MATHEMATICS OF OPERATIONS RESEARCH
Vol. 31, No. 2, May 2006, pp. 351–368
ISSN 0364-765X | EISSN 1526-5471 | 06 | 3102 | 0351
DOI 10.1287/moor.1050.0183

Permuted Standardized Time Series for Steady-State Simulations

James M. Calvin, Marvin K. Nakayama
Department of Computer Science, New Jersey Institute of Technology, Newark, New Jersey 07102-1982
{calvin@njit.edu, marvin@njit.edu}

We describe an extension procedure for constructing new standardized time series procedures from existing ones. The approach is based on averaging over sample paths obtained by permuting path segments. Analytical and empirical results indicate that permuting improves standardized time series methods. We compare permuting to an alternative extension procedure known as batching. We demonstrate the permuting method by applying it to estimators based on the maximum and the area of a normalized path.

Key words: steady-state simulation; confidence intervals; functional central limit theorem
MSC2000 subject classification: Primary: 65C05; secondary: 62M10
OR/MS subject classification: Primary: Simulation/statistical analysis; secondary: statistics/time series
History: Received July 25, 2002; revised September 10, 2004, and July 5, 2005.

1. Introduction. The goal of many steady-state simulations is to estimate and construct confidence intervals for the steady-state mean μ of a stochastic process. Under mild conditions, the time average of a process converges to μ with probability one, so constructing a point estimator for μ is straightforward: run a long simulation of the process, and use the time average of the process over the simulation as a point estimate for μ. Time averages often satisfy a central limit theorem, but the autocorrelations typically present in simulation output make constructing confidence intervals for μ based on a central limit theorem a delicate undertaking. A number of techniques have been developed to do this under various assumptions; e.g., see Law and Kelton [10, §9.5].

One such approach is to apply a standardized time series (STS) technique (Schruben [11]). The validity of STS methods depends on the simulated process satisfying a functional central limit theorem (e.g., Billingsley [1]); that is, a suitably centered and scaled version of the process converges to a Brownian motion as the run length grows large. One implements an STS method by applying a "scaling function" b to a standardized version of the original process. Glynn and Iglehart [7] generalize the class of STS methods, study some of their theoretical asymptotic properties, and provide references that establish sufficient conditions under which a functional central limit theorem holds.

If we break the original path into a fixed number m of (nonoverlapping) batches, then each of the batches also satisfies a functional central limit theorem if the original process does. Hence, we can apply an STS method to each of the batches independently, and then combine the results using a p-norm, $p \geq 1$. Proposed by Schruben [11], this approach, which we call *batching*, is thus a way to take a given STS scaling function and derive new ones from it. Although batching is asymptotically valid (as run lengths grow to infinity), in practice, finite sample sizes can significantly degrade the performance (e.g., lower coverage) compared to no batching for the same basic STS scaling function. Because each batch is much shorter than the entire run length, the Brownian approximation will typically be worse for each batch than it is for the entire sample path.

In this paper, we introduce an approach for developing new STS methods from existing ones. The idea is based on dividing a sample path into segments and averaging an estimator over all paths obtained by permuting the path segments. (This is similar to an approach we developed for regenerative processes in Calvin and Nakayama [3, 4].) Specifically, suppose that the sample path is split into m equal-length segments, $m \geq 2$. Then, permuting the segments and piecing them together yields another sample path, which satisfies a functional central limit theorem if the original process does. (Note that we differentiate between the terms *segments* and *batches*.) We apply a scaling function b to each permuted sample path, and we obtain the permuted scaling function $\tilde{b}_{m,p}$ by combining the values of b over all $m!$ permutations using a p-norm, $p \geq 1$.

When permutations are combined with the 1-norm (i.e., $p = 1$), the expected value (respectively, variance) of $\tilde{b}_{m,p}$ applied to the Brownian limit is the same as (respectively, no greater than) that for the nonpermuted scaling function b (see Theorem 5.1). Theorem 5.1 also shows that when $p = 2$, the mean (respectively, variance) of $\tilde{b}_{m,p}$ applied to the Brownian limit is no less than (respectively, no greater than) that for the nonpermuted scaling function b. However, because the expected width of the limiting confidence interval also depends on an appropriate quantile point, as well as the expected value of the scaling function applied to the Brownian limit, the results in Theorem 5.1 do not tell the entire story about expected asymptotic confidence-interval widths. But we present theoretical and empirical results showing that permuting typically leads to smaller and less variable confidence intervals compared to the *regular* (i.e., nonpermuted and nonbatched) STS estimator. While empirical evidence shows that there is some degradation in coverage of confidence intervals based on permuted STS estimators (compared to the regular STS confidence intervals) when sample sizes are small, the effect is less pronounced than for batching for each value of m.

We demonstrate our approach by applying it to two examples of STS methods: the maximum estimator and the area estimator (Schruben [11]). We derive a simple representation of the permuted maximum estimator with $m = 2$ segments and $p = 1$. We also consider the case of the permuted maximum estimator for $m > 2$ and $p = 1$, but the resulting estimator requires a summation over $m!$ terms, thus limiting the size of m that is feasible in practice. For the permuted area estimator with $m \geq 2$ and $p = 2$, we derive a representation that does not require explicitly summing over $m!$ terms, and Calvin and Nakayama [6] present the same for the permuted version of Goldsman and Schruben's [9] weighted area estimator.

For the maximum estimator, analytical calculations show that permuting with $m = 2$ segments and $p = 1$ results in roughly a 25% reduction in the mean asymptotic half width of 90% confidence intervals, relative to that of the regular maximum estimator. We also compare analytically the expected asymptotic half widths when permuting $m = 2$ segments and batching with $m = 2$ batches, each with $p = 1$, and show that permuting leads to a slight reduction in expected half width. The asymptotic variability of the half widths is also reduced.

In addition, we show that asymptotically, the permuted area estimator ($m \geq 2$, $p = 2$) is equal in distribution to a function of two independent chi-square random variables. We use simulation to estimate the quantile points needed for constructing confidence intervals.

We also ran experiments with our estimators. For the maximum estimator, empirical results for up to $m = 5$ batches or segments indicate that permuting gives rise to shorter and less variable confidence intervals, compared to the regular maximum, but with virtually no loss in coverage. In contrast, batching also leads to shorter and less variable intervals than the regular maximum, but there is significant degradation in coverage. For the area and weighted area estimators, permuting and batching also improve on the regular estimator, but now with some loss in coverage for both. However, for each value of $m \geq 2$, batching consistently suffers worse degradation in coverage than permuting (with about the same average and variance of half width), so it appears that permuting outperforms batching in the small-sample context.

The maximum estimator we study has apparently not been considered before in the literature. Schruben [11] developed a similar approach, the standardized maximum, which is based on the maximum of a function divided by a function of the location at which the maximum occurs. In contrast, the maximum estimator we study is based only on the value of the maximum. Tokol et al. [13] propose an STS method based on the maximum of the absolute value.

Our permuting approach can also be used with other STS methods not considered here, but the main purpose of this paper is to introduce and illustrate the use of permuting and to compare properties of the resulting permuted scaling function to the original and batched scaling functions.

The rest of this paper is organized as follows: In §2, we review the general method of standardized time series. We introduce the maximum estimator in §3, and discuss in §4 the general approach of batching. In §5, we present the permuted STS approach and apply it to the maximum estimator ($m \geq 2$, $p = 1$). We also study some asymptotic properties of the permuted maximum for the special case when $m = 2$, and compare it to the batched maximum for $m = 2$. We derive the permuted area estimator ($m \geq 2$, $p = 2$) in §6. Empirical results are discussed in §7, and we give concluding comments in §8. An appendix contains some of the longer proofs. Calvin and Nakayama [5] presents some of the results from this paper without proofs.

2. Background on standardized time series. In this section, we summarize some facts about standardized time series; see Schruben [11] or Glynn and Iglehart [7] for additional details. Let $Y = [Y(t): t \geq 0]$ be a real-valued right-continuous stochastic process with left limits representing the output of a simulation experiment. For each $n \geq 1$, define the scaled process Y_n, $n \geq 1$, by

$$Y_n(t) = \frac{1}{n} \int_{s=0}^{nt} Y(s)\, ds, \quad 0 \leq t \leq 1.$$

Note that $Y_n(1)$ is the sample mean of the process Y up to time n.

Let Ω_0 be the space of continuous real-valued functions x on the unit interval that vanish at 0, and let Ω_{00} denote $\{x \in \Omega_0: x(1) = 0\}$.

Suppose that there is a real number μ and a positive number σ such that if we define a sequence of processes X_n, $n \geq 1$, by

$$X_n(t) = \sqrt{n}(Y_n(t) - \mu t), \quad 0 \leq t \leq 1,$$

then

$$X_n \xrightarrow{\mathcal{D}} \sigma W \tag{1}$$

in Ω_0 endowed with the uniform metric as $n \to \infty$, where W is a standard Brownian motion, and $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Note that Y_n, X_n, and W are elements of Ω_0, although Y need not be. The functional central limit theorem (FCLT) (1) has been shown to hold under a variety of assumptions. Often the conditions require a type of asymptotic independence in the form of mixing conditions, which assert that two events far apart in time are almost independent (Billingsley [1]).

We are interested in constructing confidence intervals for the unknown parameter μ, the steady-state mean of Y. One way of accomplishing this is to apply a technique from the class of standardized time series methodologies, which we now describe.

Let Γ be the projection from Ω_0 onto Ω_{00} given by

$$\Gamma(x)(t) = x(t) - tx(1), \quad 0 \le t \le 1,$$

for $x \in \Omega_0$. For $\alpha \in \mathbb{R}$ and $y \in \Omega_0$ defined by $y(t) = \alpha t$, $0 \le t \le 1$, $\Gamma(y) = 0$. Let B denote a standard Brownian bridge process on [0, 1], and note that $\Gamma(W) \stackrel{\mathcal{D}}{=} B$, where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution. The limit theorem (1) implies that

$$\big(\Gamma(X_n), X_n(1)\big) \xrightarrow{\mathcal{D}} \big(\sigma B, \sigma W(1)\big) \tag{2}$$

as $n \to \infty$.

Let $\mathbb{R}_+$ be the nonnegative real numbers. An STS method is based on a continuous function b: $\Omega_{00} \to \mathbb{R}_+$ satisfying the following conditions:

(i) $b(\alpha x) = \alpha b(x)$ for $\alpha \in \mathbb{R}$ with $\alpha > 0$ and $x \in \Omega_{00}$,

(ii) $P\{b(B) > 0\} = 1$.

We call any continuous function b satisfying the conditions above a *scaling function*. Glynn and Iglehart [7] present several examples of scaling functions b, including ones for the area estimator, the standardized maximum, and the method of batch means.

When (1) holds and b is a scaling function,

$$\frac{Y_n(1) - \mu}{b \circ \Gamma(Y_n)} = \frac{X_n(1)}{\sqrt{n}b \circ \Gamma(Y_n)} = \frac{X_n(1)}{b \circ \Gamma(X_n)} \xrightarrow{\mathcal{D}} \frac{W(1)}{b(B)} \tag{3}$$

as $n \to \infty$, where $b \circ \Gamma(x) = b(\Gamma(x))$. In the second equality we used the scaling property (i) of b and the linearity of Γ, and the convergence follows from the continuity of b and (2). Let $\Phi(x) = P(W(1) \le x)$ (i.e., Φ is the distribution function of a standard normal), $\phi(x) = d\Phi(x)/dx$, $G(x) = P(b(B) \le x)$, and $H(x) = P(W(1)/b(B) \le x)$. Then,

$$H(x) = \int_0^\infty \Phi(xy)\, G(dy)$$

because $W(1)$ and $b(B)$ are independent (e.g., see Glynn and Iglehart [7, Proposition 2.8]). By (3),

$$P\left(\frac{Y_n(1) - \mu}{b \circ \Gamma(Y_n)} \le x\right) \to H(x)$$

as $n \to \infty$.

To construct $100(1-\delta)\%$ confidence intervals, select $\gamma_{b,\,1-\delta/2} \in \mathbb{R}$ such that $H(\gamma_{b,\,1-\delta/2}) = 1 - \delta/2$, and note that H is a symmetric distribution. Then, as the run length $n \to \infty$, the interval

$$\big[Y_n(1) - \gamma_{b,\,1-\delta/2} b \circ \Gamma(Y_n),\, Y_n(1) + \gamma_{b,\,1-\delta/2} b \circ \Gamma(Y_n)\big]$$

is an asymptotic $100(1-\delta)\%$ confidence interval for μ. The half width of the confidence interval is $L_n \equiv \gamma_{b,\,1-\delta/2} b \circ \Gamma(Y_n)$, and the scaling property of b and linearity of Γ imply that $L_n = \gamma_{b,\,1-\delta/2} b \circ \Gamma(X_n)/\sqrt{n}$. Thus, when (1) holds, $\sqrt{n}L_n \xrightarrow{\mathcal{D}} \gamma_{b,\,1-\delta/2} b(\sigma B) = \gamma_{b,\,1-\delta/2} \sigma b(B)$ as $n \to \infty$ by the continuous mapping theorem (e.g., see Billinglsey [1]). Moreover, if $\{b \circ \Gamma(X_n)\colon n \ge 1\}$ is uniformly integrable (e.g., see Billingsley [1]), then $\sqrt{n}E[L_n] \to \gamma_{b,\,1-\delta/2} \sigma E[b(B)]$ as $n \to \infty$. Note that σ depends only on the original process Y and not on the choice of the function b, so we will use

$$\psi_1(b, 1-\delta) \equiv \gamma_{b,\,1-\delta/2} E[b(B)] \tag{4}$$

as a measure of the limiting expected half width of a $100(1-\delta)\%$ confidence interval obtained using the STS method based on scaling function b. We also consider

$$\psi_2(b, 1-\delta) \equiv \gamma^2_{b,1-\delta/2} \operatorname{Var}[b(B)] \tag{5}$$

as a measure of the asymptotic variability of the confidence interval half width.

3. Standardized time series based on maximum. Define $b_{\max}: \Omega_{00} \to \mathbb{R}_+$ by

$$b_{\max}(x) = \max_{0\le t\le 1} x(t). \tag{6}$$

Because the maximum of a Brownian bridge is almost surely positive, $b_{\max}$ defines an STS scaling function. The distribution function H of $W(1)/b_{\max}(B)$ is given in the following result.

Proposition 3.1. *For the maximum scaling function $b_{\max}$ defined in (6),*

$$H(x) = \frac{1}{2}\left(1 + \frac{x}{\sqrt{4+x^2}}\right),$$

and

$$E[b_{\max}(B)] = \frac{\sqrt{2\pi}}{4}, \qquad \mathrm{Var}[b_{\max}(B)] = \frac{4-\pi}{8}.$$

Proof. The distribution of the maximum of the Brownian bridge is given by

$$P(b_{\max}(B) > y) = \exp\{-2y^2\}, \quad y \ge 0; \tag{7}$$

e.g., see Breiman [2, p. 290]. Thus, for $x > 0$, we have

$$\begin{aligned} H(x) &= \frac{1}{2} + \int_{y=0}^{\infty}\int_{w=0}^{xy} \phi(w)\,dw\,G(dy) \\ &= \frac{1}{2} + \int_{w=0}^{\infty} \phi(w) \int_{y=w/x}^{\infty} G(dy)\,dw \\ &= \frac{1}{2} + \int_{w=0}^{\infty} \phi(w)(1 - G(w/x))\,dw \\ &= \frac{1}{2} + \int_{w=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} e^{-2(w/x)^2}\,dw \\ &= \frac{1}{2} + \int_{w=0}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-w^2}{2/(1+4/x^2)}\right\} dw \\ &= \frac{1}{2}\left(1 + \frac{x}{\sqrt{4+x^2}}\right). \end{aligned}$$

Also, (7) implies

$$E[b_{\max}(B)] = \int_{y=0}^{\infty} P(b_{\max}(B) > y)\,dy = \int_{y=0}^{\infty} e^{-2y^2}\,dy = \frac{\sqrt{2\pi}}{4}.$$

Similarly,

$$E[b_{\max}(B)^2] = 2\int_{y=0}^{\infty} yP(b_{\max}(B) > y)\,dy = 2\int_{y=0}^{\infty} ye^{-2y^2}\,dy = \frac{1}{2}. \quad \square$$

Setting $H(\gamma_{b_{\max},1-\delta/2}) = 1 - \delta/2$ and solving for $\gamma_{b_{\max},1-\delta/2}$, we obtain the critical points of the distribution H for $b_{\max}$ as

$$\gamma_{b_{\max},1-\delta/2} = \frac{2(1-\delta)}{\sqrt{1-(1-\delta)^2}}.$$

The values of the critical points for $1-\delta = 0.90$, 0.95, and 0.99 are given in the first row of Table 1. (The parameters m and p, which are used in batching and permuting, will be defined later.) Table 2 gives the limiting expected half-width measure ψ_1 and the limiting variance measure ψ_2 for various confidence levels.

Table 1. Critical points for the maximum estimator and the batched and permuted maximum estimators with $m = 2$ and $p = 1$.

Method	$E[b(B)]$	$\gamma_{b,0.95}$	$\gamma_{b,0.975}$	$\gamma_{b,0.995}$
Maximum	$\sqrt{2\pi}/4$	4.129	6.085	14.036
Batched maximum	$\sqrt{2\pi}/4$	3.218	4.198	6.979
Permuted maximum	$\sqrt{2\pi}/4$	3.096	3.961	6.244

Table 2. Confidence interval properties of the maximum estimator and the batched and permuted maximum estimators with $m=2$ and $p=1$.

Method	$\psi_1(b, 0.90)$	$\psi_1(b, 0.95)$	$\psi_1(b, 0.99)$	$\psi_2(b, 0.90)$	$\psi_2(b, 0.95)$	$\psi_2(b, 0.99)$
Maximum	2.588	3.813	8.796	1.829	4.969	21.139
Batched maximum	2.016	2.631	4.374	0.556	0.945	2.613
Permuted maximum	1.940	2.482	3.913	0.429	0.703	1.747

4. Batching. We now describe a standard way to extend standardized time series methods to multiple "batches." The idea is rather than approximate the entire sample path by a single Brownian bridge, we break up the path into m (nonoverlapping) batches and approximate each batch by a Brownian bridge. The scaling function b is then applied to each centered and scaled batch, and we combine the m b-function values to come up with the overall estimator. We now discuss in detail this approach, which is discussed in Schruben [11] and Glynn and Iglehart [7].

We describe first how to break up a sample path into m batches, and then discuss combining the scaling-function values of the m batches. Let $L_m(x)\Lambda_m(x)$ denote the linear interpolation of the function x based on the function values at the points $0/m, 1/m, 2/m, \ldots, m/m$:

$$\Lambda_m(x)(s) = (j+1-sm)x\left(\frac{j}{m}\right) + (sm-j)x\left(\frac{j+1}{m}\right), \quad \frac{j}{m} \le s \le \frac{j+1}{m},$$

for $j = 0, 1, \ldots, m-1$. For $i = 1, 2, \ldots, m$, define $\Gamma_{i,m}\colon \Omega_0 \to \Omega_{00}$ by

$$\begin{aligned}\Gamma_{i,m}(x)(t) &= \sqrt{m}\left[x\left(\frac{i-1+t}{m}\right) - \Lambda_m(x)\left(\frac{i-1+t}{m}\right)\right]\\ &= \sqrt{m}\left[x\left(\frac{i-1+t}{m}\right) - (1-t)x\left(\frac{i-1}{m}\right) - tx\left(\frac{i}{m}\right)\right]\end{aligned}$$

for $0 \le t \le 1$. Note that $\Gamma_{i,m}$ takes the ith batch and stretches it to the unit interval with $\Gamma_{i,m}(x)$ vanishing at the endpoints. Moreover, it can be shown (see Glynn and Iglehart [7, pp. 6–7]) that $B_i \equiv \Gamma_{i,m}(B) = \Gamma_{i,m}(W)$, $i = 1, 2, \ldots, m$, are independent standard Brownian bridges over the unit interval. Hence, if b is an STS scaling function, then $P\{b \circ \Gamma_{i,m}(B) > 0\} = P\{b(B_i) > 0\} = 1$. This fact, together with the linearity of $\Gamma_{i,m}$, imply that $b \circ \Gamma_{i,m}$ is an STS scaling function whenever b is. Then, we combine the m functions $b \circ \Gamma_{i,m}$ to get the batched scaling function; i.e., for $x \in \Omega_{00}$, set

$$b^*_{m,p}(x) = \left(\frac{1}{m}\sum_{i=1}^{m}\left[b \circ \Gamma_{i,m}(x)\right]^p\right)^{1/p}, \tag{8}$$

which is the *batched scaling function* for m batches using the p-norm constructed from scaling function b. In this paper, we consider the case of $p = 1, 2$, but other values of p, or more generally, any norm, can be used.

The following result compares some of the moments of batched and unbatched scaling functions applied to the Brownian limit.

Proposition 4.1. *Suppose that b is a scaling function. Then, for any number of batches $m \ge 1$,*
(i) $E[b^*_{m,p}(B)^p] = E[b(B)^p]$ *when* $p \ge 1$,
(ii) $\mathrm{Var}[b^*_{m,p}(B)^p] = \mathrm{Var}[b(B)^p]/m$ *when* $p \ge 1$,
(iii) $E[b^*_{m,p}(B)] \ge E[b(B)]$ *when* $p > 1$,
(iv) $\mathrm{Var}[b^*_{m,p}(B)] \le \mathrm{Var}[b(B)]$ *when* $p = 2$.

Proof. Parts (i) and (ii) follow from (8) and the fact that $\Gamma_{i,m}(B)$, $i = 1, \ldots, m$, are independent and identically distributed (i.i.d.) standard Brownian bridges. Part (iii) holds by Hölder's inequality, and using (i) and (iii) with $p = 2$ yields (iv). □

5. Permuted standardized time series. We now present a new method for developing STS methods from existing ones. The basic idea entails dividing the sample path into m nonoverlapping equal-length segments. Permute the m segments to generate another sample path, and apply a scaling function b to the entire permuted path. Averaging over all $m!$ permutations results in the permuted estimator.

More precisely, fix $m \geq 2$ and let S_m denote the set of permutations of $\{1, 2, \ldots, m\}$. For permutation $\pi = (\pi(1), \pi(2), \ldots, \pi(m)) \in S_m$, define $T_\pi\colon \Omega_0 \to \Omega_0$ so that $T_\pi(x)$ is the function obtained by gluing together the m segments of x according to the permutation π. Specifically, for $1 \leq i \leq m$, define $\Delta_i\colon \Omega_0 \to \Omega_0$ by

$$\Delta_i(x)(s) = x\left(\frac{i-1}{m} + s\right) - x\left(\frac{i-1}{m}\right), \quad 0 \leq s \leq \frac{1}{m},$$

which is the process over the ith batch, shifted to start at the origin, and set

$$T_\pi(x)(s) = \sum_{j=1}^{i-1} \Delta_{\pi(j)}(x)\left(\frac{1}{m}\right) + \Delta_{\pi(i)}(x)\left(s - \frac{i-1}{m}\right), \quad \frac{i-1}{m} \leq s < \frac{i}{m},$$

for $1 \leq i \leq m$. For each $\pi \in S_m$, $T_\pi(B) \stackrel{\mathcal{D}}{=} B$, so $b \circ T_\pi$ is an STS scaling function if b is. As for the batching scheme, all $m!$ scaling function values so obtained can be combined by applying a norm. Thus, we define

$$\tilde{b}_{m,p}(x) = \left(\frac{1}{m!} \sum_{\pi \in S_m} \big(b \circ T_\pi(x)\big)^p\right)^{1/p}, \tag{9}$$

which is the *permuted scaling function* with m segments using scaling function b and the p-norm.

A slightly different way of looking at the permuted scaling function is as follows. Let $x \in \Omega_{00}$. For a fixed number of segments m, let $\widetilde{\Omega}_{x,m} = \{T_\pi(x)\colon \pi \in S_m\}$ be the set of $m!$ permuted paths that can be constructed from x. Let $\widetilde{P}(\cdot \mid x)$ denote the uniform probability distribution on $\widetilde{\Omega}_{x,m}$; i.e., $\widetilde{P}(\cdot \mid x)$ assigns probability $1/m!$ to each $T_\pi(x) \in \widetilde{\Omega}_{x,m}$. Let $\widetilde{E}(\cdot \mid x)$ and $\widetilde{\mathrm{Var}}(\cdot \mid x)$ denote the corresponding expectation and variance, respectively, and let $\widetilde{X}$ be a random draw from $\widetilde{\Omega}_{x,m}$ under distribution $\widetilde{P}(\cdot \mid x)$. Then, we can write the permuted scaling function $\tilde{b}_{m,p}$ applied to path x as

$$\tilde{b}_{m,p}(x) = \big(\widetilde{E}[b(\widetilde{X})^p \mid x]\big)^{1/p}.$$

Now replace x with a Brownian bridge B, so $\widetilde{\Omega}_{B,m} = \{T_\pi(B)\colon \pi \in S_m\}$ is the set of permuted paths of B for m segments. Let $\widetilde{B}$ be a random sample path in Ω_{00} obtained by first sampling a sample path of B and then uniformly randomly selecting a path from $\Omega_{B,m}$. Because $T_\pi(B) \stackrel{\mathcal{D}}{=} B$ for each $\pi \in S_m$, it follows that $\widetilde{B} \stackrel{\mathcal{D}}{=} B$. Let $\widetilde{P}(\cdot \mid B)$, $\widetilde{E}(\cdot \mid B)$, and $\widetilde{\mathrm{Var}}(\cdot \mid B)$ denote the conditional probability, expectation, and variance, respectively, given B.

We now use this framework to study the moments of regular and permuted scaling functions applied to B. First, suppose that $p \geq 1$. Then,

$$E\big[\tilde{b}_{m,p}(B)^p\big] = E\big[\widetilde{E}[b(\widetilde{B})^p \mid B]\big] = E[b(\widetilde{B})^p] = E[b(B)^p] \tag{10}$$

because $\widetilde{B} \stackrel{\mathcal{D}}{=} B$. Moreover, a variance decomposition yields

$$\mathrm{Var}[b(B)^p] = \mathrm{Var}[b(\widetilde{B})^p] = \mathrm{Var}[\widetilde{E}[b(\widetilde{B})^p \mid B]] + E[\widetilde{\mathrm{Var}}[b(\widetilde{B})^p \mid B]]. \tag{11}$$

Thus, $\mathrm{Var}[\tilde{b}_{m,p}(B)^p] \leq \mathrm{Var}[b(B)^p]$ because $E[\widetilde{\mathrm{Var}}[b(\widetilde{B})^p \mid B]] \geq 0$ and $\tilde{b}_{m,p}(B)^p = \widetilde{E}[b(\widetilde{B})^p \mid B]$.

For the case when $p > 1$, Jensen's inequality ensures

$$E\big[\tilde{b}_{m,p}(B)\big] = E\big[(\widetilde{E}[b(\widetilde{B})^p \mid B])^{1/p}\big] \geq E\big[\widetilde{E}[b(\widetilde{B}) \mid B]\big] = E[b(B)]. \tag{12}$$

Therefore, when $p = 2$, (10) and (12) imply $\mathrm{Var}[\tilde{b}_{m,p}(B)] \leq \mathrm{Var}[b(B)]$. We summarize the above discussion in the following theorem.

THEOREM 5.1. *Suppose that b is a scaling function. Then, for any number of segments $m \geq 1$,*
(i) $E[\tilde{b}_{m,p}(B)^p] = E[b(B)^p]$ *when* $p \geq 1$,
(ii) $\mathrm{Var}[\tilde{b}_{m,p}(B)^p] \leq \mathrm{Var}[b(B)^p]$ *when* $p \geq 1$,
(iii) $E[\tilde{b}_{m,p}(B)] \geq E[b(B)]$ *when* $p > 1$,
(iv) $\mathrm{Var}[\tilde{b}_{m,p}(B)] \leq \mathrm{Var}[b(B)]$ *when* $p = 2$.

While Theorem 5.1 (also see Proposition 4.1) compares some of the moments of the limiting permuted estimator $\tilde{b}_{m,p}(B)$ to those of the asymptotic regular estimator $b(B)$, these results do not allow us to directly determine the impact of permuting on the mean and variance of the widths of the limiting confidence interval. As seen in (4) and (5), these measures of confidence-interval width also depend on the quantile points $\gamma_{\tilde{b}_{m,p},1-\delta/2}$ and $\gamma_{b,1-\delta/2}$, whose values Theorem 5.1 does not address. However, we later provide theoretical and empirical results for specific STS methods (see also Calvin and Nakayama [5, 6]) showing that permuting appears to reduce both the mean and variance of confidence-interval widths compared to the regular estimator.

5.1. Permuted maximum estimator. We now apply permutations to the maximum estimator discussed in §3. Fix $m \geq 2$ and $p = 1$. For $x \in \Omega_{00}$ and $i = 1, 2, \ldots, m$, let $Z_i(x) = x(i/m) - x((i-1)/m)$, which is the increment of the ith segment. Then, straightforward calculations show that setting $b = b_{\max}$ in (9) yields the permuted maximum scaling function for m segments and $p = 1$ as

$$\tilde{b}_{\max,\,m,\,1}(x) = \frac{1}{m!}\sum_{\pi \in S_m} \max_{i=1,\ldots,m}\left\{ \max_{0\leq s\leq 1/m}\left[x\left(s + \frac{\pi(i)-1}{m}\right) - x\left(\frac{\pi(i)-1}{m}\right)\right] + \sum_{j=1}^{i-1} Z_{\pi(j)}(x)\right\}. \tag{13}$$

For general $m > 2$, it appears that this cannot be simplified further to avoid the sum over the $m!$ terms in S_m, thereby limiting the size of m that can be used in practice. However, in the case when $m = 2$, the permuted estimator has a simple form, which we now describe.

For $x \in \Omega_{00}$, let $\tilde{x}$ be the sample path obtained by swapping the two halves of x. Specifically, let $x = [x(t)\colon 0 \leq t \leq 1]$, and define $\tilde{x} = [\tilde{x}(t)\colon 0 \leq t \leq 1]$ such that

$$\tilde{x}(t) = \begin{cases} x\left(t + \frac{1}{2}\right) - x\left(\frac{1}{2}\right) & \text{if } 0 \leq t \leq \frac{1}{2}, \\ x\left(t - \frac{1}{2}\right) + x(1) - x\left(\frac{1}{2}\right) & \text{if } \frac{1}{2} < t \leq 1. \end{cases}$$

But $x(1) = 0$ for $x \in \Omega_{00}$, so $\tilde{x}(t) = x'(t) - x(1/2)$, where

$$x'(t) = \begin{cases} x\left(t + \frac{1}{2}\right) & \text{if } 0 \leq t \leq \frac{1}{2}, \\ x\left(t - \frac{1}{2}\right) & \text{if } \frac{1}{2} < t \leq 1. \end{cases}$$

Hence, for any scaling function b, the corresponding permuted scaling function when $m = 2$ and $p = 1$ is

$$\tilde{b}_{2,\,1}(x) = \frac{1}{2}\big(b(x) + b(\tilde{x})\big) = \frac{1}{2}\big(b(x) + b(x' - x(1/2))\big).$$

For the maximum function $b_{\max}$ defined in (6), note that $b_{\max}(x' - x(1/2)) = b_{\max}(x) - x(1/2)$ for $x \in \Omega_{00}$. Hence, the permuted maximum scaling function with $m = 2$ and $p = 1$ is

$$\tilde{b}_{\max,\,2,\,1}(x) = b_{\max}(x) - \frac{1}{2}x\left(\frac{1}{2}\right). \tag{14}$$

Applying this to the limiting Brownian bridge B, we get

$$\tilde{b}_{\max,\,2,\,1}(B) = b_{\max}(B) - \frac{1}{2}B\left(\frac{1}{2}\right), \tag{15}$$

and we can view the permuted estimator in the limit as the original minus a control variate, $B(1/2)/2$, because $E[B(1/2)] = 0$.

From the representation (14) we see that the additional work to compute the permuted estimator is just subtracting half the midpoint of the normalized process. The proof of the following proposition appears in the appendix.

Proposition 5.1. *For $b_{\max}$ defined as in (6) and permuting with $m = 2$ segments and $p = 1$,*

$$H(x) = 1 - \frac{1}{\pi}\tan^{-1}\left(\frac{4}{x}\right) + \frac{2\sqrt{2}x}{\pi\sqrt{8+x^2}}\tan^{-1}\left(\sqrt{\frac{8}{8+x^2}}\right) - \frac{4x}{\pi(16+x^2)}$$

and

$$E\big[\tilde{b}_{\max},(B)\big] = \frac{\sqrt{2\pi}}{4}, \qquad \mathrm{Var}\big[\tilde{b}_{\max,\,2,\,1}(B)\big] = \frac{7-2\pi}{16}.$$

Some of the critical points of the distribution H for $\tilde{b}_{\max,\,2,\,1}$ are given in the last row of Table 1. These values were calculated numerically. Table 2 provides the limiting half widths' means and variabilities of the corresponding confidence intervals.

For the sake of comparison, we now consider the batched maximum estimator. The proof of the following proposition appears in the appendix.

PROPOSITION 5.2. *For $b_{\max}$ as defined in (6) and batching with $m = 2$ batches and $p = 1$,*

$$H(x) = \frac{1}{2}\left(1 + \frac{x\sqrt{x^2+16}}{x^2+8}\right), \quad x \geq 0. \tag{16}$$

Also,

$$E\left[b^*_{\max,2,1}(B)\right] = \frac{\sqrt{2\pi}}{4}, \qquad \mathrm{Var}\left[b^*_{\max,2,1}(B)\right] = \frac{4-\pi}{16}.$$

Setting $H(\gamma) = 1 - \delta/2$ for $b^*_{\max,2,1}$ implies that

$$\gamma = \left(-8 + \frac{8}{\sqrt{1-(1-\delta)^2}}\right)^{1/2}.$$

The values of the critical points for $1-\delta = 0.90$, 0.95, and 0.99 are given in the second row of Table 1, and Table 2 gives the limiting half widths and variabilities of the corresponding confidence intervals.

Comparing in Table 2 the various maximum estimators in terms of limiting expected half widths and variabilities of half widths, we see that when $p = 1$, permuting with $m = 2$ is better than not permuting, and permuting with $m = 2$ segments slightly outperforms batching with $m = 2$ batches.

In §7, we empirically compare coverages of the regular, batched, and permuted maximum confidence intervals. Also, we compare batching and permuting for up to $m = 5$.

We now compare our maximum estimator with Schruben's standardized maximum estimator, which is based on the scaling function

$$b_{\mathrm{sm}}(x) = \frac{x(t^*)}{(t^*(1-t^*))^{1/2}},$$

where $t^* = \inf\{t \geq 0\colon x(t) = M\}$ and $M = \max\{x(t)\colon 0 \leq t \leq 1\}$. Schruben [11] showed that $W(1)/b_{\mathrm{sm}}(B) \stackrel{\mathcal{D}}{=} t_3/\sqrt{3}$, where t_d is a student-t random variable with d degrees of freedom. Thus, it can be shown that $E[b_{\mathrm{sm}}(B)] = 2\sqrt{2/\pi}$. Consequently, the mean half width of a 90% confidence interval corresponding to the maximum estimator is asymptotically about 19% larger than that for the standardized maximum. However, empirical results in Calvin and Nakayama [5] seem to indicate that the coverage of confidence intervals constructed using the maximum method is closer to the nominal level than those for the standardized maximum.

6. Area estimators. In this section, we apply the permutation method to the STS method based on the integral of the Brownian bridge. Known as the "area" estimator (Schruben [11]), the method has scaling function

$$b_{\mathrm{area}}(x) = \left|\int_{t=0}^{1} x(t)\,dt\right|.$$

Schruben [11] shows that $b_{\mathrm{area}}(B) \stackrel{\mathcal{D}}{=} |N(0,1/12)|$, $E[b_{\mathrm{area}}(B)^2] = 1/12$, and $\mathrm{Var}[b_{\mathrm{area}}(B)^2] = 1/72$, where $N(\alpha,\beta)$ denotes a normal random variable with mean α and variance β. For the area scaling function, we will compare batching and permuting for $m \geq 2$ and $p = 2$.

Let

$$\begin{aligned} A_i(x) &= \int_{s=(i-1)/m}^{i/m} [x(s) - \Lambda_m(x)(s)]\,ds \\ &= \int_{s=(i-1)/m}^{i/m} x(s)\,ds - \frac{1}{2m}\left[x\left(\frac{i-1}{m}\right) + x\left(\frac{i}{m}\right)\right] \qquad (17) \\ &= \frac{1}{m^{3/2}}\int_{s=0}^{1} \Gamma_{i,m}(x)(s)\,ds. \qquad (18) \end{aligned}$$

Thus, because $\Gamma_{i,m}(B)$ is a standard Brownian bridge, the $A_i(B)$, $i = 1, 2, \ldots, m$, are i.i.d. $N(0, 1/(12m^3))$.

6.1. Batching. We consider first the batched area estimator with m batches. Let χ^2_k denote a chi-squared random variable with k degrees of freedom. Schruben [11] established (20) below, and (21) is shown in Goldsman and Schruben [8].

Proposition 6.1. *The batched area scaling function $b^*_{\text{area},\,m,\,2}$ with $m \ge 1$ batches and $p = 2$ is given by*

$$b^*_{\text{area},\,m,\,2}(x) = \left(\frac{1}{m}\sum_{i=1}^{m}(b_{\text{area}} \circ \Gamma_{i,\,m}(x))^2\right)^{1/2} = \left(m^2 \sum_{i=1}^{m} A_i(x)^2\right)^{1/2} \tag{19}$$

for $x \in \Omega_{00}$. Also,

$$b^*_{\text{area},\,m,\,2}(B) \stackrel{\mathcal{D}}{=} \left(\frac{\chi^2_m}{12m}\right)^{1/2} \qquad \text{and} \tag{20}$$

$$E[b^*_{\text{area},\,m,\,2}(B)^2] = \frac{1}{12}, \qquad \operatorname{Var}[b^*_{\text{area},\,m,\,2}(B)^2] = \frac{1}{72m}. \tag{21}$$

Furthermore, the limiting random variable

$$\frac{W(1)}{b^*_{\text{area},\,m,\,2}(B)} \stackrel{\mathcal{D}}{=} \sqrt{12}\,t_m.$$

6.2. Permuting. We next consider the permuted area estimator with m segments and $p = 2$, which has scaling function

$$\tilde{b}_{\text{area},\,m,\,2}(x) = \left[E_\pi\left(\int_{s=0}^{1} T_\pi(x)(s)\,ds\right)^2\right]^{1/2}, \tag{22}$$

where the expectation E_π is with respect to randomly choosing a permutation π uniformly from S_m.

Theorem 6.1. *The permuted area scaling function $\tilde{b}_{\text{area},\,m,\,2}$ with $m \ge 1$ segments and $p = 2$ is given by*

$$\tilde{b}_{\text{area},\,m,\,2}(x) = \left[\left(\sum_{i=1}^{m} A_i(x)\right)^2 + \frac{m+1}{12m}\sum_{i=1}^{m} Z_i(x)^2\right]^{1/2} \tag{23}$$

for $x \in \Omega_{00}$. Also,

$$\tilde{b}_{\text{area},\,m,\,2}(B) \stackrel{\mathcal{D}}{=} \left(\frac{\chi^2_1 + (m+1)\chi^2_{m-1}}{12m^2}\right)^{1/2}, \tag{24}$$

where χ^2_1 and χ^2_{m-1} are independent, so

$$E\big[\tilde{b}_{\text{area},\,m,\,2}(B)^2\big] = \frac{1}{12}, \qquad \operatorname{Var}\big[\tilde{b}_{\text{area},\,m,\,2}(B)^2\big] = \frac{m^3 + m^2 - m}{72m^4}.$$

Furthermore, the limiting random variable

$$\frac{W(1)}{\tilde{b}_{\text{area},\,m,\,2}(B)} \stackrel{\mathcal{D}}{=} \frac{N(0,1)}{[(\chi^2_1 + (m+1)\chi^2_{m-1})/(12m^2)]^{1/2}}, \tag{25}$$

where $N(0,1)$, χ^2_1, and χ^2_{m-1} are mutually independent.

Proof. For $x \in \Omega_{00}$, $x(0) = x(1) = 0$, so (17) implies

$$\begin{aligned}
\int_{s=0}^{1} x(s)\,ds &= \sum_{i=1}^{m} A_i(x) + \frac{1}{2m}\sum_{i=1}^{m}\left[x\left(\frac{i-1}{m}\right) + x\left(\frac{i}{m}\right)\right] \\
&= \sum_{i=1}^{m} A_i(x) + \frac{1}{m}\sum_{i=1}^{m-1} x\left(\frac{i}{m}\right) \\
&= \sum_{i=1}^{m} A_i(x) + \frac{1}{m}\sum_{i=1}^{m-1}\sum_{j=1}^{i} Z_j(x) \\
&= \sum_{i=1}^{m} A_i(x) + \frac{1}{m}\sum_{j=1}^{m-1}(m-j)Z_j(x).
\end{aligned} \tag{26}$$

For any $\pi \in S_m$, $\sum_{i=1}^m A_{\pi(i)}(x) = \sum_{i=1}^m A_i(x)$. Thus, to compute the square of (22), we average the square of (26) over all $m!$ permutations $\pi \in S_m$ to get

$$\begin{aligned}
E_\pi\left(\int_{s=0}^1 T_\pi(x)(s)\,ds\right)^2 &= \frac{1}{m!}\sum_{\pi\in S_m}\left(\sum_{i=1}^m A_{\pi(i)}(x) + \frac{1}{m}\sum_{j=1}^{m-1}(m-j)Z_{\pi(j)}(x)\right)^2 \\
&= \left(\sum_{i=1}^m A_i(x)\right)^2 + \frac{1}{m!}\sum_{\pi\in S_m}\left(\frac{1}{m}\sum_{j=1}^{m-1}(m-j)Z_{\pi(j)}(x)\right)^2 \\
&\quad + 2\sum_{i=1}^m A_i(x)\frac{1}{m!\,m}\sum_{\pi\in S_m}\sum_{j=1}^{m-1}(m-j)Z_{\pi(j)}(x) \\
&= \left(\sum_{i=1}^m A_i(x)\right)^2 + \frac{1}{m!}\sum_{\pi\in S_m}\left(\frac{1}{m}\sum_{j=1}^{m-1}(m-j)Z_{\pi(j)}(x)\right)^2 \qquad (27)
\end{aligned}$$

because $\sum_{\pi\in S_m} Z_{\pi(j)}(x) = 0$ for each j. But, as we show below, this last expression equals

$$\left(\sum_{i=1}^m A_i(x)\right)^2 + \frac{m+1}{12}E_\pi[Z_{\pi(1)}(x)^2] \qquad (28)$$

$$= \left(\sum_{i=1}^m A_i(x)\right)^2 + \frac{m+1}{12m}\sum_{i=1}^m Z_i(x)^2, \qquad (29)$$

which yields (23). To see why (27) is the same as (28), observe that because $Z_1(x) + \cdots + Z_m(x) = 0$,

$$0 = E_\pi\left(\sum_{i=1}^m Z_{\pi(i)}(x)\right)^2 = mE_\pi(Z_{\pi(1)}(x)^2) + m(m-1)E_\pi[Z_{\pi(1)}(x)Z_{\pi(2)}(x)],$$

and so

$$E_\pi[Z_{\pi(1)}(x)\,Z_{\pi(2)}(x)] = -\frac{1}{m-1}\,E_\pi[Z_{\pi(1)}(x)^2].$$

Thus,

$$\begin{aligned}
E_\pi\left[\frac{1}{m}\sum_{j=1}^{m-1}(m-j)Z_{\pi(j)}(x)\right]^2 &= \frac{1}{m^2}E_\pi\left[\sum_{j=1}^{m-1}(m-j)^2 Z_{\pi(j)}(x)^2 + 2\sum_{j=1}^{m-2}\sum_{k=j+1}^{m-1}(m-j)Z_{\pi(j)}(x)(m-k)Z_{\pi(k)}(x)\right] \\
&= \frac{1}{m^2}\left(\sum_{j=1}^{m-1}(m-j)^2 E_\pi[Z_{\pi(j)}(x)^2] + 2\sum_{j=1}^{m-2}\sum_{k=j+1}^{m-1}(m-j)(m-k)E_\pi[Z_{\pi(j)}(x)Z_{\pi(k)}(x)]\right) \\
&= \frac{1}{m^2}\left(E_\pi[Z_{\pi(1)}(x)^2]\sum_{j=1}^{m-1}(m-j)^2 + 2E_\pi[Z_{\pi(1)}(x)Z_{\pi(2)}(x)]\sum_{j=1}^{m-2}\sum_{k=j+1}^{m-1}(m-j)(m-k)\right) \\
&= \frac{1}{m^2}E_\pi[Z_{\pi(1)}(x)^2]\left(\sum_{j=1}^{m-1}(m-j)^2 - \frac{2}{m-1}\sum_{j=1}^{m-2}\sum_{k=j+1}^{m-1}(m-j)(m-k)\right) \\
&= \frac{1}{m^2}E_\pi[Z_{\pi(1)}(x)^2]\frac{m^2(m+1)}{12} \\
&= \frac{m+1}{12}E_\pi[Z_{\pi(1)}(x)^2],
\end{aligned}$$

thereby establishing the equality of (27) and (28).

We now evaluate (29) at $x = B$. The $A_i(B) \stackrel{\mathcal{D}}{=} N(0, 1/(12m^3))$, $i = 1, \ldots, m$, are independent, so (29) with $x = B$ is equal in distribution to

$$\frac{1}{12m^2}\left(N(0,1)^2 + m(m+1)\sum_{i=1}^m Z_i(B)^2\right).$$

Glynn and Iglehart [7] show that $m\sum_{i=1}^m Z_i(B)^2 \stackrel{\mathscr{D}}{=} \chi^2_{m-1}$, so

$$\left[E_\pi\left(\int_{s=0}^1 T_\pi(B)(s)\,ds\right)^2\right]^{1/2} \stackrel{\mathscr{D}}{=} \frac{1}{m\sqrt{12}}\left(\chi_1^2+(m+1)\chi^2_{m-1}\right)^{1/2},$$

where χ_1^2 and χ^2_{m-1} are independent. Also, it is clear that (25) holds, and the expectation and variance of $\tilde{b}_{\text{area},\,m,\,2}(B)^2$ follow from the fact that $E[\chi_k^2]=k$ and $\text{Var}[\chi_k^2]=2k$. □

From the representation (23), we see that for a simulation of length n and with m segments, the permuted estimator can be computed in time $\Theta(n+m)$, which is similar to that for the batched estimator for m batches. Thus, it seems reasonable to choose $m=o(n)$ so that the computational overhead is negligible as $n\to\infty$.

6.3. Comparisons. We now compare the batched and permuted area scaling functions applied to a Brownian bridge as the number m of batches or segments grows large. Glynn and Iglehart [7, Proposition 3.6] shows that for the batched scaling function $b^*_{m,\,2}$ with m batches and $p=2$ based on *any* scaling function b,

$$b^*_{m,\,2}(B) \xrightarrow{\mathscr{D}} \sqrt{E[b(B)^2]}$$

as $m\to\infty$ whenever $E[b(B)^2]<\infty$. Thus, in the case of the batched area estimator, because $E[b_{\text{area}}(B)^2]=1/12$,

$$b^*_{\text{area},\,m,\,2}(B) \xrightarrow{\mathscr{D}} \sqrt{1/12}$$

as $m\to\infty$. Now for the permuted area estimator, (24) and the law of large numbers imply that

$$\tilde{b}_{\text{area},\,m,\,2}(B) \xrightarrow{\mathscr{D}} \sqrt{1/12} \tag{30}$$

as $m\to\infty$. Hence, the batched and permuted area estimators applied to the Brownian limit become asymptotically equivalent in distribution as the number m of batches and segments grows large.

We now discuss some connections between the permuted area and batch means estimators. Glynn and Iglehart [7] show that the batch means method with $m\geq 2$ batches has scaling function

$$b_{\text{bm},\,m}(x)=\left(\frac{m}{m-1}\sum_{i=1}^m Z_i(x)^2\right)^{1/2} \tag{31}$$

for $x\in\Omega_{00}$, and $b_{\text{bm},\,m}(B)\stackrel{\mathscr{D}}{=}(\chi^2_{m-1}/(m-1))^{1/2}$. Hence,

$$b_{\text{bm},\,m}(B) \xrightarrow{\mathscr{D}} 1 \tag{32}$$

as $m\to\infty$ by the law of large numbers, and

$$\frac{W(1)}{b_{\text{bm},\,m}(B)} \stackrel{\mathscr{D}}{=} t_{m-1} \xrightarrow{\mathscr{D}} N(0,1)$$

as $m\to\infty$. From (19) and (31), we then see that the permuted area scaling function in (23) can be expressed as

$$\begin{aligned}\tilde{b}_{\text{area},\,m,\,2}(x) &= \left[\left(\sum_{i=1}^m A_i(x)\right)^2+\frac{m^2-1}{12m^2}b_{\text{bm},\,m}(x)^2\right]^{1/2} \qquad (33)\\ &= \left[\frac{b^*_{\text{area},\,m,\,2}(x)^2}{m^2}+\sum_{i\neq j}A_i(x)A_j(x)+\frac{m^2-1}{12m^2}b_{\text{bm},\,m}(x)^2\right]^{1/2},\end{aligned}$$

so the square of the permuted area estimator can be expressed as a linear combination of the squares of the batched area and batch means estimators and some other terms. Now evaluating (33) at $x=B$ and using (18) yields

$$\tilde{b}_{\text{area},\,m,\,2}(B)=\left[\left(\sum_{i=1}^m\frac{1}{m^{3/2}}\int_0^1\Gamma_{i,\,m}(B)(s)\,ds\right)^2+\frac{m^2-1}{12m^2}b_{\text{bm},\,m}(B)^2\right]^{1/2}. \tag{34}$$

Because $\int_0^1\Gamma_{i,\,m}(B)(s)\,ds$, $i=1,2,\ldots,m$, are i.i.d. $N(0,1/12)$, we see that

$$\sum_{i=1}^m\frac{1}{m^{3/2}}\int_0^1\Gamma_{i,\,m}(B)(s)\,ds \xrightarrow{\mathscr{D}} 0$$

as $m \to \infty$ by the law of large numbers. Also,

$$\frac{m^2-1}{12m^2} b_{\mathrm{bm},\,m}(B)^2 \xrightarrow{\mathscr{D}} \frac{1}{12}$$

as $m \to \infty$ by (32). Thus, (30) and the representation in (34) show that the permuted area estimator applied to a Brownian bridge becomes more and more like the batch means estimator, appropriately scaled, as m gets large.

7. Numerical experiments. We ran simulations to study the coverages, as well as average and variance of half width, of 90% confidence intervals based on the methods discussed in this paper. The model simulated is the embedded discrete-time Markov chain of the number of customers in an $M/M/1/100$ queue, i.e., an $M/M/1$ queue on a truncated state space $S = \{0, 1, 2, \ldots, 100\}$. For each experiment, we ran 10^5 independent replications, each time selecting the initial state using the steady-state distribution.

For the first set of experiments, we used traffic intensities $\rho = 0.3$, 0.5, 0.7, and 0.9. We considered three methods: the regular maximum (i.e., with no permuting and no batching), the permuted maximum, and the batched maximum. The permuted estimators were constructed using $m = 2$ segments and the batched estimator with $m = 2$ batches, and both with $p = 1$. In each replication, the run length n corresponds to the number of transitions, which we varied between 500 and 5,000.

Figure 1 shows the observed coverages for the confidence intervals. Note that the coverages for the regular maximum and for permuting are always about the same, and the coverage for batching is consistently less. Moreover, as ρ increases, the difference between batching and the other two methods grows larger.

Figure 2 shows the average half widths of the confidence intervals. For all values of ρ, permuting and batching lead to smaller intervals than the regular maximum. For small values of ρ, permuting and batching perform about equally. However, for larger ρ, batching yields smaller intervals, but as we see in Figure 1, the coverage for batching is significantly compromised. Thus, it appears that the higher autocorrelations for large ρ result

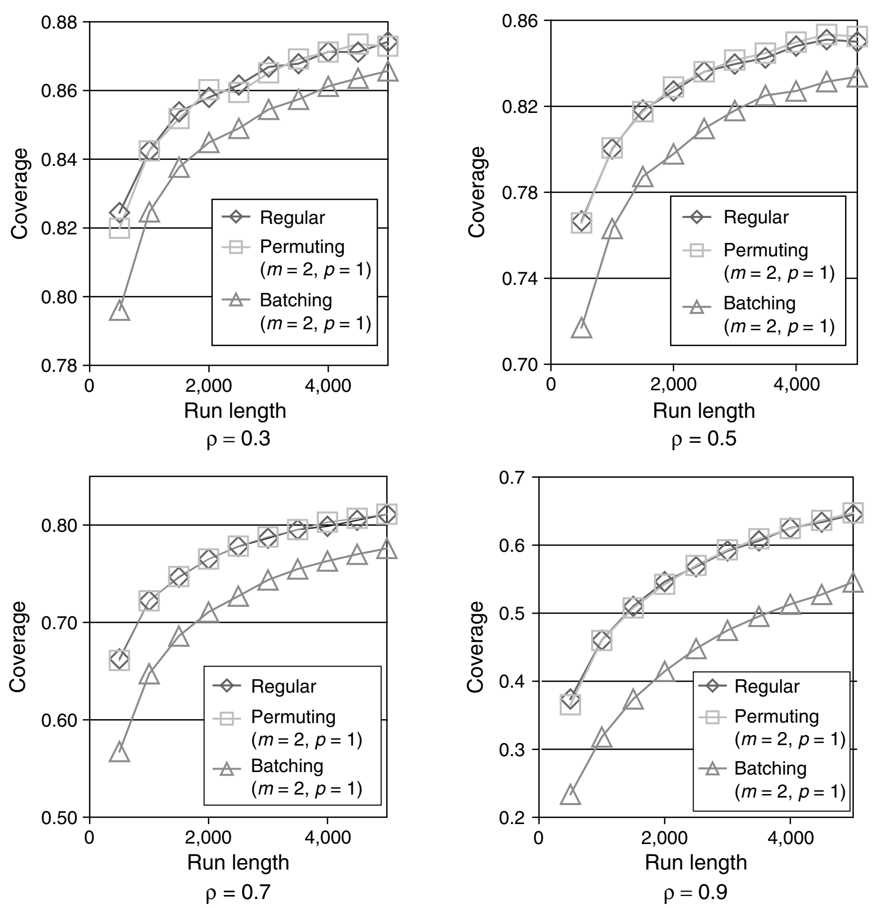

FIGURE 1. Coverage for maximum estimators.

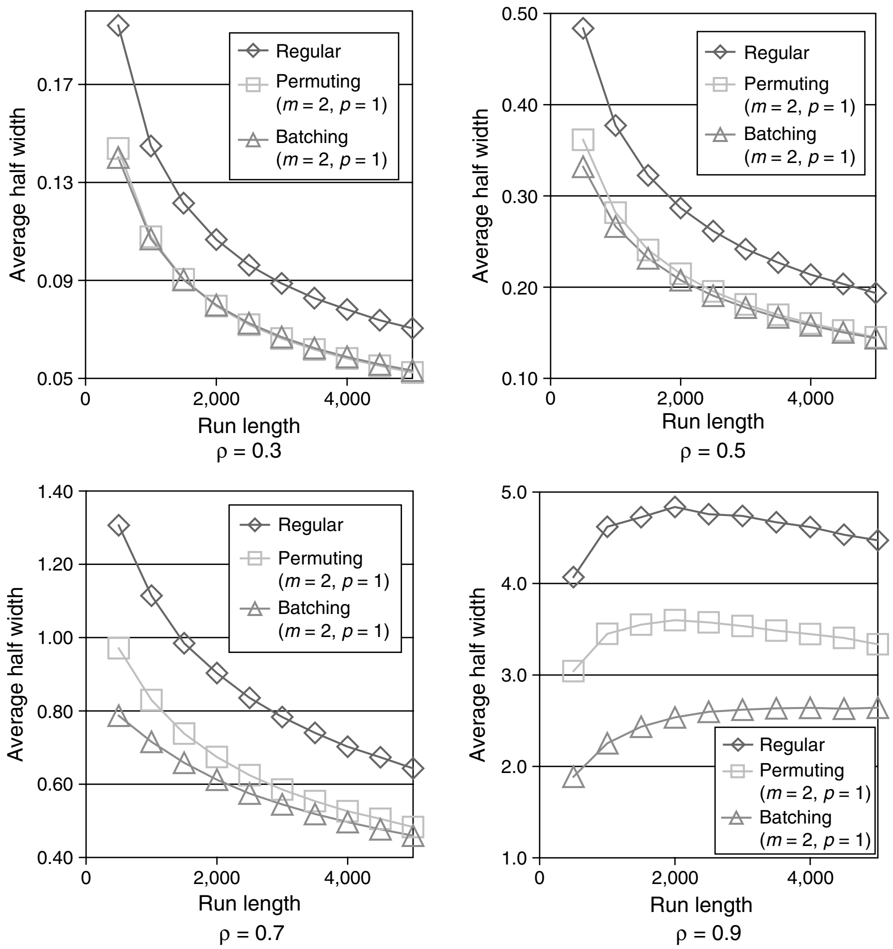

FIGURE 2. Average half widths for maximum estimators.

in batching producing confidence intervals that are too short, which causes poor coverage. Figure 3 displays similar results for the variability of the half widths.

We chose to run simulations with short run lengths so that we can easily discern differences in coverage. We also ran the same experiments with run lengths varying between 10^5 and 10^6 transitions (results not included in this paper), and the results are similar to the ones discussed above, except that it is difficult to discern differences in coverage because they then are all close to the nominal level. Also, for the longer run lengths, the results for the average and variance of half width when $\rho = 0.9$ are in line with those for the smaller values of ρ.

We also performed experiments with the $M/M/1/100$ model with traffic intensity $\rho = 0.8$ to compare constructed 90% confidence intervals using permuted and batched estimators as the number m of batches and segments increases with the same fixed total run length. The run length in all experiments was $n = 10^4$ transitions, and we varied m from 1 to 5, where $m = 1$ corresponds to the regular maximum. Tables 3 and 4 give the results for the maximum and area estimators, respectively. Also, Table 5 presents experimental results applying batching and permuting to Goldsman and Schruben's [9] weighted area STS method with weighting function $q(t) = \sqrt{840}(3t^2 - 3t + 1/2)$. (See Calvin and Nakayama [6] for other empirical results using this estimator.)

The critical points for the regular maximum, permuted maximum with $m = 2$, and batched maximum with $m = 2$ are given in Table 1. From Proposition 6.1, the critical points for the batched area estimator with m batches are just those of a student-t_m distribution multiplied by $\sqrt{12}$. Goldsman and Schruben [9] show that the critical points of the batched weighted area estimator have a similar form. For all of the other estimators, we estimated the critical points using simulations with 10^6 independent replications. In each replication, we generated samples of appropriately distributed random variables, and substituted these into the appropriate formulae. In most cases, the random variables were multivariate normals with an appropriate covariance structure, but for the maximum estimators, we generated observations from the distribution of the maximum of a Brownian bridge, which can be determined from Shepp [12].

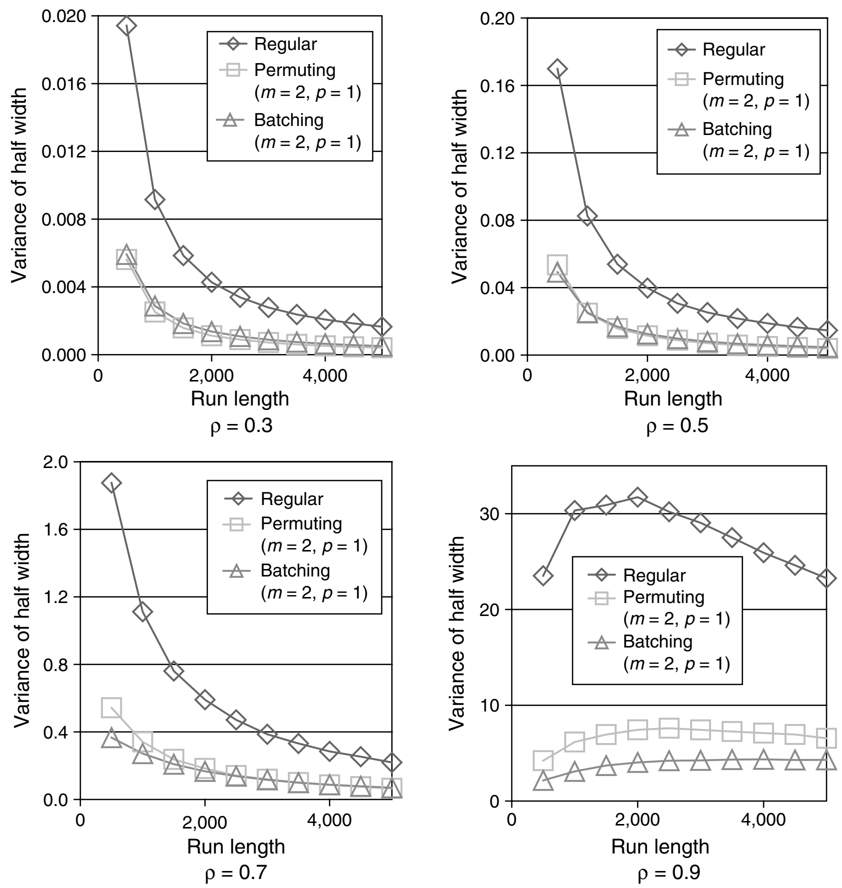

FIGURE 3. Variance of half widths for maximum estimators.

For the maximum estimators (Table 3), there is almost no degradation in coverage for permuting as m increases, at least up to 5, but there is significant degradation for batching. For each m, batching always gives confidence intervals whose widths are smaller and less variable than permuting. However, comparing permuting with $m = 3$ and batching with $m = 2$, the average half widths are roughly the same, but permuting has less variable half widths and better coverage.

Now looking at the results for the area estimators (Table 4), we see that as m increases, coverages for both permuting and batching decrease. The reason for this is that as the number m of segments, or batches, increases, the estimators are based on more values of the process, so the Brownian approximation for both permuting and batching becomes more sensitive. This suggests that in practice, one should not choose m too large. However, coverage degrades more rapidly for batching than for permuting. As with the maximum estimators, for each m, batching yields confidence-interval half widths that are smaller and less variable than those for permuting. But choosing different values of m for batching and permuting so that the average half widths are roughly the same (e.g., $m = 3$ for batching and $m = 4$ for permuting), we see that permuting yields better coverage and less variability.

The results for the weighted area estimators (Table 5) are mostly similar to those of the area estimators (Table 4), but with some slight differences. Comparing batching for the area and weighted area scaling functions, we see that the batched weighted area estimators have slightly better coverages, but the intervals are slightly wider on average and more variable. Comparing the results for the area and weighted area estimators (Tables 4 and 5) with those for the maximum estimators (Table 3), we see that the area estimators consistently have higher coverage but also larger and more variable half widths.

For comparison, we also include in Table 6 empirical results for the same simulation runs when applying batch means, which has scaling function given in (31). For each m, batch means seems to give slightly better coverage than either permuting or batching with the area and weighted area estimators, but this is at the expense of larger and more variable confidence intervals.

In all of our experiments, permuting outperforms batching in terms of coverage. This may be due to the fact that when permuting, we apply the scaling function b to each entire permuted sample path, whereas batching

TABLE 3. Empirical results for maximum estimators.

m	Method	Cov (%)	Avg HW	Var HW
1	Regular	79.8	1.126	0.732
2	Permuting	79.9	0.844	0.221
2	Batching	76.0	0.794	0.227
3	Permuting	79.7	0.787	0.151
3	Batching	72.7	0.686	0.128
4	Permuting	79.5	0.762	0.126
4	Batching	70.1	0.620	0.086
5	Permuting	79.4	0.749	0.113
5	Batching	67.4	0.574	0.062

TABLE 4. Empirical results for area estimators.

m	Method	Cov (%)	Avg HW	Var HW
1	Regular	88.5	2.618	5.614
2	Permuting	87.2	1.408	0.999
2	Batching	85.9	1.304	0.872
3	Permuting	86.5	1.166	0.531
3	Batching	84.3	1.074	0.478
4	Permuting	86.1	1.060	0.375
4	Batching	83.4	0.977	0.339
5	Permuting	85.8	1.001	0.300
5	Batching	82.7	0.921	0.273

applies the b function to each of the smaller batches individually. Thus, under batching, b is applied to shorter pieces of the sample path, which requires that the Brownian approximation implied by the FCLT in (1) holds for each of the smaller batches. But for permuting, because b is applied to each entire permuted path, the FCLT approximation is based on a larger sample.

We also reran the experiments in Tables 3–6 with $n = 2 \times 10^5$ transitions, which we now discuss but do not include the explicit results here. As one might expect, the coverages for the area, weighted area, and batch means confidence intervals, which were already fairly close to the nominal level with the shorter run length of $n = 10^4$, are now all just about at the nominal level, and it is difficult to discern differences. Moreover, the results for the average and variance of half width for these methods with the longer run length exhibit the same trends that we see in Tables 4–6.

In our experiments with the maximum estimator for the longer run length, the coverages are much closer to the nominal level than what Table 3 shows. There are still distinguishable differences in coverage, and the qualitative results are similar to those in Table 3. In addition, the average half widths for batching and permuting for each m are now much closer (within 6% of each other). Finally, for the longer run length, permuting leads to less variability in half width than batching for each m, which is the opposite of what Table 3 shows. One reason for this may be that for batching, the asymptotics were not yet in effect for the shorter run length, as seen by the low coverages in Table 3, and this led to intervals that were too small and having low variability.

8. Conclusions. In this paper, we showed how to construct a new STS method from an existing one by averaging the STS scaling function over all paths obtained by permuting path segments. Analytical results and numerical experiments indicate that permuting improves on the original STS method on which they are based. Compared to batching, an alternative extension procedure, permuting seems to always yield confidence intervals with better coverage. We carried out our analyses on three STS methods: a new maximum estimator, Schruben's [11] area estimator, and Goldsman and Schruben's [9] weighted area estimator. Permuting can also be applied to other STS methods, but the main purpose of this paper is to introduce and illustrate the use of permuting on a few examples and to show that it can improve upon existing STS methods.

There are several possible directions for future work. While we were able to derive representations of the permuted area and permuted weighted area estimators that avoid explicitly summing over all permutations, we could only provide an expression for the permuted maximum estimator for $m > 2$ segments that sums over $m!$ terms. It would be interesting to develop a characterization for when an arbitrary STS method will admit an efficient representation for its permuted scaling function. Also, if certain permuted estimators do not have an efficient representation, can significant benefits still be obtained by averaging over only a subset of the permutations?

TABLE 5. Empirical results for weighted area estimators.

m	Method	Cov (%)	Avg HW	Var HW
1	Regular	88.9	2.700	5.694
2	Permuting	88.2	2.071	3.202
2	Batching	86.6	1.341	0.914
3	Permuting	86.4	1.095	0.430
3	Batching	85.3	1.108	0.502
4	Permuting	85.8	1.006	0.313
4	Batching	84.6	1.014	0.364
5	Permuting	85.5	0.966	0.265
5	Batching	84.0	0.958	0.288

TABLE 6. Empirical results for batch means estimators.

m	Method	Cov (%)	Avg HW	Var HW
2	Batch means	89.2	2.675	5.282
3	Batch means	87.7	1.349	0.813
4	Batch means	86.7	1.119	0.445
5	Batch means	86.1	1.026	0.326

Appendix.

Proof of Proposition 5.1. Conditional on $B(1/2)$, the maxima over $[0,1/2]$ and $[1/2,1]$ are independent, each with distribution

$$P(\max \le y) = 1 - \exp(-4y(y - B(1/2)))$$

for $y > \max(B(1/2), 0)$ (see, for example, Shepp [12]). Also, $B(1/2) \sim N(0, 1/4)$. By (15), $\tilde{b}_{\max,2,1}(B) = b_{\max}(B) - B(1/2)/2$. Therefore, for $y \ge 0$,

$$\begin{aligned}
G(y) &= P\big(\tilde{b}_{\max,2,1}(B) \le y\big) \\
&= \int_{x=-\infty}^{\infty} P\big(\tilde{b}_{\max,2,1}(B) \le y \mid B(1/2) = x\big)\sqrt{\frac{2}{\pi}}\exp(-2x^2)\,dx \\
&= \int_{x=-2y}^{2y} \big(1 - \exp(-4(y + x/2)(y - x/2))\big)^2 \sqrt{\frac{2}{\pi}}\exp(-2x^2)\,dx \\
&= \int_{x=-2y}^{2y} \sqrt{\frac{2}{\pi}}\exp(-2x^2)\,dx - 2\int_{x=-2y}^{2y} \exp(-4(y+x/2)(y-x/2))\sqrt{\frac{2}{\pi}}\exp(-2x^2)\,dx \\
&\quad + \int_{x=-2y}^{2y} \exp(-8(y+x/2)(y-x/2))\sqrt{\frac{2}{\pi}}\exp(-2x^2)\,dx \\
&= \int_{x=-2y}^{2y} \sqrt{\frac{2}{\pi}}\exp(-2x^2)\,dx - 2\exp(-4y^2)\sqrt{\frac{2}{\pi}}\int_{x=-2y}^{2y}\exp(-x^2)\,dx + \sqrt{\frac{2}{\pi}}\exp(-8y^2)\int_{x=-2y}^{2y} dx \\
&= 2\Phi(4y) - 4\sqrt{2}e^{-4y^2}\Phi(2\sqrt{2}y) + 2\sqrt{2}e^{-4y^2} + 4\sqrt{2/\pi}\,ye^{-8y^2} - 1,
\end{aligned}$$

and so

$$1 - G(y) = 2\big(\bar{\Phi}(4y) - 2\sqrt{2}e^{-4y^2}\bar{\Phi}(2\sqrt{2}y) + \sqrt{2}e^{-4y^2} - 2\sqrt{2/\pi}\,ye^{-8y^2}\big), \tag{35}$$

where $\bar{\Phi}(x) = 1 - \Phi(x)$. Then, for $x > 0$,

$$\begin{aligned}
H(x) &= \frac{1}{2} + \int_{w=0}^{\infty} \phi(w)(1 - G(w/x))\,dw \\
&= \frac{1}{2} + 2\int_{w=0}^{\infty} \phi(w)\big(\bar{\Phi}(4w/x) - 2\sqrt{2}\exp(-4(w/x)^2)\bar{\Phi}(2\sqrt{2}w/x) \\
&\qquad + \sqrt{2}\exp(-4(w/x)^2) - 2\sqrt{2/\pi}(w/x)\exp(-8(w/x)^2)\big)\,dw \\
&= \frac{1}{2} + \frac{1}{\pi}\int_{w=0}^{\infty}\int_{z=4w/x}^{\infty} \exp\left(-\frac{w^2}{2}\right)\exp\left(-\frac{z^2}{2}\right)dz\,dw \\
&\qquad - \frac{2\sqrt{2}}{\pi}\int_{w=0}^{\infty}\int_{z=2\sqrt{2}w/x}^{\infty} \exp\left(-\frac{w^2}{2}\left(1 + \frac{8}{x^2}\right)\right)\exp\left(-\frac{z^2}{2}\right)dz\,dw \\
&\qquad + \frac{2}{\sqrt{\pi}}\int_{w=0}^{\infty} \exp\left(-\frac{w^2}{2}\left(1 + \frac{8}{x^2}\right)\right)dw - \frac{4}{\pi x}\int_{w=0}^{\infty} w\exp\left(-\frac{w^2}{2}\left(1 + \frac{16}{x^2}\right)\right)dw.
\end{aligned}$$

Now, use the identity

$$\int_{w=0}^{\infty} e^{-aw^2}\int_{z=cw/x}^{\infty} e^{-z^2/2}\,dz\,dw = \frac{1}{\sqrt{2a}}\left(\frac{\pi}{2} - \tan^{-1}\left(\frac{c}{x\sqrt{2a}}\right)\right)$$

to obtain

$$\begin{aligned}
H(x) &= \frac{1}{2} + \frac{1}{\pi}\left(\frac{\pi}{2} - \tan^{-1}(4/x)\right) - \frac{2\sqrt{2}}{\pi}\frac{1}{\sqrt{1 + 8/x^2}}\left(\frac{\pi}{2} - \tan^{-1}\left(\frac{2\sqrt{2}}{x\sqrt{1 + 8/x^2}}\right)\right) \\
&\qquad + \frac{2}{\sqrt{\pi}}\sqrt{\frac{\pi}{2 + 16/x^2}} - \frac{4}{\pi x}\left(\frac{1}{1 + 16/x^2}\right) \\
&= 1 - \frac{1}{\pi}\tan^{-1}(4/x) + \frac{2\sqrt{2}x}{\pi\sqrt{8 + x^2}}\tan^{-1}\left(\sqrt{\frac{8}{8 + x^2}}\right) - \frac{4x}{\pi(16 + x^2)}.
\end{aligned}$$

The mean of $\tilde{b}_{\max,2,1}(B)$ is determined by applying Theorem 5.1(i) with $p=1$ and Proposition 3.1. To compute its second moment, note that

$$\begin{aligned}E[\tilde{b}^2_{\max,2,1}(B)] &= 2\int_{y=0}^{\infty} y(1-G(y))\,dy\\ &= 4\int_{y=0}^{\infty} y\overline{\Phi}(4y)\,dy + 8\sqrt{2}\int_{y=0}^{\infty} y\exp(-4y^2)\Phi(2\sqrt{2}y)\,dy - 4\sqrt{2}\int_{y=0}^{\infty} y\exp(-4y^2)\,dy\\ &\quad - 8\sqrt{2/\pi}\int_{y=0}^{\infty} y^2\exp(-8y^2)\,dy = \frac{7}{16}. \quad \square\end{aligned}$$

Proof of Proposition 5.2. Let $M_i = b_{\max}(\Gamma_{i,2}(B))$, $i=1,2$, which are independent. Because $\Gamma_{i,2}(B) \overset{\mathcal{D}}{=} B$, (7) implies

$$\begin{aligned}G(y) &= P\left(\frac{1}{2}(M_1+M_2)\le y\right)\\ &= \int_{z=0}^{2y} P(M_1+M_2\le 2y \mid M_2 = z)4z\exp(-2z^2)\,dz\\ &= \int_{z=0}^{2y}(1-\exp(-2(2y-z)^2))4z\exp(-2z^2)\,dz\\ &= 1-\exp(-8y^2)\left(1+\int_{z=0}^{2y}4z\exp(-4z^2+8yz)\,dz\right).\end{aligned}$$

Therefore, for $x>0$,

$$\begin{aligned}H(x) &= \frac{1}{2}+\int_{w=0}^{\infty}\phi(w)(1-G(w/x))\,dw\\ &= \frac{1}{2}+\int_{w=0}^{\infty}\phi(w)\exp(-8(w/x)^2)\left(1+\int_{z=0}^{2w/x}4z\exp(-4z^2+8(w/x)z)\,dz\right)dw\\ &= \frac{1}{2}+\frac{1}{\sqrt{2\pi}}\int_{w=0}^{\infty}\exp\left(-\frac{w^2}{2}-\frac{8w^2}{x^2}\right)dw\\ &\quad +\frac{1}{\sqrt{2\pi}}\int_{w=0}^{\infty}\exp\left(-\frac{w^2}{2}-\frac{8w^2}{x^2}\right)\int_{z=0}^{2w/x}4z\exp(-4z^2+8(w/x)z)\,dz\,dw\\ &= \frac{1}{2}+\frac{1}{\sqrt{2\pi}}\int_{w=0}^{\infty}\exp\left(-\frac{w^2}{2}\left(1+\frac{16}{x^2}\right)\right)dw\\ &\quad +\frac{1}{\sqrt{2\pi}}\int_{w=0}^{\infty}\exp\left(-\frac{w^2}{2}\left(1+\frac{16}{x^2}\right)\right)\int_{z=0}^{2w/x}4z\exp(-4z^2+8(w/x)z)\,dz\,dw.\end{aligned}$$

Now,

$$\int_{z=0}^{2w/x}4z\exp(-4z^2+8(w/x)z)\,dz = \exp\left(4\frac{w^2}{x^2}\right)\int_{z=0}^{2w/x}4z\exp\left(-4\left(z-\frac{w}{x}\right)^2\right)dz.$$

Make the substitution $v = z-w/x$ to get

$$=\exp\left(4\frac{w^2}{x^2}\right)\int_{v=-w/x}^{w/x}4(v+w/x)\exp(-4v^2)\,dv = \exp\left(4\frac{w^2}{x^2}\right)4(w/x)\int_{v=-w/x}^{w/x}\exp(-4v^2)\,dv.$$

Therefore,

$$\begin{aligned}&\frac{1}{\sqrt{2\pi}}\int_{w=0}^{\infty}\exp\left(-\frac{w^2}{2}\left(1+\frac{16}{x^2}\right)\right)\int_{z=0}^{2w/x}4z\exp(-4z^2+8(w/x)z)\,dz\,dw\\ &\quad=\frac{1}{\sqrt{2\pi}}\int_{w=0}^{\infty}\exp\left(-\frac{w^2}{2}\left(1+\frac{16}{x^2}\right)\right)\exp\left(4\frac{w^2}{x^2}\right)4(w/x)\int_{v=-w/x}^{w/x}\exp(-4v^2)\,dv\,dw\\ &\quad=\frac{8}{x\sqrt{2\pi}}\int_{w=0}^{\infty}\exp\left(-\frac{w^2}{2}\left(1+\frac{16}{x^2}\right)\right)\exp\left(4\frac{w^2}{x^2}\right)w\int_{v=0}^{w/x}\exp(-4v^2)\,dv\,dw\end{aligned}$$

$$= \frac{8}{x\sqrt{2\pi}} \int_{v=0}^{\infty} e^{-4v^2} \int_{w=xv}^{\infty} w \exp\left(-\frac{w^2}{2}\left(1+\frac{8}{x^2}\right)\right) dw\, dv$$
$$= \frac{8}{x\sqrt{2\pi}} \int_{v=0}^{\infty} e^{-4v^2} \frac{1}{1+8/x^2} \exp\left(-\frac{x^2v^2}{2}\left(1+\frac{8}{x^2}\right)\right) dv$$
$$= \frac{8}{x(1+8/x^2)} \frac{1}{\sqrt{2\pi}} \int_{v=0}^{\infty} \exp\left(-\frac{v^2}{2}(16+x^2)\right) dv$$
$$= \frac{8}{x(1+8/x^2)} \frac{1}{2\sqrt{16+x^2}}.$$

Therefore, for $x > 0$,

$$H(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{w=0}^{\infty} \exp\left(-\frac{w^2}{2}\left(1+\frac{16}{x^2}\right)\right) dw + \frac{8}{x(1+8/x^2)} \frac{1}{2\sqrt{16+x^2}}$$
$$= \frac{1}{2} + \frac{1}{2\sqrt{1+16/x^2}} + \frac{8}{x(1+8/x^2)} \frac{1}{2\sqrt{16+x^2}}$$
$$= \frac{1}{2}\left(1 + \frac{x\sqrt{x^2+16}}{x^2+8}\right).$$

Because M_1 and M_2 are i.i.d. and have the same distribution as $b_{\max}(B)$,

$$E\left[b^*_{\max,2,1}(B)\right] = E\left[\frac{1}{2}(M_1+M_2)\right] = E\left[b_{\max}(B)\right] = \frac{\sqrt{2\pi}}{4}$$

by Proposition 3.1. Moreover,

$$E\left[(b^*_{\max,2,1}(B))^2\right] = 2\int_{y=0}^{\infty} y(1-G(y))\, dy$$
$$= 2\int_{x=0}^{\infty} x\exp(-8x^2)\left(1 + \int_{z=0}^{2x} 4z\exp(-4z^2+8xz)\, dz\right) dx$$
$$= \frac{4+\pi}{16}. \quad \square$$

Acknowledgments. The authors thank David Goldsman for some discussions related to this work. The work of the first author was supported in part by the National Science Foundation under Grant DMI-9900117. The work of the second author was supported in part by the National Science Foundation under Grants DMI-9624469 and DMI-9900117.

References

[1] Billingsley, P. 1968. *Convergence of Probability Measures*, 2nd ed. John Wiley & Sons, New York.
[2] Breiman, L. 1992. *Probability*, SIAM, Philadelphia, PA.
[3] Calvin, J. M., M. K. Nakayama. 1998. Using permutations in regenerative simulations to reduce variance. *ACM Trans. Modeling Comput. Simulation* **8** 153–193.
[4] Calvin, J. M., M. K. Nakayama. 2000. Central limit theorems for permuted regenerative estimators. *Oper. Res.* **48** 776–787.
[5] Calvin, J. M., M. K. Nakayama. 2001. Improving standardized time series methods by permuting path segments. D. J. Medeiros, B. A. Peters, J. S. Smith, and M. W. Rohrer, eds. *Proc. 2001 Winter Simulation Conf.* IEEE, Piscataway, NJ, pp. 348–353.
[6] Calvin, J. M., M. K. Nakayama. 2004. Permuted weighted area estimators. J. S. Smith, R. G. Ingalls, M. D. Rossetti, and B. A. Peter, eds. *Proc. 2004 Winter Simulation Conf.* IEEE, Piscataway, NJ, pp. 721–727.
[7] Glynn, P. W., D. L. Iglehart. 1990. Simulation output analysis using standardized time series. *Math. Oper. Res.* **14** 1–16.
[8] Goldsman, D., L. W. Schruben. 1984. Properties of standardized time series weighted area variance estimators. *Management Sci.* **30** 1217–1225.
[9] Goldsman, D., L. W. Schruben. 1990. New confidence interval estimators using standardized time series. *Management Sci.* **36** 393–397.
[10] Law, A. M., W. D. Kelton. 1992. Simulation modeling and analysis, 3rd ed. McGraw-Hill, New York.
[11] Schruben, L. W. 1983. Confidence interval estimation using standardized time series. *Oper. Res.* **31** 1090–1108.
[12] Shepp, L. A. 1979. The joint density of the maximum and its location for a Wiener process with drift. *J. Appl. Probab.* **31** 423–427.
[13] Tokol, G., D. Goldsman, D. H. Ockerman, J. J. Swain. 1998. Standardized time series L_p-norm variance estimators for simulation. *Management Sci.* **44** 234–245.