Submitted to *Management Science*
manuscript (Please, provide the manuscript number!)

Steady-State Quantile Estimation Using Standardized Time Series

Christos Alexopoulos, David Goldsman, Athanasios Lolos
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, {christos@gatech.edu, sman@gatech.edu, thnlolos@gatech.edu}

Kemal Dinçer Dingeç
Department of Industrial Engineering, Gebze Technical University, 41400 Gebze, Kocaeli, TURKEY, kdingec@yahoo.com

James R. Wilson
Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695-7906, jwilson@ncsu.edu

We construct point estimators and confidence intervals (CIs) for a selected quantile of a steady-state simulation output process using nonsequential standardized time series (STS) procedures based on a selected number of nonoverlapping batches of outputs. Designed for analyzing dependent outputs, these procedures extend the work of Calvin and Nakayama in 2013 for analyzing independent outputs. We derive key asymptotic properties of batched-STS quantile-estimation processes as the batch size increases under mild, empirically verifiable conditions, including a geometric-moment contraction condition assumed for the output process and a functional central limit theorem assumed for a related indicator process. We show asymptotic exactness of CIs centered on the full-sample empirical quantile using the following variance-parameter estimators: (i) the mean of the STS area estimators computed from each batch; (ii) the batch size times the variance of the empirical quantiles computed from each batch; (iii) the batch size times the mean squared deviation of the empirical quantiles in (ii) away from the full-sample empirical quantile; and (iv) a linear combination of (i) with (ii) or (iii) but not both. An experimental study illustrates the performance these variance-parameter estimators and the associated CIs.

Key words: steady-state simulation; output analysis; quantile estimation; nonsequential procedure; standardized time series; method of batching

History: **Current version: stsmsr7-dg–Jan/27/2024–11:35**

---

1. Introduction

For almost seven decades, research on simulation-based estimation of the steady-state mean has grown rapidly. On the other hand, for the past four decades research on simulation-based estimation of steady-state quantiles has grown much more slowly because of the theoretical and computational difficulties encountered when the output process exhibits any of the following anomalies: (i) contamination by warm-up effects arising from non-steady-state initialization of the simulation; (ii) stochastic dependence (for example, autocorrelation) between successive outputs; (iii) a heavy-tailed marginal cumulative distribution function (c.d.f.); (iv) the lack of a marginal probability density function (p.d.f.); (v) a marginal p.d.f. that vanishes on some interval(s) of $\mathbb{R}$; (vi) a marginal p.d.f. with multiple modes or large variation; or (vii) a marginal p.d.f. with discontinuities, vertical asymptotes, or other departures from global smoothness. While problems (iv)–(vii) have relatively little impact on steady-state mean estimation, they complicate steady-state quantile estimation significantly. To address problems (i)–(iii), several steady-state quantile-estimation procedures have been developed; for a survey of these procedures, see Alexopoulos et al. (2017, p. 22:3; 2019b, §1).

This article builds on two recent sequential procedures for estimating a given steady-state p-quantile for $p \in (0, 1)$ that provide improved techniques for handling all of the problems (i)–(vii). Sequest (Alexopoulos et al. 2019b) is designed for estimation of nonextreme quantiles—i.e., $p \in [0.05, 0.95]$; and Sequem (Alexopoulos et al. 2017) is designed for estimation of near-extreme quantiles—i.e., $p \in (0.0005, 0.05) \cup (0.95, 0.9995)$. Implemented in public-domain software (Alexopoulos et al. 2019a), these automated procedures are based on batch quantile estimators (BQEs) computed from a bounded number of nonoverlapping batches to deliver confidence intervals (CIs) for a given quantile. When applied to a suite of difficult test processes, Sequest and Sequem have exhibited ease of use, close conformance to user-specified requirements on the coverage probability and precision of the CI, and substantial reductions in the sample sizes needed to satisfy the CI requirements.

In this article we develop nonsequential quantile-estimation procedures based on the method of standardized time series (STS) that can be used as alternatives or complements to nonsequential BQE-based procedures. Nonsequential STS-based mean-estimation procedures have been studied extensively (Schruben 1983, Glynn and Iglehart 1990, Goldsman et al. 1990, Alexopoulos et al. 2016). To construct a CI for a steady-state quantile, STS-based procedures use a pivotal quantity that is the ratio of two random variables (r.v.'s), where the numerator is the quantile estimator minus the true quantile, and the denominator is the square root of an STS-based estimator of the associated variance parameter divided by the sample size. The *variance parameter* (also called the time-average variance constant) is the limit of the sample size times the variance of the quantile estimator as the sample size tends to infinity. Thus the pivotal quantity is similar to Student's t-statistic. As in applications of the t-statistic to build CIs for the true mean, if the asymptotic distribution of the pivotal quantity is known, then we can build an *asymptotically exact* CI for the true quantile—i.e., a CI whose asymptotic coverage probability equals the user-specified nominal value when the sample size tends to infinity—without needing a consistent estimator of the variance parameter, which is often difficult to obtain (Alexopoulos et al. 2007; Asmussen and Glynn 2007, Chap. IV, §3; Wu 2009).

Calvin and Nakayama (2013) use Kiefer processes to develop a nonsequential STS-based quantile-estimation procedure for processes consisting of independent and identically distributed (i.i.d.) r.v.'s. Dong and Glynn (2019) formulate a class of sequential STS-based mean-estimation procedures for dependent processes such that each procedure in the class is *asymptotically valid*—i.e., each CI delivered by such a procedure is asymptotically exact—provided the following conditions hold: (i) the output process satisfies the assumption of strong approximation (Damerdji 1994, p. 496); (ii) the procedure can be represented as a function of the current simulation time and the output process observed so far, and that function satisfies certain regularity conditions (Dong and Glynn 2019, p. 334); and (iii) as the CI's precision requirement tends to zero, the pivotal quantity's denominator evaluated at the procedure's stopping time must converge weakly to an r.v. that is positive almost surely (a.s.). Using a fixed number of batches, the authors formulate sequential procedures in which the STS method is combined with the method of batch means or a variant of batch means that is derived from the standardized sum process or the standardized maximum interval process. When conditions (i) and (ii) hold, the authors conjecture that with at least four batches, condition (iii) also holds so that each of these batched-STS sequential mean-estimation procedures is asymptotically valid. To the best of our knowledge, this conjecture has not yet been proved.

In this article we merge two threads of research concerning BQE-based sequential procedures for quantile estimation (Alexopoulos et al. 2019b) and batched-STS nonsequential procedures for mean estimation (Alexopoulos et al. 2016) in order to make the following contributions:

1. We show the applicability of nonsequential batched-STS quantile-estimation procedures to handle dependent output processes. We establish basic asymptotic properties of the batched-STS quantile-estimation process under mild assumptions, including a geometric-moment contraction (GMC) Condition on the output process and a functional central limit theorem (FCLT) for an associated indicator process. The batched-STS area estimator for the variance parameter is the average of the squared weighted areas under the STS quantile-estimation process computed within each batch. Under substantially milder conditions on the user-specified weight function than those found in the current literature (Calvin and Nakayama 2013, Alexopoulos et al. 2016), we prove that as the batch size tends to infinity the batched-STS area estimator converges weakly to a properly scaled chi-squared r.v. Then we establish the asymptotic exactness of the corresponding CI centered at the full-sample empirical quantile.
2. We close a theoretical gap in Alexopoulos et al. (2019b) by showing weak convergence to a properly scaled chi-squared r.v. for the variance-parameter estimator that is the product of the batch size times the mean squared deviation of the BQEs away from the full-sample empirical quantile. Further, we establish the asymptotic exactness of the CI centered at the full-sample empirical quantile with half-length based on the latter variance-parameter estimator or on the variance-parameter estimator that is the product of the batch size times the sample variance of the BQEs.
3. We form combined variance-parameter estimators by computing a linear combination of the batched-STS area estimator described in item 1 with one of the BQE-based variance-parameter estimators described in item 2, and we show that each of these combined estimators converges weakly to a properly scaled chi-squared r.v. with nearly double the degrees of freedom (d.o.f.) of each of its components. Finally, we establish the asymptotic exactness of the corresponding CIs centered at the full-sample empirical quantile. In essence, the combined variance-parameter estimator *reuses* the sample

to yield a CI with smaller and less variable half-length compared with that of either of the constituent variance-parameter estimators. While similar combined variance-parameter estimators have been used in batched-STS mean-estimation procedures (Schruben 1983), in that context complete proofs have not been given to establish (i) the weak-convergence limits of the corresponding combined variance-parameter estimators, or (ii) the asymptotic validity of the corresponding CIs for the steady-state mean. However, the arguments used in the proofs of Theorems 4–6 below can be adapted to show (i) and (ii).

The ultimate goal of this endeavor is the construction of asymptotically valid sequential procedures that borrow elements of the batched-STS mean-estimation procedure SPSTS (Alexopoulos et al. 2016) and improve upon the BQE-based quantile-estimation procedures Sequest (Alexopoulos et al. 2019b) and Sequem (Alexopoulos et al. 2017) with regard to statistical efficiency.

The rest of this article is organized as follows. In Section 2 we formulate our main assumptions, which are empirically verifiable and weaker than those of Calvin and Nakayama (2013, p. 603) and Dong and Glynn (2019, §2). In Section 3 we establish the basic asymptotic properties of the STS quantile-estimation process, which are formalized in Theorem 2 below and are also discussed briefly by Alexopoulos et al. (2019c). Similarly in Sections 4 and 5 we establish the main asymptotic properties of the batched-STS variance-parameter estimators, which are formalized in Theorems 3 and 4 and are discussed briefly by Alexopoulos et al. (2020). Moreover, in Section 5 we prove Theorems 5 and 6, which ensure the asymptotic exactness of all the CIs examined in this article. Section 6 outlines an efficient algorithm for computing the batched-STS area estimator with time complexity dominated by the effort to sort each batch. Section 7 details some experimentation illustrating the asymptotic and finite-sample performance of all the aforementioned variance-parameter estimators and the associated CIs for selected nonextreme and near-extreme quantiles. Finally in Section 8 we summarize our findings and discuss directions for future work.

2. Preliminaries

2.1. Notation and Terminology

For given $p \in (0,1)$, we formulate point and CI estimators of the p-quantile y_p of a steady-state simulation-generated response $\{Y_k : k \geq 1\}$. Here $\mathbb{R} \equiv (-\infty, \infty)$ denotes the real numbers; $\mathbb{R}^+ \equiv [0, \infty)$ denotes the nonnegative real numbers; $\mathbb{Z} \equiv \{0, \pm 1, \pm 2, \ldots\}$ denotes the integers; and $\mathbb{Z}^+ \equiv \{1, 2, \ldots\}$ denotes the positive integers. For each $y \in \mathbb{R}$, we let $F(y) \equiv \Pr(Y \leq y)$ denote the c.d.f. of the steady-state response so that $y_p = F^{-1}(p) \equiv \inf\{y : F(y) \geq p\}$; and we let $f(y)$ denote the p.d.f. of $F(y)$.

Using the full sample $\{Y_1, \ldots, Y_n\}$ of size $n \geq 1$, we sort the outputs in ascending order to obtain the order statistics $Y_{(1)} \leq \cdots \leq Y_{(n)}$. Our point estimator of y_p is the full-sample empirical p-quantile $\widetilde{y}_p(n) \equiv Y_{(\lceil np \rceil)}$, where $\lceil \cdot \rceil$ denotes the ceiling function; and when $n = 0$ in any expression, we take $\widetilde{y}_p(n) \equiv 0$. In the rest of this subsection, $y \in \mathbb{R}$ is arbitrary. For each $k \geq 1$, we define the indicator r.v. $I_k(y) \equiv 1$ if $Y_k \leq y$, and $I_k(y) \equiv 0$ otherwise. For a sample of size $n \geq 1$, we let $\overline{I}_n(y) \equiv n^{-1} \sum_{k=1}^{n} I_k(y)$; and when $n = 0$ in any expression, we take $\overline{I}_n(y) \equiv 0$. For each $\ell \in \mathbb{Z}$, we let $\rho_{I(y)}(\ell) \equiv \mathrm{Corr}[I_k(y), I_{k+\ell}(y)]$ denote the autocorrelation at lag ℓ in the indicator process $\{I_k(y) : k \geq 1\}$.

We assume a fixed batch count $b \geq 2$ so that for $j = 1, \ldots, b$, the jth nonoverlapping batch of size $m \geq 1$ is the response subsequence $\{Y_{(j-1)m+1}, \ldots, Y_{jm}\}$. If $m \geq 1$, then from the jth batch we compute the mean of the associated indicator r.v.'s, $\overline{I}_{j,m}(y) \equiv m^{-1} \sum_{\ell=1}^{m} I_{(j-1)m+\ell}(y)$; and if $m = 0$ in any expression, then we take $\overline{I}_{j,m}(y) \equiv 0$. If $m \geq 1$, we sort the responses from the jth batch in ascending order to obtain the order statistics $Y_{j,(1)} \leq \cdots \leq Y_{j,(m)}$; and the jth BQE for y_p is the empirical p-quantile computed from the jth batch, $\widehat{y}_p(j,m) \equiv Y_{j,(\lceil mp \rceil)}$. If $m = 0$ in any expression, we take $\widehat{y}_p(j,m) \equiv 0$.

If the r.v.'s $\mathscr{S}$ and $\mathscr{U}$ have the same distribution, then we write $\mathscr{S} \stackrel{\mathrm{d}}{=} \mathscr{U}$ or $\mathscr{S} \sim \mathscr{U}$. Let $N(0,1)$ denote the standard normal distribution so the r.v. $Z \sim N(0,1)$ has c.d.f. $\Phi(z) \equiv \int_{-\infty}^{z} (2\pi)^{-1/2} \exp(-u^2/2)\,\mathrm{d}u$ for $z \in \mathbb{R}$. For $\nu \geq 1$, (i) $\mathbf{Z}_\nu \equiv [Z_1, \ldots, Z_\nu]^\mathsf{T}$ denotes the $\nu \times 1$ random vector whose elements $\{Z_k : k = 1, \ldots, \nu\}$ are i.i.d. $N(0,1)$; (ii) t_ν denotes an r.v. having Student's t-distribution with ν d.o.f.; (iii) $t_{\beta,\nu}$ denotes the β-quantile of Student's t-distribution with ν d.o.f. for $\beta \in (0,1)$; and (iv) χ^2_ν denotes a chi-squared r.v. with ν d.o.f. If $\{\mathscr{U}_n : n \geq 1\}$ is a sequence of r.v.'s and $\{a_n : n \geq 1\}$ is a sequence of nonnegative constants, then the relation $\mathscr{U}_n = O_{\text{a.s.}}(a_n)$ means there are r.v.'s $\mathfrak{U} \in \mathbb{R}^+$ and $\mathfrak{N} \in \mathbb{Z}^+$ that are bounded a.s. and satisfy $|\mathscr{U}_n| \leq \mathfrak{U} a_n$ for $n \geq \mathfrak{N}$ a.s. (Wu 2005, Equation (2))—i.e., there are finite constants $M_{\mathfrak{U}}$ and $M_{\mathfrak{N}}$ such that

$$\Pr\Big\{\mathfrak{U} \leq M_{\mathfrak{U}},\ \mathfrak{N} \leq M_{\mathfrak{N}},\ \text{and } |\mathscr{U}_n| \leq \mathfrak{U} a_n \text{ for } n \geq \mathfrak{N}\Big\} = 1. \tag{1}$$

We use the following notation and key properties of D, the space of real-valued functions on $[0,1]$ that are right-continuous with left-hand limits. Each $\zeta \in D$ is bounded on $[0,1]$ with at most countably many discontinuities; thus ζ is continuous almost everywhere (a.e.) on $[0,1]$ (Billingsley 1999, p. 122; Kolmogorov and Fomin 1975, §§28.3–28.4). Let $\|\zeta\| \equiv \sup\{|\zeta(t)| : t \in [0,1]\}$. Let Λ denote the class of strictly increasing, continuous mappings of $[0,1]$ onto itself, where $\mathbb{I} \in \Lambda$ denotes the identity map. Thus each $\lambda \in \Lambda$ must have $\lambda(0) = 0$ and $\lambda(1) = 1$. For $\zeta, \omega \in D$, let $d(\zeta,\omega) \equiv \inf_{\lambda \in \Lambda} \max\{\|\lambda - \mathbb{I}\|, \|\zeta - \omega \circ \lambda\|\}$ denote the distance between ζ and ω in the Skorohod J_1 metric on D, where $\omega \circ \lambda(t) \equiv \omega[\lambda(t)]$ for each $t \in [0,1]$ (Billingsley 1999, pp. 121–129; Whitt 2002, §3.3). Hence with the metric $d(\zeta,\omega)$, the space D is separable—i.e., it contains a countable dense subset (Billingsley 1999, Theorem 12.2). Since the definition of $d(\zeta,\omega)$ includes the case where $\lambda(t) = \mathbb{I}(t) \equiv t$ for $t \in [0,1]$, we have $d(\zeta,\omega) \leq \|\zeta - \omega\|$ for $\zeta, \omega \in D$.

2.2. Key Assumptions and Previous Results

We assume that $\{Y_k : k \geq 1\}$ and $\{I_k(y_p) : k \geq 1\}$ satisfy the following conditions.

2.2.1. Geometric-Moment Contraction (GMC) Condition. The process $\{Y_k : k \geq 0\}$ is defined by a function $\xi(\cdot)$ of a sequence of i.i.d. r.v.'s $\{\varepsilon_j : j \in \mathbb{Z}\}$ such that $Y_k = \xi(\ldots, \varepsilon_{k-1}, \varepsilon_k)$ for $k \geq 0$. Moreover, there exist constants $\psi > 0$, $C > 0$, and $r \in (0,1)$ such that for two independent sequences $\{\varepsilon_j : j \in \mathbb{Z}\}$ and $\{\varepsilon'_j : j \in \mathbb{Z}\}$ each consisting of i.i.d. r.v.'s distributed like ε_0, we have

$$\mathrm{E}\Big[\Big|\xi(\ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_k) - \xi\big(\ldots, \varepsilon'_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_k\big)\Big|^{\psi}\Big] \leq C r^k \text{ for } k = 0, 1, 2, \ldots \tag{2}$$

(Wu 2005, Equation (17)).

2.2.2. Density-Regularity (DR) Condition. The p.d.f. $f(\cdot)$ is bounded on $\mathbb{R}$ and continuous a.e. on $\mathbb{R}$; moreover, $f(y_p) > 0$, and the derivative $f'(y_p)$ exists.

REMARK 1. Theorem 4 of Wu (2005) is key, and it requires a much-stronger condition: $f(y)$ has a derivative $f'(y)$ at each $y \in \mathbb{R}$; moreover, $f(y)$ and $|f'(y)|$ are bounded on $\mathbb{R}$, and $f(y_p) > 0$. Similarly, Theorem 1 of Alexopoulos et al. (2019b) requires a somewhat-stronger condition: $f(y)$ is bounded and continuous on $\mathbb{R}$; moreover, $f(y_p) > 0$ and $f'(y_p)$ exists. In Proposition EC.1 of the Electronic Companion, we show that these theorems hold under the DR Condition stated above, which is satisfied when $f(y_p) > 0$, $f'(y_p)$ exists, and $f(y)$ is bounded on $\mathbb{R}$; however, $f(y)$ may have a countable set of discontinuities in $\mathbb{R}$, such as support end-points. Thus the DR Condition applies to the following p.d.f.'s: (i) Cauchy, exponential, extreme value (type I), inverse Gaussian, Johnson (all types), Laplace, logistic, lognormal, normal, Pareto, Pearson (type V), Rayleigh, Student's t, triangular, truncated normal, uniform, and von Mises; and (ii) beta, extreme value (types II and III), gamma, power function, and Weibull—provided the p.d.f.'s listed in (ii) have shape parameters at least equal to one. Thus the DR Condition applies to a large class of distributions that frequently arise in simulation studies. ◄

2.2.3. Short-Range Dependence (SRD) of the Indicator Process. The indicator process $\{I_k(y_p) : k \geq 1\}$ has the short-range dependence (SRD) property so that

$$0 < \sum_{\ell \in \mathbb{Z}} \rho_{I(y_p)}(\ell) \leq \sum_{\ell \in \mathbb{Z}} |\rho_{I(y_p)}(\ell)| < \infty \tag{3}$$

(Beran 1994, p. 7). Thus the variance parameters for the r.v.'s $\bar{I}_n(y_p)$ and $\widetilde{y}_p(n)$ satisfy the relations

$$\left.\begin{aligned} \sigma^2_{I(y_p)} &\equiv \lim_{n\to\infty} n\mathrm{Var}\big[\bar{I}_n(y_p)\big] = p(1-p)\sum_{\ell\in\mathbb{Z}} \rho_{I(y_p)}(\ell) \in (0,\infty), \\ \sigma^2_{\widetilde{y}_p} &\equiv \lim_{n\to\infty} n\mathrm{Var}\big[\widetilde{y}_p(n)\big] = \frac{\sigma^2_{I(y_p)}}{f^2(y_p)} \in (0,\infty). \end{aligned}\right\} \tag{4}$$

2.2.4. Functional Central Limit Theorem (FCLT) for the Indicator Process. We define the following sequence of random functions $\{\mathscr{I}_n : n \geq 1\}$ in D,

$$\mathscr{I}_n(t) \equiv \frac{\lfloor nt \rfloor}{\sigma_{I(y_p)} n^{1/2}} \big[\, \bar{I}_{\lfloor nt \rfloor}(y_p) - p\big] \text{ for } t \in [0,1] \text{ and } n \geq 1,$$

where $\lfloor \cdot \rfloor$ denotes the floor function. We assume that the random-function sequence $\{\mathscr{I}_n : n \geq 1\}$ satisfies the FCLT

$$\mathscr{I}_n \underset{n\to\infty}{\Longrightarrow} \mathscr{W}$$

in D, where $\mathscr{W}$ denotes standard Brownian motion on $[0,1]$; and $\underset{n\to\infty}{\Longrightarrow}$ denotes weak convergence as $n \to \infty$ ((Billingsley 1999, pp. 1–6, §§2–3, §8); (Whitt 2002, §3.2, §§4.2–4.4, §11.3)).

2.2.5. Previous Results on Quantile Estimation. Under the GMC, DR, SRD, and FCLT Conditions, we derive the following results, which are used throughout this article and which are based on minor modifications to the proofs of Theorem 4 of Wu (2005) and Theorem 1 of Alexopoulos et al. (2019b) as detailed in Proposition EC.1 of the Electronic Companion.

Theorem 1. *If $b \geq 2$, $m \geq 1$, $\{Y_k : k \geq 1\}$ satisfies the GMC and DR Conditions, and $\{I_k(y_p) : k \geq 1\}$ satisfies the SRD and FCLT Conditions, then*

$$\widehat{y}_p(j,m) = y_p - \frac{\overline{I}_{j,m}(y_p) - p}{f(y_p)} + O_{\text{a.s.}}\left[\frac{(\log m)^{3/2}}{m^{3/4}}\right] \text{ as } m \to \infty \text{ for } j = 1,\ldots,b; \text{ and} \tag{5}$$

$$m^{1/2}\left[\widehat{y}_p(1,m) - y_p, \ldots, \widehat{y}_p(b,m) - y_p\right]^{\mathsf{T}} \underset{m\to\infty}{\Longrightarrow} \sigma_{\widetilde{y}_p}\mathbf{Z}_b \tag{6}$$

in $\mathbb{R}^b$ with the standard Euclidean metric.

Equation (5) is the Bahadur representation for $\widehat{y}_p(j,m)$. To elaborate the significance of Theorem 1 as it is used in what follows, we state an immediate consequence of Theorem 1 as it applies to the Bahadur representation for the full-sample quantile estimator $\widetilde{y}_p(n)$ and its associated mean indicator r.v. $\overline{I}_n(y_p)$.

Corollary 1. *If the assumptions of Theorem 1 hold, then associated with the Bahadur representation for $\widetilde{y}_p(n)$ there are r.v.'s $\mathfrak{U} \in \mathbb{R}^+$ and $\mathfrak{N} \in \mathbb{Z}^+$, which are bounded a.s. such that the truncated remainder,*

$$Q_n \equiv \begin{cases} \widetilde{y}_p(n) - y_p + \dfrac{\overline{I}_n(y_p) - p}{f(y_p)} & \text{for } n \geq \mathfrak{N}, \\ 0 & \text{for } 0 \leq n \leq \mathfrak{N} - 1, \end{cases} \tag{7}$$

satisfies the relation

$$|Q_n| \leq \mathfrak{U}\frac{(\log n)^{3/2}}{n^{3/4}} \text{ for } n \geq 1 \text{ a.s.} \tag{8}$$

Thus the Bahadur representation for $\widetilde{y}_p(n)$ has the form

$$\widetilde{y}_p(n) = y_p - \frac{\overline{I}_n(y_p) - p}{f(y_p)} + O_{\text{a.s.}}\left[\frac{(\log n)^{3/2}}{n^{3/4}}\right] \text{ as } n \to \infty, \tag{9}$$

and the central limit theorem (CLT) for quantile estimation in dependent sequences has the form

$$n^{1/2}\left[\widetilde{y}_p(n) - y_p\right] \underset{n\to\infty}{\Longrightarrow} \sigma_{\widetilde{y}_p} Z. \tag{10}$$

2.3. Methods for Verifying the Key Assumptions

In this subsection we discuss the theoretical and practical considerations underlying the GMC, DR, SRD, and FCLT assumptions together with methods for verifying those assumptions in practice.

2.3.1. Verifying the GMC Condition. Alexopoulos et al. (2012) propose a graphical-statistical method to check the GMC Condition in the response process generated by a given simulation model. For a trial value of n, we let $\{Y_k : k = 0, 1, \ldots, n\}$ and $\left\{Y'_k : k = 0, 1, \ldots, n\right\}$ denote the response sequences generated by two simulation runs, where those runs are paired as specified by Equation (2) so that (i) the runs are initialized independently in steady-state operation at time 0, perhaps using preliminary warm-up periods for each run that are respectively based on the independent streams of random inputs $\{\ldots,\varepsilon_{-1},\varepsilon_0\}$ and $\{\ldots,\varepsilon'_{-1},\varepsilon'_0\}$; and (ii) subsequently the runs share the common random inputs $\{\varepsilon_1,\ldots,\varepsilon_k\}$ from time 1 to the current time k for $k=1,\ldots,n$. Then for a trial value of the exponent ψ, we compute an estimate $\widehat{\mathrm{E}}\big[|Y_k-Y'_k|^\psi\big]$ of the ψth moment $\mathrm{E}\big[|Y_k-Y'_k|^\psi\big]$ in Equation (2) for $k=0,1,\ldots,n$ by performing, say, 1,000 independent replications of the paired runs and averaging the exponentiated absolute difference $|Y_k-Y'_k|^\psi$ over all 1,000 independent replications to obtain $\widehat{\mathrm{E}}\big[|Y_k-Y'_k|^\psi\big]$ for $k=0,1,\ldots,n$. If a plot of k versus $\log\{\widehat{\mathrm{E}}\big[|Y_k-Y'_k|^\psi\big]\}$ for $k=0,1,\ldots,n$ displays an approximately linear relationship that is supported by the associated linear regression analysis, then we have substantial evidence that the GMC Condition is satisfied for the current values of ψ and n; see Alexopoulos et al. (2012, Example 1). In our experience, trial-and-error experimentation with different values of ψ and n quickly yields insights into the potential validity of the GMC assumption in the given response process.

Moreover as Wu (2005, p. 1941) and Alexopoulos et al. (2019b, p. 1165, last para.) discuss, a broad diversity of time-series models and Markov-chain models have been proved to satisfy the GMC Condition. Recently Dingeç et al. (2022, h)ave proved that in the steady-state G/G/1 queueing system with non-heavy-tailed service-time distributions, successive customer waiting times in the queue satisfy the GMC Condition. Thus some analytical tools are also available for verifying the GMC Condition in time-series and queueing models with appropriate structure. This is a subject of ongoing research.

Remark 2. Results comparable to those in this article can be derived using Theorem 21.1 of Billingsley (1968) if (a) the GMC Condition on $\{Y_k : k\ge 0\}$ is replaced by a ϕ-mixing condition (Bradley 2005, Equations (1.2) and (2.2)) such that the mixing coefficients $\{\phi(\ell):\ell\ge 1\}$ satisfy the summability condition

$$\sum_{\ell=1}^{\infty}\phi^{1/2}(\ell)<\infty; \tag{11}$$

(b) certain density-regularity conditions and other technical conditions are satisfied; and (c) the remainder in the Bahadur representation for $\widetilde{y}_p(n)$ tends to zero a.s. at a sufficiently fast rate as $n\to\infty$. Specifically in the Big-$O_{\text{a.s.}}$ expression for the remainder, the denominator must have the form $n^{(1/2)+\delta}$ for some $\delta>0$, while the numerator is a product of powers of $\log n$ and $\log\log n$.

For example, Yang et al. (2019) derive a Bahadur representation for $\widetilde{y}_p(n)$ with the remainder $O_{\text{a.s.}}\big[(\log n)/n^{3/4}\big]$ under the assumptions that (i) the response process $\{Y_k : k\ge 1\}$ is ϕ-mixing with mixing coefficients satisfying $\phi(\ell)=O(\ell^{-3})$ for $\ell\ge 1$; (ii) $f(y)$ is positive, continuous, and bounded in a neighborhood of y_p; (iii) $f'(y)$ is bounded in a neighborhood of y_p; and (iv) $\Pr(Y_j=Y_k)=0$ for $j\ne k$. Subsubsection 2.3.2 below describes methods for checking assumptions (ii) and (iii). Although verifying assumption (iv) seems problematic, perhaps the main difficulty in applying the result of Yang et al. (2019) is that to the best of our knowledge, *there are no analytical or empirical tools for verifying that the response process is ϕ-mixing, much less that the mixing coefficients satisfy a Big-O condition like $\phi(\ell)=O(\ell^{-3})$ for $\ell\ge 1$ so as to ensure the summability condition (11)—see Whitt (2002, para. after Theorem 4.4.1).* ◄

There is strong evidence to show that the GMC Condition is a more tractable approach to the analysis of STS-based quantile estimation methods compared with ϕ-mixing conditions. The same conclusion holds for α- or ρ-mixing conditions (Bradley 2005, Equations (1.1), (1.4), and (2.2); Whitt 2002, Theorem 4.4.1.).

2.3.2. Verifying the DR Condition. A graphical approach to verifying the DR Condition can be based on visual inspection of a histogram that has been calibrated for robust, approximately optimal estimation of the p.d.f. based on a preliminary sequence of responses $\{Y_k : k = 1, \ldots n\}$, where, say, $n \geq 1{,}000$ (Scott 1992, p. 64). The histogram's cell width (class interval) is taken to be $h = 2\big[\widetilde{y}_{0.75}(n) - \widetilde{y}_{0.25}(n)\big]n^{1/3}$; and the number of cells (class intervals) in the histogram is $\big\lceil (Y_{(n)} - Y_{(1)})/h \big\rceil$ as described by Freedman and Diaconis (1981, Equation (1.8)) and Scott (1992, §§3.2–3.3). Verifying the DR Condition requires good evidence that the p.d.f. is bounded on $\mathbb{R}$, continuous a.e. on $\big[Y_{(1)}, Y_{(n)}\big]$, and positive and smooth (differentiable) in the immediate vicinity of $\widetilde{y}_p(n)$. In our experience, definitive verification of the DR Condition often requires taking $n \geq 20{,}000$ if $p \in [0.05, 0.95]$ and $n \geq 100{,}000$ otherwise.

2.3.3. Verifying the SRD Condition. Based on a preliminary sequence of responses $\{Y_k : k = 1, \ldots, n\}$ and the associated quantile estimator $\widetilde{y}_p(n)$, an empirical approach to verifying the SRD Condition can be based on visual inspection of the sample autocorrelation function (ACF) $\big\{\widehat{\rho}_{I(\widetilde{y}_p)}(\ell) : \ell = 1, \ldots, L\big\}$ computed from the estimated indicator process $\big\{I_k(\widetilde{y}_p(n)) : k = 1, \ldots, n\big\}$ for trial values of n and L. **[Suggest formally defining $\widehat{\rho}_{I(\widetilde{y}_p)}(\ell)$.]** If the sample ACF appears to decline geometrically or exponentially fast with increasing lag ℓ while $1 + 2\sum_{\ell=1}^{L}\big|\widehat{\rho}_{I(\widetilde{y}_p)}(\ell)\big|$ appears to converge to a finite, nonzero value for increasing trial values of n and L, then these observations provide good empirical evidence for the validity of the SRD Condition in the given response process; see Beran (1994, §1.1), Box et al. (2008, §§6.2.1–6.2.3), and Priestley (1981, §5.3.4).

2.3.4. Verifying the FCLT Condition. For the indicator process $\{I_k(y_p) : k \geq 1\}$, the FCLT assumption can be checked empirically by performing independent replications of the estimated indicator process $\big\{I_k(\widetilde{y}_p(n)) : k = 1, \ldots, n\big\}$ and the corresponding element of D,

$$\mathscr{H}_n(t) \equiv \frac{\lfloor nt \rfloor}{n^{1/2}}\big[\bar{I}_{\lfloor nt \rfloor}(\widetilde{y}_p(n)) - p\big] \text{ for } t \in [0,1], \tag{12}$$

in order to inspect displays of selected nonoverlapping increments of the resulting approximate realizations of $\sigma_{I(\widetilde{y}_p(n))}\mathscr{I}_n(\cdot)$ using, for example, P–P and Q–Q plots for the normal distribution; see Law (2015, pp. 339–344) and Hahn and Shapiro (1967, pp. 261–282). Such visual inspection should be accompanied by statistical tests for independence and normality of those increments using, for example, the von Neumann randomness test (Fishman 1978, §2.10) and the Shapiro-Wilk normality test (Hahn and Shapiro 1967, pp. 295–298) to assess conformance to the basic properties of scaled Brownian motion $\sigma_{I(\widetilde{y}_p(n))}\mathscr{W}(\cdot)$ as n increases (Karlin and Taylor 1975, Definition 2.1, p. 343).

Remark 3. In practical applications, it is usually reasonable to assume that the FCLT Condition holds if we have $0 < \sigma^2_{I(\widetilde{y}_p(n))} < \infty$ as required by Equation (4) of the SRD Condition (Whitt 2002, p. 107, last para.). Thus another approach to verifying the FCLT Condition might be based on efficient estimation of $\sigma^2_{I(\widetilde{y}_p(n))}$ from a single prolonged realization of the estimated indicator process $\{I_k(\widetilde{y}_p(n)) : k = 1, \ldots, n\}$ using, for example, one of the STS variance-estimation techniques detailed in Alexopoulos et al. (2007, §§3–4). ◄

3. Basic Asymptotic Properties of the STS-Based Quantile-Estimation Process

In this section we build on Theorem 1 and Corollary 1 to establish the asymptotic behavior of the full-sample STS-based quantile-estimation process in D. The analysis starts by examining the remainder terms $\{Q_u : u \geq 0\}$ associated with the Bahadur representation for $\widetilde{y}_p(u)$ as defined by Equations (7) and (8). If $n \geq \mathfrak{N}$ and $t \in [n^{-1}, 1]$ so that $\lfloor nt \rfloor \geq 1$, then by Equation (8) we have

$$\left|\frac{\lfloor nt \rfloor}{n^{1/2}} Q_{\lfloor nt \rfloor}\right| \leq \frac{\lfloor nt \rfloor}{n^{1/2}} \left\{\mathfrak{U}\frac{(\log\lfloor nt \rfloor)^{3/2}}{\lfloor nt \rfloor^{3/4}}\right\} = \mathfrak{U}\frac{\lfloor nt \rfloor^{1/4}(\log\lfloor nt \rfloor)^{3/2}}{n^{1/2}} \leq \mathfrak{U}\frac{(\log n)^{3/2}}{n^{1/4}} \text{ for } n \geq \mathfrak{N} \text{ and } t \in [n^{-1}, 1] \text{ a.s.}; \quad (13)$$

and if we have $n \geq \mathfrak{N}$ and $t \in [0, n^{-1})$ so that $\lfloor nt \rfloor = 0$, then

$$\left|\frac{\lfloor nt \rfloor}{n^{1/2}} Q_{\lfloor nt \rfloor}\right| = 0 \leq \mathfrak{U}\frac{(\log n)^{3/2}}{n^{1/4}} \text{ for } n \geq \mathfrak{N} \text{ and } t \in [0, n^{-1}) \text{ a.s.} \quad (14)$$

Combining Equations (13) and (14), we have

$$\left|\frac{\lfloor nt \rfloor}{n^{1/2}} Q_{\lfloor nt \rfloor}\right| \leq \mathfrak{U}\frac{(\log n)^{3/2}}{n^{1/4}} \text{ for } n \geq \mathfrak{N} \text{ and } t \in [0, 1] \text{ a.s.} \quad (15)$$

The full-sample STS-based quantile-estimation process is

$$T_n(t) \equiv \frac{\lfloor nt \rfloor}{n^{1/2}} \left[\widetilde{y}_p(n) - \widetilde{y}_p(\lfloor nt \rfloor)\right] \text{ for } n \geq 1 \text{ and } t \in [0, 1]. \quad (16)$$

We let $\mathscr{B}(t) \equiv \mathscr{W}(t) - t\mathscr{W}(1)$ for $t \in [0, 1]$ denote the standard Brownian bridge, which is independent of $\mathscr{W}(1)$ (Resnick 2002, Lemma 6.10.4). We define $\mathfrak{B} : D \mapsto \mathbb{R} \times D$ so that for $\zeta \in D$, we take

$$\mathfrak{B}\{\zeta\} \equiv [\mathfrak{B}_1\{\zeta\}, \mathfrak{B}_2\{\zeta\}], \text{ where } \mathfrak{B}_1\{\zeta\} \equiv \zeta(1) \text{ and } \mathfrak{B}_2\{\zeta\}(t) \equiv \zeta(t) - t\zeta(1) \text{ for } t \in [0, 1]. \quad (17)$$

On the product space $\mathbb{R} \times D$, we define the metric τ as follows:

$$\tau\left[(a_1, \zeta_1), (a_2, \zeta_2)\right] \equiv \max\{|a_1 - a_2|, d(\zeta_1, \zeta_2)\}, \text{ where } (a_i, \zeta_i) \in \mathbb{R} \times D \text{ for } i = 1, 2 \quad (18)$$

(Billingsley 1999, §M10, Equation (9)). Since the rational numbers form a countable dense subset of the real numbers (Kolmogorov and Fomin 1975, p. 48, Example 1), we see that $\mathbb{R}$ with the standard Euclidean metric is separable; and since D with the Skorohod J_1 metric d is separable, the product space $\mathbb{R} \times D$ with the metric τ is also separable (Whitt 2002, Theorem 11.4.1).

Theorem 2. *If $\{Y_k : k \geq 1\}$ satisfies the assumptions of Theorem 1, then in $\mathbb{R} \times D$ with the metric τ,*

$$\left[n^{1/2}\big(y_p - \widetilde{y}_p(n)\big), T_n\right] \underset{n \to \infty}{\Longrightarrow} \sigma_{\widetilde{y}_p}[\mathscr{W}(1), \mathscr{B}]. \quad (19)$$

Proof First we show that the functional $\mathfrak{B}\{\cdot\}$ is continuous at every $\zeta \in D$. Pick $\zeta \in D$ and $\varepsilon > 0$ arbitrarily. If $\omega \in D$ and $d(\zeta,\omega) < \varepsilon/(3+|\zeta(1)|)$, then by the definition of $d(\zeta,\omega)$ there exists $\lambda \in \Lambda$ such that

$$|\lambda(t)-t| < \varepsilon/(3+|\zeta(1)|) \text{ and } |\zeta(t)-\omega\circ\lambda(t)| < \varepsilon/(3+|\zeta(1)|) \text{ for each } t \in [0,1]. \tag{20}$$

Because $\lambda(1) = 1$, the second inequality in Equation (20) ensures that $|\zeta(1)-\omega(1)| < \varepsilon/(3+|\zeta(1)|)$; moreover, Equations (17) and (18) ensure that

$$\begin{aligned}\tau[\mathfrak{B}\{\zeta\},\mathfrak{B}\{\omega\}] &= \max\big[\big|\mathfrak{B}_1\{\zeta\}-\mathfrak{B}_1\{\omega\}\big|, d\big(\mathfrak{B}_2\{\zeta\},\mathfrak{B}_2\{\omega\}\big)\big]\\ &= \max\big[\big|\zeta(1)-\omega(1)\big|, d\big(\zeta(t)-t\zeta(1), \omega(t)-t\omega(1)\big)\big],\end{aligned} \tag{21}$$

where the abuse of notation in the expression $d\big(\zeta(t)-t\zeta(1),\omega(t)-t\omega(1)\big)$ should clarify the intermediate calculations given below, which require applying the definition of the proper expression $d\big(\mathfrak{B}_2\{\zeta\},\mathfrak{B}_2\{\omega\}\big)$ followed by applying the triangle inequality for $\|\cdot\|$ twice:

$$\begin{aligned}d\big(\zeta(t)-t\zeta(1),\omega(t)-t\omega(1)\big) &\le \max\big\{\|\lambda(t)-t\|, \|[\zeta(t)-t\zeta(1)]-[\omega\circ\lambda(t)-\lambda(t)\omega(1)]\|\big\}\\ &\le \|\lambda(t)-t\| + \|\zeta(t)-\omega\circ\lambda(t)\| + \|[\lambda(t)\zeta(1)-t\zeta(1)]-[\lambda(t)\zeta(1)-\lambda(t)\omega(1)]\|\\ &\le \frac{2\varepsilon}{3+|\zeta(1)|} + \|\lambda(t)\zeta(1)-t\zeta(1)\| + \|\lambda(t)\zeta(1)-\lambda(t)\omega(1)\|\\ &\le \frac{2\varepsilon}{3+|\zeta(1)|} + \frac{|\zeta(1)|\varepsilon}{3+|\zeta(1)|} + |\zeta(1)-\omega(1)| \le \varepsilon.\end{aligned} \tag{22}$$

Equations (21) and (22) imply that $\tau[\mathfrak{B}\{\zeta\},\mathfrak{B}\{\omega\}] \le \varepsilon$; and since ζ and ε were chosen arbitrarily, we see that $\mathfrak{B}\{\cdot\}$ is continuous on D.

For $n \ge 1$ and $t \in [0,1]$, we define the STS-based indicator process

$$\mathscr{T}_n(t) \equiv \sigma_{\widetilde{y}_p}\mathscr{I}_n(t) - t\sigma_{\widetilde{y}_p}\mathscr{I}_n(1) = \frac{\lfloor nt\rfloor}{n^{1/2}}\left[\frac{\bar{I}_{\lfloor nt\rfloor}(y_p)-p}{f(y_p)}\right] - t\left\{n^{1/2}\left[\frac{\bar{I}_n(y_p)-p}{f(y_p)}\right]\right\} \tag{23}$$

in D. Because $\mathfrak{B}\{\cdot\}$ is continuous on D while both D and $\mathbb{R}\times D$ are separable metric spaces, the FCLT (**??**) and the continuous-mapping theorem (Whitt 2002, p. 85, last para. to Theorem 3.4.3) imply that

$$\mathfrak{B}\{\sigma_{\widetilde{y}_p}\mathscr{I}_n\} = [\sigma_{\widetilde{y}_p}\mathscr{I}_n(1), \mathscr{T}_n] \underset{n\to\infty}{\Longrightarrow} \mathfrak{B}\{\sigma_{\widetilde{y}_p}\mathscr{W}\} = \sigma_{\widetilde{y}_p}[\mathscr{W}(1), \mathscr{B}]. \tag{24}$$

To prove Equation (19), we build on Equation (24) by applying the convergence-together theorem (Billingsley 1999, Theorem 3.1; Whitt 2002, Theorem 11.4.7) to two selected random elements $\mathfrak{s}_1(n)$, $\mathfrak{s}_2(n) \in \mathbb{R}\times D$ as $m \longrightarrow \infty$. We let

$$\left.\begin{aligned}\mathfrak{s}_1(n) &\equiv \big[n^{1/2}\big(\bar{I}_n(y_p)-p\big)/f(y_p), \mathscr{T}_n\big)\big] = \mathfrak{B}\{\sigma_{\widetilde{y}_p}\mathscr{I}_n\} \text{ and}\\ \mathfrak{s}_2(n) &\equiv \big[n^{1/2}\big(y_p-\widetilde{y}_p(n)\big), T_n\big]\end{aligned}\right\}\text{ so that} \tag{25}$$

$$\tau[\mathfrak{s}_1(n),\mathfrak{s}_2(n)] = \max\left\{\left|n^{1/2}\left[\frac{\bar{I}_n(y_p)-p}{f(y_p)}\right] - n^{1/2}\big[y_p-\widetilde{y}_p(n)\big]\right|, d(\mathscr{T}_n, T_n)\right\}; \tag{26}$$

and in terms of the Bahadur remainder Q_n given by Equations (7) and (8), for $n \ge \mathfrak{N}$ a.s. we have

$$\left|n^{1/2}\left[\frac{\bar{I}_n(y_p)-p}{f(y_p)}\right] - n^{1/2}\big[y_p-\widetilde{y}_p(n)\big]\right| = \big|n^{1/2}Q_n\big| \le \mathfrak{U}\frac{(\log n)^{3/2}}{n^{1/4}} \underset{n\to\infty}{\Longrightarrow} 0 \tag{27}$$

by Slutsky's Theorem (Bickel and Doksum 2007, Theorem A.14.9). Equations (7) and (8) also ensure that

$$
\begin{aligned}
\|\mathscr{T}_n - T_n\| &= \sup_{t\in[0,1]}\left|\left\{\frac{\lfloor nt\rfloor}{n^{1/2}}\left[\frac{\overline{I}_{\lfloor nt\rfloor}(y_p)-p}{f(y_p)}\right]-tn^{1/2}\left[\frac{\overline{I}_n(y_p)-p}{f(y_p)}\right]\right\}-\frac{\lfloor nt\rfloor}{n^{1/2}}\left\{\left[\widetilde{y}_p(n)-y_p\right]-\left[\widetilde{y}_p(\lfloor nt\rfloor)-y_p\right]\right\}\right| \\
&= \sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor}{n^{1/2}}\left[\frac{\overline{I}_{\lfloor nt\rfloor}(y_p)-p}{f(y_p)}+\widetilde{y}_p(\lfloor nt\rfloor)-y_p\right]\right. \\
&\qquad\left.-\frac{\lfloor nt\rfloor}{n^{1/2}}\left[\frac{\overline{I}_n(y_p)-p}{f(y_p)}+\widetilde{y}_p(n)-y_p\right]-\left(\frac{nt-\lfloor nt\rfloor}{n}\right)n^{1/2}\left[\frac{\overline{I}_n(y_p)-p}{f(y_p)}\right]\right| \\
&= \sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor}{n^{1/2}}(Q_{\lfloor nt\rfloor}-Q_n)-\left(\frac{nt-\lfloor nt\rfloor}{n}\right)n^{1/2}\left[\overline{I}_n(y_p)-p\right]/f(y_p)\right| \\
&\le \sup_{t\in[0,1]}\left\{\frac{\lfloor nt\rfloor}{n^{1/2}}\left(\left|Q_{\lfloor nt\rfloor}\right|+\left|Q_n\right|\right)+n^{-1}\left|n^{1/2}\left[\overline{I}_n(y_p)-p\right]/f(y_p)\right|\right\} \\
&\le 2\mathfrak{U}\frac{(\log n)^{3/2}}{n^{1/4}}+n^{-1}\left|\sigma_{\widetilde{y}_p}\mathscr{I}_n(1;y_p)\right| \text{ for each } n\ge\mathfrak{N} \text{ a.s.}
\end{aligned}
\tag{28}
$$

Equation (24) shows that $\sigma_{\widetilde{y}_p}\mathscr{I}_n(1;y_p) \underset{n\to\infty}{\Longrightarrow} \sigma_{\widetilde{y}_p}\mathscr{W}(1)$; hence by Slutsky's theorem, we have

$$
n^{-1}\left|\sigma_{\widetilde{y}_p}\mathscr{I}_n(1;y_p)\right| \underset{n\to\infty}{\Longrightarrow} 0. \tag{29}
$$

Equations (28), and (29) together with Slutsky's Theorem ensure that $\|\mathscr{T}_n - T_n\| \underset{n\to\infty}{\Longrightarrow} 0$, from which it follows immediately that

$$
d(\mathscr{T}_n, T_n) \underset{n\to\infty}{\Longrightarrow} 0. \tag{30}
$$

Equations (26), (27), and (30) yield

$$
\tau[\mathfrak{s}_1(n), \mathfrak{s}_2(n)] \underset{n\to\infty}{\Longrightarrow} 0 \tag{31}
$$

so that Equation (19) follows from Equations (24) and (31) and the convergence-together theorem. □

4. Weak Convergence of the Full-Sample STS Area Estimator

Let $w \in D$ denote a given deterministic weighting function. Thus w is bounded on $[0,1]$ and continuous a.e. on $[0,1]$; moreover, for each deterministic or probabilistic element $\zeta \in D$, the function $w(t)\zeta(t)$, $t\in[0,1]$, has the same properties and hence is Riemann integrable over $[0,1]$ (Rudin 1976, Theorem 11.33(b)). The full-sample STS area estimator of the variance parameter $\sigma^2_{\widetilde{y}_p}$ is $\mathscr{A}_n^{\,2}(w)$, where

$$
\mathscr{A}_n(w) \equiv n^{-1}\sum_{k=1}^{n} w(k/n)T_n(k/n) \text{ for } n \ge 1. \tag{32}
$$

In terms of the related r.v.

$$
\mathbb{A}(w) \equiv \int_0^1 w(t)\mathscr{B}(t)\,\mathrm{d}t,
$$

we select w to yield $\mathrm{E}[\mathbb{A}^2(w)] = 1$ so that $\mathbb{A}(w) \stackrel{\mathrm{d}}{=} Z$; see Alexopoulos et al. (2007, Equation (4)). Let

$$
M_w \equiv \|w\| \in (0,\infty). \tag{33}
$$

To establish the asymptotic distribution of $\mathscr{A}_n^2(w)$ as $n \to \infty$, we let

$$\Delta_n(\zeta) \equiv n^{-1} \sum_{k=1}^{n} w(k/n)\zeta(k/n) \text{ and } \Delta(\zeta) \equiv \int_0^1 w(t)\zeta(t)\mathrm{d}t \text{ for } \zeta \in D.$$

A key property in the analysis of the STS area estimator is that the relevant weak-convergence limits $\mathscr{W}$ and $\mathscr{B}$ in D also belong to C, the subspace of continuous functions on $[0,1]$, because in the underlying probability space, *every* realization of the random processes $\{\mathscr{W}(t) : t \in [0,1]\}$ and $\{\mathscr{B}(t) = \mathscr{W}(t) - t\mathscr{W}(1) : t \in [0,1]\}$ is a continuous function on $[0,1]$ (Billingsley 1968, p. 64; Resnick 2002, §6.4, Property 5).

Theorem 3. *If $\{Y_k : k \geq 1\}$ satisfies the assumptions of Theorem 1, then*

$$\mathscr{A}_n^2(w) \underset{n\to\infty}{\Longrightarrow} \sigma^2_{\tilde{y}_p}\chi^2_1.$$

Proof Pick a deterministic element $\eta \in C$ arbitrarily. Let $\{\eta_n : n \geq 1\}$ denote an arbitrary deterministic sequence in D for which $d(\eta_n, \eta)$ converges to zero as $n \to \infty$. Let

$$\delta_n \equiv d(\eta_n, \eta) + n^{-1} \text{ for } n \geq 1 \text{ so that } \lim_{n\to\infty} \delta_n = 0. \tag{34}$$

By the triangle inequality, we see that

$$|\Delta_n(\eta_n) - \Delta(\eta)| \leq |\Delta_n(\eta_n) - \Delta_n(\eta)| + |\Delta_n(\eta) - \Delta(\eta)| \text{ for each } n \geq 1. \tag{35}$$

Since $w(\cdot)\eta(\cdot)$ is Riemann integrable on $[0,1]$, we have

$$\lim_{n\to\infty} \Delta_n(\eta) = \Delta(\eta). \tag{36}$$

Equation (34) ensures that $d(\eta_n, \eta) < \delta_n$ for every $n \geq 1$; hence by the definition of the Skorohod J_1 metric on D, for every $n \geq 1$ there exists $\lambda_n \in \Lambda$ such that

$$\|\lambda_n - \mathbb{I}\| = \sup_{t\in[0,1]} |\lambda_n(t) - t| < \delta_n \quad \text{and} \quad \|\eta_n - \eta\circ\lambda_n\| = \sup_{t\in[0,1]} |\eta_n(t) - \eta\circ\lambda_n(t)| < \delta_n. \tag{37}$$

By the triangle inequality and Equation (33), we also have

$$\begin{aligned} |\Delta_n(\eta_n) - \Delta_n(\eta)| &\leq n^{-1} \sum_{k=1}^{n} |w(k/n)[\eta_n(k/n) - \eta(k/n)]| \\ &\leq M_w \left\{ n^{-1} \sum_{k=1}^{n} |\eta_n(k/n) - \eta\circ\lambda_n(k/n)| + n^{-1} \sum_{k=1}^{n} |\eta(k/n) - \eta\circ\lambda_n(k/n)| \right\}. \end{aligned} \tag{38}$$

Since both λ_n and η are continuous on $[0,1]$, those functions are uniformly continuous on $[0,1]$ (Rudin 1976, Theorem 4.19). Moreover since the first part of Equation (37) ensures that λ_n converges uniformly to $\mathbb{I}$ on $[0,1]$, we see that $\eta - \eta\circ\lambda_n$ converges uniformly to zero on $[0,1]$ so that we have

$$\lim_{n\to\infty} \|\eta - \eta\circ\lambda_n\| = 0 \tag{39}$$

(Rudin 1976, Theorem 7.9); and by the second part of Equation (37) and Equations (38)–(39), we have

$$\lim_{n\to\infty} |\Delta_n(\eta_n) - \Delta_n(\eta)| \leq \lim_{n\to\infty} M_w(\delta_n + \|\eta - \eta\circ\lambda_n\|) = 0. \tag{40}$$

Equations (35), (36), and (40) ensure that

$$\lim_{n\to\infty} \Delta_n^2(\eta_n) = \Delta^2(\eta)\,. \tag{41}$$

To complete the proof of Theorem 3, we verify the assumptions of the generalized continuous-mapping theorem (Whitt 2002, Theorem 3.4.4) before applying that theorem. In the following definition, each element of D is deterministic:

$$\mathfrak{F} \equiv \Big\{ \gamma \in D : \text{There is } \{\gamma_n : n \ge 1\} \subset D \text{ with } \lim_{n\to\infty} d(\gamma_n, \gamma) = 0, \text{ but } \Delta_n^2(\gamma_n) \nrightarrow \Delta^2(\gamma) \Big\}. \tag{42}$$

Hence $\mathfrak{F}$ is the set consisting of every deterministic element γ in D to which some deterministic sequence of elements $\{\gamma_n : n \ge 1\}$ in D converges with respect to d, but the associated real sequence $\big\{\Delta_n^2(\gamma_n) : n \ge 1\big\}$ does not converge to $\Delta^2(\gamma)$. Since in Equation (41) the deterministic elements $\eta \in C$ and $\{\eta_n : n \ge 1\} \subset D$ were chosen arbitrarily, we have $C \cap \mathfrak{F} = \varnothing$; and since the random element $\mathscr{B}$ belongs to C always, no realization of $\mathscr{B}$ belongs to $\mathfrak{F}$ so that in the underlying probability space,

$$\Pr\{\mathscr{B} \in \mathfrak{F}\} = \Pr(\varnothing) = 0. \tag{43}$$

Theorem 2 ensures that $T_n \underset{n\to\infty}{\Longrightarrow} \sigma_{\widetilde{y}_p}\mathscr{B}$; therefore Equations (42)–(43) and the generalized continuous-mapping theorem ensure that

$$\mathscr{A}_n^2(w) = \Delta_n^2(T_n) \underset{n\to\infty}{\Longrightarrow} \Delta^2(\sigma_{\widetilde{y}_p}\mathscr{B}) = \sigma^2_{\widetilde{y}_p}\mathbb{A}^2(w) \stackrel{\text{d}}{=} \sigma^2_{\widetilde{y}_p}\chi_1^2. \quad \square \tag{44}$$

5. Asymptotically Exact CIs Using Batched-STS Area Estimators

First we sketch the proofs of Theorems 4–6 so as to motivate the asymptotically exact CIs developed in this article. For $j = 1, \ldots, b$, we will define the *batched-STS processes* $\mathscr{I}_{j,m}$, $\mathscr{T}_{j,m}$, and $T_{j,m}$, which are analogous to the full-sample processes $\mathscr{I}_n$, $\mathscr{T}_n$, and T_n, respectively, but depend solely on the subsequence of random vectors $\{[Y_k, I_k(y_p)] : k = (j-1)m+1, \ldots, jm\}$ occurring in the jth batch. We will use the batched-STS processes and the Bahadur representations for the related BQEs to derive the asymptotic joint probability distribution of the r.v.'s $\big\{m^{1/2}\big[\widehat{y}_p(j,m) - y_p\big] : j = 1, \ldots, b\big\}$ and the batched-STS quantile-estimation processes $\{T_{j,m} : j = 1, \ldots, b\}$ as $m \to \infty$. Thus for $j = 1, \ldots, b$, $m \ge 1$, and $t \in [0,1]$, we let

$$\left.\begin{aligned}
\mathscr{I}_{j,m}(t) &\equiv \frac{\lfloor mt \rfloor}{\sigma_{I(y_p)} m^{1/2}}\big(\bar{I}_{j,\lfloor mt \rfloor}(y_p) - p\big), \\
\mathscr{T}_{j,m}(t) &\equiv \sigma_{\widetilde{y}_p}\big[\mathscr{I}_{j,m}(t) - t\,\mathscr{I}_{j,m}(1)\big], \text{ and} \\
T_{j,m}(t) &\equiv \frac{\lfloor mt \rfloor}{m^{1/2}}\big[\widehat{y}_p(j,m) - \widehat{y}_p(j, \lfloor mt \rfloor)\big].
\end{aligned}\right\} \tag{45}$$

Starting from Equation (45), we will adapt the proofs of Theorems 2 and 3 to handle batches of size m, thereby establishing in the proof of Theorem 4 that in the appropriate metric spaces, we have

$$\left.\begin{aligned}
\Big\{\Big[\frac{m^{1/2}}{f(y_p)}\big(\bar{I}_{j,m}(y_p) - p\big), \mathscr{T}_{j,m}\Big] : j = 1, \ldots, b\Big\} &\underset{m\to\infty}{\overset{\text{i.i.d.}}{\Longrightarrow}} \sigma_{\widetilde{y}_p}\big[\mathscr{W}(1), \mathscr{B}\big], \\
\big\{\big[m^{1/2}\big(\widehat{y}_p(j,m) - y_p\big), T_{j,m}\big] : j = 1, \ldots, b\big\} &\underset{m\to\infty}{\overset{\text{i.i.d.}}{\Longrightarrow}} \sigma_{\widetilde{y}_p}\big[\mathscr{W}(1), \mathscr{B}\big], \text{ and} \\
\big\{\big[m^{1/2}\big(\widehat{y}_p(j,m) - y_p\big), \Delta_m(T_{j,m})\big] : j = 1, \ldots, b\big\} &\underset{m\to\infty}{\overset{\text{i.i.d.}}{\Longrightarrow}} \sigma_{\widetilde{y}_p}\big[\mathscr{W}(1), \Delta(\mathscr{B})\big],
\end{aligned}\right\} \tag{46}$$

where in general the shorthand notation $\{\mathscr{U}_{j,m} : j = 1, \ldots, b\} \underset{m\to\infty}{\overset{\text{i.i.d.}}{\Longrightarrow}} \mathscr{S}$ means that (i) $\mathscr{U}_{j,m} \underset{m\to\infty}{\Longrightarrow} \mathscr{S}$ for $j = 1, \ldots, b$; and (ii) the $\{\mathscr{U}_{j,m} : j = 1, \ldots, b\}$ are asymptotically i.i.d. as $m \to \infty$. In the proof of Theorem 4, a more-detailed specification of (i) and (ii) will be given in the appropriate product-form metric spaces.

In Theorem 4 the first asymptotically exact $100(1-\alpha)\%$ CI for y_p has the form

$$\overline{\overline{y}}_p(b,m) \pm t_{1-\alpha/2,b}\left(\mathscr{A}^2_{b,m}/n\right)^{1/2} \tag{47}$$

based on the sample means of the BQEs and the batched-STS area estimators, which are given by

$$\overline{\overline{y}}_p(b,m) \equiv b^{-1}\sum_{j=1}^{b}\widehat{y}_p(j,m) \ \text{ and } \ \mathscr{A}^2_{b,m} \equiv b^{-1}\sum_{j=1}^{b}\Delta^2_m(T_{j,m}), \tag{48}$$

respectively. In the proof of Theorem 4, the third relation in Equation (46) will be used to show that

$$\left[n^{1/2}\left(\overline{\overline{y}}_p(b,m) - y_p\right), \mathscr{A}_{b,m}\right] \underset{m\to\infty}{\Longrightarrow} \sigma_{\widetilde{y}_p}\left[Z, \left(\chi^2_b/b\right)^{1/2}\right], \text{ where } Z \text{ and } \left(\chi^2_b/b\right)^{1/2} \text{ are independent.} \tag{49}$$

Equation (49) will be shown to ensure that

$$n^{1/2}\left[\overline{\overline{y}}_p(b,m) - y_p\right]/\mathscr{A}_{b,m} \underset{m\to\infty}{\Longrightarrow} t_b, \tag{50}$$

and hence the CI in Equation (47) has asymptotic coverage probability $1-\alpha$ as $m \to \infty$.

The preceding discussion naturally leads to another asymptotically exact CI for y_p in which the estimator of $\sigma^2_{\widetilde{y}_p}$ is a suitable linear combination of $\mathscr{A}^2_{b,m}$ and the sample variance of the BQEs,

$$S^2_{b,m} \equiv (b-1)^{-1}\sum_{j=1}^{b}\left[\widehat{y}_p(j,m) - \overline{\overline{y}}_p(b,m)\right]^2. \tag{51}$$

In the proof of Theorem 4 we will show that

$$n^{1/2}\left[\overline{\overline{y}}_p(b,m) - y_p\right] \underset{m\to\infty}{\Longrightarrow} \sigma_{\widetilde{y}_p} Z; \tag{52}$$

$$\mathscr{A}^2_{b,m} \underset{m\to\infty}{\Longrightarrow} \sigma^2_{\widetilde{y}_p}\chi^2_b/b; \ \ mS^2_{b,m} \underset{m\to\infty}{\Longrightarrow} \sigma^2_{\widetilde{y}_p}\chi^2_{b-1}/(b-1); \ \text{ and} \tag{53}$$

$$\sigma_{\widetilde{y}_p}Z, \ \ \sigma^2_{\widetilde{y}_p}\chi^2_b/b, \ \text{ and } \ \sigma^2_{\widetilde{y}_p}\chi^2_{b-1}/(b-1) \ \text{ are independent.} \tag{54}$$

Based on Equations (52)–(54), we will show that $\mathscr{V}_{b,m}$, the combined estimator of $\sigma^2_{\widetilde{y}_p}$, satisfies

$$\mathscr{V}_{b,m} \equiv \frac{b\mathscr{A}^2_{b,m} + (b-1)mS^2_{b,m}}{2b-1} \underset{m\to\infty}{\Longrightarrow} \sigma^2_{\widetilde{y}_p}\chi^2_{2b-1}/(2b-1), \ \text{ where} \tag{55}$$

$$\sigma_{\widetilde{y}_p}Z \ \text{ and } \ \sigma^2_{\widetilde{y}_p}\chi^2_{2b-1}/(2b-1) \ \text{ are independent.} \tag{56}$$

It follows from Equations (55) and (56) that as $m \to \infty$, an asymptotically exact $100(1-\alpha)\%$ CI for y_p is

$$\overline{\overline{y}}_p(b,m) \pm t_{1-\alpha/2,2b-1}\left(\mathscr{V}_{b,m}/n\right)^{1/2}. \tag{57}$$

Building on Theorem 4 and seeking asymptotically exact CIs with better small-sample performance, in Theorem 5 we will show the asymptotic exactness of the CIs (92) and (93) below, which are centered on $\widetilde{y}_p(n)$ but have the same half-lengths as in Equations (47) and (57), respectively. Finally building on Theorem 5, in Theorem 6 we will show the asymptotic exactness of the CIs (103) and (104) below, which are centered on $\widetilde{y}_p(n)$ but their half-lengths depend respectively on (i) m times the mean squared deviation of the BQEs away from $\widetilde{y}_p(n)$; or (ii) a linear combination of (i) and $\mathscr{A}^2_{b,m}$.

Theorem 4. *If $\{Y_k : k \geq 1\}$ satisfies the assumptions of Theorem 1, then*

$$\left.\begin{array}{l}\textit{Equation (46) holds, and Equations (47) and (57) are asymptotically}\\ \textit{exact } 100(1-\alpha)\% \textit{ CIs for } y_p \textit{ as } m\to\infty.\end{array}\right\} \tag{58}$$

Proof For the sequences of random elements in D defined in Equation (45), we seek an easily verified sufficient condition ensuring the validity of the associated weak-convergence results in Equation (46). Deriving those results requires repeated applications of the generalized continuous-mapping and convergence-together theorems; and to facilitate the analysis, for $1 \leq j \leq b$ and for each probabilistic or deterministic element $\zeta \in D$, we define the functionals $\mathfrak{X}_j\{\zeta\} \in D$, $\mathfrak{A}_j(\zeta) \in \mathbb{R}$, and $\mathfrak{B}_j\{\zeta\} \in D$, respectively, as follows:

$$\left.\begin{aligned}\mathfrak{X}_j\{\zeta\}(t) &\equiv b^{1/2}\left[\zeta\left(\tfrac{j+t-1}{b}\right)-\zeta\left(\tfrac{j-1}{b}\right)\right] \text{ for } t\in[0,1];\\ \mathfrak{A}_j(\zeta) &\equiv \mathfrak{X}_j\{\zeta\}(1); \text{ and}\\ \mathfrak{B}_j\{\zeta\}(t) &\equiv \mathfrak{X}_j\{\zeta\}(t)-t\mathfrak{A}_j(\zeta) \text{ for } t\in[0,1].\end{aligned}\right\} \tag{59}$$

To infer the desired sufficient condition, we take the approach used in the proof of Theorem 3, where without loss of generality and for simplicity we assume that the relevant elements of C and D are deterministic. Pick the deterministic element $\eta \in C$ arbitrarily. Let $\{\eta_n : n \geq 1\} \subset D$ denote an arbitrary deterministic sequence such that $\lim_{n\to\infty} d(\eta_n, \eta) = 0$. We define the sequence $\{\delta_n : n \geq 1\}$ according to Equation (34). Thus for every $n \geq 1$, there exists $\lambda_n \in \Lambda$ such that Equations (37) and (39) hold, and we have

$$\lim_{n\to\infty}\|\eta_n-\eta\circ\lambda_n\| = \lim_{n\to\infty}\|\eta-\eta\circ\lambda_n\| = 0. \tag{60}$$

Next we show that $d(\mathfrak{X}_j\{\eta_n\}, \mathfrak{X}_j\{\eta\}) \to 0$; $\mathfrak{A}_j(\eta_n) \to \mathfrak{A}_j(\eta)$; and $d(\mathfrak{B}_j\{\eta_n\}, \mathfrak{B}_j\{\eta\}) \to 0$ for $j = 1, \ldots, b$ as $n \to \infty$. For each $t \in [0,1]$ and $n = bm$ when $m \geq 1$, by the triangle inequality and the definition of $\|\cdot\|$, we have

$$\begin{aligned}\left|b^{1/2}\left[\eta_n(t)-\eta(t)\right]\right| &\leq \left|b^{1/2}\left[\eta_n(t)-\eta\circ\lambda_n(t)\right]\right|+\left|b^{1/2}\left[\eta(t)-\eta\circ\lambda_n(t)\right]\right|\\ &\leq b^{1/2}\left[\|\eta_n-\eta\circ\lambda_n\|+\|\eta-\eta\circ\lambda_n\|\right],\end{aligned} \tag{61}$$

$$\begin{aligned}|\mathfrak{X}_j\{\eta_n\}(t)-\mathfrak{X}_j\{\eta\}(t)| &= \left|b^{1/2}\left[\eta_n\left(\tfrac{j+t-1}{b}\right)-\eta\left(\tfrac{j+t-1}{b}\right)\right]-b^{1/2}\left[\eta_n\left(\tfrac{j-1}{b}\right)-\eta\left(\tfrac{j-1}{b}\right)\right]\right|\\ &\leq 2b^{1/2}\left[\|\eta_n-\eta\circ\lambda_n\|+\|\eta-\eta\circ\lambda_n\|\right], \text{ and}\end{aligned} \tag{62}$$

$$\begin{aligned}|\mathfrak{B}_j\{\eta_n\}(t)-\mathfrak{B}_j\{\eta\}(t)| &\leq |\mathfrak{X}_j\{\eta_n\}(t)-\mathfrak{X}_j\{\eta\}(t)|+t|\mathfrak{X}_j\{\eta_n\}(1)-\mathfrak{X}_j\{\eta\}(1)|\\ &\leq 4b^{1/2}\left[\|\eta_n-\eta\circ\lambda_n\|+\|\eta-\eta\circ\lambda_n\|\right].\end{aligned} \tag{63}$$

Equations (62)–(63) and the definition of $\|\cdot\|$ imply that

$$\left.\begin{aligned}\|\mathfrak{X}_j\{\eta_n\}-\mathfrak{X}_j\{\eta\}\| &\leq 2b^{1/2}\left[\|\eta_n-\eta\circ\lambda_n\|+\|\eta-\eta\circ\lambda_n\|\right],\\ |\mathfrak{A}_j(\eta_n)-\mathfrak{A}_j(\eta)| &\leq 2b^{1/2}\left[\|\eta_n-\eta\circ\lambda_n\|+\|\eta-\eta\circ\lambda_n\|\right], \text{ and}\\ \|\mathfrak{B}_j\{\eta_n\}-\mathfrak{B}_j\{\eta\}\| &\leq 4b^{1/2}\left[\|\eta_n-\eta\circ\lambda_n\|+\|\eta-\eta\circ\lambda_n\|\right]\end{aligned}\right\} \text{ for } j=1,\ldots b. \tag{64}$$

Results similar to Equation (64) are needed for $|\Delta_m(\mathfrak{B}_j\{\eta_n\}) - \Delta(\mathfrak{B}_j\{\eta\})|$, where $j = 1, \ldots, b$. By the triangle inequality,

$$|\Delta_m(\mathfrak{B}_j\{\eta_n\}) - \Delta(\mathfrak{B}_j\{\eta\})| \le |\Delta_m(\mathfrak{B}_j\{\eta_n\}) - \Delta_m(\mathfrak{B}_j\{\eta\})| + |\Delta_m(\mathfrak{B}_j\{\eta\}) - \Delta(\mathfrak{B}_j\{\eta\})| \tag{65}$$

for $m \ge 1$. Since $w(\cdot)\mathfrak{B}_j\{\eta\}(\cdot)$ is Riemann integrable on $[0,1]$, we have

$$\lim_{m\to\infty} \Delta_m(\mathfrak{B}_j\{\eta\}) = \Delta(\mathfrak{B}_j\{\eta\}). \tag{66}$$

By the triangle inequality, the definition of $\|\cdot\|$, Equation (33), and Equation (63), we also have

$$\begin{aligned}|\Delta_m(\mathfrak{B}_j\{\eta_n\}) - \Delta_m(\mathfrak{B}_j\{\eta\})| &\le m^{-1}\sum_{k=1}^{m} |w(k/m)[\mathfrak{B}_j\{\eta_n\}(k/m) - \mathfrak{B}_j\{\eta\}(k/m)]| \\ &\le 4M_w b^{1/2}\big[\|\eta_n - \eta\circ\lambda_n\| + \|\eta - \eta\circ\lambda_n\|\big].\end{aligned} \tag{67}$$

Equations (60) and (64)–(67) show that

$$\left.\begin{aligned}\lim_{n\to\infty} d(\mathfrak{X}_j\{\eta_n\}, \mathfrak{X}_j\{\eta\}) &= 0, \\ \lim_{n\to\infty} \mathfrak{A}_j(\eta_n) &= \mathfrak{A}_j(\eta), \\ \lim_{n\to\infty} d(\mathfrak{B}_j\{\eta_n\}, \mathfrak{B}_j\{\eta\}) &= 0, \text{ and} \\ \lim_{m\to\infty} \Delta_m(\mathfrak{B}_j\{\eta_n\}) &= \Delta(\mathfrak{B}_j\{\eta\})\end{aligned}\right\} \text{ for } j = 1, \ldots, b.$$

Therefore we have the following sufficient condition for applying the generalized continuous-mapping theorem and the convergence-together theorem to derive the weak-convergence limits of the random sequences $\{\mathfrak{X}_j\{\mathscr{I}_n\} : n \ge 1\}$, $\{\mathfrak{A}_j(\mathscr{I}_n) : n \ge 1\}$, $\{\mathfrak{B}_j\{\mathscr{I}_n\} : n \ge 1\}$, and $\{\Delta_m(T_{j,m}) : m \ge 1\}$ for $j = 1, \ldots, b$:

$$\left.\begin{array}{l}\text{At every } \zeta \in C, \text{ the functionals } \mathfrak{X}_j\{\zeta\}, \mathfrak{A}_j(\zeta), \text{ and } \mathfrak{B}_j\{\zeta\} \text{ are continuous} \\ \text{in } D; \text{ and for every sequence } \{\zeta_n : n \ge 1\} \text{ in } D \text{ converging to } \zeta, \text{ we have} \\ \lim_{m\to\infty} \Delta_m(\mathfrak{B}_j\{\zeta_n\}) = \Delta(\mathfrak{B}_j\{\zeta\}), \text{ where } j = 1, \ldots, b, \text{ and } n = bm \text{ for } m \ge 1.\end{array}\right\} \tag{68}$$

In terms of the random elements $\{\mathscr{I}_n : n = bm \text{ and } m \ge 1\}$ in C defined by Equation (**??**), we can apply Equation (68), the FCLT (**??**), and the generalized continuous-mapping theorem to conclude that

$$\left.\begin{aligned}\mathfrak{X}_j\{\mathscr{I}_n\} &\underset{n\to\infty}{\Longrightarrow} \mathfrak{X}_j\{\mathscr{W}\} \text{ in } D, \\ \mathfrak{A}_j(\mathscr{I}_n) &\underset{n\to\infty}{\Longrightarrow} \mathfrak{A}_j(\mathscr{W}) \text{ in } \mathbb{R}, \\ \mathfrak{B}_j\{\mathscr{I}_n\} &\underset{n\to\infty}{\Longrightarrow} \mathfrak{B}_j\{\mathscr{W}\} \text{ in } D, \text{ and} \\ \Delta_m(\mathfrak{B}_j\{\mathscr{I}_n\}) &\underset{n\to\infty}{\Longrightarrow} \Delta[\mathfrak{B}_j\{\mathscr{W}\}] \text{ in } \mathbb{R}\end{aligned}\right\} \text{ for } j = 1, \ldots, b. \tag{69}$$

Basic properties of $\mathscr{W}$ ensure that

$$\left.\begin{aligned}\{\mathfrak{X}_j\{\mathscr{W}\} : j = 1, \ldots, b\} &\overset{\text{i.i.d.}}{\sim} \mathscr{W} \text{ in } D, \\ \{\mathfrak{A}_j(\mathscr{W}) : j = 1, \ldots, b\} &\overset{\text{i.i.d.}}{\sim} \mathscr{W}(1) \overset{\text{d}}{=} Z \text{ in } \mathbb{R}, \\ \{\mathfrak{B}_j\{\mathscr{W}\} : j = 1, \ldots, b\} &\overset{\text{i.i.d.}}{\sim} \mathscr{B} \text{ in } D, \text{ and} \\ \{\Delta(\mathfrak{B}_j\{\mathscr{W}\}) : j = 1, \ldots, b\} &\overset{\text{i.i.d.}}{\sim} \sigma_{\widetilde{y}_p} Z \text{ in } \mathbb{R},\end{aligned}\right\} \tag{70}$$

where, for example, the first relation in Equation (70) means that the random elements $\{\mathfrak{X}_j\{\mathscr{W}\} : j = 1, \ldots, b\}$ are i.i.d. with the probability distribution of $\mathscr{W}$. The latter property holds because (i) Brownian motion is self-similar with Hurst index $1/2$ so that $\mathfrak{X}_j\{\mathscr{W}\} \overset{\text{d}}{=} \mathscr{W}$ for $j = 1, \ldots, b$ (Whitt 2002, §4.2.2); and (ii) by the independent-increments property of Brownian motion, the random elements $\{\mathfrak{X}_j\{\mathscr{W}\} : j = 1, \ldots, b\}$ are independent since they are respectively defined as rescaled increments of $\mathscr{W}$ on the disjoint subintervals $\big\{\big(\frac{j-1}{b},\frac{j}{b}\big]:j=1,\ldots,b\big\}$ of $[0,1]$ (Whitt 2002, §1.2.3). By a similar argument we see that

$$\big\{[\mathfrak{A}_j(\mathscr{W}),\mathfrak{B}_j\{\mathscr{W}\}]:j=1,\ldots,b\big\}\overset{\text{i.i.d.}}{\sim}[\mathscr{W}(1),\mathscr{B}]\text{ in }\mathbb{R}\times D. \tag{71}$$

The next step in completing the proof of Theorem 4 is to verify Equation (46). On the product space $\bigtimes_{j=1}^{b}(\mathbb{R}\times D)_j$, we define the metric $\mathfrak{h}(\mathfrak{s}_1,\mathfrak{s}_2)$ thus: for $\mathfrak{s}_i=[(a_{i,1},\zeta_{i,1}),\ldots,(a_{i,b},\zeta_{i,b})]\in\bigtimes_{j=1}^{b}(\mathbb{R}\times D)_j$ and $i\in\{1,2\}$, let

$$\mathfrak{h}(\mathfrak{s}_1,\mathfrak{s}_2)\equiv\max_{1\le j\le b}\big\{\max\{|a_{1,j}-a_{2,j}|,d(\zeta_{1,j},\zeta_{2,j})\}\big\}; \tag{72}$$

see Billingsley (1999, §M10, Equation (9)). In the proof of Theorem 2, we showed that $R\times D$ with the metric τ is separable; the same argument shows that $\bigtimes_{j=1}^{b}(\mathbb{R}\times D)_j$ with the metric $\mathfrak{h}$ is also separable. For $m\ge1$, we let

$$\left.\begin{aligned}\mathfrak{s}_1(m)&\equiv\sigma_{\widetilde{y}_p}\big[\big(\mathfrak{A}_1(\mathscr{I}_n),\mathfrak{B}_1\{\mathscr{I}_n\}\big),\ldots,\big(\mathfrak{A}_b(\mathscr{I}_n),\mathfrak{B}_b\{\mathscr{I}_n\}\big)\big]\\&=\bigg[\Big(\tfrac{m^{1/2}}{f(y_p)}\big(\overline{I}_{1,m}(y_p)-p\big),\mathscr{T}_{1,m}\Big),\ldots,\Big(\tfrac{m^{1/2}}{f(y_p)}\big(\overline{I}_{b,m}(y_p)-p\big),\mathscr{T}_{b,m}\Big)\bigg],\\\mathfrak{s}_2(m)&\equiv\Big[\Big(m^{1/2}\big(y_p-\widehat{y}_p(1,m)\big),T_{1,m}\Big),\ldots,\Big(m^{1/2}\big(y_p-\widehat{y}_p(b,m)\big),T_{b,m}\Big)\Big].\end{aligned}\right\} \tag{73}$$

To obtain the weak-convergence limit for $\mathfrak{s}_2(m)$, first we invoke the FCLT (**??**), Equation (68), and the continuous-mapping theorem to conclude that

$$\mathfrak{s}_1(m)\underset{m\to\infty}{\Longrightarrow}\sigma_{\widetilde{y}_p}\big[\big(\mathfrak{A}_1(\mathscr{W}),\mathfrak{B}_1\{\mathscr{W}\}\big),\ldots,\big(\mathfrak{A}_b(\mathscr{W}),\mathfrak{B}_b\{\mathscr{W}\}\big)\big] \tag{74}$$

in the product space $\bigtimes_{j=1}^{b}(\mathbb{R}\times D)_j$ with the metric $\mathfrak{h}$. From Equations (72)–(73) we see that

$$\mathfrak{h}[\mathfrak{s}_1(m),\mathfrak{s}_2(m)]=\max_{1\le j\le b}\left\{\max\left(\left|\frac{m^{1/2}}{f(y_p)}\big[\overline{I}_{j,m}(y_p)-p\big]-m^{1/2}\big[y_p-\widehat{y}_p(j,m)\big]\right|,d(\mathscr{T}_{j,m},T_{j,m})\right)\right\}. \tag{75}$$

For $j=1,\ldots,b$, let $Q_{j,m}$ denote the remainder in the Bahadur representation (5) of the BQE $\widehat{y}_p(j,m)$,

$$Q_{j,m}\equiv\widehat{y}_p(j,m)-y_p+\frac{\overline{I}_{j,m}(y_p)-p}{f(y_p)}\text{ for }m\ge1; \tag{76}$$

hence Theorem 1 ensures there are a.s. bounded r.v.'s $\mathfrak{U}_j\in\mathbb{R}^+$ and $\mathfrak{N}_j\in\mathbb{Z}^+$ such that

$$|Q_{j,m}|\le\mathfrak{U}_j\frac{(\log m)^{3/2}}{m^{3/4}}\text{ for }m\ge\mathfrak{N}_j\text{ a.s.} \tag{77}$$

Therefore in Equation (75), for $m\ge\mathfrak{N}_j$ a.s. Equation (77) yields

$$\left|\frac{m^{1/2}}{f(y_p)}\big[\overline{I}_{j,m}(y_p)-p\big]-m^{1/2}\big[y_p-\widehat{y}_p(j,m)\big]\right|=\big|m^{1/2}Q_{j,m}\big|\le\mathfrak{U}_j\frac{(\log m)^{3/2}}{m^{1/4}}\underset{m\to\infty}{\Longrightarrow}0. \tag{78}$$

By an analysis similar to Equation (28) that has been adapted to the Bahadur representation for BQEs based on batches of size m, we also have the following result for $j=1,\ldots,b$:

$$\begin{aligned}\big|\mathscr{T}_{j,m}(t)-T_{j,m}(t)\big|&=\left|\frac{\lfloor mt\rfloor}{m^{1/2}}\big(Q_{j,\lfloor mt\rfloor}-Q_{j,m}\big)-\left(\frac{mt-\lfloor mt\rfloor}{m}\right)m^{1/2}\big[\overline{I}_{j,m}(y_p)-p\big]/f(y_p)\right|\\&\le\frac{\lfloor mt\rfloor}{m^{1/2}}\big(|Q_{j,\lfloor mt\rfloor}|+|Q_{j,m}|\big)+m^{-1}\big|m^{1/2}\big[\overline{I}_{j,m}(y_p)-p\big]/f(y_p)\big|\\&\le2\mathfrak{U}_j\frac{(\log m)^{3/2}}{m^{1/4}}+m^{-1}\big|m^{1/2}\big[\overline{I}_{j,m}(y_p)-p\big]/f(y_p)\big|\text{ for each }t\in[0,1]\text{ and }m\ge\mathfrak{N}_j\text{ a.s.}\end{aligned} \tag{79}$$

By adapting to batches of size m the FCLT (**??**) and the analysis used in Equations (29)–(30), we can apply Equation (79) to obtain

$$d(\mathscr{T}_{j,m}, T_{j,m}) \underset{m\to\infty}{\Longrightarrow} 0 \text{ for } j = 1, \dots, b; \tag{80}$$

hence Equations (75), (78), and (80) ensure that

$$\mathfrak{h}[\mathfrak{s}_1(m), \mathfrak{s}_2(m)] \underset{m\to\infty}{\Longrightarrow} 0. \tag{81}$$

Equations (74), (75), and (81) as well as the convergence-together theorem ensure that

$$\mathfrak{s}_2(m) \underset{m\to\infty}{\Longrightarrow} \sigma_{\widetilde{y}_p} \big[\big(\mathfrak{A}_1(\mathscr{W}), \mathfrak{B}_1\{\mathscr{W}\}\big), \dots, \big(\mathfrak{A}_b(\mathscr{W}), \mathfrak{B}_b\{\mathscr{W}\}\big) \big] \tag{82}$$

in $\times_{j=1}^{b}(\mathbb{R}\times D)_j$ with the metric $\mathfrak{h}$. Equations (71), (73), and (82) ensure the validity of the first two relations in Equation (46).

To verify the third relation in Equation (46) by the same general approach used to derive the first two relations, on the product space $\times_{j=1}^{b}(\mathbb{R}^2)_j$, we define the metric $\mathfrak{d}(\mathfrak{r}_1, \mathfrak{r}_2)$ as follows: for $\mathfrak{r}_i = [(a_{i,1}, u_{i,1}), \dots, (a_{i,b}, u_{i,b})] \in \times_{j=1}^{b}(\mathbb{R}^2)_j$ and $i \in \{1, 2\}$, let

$$\mathfrak{d}(\mathfrak{r}_1, \mathfrak{r}_2) \equiv \max_{1\le j\le b} \big\{ \max\{|a_{1,j} - a_{2,j}|, |u_{1,j} - u_{2,j}|\} \big\}, \tag{83}$$

which yields the same product topology on $\mathbb{R}^{2b}$ as the standard Euclidean metric (Billingsley 1999, Equations (8)–(10)) so that the resulting metric space is separable. For $m \ge 1$, we let

$$\left.\begin{aligned} \mathfrak{r}_1(m) &\equiv \sigma_{\widetilde{y}_p} \big[\big(\mathfrak{A}_1(\mathscr{I}_n), \Delta_m(\mathfrak{B}_1\{\mathscr{I}_n\})\big), \dots, \big(\mathfrak{A}_b(\mathscr{I}_n), \Delta_m(\mathfrak{B}_b\{\mathscr{I}_n\})\big) \big] \\ &= \bigg[\Big(\tfrac{m^{1/2}}{f(y_p)} \big(\bar{I}_{1,m}(y_p) - p\big), \Delta_m(\mathscr{T}_{1,m}) \Big), \dots, \Big(\tfrac{m^{1/2}}{f(y_p)} \big(\bar{I}_{b.m}(y_p) - p\big), \Delta_m(\mathscr{T}_{b,m}) \Big) \bigg] \\ \mathfrak{r}_2(m) &\equiv \bigg[\Big(m^{1/2}\big(y_p - \widehat{y}_p(1,m)\big), \Delta_m(T_{1,m}) \Big), \dots, \Big(m^{1/2}\big(y_p - \widehat{y}_p(b,m)\big), \Delta_m(T_{b,m}) \Big) \bigg]. \end{aligned}\right\} \tag{84}$$

To obtain the weak-convergence limit for $\mathfrak{r}_2(m)$, we note first by the FCLT (**??**), Equation (68), and the generalized continuous-mapping theorem that

$$\mathfrak{r}_1(m) \underset{m\to\infty}{\Longrightarrow} \sigma_{\widetilde{y}_p} \big[\big(\mathfrak{A}_1(\mathscr{W}), \Delta(\mathfrak{B}_1\{\mathscr{W}\})\big), \dots, \big(\mathfrak{A}_b(\mathscr{W}), \Delta(\mathfrak{B}_b\{\mathscr{W}\})\big) \big] \tag{85}$$

in the product space $\times_{j=1}^{b}(\mathbb{R}^2)_j$ with the metric $\mathfrak{d}$. Using an analysis similar to Equations (75)–(82) together with the (final) limit result in Equation (68), we conclude that in $\times_{j=1}^{b}(\mathbb{R}^2)_j$ with the metric $\mathfrak{d}$,

$$\begin{aligned} &\big[\big(m^{1/2}\big(y_p - \widehat{y}_p(1,m)\big), \Delta_m(T_{1,m})\big), \dots, \big(m^{1/2}\big(y_p - \widehat{y}_p(b,m)\big), \Delta_m(T_{b,m})\big) \big] \\ &\qquad \underset{m\to\infty}{\Longrightarrow} \sigma_{\widetilde{y}_p} \big[\big(\mathfrak{A}_1(\mathscr{W}), \Delta(\mathfrak{B}_1\{\mathscr{W}\})\big), \dots, \big(\mathfrak{A}_b(\mathscr{W}), \Delta(\mathfrak{B}_b\{\mathscr{W}\})\big) \big]. \end{aligned} \tag{86}$$

Equations (71) and (86) ensure the validity of the third relation in Equation (46).

Next we verify Equation (52) and the first part of Equation (53). By an analysis similar to the derivation of Equation (86), we see that

$$\left.\begin{aligned} &\Delta_m^2(T_{j,m}) \underset{m\to\infty}{\Longrightarrow} \sigma_{\widetilde{y}_p}^2 \Delta^2(\mathfrak{B}_j\{\mathscr{W}\}) \overset{\mathrm{d}}{=} \sigma_{\widetilde{y}_p}^2 \chi_1^2 \;\text{ for }\; j = 1, \dots, b; \;\text{ and} \\ &\big\{ \big[\mathfrak{A}_j(\mathscr{W}), \Delta^2(\mathfrak{B}_j\{\mathscr{W}\})\big] : j = 1, \dots, b \big\} \overset{\text{i.i.d.}}{\sim} [\mathscr{W}(1), \Delta^2(\mathscr{B})] \end{aligned}\right\} \tag{87}$$

so the first relation in Equation (53) follows from the definition (48) of $\mathscr{A}^2_{b,m}$ and Equation (87). Moreover, the last relation in Equation (46) and Equation (87) imply that the sequences $\{m^{1/2}[\widehat{y}_p(j,m) - y_p] : j = 1,\ldots,b\}$ and $\{\Delta^2_m(T_{j,m}) : j = 1,\ldots,b\}$ are asymptotically independent as $m \to \infty$. Since $n^{1/2}[\overline{\widehat{y}}(b,m) - y_p]$ is the average of the $\{m^{1/2}[\widehat{y}_p(j,m) - y_p] : j = 1,\ldots,b\}$ times $b^{1/2}$ and $\mathscr{A}^2_{b,m}$ is the average of the $\{\Delta^2_m(T_{j,m}) : j = 1,\ldots,b\}$, a straightforward application of the continuous-mapping theorem yields Equation (49), showing that $n^{1/2}[\overline{\widehat{y}}_p(b,m) - y_p]$ and $\mathscr{A}^2_{b,m}$ are asymptotically independent as $m \to \infty$.

If the r.v.'s $\{\chi^2_1(j) : 1 \le j \le \nu\} \overset{\text{i.i.d.}}{\sim} \chi^2_1$, then we have

$$\mathscr{L}_\nu \equiv \left[\nu^{-1}\sum_{j=1}^{\nu}\chi^2_1(j)\right]^{1/2} \overset{\text{d}}{=} \left(\chi^2_\nu/\nu\right)^{1/2} > 0 \text{ for } \nu \ge 1 \text{ a.s.;} \tag{88}$$

and from Theorem 1, Equations (52)–(53), Equation (88), and the continuous-mapping theorem, we see that the pivotal quantity

$$\mathscr{G}_{b,m} \equiv \frac{\overline{\widehat{y}}_p(b,m) - y_p}{\left(\mathscr{A}^2_{b,m}/n\right)^{1/2}} = \frac{n^{1/2}\left[\overline{\widehat{y}}_p(b,m) - y_p\right]/\sigma_{\widehat{y}_p}}{\mathscr{A}_{b,m}/\sigma_{\widehat{y}_p}} \underset{m\to\infty}{\Longrightarrow} \frac{Z}{\mathscr{L}_b} \overset{\text{d}}{=} t_b, \tag{89}$$

where Z is independent of $\mathscr{L}_b$. For $\nu \ge 1$, let $G_\nu(u) \equiv \Pr\{t_\nu \le u\}$, $u \in \mathbb{R}$, denote the c.d.f. of t_ν. Since $G_\nu(u)$ is continuous at each $u \in \mathbb{R}$ for $\nu \ge 1$, Equation (89) and the standard definition of convergence in distribution for a sequence of r.v.'s (Whitt 2002, Equation (2.7)) ensure that

$$\lim_{m\to\infty} \Pr\{\mathscr{G}_{b,m} \le u\} = G_b(u) \text{ at each } u \in \mathbb{R}; \text{ hence} \tag{90}$$

$$\begin{aligned}
&\lim_{m\to\infty} \Pr\left\{\overline{\widehat{y}}_p(b,m) - t_{1-\alpha/2,b}\left(\mathscr{A}^2_{b,m}/n\right)^{1/2} \le y_p \le \overline{\widehat{y}}_p(b,m) + t_{1-\alpha/2,b}\left(\mathscr{A}^2_{b,m}/n\right)^{1/2}\right\} \\
&\quad = \lim_{m\to\infty} \Pr\{-t_{1-\alpha/2,b} \le \mathscr{G}_{b,m} \le t_{1-\alpha/2,b}\} \\
&\quad = G_b(t_{1-\alpha/2,b}) - G_b(-t_{1-\alpha/2,b}) = (1-\alpha/2) - \alpha/2 = 1-\alpha.
\end{aligned} \tag{91}$$

Thus the CI (47) is asymptotically exact as $m \to \infty$.

To finish proving asymptotic exactness of the CI (57), we must show (i) $mS^2_{b,m} \underset{m\to\infty}{\Longrightarrow} \sigma^2_{\widehat{y}_p}\chi^2_{b-1}/(b-1)$; (ii) $n^{1/2}[\overline{\widehat{y}}_p(b,m) - y_p]$ and $mS^2_{b,m}$ are asymptotically independent; (iii) $\mathscr{A}^2_{b,m}$ and $mS^2_{b,m}$ are asymptotically independent; (iv) $\mathscr{V}_{b,m} \underset{m\to\infty}{\Longrightarrow} \sigma^2_{\widehat{y}_p}\chi^2_{2b-1}/(2b-1)$; and (v) $n^{1/2}[\overline{\widehat{y}}_p(b,m) - y_p]$ and $\mathscr{V}_{b,m}$ are asymptotically independent. Item (i) follows from the CLT (6) for the BQEs and the continuous-mapping theorem since $mS^2_{b,m}$ is a continuous function of the BQEs. Item (ii) follows from the CLT (6) for BQEs and the basic property that for a random sample from a nondegenerate normal distribution, the sample mean and the sample variance are independent (Wilks 1962, Theorem 8.4.2). Item (iii) follows by noting that $mS^2_{b,m}$ is a continuous function of only the r.v.'s $\{m^{1/2}[\widehat{y}_p(j,m) - y_p] : j = 1,\ldots,b\}$, whereas $\mathscr{A}^2_{b,m}$ is a continuous function of only the r.v.'s $\{\Delta^2_m(T_{j,m}) : j = 1,\ldots,b\}$; and since these two sets of r.v.'s are asymptotically independent, so are $mS^2_{b,m}$ and $\mathscr{A}^2_{b,m}$. Item (iv) follows from items (i) and (iii) and Equation (49). Item (v) follows from Equation (49) and item (ii). For completeness, we note that items (i)–(v) imply Equation (54).

Finally using items (i)–(v) in an analysis similar to that in Equations (88)–(91), we see that the CI (57) is asymptotically exact as $m \to \infty$. $\square$

Next we establish the asymptotic exactness of CIs that are comparable to Equations (47) and (57), respectively, but are centered on $\widetilde{y}_p(n)$,

$$\widetilde{y}_p(n) \pm t_{1-\alpha/2,b}\big(\mathscr{A}^2_{b,m}/n\big)^{1/2} \text{ and} \tag{92}$$

$$\widetilde{y}_p(n) \pm t_{1-\alpha/2,2b-1}\big(\mathscr{V}_{b,m}/n\big)^{1/2}. \tag{93}$$

Theorem 5. *If $\{Y_k : k \geq 1\}$ satisfies the assumptions of Theorem 1, then Equations (92) and (93) are asymptotically exact 100$(1-\alpha)$% CI estimators of y_p as $m \to \infty$.*

Proof We show the asymptotic exactness of the CI (92) by proving that

$$\left|\frac{\widetilde{y}_p(n) - y_p}{\big(\mathscr{A}^2_{b,m}/n\big)^{1/2}} - \frac{\bar{\bar{y}}_p(b,m) - y_p}{\big(\mathscr{A}^2_{b,m}/n\big)^{1/2}}\right| = \frac{n^{1/2}\big|\widetilde{y}_p(n) - \bar{\bar{y}}_p(b,m)\big|}{\mathscr{A}_{b,m}} \underset{m\to\infty}{\Longrightarrow} 0; \tag{94}$$

then the result will follow from Equation (89) and the convergence-together theorem. Equation (76) implies

$$\bar{\bar{y}}_p(b,m) = y_p - \frac{1}{b}\sum_{j=1}^{b}\frac{\bar{I}_{j,m}(y_p) - p}{f(y_p)} + \frac{1}{b}\sum_{j=1}^{b} Q_{j,m} = y_p - \frac{\bar{I}_n(y_p) - p}{f(y_p)} + \overline{Q}_m \text{ for } m \geq 1, \tag{95}$$

where $\overline{Q}_m \equiv b^{-1}\sum_{j=1}^{b} Q_{j,m}$. By Equation (77), the r.v.'s $\overline{\mathfrak{U}} \equiv b^{-1}\sum_{j=1}^{b}\mathfrak{U}_j$ and $\mathfrak{N}^* \equiv \max\{\mathfrak{N}_j : 1 \leq j \leq b\}$ are bounded a.s. so that

$$\big|\overline{Q}_m\big| \leq \overline{\mathfrak{U}}\frac{(\log m)^{3/2}}{m^{3/4}} \text{ for } m \geq \mathfrak{N}^* \text{ a.s.} \tag{96}$$

By Equations (7)–(9) and (95)–(96), for $m \geq \mathfrak{N}^*$ we have

$$n^{1/2}\big|\widetilde{y}_p(n) - \bar{\bar{y}}_p(b,m)\big| \leq (bm)^{1/2}\left[\mathfrak{U}\frac{(\log bm)^{3/2}}{(bm)^{3/4}} + \overline{\mathfrak{U}}\frac{(\log m)^{3/2}}{m^{3/4}}\right] \underset{m\to\infty}{\Longrightarrow} 0. \tag{97}$$

Moreover, from Equation (53), Equation (88), and the continuous-mapping theorem, we have

$$\mathscr{A}^{-1}_{b,m} \underset{m\to\infty}{\Longrightarrow} \big(\sigma^2_{\bar{y}_p}\chi^2_b/b\big)^{-1/2}, \tag{98}$$

which is a square-root inverted-gamma r.v. (Bernardo and Smith 2000, p. 119 and 431). From Equations (97)–(98) and Slutsky's theorem, we obtain Equation (94). Equations (89) and (94) as well as the convergence-together theorem ensure that

$$\frac{n^{1/2}\big[\widetilde{y}_p(n) - y_p\big]}{\mathscr{A}_{b,m}} \underset{m\to\infty}{\Longrightarrow} t_b; \tag{99}$$

and the asymptotic exactness of the CI (92) follows by an argument similar to Equations (88)–(91).

Proving the asymptotic exactness of the CI (93) parallels the analysis leading to Equations (94), (98), and (99). Equation (52) and items (iv)–(v) in the last paragraph of the proof of Theorem 4 ensure that

$$\frac{n^{1/2}\big[\bar{\bar{y}}_p(b,m) - y_p\big]}{\mathscr{V}^{1/2}_{b,m}} \underset{m\to\infty}{\Longrightarrow} t_{2b-1}. \tag{100}$$

Moreover Equation (88), Equation (55), and the continuous-mapping theorem imply that

$$\mathscr{V}_{b,m}^{-1/2} \underset{m\to\infty}{\Longrightarrow} \left[\sigma_{\widetilde{y}_p}^2 \chi_{2b-1}^2 / (2b-1)\right]^{-1/2},$$

which is another square-root inverted-gamma r.v. Hence Equation (97) and Slutsky's theorem imply that

$$\left|\frac{n^{1/2}\left[\widetilde{y}_p(n) - y_p\right]}{\mathscr{V}_{b,m}^{1/2}} - \frac{n^{1/2}\left[\overline{\overline{y}}_p(b,m) - y_p\right]}{\mathscr{V}_{b,m}^{1/2}}\right| = \left|n^{1/2}\left[\widetilde{y}_p(n) - \overline{\overline{y}}(b,n)\right]\mathscr{V}_{b,m}^{-1/2}\right| \underset{m\to\infty}{\Longrightarrow} 0. \tag{101}$$

Then Equation (100), Equation (101), and the convergence-together theorem ensure that

$$\frac{n^{1/2}\left[\widetilde{y}_p(n) - y_p\right]}{\mathscr{V}_{b,m}^{1/2}} \underset{m\to\infty}{\Longrightarrow} t_{2b-1}; \tag{102}$$

and the asymptotic exactness of the CI (93) follows by an argument similar to Equations (88)–(91). □

Finally we establish the asymptotic exactness of CIs that use the full-sample quantile estimator $\widetilde{y}_p(n)$ rather than $\overline{\overline{y}}_p(b,m)$ in both the numerator and denominator of the relevant pivotal quantities in an attempt to improve the small-sample precision and coverage probability of the resulting CIs,

$$\widetilde{y}_p(n) \pm t_{1-\alpha/2,b-1}\left(\widetilde{S}_{b,m}^2/b\right)^{1/2} \quad \text{and} \tag{103}$$

$$\widetilde{y}_p(n) \pm t_{1-\alpha/2,2b-1}\left(\widetilde{\mathscr{V}}_{b,m}/n\right)^{1/2}, \quad \text{where} \tag{104}$$

$$\widetilde{S}_{b,m}^2 \equiv (b-1)^{-1}\sum_{j=1}^{b}\left[\widehat{y}_p(j,m) - \widetilde{y}_p(n)\right]^2 \quad \text{and} \quad \widetilde{\mathscr{V}}_{b,m} \equiv \frac{b\mathscr{A}_{b,m}^2 + (b-1)m\widetilde{S}_{b,m}^2}{2b-1}. \tag{105}$$

Theorem 6. *If $\{Y_k : k \ge 1\}$ satisfies the assumptions of Theorem 1, then Equations (103) and (104) are asymptotically exact 100$(1-\alpha)$% CI estimators of y_p as $m \to \infty$.*

Proof Equations (52)–(54) ensure that

$$\left[n^{1/2}\left(\overline{\overline{y}}_p(b,m) - y_p\right), mS_{b,m}^2\right] \underset{m\to\infty}{\Longrightarrow} \left[\sigma_{\widetilde{y}_p} Z, \frac{\sigma_{\widetilde{y}_p}^2 \chi_{b-1}^2}{b-1}\right]; \text{ and } Z \text{ and } \chi_{b-1}^2 \text{ are independent.} \tag{106}$$

We have

$$\begin{aligned} m\widetilde{S}_{b,m}^2 &= \frac{m}{b-1}\sum_{j=1}^{b}\left\{\left[\widehat{y}_p(j,m) - \overline{\overline{y}}_p(b,m)\right] + \left[\overline{\overline{y}}_p(b,m) - \widetilde{y}_p(n)\right]\right\}^2 \\ &= mS_{b,m}^2 + \frac{2m}{b-1}\left[\overline{\overline{y}}_p(b,m) - \widetilde{y}_p(n)\right]\sum_{j=1}^{b}\left[\widehat{y}_p(j,m) - \overline{\overline{y}}_p(b,m)\right] + \frac{m}{b-1}\sum_{j=1}^{b}\left[\widetilde{y}_p(n) - \overline{\overline{y}}_p(b,m)\right]^2 \\ &= mS_{b,m}^2 + (b-1)^{-1}\left\{n^{1/2}\left[\widetilde{y}_p(n) - \overline{\overline{y}}_p(b,m)\right]\right\}^2; \end{aligned} \tag{107}$$

thus Equations (97) and (107) imply that

$$\mathfrak{g}_m \equiv n^{1/2}\left[\widetilde{y}_p(n) - \overline{\overline{y}}_p(b,m)\right] \underset{m\to\infty}{\Longrightarrow} 0 \quad \text{and} \quad \vartheta_m \equiv m\widetilde{S}_{b,m}^2 - mS_{b,m}^2 = (b-1)^{-1}\mathfrak{g}_m^2 \underset{m\to\infty}{\Longrightarrow} 0. \tag{108}$$

We observe that

$$\left[n^{1/2}\left(\widetilde{y}_p(n) - y_p\right), m\widetilde{S}_{b,m}^2\right] \quad = \quad \left[n^{1/2}\left(\overline{\overline{y}}_p(b,m) - y_p\right), mS_{b,m}^2\right] + \left(\mathfrak{g}_m, \vartheta_m\right) \tag{109}$$

$$\underset{m\to\infty}{\Longrightarrow} \left[\sigma_{\widetilde{y}_p} Z, \frac{\sigma^2_{\widetilde{y}_p}\chi^2_{b-1}}{b-1}\right]; \text{ and } Z \text{ and } \chi^2_{b-1} \text{ are independent} \tag{110}$$

by Equations (106), (108), and (109) as well as the convergence-together theorem. Thus by Equation (110), Equation (88), and the continuous-mapping theorem, we have

$$\frac{n^{1/2}[\widetilde{y}_p(n) - y_p]}{m^{1/2}\widetilde{S}_{b,m}} \underset{m\to\infty}{\Longrightarrow} \frac{Z}{[\chi^2_{b-1}/(b-1)]^{1/2}} \stackrel{\mathrm{d}}{=} t_{b-1};$$

and the exactness of the CI (103) follows by an argument similar to Equations (88)–(91).

Proving the asymptotic exactness of the CI (104) parallels the analysis leading to Equations (106)–(110). From Equations (52), (55), and (56) we see that

$$\left[n^{1/2}(\overline{\overline{y}}_p(b,m) - y_p), \mathscr{V}_{b,m}\right] \underset{m\to\infty}{\Longrightarrow} \left[\sigma_{\widetilde{y}_p} Z, \frac{\sigma^2_{\widetilde{y}_p}\chi^2_{2b-1}}{2b-1}\right]; \text{ and } Z \text{ and } \chi^2_{2b-1} \text{ are independent.} \tag{111}$$

By Equations (107) and (108), we have

$$\vartheta^*_m \equiv \widetilde{\mathscr{V}}_{b,m} - \mathscr{V}_{b,m} = \left(\frac{b-1}{2b-1}\right)\vartheta_m \underset{m\to\infty}{\Longrightarrow} 0.$$

We observe that

$$\begin{aligned} \left[n^{1/2}(\widetilde{y}_p(n) - y_p), \widetilde{\mathscr{V}}_{b,m}\right] &= \left[n^{1/2}(\overline{\overline{y}}_p(b,m) - y_p), \mathscr{V}_{b,m}\right] + (\mathfrak{g}_m, \vartheta^*_m) \\ &\underset{m\to\infty}{\Longrightarrow} \left[\sigma_{\widetilde{y}_p} Z, \frac{\sigma^2_{\widetilde{y}_p}\chi^2_{2b-1}}{2b-1}\right]; \text{ and } Z \text{ and } \chi^2_{2b-1} \text{ are independent.} \end{aligned} \tag{112}$$

Thus by Equation (112), Equation (88), and the continuous-mapping theorem, we have

$$\frac{n^{1/2}[\widetilde{y}_p(n) - y_p]}{\widetilde{\mathscr{V}}^{1/2}_{b,m}} \underset{m\to\infty}{\Longrightarrow} \frac{Z}{[\chi^2_{2b-1}/(2b-1)]^{1/2}} \stackrel{\mathrm{d}}{=} t_{2b-1};$$

and the exactness of the CI (104) follows by an argument similar to Equations (88)–(91). □

6. Computational Complexity

[I'm thinking that a lot of the next paragraph or two has been cut-and-pasted into the JoS paper.] In this section we show that the effort required to compute the batched-STS area estimator $\mathscr{A}^2_{b,m}$ in Equation (48) is dominated by the effort to sort the data (each batch and the entire sample). To simplify the discussion, we first consider the case with a single batch of size n. Since the evaluation of the STS quantile-estimation process $\{T_n(t) : t \in [0,1]\}$ defined by Equation (16) at the points $t \in \{1/n, 2/n, \dots, (n-1)/n, 1\}$ involves the computation of p-quantile estimates from all partial samples of sizes $1, \dots, n$, one practically needs to start with a complete sort of the sample $\{Y_1, \dots, Y_n\}$. Since we use an object-oriented paradigm in Java, the nonprimitive list is sorted using the default timsort algorithm of Tim Peters, a stable hybrid between merge sort and insertion sort with $O(n \log_2 n)$ average and worst-time complexity based on techniques from McIlroy (1993).

It should be clear that once we have evaluated the STS quantile-estimation process $\{T_n(t) : t \in [0,1]\}$ defined by Equation (16) at the points $t \in \{1/n, 2/n, \dots, (n-1)/n, 1\}$, the evaluation of $\mathscr{A}^2_{1,n}$ using Equation

(32) takes $O(n)$ extra time. For clarity, we temporarily adopt the classical notation $Y_{\ell:k}$ for the ℓth order statistic from the kth partial sample $\{Y_1,\ldots,Y_k\}$ for $1\le\ell\le k\le n$ so that $\widetilde{y}_p(k)=Y_{\lceil kp\rceil:k}$ for $1\le k\le n$. Then the evaluation of $T_n(k/n)$ reduces to the computation of $Y_{\lceil kp\rceil:k}$ for $k=1,\ldots,n$. Below we show how this task can be accomplished recursively in $O(n)$ time using object orientation and proceeding backwards to compute $Y_{\lceil kp\rceil:k}$ in stage k for $k=n,n-1,\ldots,1$.

We store the original dataset $\{Y_1,\ldots,Y_n\}$ in a list comprised of n instances of an object. The kth instance has the following properties: the value Y_k, a reference (property) to the predecessor of that object in the original list having the value Y_{k-1}, and references to the predecessor and successor of that object in the sorted list. For brevity, we will often refer to the kth object by the usual symbol Y_k for its value.

We proceed by sorting the original list to obtain the sorted list $Y_{1:n}\le Y_{2:n}\le\cdots\le Y_{n:n}$ and setting the predecessor/successor references for each object in the sorted list (essentially forming a doubly linked list of object instances). Starting at stage n, we obtain the value $Y_{\lceil np\rceil:n}$ from the $\lceil np\rceil$th object in the sorted list in $O(n)$ time.

We now focus on the recursive computation of $Y_{\lceil kp\rceil:k}$ from $Y_{\lceil (k+1)p\rceil:k+1}$ for $k\le n-1$. The location of Y_{k+1} in the sorted list can be identified directly (in $O(1)$ time) using the predecessor reference of Y_{k+2} in the original list. Since $p\in(0,1)$, we have only two potential cases:

- $\lceil kp\rceil=\lceil (k+1)p\rceil$: If the value $Y_{k+1}\le Y_{\lceil (k+1)p\rceil:k+1}$, then we set $Y_{\lceil kp\rceil:k}$ equal to the successor of $Y_{\lceil (k+1)p\rceil:k+1}$ in the sorted list; otherwise, we set $Y_{\lceil kp\rceil:k}=Y_{\lceil (k+1)p\rceil:k+1}$.
- $\lceil kp\rceil=\lceil (k+1)p\rceil-1$: If the value $Y_{k+1}\ge Y_{\lceil (k+1)p\rceil:k+1}$, then we set $Y_{\lceil kp\rceil:k}$ equal to the predecessor of $Y_{\lceil (k+1)p\rceil:k+1}$ in the sorted list; otherwise, we set $Y_{\lceil kp\rceil:k}=Y_{\lceil (k+1)p\rceil:k+1}$.

After the update, we "remove" Y_{k+1} from the sorted list by adjusting the predecessor and successor references from and to its previous successor and predecessor elements, respectively, in the sorted list (essentially, the list now contains k items because there are no references to/from Y_{k+1}). Since this recursive evaluation of $Y_{\lceil kp\rceil:k}$ from $Y_{\lceil (k+1)p\rceil:k+1}$ takes $O(1)$ time, the evaluation of $Y_{\lceil kp\rceil:k}$ for $k=n,n-1,\ldots,1$ takes a total of $O(n)$ time. It follows that the computation of $\mathscr{A}^2_{1,n}$ takes a total of $O(n)$ time on top of the time to sort the entire sample.

Remark 4. Clearly, the use of objects results in higher memory usage. If one uses traditional (primitive) arrays instead of objects, the location of Y_{k+1} in the sorted array can be found in $O(\log_2(k+1))$ time (e.g., using a binary search); therefore the total time required for the evaluation of the values $Y_{\lceil kp\rceil:k}$ jumps to $O(n\log_2 n)$.

In the case of $b>1$ batches, the average and worst-case time for sorting the batches and computing the full-sample point estimator remains $O(n\log_2 n)$ and the additional time for computing $\mathscr{A}^2_{b,m}$ remains linear in n because $bO(m)=O(n)$. It should be clear that variance estimators based solely on BQEs (e.g., $m\widetilde{S}^2_{b,m}$ defined by Equation (105)) can be computed in parallel with $\mathscr{A}^2_{b,m}$.

Remark 5. We close this section by noting that the calculation of a BQE-based estimator alone can be achieved in $O(n)$ average time using a quickselect algorithm that does not sort observations that are less than a desired order statistic; cf. Section 9.2 of Cormen et al. (2009). ◂

7. Experimental Results

In this section we conduct an empirical evaluation of the performance of the following variance-parameter estimators:

- the batched-STS area estimator $\mathscr{A}^2_{b,m}$ defined by Equation (48);
- the modified estimator $m\widetilde{S}^2_{b,m}$ defined by Equation (105) based on the BQEs $\{\widehat{y}_p(j,m)\}$ and the full-sample point estimator $\widetilde{y}_p(n)$, and referred to as the nonoverlapping batched quantile (NBQ) estimator;
- the combined estimator $\widetilde{\mathscr{V}}_{b,m}$ defined by Equation (105) based on $\mathscr{A}^2_{b,m}$ and $m\widetilde{S}^2_{b,m}$; and
- the $100(1-\alpha)\%$ CIs for y_p defined by Equations (92), (103), and (104), respectively.

We skipped the alternative variance-parameter estimator $mS^2_{b,m}$ based on Equation (51) for two reasons: (i) preliminary evaluation [check this statement] revealed that it is substantially more biased than the NBQ estimator $m\widetilde{S}^2_{b,m}$ for extreme values of p and small batch sizes m; and (ii) the NBQ estimator $m\widetilde{S}^2_{b,m}$ is used in the Sequest procedure of Alexopoulos et al. (2019b).

The goal of this study is the validation of our theoretical findings and, in particular, to showcase the superiority of the combined estimator $\widetilde{\mathscr{V}}_{b,m}$ with regard to its efficiency, as its asymptotic variance $\lim_{m\to\infty}\mathrm{Var}\big[\widetilde{\mathscr{V}}_{b,m}\big]$ is nearly 50% smaller than the asymptotic variances $\lim_{m\to\infty}\mathrm{Var}\big[\mathscr{A}^2_{b,m}\big]$ and $\lim_{m\to\infty}\mathrm{Var}\big[m\widetilde{S}^2_{b,m}\big]$ of its respective constituents. (Note that the three asymptotic variances in the preceding statement must be carefully distinguished from the variance parameter $\sigma^2_{\widetilde{y}_p}$.) We anticipate that the combined estimator $\widetilde{\mathscr{V}}_{b,m}$ will play a key role in the development of sequential procedures for steady-state quantile estimation that will improve upon the BQE-based methods of Alexopoulos et al. (2017) and Alexopoulos et al. (2019b). Alexopoulos et al. (2016) provides the groundwork for batched-STS sequential procedures based on batching and STS to estimate the steady-state mean.

Our analysis focuses on the constant weight function $w_0(t) = \sqrt{12}$, $t \in [0,1]$ (Schruben 1983). Although the literature includes other weight functions such as the quadratic weight $w_2(t) = \sqrt{840}(3t^2 - 3t + 1/2)$ (Goldsman et al. 1990) and the family of weights $\big\{w_{\cos,j}(t) = \sqrt{8}\pi j\cos(2\pi jt) : j = 1,2,\ldots\big\}$ (Foley and Goldsman 1999), these functions were designed to yield first-order unbiased estimators for the variance parameter $\lim_{n\to\infty} n\,\mathrm{Var}\big(\overline{Y}_n\big)$ associated with the sample mean $\overline{Y}_n = n^{-1}\sum_{k=1}^n Y_k$ of the stationary process $\{Y_k : k \ge 1\}$ (cf. Aktaran-Kalaycı et al. 2007); hence they were tailored to the computation of valid CIs for the steady-state mean. Since preliminary experimentation with these alternative weight functions in our quantile-estimation setting did not reveal any superiority over the constant weight function, ongoing work includes further evaluation as well as the identification of alternative weight functions aimed at reduction of the bias of the batched-STS area estimator $\mathscr{A}^2_{b,m}$ for small batch sizes m.

We consider two stationary test processes: a Gaussian first-order autoregressive (AR(1)) process with correlation coefficient 0.9 and a queue-waiting-time process from an M/M/1 queueing system with traffic intensity 0.8. For each process and value of p under study, we fix the number of batches at $b = 32$ and consider an increasing sequence of batch sizes $m = 2^{\mathscr{L}}$, where $\mathscr{L} \in \{10, 11, \ldots, 20\}$. We note that batch sizes with $\mathscr{L} \leq 15$ are often inadequate for variance-parameter estimation in these problems (Alexopoulos et al. 2019b).

All experiments were coded in Java using common random numbers generated by the package of L'Ecuyer et al. (2002). The numerical results were computed from 2,500 i.i.d. realizations of the AR(1) process and the M/M/1 queue-waiting-time process; and those results are summarized in two tables that appear in Sections 7.1 and 7.2, respectively. In each table, column 1 contains the values of p, y_p, and $\sigma^2_{\widetilde{y}_p}$ (the latter quantity is set in **bold red type**); column 2 contains the value of $\mathscr{L} = \log_2(m)$; columns 3, 8, and 13 respectively contain the average values of the selected variance-parameter estimators computed from 2,500 i.i.d. observations of those estimators; columns 4, 9, and 14 respectively contain the average bias of the selected variance-parameter estimators; and columns 5, 10, and 15 respectively contain the standard deviations of the selected variance-parameter estimators. For nominal 95% CI estimators of y_p that are respectively defined by Equations (92), (103), and (104), in each table columns 6, 11, and 16 have the heading "95% CI $\overline{H}$" and respectively contain the average CI half-lengths computed from 2,500 i.i.d. realizations of those CIs; moreover columns 7, 12, and 17 have the heading "95% CI Cover." and respectively contain the corresponding empirical CI coverage probabilities. Finally, the figures that appear in Sections 7.1 and 7.2 summarize the accuracy and precision of each variance-parameter estimator as the batch size increases by plotting estimates of the respective relative biases (as a percentage) and root mean squared errors (RMSEs).

7.1. AR(1) Process

The first test process is a stationary AR(1) time-series model. This process is characterized by the linear regression equation $Y_k = \phi Y_{k-1} + \varepsilon_k$ for $k \geq 1$, where the autoregressive parameter is $\phi \in (-1, 1)$, the initial state Y_0 follows the $N(0, 1)$ distribution, and the residuals $\{\varepsilon_k : k \geq 1\}$ are i.i.d. $N(0, 1 - \phi^2)$ and independent of Y_0. Since the marginal distribution of the Y_k is $N(0, 1)$, the p-quantile can be computed by $y_p = \Phi^{-1}(p)$, where $\Phi(\cdot)$ denotes the standard normal c.d.f. The AR(1) process is one in a large class of widely used time-series models that satisfy the GMC Condition; see Alexopoulos et al. (2019b, p. 1165).

[Some of the following text is similar to that of GMC1B, but I'm not particularly concerned with that.]

The asymptotic variance parameter for the AR(1) process was evaluated as follows. Let $\mathscr{T}(h, a)$ denote Owen's T-function:

$$\mathscr{T}(h, a) = \frac{1}{2\pi} \int_0^a \frac{\exp\left[-\frac{1}{2}h^2(1 + x^2)\right]}{1 + x^2}\, \mathrm{d}x \text{ for } h, a \in \mathbb{R}.$$

For two standard normal variates Z_1 and Z_2 with correlation $\varphi = \text{Corr}(Z_1, Z_2) \in (-1, 1)$, one has

$$\Pr\{Z_1 \le \Phi^{-1}(p), Z_2 \le \Phi^{-1}(p)\} = p - 2\mathscr{T}\left[\Phi^{-1}(p), \left(\frac{1-\varphi}{1+\varphi}\right)^{1/2}\right] \quad \text{for } p \in (0,1);$$

see Meyer (2013, Equation (3.12)). Since $\text{Corr}(Y_k, Y_{k+\ell}) = \phi^\ell$ for $\ell \ge 0$, we have

$$\Pr\{Y_k \le y_p, Y_{k+\ell} \le y_p\} = p - 2\mathscr{T}\left[\Phi^{-1}(p), \left(\frac{1-\phi^\ell}{1+\phi^\ell}\right)^{1/2}\right] \quad \text{for } \ell \ge 0.$$

Using the definition of correlation, one can obtain the following expression for the autocorrelation function $\{\rho_{I(y_p)}(\ell) : \ell \ge 0\}$ of the indicator process $\{I_k(y_p) : k \ge 1\}$ for the AR(1) model at lag ℓ:

$$\rho_{I(y_p)}(\ell) = 1 - \frac{2}{p(1-p)}\mathscr{T}\left[\Phi^{-1}(p), \left(\frac{1-\phi^\ell}{1+\phi^\ell}\right)^{1/2}\right] \quad \text{for } p \in (0,1) \text{ and } \ell \ge 0.$$

Owen's T-function was computed using the R package and the implementation of Azzalini (2020), which is based on a series expansion. Then the variance parameter $\sigma^2_{I(y_p)}$ for the indicator process $\{I_k(y_p) : k \ge 1\}$ was approximated by truncating the infinite sum $\sigma^2_{I(y_p)} = p(1-p)\left[1 + 2\sum_{\ell=1}^{\infty} \rho_{I(y_p)}(\ell)\right]$. Since for the $N(0,1)$ p.d.f. we have $f(y_p) = (2\pi)^{-1/2}\exp\left(-y_p^2/2\right)$, the approximation of $\sigma^2_{\widetilde{y}_p} = \sigma^2_{I(y_p)}/f^2(y_p)$ follows immediately.

For experimentation we selected the values $\phi = 0.9$ and $p \in \{0.75, 0.95, 0.99\}$. Because of the symmetry of the marginal $N(0,1)$ distribution, we did not consider values of $p < 1/2$. The results are summarized in Table 1, which clearly indicated that all three estimators of the variance parameter for $\widetilde{y}_p(n)$ and their respective standard deviations converged to their asymptotic limits reasonably fast, albeit with speed that diminishes as p approaches 1. Further, the estimated coverage probabilities of the three CIs for y_p respectively based on Equations (92), (103), and (104) hovered near the nominal value of 0.95. Of equal importance, the lower standard deviation of the combined estimator $\widetilde{\mathscr{V}}_{b,m}$ became evident from the plots of the RMSEs in Figure 1. Among the three values of p, the near-extreme case of $p = 0.99$ provided a few insights, the first of which will become more prominent with the second example in Section 7.2.

- For small batch sizes, the batched-STS area estimator $\mathscr{A}^2_{b,m}$ had significantly more bias than the NBQ estimator $m\widetilde{S}^2_{b,m}$, while the bias of the combined estimator $\widetilde{\mathscr{V}}_{b,m}$ typically fell between the biases of its constituents (see Figure 1).
- For small batch sizes, the batched-STS area estimator $\mathscr{A}^2_{b,m}$ had noticeably larger standard deviation than the NBQ estimator $m\widetilde{S}^2_{b,m}$. If $\{Y_k : k \ge 1\}$ is normal, then the asymptotic standard deviation of the batched-STS area estimator, namely, $\lim_{m\to\infty}\left(\text{Var}\left[\mathscr{A}^2_{b,m}\right]\right)^{1/2} = \left(2\sigma^4_{\widetilde{y}_p}/b\right)^{1/2}$, is a bit smaller than the respective value $\lim_{m\to\infty}\left(\text{Var}\left[m\widetilde{S}^2_{b,m}\right]\right)^{1/2} = \left(2\sigma^4_{\widetilde{y}_p}/(b-1)\right)^{1/2}$ for the NBQ estimator.

7.2. M/M/1 Queue Waiting Times

Our second test process $\{Y_k : k \ge 1\}$ was generated by a stationary M/M/1 queueing system with the first-in-first-out (FIFO) service discipline, arrival rate $\lambda = 0.8$, and service rate $\omega = 1$. Specifically, Y_k is the time spent in the queue by the kth arriving customer prior to service. In this system the steady-state server utilization is $\rho = \lambda/\omega = 0.8$, and the steady-state c.d.f. of Y_k is [using ρ for correlation in this paper]

$$F(y) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - \rho e^{-\omega(1-\rho)x} & \text{if } x \geq 0; \end{cases} \tag{113}$$

hence the steady-state distribution of X_k has mean $\mu_X = \rho/(\omega - \lambda) = 4$, and the quantiles of this distribution are easily evaluated by inverting Equation (113).

The steady-state distribution (113) is markedly nonnormal, having an atom at zero, an exponential tail, and a skewness of $2(3 - 3\rho + \rho^2)/[\rho^{1/2}(2 - \rho)^{3/2}] \approx 2.1093$. These properties can induce a significant skewness in the corresponding BQEs $\left\{\widehat{y}_p(j, m) : j = 1, \ldots, b\right\}$ that can degrade the performance of the CI defined by Equation (103), resulting in a coverage probability that can be substantially below the nominal level (Alexopoulos et al. 2017). Because of the atom at zero in the c.d.f. (113), we only considered values of $p > 1 - \rho = 0.20$.

Table 1 **Experimental results for the AR(1) process with $\phi = 0.9$. All estimates are based on 2,500 independent replications with $b = 32$ batches and batch sizes $m = 2^{\mathscr{L}}$, $\mathscr{L} \in \{10, 11, \ldots, 20\}$, where for nominal 95% CI estimators of y_p, the average CI half-length and coverage probability are denoted by "95% CI $\overline{H}$" and "95% CI Cover.", respectively.**

		Batched STS Area Estimator					NBQ Estimator					Combined Estimator				
p (y_p) **Var. Par.**	$\mathscr{L}$	Avg.	Bias	Std. Dev.	95% CI $\overline{H}$	95% CI Cover.	Avg.	Bias	Std. Dev.	95% CI $\overline{H}$	95% CI Cover.	Avg.	Bias	Std. Dev.	95% CI $\overline{H}$	95% CI Cover.
0.75	10	22.3	−0.6	6.0	0.0527	93.64	22.8	−0.1	5.8	0.0533	94.36	22.5	−0.4	4.2	0.0522	94.44
(0.6745)	11	22.7	−0.2	5.8	0.0376	94.60	22.8	−0.1	5.8	0.0377	94.32	22.8	−0.1	4.0	0.0371	94.44
22.9	12	22.9	0.0	5.7	0.0267	95.32	22.7	−0.2	5.8	0.0266	95.24	22.8	−0.1	4.1	0.0263	95.60
	13	22.7	−0.2	5.9	0.0188	94.76	22.8	−0.1	5.9	0.0189	94.96	22.7	−0.2	4.1	0.0185	95.24
	14	22.9	0.0	5.8	0.0134	95.12	22.8	−0.1	5.8	0.0134	95.44	22.9	0.0	4.1	0.0131	95.16
	15	22.8	−0.1	5.8	0.0094	94.80	22.8	−0.1	5.9	0.0094	95.12	22.8	−0.1	4.1	0.0093	94.80
	16	22.8	−0.1	5.7	0.0067	94.76	22.9	0.0	5.8	0.0067	95.08	22.9	0.0	4.1	0.0066	95.00
	17	22.7	−0.2	5.7	0.0047	95.24	22.9	0.0	5.8	0.0047	95.08	22.8	−0.1	4.0	0.0046	95.48
	18	22.9	0.0	5.7	0.0033	95.68	23.0	0.1	5.9	0.0034	96.16	22.9	0.0	4.1	0.0033	95.76
	19	22.8	−0.1	5.6	0.0024	94.92	23.0	0.1	5.8	0.0024	94.00	22.9	0.0	4.0	0.0023	94.12
	20	23.0	0.1	5.8	0.0017	95.00	22.8	−0.1	5.7	0.0017	95.52	22.9	0.0	4.0	0.0016	95.16
0.95	10	37.8	−0.5	11.4	0.0684	94.32	38.1	−0.2	10.1	0.0689	95.00	38.0	−0.3	7.9	0.0677	94.88
(1.6449)	11	38.4	0.1	10.8	0.0488	93.96	38.0	−0.3	9.8	0.0487	94.40	38.2	−0.1	7.4	0.0480	94.40
38.3	12	38.7	0.4	10.4	0.0347	95.28	38.1	−0.2	9.8	0.0345	94.56	38.4	0.1	7.2	0.0340	94.84
	13	38.1	−0.2	10.0	0.0243	94.36	38.1	−0.2	9.9	0.0244	95.00	38.1	−0.2	7.1	0.0240	94.64
	14	38.3	0.0	9.9	0.0173	95.12	38.3	0.0	9.7	0.0173	95.68	38.3	0.0	7.0	0.0170	95.40
	15	38.4	0.1	9.9	0.0122	94.68	38.1	−0.2	10.0	0.0122	94.40	38.3	0.0	7.0	0.0120	94.36
	16	38.2	−0.1	9.9	0.0086	95.32	38.3	0.0	9.8	0.0086	95.44	38.2	−0.1	7.1	0.0085	95.40
	17	38.2	−0.1	9.5	0.0061	95.72	38.4	0.1	9.9	0.0061	95.52	38.3	0.0	6.8	0.0060	95.56
	18	38.5	0.2	9.6	0.0043	95.16	38.7	0.4	9.9	0.0043	95.44	38.6	0.3	6.9	0.0043	95.24
	19	38.4	0.1	9.5	0.0031	94.40	38.7	0.4	9.8	0.0031	94.28	38.5	0.2	6.7	0.0030	94.52
	20	38.8	0.5	9.7	0.0022	94.96	38.3	0.0	9.7	0.0022	95.12	38.6	0.3	6.8	0.0021	95.08
0.99	10	76.4	−5.2	32.0	0.0964	92.92	79.6	−2.0	22.8	0.0995	94.52	77.9	−3.7	21.4	0.0966	93.96
(2.3263)	11	81.8	0.2	29.7	0.0709	94.32	81.2	−0.4	21.8	0.0712	94.84	81.5	−0.1	19.9	0.0700	94.56
81.6	12	84.0	2.4	26.1	0.0509	94.60	81.8	0.2	21.6	0.0505	94.32	82.9	1.3	17.4	0.0500	94.84
	13	82.6	1.0	23.8	0.0358	94.00	81.4	−0.2	21.7	0.0356	94.48	82.0	0.4	16.6	0.0352	94.12
	14	82.4	0.8	22.7	0.0253	95.36	81.4	−0.2	20.8	0.0252	95.44	81.9	0.3	15.4	0.0249	95.40
	15	81.8	0.2	21.2	0.0178	94.88	81.5	−0.1	20.3	0.0178	95.28	81.6	0.0	14.9	0.0176	95.12
	16	82.1	0.5	21.3	0.0126	94.92	81.1	−0.5	20.4	0.0126	94.92	81.6	0.0	14.7	0.0124	95.00
	17	81.8	0.2	20.7	0.0089	95.72	81.5	−0.1	20.6	0.0089	95.00	81.7	0.1	14.5	0.0088	95.04
	18	81.5	−0.1	21.3	0.0063	95.04	82.3	0.7	20.8	0.0063	94.92	81.9	0.3	14.9	0.0062	94.80
	19	82.1	0.5	20.9	0.0045	95.28	82.6	1.0	20.9	0.0045	94.52	82.3	0.7	14.7	0.0044	94.68
	20	83.0	1.4	20.8	0.0032	94.80	82.2	0.6	20.9	0.0032	95.00	82.6	1.0	14.4	0.0031	95.04

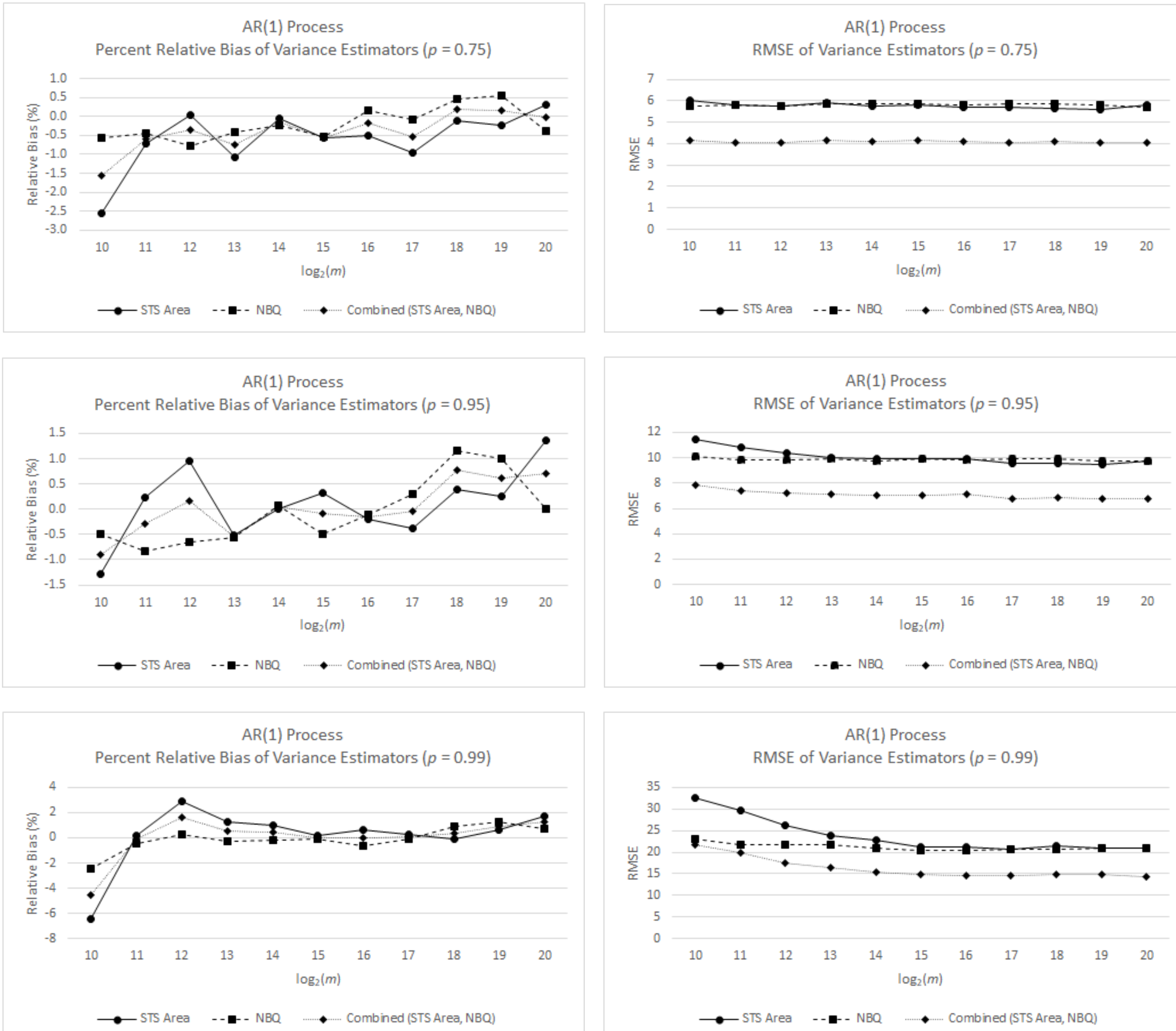

Figure 1 **Estimated percent relative bias and RMSE of the variance-parameter estimators for selected marginal quantiles of a stationary AR(1) process with lag-1 correlation $\phi = 0.9$. All estimates are based on 2,500 independent replications with $b = 32$ batches and batch sizes $m = 2^{\mathscr{L}}$, $\mathscr{L} \in \{10, 11, \ldots, 20\}$.**

We have recently proved that the GMC Condition is satisfied by queue-waiting times in the stable G/G/1 queueing system with non-heavy-tailed service times (i.e., the moment generating function of the service-time distribution exists in a neighborhood of zero); see Dingeç et al. (2022). Thus in steady-state operation, the M/M/1 queue-waiting time process satisfies the GMC Condition. Relatively elementary methods were used to verify the GMC Condition for G/G/1 queue-waiting times. By contrast, it seems that more-advanced methods are required to show that G/G/1 waiting times satisfy the assumption of strong approximation; see, for example, Damerdji (1994, Example 2.3).

[Some of this content is similar to GMC1B, but I'm not too concerned.] The variance parameter $\sigma^2_{I(y_p)}$ of the indicator process was computed from Equation (22) of Blomqvist (1967). After some algebra, we obtained the following analytical expression for the asymptotic variance parameter corresponding to $\widetilde{y}_p(n)$:

$$\sigma^2_{\widetilde{y}_p} = \frac{1}{\omega^2(1-\rho)^4}\left\{\frac{[-2+p(3-\rho)+2\rho](1+\rho)}{1-p} - 4\rho\,\ell\mathrm{n}\Big(\frac{\rho}{1-p}\Big)\right\}.$$

We generated the stationary version $\{Y_k : k \geq 1\}$ of this waiting-time process by sampling Y_1 using Equation (113), and then using Lindley's recursion. Table 2 below lists the numerical experimental outcomes. We selected three values of p, $p = 0.25$ (corresponding to the lower quartile and near the value $1-\rho = 0.2$), $p = 0.75$, and the extreme value $p = 0.99$.

A careful examination of Table 2 confirmed that all three variance-parameter estimators and their standard deviations converged to the respective theoretical limits, but at a significantly lower rate than for the AR(1) process in Section 7.1. Most importantly, it revealed the presence of substantial bias in the variance-parameter estimators for small batch sizes m; this bias apparently became more prominent for large values of p (near-extreme quantiles). We believe that this bias is primarily explained by the bias of the point estimator $\widetilde{y}_p(n)$ that is evident in the Bahadur representation (9). Ongoing work includes a comprehensive study of the relationship between the bias of $\widetilde{y}_p(n)$ and the bias of the batched-STS area estimator $\mathscr{A}^2_{b,m}$. The magnitude of this small-batch bias of the variance-parameter estimators corresponding to the full-sample quantile estimator $\widetilde{y}_p(n)$ was more pronounced than the bias of the respective variance-parameter estimators corresponding to the sample mean $\overline{Y}_n$; see Table 4 of Alexopoulos et al. (2007). [← There may be some relevant discussion in GMC1A and GMC1B regarding the "Hammersley" ratios that we looked at.]

Among the three variance-parameter estimators corresponding to $\widetilde{y}_p(n)$, the NBQ estimator $m\widetilde{S}^2_{b,m}$ exhibited the lowest small-sample bias, while the batched-STS area estimator $\mathscr{A}^2_{b,m}$ exhibited the largest. Since the combined estimator $\widetilde{\mathscr{V}}_{b,m}$ is roughly the average of its constituents, its average bias tends to fall in the middle; see Figure 2. For example, when $p = 0.25$, all three estimators exhibited substantial positive bias for small batch sizes ($m \leq 2^{14}$): the average percent relative bias of the batched-STS area estimator decreased from an overwhelming 272.43% for $m = 2^{10}$ to under 1% at approximately $m = 2^{17}$; the relative bias of the NBQ estimator dipped from roughly 40.83% at $m = 2^{10}$ to below 1% at $m = 2^{15}$; and the relative bias of the combined estimator dropped from roughly 158.47% at $m = 2^{10}$ to under 1% near $m = 2^{17}$.

When $p = 0.75$ all three variance-parameter estimators exhibited bias with nearly similar behavior. In particular, the average relative bias of the batched-STS area estimator decreased slowly from 47.12% above the asymptotic variance parameter for $m = 2^{10}$ to about 0.19% below for $m = 2^{20}$. When $p = 0.99$, the variance-parameter estimators approached their limit [← Maybe "limits" (plural since various estimators)? Also, do you mean the *expected values* of the variance estimators?] more slowly, with a relative bias that started at nearly 86% below the asymptotic variance parameter for $m = 2^{10}$, became positive near $m = 2^{15}$, and then dropped slowly as m increased.

Table 2 Experimental results for a stationary waiting-time process in an M/M/1 queueing system with traffic intensity $\rho = 0.8$. All estimates are based on 2,500 independent replications with $b = 32$ batches and batch sizes $m = 2^{\mathscr{L}}$, $\mathscr{L} = 10, 11, \ldots, 20$, where for nominal 95% CI estimators of y_p, the average CI half-length and coverage probability are denoted by "95% CI $\overline{H}$" and "95% CI Cover.", respectively.

		STS Area Estimator					NBQ Estimator					Combined Estimator				
p (y_p) Var. Par.	$\mathscr{L}$	Avg.	Bias	Std. Dev.	95% CI $\overline{H}$	95% CI Cover.	Avg.	Bias	Std. Dev.	95% CI $\overline{H}$	95% CI Cover.	Avg.	Bias	Std. Dev.	95% CI $\overline{H}$	95% CI Cover.
0.25	10	357.2	261.3	522.4	0.1923	99.24	135.1	39.2	92.6	0.1266	96.04	247.9	152.0	278.4	0.1626	98.48
(0.3227)	11	192.1	96.1	175.4	0.1047	97.52	113.2	17.3	44.0	0.0834	96.00	153.3	57.3	96.4	0.0939	97.56
95.9	12	127.2	31.3	57.0	0.0622	96.84	104.7	8.8	31.4	0.0570	96.12	116.1	20.2	34.8	0.0589	96.60
	13	108.8	12.9	34.5	0.0410	95.96	100.2	4.3	28.0	0.0395	95.84	104.6	8.7	22.8	0.0397	96.28
	14	102.4	6.5	28.4	0.0282	95.88	97.4	1.5	25.7	0.0276	94.96	100.0	4.0	19.5	0.0275	95.60
	15	98.6	2.6	26.6	0.0196	94.64	96.8	0.8	25.3	0.0194	95.04	97.7	1.8	18.6	0.0192	94.88
	16	97.5	1.6	25.3	0.0138	95.00	96.3	0.4	24.2	0.0137	95.04	96.9	1.0	17.4	0.0135	95.12
	17	96.7	0.8	24.4	0.0097	94.72	96.3	0.4	23.9	0.0097	94.84	96.5	0.6	17.0	0.0096	94.96
	18	96.4	0.5	24.4	0.0069	93.84	95.4	−0.5	23.8	0.0068	94.16	95.9	0.0	17.0	0.0067	93.80
	19	96.5	0.6	24.6	0.0048	94.76	95.3	−0.7	24.6	0.0048	95.12	95.9	0.0	17.3	0.0048	94.76
	20	95.5	−0.4	23.6	0.0034	94.84	95.9	0.0	24.5	0.0034	94.96	95.7	−0.2	16.9	0.0034	94.88
0.75	10	4853.0	1554.3	3419.9	0.7503	95.92	4798.4	1499.7	3211.7	0.7495	96.12	4826.1	1527.4	2831.1	0.7425	96.52
(5.8158)	11	4992.9	1694.2	3657.9	0.5402	96.56	4113.1	814.4	2162.2	0.4981	95.80	4560.0	1261.3	2449.3	0.5143	96.60
3298.7	12	4242.5	943.8	2046.1	0.3583	96.16	3703.1	404.4	1329.6	0.3379	95.96	3977.1	678.4	1361.7	0.3438	96.20
	13	3819.2	520.5	1402.5	0.2423	96.32	3466.6	167.9	1036.7	0.2320	95.68	3645.7	347.0	944.1	0.2338	96.20
	14	3547.5	248.8	1045.6	0.1658	95.36	3366.1	67.4	905.0	0.1620	95.16	3458.3	159.6	726.8	0.1614	95.24
	15	3412.5	113.8	936.5	0.1152	94.64	3345.8	47.1	878.2	0.1142	94.72	3379.7	81.0	652.7	0.1129	94.76
	16	3356.4	57.7	873.3	0.0808	94.60	3337.2	38.5	861.3	0.0807	94.80	3347.0	48.3	617.7	0.0795	94.64
	17	3332.1	33.4	859.7	0.0569	94.48	3327.3	28.6	839.2	0.0570	94.68	3329.7	31.0	605.0	0.0561	94.48
	18	3316.1	17.4	814.8	0.0402	94.60	3312.8	14.1	829.5	0.0402	94.76	3314.5	15.8	578.2	0.0396	94.60
	19	3310.2	11.5	838.5	0.0284	94.36	3306.4	7.7	856.2	0.0284	94.68	3308.3	9.6	593.9	0.0279	94.88
	20	3292.4	−6.3	813.3	0.0200	94.64	3316.4	17.7	853.0	0.0201	95.04	3304.2	5.5	580.6	0.0198	94.76
0.99	10	27618.0	−163642.9	17700.9	1.7889	54.88	53584.8	−137676.1	30487.4	2.5205	71.00	40395.4	−150865.5	20481.1	2.1576	62.72
(21.9101)	11	54706.7	−136554.2	37687.2	1.7710	67.96	80128.0	−111132.9	37863.4	2.2007	80.12	67215.6	−124045.3	31916.3	1.9742	74.88
191260.9	12	92768.8	−98492.1	66686.8	1.6278	79.08	123087.5	−68173.4	56786.8	1.9309	89.04	107687.5	−83573.4	52972.3	1.7657	85.16
	13	135781.4	−55479.5	93622.5	1.3998	87.72	179439.6	−11821.3	87474.8	1.6448	93.52	157264.0	−33996.9	76578.9	1.5096	91.16
	14	179612.2	−11648.7	128351.8	1.1440	91.20	218074.5	26813.6	117142.2	1.2789	94.56	198538.1	7277.2	106052.3	1.1979	93.00
	15	204721.9	13461.0	110567.3	0.8759	94.40	213376.8	22115.9	109485.9	0.9002	94.88	208980.7	17719.8	91853.7	0.8764	94.92
	16	209708.5	18447.6	106715.1	0.6301	95.44	202610.7	11349.8	77118.1	0.6250	95.84	206215.9	14955.0	75154.4	0.6190	95.80
	17	203575.5	12314.6	70787.1	0.4429	95.32	195545.3	4284.4	55570.2	0.4361	95.00	199624.1	8363.2	48901.2	0.4329	95.00
	18	199606.7	8345.8	57125.7	0.3111	95.24	193632.8	2371.9	53765.4	0.3070	94.84	196667.2	5406.3	41549.9	0.3043	95.00
	19	196112.6	4851.7	52084.9	0.2183	95.52	192647.7	1386.8	51991.8	0.2166	95.20	194407.7	3146.8	38005.1	0.2141	95.04
	20	193492.2	2231.3	49780.2	0.1534	95.52	193054.5	1793.6	48724.2	0.1535	95.16	193276.8	2015.9	35608.7	0.1510	95.16

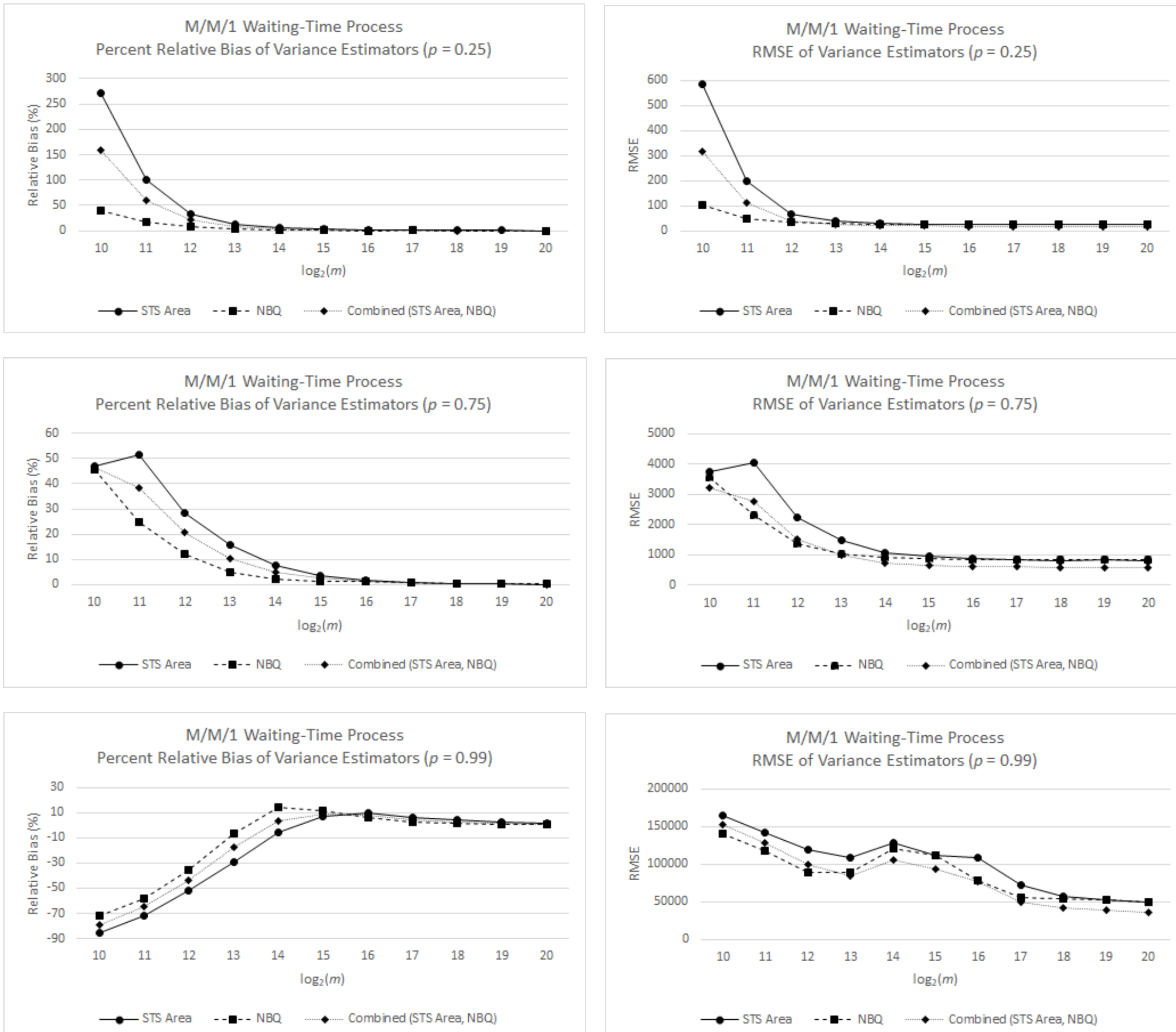

Figure 2 Estimated percent relative bias and RMSE of the variance-parameter estimators for selected marginal quantiles of a stationary waiting-time process in an M/M/1 queueing system with traffic intensity $\rho = 0.8$. All estimates are based on 2500 independent replications with $b = 32$ batches and batch sizes $m = 2^{\mathscr{L}}$, $\mathscr{L} = 10, 11, \ldots, 20$.

Notice that for $m = 2^{20}$ (total sample size $n = 2^{25} \approx 33$ million), the average relative bias of the batched-STS area estimator was 1.17%, while the average relative bias of the NBQ estimator was a bit lower (0.94%) and the average relative bias of the combined estimator was about 1.05%. Overall, the behavior of the bias of the three estimators exhibited no clear patterns as the batch size increased. Detailed analysis of the bias is a hard problem and a subject of ongoing research.

At this juncture, we would like to caution the reader that for this output process and $p = 0.99$, the Sequest procedure of Alexopoulos et al. (2019b), which is based on the NBQ estimator $m\widetilde{S}^2_{b,m}$ defined by Equation (105), often delivered CIs that exhibited significant undercoverage while requiring excessive sample sizes. This discovery motivated the development of the Sequem procedure (Alexopoulos et al. 2017) for the more-challenging problem of estimating near-extreme quantiles.

We now turn to the remaining statistics in Table 2. The standard deviation of each variance-parameter estimator converged to its respective theoretical limit. In particular, the standard deviation of the batched-STS area estimator (column 5) converged to $(2\sigma^4_{\tilde{y}_p}/b)^{1/2}$, based on [← "which is what we would expect from" instead of "based on"?] Equation (53). For instance, when $p = 0.99$ and $m = 2^{20}$, the average standard deviation of 49780.2 was only 4.11% larger than the theoretical limit $\sigma^2_{\tilde{y}_p}/4 = 47815.2$. In comparison, the average standard deviation 35608.7 of the combined estimator was only 4.49% larger than the theoretical limit $\left[2\sigma^4_{\tilde{y}_p}/(2b-1)\right]^{1/2} = 34077.8$. The dominance of the combined estimator with respect to its variance, and hence its mean squared error (MSE), was evident from the plots of the estimated RMSEs in Figure 2, in particular once the variance-parameter estimates approached the value $\sigma^2_{\tilde{y}_p}$.

The estimated coverage probabilities of the CIs obtained from Equations (92), (103), and (104) echoed the respective small-batch-size issues. When $p = 0.25$ or 0.75, the estimated coverage probability of the approximate 95% CIs was near the nominal level for all batch sizes; this is due to the convergence of the variance-parameter estimators to $\sigma^2_{\tilde{y}_p}$ from above. Unfortunately, this was not the case for $p = 0.99$, when the approximate 95% CIs exhibited substantial undercoverage for moderate sample sizes; indeed, the estimated coverage probabilities started approaching 0.95 only as $m \geq 2^{15}$. Overall, all three variance-parameter estimators appeared to be equally competitive when $p = 0.75$, while the conventional NBQ estimator appeared to dominate with regard to CI estimated coverage probability when $p = 0.99$ and $m \leq 2^{14}$ followed by the combined estimator and the batched-STS area estimator. As we stated earlier, such batch sizes are grossly inadequate for estimating such extreme quantiles.

8. Conclusions and Directions for Future Work

In this article we laid out an approach to formulating asymptotically exact CIs for a user-selected marginal quantile y_p, where $p \in (0, 1)$, of a stationary simulation output process using the STS method of output analysis. The core of our development was constructing an estimator of the associated variance parameter $\sigma^2_{\tilde{y}_p}$ based on STS quantile-estimation processes computed within each of a user-selected number b of nonoverlapping batches of size m. The key assumptions underlying our approach are the GMC Condition specified by Equation (2) for the output process and the FCLT specified by Equation (**??**) for the related indicator process. In contrast to output-analysis approaches requiring the assumption of strong-mixing conditions or other restrictive conditions that are difficult to verify, in our approach both the GMC and FCLT assumptions are at least empirically verifiable. Moreover, the GMC Condition has been theoretically verified for many widely used stationary time-series models as well as the queue-waiting-time processes generated by G/G/1 queueing systems in steady-state operation.

In the first step of this work, we established the following weak-convergence results as the batch size $m \to \infty$: (i) the batched-STS area estimator $\mathscr{A}^2_{b,m}$ converges weakly to the r.v. $\sigma^2_{\tilde{y}_p}\chi^2_b \big/ b$; (ii) the conventional batching-based estimator $mS^2_{b,m}$ converges weakly to the r.v. $\sigma^2_{\tilde{y}_p}\chi^2_{b-1} \big/ (b-1)$; (iii) the NBQ estimator

$m\widetilde{S}^2_{b,m}$ converges weakly to the r.v. $\sigma^2_{\widetilde{y}_p}\chi^2_{b-1}\big/(b-1)$; and (iv) the combined estimators $\mathscr{V}_{b,m}$ and $\widetilde{\mathscr{V}}_{b,m}$ respectively defined by Equations (55) and (105) converge weakly to the r.v. $\sigma^2_{\widetilde{y}_p}\chi^2_{2b-1}\big/(2b-1)$. Establishing the asymptotic distribution of $m\widetilde{S}^2_{b,m}$ filled a theoretical gap in our previous work (Alexopoulos et al. 2019b) on estimation of nonextreme steady-state quantiles, where $p \in [0.05, 0.95]$.

In the second step of this work, we showed the asymptotic exactness of the CIs for y_p that are centered at the full-sample point estimator $\widetilde{y}_p(n)$ with half-length based on each of the aforementioned variance-parameter estimators and the associated critical value of the t-distribution with the appropriate d.o.f. Naturally, the combined estimator $\widetilde{\mathscr{V}}_{b,m}$ dominated its constituent estimators with regard to the MSE incurred in estimating $\sigma^2_{\widetilde{y}_p}$; and the combined estimators yielded CIs for y_p with smaller and less-variable half-lengths for the same batch size and batch count.

In the third and final step of this work, a preliminary experimental performance evaluation based on two widely used test processes satisfying the GMC and FCLT assumptions illustrated the theoretical properties of the variance-parameter estimators $\mathscr{A}^2_{b,m}$, $m\widetilde{S}^2_{b,m}$, and $\widetilde{\mathscr{V}}_{b,m}$. We found that the batched-STS area estimator $\mathscr{A}^2_{b,m}$ based on the constant weight function often had significantly larger absolute bias than the NBQ estimator $m\widetilde{S}^2_{b,m}$ for small batch sizes and near-extreme quantiles, where $p \in (0.0005, 0.05) \cup (0.95, 0.9995)$. Such small-sample bias was also pronounced in the batched-STS variance-parameter estimators used to construct CIs for the steady-state mean (Alexopoulos et al. 2007). However, small batch sizes are often grossly inadequate for constructing valid CIs for steady-state quantiles (Alexopoulos et al. 2019b). As dictated by theory, the combined estimator $\widetilde{\mathscr{V}}_{b,m}$ typically had bias that fell between the biases of its constituents $\mathscr{A}^2_{b,m}$ and $m\widetilde{S}^2_{b,m}$; thus its lower asymptotic variance made it the clear winner among the three estimators with regard to statistical efficiency.

Our ongoing work includes the following threads: (i) provide details on the small-sample properties of batched-STS area estimators; (ii) study batched STS area estimators that use alternative weight functions tailored to quantile estimation; (iii) study the effects of the batch size and the batch count on the MSEs of our variance-parameter estimators; (iv) consider other batched-STS variance-parameter estimators besides the area estimator, e.g., the CvM estimator; (v) carry out a comprehensive experimental study designed to stress-test our procedures so as to better understand the performance of those procedures in practice; and (vi) study overlapping versions of the area and CvM estimators in the context of quantile estimation. Our ultimate goal is to develop automated, asymptotically valid sequential procedures to deliver CIs for user-specified steady-state quantiles that satisfy absolute- or relative-precision requirements and exhibit improved performance when compared with the recent BQE-based procedures by Alexopoulos et al. (2017, 2019b).

Acknowledgments

The authors gratefully acknowledge Joseph Haden Boone for his able programming assistance.

References

Aktaran-Kalaycı T, Alexopoulos C, Argon NT, Goldsman D, Wilson JR (2007) Exact expected values of variance estimators in steady-state simulation. *Naval Research Logistics* 54(4):397–410.

Alexopoulos C, Argon NT, Goldsman D, Tokol G, Wilson JR (2007) Overlapping variance estimators for simulation. *Operations Research* 55(6):1090–1103.

Alexopoulos C, Boone JH, Goldsman D, Lolos A, Dingeç KD, Wilson JR (2020) Steady-state quantile estimation using standardized time series. Bae KH, Feng B, Kim S, Lazarova-Molnar L, Zheng Z, Roeder T, Thiesing R, eds., *Proceedings of the 2020 Winter Simulation Conference*, 289–300 (Piscataway, New Jersey: Institute of Electrical and Electronics Engineers).

Alexopoulos C, Goldsman D, Mokashi AC, Tien KW, Wilson JR (2019a) Online availability of the Sequest software for Linux, MacOS, and Windows. URL https://people.engr.ncsu.edu/jwilson/files/sequest-availability.pdf, accessed April 30, 2021.

Alexopoulos C, Goldsman D, Mokashi AC, Tien KW, Wilson JR (2019b) Sequest: A sequential procedure for estimating quantiles in steady-state simulations. *Operations Research* 67(4):1162–1183.

Alexopoulos C, Goldsman D, Mokashi AC, Wilson JR (2017) Automated estimation of extreme steady-state quantiles via the maximum transformation. *ACM Transactions on Modeling and Computer Simulation* 27(4):22:1–22:29.

Alexopoulos C, Goldsman D, Mokashi AC, Wilson JR (2019c) Sequential estimation of steady-state quantiles: Some new developments in methods and software. Mustafee N, Bae KHG, Lazarova-Molnar S, Rabe M, Szabo C, Haas P, Son YJ, eds., *Proceedings of the 2019 Winter Simulation Conference*, 3774–3785 (Piscataway, New Jersey: Institute of Electrical and Electronics Engineers).

Alexopoulos C, Goldsman D, Tang P, Wilson JR (2016) SPSTS: A sequential procedure for estimating the steady-state mean using standardized time series. *IIE Transactions* 48(9):864–880.

Alexopoulos C, Goldsman D, Wilson JR (2012) A new perspective on batched quantile estimation. Laroque C, Himmelspach J, Pasupathy R, Rose O, Uhrmacher AM, eds., *Proceedings of the 2012 Winter Simulation Conference*, 190–200 (Piscataway, New Jersey: Institute of Electrical and Electronics Engineers).

Asmussen S, Glynn PW (2007) *Stochastic Simulation: Algorithms and Analysis* (New York: Springer Science+Business Media).

Azzalini A (2020) *The R package 'sn': The skew-normal and related distributions such as the skew-t (version 1.6–2).* University of Padua, Padua, Italy, URL https://cran.r-project.org/web/packages/sn/, accessed April 30, 2021.

Beran J (1994) *Statistics for Long-Memory Processes* (Boca Raton, Florida: Chapman & Hall/CRC).

Bernardo JM, Smith AFM (2000) *Bayesian Theory* (Chichester, England: John Wiley & Sons, Ltd).

Bickel PJ, Doksum KA (2007) *Mathematical Statistics: Basic Ideas and Selected Topics* (Upper Saddle River, NJ: Pearson Prentice Hall), 2nd edition.

Billingsley P (1968) *Convergence of Probability Measures* (New York: John Wiley & Sons).

Billingsley P (1999) *Convergence of Probability Measures* (New York: John Wiley & Sons), 2nd edition.

Blomqvist N (1967) The covariance function of the M/G/1 queuing system. *Skandinavisk Aktuarietidskrift* 50:157–174.

Box GE, Jenkins GM, Reinsel GC (2008) *Time Series Analysis* (New York: John Wiley & Sons), 4th edition.

Bradley RC (2005) Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* 2:107–144.

Calvin JM, Nakayama MK (2013) Confidence intervals for quantiles with standardized time series. Pasupathy R, Kim SH, A Tolk RH, Kuhl ME, eds., *Proceedings of the 2013 Winter Simlation Conference*, 601–612 (Piscataway, New Jersey: Institute of Electrical and Electronics Engineers).

Cormen TH, Leiserson CE, Rivest RL, Stein C (2009) *Introduction to Algorithms* (Cambridge, Massachusetts: MIT Press), 3rd edition.

Damerdji H (1994) Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research* 19(2):494–512.

Dingeç KD, Alexopoulos C, Goldsman D, Lolos A, Wilson JR (2022) Geometric moment-contraction of G/G/1 waiting times. Botev Z, Keller A, Lemieux C, Tuffin B, eds., *Advances in Modeling and Simulation: Festschrift for Pierre L'Ecuyer*, to appear (Springer), URL https://people.engr.ncsu.edu/jwilson/files/gg1-gmc.pdf.

Dong J, Glynn PW (2019) The Asymptotic Validity of Sequential Stopping Rules for Confidence Interval Construction Using Standardized Time Series. Mustafee N, Bae KHG, Lazarova-Molnar S, Rabe M, Szabo C, Haas P, Son YJ, eds., *Proceedings of the 2019 Winter Simulation Conference*, 332–343 (Piscataway, New Jersey: Institute of Electrical and Electronic Engineers).

Fishman GS (1978) *Principles of Discrete Event Simulation* (New York: John Wiley & Sons).

Foley RD, Goldsman D (1999) Confidence intervals using orthonormally weighted standardized time series. *ACM Transactions on Modeling and Simulation* 9:297–325.

Freedman D, Diaconis P (1981) On the histogram as a density estimator: L_2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57:452–476.

Glynn PW, Iglehart DL (1990) Simulation output analysis using standardized time series. *Mathematics of Operations Research* 15:1–16.

Goldsman D, Meketon MS, Schruben LW (1990) Properties of standardized time series weighted area variance estimators. *Management Science* 36:602–612.

Hahn GJ, Shapiro SS (1967) *Statistical Models in Engineering* (New York: John Wiley & Sons).

Karlin S, Taylor HM (1975) *A First Course in Stochastic Processes* (New York: Academic Press), 2nd edition.

Kolmogorov AN, Fomin SV (1975) *Introductory Real Analysis* (New York: Dover Publications Inc.).

Law AM (2015) *Simulation Modeling and Analysis* (New York: McGraw-Hill), 5th edition.

L'Ecuyer P, Simard R, Chen EJ, Kelton WD (2002) An object-oriented random number package with many long streams and substeams. *Operations Research* 90(6):1073–1075.

McIlroy P (1993) Optimistic sorting and information theoretic complexity. *Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, 467–474.

Meyer C (2013) The bivariate normal copula. *Communications in Statistics-Theory and Methods* 42(13):2402–2422.

Priestley MB (1981) *Spectral Analysis and Time Series* (London: Academic Press).

Resnick SI (2002) *Adventures in Stochastic Processes* (Boston: Birkhäuser).

Rudin W (1976) *Principles of Mathematical Analysis* (New York: McGraw-Hill), 3rd edition.

Schruben LW (1983) Confidence interval estimation using standardized time series. *Operations Research* 31:1090–1108.

Scott DW (1992) *Multivariate Density Estimation: Theory, Practice, and Visualization* (New York: Wiley-Interscience).

Whitt W (2002) *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues* (New York: Springer).

Wilks SS (1962) *Mathematical Statistics* (New York: John Wiley & Sons).

Wu WB (2005) On the Bahadur representation of sample quantiles for dependent sequences. *The Annals of Statistics* 33(4):1934–1963.

Wu WB (2009) Recursive estimation of time-average variance constants. *The Annals of Applied Probability* 19(4):1529–1552.

Yang WZ, Hu SH, Wang XJ (2019) The Bahadur representation for sample quantiles under dependent sequence. *Acta Mathematicae Applicatae Sinica* 35(3):521–53.