

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Overlapping Batch Confidence Intervals on Statistical Functionals Constructed from Time Series: Application to Quantiles, Optimization, and Estimation

ZIWEI SU, Department of Statistics, Purdue University, West Lafayette, IN; Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA

RAGHU PASUPATHY, Department of Statistics, Purdue University, West Lafayette, IN, USA; Department of Computer Science & Engineering, IIT Madras, Chennai, India

YINGCHIEH YEH, Institute of Industrial Management, National Central University, Taiwan

PETER W. GLYNN, Department of Management Science and Engineering, Stanford University, USA

We propose a general purpose confidence interval procedure (CIP) for statistical functionals constructed using data from a stationary time series. The procedures we propose are based on derived distribution-free analogues of the χ^2 and Student's t random variables for the statistical functional context, and hence apply in a wide variety of settings including quantile estimation, gradient estimation, M-estimation, CVAR-estimation, and arrival process rate estimation, apart from more traditional statistical settings. Like the method of subsampling, we use overlapping batches of time series data to estimate the underlying variance parameter; unlike subsampling and the bootstrap, however, we assume that the implied point estimator of the statistical functional obeys a central limit theorem (CLT) to help identify the weak asymptotics (called OB-x limits, $x=I,II,III$) of batched Studentized statistics. The OB-x limits, certain functionals of the Wiener process parameterized by the size of the batches and the extent of their overlap, form the essential machinery for characterizing dependence, and consequently the correctness of the proposed CIPs. The message from extensive numerical experimentation is that in settings where a functional CLT on the point estimator is in effect, using *large overlapping batches* alongside OB-x critical values yields confidence intervals that are often of significantly higher quality than those obtained from more generic methods like subsampling or the bootstrap. We illustrate using examples from CVaR estimation, ARMA parameter estimation, and NHPP rate estimation; R and MATLAB code for OB-x critical values is available at web.ics.purdue.edu/~pasupath.

CCS Concepts: • ;

Additional Key Words and Phrases: to be filled

ACM Reference Format:

Ziwei Su, Raghu Pasupathy, Yingchieh Yeh, and Peter W. Glynn. 2018. Overlapping Batch Confidence Intervals on Statistical Functionals Constructed from Time Series: Application to Quantiles, Optimization, and Estimation. 1, 1 (May 2018), 40 pages. <https://doi.org/XXXXXXX.XXXXXXX>

Authors' addresses: Ziwei Su, su230@purdue.edu, Department of Statistics, Purdue University, West Lafayette, IN; Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, 150 North University Street, West Lafayette, Indiana, USA, 47907; Raghu Pasupathy, pasupath@purdue.edu, Department of Statistics, Purdue University, West Lafayette, IN, USA; Department of Computer Science & Engineering, IIT Madras, Chennai, 150 North University Street, West Lafayette, Indiana, India, 47907; Yingchieh Yeh, Institute of Industrial Management, National Central University, Taoyuan, Taiwan, yeh@mgt.ncu.edu.tw; Peter W. Glynn, Department of Management Science and Engineering, Stanford University, Palo Alto, CA, USA, glynn@stanford.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Association for Computing Machinery.

XXXX-XXXX/2018/5-ART \$15.00

<https://doi.org/XXXXXXX.XXXXXXX>

1 INTRODUCTION

Let $\{X_j, j \geq 1\}$ be an \mathcal{S} -valued discrete-time, stationary, observable stochastic process defined on some probability space (Ω, \mathcal{F}, P) , and let $\theta : \mathcal{W} \rightarrow \Theta \subseteq \mathbb{R}$ denote a known *statistical functional* (see [71, Chapter 6] or [55, Section 6.2]) defined on some space \mathcal{W} of probability measures. In this paper, we propose an overlapping batch (OB) confidence interval procedure (CIP) to construct an interval $I_n \subset \mathbb{R}$ using $X_j, j = 1, 2, \dots, n$ such that

$$\lim_{n \rightarrow \infty} P(\omega : \theta(P) \in I_n(\omega)) = 1 - \alpha, \quad (1)$$

for any specified constant $\alpha \in (0, 1)$. Importantly, since $\{X_n, n \geq 1\}$ is a time series, the dependence between random variables $X_n, n \geq 1$ is a key feature requiring careful treatment.

REMARK. *The initial segment $X_j, 1 \leq j \leq n$ of the observable process $\{X_j, j \geq 1\}$ is assumed to be a “collected dataset” or the output of a simulation that is exogenous to problem at hand. We assume no facility for variance reduction, e.g., by changing the measure governing the process $\{X_j, j \geq 1\}$, as is sometimes possible in simulation settings. See [13, 25, 26, 49, 61, 62] for variance reduced confidence interval problems in the quantile context.*

1.1 Motivation

Statistical functionals subsume a variety of interesting quantities arising in modern data settings, and are thus useful mathematical objects on which to construct confidence intervals. Consider, for instance, the following examples of statistical functionals. In each case, let $X : \Omega \rightarrow \mathcal{S}$ be an \mathcal{S} -valued random variable distributed according to P , and let $s \in \mathcal{S}$ denote an “outcome” in \mathcal{S} .

(a) Expectation. For $g : \mathcal{S} \rightarrow \mathbb{R}$, define the expectation

$$\theta(P) := \mathbb{E}[g(X)] = \int_{\mathcal{S}} g(s) dP.$$

(b) Quantile. For $g : \mathcal{S} \rightarrow \mathbb{R}$ and $\gamma \in (0, 1)$, define the $(1 - \gamma)$ -quantile

$$\theta(P) = \theta_\gamma(P) := \inf_{y \in \mathbb{R}} \{P(g(X) \leq y) \geq 1 - \gamma\}.$$

(c) Finite Difference Approximation. For $g : \mathcal{X} \times \mathcal{S} \rightarrow \mathbb{R}$ where $\mathcal{X} := \text{dom}(g) \subseteq \mathbb{R}^d$, define the finite-difference approximation of the directional derivative (assumed to exist) at $x \in \text{int}(\mathcal{X})$ along $u \in \mathbb{R}^d$:

$$\theta(P) = \theta_{x,u,\epsilon}(P) := \frac{1}{\epsilon} (\mathbb{E}[g(x + \epsilon u, X)] - g(x, X)) = \frac{1}{\epsilon} \left(\int_{\mathcal{S}} (g(x + \epsilon u, s) - g(x, s)) dP \right).$$

(d) General Optimization. For $g : \mathcal{X} \times \mathcal{S} \rightarrow \mathbb{R}$, where (\mathcal{X}, d) is a metric space,

$$\theta(P) := \inf_{x \in \mathcal{X}} \mathbb{E}[g(x, X)] = \inf_{x \in \mathcal{X}} \int_{\mathcal{S}} g(x, s) dP.$$

(e) Root Finding. For $g : \mathcal{X} \times \mathcal{S} \rightarrow \mathbb{R}$, $\theta(P) = \theta \in \mathcal{X}$ is such that

$$\int_{\mathcal{S}} g(\theta, s) dP = 0.$$

(f) Conditional Value at Risk (CVaR). For $0 < \gamma < 1$, and $g : \mathcal{S} \rightarrow \mathbb{R}$,

$$\theta(P) = \theta_\gamma(P) := \mathbb{E}[g(S) | g(S) > q_\gamma] = \frac{1}{P(g(S) > q_\gamma)} \left(\int_{\mathcal{S}} g(s) \mathbb{I}\{g(s) > q_\gamma\} dP \right),$$

where the γ -quantile $q_\gamma := \inf \left\{ t \in \mathbb{R} : \int_{\mathcal{S}} \mathbb{I}\{g(s) \leq t\} dP \geq \gamma \right\}$.

(g) ARMA(p, q). The ARMA(p, q) process is a discrete-time real-valued process $\{Y_t, t \geq 1\}$ having p “autoregressive” parameters, $\phi_j, j = 1, 2, \dots, p$, and q “moving average” parameters, $\theta_j, j = 1, 2, \dots, q$, and is expressed as

$$Y_t = c + \sum_{j=1}^p \phi_j Y_{t-j} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t,$$

where $\{\epsilon_j, j \geq 1\}$ are independent and identically distributed (iid) random variables having mean zero and unit variance. Given observations $y_i, i = 1, 2, \dots, n$ of the process $\{Y_t, t \geq 1\}$, the estimators $\hat{c}, \hat{\phi}_j, j = 1, 2, \dots, p$ and $\hat{\theta}_j, j = 1, 2, \dots, q$ of the parameters $c, \phi_j, j = 1, 2, \dots, p$ and $\theta_j, j = 1, 2, \dots, q$, are statistical functionals that can be estimated by minimizing the sum of squared residuals [16]:

$$\underset{c, \{\phi_j, 1 \leq j \leq p\}, \{\theta_j, 1 \leq j \leq q\}}{\text{minimize:}} \sum_{i=p+1}^n \hat{\epsilon}_i^2,$$

where the residuals are given by:

$$\hat{\epsilon}_i := \begin{cases} y_i - c - \sum_{j=1}^p \phi_j y_{i-j} - \sum_{j=1}^q \theta_j \hat{\epsilon}_{i-j} & \text{if } i \geq p + 1, \\ 0 & \text{otherwise.} \end{cases}$$

In addition to the above examples, a wide variety of quantities arising within classical statistics, e.g., higher order moments, ratio of moments, k -means clustering, α -trimmed mean, Mann-Whitney functional, and the simplicial depth functional are all statistical functionals, making the question of constructing confidence intervals on statistical functionals of wide interest. (See [74, Chapter 7] and [55, Chapter 6] for other examples and a full treatment of statistical functionals.)

REMARK. Whereas $\theta(P)$ in some of the examples listed above are naturally \mathbb{R}^d -valued with $d > 1$, e.g., (g), the treatment in this paper is entirely real-valued, that is, $\theta(P) \in \mathbb{R}$. Extending our methods from \mathbb{R} to \mathbb{R}^d is straightforward but further extension into a function space involves non-trivial technical aspects.

1.2 Notation and Terminology

(i) \mathbb{N} refers to the set $\{1, 2, \dots\}$ of natural numbers. (ii) $\mathbb{I}_A(x)$ is the indicator variable taking the value 1 if $x \in A$ and 0 otherwise. Also, depending on the context, we write $\mathbb{I}(A)$ where $\mathbb{I}(A) = 1$ if the event A is true and 0 otherwise. (iii) $Z(0, 1)$ denotes the standard normal random variable, and χ_ν^2 refers to the chi-square random variable with ν degrees of freedom. (iv) $\{W(t), t \geq 0\}$ refers to the standard Wiener process [5, Section 37], and $\{B(t), t \in [0, T]\}, B(t) = W(t) - \frac{t}{T}W(1)$ refers to the Brownian bridge on $[0, T]$. (v) For a random sequence $\{X_n, n \geq 1\}$, we write $X_n \xrightarrow{wp1} X$ to refer to almost sure convergence, $X_n \xrightarrow{p} X$ to refer to convergence in probability, and $X_n \xrightarrow{d} X$ to refer to convergence in distribution (or weak convergence). (vi) We write $X \stackrel{d}{=} Y$ to mean that random variables X and Y have the same distribution. (vii) The empirical measure P_n constructed from the sequence $\{X_n, n \geq 1\}$ is given by $P_n(A) = n^{-1} \sum_{j=1}^n \mathbb{I}_A(X_j)$ for appropriate sets A .

1.3 Organization of the Paper

In the following section, we discuss literature on confidence intervals with a view toward providing perspective on how the proposed methods fit within the existing literature. This is followed by Section 3 where we present the main idea underlying the interval estimators we propose, along

with a synopsis of results. Section 4 includes key assumptions, followed by Section 5–7 which present the theorems corresponding to the OB-I, OB-II, and OB-III limits. In Section 8, we present brief discussion on questions that seem to arise frequently. We end with Section 9 where numerical illustration using three different contexts illustrate the effectiveness of using large batch OB-I and OB-II confidence intervals.

2 EXISTING LITERATURE, PERSPECTIVE, AND CONTRIBUTION

In this section, we provide an overview of CIPs in general. While we provide a large number of references, our intent is not an encyclopedic account but to provide perspective on how the proposed method fits within the literature.

2.1 CLT-based Methods

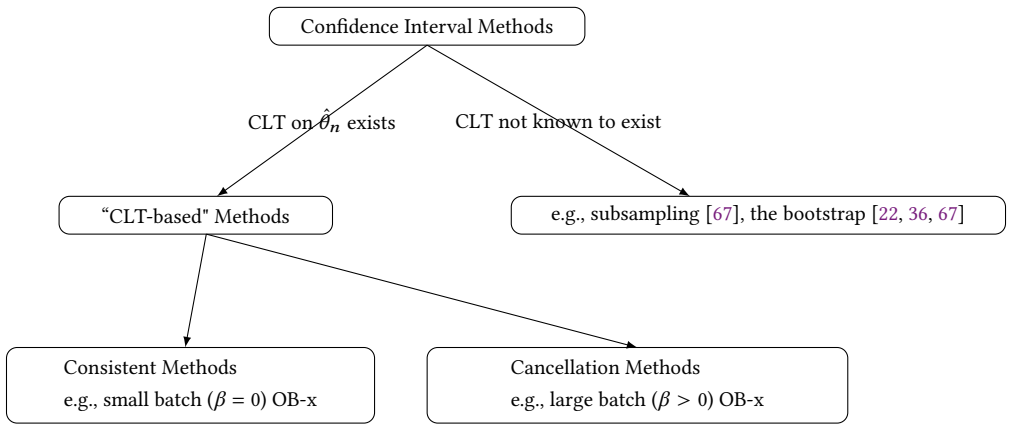


Fig. 1. A taxonomy of methods for constructing confidence intervals on statistical functionals. Consistent methods construct a consistent estimator of the variance constant σ , while cancellation methods allow the use of large batches and construct ratio estimators that “cancel out” the variance constant σ .

Analogous to the taxonomy [42] of CIPs on the steady-state mean of a real-valued process, it is instructive to categorize CIPs for statistical functionals based on whether a central limit theorem of the form

$$\sqrt{n}(\hat{\theta}_n - \theta(P)) \xrightarrow{d} \sigma N(0, 1) \quad (2)$$

exists. In (2), $\hat{\theta}_n$ is an implied point estimator of $\theta(P)$ constructed from the time series $\{X_n, n \geq 1\}$, $N(0, 1)$ is the standard normal random variable, and $\sigma \in (0, \infty)$ is an unknown parameter often called the variance constant. Further, and as depicted in Figure 1, a CIP that assumes (2) may either be a *consistent method* by which we mean that the CIP constructs another observable process $\{\hat{\sigma}_n, n \geq 1\}$ from $\{X_n, n \geq 1\}$ to consistently estimate σ , that is,

$$\hat{\sigma}_n \xrightarrow{p} \sigma \text{ as } n \rightarrow \infty; \quad (3)$$

or a *cancellation method* by which we mean that the CIP constructs a non-vanishing process $\{Y_n, n \geq 1\}$ such that

$$(\sqrt{n}(\hat{\theta}_n - \theta(P)), Y_n) \xrightarrow{d} (\sigma N(0, 1), \sigma Y) \text{ as } n \rightarrow \infty, \quad (4)$$

and Y is a well-defined random variable whose distribution can be computed, implying, among other things, that Y does not depend on the unknown quantities σ and $\theta(P)$.

In consistent methods, since (2) and (3) hold, Slutsky’s theorem (A.2) assures us that an asymptotically valid two-sided $(1 - \alpha)$ confidence interval on $\theta(P)$ is

$$\left(\hat{\theta}_n - z_{1-\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}}, \hat{\theta}_n + z_{1-\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}}\right),$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution. It is in this sense that a consistent method essentially reduces the confidence interval construction problem into the often nontrivial problem [3, 13, 43] of consistently estimating the variance parameter σ . Various consistent methods exist in the steady-state mean context. For example, the regenerative method [15, 53], the spectral procedure [4, 18–20, 73] with certain restrictions on the bandwidth, and the batch means procedure where the variance parameter is estimated using one of various well-established methods (e.g., NBM, OBM, weighted CvM or weighted area estimators) provided the batch size tends to infinity in a way that the batch size expressed as a fraction of the total data size tends to zero. See [1, 2] and references therein for a thorough account on estimating the variance parameter associated with a steady-state real-valued process.

In contrast to consistent methods, cancellation methods are based on the important idea that σ need not be estimated consistently to construct a valid confidence interval on $\theta(P)$. This seems to have been first observed in the seminal account [70] introducing standardized time series in the context of constructing confidence intervals on the steady state mean. Specifically, in cancellation methods, since (2) and (4) hold, and Y is non-vanishing, applying the continuous mapping theorem [6] leads to “cancellation” of σ in the sense that

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta(P))}{Y_n} \xrightarrow{d} \frac{\phi N(0, 1)}{\phi Y}, \tag{5}$$

leading to the two-sided $(1 - \alpha)$ confidence interval

$$\left(\hat{\theta}_n - y_{\alpha/2} \frac{Y_n}{\sqrt{n}}, \hat{\theta}_n + y_{1-\alpha/2} \frac{Y_n}{\sqrt{n}}\right),$$

where y_q is the q -quantile of $N(0, 1)/Y$. If constructing a consistent estimator of σ is the principal challenge in consistent methods, selecting Y_n and characterizing Y turns out to be the principal challenge in cancellation methods. Cancellation methods have been studied [11, 46, 48, 60, 70] in the context of constructing confidence intervals on the steady-state mean, and more recently for quantiles — see the exceptionally well-written articles [10, 27].

2.2 Subsampling

Subsampling and the bootstrap are examples of methods that are not CLT-based methods in the sense that they do not assume knowledge of the normal weak limit in (2), although they assume the existence of a weak limit. Subsampling is the culmination of decades of thought on using batches for confidence intervals, and was formalized in a 1992 paper by Politis and Romano [66] (henceforth abbreviated PR1992). See [67] for a book-length treatment that includes situations where $\theta(P)$ resides in a separable Banach space.

REMARK. *There is a long history of using batches within the classical statistics literature in the context of constructing a confidence interval from time series data, e.g., interpenetration samples by Mahalanobis [57], the jackknife by Quenouille [68], pseudoreplication by McCarthy [58], and subsampling by Hartigan [51, 52]. There is also a corresponding history in the simulation literature dating back to Conway [14], Mechanic and McKay [59], and Fishman [38] — precursors to the now mature methods to construct confidence intervals on the steady-state mean using batched simulation output.*

In the interest of easily conveying the essence of subsampling and the bootstrap, the ensuing discussion assumes use of the Studentized statistic $\hat{\sigma}_n^{-1}(\hat{\theta}_n - \theta(P))$, where $\hat{\theta}_n$ is the point estimator of $\theta(P)$, and $\hat{\sigma}_n^2$ is a point estimator of the variance parameter σ^2 . Suppose

$$\tau_n \frac{(\hat{\theta}_n - \theta(P))}{\hat{\sigma}_n} \xrightarrow{d} J(P); \quad (6)$$

and define

$$\begin{aligned} L_n(x) &:= \frac{1}{n - m_n + 1} \sum_{j=1}^{n-m_n+1} \mathbb{I} \left\{ \tau_{m_n} \frac{(\hat{\theta}_{j,m_n} - \hat{\theta}_n)}{\hat{\sigma}_{j,m_n}} \leq x \right\}, \quad x \in \mathbb{R} \\ &= \frac{1}{n - m_n + 1} \sum_{j=1}^{n-m_n+1} \mathbb{I} \left\{ \frac{\tau_{m_n}(\hat{\theta}_{m_n,j} - \theta(P))}{\hat{\sigma}_{m_n,j}} + \frac{\tau_{m_n}(\theta(P) - \hat{\theta}_n)}{\hat{\sigma}_{m_n,j}} \leq x \right\}, \end{aligned} \quad (7)$$

where $\hat{\theta}_n$ is a point estimator of $\theta(P)$ constructed from the entire data set (X_1, X_2, \dots, X_n) , θ_{j,m_n} is the estimator of $\theta(P)$ constructed from the j -th subsample $(X_{(j-1)m_n+1}, X_{(j-1)m_n+2}, \dots, X_{(j-1)m_n+m_n})$ (see Figure 2 with offset $d_n = 1$), $\{\tau_n, n \geq 1\}$ is a "scaling" sequence, and $\hat{\sigma}_n$ is an estimate of what is called the *scale* σ in PR1992, and what we call the *variance constant* in this paper. Also define

$$U_n(x) := \frac{1}{n - m_n + 1} \sum_{j=1}^{n-m_n+1} \mathbb{I} \left\{ \frac{\tau_{m_n}(\hat{\theta}_{m_n,j} - \theta(P))}{\hat{\sigma}_{m_n,j}} \leq x \right\}; \quad E_n := \left(\frac{\tau_{m_n}(\theta(P) - \hat{\theta}_n)}{\hat{\sigma}_{m_n,j}} \leq x \right). \quad (8)$$

Suppose that in addition to the assumption of the weak limit existence in (6), the following assumptions hold.

(A.1) the sequence $\{X_n, n \geq 1\}$ is stationary and strong-mixing (defined in Section 4);

(A.2) $\tau_{m_n}/\tau_n \rightarrow 0$, $m_n/n \rightarrow 0$; and

(A.3) the cdf $J_n(\cdot, P)$ of the Studentized statistic $\tau_n(\hat{\theta}_n - \theta(P))/\hat{\sigma}_n$ is continuous (in its first argument).

The crucial insight of subsampling is that the empirical cdf L_n can be used to approximate the sampling distribution of the Studentized statistic by replacing $\hat{\theta}_n$ and $\hat{\sigma}_n$ in (6) by their subsample counterparts $\hat{\theta}_{j,m_n}$, $\hat{\sigma}_{j,m_n}$, and by replacing $\theta(P)$ in (6) by $\hat{\theta}_n$. PR1992 formalize this by demonstrating the intuitive result

$$\sup_{x \in \mathbb{R}} |L_n(x) - J_n(x, P)| \xrightarrow{P} 0. \quad (9)$$

The assertion in (9) motivates constructing the following (two-sided) subsampling confidence interval $I_{n,\alpha}$ on $\theta(P)$, that can be shown to be asymptotically valid:

$$I_{n,\alpha} := \left(\hat{\theta}_n - c_{n,\alpha/2} \frac{\hat{\sigma}_n}{\tau_n}, \hat{\theta}_n + c_{n,\alpha/2} \frac{\hat{\sigma}_n}{\tau_n} \right); \quad c_{n,q} := \inf \{x : L_n(x) \geq q\}. \quad (10)$$

REMARK. Notice from (10) that subsampling assumes knowledge of the scaling τ_n but not the weak limit J . This will be true about the bootstrap as well.

We emphasize that PR1992's main instrument for establishing that $L_n(x) \xrightarrow{P} J(x, P)$ is the "small batch size" assumption in (A.2) above, which ensures that the probability of the event E_n in (8) tends to one, and the variance of $U_n(x)$ in (8) tends to zero.

2.3 The Bootstrap

In the service of precisely explaining the bootstrap [22, 29, 35, 50], let's enhance the notation introduced previously to view point estimators as functionals, that is, $\hat{\theta} : \mathcal{D}_n \rightarrow \mathbb{R}$ and $\hat{\sigma}^2 : \mathcal{D}_n \rightarrow \mathbb{R}$, where \mathcal{D}_n is the space of datasets of size n with S -valued observations. So, the point estimator $\hat{\theta}_n$ of the statistical functional $\theta(P)$ and the point estimator $\hat{\sigma}_n^2$ of the variance constant σ^2 constructed using the given dataset X_1, X_2, \dots, X_n are

$$\hat{\theta}_n \equiv \hat{\theta}(\{X_1, X_2, \dots, X_n\}); \quad \hat{\sigma}_n^2 \equiv \hat{\sigma}_n^2(\{X_1, X_2, \dots, X_n\}).$$

The bootstrap's central idea is a method for approximating the sampling distribution of the Studentized statistic $\sqrt{n} (\hat{\theta}_n - \theta(P)) / \hat{\sigma}_n$. (The bootstrap assumes that the weak limit in (6) holds with $\tau_n = \sqrt{n}$, although the assumption of the existence of a weak limit is often not stated explicitly.) And, whereas subsampling constructs the empirical cdf L_n in (7) using subsamples, the bootstrap accomplishes the objective of estimating the sampling distribution of $\sqrt{n} (\hat{\theta}_n - \theta(P)) / \hat{\sigma}_n$ through the following two steps:

- (1) *resample*, that is, use a "resampling measure" \hat{P}_n to generate B datasets $\{X_{j,i}^*, 1 \leq j \leq n\}, i = 1, 2, \dots, B$, e.g., iid draws with replacement from the original dataset; and
- (2) *compute*, that is, use the generated datasets $\{X_{j,i}^*, 1 \leq j \leq n\}, i = 1, 2, \dots, B$ to compute "bootstrap realizations" $\hat{\theta}(\{X_{j,i}^*, 1 \leq j \leq n\}), i = 1, 2, \dots, B$ of the point estimator and "bootstrap realizations"

$$S_{n,i} := \sqrt{n} \frac{\left(\hat{\theta}(\{X_{j,i}^*, 1 \leq j \leq n\}) - \hat{\theta}_n \right)}{\hat{\sigma}_n}, i = 1, 2, \dots, B$$

of the Studentized statistic.

The bootstrap then uses observations $S_{n,i}, i = 1, 2, \dots, B$ to compute the empirical cdf $\hat{J}_n(\cdot, \hat{P}_n)$ used to approximate the sampling distribution function $J_n(\cdot, P)$ of $\sqrt{n} (\hat{\theta}_n - \theta(P)) / \hat{\sigma}_n$, yielding the following two-sided $(1 - \alpha)$ confidence interval on $\theta(P)$:

$$(\hat{\theta}_n - \hat{J}_n^{-1}(1 - \alpha/2), \hat{\theta}_n + \hat{J}_n^{-1}(1 - \alpha/2)). \tag{11}$$

As can be observed in the statement and proof of the bootstrap's main theorem [67, Theorem 1.2.1], consistency follows upon assuming that the resampling and compute steps above are such that (i) the resulting approximation \hat{J}_n in a sense consistently approximates $J_n(\cdot, P)$, e.g., $\rho_L(J_n(\cdot, P), \hat{J}_n(\cdot, \hat{P}_n)) \xrightarrow{\text{wp1}} 0$ as $n \rightarrow \infty$, where ρ_L is the Lévy metric [21, Section 15.1]; and (ii) the distribution $J(\cdot, P)$ of the weak limit $J(P)$ in (6) is continuous and strictly increasing at $\inf\{x : J(x, P) \geq 1 - \alpha\}$.

Since its original introduction in 1979 [34], the bootstrap has received tremendous attention due to its simplicity and wide applicability, resulting in popular refinements [30–33], the ability to handle time series [7, 8], extensions to the functional context [24], higher-order corrections [50, 56] to improve coverage accuracy, and most recently a computationally "cheap" version [54]. Debates on whether subsampling or the bootstrap is better have continued, but it is now known that subsampling is more general in that the bootstrap requires the behavior of the bootstrap distribution $\hat{J}_n(\cdot, \hat{P}_n)$ to be smooth (around P) when seen as a function of its second argument. We go into no further detail on this point but see [67, Section 2.3]. Also see [41] for an interesting theorem on the sense in which the bootstrap in its basic form is not valid if the variance parameter does not exist.

2.4 Further Perspective and Summary of Contribution

The uniqueness of any CIP (including subsampling, the bootstrap, and what we propose here) stems from the manner in which the procedure approximates the sampling distribution of its chosen statistic. So, while subsampling uses the empirical cdf L_n in (7) formed from subsamples, and the bootstrap uses resampling, the methods proposed in this paper approximate the sampling distribution of the Studentized statistic $(\hat{\theta}_n - \theta(P))/\hat{\sigma}_n$ by characterizing its weak limit. In particular, we assume the existence of a functional CLT governing $\hat{\theta}_n$ and exploit the resulting structure to characterize the weak limit of $(\hat{\theta}_n - \theta(P))/\hat{\sigma}_n$.

To be clear, neither subsampling nor the bootstrap assume a CLT (let alone a functional CLT) on $\hat{\theta}_n$, and this is their strength. However, our argument is that there exist numerous important contexts where a functional CLT on $\hat{\theta}_n$ holds and can be usefully exploited if we can identify the weak limit of the statistic in use. For example, vis-à-vis subsampling, knowledge of the weak limit allows replacing the empirical quantiles $c_{n,q}$ in (10) by their limiting counterparts, in the process allowing to dispense with subsampling's key stipulation that batch sizes be small, that is, $m_n/n \rightarrow 0$.

To further clarify, we now provide a summary of contribution.

- (1) This work presents CLT-based overlapping batch CIPs for constructing confidence intervals on *statistical functionals*. There exists a well-developed literature on CLT-based OB CIPs for the steady-state mean, and more recently for quantiles, but the only treatment of statistical functionals through CLT-based methods that we are aware of is [60, Section 2.4].
- (2) We derive the weak limits (called OB-x limits, x=I,II,III) of the statistic underlying each of the proposed OB CIPs. Of these, the OB-II limit and its bias-correction factor (Theorem 6.1) have not appeared in the literature even in the steady-state mean context to the best of our knowledge; OB-II might prove to be especially relevant in computationally intensive settings. The OB-I and OB-III limits (Theorem 5.1 and Theorem 7.1, respectively) have appeared in the literature but in the steady-state mean [1, 2] and the quantile [10] contexts. The asymptotic moment expression for the OB-I limit (Theorem 5.2) has not appeared elsewhere but the corresponding result for the special case of fully overlapping batches in the steady-state mean context appeared in [20].
- (3) To aid future investigation of computationally intensive contexts, our analysis of overlapping batches is general in the sense that it introduces an offset parameter d_n whose value connotes the extent of batching, e.g., $d_n = 1$ connotes fully overlapping batches and $d_n \geq m_n$ connotes non-overlapping batches. We shall see (Theorem 5.2) that the effect of the extent of overlap features prominently in the asymptotic variance of the variance estimator.
- (4) Extensive numerical experimentation over a variety of applications indicates that cancellation methods resulting from the use of large batches, that is, when $m_n/n \rightarrow \beta > 0$, exhibits behavior that is consistently better. Aspects responsible for such better behavior are not yet fully understood and should form the topic of future investigation.
- (5) We provide access to code (that includes a critical value calculation module for OB-I, OB-II, and OB-III) for constructing confidence intervals on a statistical functional using our recommended OB-x methods.

3 MAIN IDEA AND SYNOPSIS OF RESULTS

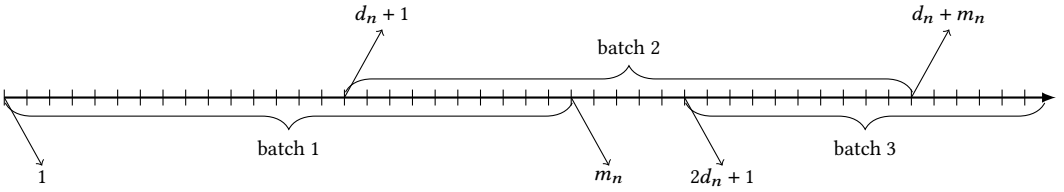
To set the stage for precisely describing the proposed confidence interval procedure, consider partitioning the available "data" X_1, X_2, \dots, X_n into b_n possibly overlapping batches each of size m_n as shown in Figure 2. The first of these batches consists of observations X_1, X_2, \dots, X_{m_n} , the second consists of observations $X_{d_n+1}, X_{d_n+2}, \dots, X_{d_n+m_n}$, and so on, and the last batch consists of

393 observations $X_{(b_n-1)d_n+1}, X_{(b_n-1)d_n+2}, \dots, X_n$. The quantity $d_n \geq 1$ represents the offset between
 394 batches, with the choice $d_n = 1$ corresponding to “fully-overlapping” batches and any choice
 395 $d_n \geq m_n$ corresponding to “non-overlapping” batches. Notice then that the offset d_n and the
 396 number of batches b_n are related as

$$397 \quad d_n = \frac{n - m_n}{b_n - 1}. \quad (12)$$

398 Suppose that the batch size m_n and the number of batches b_n are chosen so that the following limits
 399 exist:
 400

$$401 \quad \lim_{n \rightarrow \infty} \frac{m_n}{n} = \beta \in [0, 1); \quad \lim_{n \rightarrow \infty} b_n = b \in \{2, 3, \dots, \infty\}. \quad (13)$$



411 Fig. 2. The figure depicts partially overlapping batches. Batch 1 consists of observations $X_j, j = 1, 2, \dots, m_n$;
 412 batch 2 consists of observations $X_j, j = d_n + 1, d_n + 2, \dots, d_n + m_n$, and so on, with batch i consisting
 413 $X_j, j = (i - 1)d_n + 1, (i - 1)d_n + 2, \dots, (i - 1)d_n + m_n$. There are thus $b_n := d_n^{-1}(n - m_n) + 1$ batches in total,
 414 where n is the size of the dataset.
 415

416 Note that $\beta = 0$ and $b = \infty$ are allowed in (13). We will sometimes refer to β as the *asymptotic*
 417 *batch size* and to b as the *asymptotic number of batches*. Also, we will refer to $\beta = 0$ as the *small*
 418 *batch regime*, and to $\beta > 0$ as the *large batch regime*.
 419

420 3.1 “Centering” the Confidence Interval

421 Suppose we have at our disposal a method to construct a point estimator $\hat{\theta}(\{X_j, \ell \leq j \leq u\})$ of $\theta(P)$
 422 using any batch $(X_\ell, X_{\ell+1}, \dots, X_u), \ell, u \in \{1, 2, \dots, n\}$ of consecutive observations from the available
 423 data $X_j, 1 \leq j \leq n$. For now, we place no restrictions on $\hat{\theta}(\{X_j, \ell \leq j \leq u\})$ but a natural choice
 424 for $\hat{\theta}(\{X_j, \ell \leq j \leq u\})$, especially in the non-parametric setting, is the “plug-in” estimator $\theta(P_{\ell,u})$,
 425 where $P_{\ell,u}$ is the empirical measure constructed from $X_j, \ell \leq j \leq u$. For example, when $X_j, 1 \leq$
 426 $j \leq n$ are real-valued and $\theta(P)$ is the population mean $\mathbb{E}[X_1] = \int x P(dx)$, the point estimator
 427 $\hat{\theta}(\{X_j, \ell \leq j \leq u\})$ is the sample mean of the observations $X_j, \ell \leq j \leq u$; and likewise, when $\theta(P)$
 428 is the γ -quantile $\min\{x : P(X_1 \leq x) \geq \gamma\}$ of X_1 , the natural choice for the point estimator is the
 429 empirical quantile $F_{\ell,u}^{-1}(\gamma) := \min\{x : F_{\ell,u}(x) \geq \gamma\}$, where $F_{\ell,u}(x) = (u - \ell + 1)^{-1} \sum_{j=\ell}^u \mathbb{I}_{\{X_j \leq x\}}, x \in \mathbb{R}$
 430 is the usual empirical cumulative distribution function (cdf) constructed from the observations
 431 $X_j, \ell \leq j \leq u$.
 432

433 When $\ell = 1$ and $u = n$, that is, all available observations are utilized in constructing the point
 434 estimator of $\theta(P)$, we obtain what is often called the *sectioning estimator* [62], given special notation
 435 here since we heavily invoke this estimator throughout the rest of the paper:

$$436 \quad \hat{\theta}_n := \hat{\theta}(\{X_j, 1 \leq j \leq n\}). \quad (14)$$

437 The *asymptotic variance parameter* σ^2 , assumed to exist and defined as

$$438 \quad \sigma^2 := \lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\sqrt{n} (\hat{\theta}_n - \theta(P)) \right)^2 \right], \quad (15)$$

will play a crucial role in our later analysis. Also, owing to the manner in which we will construct batches, we use special notation for the point estimators constructed from the observations $(X_{(i-1)d_n+1}, X_{(i-1)d_n+2}, \dots, X_{(i-1)d_n+m_n})$ in the i -th batch (see Figure 2):

$$\hat{\theta}_{i,m_n} := \hat{\theta}(\{X_j, (i-1)d_n + 1 \leq j \leq (i-1)d_n + m_n\}), i = 1, 2, \dots, b_n \quad (16)$$

and $b_n = \frac{n-m_n}{d_n} + 1$.

We shall see shortly that the sectioning estimator appearing in (14) is a candidate for centering the confidence interval that we construct. An alternative to the sectioning estimator is the *batching estimator* [62], obtained by averaging the point estimators $\hat{\theta}_{i,m_n}, i = 1, 2, \dots, b_n$, that is,

$$\bar{\theta}_n := \frac{1}{b_n} \sum_{i=1}^{b_n} \hat{\theta}_{i,m_n}. \quad (17)$$

The sectioning and batching point estimators are the two natural choices for “centering” the confidence intervals on $\theta(P)$. We will see that confidence intervals constructed with the batching estimator might be especially useful in computationally intensive contexts.

3.2 Estimating the Variance Constant σ^2

Since the variance constant σ^2 (defined in (15)) is a measure of the inherent variability of the point estimator $\hat{\theta}_n$, σ^2 's estimation plays a key role in the confidence intervals we construct. The expression in (15) suggests that a natural estimator of σ^2 is the sample variance of $\hat{\theta}_{i,m_n}, i = 1, 2, \dots, b_n$ defined in (16), after appropriate scaling:

$$\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n) := \frac{1}{\kappa_1(\beta)} \frac{m_n}{b_n} \sum_{i=1}^{b_n} (\hat{\theta}_{i,m_n} - \hat{\theta}_n)^2, \quad \kappa_1(\beta) = 1 - \beta, \quad (18)$$

where β defined in (13) is the limiting batch size. It will become clear from our later analysis that $\kappa_1(\beta)$ appearing in (18) is a “bias-correction” constant introduced to make $\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)$ asymptotically unbiased.

Notice that the estimator $\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)$ of the variance constant σ^2 appearing in (18) uses the sectioning estimator $\hat{\theta}_n$ when computing the sample variance. An alternative is to use the batching estimator $\bar{\theta}_n$ in place of the sectioning estimator to obtain the second candidate estimator of the variance constant σ^2 :

$$\hat{\sigma}_{\text{OB-II}}^2(m_n, b_n) := \frac{1}{\kappa_2(\beta, b)} \frac{m_n}{b_n} \sum_{i=1}^{b_n} (\hat{\theta}_{i,m_n} - \bar{\theta}_n)^2, \quad (19)$$

where, as we shall see in Theorem 6.1, the bias-correction constant has the more complicated form

$$\kappa_2(\beta, b) := \begin{cases} 1 & \beta = 0; \\ 1 - 2 \left(\min\left\{\frac{\beta}{1-\beta}, 1\right\} \right) + \frac{1}{\beta} \left(\min\left\{\frac{\beta}{1-\beta}, 1\right\} \right)^2 - \frac{2}{3} \frac{1-\beta}{\beta} \left(\min\left\{\frac{\beta}{1-\beta}, 1\right\} \right)^3 & \beta > 0, b = \infty; \\ 1 - \frac{1}{b} - \frac{2}{b} \sum_{h=1}^b \left(1 - \frac{h}{b-1} \frac{1-\beta}{\beta} \right)^+ (1 - h/b) & b \in \mathbb{N} \setminus 1, \end{cases} \quad (20)$$

and b defined in (13) is the limiting number of batches.

A third estimator $\hat{\sigma}_{\text{OB-III}}^2(m_n, b_n)$ of the variance constant σ^2 that we consider, called the weighted area estimator [2, 47, 48, 70], is given as follows:

$$\hat{\sigma}_{\text{OB-III}}^2(m_n, b_n) := \frac{1}{b_n} \sum_{i=1}^{b_n} A_{i,m_n}, \quad (21)$$

where

$$A_{i,m_n} := \left(\frac{1}{m_n} \sum_{j=1}^{m_n} f\left(\frac{j}{m_n}\right) \sigma T_{i,m_n}\left(\frac{j}{m_n}\right) \right)^2; \quad T_{i,m_n}(t) := \frac{\lfloor m_n t \rfloor \left(\hat{\theta}_{i, \lfloor m_n t \rfloor} - \hat{\theta}_{i,m_n} \right)}{\sigma \sqrt{m_n}}, t \in [0, 1], \quad (22)$$

and $f : [0, 1] \rightarrow \mathbb{R}^+$ is a chosen weighting function that satisfies

$$\mathbb{E} \left[\left(\int_0^1 f(t) W(t) dt \right)^2 \right] = 1; \text{ and } f \in C^2[0, 1]. \quad (23)$$

The structure of the “standardized time series” $\{T_{i,m_n}(t), t \in [0, 1]\}$ in (22) hints at why $\hat{\sigma}_{\text{OB-III}}^2(m_n, b_n)/\sigma^2$ is an analogue of the classical chi-square random variable. Specifically, notice that $\{T_{\lfloor sm_n \rfloor, m_n}(t), t \in [0, 1]\}$ for each $s \in [0, \infty)$ should converge weakly (as $m_n \rightarrow \infty$), modulo some regularity conditions, to the standard Brownian bridge

$$B_s(t) := \{W(s+t) - W(s) - t(W(s+1) - W(s)), t \in [0, 1]\}, \quad s \in [0, \infty).$$

Correspondingly, and since $\int_0^1 f(t) B_s(t) dt \stackrel{d}{=} Z(0, 1)$ if f is chosen as stipulated in (23), $A_{\lfloor sm_n \rfloor, m_n}$ should converge weakly to $\sigma^2 Z^2(0, 1) \stackrel{d}{=} \sigma^2 \chi_1^2$, in effect justifying the weighted area estimator $\hat{\sigma}_{\text{OB-III}}^2(m_n, b_n)$.

The weighted area estimator appearing in (21) has been the topic of much research over the last three decades in the context of estimating the variance constant of a stationary time series. See [2, 39, 46, 70?] for a detailed account that includes treatment of other estimators of the variance constant.

3.3 Structure of the Proposed Confidence Intervals

The proposed interval has the same elements as a classical confidence interval, namely:

- (A) a “centering” variable, e.g., the sectioning estimator $\hat{\theta}_n \in \mathbb{R}$, or the batching estimator $\bar{\theta}_n \in \mathbb{R}$, as described in Section 3.1;
- (B) a point estimator of the asymptotic variance σ^2 , e.g., $\hat{\sigma}_{\text{OB-x}}^2(m_n, b_n)$, $x = \text{I, II, III}$; and
- (C) a statistic whose weak limit supplies the critical values associated with the confidence interval.

Once the elements in (A)–(C) are specified, a $(1 - \alpha)$ confidence interval on $\theta(P)$ can then be constructed in the usual way.

For example, when the sectioning estimator $\hat{\theta}_n$ is used in (A), the variance estimator $\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)$ is used in (B), and the Studentized root

$$T_{\text{OB-I}}(m_n, b_n) := \frac{\sqrt{n}(\hat{\theta}_n - \theta(P))}{\hat{\sigma}_{\text{OB-I}}(m_n, b_n)} \stackrel{d}{\rightarrow} T_{\text{OB-I}}(\beta, b), \quad (24)$$

is used in (C), we obtain the (two-sided) confidence interval

$$\left\{ y \in \mathbb{R} : -t_{\text{OB-I}, 1-\frac{\alpha}{2}}(\beta, b) \leq \frac{\sqrt{n}(\hat{\theta}_n - y)}{\hat{\sigma}_{\text{OB-I}}(m_n, b_n)} \leq t_{\text{OB-I}, 1-\frac{\alpha}{2}}(\beta, b) \right\}, \quad (25)$$

where

$$t_{\text{OB-I}, q}(\beta, b) = \inf \{r : P(T_{\text{OB-I}}(\beta, b) \leq r) = q\}, \quad q \in (0, 1)$$

is the q -quantile (or critical value) of the random variable $T_{\text{OB-I}}(\beta, b)$. (A one-sided confidence interval analogous to (25) is straightforward.)

Similarly, using the batching estimator $\bar{\theta}_n$ in (A), the variance estimator $\hat{\sigma}_{\text{OB-II}}^2(m_n, b_n)$ in (B), and the Studentized root

$$T_{\text{OB-II}}(m_n, b_n) := \frac{\sqrt{n}(\bar{\theta}_n - \theta(P))}{\hat{\sigma}_{\text{OB-II}}(m_n, b_n)} \xrightarrow{d} T_{\text{OB-II}}(\beta, b), \quad (26)$$

in (C), we obtain our second proposed (two-sided) confidence interval

$$\left\{ y \in \mathbb{R} : -t_{\text{OB-II}, 1-\frac{\alpha}{2}}(\beta, b) \leq \frac{\sqrt{n}(\bar{\theta}_n - y)}{\hat{\sigma}_{\text{OB-II}}(m_n, b_n)} \leq t_{\text{OB-II}, 1-\frac{\alpha}{2}}(\beta, b) \right\}, \quad (27)$$

where

$$t_{\text{OB-II}, q}(\beta, b) = \inf \{ r : P(T_{\text{OB-II}}(\beta, b) \leq r) = q \}, \quad q \in (0, 1)$$

is the q -quantile (or critical value) of the random variable $T_{\text{OB-II}}(\beta, b)$.

And, finally, using the sectioning estimator $\hat{\theta}_n$ in (A), the variance estimator $\hat{\sigma}_{\text{OB-III}}^2(m_n, b_n)$ in (B), and the Studentized root

$$T_{\text{OB-III}}(m_n, b_n) := \frac{\sqrt{n}(\hat{\theta}_n - \theta(P))}{\hat{\sigma}_{\text{OB-III}}(m_n, b_n)} \xrightarrow{d} T_{\text{OB-III}}(\beta, b), \quad (28)$$

in (C), we obtain our third proposed (two-sided) confidence interval

$$\left\{ y \in \mathbb{R} : -t_{\text{OB-III}, 1-\frac{\alpha}{2}}(\beta, b) \leq \frac{\sqrt{n}(\hat{\theta}_n - y)}{\hat{\sigma}_{\text{OB-III}}(m_n, b_n)} \leq t_{\text{OB-III}, 1-\frac{\alpha}{2}}(\beta, b) \right\}, \quad (29)$$

where

$$t_{\text{OB-III}, q}(\beta, b) = \inf \{ r : P(T_{\text{OB-III}}(\beta, b) \leq r) = q \}, \quad q \in (0, 1)$$

is the q -quantile (or critical value) of the random variable $T_{\text{OB-III}}(\beta, b)$.

REMARK. Sometimes θ is known to reside in a constrained set $\Theta \subset \mathbb{R}$, in which case the sectioning estimator $\hat{\theta}_n$, and all the batch estimators $\hat{\theta}_{i, m_n}$, $i = 1, 2, \dots, b_n$ should be suitably projected onto Θ , as should the intervals in (25), (27), and (29). This will cause a corresponding change in the weak limits along with the critical values; we do not detail this case since we view it as a distraction.

The preceding discussion should emphasize that the Studentized root $T_{\text{OB-x}}(m_n, b_n)$, $x = \text{I, II, III}$ forms the essential element of the confidence intervals we propose. And, since the exact distribution of $T_{\text{OB-x}}(m_n, b_n)$, $x = \text{I, II, III}$ is unknown in general, the outlined procedure approximates its distribution by the (purported) weak limit $T_{\text{OB-x}}(\beta, b)$, $x = \text{I, II, III}$.

3.4 Synopsis of Results

The proposed intervals (25), (27), and (29) rely crucially on the existence of the following weak limits:

$$T_{\text{OB-I}}(m_n, b_n) := \frac{\sqrt{n}(\hat{\theta}_n - \theta(P))}{\hat{\sigma}_{\text{OB-I}}(m_n, b_n)} \xrightarrow{d} T_{\text{OB-I}}(\beta, b); \quad (\text{OB-I Limit})$$

$$T_{\text{OB-II}}(m_n, b_n) := \frac{\sqrt{n}(\bar{\theta}_n - \theta(P))}{\hat{\sigma}_{\text{OB-II}}(m_n, b_n)} \xrightarrow{d} T_{\text{OB-II}}(\beta, b); \quad (\text{OB-II Limit})$$

$$T_{\text{OB-III}}(m_n, b_n) := \frac{\sqrt{n}(\hat{\theta}_n - \theta(P))}{\hat{\sigma}_{\text{OB-III}}(m_n, b_n)} \xrightarrow{d} T_{\text{OB-III}}(\beta, b), \quad (\text{OB-III Limit})$$

where β and b are the limiting batch size and number of batches as defined in (13). The existence of the weak limits $T_{\text{OB-x}}$, $x = \text{I, II, III}$, however, needs to be established and their characterization will occupy much of the rest of the paper. Furthermore, on our way to characterizing $T_{\text{OB-x}}$, $x = \text{I, II, III}$, we will also establish the weak limits of the estimators $\hat{\sigma}_{\text{OB-x}}^2(m_n, b_n)$, $x = \text{I, II, III}$ of the variance constant

Table 1. A synopsis of results. In the service of constructing confidence intervals on $\theta(P)$, we construct three Studentized roots $T_{\text{OB-}x}(m_n, b_n)$, $x=I,II,III$ obtained using combinations of candidates for the point estimator of $\theta(P)$ and for the point estimator of σ^2 . The three Studentized roots give rise to the OB- x , $x=I,II,III$ weak limits, whose nature depends on the limiting batch size $\beta := \lim_{n \rightarrow \infty} m_n/n$ and the limiting number of batches $b := \lim_{n \rightarrow \infty} b_n$. Expressions for the weak limits $T_{\text{OB-}x}(\beta, b)$, $x=I,II,III$ appear in Theorems 5.1–7.1. Critical values for the OB-I and OB-II distributions appear on page 17 and page 24.

	Centering Var.	Batch Regime $(m_n/n \rightarrow \beta \stackrel{?}{=} 0)$	Variance Estimator $(\hat{\sigma}_{\text{OB-}x}^2(m_n, b_n) \rightarrow ?)$	Statistic $(T_{\text{OB-}x}(m_n, b_n) \xrightarrow{d} ?)$
OB-I	$\hat{\theta}_n$	$\beta = 0$	$\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n) \xrightarrow{P} \sigma^2$	Z
		$\beta > 0$	$\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n) \xrightarrow{d} \sigma^2 \chi_{\text{OB-I}}^2(\beta, b)$	$T_{\text{OB-I}}(\beta, b)$
OB-II	$\bar{\theta}_n$	$\beta = 0$	$\hat{\sigma}_{\text{OB-II}}^2(m_n, b_n) \xrightarrow{P} \sigma^2$	Z
		$\beta > 0$	$\hat{\sigma}_{\text{OB-II}}^2(m_n, b_n) \xrightarrow{d} \sigma^2 \chi_{\text{OB-II}}^2(\beta, b)$	$T_{\text{OB-II}}(\beta, b)$
OB-III	$\hat{\theta}_n$	$\beta = 0$	$\hat{\sigma}_{\text{OB-III}}^2(m_n, b_n) \xrightarrow{P} \sigma^2$	Z
		$\beta > 0$	$\hat{\sigma}_{\text{OB-III}}^2(m_n, b_n) \xrightarrow{d} \sigma^2 \chi_{\text{OB-III}}^2(\beta, b)$	$T_{\text{OB-III}}(\beta, b)$

σ^2 . The random variables $T_{\text{OB-}x}$, $x = I, II, III$ and $\hat{\sigma}_{\text{OB-}x}$, $x = I, II, III$ should be seen as distribution-free statistical functional analogues of the Student's t and χ^2 random variables, respectively.

As summarized in Table 1, the nature of $T_{\text{OB-}x}$, $x = I, II, III$ (and those of $\hat{\sigma}_{\text{OB-}x}$, $x = I, II, III$) depend on the limiting batch size β and the limiting number of batches b . In particular, depending on whether $\beta = 0$ (small batch regime) or $\beta > 0$ (large batch regime), the statistics behave quite differently. For example, the small batch regime ($\beta = 0$) produces the normal limit (Z statistics) along with consistent estimation of σ^2 , whereas the large batch regime ($\beta > 0$) produces limits that are functionals of the Wiener process along with no consistent estimation of σ^2 . The asymptotic number of batches b affects the nature of the limiting distributions in the large batch regime. See Table 1 for a synopsis.

4 KEY ASSUMPTIONS

In this section, we state and comment on various regularity assumptions that will be invoked when proving the technical results. Not all of these assumptions are “standing assumptions” in that some of the results to follow (especially when $\beta = 0$) will need only a subset of the assumptions.

ASSUMPTION 1 (STATIONARITY). *The S -valued sequence $\{X_n, n \geq 1\}$ is stationary, that is, for any $n_j, j = 1, 2, \dots, k < \infty$ and $k < \infty$, the distribution of $(X_{n_1+\tau}, X_{n_2+\tau}, X_{n_3+\tau}, \dots, X_{n_k+\tau})$ does not depend on $\tau \in \mathbb{Z}^+$.*

ASSUMPTION 2 (STRONG MIXING CONDITION). Suppose \mathcal{G}, \mathcal{H} are sub- σ -algebras of \mathcal{F} in the probability space (Ω, \mathcal{F}, P) . Recall that the strong mixing constant $\alpha(\mathcal{G}, \mathcal{H})$ is given by

$$\begin{aligned} \alpha(\mathcal{G}, \mathcal{H}) &= \sup_{A \in \mathcal{G}} \sup_{B \in \mathcal{H}} |P(AB) - P(A)P(B)| \\ &= \frac{1}{2} \sup_{A \in \mathcal{G}} \mathbb{E} [|P(A|\mathcal{H}) - P(A)|] \\ &= \frac{1}{2} \sup_{A \in \mathcal{H}} \mathbb{E} [|P(A|\mathcal{G}) - P(A)|]. \end{aligned} \quad (30)$$

We assume that the S -valued sequence $\{X_n, n \geq 1\}$ has strong mixing [37, pp. 347] constants $\alpha_n := \alpha(\mathcal{F}_k, \mathcal{F}_{k,n})$ satisfying $\alpha_n \searrow 0$ as $n \rightarrow \infty$, where $\mathcal{F}_k := \sigma(X_1, X_2, \dots, X_k)$, $\mathcal{F}_{k,n} := \sigma(X_{k+n}, X_{k+n+1}, \dots)$ denote sub- σ -algebras of \mathcal{F} "separated by n ."

ASSUMPTION 3 (CENTRAL LIMIT THEOREM). The sequence $\{\hat{\theta}_n, n \geq 1\}$ of section estimators satisfies a central limit theorem (CLT), that is,

$$\sqrt{n}(\hat{\theta}_n - \theta(P)) \xrightarrow{d} \sigma N(0, 1), \quad (31)$$

where $N(0, 1)$ is the standard normal random variable and $\sigma < (\infty)$ is a finite, positive constant.

ASSUMPTION 4 (ASYMPTOTIC MOMENT EXISTENCE). The sequence $\{\hat{\theta}_n, n \geq 1\}$ of section estimators is such that, for some $\delta_0 > 0$,

$$\mathbb{E} \left[\left(\sqrt{n}(\hat{\theta}_n - \theta(P)) \right)^{2+\delta_0} \right] \rightarrow \sigma^{2+\delta_0} \text{ as } n \rightarrow \infty, \quad (32)$$

where σ is the constant appearing in Assumption 3.

ASSUMPTION 5 (STRONG INVARIANCE). The sequence $\{\hat{\theta}_n, n \geq 1\}$ of section estimators satisfies the following strong invariance principle. On a rich enough probability space, there exists a standard Wiener process $\{W(t), t \geq 0\}$ and a stationary stochastic process $\{\tilde{X}_n, n \geq 1\} \stackrel{d}{=} \{X_n, n \geq 1\}$ such that as $n \rightarrow \infty$,

$$\sup_{0 \leq t \leq n} \left| \sigma^{-1} \left(\hat{\theta}_{[t]} - \theta(P) \right) - t^{-1} W(t) \right| \leq \Gamma n^{-1/2-\delta} \sqrt{\log n} \quad \text{a.s.}, \quad (33)$$

where the constant $\delta > 0$ and the real-valued random variable Γ satisfies $\mathbb{E}[\Gamma] < \infty$.

Assumption 1 on the stationarity of the sequence $\{X_n, n \geq 1\}$ is mild and standard in settings where a confidence interval is sought. Assumption 2 on strong mixing is a weak asymptotic independence condition imposed to rigorize the intuitive idea that the dependence between events formed from subsets of the sequence $\{X_n, n \geq 1\}$ in the far past and the far future decays to zero as their separation diverges. Assumption 1 and Assumption 2 are used only in our results involving small batches, that is, when $m_n/n \rightarrow 0$.

As discussed in the introductory part of the paper, Assumption 3 on the existence of a CLT on $\hat{\theta}_n$, is fundamental to the methods presented here. While there are exceptions [67], a CLT holds in numerous useful settings where a confidence interval is desired, e.g., mean estimation [71], quantile estimation [71], gradient estimation [3], M-estimation [72], CVaR estimation [23], acf and spectral density estimation [9], and robust statistics [40], apart from other more standard estimation settings in statistics. Assumption 4 goes only a little further than Assumption 3 to stipulate the existence of the $(2 + \delta_0)$ -th moment of $\hat{\theta}_n$ (for some $\delta_0 > 0$) and its convergence to $\sigma^{2+\delta}$. It can be shown that Assumption 4 implies the uniform integrability of the sequence $\{\sqrt{n}(\hat{\theta}_n - \theta(P)), n \geq 1\}$.

The inequality in (33) of Assumption 5, sometimes called “strong invariance,” essentially stipulates that the scaled process $\left\{ \sqrt{n}\sigma^{-1} \left(\hat{\theta}_{\lfloor t \rfloor} - \theta(\tilde{P}) \right), t \leq n \right\}$ can be approximated uniformly to within $n^{-\delta}$ almost surely, by a suitable standard Wiener process on a rich enough probability space. As argued in Philipp and Stout [65], and Glynn and Iglehart [45], Assumption 5 holds for a variety of weakly dependent processes. See [17] for strong invariance theorems on partial sums, empirical processes, and quantile processes.

As will become evident, Assumption 5 is used only in proving results that involve large batches, that is, when $m_n/n \rightarrow \beta > 0$. We believe all these results will still hold with a functional CLT on $\hat{\theta}_n$ instead of Assumption 5. (Loosely, strong approximation \Rightarrow functional CLT \Rightarrow CLT – see, for instance, [44, 71].) Despite this increased generality that a functional CLT affords, we have chosen to remain with Assumption 5 since the resulting proofs are more intuitive.

5 THE OB-I LIMIT

In this section, we characterize the weak limit of

$$T_{\text{OB-I}}(m_n, b_n) := \frac{\sqrt{n}(\hat{\theta}_n - \theta(P))}{\hat{\sigma}_{\text{OB-I}}(m_n, b_n)}, \tag{34}$$

as described in Section 3.4. Along the way, we also characterize the asymptotic behavior of the variance estimator $\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)$. The ensuing Section 5.1 treats the $\beta := \lim_{n \rightarrow \infty} m_n/n > 0$ (large batch) regime, and Section 5.3 treats the $\beta = 0$ (small batch) regime.

5.1 Large Batch Regime for OB-I

Theorem 5.1 that follows asserts that $\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)/\sigma^2$ and $T_{\text{OB-I}}(m_n, b_n)$ converge weakly to certain functionals of the Wiener process that we denote $\chi_{\text{OB-I}}^2(\beta, b)$ and $T_{\text{OB-I}}(\beta, b)$, respectively. It is important that Theorem 5.1 needs the strong invariance Assumption 5 to hold so that the dependence across batches can be characterized precisely.

THEOREM 5.1 (OB-I LARGE BATCH REGIME). *Suppose that Assumptions 1–5 hold, and that $\beta = \lim_{n \rightarrow \infty} m_n/n \in (0, 1)$. Assume also that $b_n \rightarrow b \in \{2, 3, \dots, \infty\}$ as $n \rightarrow \infty$. Define*

$$\chi_{\text{OB-I}}^2(\beta, b) := \begin{cases} \frac{1}{\kappa_1(\beta, b)} \frac{1}{\beta(1-\beta)} \int_0^{1-\beta} (W(u+\beta) - W(u) - \beta W(1))^2 du & b = \infty; \\ \frac{1}{\kappa_1(\beta, b)} \frac{1}{\beta b} \sum_{j=1}^b (W(c_j + \beta) - W(c_j) - \beta W(1))^2 & b \in \mathbb{N} \setminus \{1\}, \end{cases} \tag{35}$$

where $\kappa_1(\beta, b) = 1 - \beta$ and $c_j := (j - 1)\frac{1-\beta}{b-1}$. Then, as $n \rightarrow \infty$,

$$\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n) \xrightarrow{d} \sigma^2 \chi_{\text{OB-I}}^2(\beta, b); \quad \text{and} \quad T_{\text{OB-I}}(m_n, b_n) \xrightarrow{d} \frac{W(1)}{\sqrt{\chi_{\text{OB-I}}^2(\beta, b)}}. \tag{36}$$

The following theorem characterizes the (asymptotic) moments of the OB-I variance estimator $\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)$.

THEOREM 5.2 (OB-I MOMENTS). *Let the postulates of Theorem 5.1 hold. If the random variable Γ appearing in Assumption 5 satisfies $\mathbb{E}[\Gamma^4] < \infty$, and $|m_n/n - \beta| = o(n^{-\delta})$, then*

$$\mathbb{E}[\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)] = \sigma^2 + O(\epsilon_{1,n}); \quad \epsilon_{1,n} := \frac{n^{-\delta}}{\beta^\delta \kappa_1(\beta, b)} (\sqrt{2 \log^2 \beta n \log^2 n} + \sqrt{2} \log^2 n) \tag{37}$$

and $\delta > 0$ is the constant appearing in Assumption 5.

Recalling that $b := \lim_n b_n \in \{2, 3, \dots, \infty\}$, suppose further that

$$\eta := \lim_n \frac{b_n}{n} \in [0, \infty),$$

implying that necessarily

$$d := \lim_n d_n = \lim_n \frac{n - m_n}{b_n - 1} = \begin{cases} \frac{1-\beta}{\eta} & \eta > 0; \\ \infty & \eta = 0. \end{cases}$$

Then, after redefining $\infty \times 0 = 0$, we have that

$$\lim_n \text{Var}(\hat{\sigma}_{\text{OB-1}}^2(m_n, b_n)) = \frac{\sigma^4}{(1-\beta)^2} \left(2(1-2\beta+3\beta^2) \tilde{\mu}_0 + 6(1-\beta)^2 \mu_0 - 8d(1-\beta)\mu_1 + 4\mu_2 \right), \quad (38)$$

where $\tilde{\mu}_0, \mu_0, \mu_1, \mu_2$ are given by

$$\tilde{\mu}_0 := \begin{cases} \frac{1}{2} \left(\frac{1-2\beta}{1-\beta} \right)^2 \mathbb{I}_{\{\beta \leq 1/2\}} & \text{if } b = \infty, \\ \frac{1}{2} \left(1 - \frac{1}{b} \lceil \frac{\beta}{1-\beta} (b-1) \rceil \right) \left(1 - \frac{1}{b} \lceil \frac{\beta}{1-\beta} (b-1) \rceil + \frac{1}{b} \right) \mathbb{I}_{\{\beta \leq 1/2\}} & \text{if } b < \infty; \end{cases} \quad (39)$$

and defining $\gamma := \frac{\beta}{1-\beta} \wedge 1$,

$$\begin{aligned} \mu_0 &:= \begin{cases} \gamma \left(1 - \frac{\gamma}{2} \right) & \text{if } b = \infty, \\ \frac{1}{b} \lfloor \gamma(b-1) \rfloor \left(1 - \frac{1}{2} \frac{\lfloor \gamma(b-1) \rfloor}{b} - \frac{1}{2} \right) & \text{if } b < \infty; \end{cases} \\ \mu_1 &:= \frac{1}{6} \gamma^2 \frac{\eta}{\beta} (3 - 2\gamma) \mathbb{I}_{\{b=\infty\}}, \\ \mu_2 &:= \frac{1}{2} \gamma^3 \left(\frac{\eta}{\beta} \right)^2 \left(\frac{2}{3} - \frac{1}{2} \gamma \right) \mathbb{I}_{\{b=\infty\}}. \end{aligned} \quad (40)$$

We make some further observations before providing the proofs of Theorem 5.1 and Theorem 5.2.

- (a) The estimator $\hat{\sigma}_{\text{OB-1}}^2(m_n, b_n)$ does not consistently estimate the variance parameter σ^2 , but converges weakly to the product of σ^2 and the random variable $\chi_{\text{OB-1}}^2(\beta, b)$ appearing in (36). As in all cancellation methods, the weak limit of $T_{\text{OB-1}}(m_n, b_n)$ does not involve σ^2 since it “cancels out.” We slightly abuse notation for ease of exposition and use (36) to define the $T_{\text{OB-1}}$ random variable:

$$T_{\text{OB-1}}(\beta, b) := \frac{W(1)}{\sqrt{\chi_{\text{OB-1}}^2(\beta, b)}}, \quad (\beta, b) \in (0, 1) \times \mathbb{N} \setminus \{1\}.$$

- (b) The factor $\kappa_1(\beta, b) = 1 - \beta$ is a “bias correction” factor introduced to ensure that $\hat{\sigma}_{\text{OB-1}}^2(m_n, b_n)$ is asymptotically unbiased.
- (c) The expression for $\chi_{\text{OB-1}}^2(\beta, b)$ in Theorem 5.1 seems to have appeared first in [1, pp. 326] for the steady-state mean context and assuming fully overlapping batches, that is, for $d_n = 1$ and $b = \infty$. (The reader should be aware that while b in the current paper refers to the limiting number of batches, b in [1] refers to the ratio $n/m_n \rightarrow \beta^{-1}$. Furthermore, a simple re-scaling of the Wiener process is needed to see that the expression appearing in Theorem 5.1 and that in [1, pp. 326] are equivalent.) Similarly, the special case of fully overlapping batches

and $b = \infty$ for $\text{Var}(\hat{\sigma}_{\text{OB-1}}^2(m_n, b_n))$ in Theorem 5.2 appears in [20, pp. 290] for the context of the steady-state mean.

- (d) We can show through calculus on (38) that $\inf_{\beta \in (0,1)} \{\lim_{n \rightarrow \infty} \text{Var}(\hat{\sigma}_{\text{OB-1}}^2(m_n, b_n))\} = 0$ is approached as $\beta \rightarrow 0$. (The infimum is not attained although there is a local minimum around $\beta = 0.467$.) This suggests using small batches but we will argue later that this is counter to what is seen in practice and that using (38) as a means to decide the quality of confidence intervals is misleading.
- (e) The offset parameter d_n comes into play through its effect on the limiting number of batches b . Specifically, notice that since $b_n = 1 + (n - m_n)/d_n$ and $m_n/n \rightarrow \beta$, the asymptotic number of batches $b = \infty$ if $d_n/n = o(1)$, and $b < \infty$ if $\lim_n d_n/n > 0$ (assuming it exists).
- (f) The table in Figure 3 displays the critical values $t_{\text{OB-1}, 1-\alpha}(\beta, b) := \min_x P(T_{\text{OB-1}}(\beta, b) \leq x) \geq 1 - \alpha$ associated with the $T_{\text{OB-1}}$ distribution as a function of $1 - \alpha$ and for different values of the parameters β, b . R and MATLAB code for calculating the critical values can be obtained through <https://web.ics.purdue.edu/~pasupath>.

5.2 Proofs of Theorem 5.1 and Theorem 5.2

PROOF OF THEOREM 5.1. Since Assumption 5 holds, we will establish the first assertion in (36) by comparing individual terms that comprise $\hat{\sigma}_{\text{OB-1}}^2(m_n, b_n)$ against corresponding terms constructed from the Wiener process. Specifically, let's define

$$\tilde{B}_{j,m_n} := m_n^{-1} \left(W((j-1) \frac{n-m_n}{b_n-1} + m_n) - W((j-1) \frac{n-m_n}{b_n-1}) \right), \quad j = 1, 2, \dots, b_n \quad (41)$$

and observe that

$$\begin{aligned} (1-\beta)\hat{\sigma}_{\text{OB-1}}^2(m_n, b_n) &= \underbrace{\frac{m_n}{b_n} \sum_{j=1}^{b_n} \left[\left(\hat{\theta}_{j,m_n} - \hat{\theta}_n \right)^2 - \sigma^2 \left(\tilde{B}_{j,m_n} - n^{-1}W(n) \right)^2 \right]}_{E_n(m_n, b_n)} \quad (42) \\ &\quad + \underbrace{\sigma^2 \frac{1}{b_n} \sum_{j=1}^{b_n} \left(\sqrt{m_n} \tilde{B}_{j,m_n} - \frac{\sqrt{m_n}}{n} W(n) \right)^2}_{I_n} \\ &=: E_n(m_n, b_n) + I_n. \quad (43) \end{aligned}$$

Noticing that

$$\hat{\theta}_{j,m_n} - \hat{\theta}_n = \underbrace{\left(\hat{\theta}_{j,m_n} - \sigma \tilde{B}_{j,m_n} \right)}_{U_{j,m_n}} + \underbrace{\sigma \left(\tilde{B}_{j,m_n} - n^{-1}W(n) \right)}_{H_{j,m_n}} + \underbrace{\left(n^{-1}\sigma W(n) - \hat{\theta}_n \right)}_{C_n}, \quad (44)$$

we can then write

$$E_n(m_n, b_n) = m_n \left(\frac{1}{b_n} \sum_{j=1}^{b_n} U_{j,m_n}^2 + 2 \frac{\sigma}{b_n} \sum_{j=1}^{b_n} U_{j,m_n} H_{j,m_n} + 2 \frac{C_n}{b_n} \sum_{j=1}^{b_n} U_{j,m_n} + 2 \frac{\sigma C_n}{b_n} \sum_{j=1}^{b_n} H_{j,m_n} + C_n^2 \right). \quad (45)$$

Now, we see that except for a set of measure zero in the probability space implied by Assumption 5, there exists $\Gamma(\omega)$ such that, uniformly in j ,

$$|U_{j,m_n}| \leq \Gamma(\omega) m_n^{-1/2-\delta} (\log^2 m_n)^{1/2}; \quad |C_n| \leq \Gamma(\omega) n^{-1/2-\delta} (\log^2 n)^{1/2}, \quad (46)$$

834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882

OB-I CRITICAL VALUES

	$1-\alpha$	2	3	5	11	b				Inf
						16	26	51	101	
$\beta = 0.010$	0.550	0.223	0.173	0.149	0.135	0.131	0.129	0.127	0.126	0.126
	0.600	0.457	0.352	0.301	0.272	0.265	0.260	0.256	0.254	0.254
	0.650	0.717	0.542	0.461	0.414	0.404	0.396	0.389	0.386	0.386
	0.700	1.022	0.752	0.633	0.565	0.551	0.539	0.530	0.526	0.526
	0.750	1.407	0.995	0.824	0.730	0.710	0.694	0.683	0.677	0.676
	0.800	1.937	1.293	1.047	0.917	0.890	0.869	0.853	0.845	0.844
	0.850	2.762	1.689	1.323	1.141	1.103	1.074	1.052	1.042	1.040
$\beta = 0.050$	0.900	4.331	2.298	1.706	1.432	1.378	1.336	1.305	1.290	1.287
	0.950	8.884	3.558	2.372	1.891	1.801	1.733	1.684	1.660	1.655
	0.550	0.218	0.170	0.146	0.132	0.129	0.127	0.127	0.127	0.127
	0.600	0.448	0.345	0.295	0.266	0.260	0.257	0.256	0.256	0.256
	0.650	0.702	0.531	0.451	0.405	0.395	0.391	0.390	0.389	0.389
	0.700	1.001	0.737	0.620	0.554	0.539	0.534	0.531	0.530	0.530
	0.750	1.378	0.975	0.807	0.715	0.696	0.688	0.684	0.683	0.683
$\beta = 0.100$	0.800	1.897	1.266	1.025	0.899	0.872	0.860	0.856	0.855	0.854
	0.850	2.705	1.655	1.296	1.117	1.081	1.064	1.058	1.056	1.055
	0.900	4.242	2.251	1.671	1.403	1.350	1.324	1.315	1.312	1.312
	0.950	8.703	3.486	2.323	1.853	1.765	1.720	1.704	1.700	1.699
	0.550	0.212	0.165	0.142	0.129	0.129	0.128	0.128	0.128	0.128
	0.600	0.436	0.335	0.287	0.261	0.260	0.259	0.258	0.258	0.258
	0.650	0.684	0.517	0.439	0.398	0.397	0.394	0.394	0.393	0.393
$\beta = 0.200$	0.700	0.975	0.717	0.603	0.544	0.541	0.538	0.537	0.537	0.537
	0.750	1.342	0.949	0.786	0.703	0.699	0.695	0.693	0.693	0.693
	0.800	1.847	1.232	0.998	0.883	0.877	0.871	0.869	0.868	0.868
	0.850	2.633	1.611	1.262	1.098	1.089	1.080	1.077	1.077	1.077
	0.900	4.129	2.191	1.626	1.379	1.364	1.351	1.346	1.345	1.345
	0.950	8.471	3.393	2.261	1.822	1.792	1.770	1.762	1.760	1.760
	0.550	0.200	0.156	0.134	0.132	0.131	0.131	0.131	0.130	0.130
$\beta = 0.300$	0.600	0.411	0.316	0.271	0.266	0.264	0.264	0.264	0.264	0.264
	0.650	0.645	0.487	0.414	0.406	0.404	0.403	0.402	0.402	0.402
	0.700	0.919	0.676	0.569	0.555	0.552	0.551	0.550	0.550	0.550
	0.750	1.265	0.894	0.741	0.720	0.716	0.714	0.713	0.713	0.713
	0.800	1.741	1.162	0.941	0.909	0.902	0.899	0.899	0.898	0.899
	0.850	2.483	1.519	1.190	1.137	1.128	1.124	1.123	1.122	1.123
	0.900	3.893	2.066	1.533	1.440	1.427	1.420	1.418	1.418	1.419
$\beta = 0.400$	0.950	7.986	3.199	2.132	1.934	1.909	1.897	1.893	1.893	1.893
	0.550	0.187	0.146	0.137	0.134	0.133	0.133	0.133	0.133	0.133
	0.600	0.384	0.296	0.277	0.271	0.270	0.269	0.269	0.269	0.269
	0.650	0.603	0.456	0.425	0.414	0.412	0.411	0.411	0.411	0.411
	0.700	0.860	0.632	0.585	0.568	0.566	0.564	0.564	0.563	0.564
	0.750	1.183	0.837	0.764	0.739	0.736	0.734	0.733	0.733	0.733
	0.800	1.629	1.087	0.973	0.937	0.933	0.930	0.929	0.928	0.929
$\beta = 0.500$	0.850	2.322	1.420	1.236	1.181	1.174	1.170	1.168	1.168	1.168
	0.900	3.642	1.932	1.603	1.511	1.500	1.494	1.491	1.491	1.491
	0.950	7.470	2.992	2.252	2.063	2.041	2.029	2.024	2.023	2.023
	0.550	0.212	0.150	0.141	0.138	0.138	0.138	0.138	0.138	0.138
	0.600	0.436	0.306	0.286	0.280	0.280	0.280	0.280	0.280	0.280
	0.650	0.684	0.471	0.439	0.431	0.430	0.430	0.430	0.430	0.431
	0.700	0.975	0.654	0.606	0.594	0.594	0.594	0.594	0.594	0.595
$\beta = 0.600$	0.750	1.342	0.865	0.797	0.780	0.779	0.779	0.780	0.780	0.781
	0.800	1.847	1.125	1.024	1.000	0.998	0.999	1.000	1.001	1.001
	0.850	2.633	1.472	1.315	1.278	1.277	1.277	1.279	1.280	1.281
	0.900	4.129	2.004	1.733	1.671	1.668	1.669	1.672	1.674	1.675
	0.950	8.471	3.106	2.496	2.359	2.351	2.352	2.356	2.359	2.362
	0.550	0.218	0.150	0.141	0.139	0.139	0.139	0.139	0.139	0.139
	0.600	0.448	0.306	0.286	0.281	0.281	0.281	0.281	0.281	0.281
$\beta = 0.750$	0.650	0.702	0.471	0.440	0.432	0.432	0.432	0.432	0.432	0.432
	0.700	1.001	0.654	0.608	0.597	0.596	0.596	0.597	0.597	0.597
	0.750	1.378	0.866	0.799	0.783	0.782	0.783	0.783	0.784	0.784
	0.800	1.897	1.126	1.028	1.005	1.004	1.005	1.006	1.007	1.007
	0.850	2.705	1.473	1.321	1.287	1.286	1.286	1.288	1.290	1.291
	0.900	4.242	2.007	1.744	1.685	1.682	1.684	1.686	1.689	1.690
	0.950	8.703	3.112	2.515	2.384	2.377	2.378	2.382	2.386	2.389

Fig. 3. T_{OB-I} critical values. The table displays critical values $t_{OB-I,1-\alpha}(\beta, b) := \inf\{r : P(T_{OB-I}(\beta, b) \leq r) = 1-\alpha\}$ associated with the T_{OB-I} distribution as a function of $1-\alpha$, the asymptotic batch size β , and the asymptotic number of batches b .

Furthermore, due to Theorem A.4, for any given $\epsilon > 0$, except for a set of measure zero in the probability space implied by Assumption 5, there exists $n_0(\omega, \epsilon)$ such that for all $n \geq n_0(\omega, \epsilon)$, and

uniformly in j ,

$$\begin{aligned}
 |H_{j,m_n}| &\leq (1 + \epsilon) \left(m_n^{-1/2} \left(2(\log^2 n - \log \frac{m_n}{n}) \right)^{1/2} + n^{-1/2} (2 \log^2 n)^{1/2} \right) \\
 &\leq (1 + \epsilon) m_n^{-1/2} \left(\left(2(\log^2 n - \log \frac{m_n}{n}) \right)^{1/2} + (2 \log^2 n)^{1/2} \right)
 \end{aligned} \tag{47}$$

after ignoring non-integralities.

Plugging (46) and (47) in (45), we get

$$\begin{aligned}
 E_n(m_n, b_n) &\leq \Gamma^2(\omega) m_n^{-2\delta} \log^2 m_n \\
 &\quad + 2\sigma(1 + \epsilon) \Gamma(\omega) m_n^{-\delta} \left((\log^2 m_n)^{1/2} + (\log^2 n)^{1/2} \right) \left((2(\log^2 n - \log \frac{m_n}{n}))^{1/2} + (2 \log^2 n)^{1/2} \right) \\
 &\quad + 2\sigma \left(\frac{m_n}{n} \right)^{\frac{1}{2}} \Gamma^2(\omega) m_n^{-2\delta} (\log^2 m_n \log^2 n)^{1/2} + \Gamma^2(\omega) \left(\frac{m_n}{n} \right) m_n^{-2\delta} \log^2 n.
 \end{aligned} \tag{48}$$

Notice that the second term appearing on the right-hand side of (48) is dominant and goes to zero almost surely.

Now lets calculate the weak limit of $I_n := \sigma^2 \frac{1}{b_n} \sum_{j=1}^{b_n} \left(\sqrt{m_n} \tilde{B}_{j,m_n} - \frac{\sqrt{m_n}}{n} W(n) \right)^2$ appearing in (42).

To facilitate calculation, define the lattice $\{0, \delta_n, 2\delta_n, \dots, \lfloor \frac{1}{\delta_n} \rfloor \delta_n\}$ having resolution $\delta_n := \frac{1 - (m_n/n)}{b_n - 1}$, and a corresponding projection operation $\lfloor u \rfloor_{\delta_n} := \max\{k\delta_n : u \geq k\delta_n, k \in \mathbb{Z}\}$, $u \in [0, 1 - \delta_n]$.

Now, recalling that $b_n = 1 + d_n^{-1}(n - m_n)$, we can rewrite

$$\begin{aligned}
 I_n &= \sigma^2 \frac{1}{b_n} \sum_{j=1}^{b_n} \left(\sqrt{m_n} \tilde{B}_{j,m_n} - \frac{\sqrt{m_n}}{n} W(n) \right)^2 \\
 &= \sigma^2 \frac{1}{b_n} \int_0^{1 - \frac{m_n}{n} + \delta_n} \left(\frac{1}{\sqrt{m_n}} (W(n \lfloor u \rfloor_{\delta_n} + m_n) - W(n \lfloor u \rfloor_{\delta_n})) - \frac{\sqrt{m_n}}{n} W(n) \right)^2 \frac{1}{\delta_n} du \\
 &= \sigma^2 \frac{b_n - 1}{b_n} \frac{n}{n - m_n} \int_0^{1 - \frac{m_n}{n} + \delta_n} \left(\frac{1}{\sqrt{m_n}} (W(n \lfloor u \rfloor_{\delta_n} + m_n) - W(n \lfloor u \rfloor_{\delta_n})) - \frac{\sqrt{m_n}}{n} W(n) \right)^2 du \\
 &\stackrel{d}{=} \sigma^2 \frac{b_n - 1}{b_n} \frac{n}{n - m_n} \frac{n}{m_n} \int_0^{1 - \frac{m_n}{n} + \delta_n} \left(W(\lfloor u \rfloor_{\delta_n} + \frac{m_n}{n}) - W(\lfloor u \rfloor_{\delta_n}) - \frac{m_n}{n} W(1) \right)^2 du \\
 &\rightarrow \sigma^2 \frac{\beta^{-1}}{1 - \beta} \int_0^{1 - \beta} (W(u + \beta) - W(u) - \beta W(1))^2 du
 \end{aligned} \tag{49}$$

if $\delta_n \rightarrow 0$ as $n \rightarrow \infty$ which happens when $b_n \rightarrow b = \infty$. This proves the $b = \infty$ case appearing in (36).

To prove the $b \in \{2, 3, \dots\}$ case, we observe that

$$\begin{aligned}
 I_n &:= \sigma^2 \frac{1}{b_n} \sum_{j=1}^{b_n} \left(\sqrt{m_n} \tilde{B}_{j,m_n} - \frac{\sqrt{m_n}}{n} W(n) \right)^2 \\
 &\stackrel{d}{=} \sigma^2 \frac{n}{m_n} \frac{1}{b_n} \sum_{j=1}^{b_n} \left(W((j - 1) \frac{1 - (m_n/n)}{b_n - 1} + \frac{m_n}{n}) - W((j - 1) \frac{1 - (m_n/n)}{b_n - 1}) - \frac{m_n}{n} W(1) \right)^2 \\
 &\rightarrow \sigma^2 \frac{1}{\beta} \frac{1}{b} \sum_{j=1}^b \left(W((j - 1) \frac{1 - \beta}{b - 1} + \beta) - W((j - 1) \frac{1 - \beta}{b - 1}) - \beta W(1) \right)^2,
 \end{aligned} \tag{50}$$

as $n \rightarrow \infty$, proving the $b < \infty$ case appearing in (36).

Let's now prove the second statement in (36) holds. From Assumption 5 we have

$$\left| \sqrt{n} \frac{(\hat{\theta}_n - \theta(P))}{\sigma} - \frac{1}{\sqrt{n}} W(n) \right| \leq \Gamma \frac{1}{n^\delta} \sqrt{\log n} \quad \text{a.s.}, \quad (51)$$

where Γ is a well-defined random variable with finite mean, and $\delta > 0$. Hence

$$\sqrt{n}(\hat{\theta}_n - \theta(P)) = \frac{1}{\sqrt{n}} W(n) + \tilde{E}_n; \quad \tilde{E}_n = o\left(\frac{1}{n^{\delta/2}}\right) \text{ a.s.} \quad (52)$$

We can then write

$$T_{\text{OB-I}}(m_n, b_n) := \frac{\sqrt{n}(\hat{\theta}_n - \theta(P))}{\hat{\sigma}_{\text{OB-I}}(m_n, b_n)} = \frac{\frac{1}{\sqrt{n}} W(n) + \tilde{E}_n}{\sqrt{\frac{1}{1-\beta}(I_n + E_n(m_n, b_n))}}, \quad (53)$$

where I_n and $E_n(m_n, b_n)$ were introduced in (42), and both $E_n(m_n, b_n)$ and \tilde{E}_n go to zero almost surely. Now apply Theorem A.5 on (53) using the derived weak limits in (49) and (50) and observing that $\frac{1}{\sqrt{n}} W(n) \stackrel{d}{=} W(1)$, to conclude that the second assertion in (36) holds. \square

PROOF OF THEOREM 5.2. Let's next prove the asymptotic expansion appearing in (38). Simple algebra yields, for all $j = 1, 2, \dots, b_n$,

$$\mathbb{E}[(\tilde{B}_{j,m_n} - n^{-1} W(n))^2] = \frac{1}{m_n} - \frac{1}{n}, \quad (54)$$

implying that

$$\mathbb{E}[I_n] = \left(1 - \frac{m_n}{n}\right) \sigma^2. \quad (55)$$

Plugging (55) and the inequality (48) in (42) (after noticing that Assumption 5 guarantees that $\mathbb{E}[\Gamma^2] < \infty$), we conclude that as $n \rightarrow \infty$,

$$\mathbb{E}[\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)] = \sigma^2 + O(\epsilon_{1,n}), \quad (56)$$

where $\epsilon_{1,n}$ is defined in (37) and we recall that δ is the constant appearing in Assumption 5. This proves the assertion in (37).

Using a similar but tedious calculation, we find that

$$\text{Var}(I_n) = \beta^4 \left(\frac{4\beta^{-3} - 11\beta^{-2} + 4\beta^{-1} + 6}{3(1-\beta)^4} \right) \sigma^4. \quad (57)$$

Again plugging (57) and the inequality (48) in (42) (after noticing that Assumption 5 guarantees that $\mathbb{E}[\Gamma^4] < \infty$), we conclude that as $n \rightarrow \infty$,

$$\text{Var}(\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)) = \beta^4 \left(\frac{4\beta^{-3} - 11\beta^{-2} + 4\beta^{-1} + 6}{3(1-\beta)^4} \right) \sigma^4 + O(\epsilon_{1,n}^2), \quad (58)$$

thus proving the assertion in (38). \square

5.3 Small Batch Regime for OB-I

Theorem 5.1 characterizes the effect of using large batch sizes, that is, $\lim_{n \rightarrow \infty} m_n/n = \beta > 0$ on the asymptotic behavior of $T_{\text{OB-I}}(m_n, b_n)$ and $\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)$. Theorem 5.3 does the same but for the small batch ($\beta = 0$) context. In particular, Theorem 5.3 asserts that when small batches are used, $\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)$ consistently estimates σ^2 , and that $T_{\text{OB-I}}(m_n, b_n)$ converges to the standard normal distribution.

981 THEOREM 5.3 (OB-ISMALL BATCH REGIME). Suppose Assumptions 1–4 hold, and that $\beta = \lim_{n \rightarrow \infty} m_n/n =$
 982 0. Assume that the asymptotic number of batches $b := \lim_n b_n = \infty$. Then, as $n \rightarrow \infty$,

983
$$\hat{\sigma}_{\text{OB-1}}^2(m_n, b_n) \xrightarrow{p} \sigma^2; \quad \text{and} \quad T_{\text{OB-1}}(m_n, b_n) \xrightarrow{d} N(0, 1). \quad (59)$$

985 PROOF. Since $\beta = 0$, $\kappa_1(\beta) = 1 - \beta = 0$ and

986
$$\hat{\sigma}_{\text{OB-1}}^2(m_n, b_n) := \frac{m_n}{b_n} \sum_{i=1}^{b_n} (\hat{\theta}_{i,m_n} - \hat{\theta}_n)^2. \quad (60)$$

987 Also, define

988
$$\tilde{\sigma}_{\text{OB-1}}^2(m_n, b_n) := \frac{1}{b_n} \sum_{i=1}^{b_n} \underbrace{m_n (\hat{\theta}_{i,m_n} - \theta(P))^2}_{R_{i,m_n}}; \quad \tilde{\sigma}_{\text{OB-1}}^2(m_n, d_n; r) := \frac{1}{b_n} \sum_{i=1}^{b_n} \underbrace{R_{i,m_n} \mathbb{I}_{[0,r]}(R_{i,m_n})}_{R_{i,m_n}(r)}, \quad (61)$$

989 We will first demonstrate that

990
$$\tilde{\sigma}_{\text{OB-1}}^2(m_n, b_n) \xrightarrow{p} \sigma^2. \quad (62)$$

991 Notice that

992
$$\mathbb{E} [|\tilde{\sigma}_{\text{OB-1}}^2(m_n, b_n) - \sigma^2|] \leq \underbrace{\mathbb{E} [|\tilde{\sigma}_{\text{OB-1}}^2(m_n, b_n) - \tilde{\sigma}_{\text{OB-1}}^2(m_n, d_n; r)|]}_{\text{I}} + \underbrace{\mathbb{E} [|\tilde{\sigma}_{\text{OB-1}}^2(m_n, d_n; r) - \sigma^2(r)|]}_{\text{II}} + \underbrace{\mathbb{E} [|\sigma^2(r) - \sigma^2|]}_{\text{III}}, \quad (63)$$

993 where

994
$$\sigma^2(r) := \mathbb{E} [\sigma^2 Z^2 \mathbb{I}_{[0,r]}(\sigma^2 Z^2)]; \quad Z \stackrel{d}{=} N(0, 1).$$

995 Let's consider the first and last terms on the right-hand side of (63). Since $\tilde{\sigma}_{\text{OB-1}}^2(m_n, b_n) - \tilde{\sigma}_{\text{OB-1}}^2(m_n, d_n; r) =$
 996 $\frac{1}{b_n} \sum_{i=1}^{b_n} R_{i,m_n} \mathbb{I}_{(r,\infty)}(R_{i,m_n}) \geq 0$, and $R_{i,m_n}, i = 1, 2, \dots, b_n$ are identically distributed, we have

997
$$\mathbb{E} [|\tilde{\sigma}_{\text{OB-1}}^2(m_n, b_n) - \tilde{\sigma}_{\text{OB-1}}^2(m_n, d_n; r)|] = \mathbb{E} [R_{i,m_n} \mathbb{I}_{(r,\infty)}(R_{i,m_n})]. \quad (64)$$

998 Furthermore, due to Assumption 4, we know that R_{i,m_n} is uniformly integrable (for each i), and
 999 hence for any given $\epsilon > 0$, there exists $r_0 = r_0(\epsilon)$ (not dependent on i) such that for $r \geq r_0$,

1000
$$\mathbb{E} [R_{i,m_n} \mathbb{I}_{(r,\infty)}(R_{i,m_n})] \leq \epsilon; \quad \mathbb{E} [\sigma^2 Z^2 \mathbb{I}_{(r,\infty)}(\sigma^2 Z^2)] \leq \epsilon. \quad (65)$$

1001 From (65), we see that the terms I and III in (63) satisfy, for $r \geq r_0$,

1002
$$\mathbb{E} [|\tilde{\sigma}_{\text{OB-1}}^2(m_n, b_n) - \tilde{\sigma}_{\text{OB-1}}^2(m_n, d_n; r)|] \leq \epsilon; \quad \mathbb{E} [|\sigma^2(r) - \sigma^2|] \leq \epsilon. \quad (66)$$

1003 Let's now analyze the term II in (63). Write

1004
$$\mathbb{E} [|\tilde{\sigma}_{\text{OB-1}}^2(m_n, d_n; r) - \sigma^2(r)|] \leq \underbrace{\sqrt{\text{Var}(\tilde{\sigma}_{\text{OB-1}}^2(m_n, d_n; r))}}_{I_3} + \underbrace{|\mathbb{E} [\tilde{\sigma}_{\text{OB-1}}^2(m_n, d_n; r)] - \sigma^2(r)|}_{I_4}. \quad (67)$$

1005 From Assumption 4 and since $R_{i,m_n}, i = 1, 2, \dots, b_n$ have identical distributions, we know that
 1006 $\mathbb{E} [\tilde{\sigma}_{\text{OB-1}}^2(m_n, b_n)] \rightarrow \sigma^2$ as $n \rightarrow \infty$. This fact and the uniform integrability of $\tilde{\sigma}_{\text{OB-1}}^2(m_n, b_n)$ mean
 1007 that for any $\epsilon > 0$, there exist $\ell = \ell(\epsilon)$ and $r_1 = r_1(\epsilon)$ such that for $n \geq \ell$ and $r \geq r_1$ the term I_4
 1008 in (67) satisfies

1009
$$|\mathbb{E} [\tilde{\sigma}_{\text{OB-1}}^2(m_n, d_n; r)] - \sigma^2(r)| \leq \epsilon. \quad (68)$$

To quantify term I_3 in (67), write

$$\begin{aligned}
 \text{Var}(\tilde{\sigma}_{OB-I}^2(m_n, d_n; r)) &= \frac{1}{b_n} \text{Var}(R_{1,m_n}(r)) + \frac{2}{b_n^2} \sum_{j=1}^{b_n} (b_n - j) \text{Cov}(R_{1,m_n}(r), R_{1+j,m_n}(r)) \\
 &\leq \frac{1}{b_n} \text{Var}(R_{1,m_n}(r)) + 16r^2 \left(\frac{1}{b_n} \sum_{j=1}^{b_n} \left(1 - \frac{j}{b_n}\right) \alpha_j \right) \\
 &\leq \frac{r^2}{4b_n} + 16r^2 \left(\frac{1}{b_n} \sum_{j=1}^{b_n-1} \alpha_j \right) \\
 &\rightarrow 0,
 \end{aligned} \tag{69}$$

where $\alpha_j := \alpha(\mathcal{F}_{1,m_n}, \mathcal{F}_{1+j,m_n})$ is the strong mixing constant associated the sigma algebras $\sigma(X_1, X_2, \dots, X_{m_n})$, $\sigma(X_{jd_n+1}, X_{jd_n+2}, \dots, X_{jd_n+m_n})$ formed by random variables in batch 1 and batch $1 + j$, the first inequality in (69) follows upon application of Corollary 2.5 in [37, pp. 347] with $u = 1, v = \infty, w = \infty$, the second inequality in (69) follows since $R_{1,m_n}(r) \in [0, r]$, and the last inequality in (69) follows since Assumption 2 implies $\alpha_j \rightarrow 0$ implying in turn that the Césaro sum $b_n^{-1} \sum_{j=1}^{b_n} \alpha_j \rightarrow 0$.

Now by applying (66), (67), (68) and (69) in (63), and since ϵ is arbitrary, we see that (62) holds, that is, $\tilde{\sigma}_{OB-I}^2(m_n, b_n) \xrightarrow{P} \sigma^2$. To complete the first part of the theorem's assertion in (86), we write

$$\begin{aligned}
 \hat{\sigma}_{OB-I}^2(m_n, b_n) &= \frac{1}{b_n} \sum_{j=1}^{b_n} m_n (\hat{\theta}_{i,m_n} - \theta(P))^2 + m_n (\hat{\theta}_n - \theta(P))^2 + \frac{2}{b_n} \sum_{j=1}^{b_n} m_n (\hat{\theta}_{i,m_n} - \theta(P)) (\hat{\theta}_n - \theta(P)) \\
 &= \tilde{\sigma}_{OB-I}^2(m_n, b_n) + \left(\frac{m_n}{n} \right) n (\hat{\theta}_n - \theta(P))^2 \\
 &\quad + 2 \left(\sqrt{\frac{m_n}{n}} \right) \sqrt{n} (\hat{\theta}_n - \theta(P)) \frac{1}{b_n} \sum_{j=1}^{b_n} \sqrt{m_n} (\hat{\theta}_{i,m_n} - \theta(P)).
 \end{aligned} \tag{70}$$

Through prior arguments, we proved that the first term on the right-hand side of (82) tends to σ^2 in probability; also, because $\sqrt{n}(\hat{\theta}_n - \theta(P)) \xrightarrow{d} \sigma N(0, 1)$, and $\beta := \lim_{n \rightarrow \infty} m_n/n = 0$, Slutsky's theorem (A.2) ensures that the second term on the right-hand side of (82) is $o_P(1)$. To see that the third term on the right-hand side of (82) also tends to zero in probability, notice again that $\sqrt{n}(\hat{\theta}_n - \theta(P)) \xrightarrow{d} \sigma N(0, 1)$ and that

$$\sqrt{\frac{m_n}{n}} \mathbb{E} \left[\frac{1}{b_n} \sum_{j=1}^{b_n} \sqrt{m_n} (\hat{\theta}_{i,m_n} - \theta(P)) \right] \leq \sqrt{\frac{m_n}{n}} \frac{1}{b_n} \sum_{j=1}^{b_n} \mathbb{E} \left[\sqrt{m_n} |\hat{\theta}_{i,m_n} - \theta(P)| \right] \rightarrow 0, \tag{71}$$

and make use of Slutsky's theorem (A.2). This proves the first assertion of the theorem in (86).

To prove the second assertion in (86), we again apply Slutsky's theorem (A.2) to

$$T_{OB-I}(m_n, b_n) := \frac{\sqrt{n}(\hat{\theta}_n - \theta(P))}{\hat{\sigma}_{OB-I}(m_n, b_n)} \tag{72}$$

after noticing that the numerator in the expression for $T_{OB-II}(m_n, b_n)$ converges weakly to $\sigma N(0, 1)$ due to Assumption 3 and the denominator converges in probability to σ from the first assertion. \square

We now make a few observations regarding Theorem 5.3.

- (a) Unlike in the large batch setting ($\beta > 0$) of Theorem 5.1, the first assertion of Theorem 5.3 guarantees that $\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)$ is a consistent estimator of σ^2 .
- (b) Unlike Theorem 5.1, Theorem 5.3 does not need Assumption 5 simply due to the fact that σ^2 is being estimated consistently, implying that the dependence between the numerator and the denominator of $T_{\text{OB-I}}(m_n, b_n)$ does not have to be explicitly modeled. This is what allows using Slutsky's theorem in Theorem 5.3.
- (c) Theorem 5.3 assumes very little about the overlapping requirement of the batches apart from requiring the number of batches to diverge.
- (d) As is evident from (82), characterizing the next order term for the mean and variance of $\hat{\sigma}_{\text{OB-I}}^2(m_n, b_n)$ (akin to Theorem 5.2) will involve assuming the nature of higher order terms in the uniform convergence assumption appearing as Assumption 4.

6 THE OB-II LIMIT

In this section, we characterize the weak limit of

$$T_{\text{OB-II}}(m_n, b_n) := \frac{\sqrt{n}(\bar{\theta}_n - \theta(P))}{\hat{\sigma}_{\text{OB-II}}(m_n, b_n)}. \quad (73)$$

As described in Section 3.4, recall that the OB-II limit $T_{\text{OB-II}}(m_n, b_n)$ differs from the OB-I limit in that it replaces the section estimator $\hat{\theta}_n$ with the batching estimator $\bar{\theta}_n$ as the centering variable. As in the OB-I context, the ensuing sections treat the large batch and small batch regimes separately.

6.1 Large Batch ($\beta > 0$) Regime for OB-II

Theorem 6.1 that follows treats the large batch setting ($\beta := \lim_{n \rightarrow \infty} m_n/n > 0$) and asserts that $\hat{\sigma}_{\text{OB-II}}^2(m_n, b_n)/\sigma^2$ and $T_{\text{OB-II}}(m_n, b_n)$ converge weakly to certain functionals of the Wiener process that we denote $\chi_{\text{OB-II}}^2(\beta, b)$ and $T_{\text{OB-II}}(\beta, b)$, respectively. The proof of Theorem 6.1 follows closely along the lines of Theorem 5.1, and we include it in Appendix B.

THEOREM 6.1 (OB-II LARGE BATCH REGIME). *Suppose that Assumptions 1 – 5 hold, and that $\beta := \lim_{n \rightarrow \infty} m_n/n > 0$. Assume also that $b_n \rightarrow b \in \{2, 3, \dots, \infty\}$ as $n \rightarrow \infty$. Define*

$$\chi_{\text{OB-II}}^2(\beta, b) := \begin{cases} \frac{1}{\kappa_2(\beta, \infty)} \frac{\beta^{-1}}{1-\beta} \int_0^{1-\beta} \left(\tilde{W}_u(\beta) - \frac{1}{1-\beta} \int_0^{1-\beta} \tilde{W}_s(\beta) ds \right)^2 du & b = \infty; \\ \frac{1}{\kappa_2(\beta, b)} \frac{1}{\beta} \frac{1}{b} \sum_{j=1}^b \left(\tilde{W}_{c_j}(\beta) - \frac{1}{b} \sum_{i=1}^b \tilde{W}_{c_i}(\beta) \right)^2 & b \in \mathbb{N} \setminus 1, \end{cases} \quad (74)$$

where $\tilde{W}_x(\beta) := W(x + \beta) - W(x)$, $x \in [0, 1 - \beta]$, $\{W(t), t \in [0, 1]\}$ is the standard Brownian motion [5], $c_i := (i - 1) \frac{1-\beta}{b-1}$, $i = 1, 2, \dots, b$, and $\kappa_2(\beta, b)$ is the "bias-correction" factor given by

$$\kappa_2(\beta, b) := \begin{cases} 1 & \beta = 0; \\ 1 - 2 \left(\frac{\beta}{1-\beta} \wedge 1 \right) + \frac{1}{\beta} \left(\frac{\beta}{1-\beta} \wedge 1 \right)^2 - \frac{2}{3} \frac{1-\beta}{\beta} \left(\frac{\beta}{1-\beta} \wedge 1 \right)^3 & \beta > 0, b = \infty; \\ 1 - \frac{1}{b} - \frac{2}{b} \sum_{h=1}^b \left(1 - \frac{h}{b-1} \frac{1-\beta}{\beta} \right)^+ (1 - h/b) & \beta > 0, b \in \mathbb{N} \setminus 1. \end{cases} \quad (75)$$

Then, as $n \rightarrow \infty$,

$$\hat{\sigma}_{\text{OB-II}}^2(m_n, b_n) \xrightarrow{d} \sigma^2 \chi_{\text{OB-II}}^2(\beta, b); \quad (76)$$

and

$$T_{\text{OB-II}}(m_n, b_n) \xrightarrow{d} \begin{cases} \frac{1}{\sqrt{\chi_{\text{OB-II}}^2(\beta, b)}} \frac{1}{\beta} \frac{1}{(1-\beta)} \int_0^{1-\beta} (W(s+\beta) - W(s)) ds & b = \infty; \\ \frac{1}{\sqrt{\chi_{\text{OB-II}}^2(\beta, b)}} \frac{1}{\beta} \frac{1}{b} \sum_{i=1}^b W(c_i + \beta) - W(c_i) & b \in \mathbb{N} \setminus 1, \end{cases} \quad (77)$$

where $c_i := (i-1)\frac{1-\beta}{b-1}$, $i = 1, 2, \dots, b$.

PROOF. Proof See Appendix B. □

We make a number of observations in light of Theorem 6.1.

- (a) As in Theorem 5.1, we see that the variance parameter σ^2 is not estimated consistently in Theorem 6.1. Instead the estimator $\hat{\sigma}_{\text{OB-II}}^2(m_n, b_n)$ converges weakly to the product of σ^2 and $\chi_{\text{OB-II}}^2(\beta, b)$. Again, we slightly abuse notation and define the weak limit appearing in (77) as the $T_{\text{OB-II}}(\beta, b)$ random variable.
- (b) Unlike the the OB-I interval estimator, the OB-II interval estimator uses $\bar{\theta}_n$ as the centering variable and when estimating the variance constant. For this reason, and as we shall briefly discuss later, this makes the OB-II estimator attractive from a computational standpoint.
- (c) Recall that Theorem 5.1 required only Assumptions 1–4 to hold, whereas Theorem 6.1 requires Assumption 5 as well. Theorem 5.1 did not need Assumption 5 because $\hat{\theta}_n$ is the centering variable for the confidence intervals considered in Theorem 5.1 and we have assumed the existence of a CLT on $\hat{\theta}_n$ through Assumption 3. No such (direct) assumption has been made about the centering variable $\bar{\theta}_n$ forcing us to analyze the effect of dependence due to averaging adjacent batches.
- (d) As can be seen, the “bias correction” factor $\kappa_2(\beta, b)$ in (75) for the OB-II context is much more complicated. The OB-II analogue of the OB-I asymptotic variance appearing in (38) of Theorem 5.2 has been elusive.
- (e) The table in Figure 4 displays the critical values $t_{\text{OB-II}, 1-\alpha}(\beta, b) := \min_x P(T_{\text{OB-II}}(\beta, b) \leq x) \geq 1 - \alpha$ associated with the $T_{\text{OB-II}}$ distribution as a function of $1 - \alpha$ and for different values of the parameters β, b . R and MATLAB code for calculating the critical values can be obtained through <https://web.ics.purdue.edu/~pasupath>.

6.2 Small Batch ($\beta = 0$) Regime for OB-II

We now treat the small batch regime ($\beta := \lim_{n \rightarrow \infty} m_n/n = 0$) for OB-II. Like Theorem 5.1, Theorem 6.1 needs the strong invariance Assumption 5 to hold so that the dependence across batches can be characterized.

THEOREM 6.2 (OB-II SMALL BATCH REGIME). *Suppose Assumptions 1–5 hold, and that $\beta = \lim_{n \rightarrow \infty} m_n/n = 0$. Assume that the number of batches $b_n \rightarrow \infty$. Then, as $n \rightarrow \infty$,*

$$\hat{\sigma}_{\text{OB-II}}^2(m_n, b_n) \xrightarrow{p} \sigma^2; \quad \text{and} \quad T_{\text{OB-II}}(m_n, b_n) \xrightarrow{d} N(0, 1). \quad (78)$$

PROOF. Since $\beta = 0$, $\kappa_2(\beta) = 1$ and recall that

$$\hat{\sigma}_{\text{OB-II}}^2(m_n, b_n) := \frac{m_n}{b_n} \sum_{i=1}^{b_n} (\hat{\theta}_{i, m_n} - \bar{\theta}_n)^2; \quad \bar{\theta}_n := \frac{1}{b_n} \sum_{i=1}^{b_n} \hat{\theta}_{i, m_n}. \quad (79)$$

1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225

OB-II CRITICAL VALUES

	1- α	2	3	5	11	b				Inf
						16	26	51	101	
$\beta = 0.010$	0.550	0.158	0.142	0.134	0.129	0.128	0.127	0.126	0.126	0.126
	0.600	0.325	0.289	0.271	0.260	0.258	0.256	0.255	0.254	0.254
	0.650	0.510	0.445	0.414	0.397	0.393	0.390	0.388	0.386	0.386
	0.700	0.727	0.617	0.569	0.542	0.536	0.531	0.528	0.526	0.526
	0.750	1.000	0.816	0.741	0.700	0.691	0.684	0.679	0.677	0.676
	0.800	1.376	1.061	0.941	0.879	0.866	0.856	0.849	0.845	0.844
	0.850	1.963	1.386	1.190	1.093	1.074	1.058	1.047	1.042	1.040
	0.900	3.078	1.886	1.533	1.372	1.341	1.316	1.299	1.290	1.287
	0.950	6.314	2.920	2.132	1.812	1.753	1.708	1.676	1.660	1.655
$\beta = 0.050$	0.550	0.158	0.142	0.134	0.129	0.128	0.127	0.127	0.127	0.127
	0.600	0.325	0.289	0.271	0.260	0.258	0.256	0.256	0.256	0.256
	0.650	0.510	0.445	0.414	0.397	0.393	0.390	0.389	0.389	0.389
	0.700	0.727	0.617	0.569	0.542	0.536	0.532	0.531	0.530	0.530
	0.750	1.000	0.816	0.741	0.700	0.691	0.685	0.684	0.683	0.683
	0.800	1.376	1.061	0.941	0.879	0.866	0.858	0.855	0.854	0.854
	0.850	1.963	1.386	1.190	1.093	1.074	1.061	1.057	1.056	1.056
	0.900	3.078	1.886	1.533	1.372	1.341	1.320	1.314	1.312	1.312
	0.950	6.314	2.920	2.132	1.812	1.753	1.715	1.703	1.700	1.699
$\beta = 0.100$	0.550	0.158	0.142	0.134	0.129	0.128	0.128	0.128	0.128	0.128
	0.600	0.325	0.289	0.271	0.260	0.259	0.259	0.258	0.258	0.258
	0.650	0.510	0.445	0.414	0.397	0.395	0.394	0.394	0.393	0.393
	0.700	0.727	0.617	0.569	0.542	0.539	0.537	0.537	0.537	0.537
	0.750	1.000	0.816	0.741	0.700	0.696	0.694	0.693	0.693	0.693
	0.800	1.376	1.061	0.941	0.880	0.873	0.870	0.869	0.868	0.869
	0.850	1.963	1.386	1.190	1.094	1.084	1.079	1.077	1.077	1.077
	0.900	3.078	1.886	1.533	1.373	1.358	1.349	1.346	1.346	1.346
	0.950	6.314	2.920	2.132	1.815	1.784	1.768	1.762	1.761	1.761
$\beta = 0.200$	0.550	0.158	0.142	0.134	0.131	0.131	0.131	0.131	0.131	0.131
	0.600	0.325	0.289	0.271	0.265	0.264	0.264	0.264	0.264	0.264
	0.650	0.510	0.445	0.414	0.404	0.403	0.403	0.403	0.403	0.403
	0.700	0.727	0.617	0.569	0.553	0.552	0.551	0.551	0.551	0.551
	0.750	1.000	0.816	0.741	0.717	0.715	0.714	0.714	0.714	0.715
	0.800	1.376	1.061	0.941	0.905	0.902	0.901	0.901	0.901	0.901
	0.850	1.963	1.386	1.190	1.132	1.128	1.126	1.126	1.126	1.127
	0.900	3.078	1.886	1.533	1.435	1.428	1.424	1.423	1.424	1.425
	0.950	6.314	2.920	2.132	1.928	1.912	1.904	1.902	1.903	1.905
$\beta = 0.300$	0.550	0.158	0.142	0.135	0.134	0.133	0.133	0.133	0.133	0.133
	0.600	0.325	0.289	0.274	0.270	0.270	0.270	0.270	0.270	0.270
	0.650	0.510	0.445	0.420	0.413	0.413	0.413	0.413	0.413	0.413
	0.700	0.727	0.617	0.578	0.567	0.566	0.566	0.566	0.567	0.567
	0.750	1.000	0.816	0.755	0.739	0.737	0.737	0.737	0.738	0.738
	0.800	1.376	1.061	0.963	0.938	0.935	0.935	0.935	0.935	0.936
	0.850	1.963	1.386	1.224	1.182	1.179	1.178	1.178	1.179	1.180
	0.900	3.078	1.886	1.589	1.515	1.508	1.507	1.507	1.508	1.509
	0.950	6.314	2.920	2.234	2.072	2.056	2.051	2.051	2.052	2.054
$\beta = 0.900$	0.550	0.158	0.145	0.139	0.136	0.136	0.136	0.136	0.136	0.136
	0.600	0.325	0.294	0.282	0.276	0.276	0.275	0.275	0.275	0.275
	0.650	0.510	0.454	0.433	0.424	0.423	0.422	0.422	0.422	0.422
	0.700	0.727	0.633	0.598	0.584	0.582	0.581	0.581	0.581	0.581
	0.750	1.000	0.841	0.786	0.764	0.762	0.760	0.759	0.759	0.759
	0.800	1.376	1.099	1.009	0.976	0.972	0.969	0.968	0.968	0.968
	0.850	1.963	1.447	1.295	1.240	1.233	1.230	1.228	1.227	1.227
	0.900	3.078	1.985	1.701	1.605	1.592	1.586	1.583	1.582	1.582
	0.950	6.314	3.104	2.430	2.221	2.195	2.181	2.175	2.174	2.173
$\beta = 0.950$	0.550	0.158	0.145	0.139	0.136	0.136	0.136	0.136	0.136	0.136
	0.600	0.325	0.294	0.282	0.276	0.276	0.275	0.275	0.275	0.275
	0.650	0.510	0.454	0.433	0.424	0.423	0.422	0.422	0.422	0.422
	0.700	0.727	0.633	0.598	0.584	0.582	0.581	0.581	0.581	0.581
	0.750	1.000	0.841	0.786	0.764	0.762	0.760	0.759	0.759	0.759
	0.800	1.376	1.099	1.009	0.976	0.972	0.969	0.968	0.968	0.968
	0.850	1.963	1.447	1.295	1.240	1.233	1.230	1.228	1.227	1.227
	0.900	3.078	1.985	1.701	1.605	1.593	1.586	1.583	1.582	1.582
	0.950	6.314	3.104	2.429	2.221	2.195	2.181	2.175	2.174	2.173

Fig. 4. T_{OB-II} Critical Values. The table displays critical values $t_{OB-II,1-\alpha} := \inf\{r : P(T_{OB-II}(\beta, b) \leq r) = 1 - \alpha\}$ associated with the OB-II distribution as a function of $1 - \alpha$, the asymptotic batch size β and the asymptotic number of batches b .

Also, define

$$\tilde{\sigma}_{OB-II}^2(m_n, b_n) := \frac{1}{b_n} \sum_{i=1}^{b_n} \underbrace{m_n(\hat{\theta}_{i,m_n} - \theta(P))^2}_{R_{i,m_n}}; \quad \tilde{\sigma}_{OB-II}^2(m_n, d_n; r) := \frac{1}{b_n} \sum_{i=1}^{b_n} \underbrace{R_{i,m_n} \mathbb{I}_{[0,r]}(R_{i,m_n})}_{R_{i,m_n}(r)} \quad (80)$$

From arguments identical to that in the proof of Theorem 5.3 (specifically, (63)–(69)), we see that $\tilde{\sigma}_{\text{OB-II}}^2(m_n, b_n)$ consistently estimates σ^2 , that is,

$$\tilde{\sigma}_{\text{OB-II}}^2(m_n, b_n) \xrightarrow{P} \sigma^2. \quad (81)$$

To complete the first part of the theorem's assertion in (78), we write

$$\begin{aligned} \hat{\sigma}_{\text{OB-II}}^2(m_n, b_n) &= \frac{1}{b_n} \sum_{j=1}^{b_n} m_n (\hat{\theta}_{i, m_n} - \theta(P))^2 + m_n (\bar{\theta}_n - \theta(P))^2 + \frac{2}{b_n} \sum_{j=1}^{b_n} m_n (\hat{\theta}_{i, m_n} - \theta(P)) (\bar{\theta}_n - \theta(P)) \\ &= \tilde{\sigma}_{\text{OB-II}}^2(m_n, b_n) + \left(\frac{m_n}{m_n b_n} \right) m_n b_n (\bar{\theta}_n - \theta(P))^2 \\ &\quad + 2 \left(\sqrt{\frac{m_n}{m_n b_n}} \right) \sqrt{m_n b_n} (\bar{\theta}_n - \theta(P)) \frac{1}{b_n} \sum_{j=1}^{b_n} \sqrt{m_n} (\hat{\theta}_{i, m_n} - \theta(P)). \end{aligned} \quad (82)$$

From (81), we see that the first term on the right-hand side of (82) tends to σ^2 in probability; also, because $\sqrt{n}(\hat{\theta}_n - \theta(P)) \xrightarrow{d} \sigma N(0, 1)$, and $\beta := \lim_{n \rightarrow \infty} m_n/n = 0$, Slutsky's theorem (A.2) ensures that the second term on the right-hand side of (82) is $o_P(1)$. To see that the third term on the right-hand side of (82) also tends to zero in probability, notice again that $\sqrt{n}(\hat{\theta}_n - \theta(P)) \xrightarrow{d} \sigma N(0, 1)$ and that

$$\sqrt{\frac{m_n}{n}} \mathbb{E} \left[\frac{1}{b_n} \sum_{j=1}^{b_n} \sqrt{m_n} (\hat{\theta}_{i, m_n} - \theta(P)) \right] \leq \sqrt{\frac{m_n}{n}} \frac{1}{b_n} \sum_{j=1}^{b_n} \mathbb{E} \left[\sqrt{m_n} |\hat{\theta}_{i, m_n} - \theta(P)| \right] \rightarrow 0, \quad (83)$$

and make use of Slutsky's theorem (A.2). This proves the first assertion of the theorem in (86).

To prove the second assertion in (86), we again apply Slutsky's theorem (A.2) to

$$T_{\text{OB-II}}(m_n, b_n) := \frac{\sqrt{n}(\hat{\theta}_n - \theta(P))}{\hat{\sigma}_{\text{OB-II}}(m_n, b_n)} \quad (84)$$

after noticing that the numerator in the expression for $T_{\text{OB-II}}(m_n, b_n)$ converges weakly to $\sigma N(0, 1)$ due to Assumption 3 and the denominator converges in probability to σ from the first assertion. \square

7 THE OB-III LIMIT

Theorem 7.1 that follows treats the large batch setting ($\beta := \lim_{n \rightarrow \infty} m_n/n > 0$) and asserts that $\hat{\sigma}_{\text{OB-III}}^2(m_n, b_n)/\sigma^2$ and $T_{\text{OB-III}}(m_n, b_n)$ converge weakly to certain functionals of the Wiener process that we denote $\chi_{\text{OB-III}}^2(\beta, b)$ and $T_{\text{OB-III}}(\beta, b)$, respectively. Since the proof of Theorem 7.1 follows closely along the lines of Theorem 5.1 and Theorem 6.1, we do not provide a proof.

THEOREM 7.1 (OB-III LARGE BATCH REGIME). *Suppose that Assumptions 1–5 hold, and that $\beta = \lim_{n \rightarrow \infty} m_n/n \in (0, 1)$. Assume also that $b_n \rightarrow b \in \{2, 3, \dots, \infty\}$ as $n \rightarrow \infty$, and that the weighting function $f : [0, 1] \rightarrow \mathbb{R}^+$ satisfies the stipulations in (23). Define*

$$\chi_{\text{OB-III}}^2(\beta, b) := \begin{cases} \frac{1}{\beta^{-1} - 1} \int_0^{\beta^{-1}-1} \left(\int_0^1 f(v) B_u(v) dv \right)^2 du & b = \infty; \\ \frac{1}{b} \sum_{j=1}^b \left(\int_0^1 f(v) B_{c_j/\beta}(v) dv \right)^2 du & b \in \mathbb{N} \setminus \{1\}, \end{cases}$$

where $c_j := (j - 1) \frac{1-\beta}{b-1}$ and

$$B_s(t) := W(s+t) - W(s) - t(W(s+1) - W(s)), \quad s \in [0, 1-t], t \in [0, 1].$$

Then, as $n \rightarrow \infty$,

$$\hat{\sigma}_{\text{OB-III}}^2(m_n, b_n) \xrightarrow{d} \sigma^2 \chi_{\text{OB-III}}^2(\beta, b); \quad \text{and} \quad T_{\text{OB-III}}(m_n, b_n) \xrightarrow{d} \frac{W(1)}{\sqrt{\chi_{\text{OB-III}}^2(\beta, b)}}. \quad (85)$$

We conclude with a corresponding result in the small batch regime.

THEOREM 7.2 (OB-III SMALL BATCH REGIME). *Suppose Assumptions 1–4 hold, and that $\beta = \lim_{n \rightarrow \infty} m_n/n = 0$. Assume that the asymptotic number of batches $b := \lim_n b_n = \infty$, and that the weighting function $f : [0, 1] \rightarrow \mathbb{R}^+$ satisfies the stipulations in (23). Then, as $n \rightarrow \infty$,*

$$\hat{\sigma}_{\text{OB-III}}^2(m_n, b_n) \xrightarrow{p} \sigma^2; \quad \text{and} \quad T_{\text{OB-III}}(m_n, b_n) \xrightarrow{d} N(0, 1). \quad (86)$$

8 CONSIDERATIONS DURING IMPLEMENTATION

In this section, we discuss “practitioner” questions that seem to arise repeatedly.

8.1 OB Critical Values versus Gaussian or Student’s t Critical Values.

In the absence of the OB-I and OB-II critical value tables on page 17 and page 24 respectively, it has been customary to use critical values from the z -table or the Student’s t table with an appropriate number of degrees of freedom. From a practical standpoint, how much difference does it make if one uses the z -table or the Student’s t table versus the OB critical value table?

When the batch size is large, that is, if $\beta := \lim_n m_n/n > 0$, and when the limiting number of batches $b = \infty$, the OB-I and OB-II critical values correspond to the rightmost columns of the tables appearing on page 17 and page 24, respectively. Looking at these columns, it should be immediately clear that the OB-I, OB-II critical values can be quite different from those of the standard normal distribution. For instance, when $\beta = 0.1$, the 0.95-quantile of the OB-I and OB-II distributions are each around 1.76 whereas the corresponding standard normal quantile $\Phi^{-1}(0.95) = 1.645$, a difference of more than 7%. This difference increases as β increases, and vanishes as $\beta \rightarrow 0$.

When $\beta := \lim_n m_n/n > 0$ but the limiting number of batches $b < \infty$, the natural temptation, in absence of the OB-I and OB-II distributions, might be to use the Student’s t critical value with $b - 1$ degrees of freedom. (Some algebra reveals that when $\beta > 0$, $b < \beta^{-1}$ results in non-overlapping batches and $b \geq \beta^{-1}$ results in overlapping batches.) However, notice again the quantiles reported on pages 17 and 24 can be quite different from the corresponding Student’s t critical value with $b - 1$ degrees of freedom. For instance, when $\beta = 0.2$ and $b = 51$, the 0.95-quantile for the OB-I and OB-II distributions are 1.893 and 1.902 respectively, whereas the 0.95-quantile of the Student’s t distribution with 50 degrees of freedom is 1.6749, a difference of more than 11%. As $\beta \rightarrow 0$ and assuming $b < \infty$, the quantiles of the OB-II distribution converge to those of the Student’s t distribution with $b - 1$ degrees of freedom; the difference between the quantiles of the $T_{\text{OB-I}}$ distribution and those of the Student’s t distribution with $b - 1$ degrees of freedom persist even as $\beta \rightarrow 0$.

In summary, substituting the normal or Student’s t critical value for the OB critical values will not provide the correct coverage unless $\beta = 0$. And, the deviation from the nominal coverage with such substitution can become substantial as the asymptotic batch size β becomes large.

8.2 Which OB CIP?

We’ve presented three statistics along with their weak convergence limits OB-x, x=I,II,III, amounting to three possible CIPs. Numerical evidence to be provided in the ensuing section suggests that

Table 2. Time complexities of the three OB CIPs. Recall that n represents the size of the dataset, m_n represents the batch size, b_n represents the number of batches, and $\tau(i)$ is the time complexity of constructing the estimator of the statistical functional $\theta(P)$ using a batch of size i .

CIP	Time Complexity
OB-I	$O(\tau(n) + b_n \tau(m_n))$
OB-II	$O(b_n \tau(m_n))$
OB-III	$O(\tau(n) + b_n \sum_{i=2}^{m_n} \tau(i))$

using these CIPs with large overlapping batches tends to result in confidence intervals having good behavior across a variety of contexts. How do the OB CIPs compare against each other?

Unfortunately, providing a satisfactory answer appears to be context-dependent and requires much further investigation, especially around the question of batch size choice. The sectioning estimator $\hat{\theta}_n$ used within the OB-I CIP typically has variance $O(\frac{1}{n})$ and bias $O(\frac{1}{n^\lambda})$ for some $\lambda \geq 1/2$, whereas the batching estimator $\hat{\theta}_n$ used within OB-II has typical variance $O(\frac{1}{b_n m_n})$ and bias $O(\frac{1}{m_n^\lambda})$. These expressions reveal that the batching estimator has lower variance (when using overlapping batches) and higher bias than the sectioning estimator; how these collude to decide the quality of the resulting confidence intervals is a context-dependent question.

In summary, from the standpoint of interval quality as assessed by coverage probability and expected half-width, little is known theoretically on the relative behavior of OB-x, $x=I,II,III$ especially when implemented with their corresponding optimal batch sizes. This should form the agenda for future investigation.

The difference between the proposed procedures is much clearer from the standpoint of computational complexity. Suppose $\tau(|u - \ell|)$ is the time complexity of calculating the estimator $\hat{\theta}(\{X_j, \ell \leq j \leq u\})$ described in Section 3.1. Then, as can be seen in Table 2, simple calculations reveal that OB-II CIP is the most computationally efficient and the OB-III CIP the least computationally efficient. The relative complexities of the three CIPs become stark when using large batches with significant overlap, that is, when $m_n/n \rightarrow \beta > 0$ and $d_n = O(1)$. This leads to $O(n\tau(n))$ complexity for OB-I and OB-II, but $O(n \sum_{i=2}^{m_n} \tau(i))$ complexity for OB-III. With sparse overlap resulting in finite number of asymptotic batches, that is, if $b < \infty$, OB-I has complexity $O(\tau(n))$, OB-II has complexity $O(\tau(m_n))$, and OB-III has complexity $O(\tau(n) + \sum_{i=2}^{m_n} \tau(i))$.

9 NUMERICAL ILLUSTRATION

We now present numerical results from three popular contexts to gain further insight on the behavior of confidence intervals produced by OB-I, OB-II, and subsampling.

9.1 Example 1 : CVaR Estimation.

Let θ_γ the CVaR associated with the standard normal random variable. From the definition of CVaR [69], we have

$$\theta_\gamma := \frac{1}{1-\gamma} \int_{q_\gamma}^{\infty} z \phi(z) dz; \quad q_\gamma := \Phi^{-1}(\gamma),$$

where $\phi(\cdot)$, $\Phi(\cdot)$ are the standard normal density and cdf, respectively. With observations from an iid sequence $\{Z_n, n \geq 1\}$ of standard normal random variables, we can construct a point estimator

Table 3. The table summarizes coverage probabilities obtained using small batch OB-I ($\beta = 0$), large batch OB-I ($\beta = 0.1, 0.25$), small batch OB-II ($\beta = 0$), large batch OB-II ($\beta = 0.1, 0.25$), and subsampling (SS) for the CVaR problem with $\gamma = 0.7$. The numbers in parenthesis are estimated expected half-widths of the confidence intervals.

	OB-I ($\beta = 0, 0.1, 0.25$)			OB-II ($\beta = 0, 0.1, 0.25$)			SS ($m_n = \sqrt{n}$)
$n = 100$	0.390 (0.184)	0.402 (0.212)	0.947 (0.233)	0.382 (0.182)	0.397 (0.213)	0.935 (0.236)	0.388 (0.205)
$n = 500$	0.897 (0.085)	0.954 (0.091)	0.950 (0.099)	0.887 (0.085)	0.944 (0.091)	0.935 (0.099)	0.907 (0.097)
$n = 1000$	0.946 (0.060)	0.952 (0.063)	0.951 (0.070)	0.938 (0.060)	0.944 (0.063)	0.938 (0.071)	0.956 (0.066)
$n = 2000$	0.950 (0.042)	0.952 (0.044)	0.951 (0.049)	0.943 (0.042)	0.945 (0.045)	0.936 (0.050)	0.961 (0.045)
$n = 3000$	0.949 (0.034)	0.950 (0.036)	0.950 (0.040)	0.945 (0.034)	0.944 (0.036)	0.937 (0.040)	0.961 (0.037)
$n = 5000$	0.950 (0.026)	0.952 (0.028)	0.954 (0.032)	0.947 (0.026)	0.943 (0.028)	0.936 (0.031)	0.961 (0.028)

for θ_γ as follows:

$$\hat{\theta}(\{X_j, \ell \leq j \leq u\}) := \frac{1}{1-\gamma} \sum_{j=\ell}^u Z_j \mathbb{I}_{[q_\gamma, \infty)}(Z_j).$$

We wish to construct a 0.95-confidence interval on θ_γ for $\gamma = 0.7, 0.9, 0.95$ with number of observations $n = 100, 500, 1000, 2000, 3000$ and 5000 .

Tables 3–5 display the estimated coverage probability (along with the estimated expected half-width in parenthesis) of confidence intervals constructed using fully-overlapping OB-I, OB-II CIPs having asymptotic batch size $\beta = 0, 0.1, 0.25$, and using subsampling with the recommended [67] sample size $m_n = \sqrt{n}$. The coverage was estimated with a large ($m = 100000$) number of replications.

Tables 3–5 display clear trends that will be repeated, more or less, across the different experiments we present. All methods seem to tend to the nominal coverage as the available data increases. However, OB-I and OB-II with $\beta > 0$ seem to get to the nominal coverage much faster than the rest. For example, in Table 3, OB-I and OB-II with $\beta = 0.25$ seem to get to the vicinity of the nominal coverage after only about $n = 100$ observations; and OB-I and OB-II with $\beta = 0.1$ seem to get to the vicinity of the nominal coverage after about $n = 500$ observations. Similarly, in Table 5, OB-I and OB-II with $\beta = 0.25$ seem to get to the vicinity of the nominal coverage after about $n = 1000$ observations, while for $\beta = 0.1$, the corresponding number is $n = 2000$. The performance of OB confidence intervals with small batches seems comparable to that of subsampling; both OB-x with $\beta = 0$ and subsampling seem to struggle on the CVaR problem with $\gamma = 0.95$.

Table 4. The table summarizes coverage probabilities obtained using small batch OB-I ($\beta = 0$), large batch OB-I ($\beta = 0.1, 0.25$), small batch OB-II ($\beta = 0$), large batch OB-II ($\beta = 0.1, 0.25$), and subsampling (SS) for the CVaR problem with $\gamma = 0.9$. The numbers in parenthesis are estimated expected half-widths of the confidence intervals.

	OB-I ($\beta = 0, 0.1, 0.25$)			OB-II ($\beta = 0, 0.1, 0.25$)			SS ($m_n = \sqrt{n}$)
$n = 100$	0.000 (0.196)	0.000 (0.227)	0.444 (0.311)	0.000 (0.195)	0.000 (0.228)	0.437 (0.314)	0.000 (0.233)
$n = 500$	0.001 (0.110)	0.755 (0.135)	0.949 (0.142)	0.001 (0.110)	0.744 (0.134)	0.932 (0.141)	0.001 (0.142)
$n = 1000$	0.014 (0.084)	0.953 (0.091)	0.950 (0.098)	0.014 (0.084)	0.944 (0.091)	0.937 (0.100)	0.014 (0.101)
$n = 2000$	0.134 (0.062)	0.952 (0.063)	0.950 (0.069)	0.132 (0.062)	0.945 (0.063)	0.935 (0.069)	0.136 (0.073)
$n = 3000$	0.350 (0.050)	0.951 (0.050)	0.948 (0.055)	0.345 (0.050)	0.944 (0.051)	0.936 (0.056)	0.356 (0.058)
$n = 5000$	0.705 (0.039)	0.952 (0.039)	0.952 (0.044)	0.698 (0.039)	0.943 (0.039)	0.935 (0.043)	0.717 (0.044)

9.2 Example 2: Parameter Estimation for AR(1).

Consider the AR(1) process given by

$$X_t = c + \phi X_{t-1} + \epsilon_t, \quad \epsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma_\epsilon^2), \quad t = 1, 2, \dots$$

With observations from the time series $\{X_n, n \geq 1\}$, the least-squares point estimator for $\theta(P) := \phi$ (after fixing $c = 0$) is

$$\hat{\theta}(\{X_j, \ell \leq j \leq u\}) := \arg \min_{\phi \in \mathbb{R}} \sum_{j=\ell}^{u-1} (X_{j+1} - \phi X_j)^2. \tag{87}$$

We wish to construct a 0.95-confidence interval on $\phi = 0.5, 0.9$ for $\sigma_\epsilon = 1$ and with number of observations $n = 100, 500, 1000, 5000$ and 10000 .

Tables 6–7 are in the same format as Tables 3–5 and display the results for the AR(1) example. The trends in coverage probabilities appear to be similar to those observed in Example 1, with large batches playing a seemingly important role in ensuring close to nominal coverage. Interestingly, Example 2 seems to do a better job in distinguishing between OB-I and OB-II for the same β , and in distinguishing between OB methods and subsampling. For example, due to the increased estimator bias associated with $\phi = 0.9$, OB-I with $\beta = 0, 0.1, 0.25$ appear to dominate OB-II with corresponding $\beta = 0, 0.1, 0.25$. Subsampling clearly generates intervals with smaller expected half-width but the coverage is substantially lower than nominal especially when $\phi = 0.9$. Such differences were not as evident in Example 1, probably because of the more muted effects of bias.

Table 5. The table summarizes coverage probabilities obtained using small batch OB-I ($\beta = 0$), large batch OB-I ($\beta = 0.1, 0.25$), small batch OB-II ($\beta = 0$), large batch OB-II ($\beta = 0.1, 0.25$), and subsampling (SS) for the CVaR problem with $\gamma = 0.95$. The numbers in parenthesis are estimated expected half-widths of the confidence intervals.

	OB-I ($\beta = 0, 0.1, 0.25$)	OB-II ($\beta = 0, 0.1, 0.25$)	SS ($m_n = \sqrt{n}$)
$n = 100$	NA NA 0.082 (NA) (NA) (0.322)	NA NA 0.080 (NA) (NA) (0.327)	NA (NA)
$n = 500$	NA 0.096 0.915 (NA) (0.161) (0.189)	NA 0.094 0.899 (NA) (0.160) (0.188)	NA (NA)
$n = 1000$	NA 0.725 0.948 (NA) (0.121) (0.129)	NA 0.715 0.936 (NA) (0.121) (0.130)	NA (NA)
$n = 2000$	NA 0.953 0.951 (NA) (0.083) (0.089)	NA 0.943 0.935 (NA) (0.083) (0.089)	NA (NA)
$n = 3000$	0.000 0.952 0.948 (0.061) (0.066) (0.071)	0.000 0.943 0.934 (0.061) (0.066) (0.072)	0.000 (0.069)
$n = 5000$	0.000 0.953 0.952 (0.050) (0.051) (0.056)	0.000 0.943 0.934 (0.050) (0.050) (0.056)	0.000 (0.059)

9.3 Example 3: Non-Homogeneous Poisson Process (NHPP) Rate Estimation

In the final example, we consider a nonhomogeneous Poisson process $\{N(t), t \in [0, 1]\}$ [12] having rate $\lambda(t) = 4 + 8t, t \in [0, 1]$. Suppose also that we have iid realizations of the process $\{N(t), t \in [0, 1]\}$, using which we wish to construct a 0.95-confidence interval on

$$\theta_t(P) := \lambda(t) \text{ for } t = 0.25, 0.5, 0.75.$$

We emphasize that this problem constructs “marginal confidence intervals” and is different from that of identifying a confidence region on the vector $(\lambda(0.25), \lambda(0.5), \lambda(0.75))$ or on the function $\lambda(t), t \in [0, 1]$. The latter two problems, while very useful, lie outside the current paper’s scope of real-valued $\theta(P)$.

Given iid realizations $\{X_j(t), t \in [0, 1]\}, j = 1, 2, \dots$, of $\{N(t), t \in [0, 1]\}$, a simple point estimator for $\lambda(t)$ (t fixed) can be constructed as follows:

$$\hat{\theta}_t(\{X_j, \ell \leq j \leq u\}) := \frac{1}{u - \ell + 1} \sum_{j=\ell}^u \frac{1}{\delta} (X_j(t + \delta) - X_j(t)).$$

The realizations $X_j, j = 1, 2, \dots$ were generated using Algorithm 6 in [63, 64], the constant δ was fixed at 10^{-4} , and the number of observations $n = 1000, 2000, 5000, 10000, 20000$ and 50000 .

Table 8 presents results on coverage probability delivered by OB-I (with $\beta = 0.1, 0.25$) and subsampling (with $m_n = \sqrt{n}$) for each of the three “marginal” confidence intervals associated with $t = 0.25, 0.5, 0.75$. As in previous examples, the numbers in parenthesis refer to the estimated expected half-width.

Table 6. The table summarizes coverage probabilities obtained using small batch OB-I ($\beta = 0$), large batch OB-I ($\beta = 0.25$), small batch OB-II ($\beta = 0$), large batch OB-II ($\beta = 0.25$), and subsampling (SS) for the AR(1) problem with $\phi = 0.5, c = 0, \sigma_\epsilon^2 = 1$. The numbers in parenthesis are estimated expected half-widths of the confidence intervals.

	OB-I ($\beta = 0, 0.1, 0.25$)	OB-II ($\beta = 0, 0.1, 0.25$)	SS ($m_n = \sqrt{n}$)
$n = 100$	0.898 0.931 0.932 (0.191) (0.222) (0.245)	0.321 0.420 0.833 (0.154) (0.180) (0.230)	0.644 (0.110)
$n = 500$	0.931 0.947 0.944 (0.089) (0.100) (0.110)	0.366 0.821 0.908 (0.081) (0.095) (0.109)	0.831 (0.071)
$n = 1000$	0.940 0.949 0.946 (0.064) (0.070) (0.078)	0.376 0.883 0.924 (0.059) (0.069) (0.079)	0.866 (0.053)
$n = 5000$	0.946 0.951 0.948 (0.029) (0.031) (0.035)	0.419 0.933 0.931 (0.028) (0.031) (0.035)	0.912 (0.026)
$n = 10000$	0.949 0.947 0.950 (0.021) (0.022) (0.025)	0.425 0.935 0.936 (0.020) (0.022) (0.025)	0.925 (0.019)

Table 7. The table summarizes coverage probabilities obtained using small batch OB-I ($\beta = 0$), large batch OB-I ($\beta = 0.25$), small batch OB-II ($\beta = 0$), large batch OB-II ($\beta = 0.25$), and subsampling (SS) for the AR(1) problem with $\phi = 0.9, c = 0, \sigma_\epsilon^2 = 1$. The numbers in parenthesis are estimated expected half-widths of the confidence intervals.

	OB-I ($\beta = 0, 0.1, 0.25$)	OB-II ($\beta = 0, 0.1, 0.25$)	SS ($m_n = \sqrt{n}$)
$n = 100$	0.698 0.745 0.799 (0.630) (0.726) (0.925)	NA NA 0.332 (0.246) (0.288) (0.593)	0.045 (0.005)
$n = 500$	0.859 0.901 0.912 (0.365) (0.437) (0.505)	NA 0.260 0.780 (0.195) (0.323) (0.452)	0.282 (0.086)
$n = 1000$	0.900 0.929 0.934 (0.272) (0.320) (0.365)	NA 0.546 0.859 (0.169) (0.276) (0.350)	0.438 (0.096)
$n = 5000$	0.924 0.942 0.945 (0.131) (0.148) (0.165)	NA 0.861 0.919 (0.105) (0.144) (0.164)	0.743 (0.083)
$n = 10000$	0.935 0.945 0.946 (0.094) (0.104) (0.116)	0.000 0.900 0.923 (0.081) (0.103) (0.117)	0.816 (0.068)

Table 8. The table summarizes coverage probabilities obtained using large batch OB-I ($\beta = 0.1, 0.25$) and subsampling for the NHPP rate estimation problem with $\lambda(t) = 4 + 8t$. The three numbers in each column indicate the coverage probability estimates corresponding to a confidence interval on $\lambda(t)$ for $t = 0.25, 0.5, 0.75$. The numbers in parenthesis are estimated expected half-widths of the confidence intervals.

	OB-I ($\beta = 0.1, 0.25$)	Subsampling ($m_n = \sqrt{n}$)
$n = 1000$	0.447, 0.446; 0.542, 0.542; 0.629, 0.628 (19.913, 78.657); (28.100, 106.502); (35.863, 135.759)	0.312; 0.337; 0.349 (NA); (NA); (NA)
$n = 2000$	0.694, 0.692; 0.793, 0.790; 0.863, 0.859 (15.505, 42.562); (21.772, 57.227); (28.120, 70.158)	NA; 0.249; 0.440 (NA); (NA); (NA)
$n = 5000$	0.937, 0.939; 0.912, 0.969; 0.962, 0.976 (11.582, 20.017); (15.546, 24.972); (19.767, 27.624)	0.606; 0.548; 0.500 (NA); (NA); (NA)
$n = 10000$	0.940, 0.966; 0.985, 0.974; 0.987, 0.974 (8.686, 11.062); (11.210, 12.073); (13.189, 12.922)	0.461; 0.629; 0.576 (NA); (NA); (NA)
$n = 20000$	0.989, 0.967; 0.989, 0.963; 0.988, 0.962 (5.917, 5.634); (6.910, 6.053); (7.730, 6.520)	0.534; 0.620; 0.574 (1.358); (1.563); (1.744)
$n = 50000$	0.978, 0.957; 0.973, 0.959; 0.970, 0.955 (2.963, 2.845); (3.319, 3.266); (3.547, 3.537)	0.595; 0.537; 0.611 (0.877); (1.007); (1.220)

The trends in Table 8 are consistent with those from the previous examples with OB-I delivering confidence intervals that are clearly better in terms of coverage, although nominal coverage seems to need a higher value of n than in previous examples. Subsampling does not reach nominal coverage even with $n = 50000$ although the generated intervals have much smaller half-widths.

ACKNOWLEDGMENTS.

This work was greatly influenced by conversations and original ideas of Bruce Schmeiser, especially in viewing the proposed statistics in analogy with the χ^2_v and Student's t distributions. Raghu Pasupathy also gratefully acknowledges the Office of Naval Research for support provided through the grants N000141712295 and 13000991.

REFERENCES

[1] T. Aktaran-Kalayci, C. Alexopoulos, D. Goldsman, and J. R. Wilson. 2009. Optimal Linear Combinations of Overlapping Variance Estimators for Steady-State Simulation. In *Advancing the Frontiers of Simulation*, C. Alexopoulos, D. Goldsman, and J. R. Wilson (Eds.). Springer, NY.

[2] Christos Alexopoulos, Nilay Tanik Argon, David Goldsman, Gamze Tokol, and James R Wilson. 2007. Overlapping Variance Estimators for Simulation. *Operations research* 55, 6 (2007), 1090–1103.

[3] S. Asmussen and P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. Springer, New York, NY.

[4] M.S. Bartlett. 1950. Periodogram analysis and continuous spectra. *Biometrika* 37 (1950), 1–16.

[5] P. Billingsley. 1995. *Probability and Measure*. Wiley, New York, NY.

[6] Patrick Billingsley. 1999. *Convergence of probability measures* (2nd ed. ed.). Wiley, New York.

[7] A Bose. 1988. Edgeworth Correction by Bootstrap in Autoregressions. *The Annals of statistics* 16, 4 (1988), 1709–1722.

- 1618 [8] A Bose. 1990. Bootstrap in moving average models. *Annals of the Institute of Statistical Mathematics* 42, 4 (1990),
1619 753–768.
- 1620 [9] D. R. Brillinger. 2001. *Time series: data analysis and theory*. SIAM.
- 1621 [10] James Calvin and Marvin Nakayama. 2013. Confidence intervals for quantiles with standardized time series. In
1622 *Proceedings of the 2013 Winter Simulation Conference (WSC '13)*. IEEE Press, 601–612.
- 1623 [11] James M Calvin and Marvin K Nakayama. 2006. Permuted Standardized Time Series for Steady-State Simulations.
1624 *Mathematics of operations research* 31, 2 (2006), 351–368.
- 1625 [12] E. Çinlar. 1975. *Introduction to Stochastic Processes*. Prentice-Hall, New Jersey.
- 1626 [13] F. Chu and M. K. Nakayama. 2012. Confidence intervals for quantiles when applying variance-reduction techniques.
1627 *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 22, 2 (2012), 1–25.
- 1628 [14] R. W. Conway. 1963. Some tactical problems in digital simulation. *Management science* 10, 1 (1963), 47–61.
- 1629 [15] M.A. Crane and A. J. Lemoine. 1977. *An Introduction to the Regenerative Method for Simulation Analysis* (1st ed. 1977.
1630 ed.). Springer Berlin Heidelberg, Berlin, Heidelberg.
- 1631 [16] J. D. Cryer and K-S. Chan. 2008. *Time Series Analysis With Applications in R* (2nd ed. 2008. ed.). Springer New York,
1632 New York, NY.
- 1633 [17] M. Csörgö and Révész. 1981. *Strong approximations in probability and statistics*. Academic Press, New York.
- 1634 [18] Halim Damerджи. 1991. Strong Consistency and Other Properties of the Spectral Variance Estimator. *Management*
1635 *Science* 37, 11 (1991), 1424–1440.
- 1636 [19] Halim Damerджи. 1994. Strong Consistency of the Variance Estimator in Steady-State Simulation Output Analysis.
1637 *Mathematics of Operations Research* 19, 2 (1994), 494–512.
- 1638 [20] Halim Damerджи. 1995. Mean-Square Consistency of the Variance Estimator in Steady-State Simulation Output Analysis.
1639 *Mathematics of Operations Research* 20, 2 (1995), 282–291.
- 1640 [21] A. DasGupta. 2011. *Probability for Statistics and Machine Learning*. Springer.
- 1641 [22] A. C. (Anthony Christopher) Davison. 1997. *Bootstrap methods and their application*. Cambridge University Press,
1642 Cambridge ;.
- 1643 [23] J. Dedecker and F Merlevede. 2022. Central limit theorem and almost sure results for the empirical estimator of
1644 superquantiles/CVaR in the stationary case. *Statistics* 0, 0 (2022), 1–20. <https://doi.org/10.1080/02331888.2022.2043325>
1645 arXiv:<https://doi.org/10.1080/02331888.2022.2043325>
- 1646 [24] Thomas J Diccio and Joseph P Romano. 1988. A review of bootstrap confidence intervals. *Journal of the Royal*
1647 *Statistical Society: Series B (Methodological)* 50, 3 (1988), 338–354.
- 1648 [25] H. Dong and M. K. Nakayama. 2014. Constructing confidence intervals for a quantile using batching and sectioning
1649 when applying Latin hypercube sampling. In *Proceedings of the Winter Simulation Conference 2014*. IEEE, 640–651.
- 1650 [26] Hui Dong and Marvin K Nakayama. 2018. A tutorial on quantile estimation via Monte Carlo. In *International Conference*
1651 *on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*. Springer, 3–30.
- 1652 [27] Hui Dong and Marvin K Nakayama. 2020. A Tutorial on Quantile Estimation via Monte Carlo. In *Monte Carlo and*
1653 *Quasi-Monte Carlo Methods*. Springer International Publishing, Cham, 3–30.
- 1654 [28] R. Durrett. 2010. *Probability: Theory and Examples*. Cambridge University Press, New York, NY.
- 1655 [29] Bradley Efron. 1979. Bootstrap Methods: Another Look at the Jackknife. In *Breakthroughs in Statistics*. Springer New
1656 York, New York, NY, 569–593.
- 1657 [30] Bradley Efron. 1981. Nonparametric standard errors and confidence intervals. *canadian Journal of Statistics* 9, 2 (1981),
1658 139–158.
- 1659 [31] Bradley Efron. 1982. *The jackknife, the bootstrap and other resampling plans*. SIAM.
- 1660 [32] Bradley Efron. 1985. Bootstrap confidence intervals for a class of parametric problems. *Biometrika* 72, 1 (1985), 45–58.
- 1661 [33] Bradley Efron. 1987. Better bootstrap confidence intervals. *Journal of the American statistical Association* 82, 397 (1987),
1662 171–185.
- 1663 [34] Bradley Efron. 1992. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*. Springer, 569–593.
- 1664 [35] B. Efron and R. Tibshirani. 1994. *An introduction to the bootstrap*. Chapman & Hall, New York.
- 1665 [36] B. Efron and R. J. Tibshirani. 1998. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca Raton, FL.
- 1666 [37] S. N. Ethier and T. G. Kurtz. 2009. *Markov processes: Characterization and Convergence*. John Wiley & Sons.
- [38] G. S. Fishman. 1978. Grouping observations in digital simulation. *Management Science* 24, 5 (1978), 510–521.
- [39] Robert Foley and David Goldsman. 1999. Confidence intervals using orthonormally weighted standardized time series.
9, 4 (1999), 297–325.
- [40] J. L. Gastwirth and H. Rubin. 1975. The behavior of robust estimators on dependent data. *The Annals of Statistics*
(1975), 1070–1100.
- [41] Evarist Giné and Joel Zinn. 1989. Necessary conditions for the bootstrap of the mean. *The annals of statistics* (1989),
684–691.

- 1667 [42] Peter Glynn and Donald Iglehart. 1985. Large-sample theory for standardized time series: an overview. In *Proceedings*
1668 *of the 17th conference on winter simulation (WSC '85)*. ACM, 129–134.
- 1669 [43] Peter W Glynn. 1996. Importance sampling for Monte Carlo estimation of quantiles. In *Mathematical Methods in*
1670 *Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation*. Citeseer,
1671 180–185.
- 1672 [44] P. W. Glynn. 1998. Strong approximations in queueing theory. In *Asymptotic methods in probability and statistics*.
1673 Elsevier, 135–150.
- 1674 [45] P. W. Glynn and D. L. Iglehart. 1988. A New Class of Strongly Consistent Variance Estimators Simulations. *Stochastic*
1675 *Processes and Their Applications* 28 (1988), 71–80.
- 1676 [46] Peter W Glynn and Donald L Iglehart. 1990. Simulation Output Analysis Using Standardized Time Series. 15, 1 (1990),
1677 1–16.
- 1678 [47] D. Goldsman, M. Meketon, and L. W. Schruben. 1990. Properties of standardized time series weighted area variance
1679 estimators. *Management Science* 36, 5 (1990), 602–612.
- 1680 [48] David Goldsman and Lee Schruben. 1990. Note—New Confidence Interval Estimators Using Standardized Time Series.
1681 36, 3 (1990), 393–397.
- 1682 [49] D. Grabaskas, M. K. Nakayama, R. Denning, and T. Aldemir. 2016. Advantages of variance reduction techniques in
1683 establishing confidence intervals for quantiles. *Reliability Engineering & System Safety* 149 (2016), 187–203.
- 1684 [50] Peter Hall. 1992. *Principles of Edgeworth Expansion*. Springer New York, New York, NY, 39–81. [https://doi.org/10.](https://doi.org/10.1007/978-1-4612-4384-7_2)
1685 [1007/978-1-4612-4384-7_2](https://doi.org/10.1007/978-1-4612-4384-7_2)
- 1686 [51] J. A. Hartigan. 1969. Using Subsample Values as Typical Values. *J. Amer. Statist. Assoc.* 64, 328 (1969), 1303–1317.
- 1687 [52] J. A. Hartigan. 1975. Necessary and Sufficient Conditions for Asymptotic Joint Normality of a Statistic and Its Subsample
1688 Values. *The Annals of statistics* 3, 3 (1975), 573–580.
- 1689 [53] D. L. Iglehart. 1978. The regenerative method for simulation analysis. In *Handbook of Optimization in Medicine*, K.M.
1690 Chandy and R.T. Yeh (Eds.). Prentice-Hall, Englewood Cliffs, N.J.
- 1691 [54] Henry Lam. 2022. A Cheap Bootstrap Method for Fast Inference. <https://doi.org/10.48550/ARXIV.2202.00090>
- 1692 [55] E. L. (Erich Leo) Lehmann. 1999. *Elements of large-sample theory*. Springer, New York.
- 1693 [56] Wei-Yin Loh. 1987. Calibrating confidence coefficients. *J. Amer. Statist. Assoc.* 82, 397 (1987), 155–162.
- 1694 [57] P. C. Mahalanobis. 1946. Sample Surveys of Crop Yields in India. *Sankhyā the Indian journal of statistics* 7, 3 (1946),
1695 269–280.
- 1696 [58] P. J McCarthy. 1969. Pseudo-Replication: Half Samples. *Revue de l'Institut international de statistique* 37, 3 (1969), 239.
- 1697 [59] H. Mechanic and W. McKay. [n.d.]. *Confidence intervals for averages of dependent data in simulations II*. Technical
1698 Report.
- 1699 [60] David Fernando Munoz. 1991. Cancellation methods in the analysis of simulation output.
- 1700 [61] M. K. Nakayama. 2011. Asymptotically valid confidence intervals for quantiles and values-at-risk when applying Latin
1701 hypercube sampling. *International Journal on Advances in Systems and Measurements* 4 (2011).
- 1702 [62] Marvin K Nakayama. 2014. Confidence intervals for quantiles using sectioning when applying variance-reduction
1703 techniques. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 24, 4 (2014), 1–21.
- 1704 [63] Raghu Pasupathy. 2011. Generating homogeneous Poisson processes. *Wiley encyclopedia of operations research and*
1705 *management science* (2011).
- 1706 [64] Raghu Pasupathy. 2011. Generating nonhomogeneous Poisson processes. *Wiley encyclopedia of operations research and*
1707 *management science* (2011).
- 1708 [65] W. Philipp and W. Stout. 1975. Almost Sure Invariance Principles for Partial Sums of Weakly Dependent Random
1709 Variables. *Mem. Amer. Math. Soc* 161 (1975).
- 1710 [66] D. N. Politis and J. P. Romano. 1994. Large Sample Confidence Regions Based on Subsamples under Minimal Assump-
1711 tions. *The Annals of statistics* 22, 4 (1994), 2031–2050.
- 1712 [67] D. N. Politis, J. P. Romano, and M. Wolf. 1999. *Subsampling* (1st ed. 1999. ed.).
- 1713 [68] M. H. Quenouille. 1949. Approximate Tests of Correlation in Time-Series. *Journal of the Royal Statistical Society. Series*
1714 *B, Methodological* 11, 1 (1949), 68–84.
- 1715 [69] S. Sarykalin, G. Serraino, and S. Uryasev. 2008. Value-at-risk vs. conditional value-at-risk in risk management and
optimization. In *State-of-the-art decision-making tools in the information-intensive age*. Informs, 270–294.
- [70] L. W. Schruben. 1983. Confidence Interval Estimation Using Standardized Time Series. *Operations Research* 31, 6 (1983),
1090–1108.
- [71] R. J. Serfling. 1980. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Inc., New York, New York.
- [72] S. A. van de Geer. 2006. *Empirical Processes in M-Estimation* (1-st ed.). Cambridge University Press, Cambridge.
- [73] P.D. Welch. 1967. The use of the Fast Fourier Transform for the estimation of spectra; a method based on time averaging
over short modified periodograms. *IEEE Transactions on Audio and Electroacoustics* 2 (1967), 70–73.

[74] Jon Wellner. 2022. Jon Wellner’s Lecture Notes in Mathematical Statistics. <https://sites.stat.washington.edu/peter/581/jaw/jaw.html>. [Online; accessed 02-March-2022].

A SOME USEFUL RESULTS

We will invoke the following useful result from [28] that provides a weak law for triangular arrays of real-valued random variables that are not necessarily identically distributed.

THEOREM A.1 (WEAK LAW FOR TRIANGULAR ARRAYS, THEOREM 2.2.6, [28]). *For each n , let $X_{n,k}$, $1 \leq k \leq n$ be independent. Let $t_n > 0$ with $t_n \rightarrow \infty$ and let $\bar{X}_{n,k} = X_{n,k} \mathbb{1}(|X_{n,k}| \leq t_n)$. Suppose that as $n \rightarrow \infty$,*

- (1) $\sum_{k=1}^n P(|X_{n,k}| > t_n) \rightarrow 0$; and
- (2) $t_n^{-2} \sum_{k=1}^n \mathbb{E}[\bar{X}_{n,k}^2] \rightarrow 0$.

If we let $S_n = X_{n,1} + X_{n,2} + \dots + X_{n,n}$ and put $\mu_n = \sum_{k=1}^n \mathbb{E}[\bar{X}_{n,k}]$, then $t_n^{-1}(S_n - \mu_n) \xrightarrow{p} 0$.

THEOREM A.2 (SLUTSKY’S THEOREM, SEE PAGE 19 IN [71]). *Suppose $\{A_n, n \geq 1\}$, $\{B_n, n \geq 1\}$ and $\{X_n, n \geq 1\}$ are real-valued random sequences so that*

$$X_n \xrightarrow{d} X; \quad A_n \xrightarrow{p} A; \quad B_n \xrightarrow{p} B.$$

Then

$$A_n X_n + B_n \xrightarrow{d} AX + B.$$

If $A \neq 0$, then

$$\frac{X_n}{A_n} \xrightarrow{d} \frac{X}{A}.$$

THEOREM A.3 (COVARIANCE BOUND, SEE COROLLARY 2.5, ETHIER AND KURTZ [37]). *Let $1 \leq u, v, w \leq \infty$, $u^{-1} + v^{-1} + w^{-1} = 1$. Then for real-valued Y, Z with $Y \in L^w(\Omega, \mathcal{G}, P)$, $Z \in L^v(\Omega, \mathcal{H}, P)$,*

$$|\mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]| \leq 2^{v \wedge w \wedge 2+1} \alpha^{1/u}(\mathcal{G}, \mathcal{H}) \|Y\|_v \|Z\|_w.$$

THEOREM A.4 (SEE THEOREM 1.2.1, [17]). *Let $\{W(t), 0 \leq t < \infty\}$ denote the Wiener process. If $\{a_n, n \geq 1\}$ is a monotonically non-decreasing sequence of n such that $0 < a_n \leq n$ and the sequence $\{n/a_n, n \geq 1\}$ is monotonically non-decreasing, then,*

$$\limsup_{n \rightarrow \infty} \sup_{0 \leq t \leq n - a_n} \sup_{0 \leq s \leq a_n} \beta_n |B(t+s) - B(t)| = 1 \text{ a.s.}$$

where

$$\beta_n = \left(2a_n \left(\log \frac{n}{a_n} + \log^2 n \right) \right)^{-\frac{1}{2}}.$$

THEOREM A.5 (SEE MAPPING THEOREM 2.7, [6]). *Suppose h is an \mathcal{S}/\mathcal{S}' -measurable mapping from \mathcal{S} to \mathcal{S}' with discontinuity set $D_h \subset \mathcal{S}$, where $(\mathcal{S}, \mathcal{S})$ and $(\mathcal{S}', \mathcal{S}')$ are metric spaces. If $\{Q_n, n \geq 1\}$ is a sequence of probability measures on $(\mathcal{S}, \mathcal{S})$ with weak limit Q , that is, $Q_n \xrightarrow{d} Q$, and $QD_h = 0$, then $Q_n h^{-1} \xrightarrow{d} Qh^{-1}$.*

B PROOF OF THEOREM 6.1

PROOF. Proof. Similar to the proof of Theorem 5.1, observe that

$$\begin{aligned} \hat{\sigma}_{\text{OB-II}}^2(m_n, b_n) &= \frac{1}{\kappa_2(\beta, b)} \frac{m_n}{b_n} \sum_{j=1}^{b_n} \underbrace{\left[(\bar{\theta}_{j,m_n} - \bar{\theta}_{m_n})^2 - \sigma^2 \left(\tilde{B}_{j,m_n} - \frac{1}{b_n} \sum_{i=1}^{b_n} \tilde{B}_{i,m_n} \right)^2 \right]}_{\bar{E}_n(m_n, b_n)} \\ &\quad + \frac{1}{\kappa_2(\beta, b)} \frac{\sigma^2}{b_n} \sum_{j=1}^{b_n} \underbrace{\left(\sqrt{m_n} \tilde{B}_{j,m_n} - \frac{\sqrt{m_n}}{b_n} \sum_{i=1}^{b_n} \tilde{B}_{i,m_n} \right)^2}_{\bar{I}_n} \\ &= \frac{1}{\kappa_2(\beta, b)} (\bar{E}_n(m_n, b_n) + \bar{I}_n). \end{aligned} \tag{88}$$

We will now individually characterize the behavior of $\bar{E}_n(m_n, b_n)$ and \bar{I}_n above.

Noticing that

$$\bar{\theta}_{j,m_n} - \bar{\theta}_{m_n} = \underbrace{(\bar{\theta}_{j,m_n} - \sigma \tilde{B}_{j,m_n})}_{U_{j,m_n}} + \underbrace{\sigma \left(\tilde{B}_{j,m_n} - \frac{1}{b_n} \sum_{i=1}^{b_n} \tilde{B}_{i,m_n} \right)}_{\bar{H}_{j,m_n}} + \underbrace{\left(\frac{1}{b_n} \sum_{i=1}^{b_n} \tilde{B}_{i,m_n} - \bar{\theta}_{m_n, d_n} \right)}_{\bar{C}_{m_n}}, \tag{89}$$

we can write

$$\bar{E}_n(m_n, b_n) = m_n \left(\frac{1}{b_n} \sum_{j=1}^{b_n} U_{j,m_n}^2 + 2 \frac{\sigma}{b_n} \sum_{j=1}^{b_n} U_{j,m_n} \bar{H}_{j,m_n} + 2 \frac{\sigma \bar{C}_{m_n}}{b_n} \sum_{j=1}^{b_n} U_{j,m_n} + 2 \frac{\sigma \bar{C}_{m_n}}{b_n} \sum_{j=1}^{b_n} \bar{H}_{j,m_n} + \bar{C}_{m_n}^2 \right). \tag{90}$$

We already know that except for a set of measure zero in the probability space implied by Assumption 5, there exists $\Gamma(\omega)$ such that, uniformly in j ,

$$|U_{j,m_n}| \leq \Gamma(\omega) m_n^{-1/2-\delta} (\log^2 m_n)^{1/2}, \tag{91}$$

and similarly,

$$\begin{aligned} |\bar{C}_{m_n}| &= \frac{\sigma}{b_n} \left| \sum_{i=1}^{b_n} \sigma^{-1} \bar{\theta}_{i,m_n} - \tilde{B}_{i,m_n} \right| \\ &\leq \Gamma(\omega) m_n^{-1/2-\delta} (\log^2 m_n)^{1/2} \end{aligned} \tag{92}$$

Furthermore, similar to (47), we use Theorem A.4 carefully again to see that there exists $n_0(\omega, \epsilon)$ such that for all $n \geq n_0(\omega, \epsilon)$, and uniformly in j ,

$$|\bar{H}_{j,m_n}| \leq 2(1 + \epsilon) m_n^{-1/2} \sqrt{2 \left(\log^2 n - \log \frac{m_n}{n} \right)} \tag{93}$$

Plugging (91), (92), and (93) in (90), we get for all $n \geq n_0(\omega, \epsilon)$ that

$$\begin{aligned} \bar{E}_n(m_n, b_n) &\leq \Gamma^2(\omega) m_n^{-2\delta} \log^2 m_n \\ &\quad + 8\sigma(1 + \epsilon) \Gamma(\omega) m_n^{-\delta} (\log^2 m_n)^{1/2} \left(2 \left(\log^2 n - \log \frac{m_n}{n} \right) \right)^{1/2} \\ &\quad + 2\sigma \Gamma^2(\omega) m_n^{-2\delta} \log^2 m_n + \Gamma^2(\omega) m_n^{-2\delta} \log^2 m_n, \end{aligned} \tag{94}$$

implying that \bar{E}_n goes to zero almost surely. Now let's calculate the weak limit of

$$\bar{I}_n := \frac{\sigma^2}{b_n} \sum_{j=1}^{b_n} \left(\sqrt{m_n} \tilde{B}_{j,m_n} - \frac{1}{b_n} \sum_{i=1}^{b_n} \sqrt{m_n} \tilde{B}_{i,m_n} \right)^2$$

appearing in (88). Similar to the proof of Theorem 5.1, define a lattice $\{0, \delta_n, 2\delta_n, \dots, \lfloor \frac{1}{\delta_n} \rfloor \delta_n\}$ having resolution

$$\delta_n := \frac{1 - (m_n/n)}{b_n - 1}$$

and a corresponding projection operation on the lattice

$$\lfloor u \rfloor_{\delta_n} := \max\{k\delta_n : u \geq k\delta_n, k \in \mathbb{Z}\}, \quad u \in [0, 1 - \frac{m_n}{n} + \delta_n].$$

Recalling that $b_n = 1 + d_n^{-1}(n - m_n)$, we can write

$$\begin{aligned} \bar{I}_n = \sigma^2 \frac{1}{b_n} \frac{b_n - 1}{1 - (m_n/n)} \int_0^{1 - \frac{m_n}{n} + \delta_n} \left(\frac{1}{\sqrt{m_n}} (W(n\lfloor u \rfloor_{\delta_n} + m_n) - W(n\lfloor u \rfloor_{\delta_n} + 1)) \right. \\ \left. - \frac{1}{b_n} \frac{b_n - 1}{1 - (m_n/n)} \int_0^{1 - \frac{m_n}{n} + \delta_n} \frac{1}{\sqrt{m_n}} (W(n\lfloor s \rfloor_{\delta_n} + m_n) - W(n\lfloor s \rfloor_{\delta_n} + 1)) ds \right)^2 du \quad (95) \end{aligned}$$

$$\begin{aligned} \stackrel{d}{=} \sigma^2 \frac{b_n - 1}{b_n} \frac{n}{n - m_n} \frac{n}{m_n} \int_0^{1 - \frac{m_n}{n} + \delta_n} \left(W(\lfloor u \rfloor_{\delta_n} + \frac{m_n}{n}) - W(\lfloor u \rfloor_{\delta_n} + \frac{1}{n}) \right. \\ \left. - \frac{b_n - 1}{b_n} \frac{n}{n - m_n} \int_0^{1 - \frac{m_n}{n} + \delta_n} \frac{1}{\sqrt{m_n}} (W(n\lfloor s \rfloor_{\delta_n} + m_n) - W(n\lfloor s \rfloor_{\delta_n} + 1)) ds \right)^2 du \quad (96) \end{aligned}$$

$$\stackrel{d}{\rightarrow} \sigma^2 \frac{\beta^{-1}}{1 - \beta} \int_0^{1 - \beta} \left(W(u + \beta) - W(u) - \frac{1}{1 - \beta} \int_0^{1 - \beta} (W(s + \beta) - W(s)) ds \right)^2 du \quad (97)$$

if $\delta_n \rightarrow 0$ as $n \rightarrow \infty$ which happens when $b_n \rightarrow b = \infty$. This proves the assertion in (76) for $b = \infty$.

For the $b \in \{1, 2, \dots\}$ case, we observe that

$$\begin{aligned} \bar{I}_n := \frac{\sigma^2}{b_n} \sum_{j=1}^{b_n} \left(\sqrt{m_n} \tilde{B}_{j,m_n} - \frac{1}{b_n} \sum_{i=1}^{b_n} \sqrt{m_n} \tilde{B}_{i,m_n} \right)^2 \\ \stackrel{d}{=} \sigma^2 \frac{n}{m_n} \frac{1}{b_n} \sum_{j=1}^{b_n} \left(W((j-1) \frac{1 - (m_n/n)}{b_n - 1} + \frac{m_n}{n}) - W((j-1) \frac{1 - (m_n/n)}{b_n - 1} + \frac{1}{n}) - \right. \\ \left. \frac{1}{b_n} \sum_{i=1}^{b_n} \left(W((i-1) \frac{1 - (m_n/n)}{b_n - 1} + \frac{m_n}{n}) - W((i-1) \frac{1 - (m_n/n)}{b_n - 1} + \frac{1}{n}) \right) \right)^2 \quad (98) \end{aligned}$$

$$\begin{aligned} \rightarrow \sigma^2 \frac{1}{\beta} \frac{1}{b} \sum_{j=1}^b \left(W((j-1) \frac{1 - \beta}{b - 1} + \beta) - W((j-1) \frac{1 - \beta}{b - 1}) - \right. \\ \left. \frac{1}{b} \sum_{i=1}^b \left(W((i-1) \frac{1 - \beta}{b - 1} + \beta) - W((i-1) \frac{1 - \beta}{b - 1}) \right) \right)^2, \quad (99) \end{aligned}$$

thus proving the assertion for finite b .

Next, observe that

$$\begin{aligned} \mathbb{E}[\bar{I}_n] &= \frac{\sigma^2}{b_n} \sum_{j=1}^{b_n} \mathbb{E} \left[\sqrt{m_n} \tilde{B}_{j,m_n} - \frac{1}{b_n} \sum_{i=1}^{b_n} \sqrt{m_n} \tilde{B}_{i,m_n} \right]^2 \\ &= \sigma^2 \left(\frac{m_n}{b_n} \sum_{j=1}^{b_n} \mathbb{E}[\tilde{B}_{j,m_n}^2] - \frac{m_n}{b_n^2} \sum_{i=1}^{b_n} \sum_{j=1}^{b_n} \mathbb{E}[\tilde{B}_{i,m_n} \tilde{B}_{j,m_n}] \right). \end{aligned} \tag{100}$$

Some algebra yields

$$\frac{m_n}{b_n} \sum_{j=1}^{b_n} \mathbb{E}[\tilde{B}_{j,m_n}^2] = \frac{m_n}{b_n} \sum_{j=1}^{b_n} \frac{m_n - 1}{m_n^2} = 1, \quad j \in \{1, 2, \dots, b_n\} \tag{101}$$

and

$$m_n \mathbb{E}[\tilde{B}_{i,m_n} \tilde{B}_{j,m_n}] = \left(1 - \frac{|i-j|}{b_n-1} \frac{n-m_n}{m_n} \right)^+ \quad i, j \in \{1, 2, \dots, b_n\}. \tag{102}$$

Plug in (101) and (102) in (100), and we have

$$\begin{aligned} \mathbb{E}[\bar{I}_n] &= \sigma^2 \left(1 - \frac{1}{b_n^2} \sum_{i=1}^{b_n} \sum_{j=1}^{b_n} \left(1 - \frac{|i-j|}{b_n-1} \frac{n-m_n}{m_n} \right)^+ \right) \\ &= \sigma^2 \left(1 - \frac{1}{b_n} - \frac{2}{b_n} \sum_{h=1}^{b_n-1} \left(1 - \frac{h}{b_n-1} \frac{n-m_n}{m_n} \right)^+ (1 - h/b_n) \right) \end{aligned} \tag{103}$$

$$= \sigma^2 \kappa_2(\beta, b) + o(n^{-\delta}), \tag{104}$$

where the last equality holds by the definition of $\kappa_2(\beta, b)$ in (75) and since we have assumed that $|m_n/n - \beta| = o(n^{-\delta})$. Also, since Γ has been assumed to have finite second moment, (94) implies that

$$\mathbb{E}[\bar{E}_n] = O \left((\beta n)^{-\delta} \sqrt{2 \log^2 \beta n \log^2 n} \right). \tag{105}$$

Using (104) and (105) in (88), we conclude that for finite $b \in \{1, 2, \dots\}$,

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\sigma}_{\text{OB-II}}^2(m_n, b_n)] = \sigma^2 + O(\epsilon_{2,n}). \tag{106}$$

Let's next consider the $b = \infty$ case. We write the summation appearing in (103) as an integral on a lattice of size $\delta_n = 1/(b_n - 1)$ as follows:

$$\begin{aligned} \frac{2}{b_n} \sum_{h=1}^{b_n-1} \left(1 - \frac{h}{b_n-1} \left(\frac{n}{m_n} - 1 \right) \right)^+ (1 - \frac{h}{b_n}) \\ = 2 \int_0^1 \left(1 - \lfloor u \rfloor_{\delta_n} \left(\frac{n}{m_n} - 1 \right) \right)^+ (1 - \lfloor u \rfloor_{\delta_n} \frac{b_n-1}{b_n}) du. \end{aligned} \tag{107}$$

Plugging (107) in (103), and since $|m_n/n - \beta| = o(n^{-\delta})$ and $b_n^{-1} = o(n^{-\delta})$, we get

$$\begin{aligned} \mathbb{E}[\bar{I}_n] &= \sigma^2 \left(1 - 2 \int_0^1 \left(1 - u \left(\frac{1}{\beta} - 1 \right) \right)^+ (1 - u) du \right) + O(\epsilon_{2,n}) \\ &= \sigma^2 \left(1 - 2 \int_0^{\frac{\beta}{1-\beta} \wedge 1} \left(1 - u \frac{1-\beta}{\beta} \right) (1 - u) du \right) + O(\epsilon_{2,n}) \\ &= \sigma^2 \left(1 - 2 \left(\frac{\beta}{1-\beta} \wedge 1 \right) + \frac{1}{\beta} \left(\frac{\beta}{1-\beta} \wedge 1 \right)^2 - \frac{2}{3} \frac{1-\beta}{\beta} \left(\frac{\beta}{1-\beta} \wedge 1 \right)^3 \right) + O(\epsilon_{2,n}) \\ &=: \sigma^2 \kappa_2(\beta, \infty) + O(\epsilon_{2,n}). \end{aligned} \tag{108}$$

Now use (108) and (105) in (88) to see that the assertion corresponding to $b = \infty$ also holds.

Let's now prove that the statement in (77) holds. Using Assumption 5 and after some algebra, we have almost surely,

$$\left| \sqrt{n} \frac{(\bar{\theta}_n - \theta(P))}{\sigma} - \frac{1}{b_n} \sum_{j=1}^{b_n} \frac{\sqrt{n}}{m_n} (W(nc_j + m_n) - W(nc_j)) \right| \leq \Gamma m_n^{-\delta-1/2} \sqrt{\log^2 m_n}, \tag{109}$$

where Γ is a well-defined random variable with finite mean, $\delta > 0$, and

$$c_j = (j - 1) \frac{n - m_n}{b_n - 1}.$$

Also, since $m_n \rightarrow \infty$ and $m_n/n \rightarrow \beta > 0$, we see that

$$\frac{1}{b_n} \sum_{j=1}^{b_n} \frac{\sqrt{n}}{m_n} (W(nc_j + m_n) - W(nc_j)) \xrightarrow{d} \begin{cases} \frac{1}{1-\beta} \int_0^{1-\beta} W(s + \beta) - W(s) ds & b = \infty; \\ \frac{1}{b\beta} \sum_{j=1}^b W(nc_j + m_n) - W(nc_j) & b \in \mathbb{N} \setminus \{1\}, \end{cases}$$

implying along with (109) and the functional central limit theorem (Theorem A.5) that the assertion in (77) holds. □