $\mathcal{V}(b, m) \equiv \frac{1}{b}\sum_{i=1}^{b} \mathcal{V}_i(b, m)$. Since the batches are stationary, we have

$$\mathrm{E}[\mathcal{V}(b, m)] \;=\; \frac{1}{b}\sum_{i=1}^{b} \mathrm{E}[\mathcal{V}_i(b, m)] \;=\; \mathrm{E}[\mathcal{V}(1, m)]. \quad (15)$$

Thus, the expected value of a batched STS estimator $\mathcal{V}(b, m)$ is the same as that of the estimator $\mathcal{V}(1, m)$ based on only one batch of size m; and so (1)–(5), (7), (9), and (12) with m in place of n show that the batched estimator likely has larger bias than the original unbatched estimator $\mathcal{V}(1, n)$ based on one batch of size $n > m$. Even so, all of the batched STS estimators examined herein are asymptotically unbiased as $m \to \infty$.

For large enough batch size m, we can assume that STS estimators from two disjoint batches are approximately uncorrelated (see [4]). Then for fixed b, as m becomes large,

$$\mathrm{Var}(\mathcal{V}(b, m)) \;\approx\; \frac{\mathrm{Var}(\mathcal{V}(1, m))}{b} \;\approx\; \frac{\mathrm{Var}(\mathcal{V}(1, n))}{b}, \quad (16)$$

where the last expression follows by the fact that $\mathrm{Var}(\mathcal{V}(1, m))$ (and hence $\mathrm{Var}(\mathcal{V}(1, n))$) converges to a constant as $n \to \infty$ (as shown in Examples 1–4 and Equations (7) and (13)). Thus, the variance of a batched STS estimator $\mathcal{V}(b, m)$ is approximately a factor of b smaller than that of the original unbatched estimator $\mathcal{V}(1, n)$.

Now let $A_i(f;b,m)$, $C_i(g;b,m)$, $D_i(b,m)$, $D_{\mathrm{J},r,i}(b,m)$, and $\widetilde{D}_{\mathrm{J},r,i}(b,m)$ denote the area, CvM, DW, JDW, and MJDW estimators applied to the ith batch of size m, respectively; and let $\mathcal{A}(f;b,m)$, $\mathcal{C}(g;b,m)$, $\mathcal{D}(b,m)$, $\mathcal{D}_{\mathrm{J},r}(b,m)$, and $\widetilde{\mathcal{D}}_{\mathrm{J},r}(b,m)$ denote the corresponding batched estimators, i.e., the sample means of the estimators from each batch. Thus, for example, $\mathcal{A}(f;b,m) = \frac{1}{b}\sum_{i=1}^{b} A_i(f;b,m)$.

For comparison purposes, we define the classic NBM estimator for σ^2 as $\mathcal{N}(b,m) \equiv \frac{m}{b-1}\sum_{i=1}^{b}(\bar{Y}_{i,m}-\bar{Y}_n)^2$, where $\bar{Y}_{i,m} \equiv \sum_{j=1}^{m} Y_{(i-1)m+j}/m$, $i=1,2,\ldots,b$, are the batch means of the process. From [7], [13], and [18], we have

$$\mathrm{E}[\mathcal{N}(b,m)] \;=\; \sigma^2 - \frac{\gamma_1(b+1)}{bm} + o(\frac{1}{m}) \quad \text{and} \quad \mathrm{Var}(\mathcal{N}(b,m)) \;\to\; \frac{2\sigma^4}{b-1} \quad \text{as } m\to\infty.$$

In addition, the OBM estimator for σ^2 is $\mathcal{O}(b,m) \equiv \frac{nm}{(n-m+1)(n-m)}\sum_{i=1}^{n-m+1}(\bar{Y}^{\mathrm{O}}_{i,m}-\bar{Y}_n)^2$, where $\bar{Y}^{\mathrm{O}}_{i,m} \equiv \sum_{k=0}^{m-1} Y_{i+k}/m$, $i=1,2,\ldots,n-m+1$, are the overlapping batch means. It can be shown ([3], [9], [15], [18]) that as $m\to\infty$,

$$\mathrm{E}[\mathcal{O}(b,m)] = \sigma^2 - \frac{\gamma_1(b^2+1)}{mb(b-1)} + o(\frac{1}{m}) \quad \text{and} \quad \mathrm{Var}(\mathcal{O}(b,m)) \to \frac{(4b^3-11b^2+4b+6)\sigma^4}{3(b-1)^4}.$$

As summarized in Table 1, all of the variance estimators studied herein are asymptotically unbiased as the batch size m increases. But for a particular variance estimator using a fixed sample size n, the act of increasing m (thus decreasing $b=n/m$) usually increases variance. Of interest here is that the NBM and OBM estimators tend to have more bias than first-order unbiased STS estimators—e.g., $\mathcal{A}(f_2;b,m)$, $\mathcal{C}(g_2^\star;b,m)$, $\mathcal{C}(g_4^\star;b,m)$, $\mathcal{D}_{\mathrm{J},0.5}(b,m)$, and our new estimator $\widetilde{\mathcal{D}}_{\mathrm{J},0.5}(b,m)$. Further, for comparable choices of b and m, the asymptotic variance of $\widetilde{\mathcal{D}}_{\mathrm{J},0.5}(b,m)$ is less than those of $\mathcal{N}(b,m)$, $\mathcal{O}(b,m)$, $\mathcal{A}(f_2;b,m)$, $\mathcal{C}(g_2^\star;b,m)$, and $\mathcal{D}_{\mathrm{J},0.5}(b,m)$.

Table 1 Approximate Bias and Asymptotic Variance for Different Estimators

Estimator	(m/γ_1) Bias	(b/σ^4) Var
$\mathcal{A}(f_0;b,m)$	3	2
$\mathcal{A}(f_2;b,m)$	$o(1)$	2
$\mathcal{C}(g_0;b,m)$	5	0.8
$\mathcal{C}(g_2^\star;b,m)$	$o(1)$	1.729
$\mathcal{C}(g_4^\star;b,m)$	$o(1)$	1.042
$\mathcal{D}(b,m)$	7	0.4
$\mathcal{D}_{\mathrm{J},0.5}(b,m)$	$o(1)$	1.4
$\widetilde{\mathcal{D}}_{\mathrm{J},0.5}(b,m)$	$o(1)$	1.2
$\mathcal{N}(b,m)$	$(b+1)/b$	2
$\mathcal{O}(b,m)$	$(b^2+1)/[b(b-1)]$	1.333

5. Monte Carlo Example and Summary

This section gives Monte Carlo results involving a stationary $M/M/1$ queue-waiting-time process to compare the performance of the new estimator with that of its competitors. The $M/M/1$ is a single-server queueing system experiencing Poisson arrivals at rate λ and first-in-first-out i.i.d. exponential service times at rate ω, so that the traffic intensity is $\rho \equiv \lambda/\omega < 1$. For the queue-waiting-time process we have from [8] that $\sigma^2 = \rho^3(2+5\rho-4\rho^2+\rho^3)/[\lambda^2(1-\rho)^4]$. Here we set $\lambda = 0.8$ and $\omega = 1.0$, which gives a highly autocorrelated process with $\sigma^2 = 1976$.

We evaluated estimator performance based on 100,000 independent replications of $\boldsymbol{Y}$, each of which ran for a sample of size $n = 32768$. For each replication, we computed various area, CvM, DW, JDW, MJDW, NBM, and OBM estimators, after having divided the run using $b = n/m = 1$, 2, 4, 8, and 16 batches (with respective batch sizes $m = 32768$, 16384, 8192, 4096, and 2048). To estimate the expectation and variance of each of these estimators, we recorded the sample mean ($\widehat{\mathrm{E}}$) and sample variance ($\widehat{\mathrm{Var}}$) from the 100,000 replications; see Table 2, where the $\widehat{\mathrm{E}}$ values have standard errors of $(\widehat{\mathrm{Var}}/100000)^{1/2}$.

Table 2
Estimated mean and variance ($\times 10^{-3}$) of various batched estimators for the variance parameter of an $M/M/1$ queue-wait-time process with $\rho = 0.8$ ($\sigma^2 = 1976$) and $n = 32768$.

Estimator	$b=1$, $m=32768$		$b=2$, $b=16384$		$b=4$, $m=8192$		$b=8$, $m=4096$		$b=16$, $m=2048$	
	$\widehat{\mathrm{E}}$	$\widehat{\mathrm{Var}}$	$\widehat{\mathrm{E}}$	$\widehat{\mathrm{Var}}$	$\widehat{\mathrm{E}}$	$\widehat{\mathrm{Var}}$	$\widehat{\mathrm{E}}$	$\widehat{\mathrm{Var}}$	$\widehat{\mathrm{E}}$	$\widehat{\mathrm{Var}}$
$\mathcal{A}(f_0;b,m)$	1975	9894	1949	7892	1938	3988	1911	2770	1839	1890
$\mathcal{A}(f_2;b,m)$	1975	9618	1986	7739	1974	3827	1964	2725	1936	2115
$\mathcal{C}(g_0;b,m)$	1967	4598	1924	4183	1918	2122	1864	1534	1754	1046
$\mathcal{C}(g_2^\star;b,m)$	1977	8181	1979	6216	1974	3106	1965	2210	1931	1665
$\mathcal{C}(g_4^\star;b,m)$	1980	5834	1978	5433	1968	2759	1949	2103	1875	1424
$\mathcal{D}(b,m)$	1959	2890	1898	2876	1898	1463	1818	1069	1669	676
$\mathcal{D}_{\mathrm{J},0.5}(b,m)$	1985	8255	1984	7182	1975	3487	1968	2532	1922	1607
$\widetilde{\mathcal{D}}_{\mathrm{J},0.5}(b,m)$	1980	6052	1978	4878	1979	2452	1966	1813	1925	1329
$\mathcal{N}(b,m)$			1968	10270	1961	3894	1949	2338	1927	1682
$\mathcal{O}(b,m)$			1951	4946	1962	3052	1953	2090	1928	1569

Not surprisingly, as m increases, all the estimators become less biased for $\sigma^2 = 1976$, though some of the estimators do so more quickly than others. In particular, $\mathcal{A}(f_0;b,m)$, $\mathcal{C}(g_0;b,m)$, and $\mathcal{D}(b,m)$ perform poorly with respect to bias and will therefore be dropped from subsequent discussion, even though, for fixed b, the latter two are the least variable estimators in the list. Among the rest, as m becomes very large, $\mathcal{C}(g_4^\star;b,m)$ achieves the smallest variance, with $\widetilde{\mathcal{D}}_{\mathrm{J},0.5}(b,m)$ following close behind—in accordance with theory. Both $\mathcal{C}(g_4^\star;b,m)$ and $\widetilde{\mathcal{D}}_{\mathrm{J},0.5}(b,m)$ are first-order unbiased, though the $M/M/1$ empirical results as well as the analytical second-order bias results from (5) and (14) suggest that $\widetilde{\mathcal{D}}_{\mathrm{J},0.5}(b,m)$ may be less biased than $\mathcal{C}(g_4^\star;b,m)$ for "moderate" batch sizes (say, $m \le 8192$). In addition, $\widetilde{\mathcal{D}}_{\mathrm{J},0.5}(b,m)$ is significantly less variable and (at least for $m \ge 4096$) is a bit less biased than the benchmark $\mathcal{N}(b,m)$ and $\mathcal{O}(b,m)$ estimators. Note that all of the estimators are highly biased for "small" batch size ($m = 2048$). Thus, for batches of at least moderate size, $\widetilde{\mathcal{D}}_{\mathrm{J},0.5}(b,m)$ receives our recommendation.

Similar Monte Carlo results for a first-order autoregressive process are given in the Online Companion, as are *exact* results for a first-order moving average process—all of which show that the MJDW estimator performs as anticipated. Thus, the bottom line is that when we compare the performance of the batched MJDW estimator to the batched area, CvM, DW, and JDW estimators and the NBM and OBM estimators, we see that the MJDW estimator is at least competitive in terms of bias. In addition, the MJDW estimator seems to be better (or nearly better) than the other low-bias estimators under consideration in terms of variance.

Acknowledgments. We thank the referees, Associate Editor, and Area Editor for their suggestions. This research was supported by National Science Foundation grant DMI-0400260.

References

[1] Aktaran-Kalaycı, T., C. Alexopoulos, N. Argon, D. Goldsman, and J. R. Wilson. 2007. Exact Expected Values of Variance Estimators for Simulation. *Naval Res. Logistics*, 54:397–410.

[2] Aktaran-Kalaycı, T., D. Goldsman, and J. R. Wilson. 2007. Linear Combinations of Overlapping Variance Estimators for Simulation. *Operations Res. Letters*, 35:439–447.

[3] Alexopoulos, C., N. Argon, D. Goldsman, G. Tokol, and J. R. Wilson. 2007. Overlapping Variance Estimators for Simulation. *Operations Res.*, 55:1090–1103.

[4] Alexopoulos, C., D. Goldsman, and G. Tokol. 2001. Properties of Batched Quadratic-Form Variance Parameter Estimators for Simulations. *INFORMS J. on Computing*, 13:149–156.

[5] Batur, D., D. Goldsman, and S.-H. Kim. 2009. Online Companion to "An Improved Standardized Time Series Durbin–Watson Variance Estimator for Steady-State Simulation." http://www.isye.gatech.edu/∼sman/publications/MJDW.pdf.

[6] Billingsley, P. 1968. *Convergence of Probability Measures*. Wiley & Sons, New York.

[7] Chien, C., D. Goldsman, and B. Melamed. 1997. Large-Sample Results for Batch Means. *Management Sci.,* 43:1288–1295.

[8] Daley, D. 1968. The Serial Correlation Coefficients of Waiting Times in a Stationary Single Server Queue. *J. of the Australian Mathematical Society,* 8:683–699.

[9] Damerdji, H. 1995. Mean-Square Consistency of the Variance Estimator in Steady-State Simulation Output Analysis. *Operations Res.*, 43:282–291.

[10] Glynn, P. and D. Iglehart. 1990. Simulation Output Analysis Using Standardized Time Series. *Mathematics of Operations Res.*, 15:1–16.

[11] Goldsman, D., K. Kang, S.-H. Kim, A. Seila, and G. Tokol. 2007. Combining Standardized Time Series Area and Cramér–von Mises Variance Estimators. *Naval Res. Logistics*, 54:384–396.

[12] Goldsman, D., K. Kang, and A. Seila. 1999. Cramér–von Mises Variance Estimators for Simulations. *Operations Res.,* 47:299–309.

[13] Goldsman, D. and M. Meketon. 1986. A Comparison of Several Variance Estimators. Tech. Rep. J-85-12, School of ISyE, Georgia Institute of Technology, Atlanta.

[14] Goldsman, D., M. Meketon, and L. Schruben. 1990. Properties of Standardized Time Series Weighted Area Variance Estimators. *Management Sci.,* 36:602–612.

[15] Meketon, M. and B. Schmeiser. 1984. Overlapping Batch Means: Something for Nothing? In *Proc. 1984 Winter Simulation Conference*, ed. S. Sheppard, U. Pooch, and C. Pegden, 227–230. Piscataway, NJ: IEEE.

[16] Schmeiser, B. 1982. Batch Size Effects in the Analysis of Simulation Output. *Operations Res.,* 30:556–568.

[17] Schruben, L. 1983. Confidence Interval Estimation Using Standardized Time Series. *Operations Res.,* 31:1090–1108.

[18] Song, W. and B. Schmeiser. 1995. Optimal Mean-Squared-Error Batch Sizes. *Management Sci.,* 41:110–123.

Online Companion for 'An Improved Standardized Time Series Durbin–Watson Variance Estimator for Steady-State Simulation'

Demet Batur[a], David Goldsman[b], Seong-Hee Kim[b]

[a] *Industrial and Management Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA*

[b] *H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA*

In this Online Companion, the numbering of equations and tables is continued from the main text [A-2]. The organization of the Online Companion is as follows. In §6, we give the proof of Theorem 1. In order to supplement the findings in the main text with additional examples, §7 gives exact results for an example involving a first-order moving average process, and §8 gives Monte Carlo results for a first-order autoregressive process. §9 discusses an anomaly involving the possibility that the MJDW estimator $\widetilde{\mathcal{D}}_{\mathrm{J},0.5}(b,m)$ can occasionally produce negative realizations — the good news is that the probability of doing so vanishes quickly as we increase the number of batches.

6. Proof of Theorem 1

First, we prove the expectation result. Equation (10) and symmetry imply

$$\mathrm{E}[\widetilde{D}_{\mathrm{J},r}(n)] \equiv \beta_1\mathrm{E}[D(n)] + \beta_2\mathrm{E}[D(rn)] + \beta_3\mathrm{E}[D((1-r)n)]. \tag{17}$$

Equation (12) follows by plugging (7) and (11) into (17).

Now we prove the variance result. From [11], we know that as $n \to \infty$,

$$\begin{aligned}
\mathrm{Cov}(A(f_0;n), A(f_0;rn)) &\to 2r^3\sigma^4, & (18)\\
\mathrm{Cov}(A(f_0;n), C(g_0;rn)) &\to 6r^3\sigma^4/5, & (19)\\
\mathrm{Cov}(C(g_0;n), A(f_0;rn)) &\to 6r^2\sigma^4/5, & (20)\\
\mathrm{Cov}(C(g_0;n), C(g_0;rn)) &\to 4r^2\sigma^4/5. & (21)
\end{aligned}$$

From the definition in (6) and Equations (18)–(21), we find that

$$\begin{aligned}
\mathrm{Cov}(D(n), D(rn)) &= \mathrm{Cov}(2C(g_0;n) - A(f_0;n), 2C(g_0;rn) - A(f_0;rn)) & (22)\\
&\to 2r^2(2-r)\sigma^4/5 \quad \text{as } n \to \infty. & (23)
\end{aligned}$$

By symmetry,

$$\mathrm{Cov}\big(D(n), \widetilde{D}((1-r)n)\big) \to 2(1-r)^2(1+r)\sigma^4/5. \tag{24}$$

Since $D(rn)$ and $\widetilde{D}((1-r)n)$ consist of disjoint batches of observations, and the limiting Brownian motion from Assumption FCLT has independent increments, we have, as $n \to \infty$,

$$\mathrm{Cov}\big(D(rn), \widetilde{D}((1-r)n)\big) \to 0, \tag{25}$$

a fact that also follows by the arguments in [4]. Further, from (7) and symmetry,

$$\mathrm{Var}(D(n)), \quad \mathrm{Var}(D(rn)), \quad \text{and} \quad \mathrm{Var}(\widetilde{D}((1-r)n)) \quad \text{all} \to 2\sigma^4/5 \quad \text{as } n \to \infty. \tag{26}$$

Finally, we put everything together. From (10) and (11), we have

$$\begin{aligned}\mathrm{Var}(\widetilde{D}_{\mathrm{J},r}(n)) &= \beta_1^2\mathrm{Var}(D(n)) + \beta_2^2\mathrm{Var}(D(rn)) + \beta_3^2\mathrm{Var}(\widetilde{D}((1-r)n)) \\ &\quad +2\beta_1\beta_2\mathrm{Cov}(D(n), D(rn)) + 2\beta_1\beta_3\mathrm{Cov}(D(n), \widetilde{D}((1-r)n)) \\ &\quad +2\beta_2\beta_3\mathrm{Cov}(D(rn), \widetilde{D}((1-r)n)), \end{aligned} \tag{27}$$

and the result (13) follows by Equations (23)–(26). ◇

7. Exact Example: Moving Average Process

A first-order moving average [MA(1)] process is defined as $Y_i = \theta\epsilon_{i-1} + \epsilon_i$, $i \geq 1$, where the ϵ_i's are i.i.d. Nor(0, 1) random variables. The MA(1) has covariance function $R_0 = 1 + \theta^2$, $R_1 = \theta$, and $R_k = 0$ elsewhere; variance parameter $\sigma^2 = (1+\theta)^2$; and constants $\gamma_j = 2\theta$, $j = 1, 2, \ldots$. We next calculate the expected value and variance of the MJDW estimator for the special case of $b = 1$ batch.

We know from [11] that for the MA(1), the following expectation holds exactly:

$$\mathrm{E}[D(n)] = \sigma^2 - \frac{7\gamma_1}{n} + \frac{12\gamma_1 - \sigma^2}{n^2} - \frac{5\gamma_1}{n^3}, \tag{28}$$

which matches with (7) after noting that all of the γ_j's are equal. Similar to the first part of Theorem 1's proof, Equation (28) immediately leads us to the first-order unbiased expression for $\mathrm{E}[\widetilde{D}_{\mathrm{J},r}(n)]$ given by (12).

Now we will derive $\mathrm{Var}(\widetilde{\mathcal{D}}_{\mathrm{J},r}(n))$ for the MA(1) by calculating the components of Equation (27). First, we know from [11] that for the MA(1),

$$\mathrm{Var}(D(n)) = \frac{2\sigma^4}{5} + O(1/n). \tag{29}$$

In addition, the arguments in [11] show that, for the MA(1),

$$\mathrm{Cov}(A(f_0; n), A(f_0; rn)) = 2r^3\sigma^4 + O(1/n), \tag{30}$$

$$\mathrm{Cov}(A(f_0; n), C(g_0; rn)) = \frac{6r^3\sigma^4}{5} + O(1/n), \tag{31}$$

$$\mathrm{Cov}(C(g_0; n), A(f_0; rn)) = \frac{6r^2\sigma^4}{5} + O(1/n), \tag{32}$$

$$\mathrm{Cov}(C(g_0; n), C(g_0; rn)) = \frac{4r^2\sigma^4}{5} + O(1/n). \tag{33}$$

Plugging (30)–(33) into (22) yields

$$\mathrm{Cov}(D(n), D(rn)) = \frac{2r^2(2-r)\sigma^4}{5} + O(1/n); \tag{34}$$

and then symmetry gives

$$\mathrm{Cov}(D(n), \widetilde{D}((1-r)n)) = \frac{2(1-r)^2(1+r)\sigma^4}{5} + O(1/n). \tag{35}$$

In order to put everything together, we have one last major task — we need to show that

$$\mathrm{Cov}(D(rn), \widetilde{D}((1-r)n)) = O(1/n^2). \tag{36}$$

Before deriving Equation (36), let $\widetilde{Y}_s \equiv \sum_{i=n-s+1}^{n} Y_i/s$, $s < n$, denote the average of the last s observations out of $Y_1, Y_2, \ldots, Y_n$. For the MA(1) process under study,

$$\mathrm{Cov}(\bar{Y}_m - \bar{Y}_j, \widetilde{Y}_{n-m} - \widetilde{Y}_k) \;=\; \begin{cases} \gamma_1/[2m(n-m)], & j < m \text{ and } k < n-m \\ 0, & \text{otherwise.} \end{cases} \tag{37}$$

Now define $\widetilde{T}_s$, $s < n$, as the STS formed from the last s observations, and $\widetilde{A}(f;s)$ and $\widetilde{C}(g;s)$ as the area and CvM estimators, respectively, calculated from the last s observations. Thus,

$$\begin{aligned}
&\mathrm{Cov}(C(g_0;m), \widetilde{C}(g_0, n-m)) \\
&= \mathrm{Cov}\Big(\frac{1}{m}\sum_{j=1}^{m} g_0(\tfrac{j}{m})\sigma^2 T_m^2(\tfrac{j}{m}), \frac{1}{n-m}\sum_{k=1}^{n-m} g_0(\tfrac{k}{n-m})\sigma^2 \widetilde{T}_{n-m}^2(\tfrac{k}{n-m})\Big) \\
&= \frac{36}{m(n-m)} \sum_{j=1}^{m}\sum_{k=1}^{n-m} \mathrm{Cov}\big(\sigma^2 T_m^2(\tfrac{j}{m}), \sigma^2 \widetilde{T}_{n-m}^2(\tfrac{k}{n-m})\big) \\
&= \frac{72}{m(n-m)} \sum_{j=1}^{m}\sum_{k=1}^{n-m} \mathrm{Cov}^2\big(\sigma T_m(\tfrac{j}{m}), \sigma \widetilde{T}_{n-m}(\tfrac{k}{n-m})\big) \\
&\qquad \text{(by [A-3] since } (T_m(\cdot), \widetilde{T}_{n-m}(\cdot)) \text{ is bivariate normal with mean (0,0))} \\
&= \frac{72}{m(n-m)} \sum_{j=1}^{m}\sum_{k=1}^{n-m} \mathrm{Cov}^2\Big(\frac{j}{\sqrt{m}}(\bar{Y}_m - \bar{Y}_j), \frac{k}{\sqrt{n-m}}(\widetilde{Y}_{n-m} - \widetilde{Y}_k)\Big) \\
&= \frac{\gamma_1^2 (m-1)(2m-1)(n-m-1)(2n-2m-1)}{2m^3(n-m)^3} \quad \text{(by (37) and algebra).}
\end{aligned}$$

Substitution of $m = rn$ yields

$$\mathrm{Cov}\big(C(g_0; rn), \widetilde{C}(g_0, (1-r)n)\big) \;=\; O(n^{-2}).$$

Similarly, we find that $\mathrm{Cov}\big(C(g_0;rn), \widetilde{A}(f_0;(1-r)n)\big)$, $\mathrm{Cov}\big(A(f_0;rn), \widetilde{C}(g_0;(1-r)n)\big)$, and $\mathrm{Cov}\big(A(f_0;rn), \widetilde{A}(f_0,(1-r)n)\big)$ are all $O(n^{-2})$. These remarks finally imply

$$\begin{aligned}
&\mathrm{Cov}\big(D(rn), \widetilde{D}((1-r)n)\big) \\
&= \mathrm{Cov}\big(2C(g_0;rn) - A(f_0;rn), 2\widetilde{C}(g_0;(1-r)n) - \widetilde{A}(f_0;(1-r)n)\big) \;=\; O(n^{-2}), \quad \diamondsuit
\end{aligned}$$

which is what we wanted to show for Equation (36).

Then from Equations (27), (29), (34)–(36), and symmetry, we obtain

$$\mathrm{Var}(\widetilde{D}_{\mathrm{J},r}(n)) \;=\; \frac{2\sigma^4}{5}\Big(\beta_1^2 + \beta_2^2 + \beta_3^2 + 2\beta_1\beta_2 r^2(2-r) + 2\beta_1\beta_3(1-r)^2(1+r)\Big) + O(1/n).$$

Substituting the β_j's from (11), we get Theorem 1's variance result.

8. Monte Carlo Results for an AR(1) Process

As a supplement to the $M/M/1$ example in §5, we also consider a first-order autoregressive [AR(1)] process. An AR(1) process is defined as $Y_i = \phi Y_{i-1} + \epsilon_i$, $i \geq 1$, where the ϵ_i's are i.i.d.

$\mathrm{Nor}(0, 1 - \phi^2)$, $Y_0 \sim \mathrm{Nor}(0, 1)$ is generated independently of the others, and $-1 < \phi < 1$. The AR(1) has $R_k = \phi^k$, $k = 0, 1, 2, \ldots$, $\sigma^2 = (1+\phi)/(1-\phi)$, and $\gamma_1 = 2\phi/(1-\phi)^2$. In this example, we set $\phi = 0.9$, which gives a positively autocorrelated process with $\sigma^2 = 19$.

We again conducted 100,000 independent replications of $\boldsymbol{Y}$, each of which ran for a sample of size $n = 4096$. For each replication, we computed the various estimators after having divided the run using $b = n/m = 1$, 2, 4, 8, and 16 batches (with respective batch sizes $m = 4096$, 2048, 1024, 512, and 256). Based on the results in Table 3, we can make the same general conclusions as those based upon the $M/M/1$ Monte Carlo runs.

Table 3
Estimated mean and variance of various batched estimators for the variance parameter of an AR(1) process with $\phi = 0.9$ (so that $\sigma^2 = 19$) and $n = 4096$.

Estimator	$b=1$ $m=4096$		$b=2$ $m=2048$		$b=4$ $m=1024$		$b=8$ $m=512$		$b=16$ $m=256$	
	$\widehat{\mathrm{E}}$	$\widehat{\mathrm{Var}}$	$\widehat{\mathrm{E}}$	$\widehat{\mathrm{Var}}$	$\widehat{\mathrm{E}}$	$\widehat{\mathrm{Var}}$	$\widehat{\mathrm{E}}$	$\widehat{\mathrm{Var}}$	$\widehat{\mathrm{E}}$	$\widehat{\mathrm{Var}}$
$A(f_0; b, m)$	18.85	720	18.75	349	18.40	167	17.94	81	16.91	36
$A(f_2; b, m)$	19.00	728	19.00	360	18.84	177	18.76	89	18.13	41
$C(g_0; b, m)$	18.78	289	18.56	141	18.09	68	17.32	34	15.79	15
$C(g_2^\star; b, m)$	19.01	626	18.97	308	18.86	150	18.73	74	18.06	34
$C(g_4^\star; b, m)$	18.98	375	18.94	182	18.77	90	18.39	43	17.13	19
$D(b, m)$	18.71	145	18.37	71	17.79	35	16.69	17	14.68	8
$D_{\mathrm{J},0.5}(b, m)$	19.03	505	19.00	248	18.88	122	18.71	58	17.96	25
$\widetilde{D}_{\mathrm{J},0.5}(b, m)$	19.05	434	18.96	213	18.88	104	18.70	51	17.96	23
$\mathcal{N}(b, m)$			18.92	717	18.80	236	18.58	98	18.24	45
$\mathcal{O}(b, m)$			18.82	241	18.76	153	18.57	69	18.23	32

9. Probability of Negative Realizations

Since it is possible for the β coefficients in (10) and (11) to be negative, it is possible to obtain undesirable negative realizations of $\widetilde{\mathcal{D}}_{\mathrm{J},r}(b, m)$. As in [A-1], this problem disappears when we apply a minimum amount of batching. For example, for the AR(1) example from §8, we obtained the following empirical probabilities of negative realizations based on 100,000 runs.

(b, m)	$(1, 4096)$	$(2, 2048)$	$(4, 1024)$	$(8, 512)$	$(16, 256)$
$\widehat{\Pr}\big(\widetilde{\mathcal{D}}_{\mathrm{J},0.5}(b, m) < 0\big)$	0.0679	0.0156	0.0010	0.0000	0.0000

Additional References for Online Companion

[A-1] Alexopoulos, C., B.-Y. Chang, D. Goldsman, S. Lee, and W. S. Marshall. 2008. Overcoming Negativity Problems for Cramér–von Mises Variance Estimators. *Int. J. of Simulation and Process Modelling*, 4:1–6.

[A-2] Batur, D., D. Goldsman and S.-H. Kim. 2008. An Improved Standardized Time Series Durbin–Watson Variance Estimator for Steady-State Simulation. Technical report, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA.

[A-3] Patel, J. K. and C. B. Read. 1996. *Handbook of the Normal Distribution*, 2nd Ed. Marcel Dekker, New York.