Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

APPLICATIONS OF STOCHASTIC SIMULATION IN TWO-STAGE MULTIPLE COMPARISONS WITH THE BEST PROBLEM AND TIME AVERAGE VARIANCE CONSTANT ESTIMATION

by
Dibyendu Chakrabarti

In this dissertation, we study two problems. In the first part, we consider the two-stage methods for comparing alternatives using simulation. Suppose there are a finite number of alternatives to compare, with each alternative having an unknown parameter that is the basis for comparison. The parameters are to be estimated using simulation, where the alternatives are simulated independently. We develop two-stage selection and multiple-comparison procedures for simulations under a general framework. The assumptions are that each alternative has a parameter estimation process that satisfies a random- time-change central limit theorem (CLT), and there is a weakly consistent variance estimator (WCVE) for the variance constant appearing in the CLT. The framework encompasses comparing means of independent populations, functions of means, and steady-state means. One problem we consider of considerable practical interest and not handled in previous work on two-stage multiple-comparison procedures is comparing quantiles of alternative populations. We establish the asymptotic validity of our procedures as the prescribed width of the confidence intervals or indifference-zone parameter shrinks to zero. Also, for the steady-state simulation context, we compare our procedures based on WCVEs with techniques that instead use standardized time series methods. In the second part, we propose a new technique of estimating the variance parameter of a wide variety of stochastic processes. This new technique is better than the existing techniques for some standard stochastic processes in terms of bias and variance properties, since it reduces bias at the cost of no significant increase in variance.

APPLICATIONS OF STOCHASTIC SIMULATION IN TWO-STAGE MULTIPLE COMPARISONS WITH THE BEST PROBLEM AND TIME AVERAGE VARIANCE CONSTANT ESTIMATION

by
Dibyendu Chakrabarti

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Computer Science

Department of Computer Science, NJIT

April 2011

APPROVAL PAGE

APPLICATIONS OF STOCHASTIC SIMULATION IN TWO-STAGE MULTIPLE COMPARISONS WITH THE BEST PROBLEM AND TIME AVERAGE VARIANCE CONSTANT ESTIMATION

Dibyendu Chakrabarti

Marvin K. Nakayama, Dissertation Co-Advisor — Date
Professor, Department of Computer Science, NJIT

James M. Calvin, Dissertation Co-Advisor — Date
Professor, Department of Computer Science, NJIT

Manish Bhattacharjee, Committee Member — Date
Professor, Department of Mathematical Sciences, NJIT

Zhi Wei, Committee Member — Date
Assistant Professor, Department of Computer Science, NJIT

Narain Gehani, Committee Member — Date
Professor, Department of Computer Science, NJIT

BIOGRAPHICAL SKETCH

Author: Dibyendu Chakrabarti

Degree: Doctor of Philosophy

Date: April 2011

- Doctor of Philosophy in Computer Science,
  New Jersey Institute of Technology, Newark, NJ, 2011
- Master of Science in Computer Science,
  Indian Statistical Institute (ISI), Calcutta, India, 1998

Major: Computer Science

D. Chakrabarti, "Application of Simulation in Designing a Secure Grid of Distributed Wireless Sensor Network," *Proceedings of IEEE International Conference on Industrial Engineering and Engineering Management (IEEM 2009),* Hong Kong.

ACKNOWLEDGMENT

To begin with, I want to express my deep gratitude to my advisors, Prof Marvin K. Nakayama and Prof James M. Calvin for their guidance throughout this research. Their profound knowledge and experience have strengthened this dissertation in many ways. I am extremely grateful for their patience and encouragement. I wish to thank my Dissertation Committee members, Prof Manish Bhattacharjee, Prof Zhi Wei and Prof Narain Gehani for their valuable suggestions and strong support. I would also like to extend my appreciation to the faculty, staff and fellow graduate students of the Department of Computer Science, NJIT, whom I have known for years. I am grateful to the department for providing the financial and technical support for this research. Finally, I want to express my love and gratitude to my family. I owe them everything.

TABLE OF CONTENTS

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Mathematical Modeling

Mathematical modeling is a valuable tool in order to assess the future evolution of a complex real world system. All possible non-physical models are mathematical in nature. These models help in understanding the laws governing the system and also in the prediction/control of the modeled system.

A mathematical model may be thought of as "a representation of the essential aspects of an existing system (or a system to be constructed) which presents knowledge of that system in usable form" (see [23, p. 240]). The model may be useful in natural sciences, engineering and social sciences. Mathematical modeling is the process of developing a mathematical model that describes a system using mathematical language.

There may be four stages of any mathematical modeling process:

1. Stating the laws relating the basic elements of the system: At this stage, the mathematical relationship between the objects of the system are explored, selecting the most important aspects of the system and omitting the irrelevant details. Thus we sometimes end up with a crude model at the expense of avoiding complexity and intractability. For example, consider the differential equation governing the motion of a clock's pendulum: $x''(t) = -kx(t)$, where $x(t)$ denotes the displacement of the pendulum at time t, $x''(t)$ denotes the second derivative of x with respect to t and k is a constant independent of time t. In this case, we often ignore the physical real conditions to make the model simple (e.g., we ignore the facts that usually the rod/thread of the pendulum is not massless, the bob is not point-size and the amplitude of the oscillations is not small, i.e., less than 4°). Here, the resulting model is a simple differential equation. However, it may be a random process having a probability distribution.

2. Investigating the underlying mathematical problem: At this stage, the main question is to solve the given direct problem; given the "input data", we need to obtain the "output data" by analyzing the model. For example, we may try to find the position of the pendulum at time $t = 100$ given $x(0) = 0$ and $x'(0) = 1$.

3. Verifying the model: We need to compare the theoretical results derived for the model with the observed results. The observed results may not be available if the system being studied does not exist. If the results are not in close agreement with each other, the model is rejected and we go back to stage 1 to rethink and formulate another model, possibly by embellishing the original model.

4. Subsequently analyzing the model: If a better description of the system is desired, the existing model may be insufficient for that purpose. In this case also, we return to stage 1, but this time, we try to construct a more precise model at a higher level of complexity.

1.2 Analytical and Simulation Methods

A set of equations or relations among mathematical variables is described by a mathematical model, whereas a computer program implements a simulation model. Simulation modeling is usually a comparatively inexpensive modeling technique and is often used as a substitute of mathematical modeling.

We shall consider an example in order to compare and contrast these two approaches. Let there be a production line and we model it as a queuing system. In the analytical approach, the system is represented by a set of equations and the solution to it. In the simulation approach, a computer program representation of the queuing system is created. The program is run to produce several sample paths. If we need to calculate the steady state performance measures (e.g., average work in the system, distribution of waiting times etc.), these may be obtained respectively from the mathematical expressions or the simulation statistics.

Suppose a given mathematical model may be solved using analytic methods or may be approximated using analytical approximations or simulation. If the analytical solution to the model is known, it is usually easier and faster to compute in comparison with the corresponding simulation model. Usually it is very difficult to model complex systems by sufficiently detailed mathematical models and the unavailability of the analytical solution to the model compells the modeler to resort to simulation.

There are several reasons behind choosing a simulation method to model a complex system. First, even though it might be possible to find an analytical solution, the derivation of the solution might be prohibitively time-consuming. Second, due to mathematical complexity, the analytical solution may not be attempted at all. Finally, even if the analytical solution is attempted, it might not be possible to come up with an appropriate model capable enough to represent the system's specific aspects of interest.

On the other hand, simulation modeling is very versatile in representing any system constrained by any kind of assumptions. It spares the modeler the pain of finding the analytical solutions. The entire task boils down to write and run a simulation program. An ideal simulation user applies analytical methods whenever possible, e.g., via variance reduction techniques (VRTs).

In practice, there is no denial of the fact that the simulation methods are more versatile than the analytical methods. These methods may be used to evaluate two or more competing system designs or evaluate the probability of extremely rare events. The application areas of simulation are numerous and the complex business and industry applications demand simulation as the only option.

For a more detailed account on this, please refer to [3].

1.3 Motivation

In this dissertation, we study two problems. In the first part, we consider the two-stage methods for comparing alternatives using simulation. Suppose there are a finite number of alternatives to compare, with each alternative having an unknown parameter

that is the basis for comparison. The parameters are to be estimated using simulation, where the alternatives are simulated independently. We develop two-stage selection and multiple-comparison procedures for simulations under a general framework. The assumptions are that each alternative has a parameter estimation process that satisfies a random- time-change central limit theorem (CLT), and there is a weakly consistent variance estimator (WCVE) for the variance constant appearing in the CLT. The framework encompasses comparing means of independent populations, functions of means, and steady-state means. One problem we consider of considerable practical interest and not handled in previous work on two-stage multiple-comparison procedures is comparing quantiles of alternative populations. We establish the asymptotic validity of our procedures as the prescribed width of the confidence intervals or indifference-zone parameter shrinks to zero. Also, for the steady-state simulation context, we compare our procedures based on WCVEs with techniques that instead use standardized time series methods.

In the second part, we propose a new technique of estimating the variance parameter of a wide variety of stochastic processes. This new technique is better than the existing techniques for some standard stochastic processes in terms of bias and variance properties, since it reduces bias at the cost of no significant increase in variance.

1.4 Organization

This dissertation has two major chapters. Chapter 2 develops two-stage methods for comparing alternatives using simulation. Chapter 3 describes a special technique for the estimation of the variance parameter of a broad family of stochastic processes.

CHAPTER 2

ASYMPTOTIC ANALYSIS OF TWO-STAGE SELECTION AND MULTIPLE-COMPARISON PROCEDURES FOR SIMULATIONS

2.1 Introduction

The goal of many simulation studies is to compare alternative system designs and identify the best one relative to some parameter. For example, consider comparing 10 designs for a computer system in terms of the 0.95 quantiles of their distributions of the times to failure.

This chapter develops two-stage methods for comparing alternatives using simulation. The methods we focus on are multiple comparisons with the best (MCB; [39]) and a selection procedure. In two-stage MCB the goal is to construct simultaneously confidence intervals with confidence level $1-\alpha$ for the difference in performance of each alternative to that of the best, where the intervals have a pre-specified width parameter δ. The selection procedure we develop uses the *indifference-zone* formulation of [5], where the user is indifferent to alternatives whose parameter values are within δ of each other. Here, the goal is to correctly choose the best alternative with a pre-specified probability of correct selection (PCS) when the difference in the performance of the best and second-best alternatives is at least δ. See [6], [72], and [46] for overviews of multiple-comparison and selection procedures.

Previous work on multiple-comparison and selection procedures mostly focuses on comparing means or steady-state means of alternative systems. In this chapter we consider a general framework that encompasses these and many other parameters of interest. Our assumptions are that the parameter estimators satisfy a random-time-change central limit theorem (CLT) and that there is a weakly consistent variance estimator (WCVE) for the variance parameter appearing in the CLT. For example, one setting of considerable interest not covered by previous two-stage multiple-comparison

procedures is comparing quantiles, which cannot be expressed as means of random variables. Other problems covered by our work include comparing means of independent populations, functions of means, steady-state means, and Kiefer-Wolfowitz stochastic approximation ([44]).

Once outside the realm of independent and identically distributed (i.i.d.) sampling of independent normal populations, it is often not possible to develop MCB and selection procedures that guarantee proper coverage and PCS for finite sample sizes. Thus, we instead establish the asymptotic validity of our MCB and selection procedures as $\delta \to 0$. In practice, however, one cannot let $\delta \to 0$ but rather must choose a specific fixed δ, and our asymptotic results gives the user confidence that if δ is chosen "small enough," then the PCS and MCB coverage will be "close" to the nominal level. Section 3.3 of [51] provides further justification of the study of asymptotic procedures.

The references [21] and [46] have shown the asymptotic validity of selection and MCB procedures for steady-state simulations when each system satisfies a functional central limit theorem (FCLT). The reference [21] analyze two-stage MCB procedures using standardized time series (STS) methods ([64], [29]). STS methods lead to estimators of the process's variance parameter that are not consistent ([29]), and many (but not all) STS variance estimators asymptotically have a chi-squared distribution. The reference [45] develop two fully sequential screening and selection procedures under different assumptions on their batched variance estimators. (Screening methods allow for early elimination of alternatives that are clearly inferior.) First they consider a procedure in which variances are estimated only once in an initial sample, and they assume the variance estimators are exactly chi-squared (not just asymptotically). Their other approach allows for updating the variance estimators, but now requires them to be strongly consistent.

Two-stage procedures enjoy some advantages over fully sequential methods. First, fully sequential methods are more complicated to implement since they require frequently stopping and restarting the simulations of each alternative and switching among

the alternatives. Moreover, fully sequential procedures typically require strongly consistent variance estimators, and [32] provide an example showing that if a weakly consistent variance estimator is used in a fully sequential stopping procedure, it may no longer be asymptotically valid. On the other hand, the two-stage procedures we consider here require only weakly consistent variance estimators. Indeed, the two-stage procedures of [21] (also see [55]) use STS variance estimators, which are not even consistent. Of course, fully sequential procedures also have desirable properties compared to two-stage methods; e.g., the former can result in smaller overall run lengths, especially when combined with screening methods.

Some heuristic selection procedures for steady-state simulations have also been studied under various assumptions on the simulated processes. The reference [41] proposes heuristic fully sequential and two-stage selection procedures for regenerative processes, and [71] develop a heuristic subset selection method for stationary normal processes by using spectral estimators of the variance parameter.

We also provide an asymptotic comparison in the context of steady-state simulations of our WCVE MCB methods and the STS MCB procedures of [21]. We show that WCVE MCB methods asymptotically have strictly smaller average and less variable total run lengths than STS MCB algorithms. However, if we use batched STS methods in which the number of batches grows to infinity, then WCVE and STS MCB methods become equivalent.

The rest of the chapter has the following organization. Section 2.2 describes our notation and assumptions. Section 2.3 details our two-stage MCB procedure with an absolute-width parameter, and we modify it to be a selection procedure in Section 2.4. We consider a relative-width MCB procedure in Section 2.5. Section 2.6 provides several examples of applications satisfying the framework of Section 2.2. Section 2.7 contains an asymptotic comparison of WCVE and STS MCB methods. Section 2.8 presents some experimental results. The reference [56] presents some of the results from this

chapter without proofs for the special case of steady-state simulations of regenerative processes.

2.2 Notation and Assumptions

Suppose there are $k < \infty$ alternatives, labeled $1, 2, \ldots, k$, where each alternative i has an associated parameter $\theta_i \in \Re$, which is unknown and needs to be estimated using simulation. We want to compare the k alternatives in terms of the parameters $\theta_1, \ldots, \theta_k$. For $i = 1, \ldots, k$, let (i) denote the index of the alternative having the ith smallest parameter, so $\theta_{(1)} \leq \theta_{(2)} \leq \cdots \leq \theta_{(k)}$. We assume that larger parameters are better, so our goal is to identify alternative (k).

For each alternative i, we assume there is a continuous-time *estimation process* $\widehat{\theta}_i = [\widehat{\theta}_i(t) : t \geq 0]$, where $\widehat{\theta}_i(t)$ denotes an estimator of θ_i based on a simulation of alternative i up to time t. (Discrete-time estimators $\widehat{\theta}_{i,n}$ fit into our framework by taking $\widehat{\theta}_i(t) = \widehat{\theta}_{i,\lfloor t \rfloor}$, where $\lfloor a \rfloor$ denotes the largest integer less than or equal to $a \in \Re$.) For example, if $X_i = [X_i(t) : t \geq 0]$ is a real-valued stochastic process having steady-state mean θ_i, a natural choice for the estimation process is $\widehat{\theta}_i(t) = (1/t) \int_0^t X_i(s)\, ds$ for each $t > 0$. We will require that the estimation processes $\widehat{\theta}_1, \ldots, \widehat{\theta}_k$ are independent, and that each satisfies a central limit theorem (CLT) with rate parameter $\eta > 0$:

$$t^\eta \left[\widehat{\theta}_i(t) - \theta_i\right] \Rightarrow N(0, \sigma_i^2) \text{ as } t \to \infty, \text{ for each } i = 1, \ldots, k, \tag{2.1}$$

where "$\Rightarrow$"means weak convergence as defined in [9], $0 < \sigma_i < \infty$ is a constant. In many applications, η assumes the canonical value of $1/2$ (although we do not require this). For instance, this is typically the case in our previous example of estimating the steady-state mean of a process X_i.

The total run lengths of our two-stage procedures are random, and we need to strengthen the CLT in (2.1) so that it still holds when applied with a random-time change. For each alternative i, let $[\widetilde{T}_i(\delta) : \delta > 0]$ be a family of positive random

variables such that

$$\delta^{1/\eta}\widetilde{T}_i(\delta) \Rightarrow \widetilde{\tau}_i \text{ as } \delta \to 0, \text{ where } 0 < \widetilde{\tau}_i < \infty \text{ is a constant.} \tag{2.2}$$

We allow $\widetilde{T}_i(\delta)$, $i = 1, \ldots, k$, to be dependent. Let $\widetilde{\tau}_{i,\delta} = \widetilde{\tau}_i \delta^{-1/\eta}$, and we assume the following:

Assumption 1 *There exists a constant* $0 < \eta < \infty$ *such that for* $\widetilde{T}_i(\delta)$, $i = 1, \ldots, k$, *satisfying (2.2),*

$$\left(\widetilde{\tau}_{i,\delta}^{\eta}\left[\widehat{\theta}_i(\widetilde{T}_i(\delta)) - \theta_i\right], i = 1, \ldots, k\right) \Rightarrow \left(N_i(0, \sigma_i^2), i = 1, \ldots, k\right) \text{ as } \delta \to 0, \tag{2.3}$$

where $N_i(0, \sigma_i^2)$, $i = 1, \ldots, k$, *are independent normals.*

The CLT in (2.1) is a special case of Assumption 1 since we can take $\widetilde{T}_i(\delta) = \delta^{-1/\eta}$. Section 2.6 provides various sufficient conditions for Assumption 1 to hold.

We call σ_i^2 in (2.1) and (2.3) the *variance parameter* associated with the estimation process $\widehat{\theta}_i$. For each alternative i, we define a *variance estimation process* $V_i = [V_i(t) : t \geq 0]$, and we assume $V_i(t)$ is a weakly consistent estimator for σ_i^2:

Assumption 2 *For each alternative* i, $V_i(t) \Rightarrow \sigma_i^2$ *as* $t \to \infty$.

For example,in the previous example where θ_i represents the steady-state mean of the process X_i, if we assume that the process X_i is regenerative, we can use the regenerative method ([42]) to construct a variance estimation process V_i satisfying Assumption 2. Section 2.6 provides more details on this and other simulation settings satisfying our two assumptions.

2.3 Absolute-Width MCB

The following two-stage procedure, which is related to a method of [59], constructs MCB intervals with absolute-width parameter $\delta > 0$.

Procedure A

1. Specify the confidence level $1-\alpha$, the desired absolute-width parameter δ of the MCB confidence intervals, and the first-stage run length $T_{0,i}$ for each alternative i.

2. Independently simulate each alternative i for a run length of $T_{0,i}$.

3. For each alternative i, compute the total run length required as

$$T_i(\delta) = \max\left(T_{0,i}, \left(\frac{\gamma\sqrt{V_i(T_{0,i})}}{\delta}\right)^{1/\eta}\right), \tag{2.4}$$

   where the constant $\gamma \equiv \gamma(k, 1-\alpha) = \sqrt{2}\, z_{(1-\alpha)^{1/(k-1)}}$, with z_β satisfying $\Phi(z_\beta) = \beta$ for $0 < \beta < 1$, Φ is the distribution function of a standard (mean 0 and variance 1) normal distribution, η is as defined in Assumption 1, and $V_i(t)$ is any estimator satisfying Assumption 2.

4. For each alternative i, continue to simulate from time $T_{0,i}$ to $T_i(\delta)$, where the k alternatives are simulated independently, and form the point estimator $\widetilde{\theta}_i(\delta) = \widehat{\theta}_i(T_i(\delta))$ of θ_i.

5. Use the absolute-width parameter δ to construct simultaneous MCB confidence intervals

$$I_i(\delta) = \left[-\left(\widetilde{\theta}_i(\delta) - \max_{\ell\neq i}\widetilde{\theta}_\ell(\delta) - \delta\right)^-,\ \left(\widetilde{\theta}_i(\delta) - \max_{\ell\neq i}\widetilde{\theta}_\ell(\delta) + \delta\right)^+\right], \quad i = 1, \ldots, k,$$

   for $\theta_i - \max_{\ell\neq i}\theta_\ell$, $i = 1, \ldots, k$, respectively, where $-(\beta)^- = \min(\beta, 0)$ and $(\beta)^+ = \max(\beta, 0)$.

Let $\bar{\gamma} \equiv \bar{\gamma}(k, 1-\alpha) = \sqrt{2}\bar{z}_{k-1,1-\alpha}$, where $\bar{z}_{p,\beta}$ is the upper-β equicoordinate point of a p-variate standard normal distribution with unit variances and common correlation coefficient 1/2. Table B.1 of [6] provides values for $\bar{z}_{p,\beta}$ for various p and β. When σ_i, $i = 1, \ldots, k$, are known, one can instead use a single-stage procedure with the total run length for alternative i being $(\bar{\gamma}\sigma_i/\delta)^{1/\eta}$ (Section 2.6 of [6]).

Theorem 1 *If the CLT in (2.1) and Assumption 2 hold and Procedure A is used with first-stage run length* $T_{0,i} = \zeta_i \delta^{-\lambda}$ *for each* i*, where* $\zeta_i > 0$ *and* $0 < \lambda \leq 1/\eta$ *are any constants, then*

(i) $\delta^{1/\eta}\, T_i(\delta) \Rightarrow \tau_i$ *as* $\delta \to 0$*, where*

$$\tau_i = \begin{cases} (\gamma\sigma_i)^{1/\eta} & \text{if } \lambda < 1/\eta \\ \max(\zeta_i, (\gamma\sigma_i)^{1/\eta}) & \text{if } \lambda = 1/\eta \end{cases} . \tag{2.5}$$

Moreover, if $\{V_i(t) : t \geq 0\}$ *is uniformly integrable, then* $E[\delta^{1/\eta}\, T_i(\delta)] \to \tau_i$ *as* $\delta \to 0$*.*

In addition, if the CLT in (2.1) is strengthened to Assumption 1, then the following also hold:

(ii) $\lim_{\delta \to 0} P\{\theta_i - \max_{\ell \neq i} \theta_\ell \in I_i(\delta),\ i = 1, \ldots, k\} > 1 - \alpha$*.*

(iii) If $\eta = 1/2$ *and* $0 < \lambda < 2$*, then (i)–(ii) still hold when* γ *in (2.4) is replaced with* $\bar{\gamma}$*, and* $\bar{\gamma} < \gamma$*. Moreover,* $T_i(\delta)/(\bar{\gamma}\sigma_i/\delta)^2 \Rightarrow 1$ *as* $\delta \to 0$*, and if, in addition,* $\{V_i(t) : t \geq 0\}$ *is uniformly integrable, then* $E[T_i(\delta)/(\bar{\gamma}\sigma_i/\delta)^2] \to 1$ *as* $\delta \to 0$*.*

Part (i) shows that $T_i(\delta)$ is asymptotically equal to $\tau_i \delta^{-1/\eta}$ to first order, where τ_i is deterministic. Part (ii) establishes the asymptotic validity of the MCB intervals. Part (iii) shows that when η in Assumption 1 has the canonical value 1/2 and the first-stage run length is asymptotically negligible compared to $1/\delta^2$ (i.e., when $\lambda < 2$), Procedure A using $\bar{\gamma}$ instead of γ in (2.4) is asymptotically efficient in the sense that the total run length is asymptotically the same (to first order) as when the σ_i are known. The last result generalizes work of [54], which shows the asymptotic efficiency of Rinott's procedure ([59]) for i.i.d. normal populations when there are only $k = 2$ populations.

2.4 Selection Procedure

As shown in [52], MCB procedures can often be modified to also be selection procedures under the "indifference zone" formulation of [5], and we now modify Procedure A to

do this. Specifically, we use δ as an indifference parameter, and assume that we are indifferent to selecting any alternative i as long as $\theta_i > \theta_{(k)} - \delta$, i.e., the parameter θ_i of the selected alternative i is within δ of the largest parameter. Define $[1], [2], \ldots, [k]$ such that $\widetilde{\theta}_{[1]}(\delta) \leq \widetilde{\theta}_{[2]}(\delta) \leq \cdots \leq \widetilde{\theta}_{[k]}(\delta)$; i.e., $[i]$ is the alternative with the ith largest estimated parameter in step 4 of Procedure A. Then we make the following modification to Procedure A:

Procedure A.2

Use steps 1–5 of Procedure A, and include the additional step:

6. Select alternative $[k]$ as the best alternative.

We define a *correct selection* to be choosing an alternative i whose parameter θ_i is within δ of the largest parameter $\theta_{(k)}$. Thus, we define $\mathrm{CS}_\theta(\delta) = \{\theta_{[k]} > \theta_{(k)} - \delta\}$, which is the event of a correct selection under parameters $\theta = (\theta_1, \ldots, \theta_k)$ when the indifference parameter is δ. Define the *preference zone* to be

$$\Omega(\delta) = \{\theta = (\theta_1, \ldots, \theta_k) : \theta_{(k)} \geq \theta_{(k-1)} + \delta\},$$

which is the space of parameter configurations such that only the best alternative is desirable. Hence,

$$\mathrm{CS}_\theta(\delta) = \{\widetilde{\theta}_{(k)}(\delta) > \widetilde{\theta}_i(\delta), \forall i \neq (k)\} \quad \text{for } \theta \in \Omega(\delta).$$

To study the asymptotic validity of Procedure A.2, we need to develop a framework such that the result proven is theoretically nontrivial. Specifically, in the setup in Theorem 1, the k alternatives do not change as we take the limit $\delta \to 0$, so the parameters θ_i are fixed as we take the limit. It is typically the case that $\widetilde{\theta}_i(\delta) = \widehat{\theta}_i(T_i(\delta)) \to \theta_i$ almost surely (a.s.) as $\delta \to 0$ by the strong law of large numbers since $\lambda > 0$ and $T_i(\delta) \geq T_{0,i} = \zeta_i \delta^{-\lambda} \to \infty$ as $\delta \to 0$. Hence, if we add step 6 to Procedure A and

take $\delta \to 0$, we will be able to select the best alternative with probability approaching 1. To avoid this theoretical triviality, we allow the vector θ of parameters to vary as $\delta \to 0$. However, this significantly complicates the analysis required ([20]), so we make the following simplifying assumption:

Assumption 3 *For each alternative i, there exists a process $Y_i = [Y_i(t) : t \geq 0]$ such that $\widehat{\theta}_i(t) = \theta_i + Y_i(t)$ for all $t > 0$, where the distribution of Y_i does not depend on θ_i, and $Y_1, \ldots, Y_k$ are independent.*

This means that θ_i is a "location parameter," and we can think of Y_i as being a "noise" process with distribution independent of θ_i. In this case, the CLT (2.1) becomes $t^\eta Y_i(t) \Rightarrow N(0, \sigma_i^2)$ as $t \to \infty$. The reference [46] adopt a similar framework for their selection procedures for steady-state simulations and provide additional justification.

Theorem 2 *If Assumptions 1–3 hold and Procedure A.2 is used with first-stage run length $T_{0,i} = \zeta_i \delta^{-\lambda}$ for each alternative i, where $\zeta_i > 0$ and $0 < \lambda \leq 1/\eta$ are any constants, then*

$$\lim_{\delta \to 0} \inf_{\theta \in \Omega(\delta)} P\left\{ CS_\theta(\delta),\ \theta_i - \max_{\ell \neq i} \theta_\ell \in I_i(\delta),\ i = 1, \ldots, k \right\} > 1 - \alpha.$$

2.5 Relative-Width MCB

Procedure A produces MCB intervals with absolute-width parameter δ. However, in many situations, one desires confidence intervals having a pre-specified relative precision, e.g., $\pm 10\%$ of the point estimate. Below we present a procedure to do this.

For $i = 1, \ldots, k$, define $\langle i \rangle$ to be the alternative with the ith smallest parameter estimate at the end of the first stage. Thus, $\widehat{\theta}_{\langle 1 \rangle}(T_{0,\langle 1 \rangle}) \leq \widehat{\theta}_{\langle 2 \rangle}(T_{0,\langle 2 \rangle}) \leq \cdots \leq \widehat{\theta}_{\langle k \rangle}(T_{0,\langle k \rangle})$.

Procedure R

1. Specify the confidence level $1 - \alpha$, the desired relative-width parameter δ of the MCB confidence intervals, and the first-stage run length $T_{0,i}$ for each alternative i.

2. Independently simulate each alternative i for a run length of $T_{0,i}$.

3. For each alternative i, compute the total run length required as

$$T_{i,r}(\delta) = \max\left(T_{0,i}, \left(\frac{\gamma\sqrt{V_i(T_{0,i})}}{\delta\,\epsilon_i(T_0)}\right)^{1/\eta}\right), \tag{2.6}$$

where $T_0 = (T_{0,1}, T_{0,2}, \ldots, T_{0,k})$,

$$\epsilon_i(T_0) = \begin{cases} \widehat{\theta}_{\langle k\rangle}(T_{0,\langle k\rangle}) - \widehat{\theta}_i(T_{0,i}) & \text{if } \widehat{\theta}_i(T_{0,i}) < \widehat{\theta}_{\langle k\rangle}(T_{0,\langle k\rangle}), \\ \widehat{\theta}_{\langle k\rangle}(T_{0,\langle k\rangle}) - \widehat{\theta}_{\langle k-1\rangle}(T_{0,\langle k-1\rangle}) & \text{if } \widehat{\theta}_i(T_{0,i}) = \widehat{\theta}_{\langle k\rangle}(T_{0,\langle k\rangle}), \end{cases} \tag{2.7}$$

the constant γ is the same as in (2.4), η is as defined in Assumption 1, and $V_i(t)$ is any estimator satisfying Assumption 2.

4. For each alternative i, continue to simulate from time $T_{0,i}$ to $T_{i,r}(\delta)$, where the k alternatives are simulated independently, and form the point estimator $\widetilde{\theta}_{i,r}(\delta) = \widehat{\theta}_i(T_{i,r}(\delta))$ of θ_i.

5. Use the relative-width parameter δ to construct the simultaneous MCB confidence intervals

$$\begin{aligned} I_{i,r}(\delta) \;=\; & \left[-\left(\widetilde{\theta}_{i,r}(\delta) - \max_{\ell\neq i}\widetilde{\theta}_{\ell,r}(\delta) - \delta\left|\widetilde{\theta}_{i,r}(\delta) - \max_{\ell\neq i}\widetilde{\theta}_{\ell,r}(\delta)\right|\right)^{-},\right. \\ & \left.\left(\widetilde{\theta}_{i,r}(\delta) - \max_{\ell\neq i}\widetilde{\theta}_{\ell,r}(\delta) + \delta\left|\widetilde{\theta}_{i,r}(\delta) - \max_{\ell\neq i}\widetilde{\theta}_{\ell,r}(\delta)\right|\right)^{+}\right], \quad i = 1, \ldots, k, \end{aligned}$$

for $\theta_i - \max_{\ell\neq i}\theta_\ell$, $i = 1, \ldots, k$, respectively.

Theorem 3 *Suppose the CLT in (2.1) and Assumption 2 hold and $\theta_{(k-1)} < \theta_{(k)}$. If Procedure R is used with first-stage run length $T_{0,i} = \zeta_i\delta^{-\lambda}$ for each alternative i, where $\zeta_i > 0$ and $0 < \lambda \leq 1/\eta$ are any constants, then*

(i) $\delta^{1/\eta}\,T_{i,r}(\delta) \Rightarrow \tau_{i,r}$ as $\delta \to 0$, where

$$\tau_{i,r} = \begin{cases} (\gamma\sigma_i/\epsilon_i)^{1/\eta} & \text{if } \lambda < 1/\eta \\ \max(\zeta_i, (\gamma\sigma_i/\epsilon_i)^{1/\eta}) & \text{if } \lambda = 1/\eta \end{cases}, \tag{2.8}$$

$\epsilon_i = \theta_{(k)} - \theta_i$ *if* $i \neq (k)$, *and* $\epsilon_i = \theta_{(k)} - \theta_{(k-1)}$ *if* $i = (k)$. *Moreover, if* $\{V_i(t) : t \geq 0\}$ *is uniformly integrable, then* $E[\delta^{1/\eta}\, T_{i,r}(\delta)] \to \tau_{i,r}$ *as* $\delta \to 0$.

In addition, if the CLT in (2.1) is strengthened to Assumption 1, then the following also hold:

(ii) $\lim_{\delta \to 0} P\left\{\theta_i - \max_{\ell \neq i} \theta_\ell \in I_{i,r}(\delta),\, i = 1, \ldots, k\right\} > 1 - \alpha$.

(iii) If $\eta = 1/2$ *and* $0 < \lambda < 2$, *then (i)–(ii) still hold when* γ *in (2.6) is replaced with* $\bar{\gamma}$ *defined just before Theorem 1. Moreover,* $T_{i,r}(\delta)/(\bar{\gamma}\sigma_i/\delta)^2 \Rightarrow 1$ *as* $\delta \to 0$, *and if, in addition,* $\{V_i(t) : t \geq 0\}$ *is uniformly integrable, then* $E[T_{i,r}(\delta)/(\bar{\gamma}\sigma_i/\delta)^2] \to 1$ *as* $\delta \to 0$.

2.6 Examples

Before presenting simulation examples that satisfy Assumptions 1 and 2, we first develop general conditions that guarantee Assumption 1 holds. First, consider the following, which is known as *Anscombe's condition* ([4])

Condition 1 *For each alternative* i *and any positive* ε *and* ψ, *there exist positive* c_i *and* t_i *such that*

$$P\left\{\sup_{s\,:\,|s-t| \leq c_i t} t^\eta \left|\widehat{\theta}_i(s) - \widehat{\theta}_i(t)\right| > \varepsilon\right\} < \psi, \qquad \textit{for all } t \geq t_i.$$

Then we have the following:

Proposition 1 *The CLT in (2.1) and Condition 1 together imply Assumption 1.*

The reference [4] demonstrates that Condition 1 holds for many estimators arising in practice, including the sample mean of i.i.d. samples, maximum likelihood estimators, and quantile estimators.

A slightly stronger assumption than the CLT in (2.1) is a functional central limit theorem (FCLT), which we now describe. Let $C[0,1]$ and $C[0,\infty)$ be the spaces of

continuous functions on $[0,1]$ and $[0,\infty)$, respectively, and let $D[0,1]$ and $D[0,\infty)$ be the spaces of right-continuous functions with left limits on $[0,1]$ and $[0,\infty)$, respectively; see [10] for details. We equip $C[0,1]$ and $D[0,1]$ with the uniform and Skorohod-J_1 topologies, respectively; e.g., see Sections 7 and 12 of [10]. Assume that each estimation process $\widehat{\theta}_i \in D[0,\infty)$. Then an FCLT is as follows:

Condition 2 *There exist finite positive constants η, υ and ω with $2\omega - \upsilon = 2\eta$ such that for each i, $U_{i,n} \Rightarrow U_i$ in $D[0,1]$ as $n \to \infty$, where $U_{i,n} = [U_{i,n}(t) : 0 \le t \le 1]$, $U_i = [U_i(t) : 0 \le t \le 1]$, $U_{i,n}(t) = n^{\eta} t^{\omega} \left[\widehat{\theta}_i(nt) - \theta_i\right]$ for $0 \le t \le 1$, $U_i(t) = \sigma_i B_i(t^{\upsilon})$ for $0 \le t \le 1$, and $B_i = [B_i(t) : t \ge 0]$ a standard Brownian motion.*

As we will see in the examples below, in many applications, η, ω and υ assume the canonical values of 1/2, 1 and 1, respectively (although we do not require this).

For any values of η, ω and υ, the limit U_i in Condition 2 satisfies $P\{U_i \in C[0,1]\} = 1$ since Brownian motion has continuous sample paths with probability 1. Thus, the FCLT in Condition 2 and the continuous-mapping theorem (e.g., Theorem 2.7 of [10]) imply $n^{\eta}\left[\widehat{\theta}_i(n) - \theta_i\right] = U_{i,n}(1) \Rightarrow U_i(1) = \sigma_i B_i(1) \sim N(0, \sigma_i^2)$ as $n \to \infty$, so the FCLT implies the ordinary CLT in (2.1). The reference [8] shows that Condition 1 is true when the FCLT in Condition 2 holds in $C[0,1]$, and his arguments can also be applied to show the same when the FCLT holds in $D[0,1]$ by applying Lemma 1.6.4 and Theorem 1.6.4 of [67]. Therefore, Proposition 1 implies the following:

Proposition 2 *The FCLT in Condition 2 implies the CLT in (2.1) and Condition 1, so Assumption 1 holds.*

We now provide several examples of simulation settings that satisfy Assumptions 1 and 2. In all but the last example, $\eta = 1/2$ in Assumption 1. Other examples for which Assumption 1 holds with $\eta \neq 1/2$ are discussed in Glynn and Whitt [31] and [27].

Example 1 (Means of independent populations) Suppose there are k independent (not necessarily normally distributed) populations and we perform i.i.d. sampling

within each population. The goal is to determine the population i with the largest mean θ_i. Let $Z_{i,1}, Z_{i,2}, \ldots$ be the i.i.d. samples from the ith population, and let $\sigma_i^2 = \mathrm{Var}(Z_{i,1})$, which we assume satisfies $0 < \sigma_i^2 < \infty$. Then we can define the discrete-time estimation process $[\widehat{\theta}_{i,n} : n = 1, 2, \ldots]$ and the discrete-time variance estimation process $[V_{i,n} : n = 2, 3, \ldots]$ with $\widehat{\theta}_{i,n} = (1/n)\sum_{j=1}^{n} Z_{i,j}$ and $V_{i,n} = (1/(n-1))\sum_{j=1}^{n}\left(Z_{i,j} - \widehat{\theta}_{i,n}\right)^2$, respectively. Donsker's theorem (e.g., Theorem 14.1 of [10]) implies that each $\widehat{\theta}_{i,n}$ satisfies the FCLT in Condition 2 with $\omega = \upsilon = 1$, so Assumption 1 holds by Proposition 2. Also, $V_{i,n}$ satisfies Assumption 2; e.g., see p. 73 of [65]. This scenario covers simulating i.i.d. replications in a transient simulation.

Example 2 (Functions of means) For each alternative i, suppose $Z_i = (Z_i^{(1)}, \ldots, Z_i^{(d)}) \in \Re^d$ is a random vector having mean $\nu_i = (\nu_i^{(1)}, \ldots, \nu_i^{(d)}) \in \Re^d$. Assume that $E[\|Z_i\|^2] < \infty$, where $\|\cdot\|$ denotes the Euclidean norm on $\Re^d$, so the covariance matrix Σ_i of Z_i is finite. Also, assume Σ_i is positive definite. Let $\theta_i = g(\nu_i)$, where $g : \Re^d \to \Re$, and assume g is continuously differentiable in a neighborhood of ν_i with $\nabla g(\nu_i) \neq 0$, where ∇g denotes the gradient of g. Again, the goal is to compare $\theta_1, \ldots, \theta_k$. Let $Z_{i,1}, Z_{i,2}, \ldots$ be i.i.d. copies of Z_i. Also, let $\bar{Z}_{i,n} = (\bar{Z}_{i,n}^{(1)}, \ldots, \bar{Z}_{i,n}^{(d)})$, where $\bar{Z}_{i,n}^{(p)} = (1/n)\sum_{j=1}^{n} Z_{i,j}^{(p)}$ is the sample mean of $Z_{i,1}^{(p)}, \ldots, Z_{i,n}^{(p)}$, for $p = 1, \ldots, d$. Then if we take $\widehat{\theta}_{i,n} = g(\bar{Z}_{i,n})$, each $\widehat{\theta}_i$ satisfies the FCLT in Condition 2 with $\upsilon = \omega = 1$ and $\sigma_i^2 = (\nabla g(\nu_i))^\top \Sigma_i (\nabla g(\nu_i))$, where superscript $\top$ denotes transpose (e.g., see Theorem 3 of [31]), so Assumption 1 holds by Proposition 2. Let $\widehat{\Sigma}_{i,n} = (\widehat{\Sigma}_{i,n}^{(p,q)} : p, q = 1, \ldots, d)$ be the sample covariance matrix of Z_i based on $Z_{i,1}, \ldots, Z_{i,n}$; i.e., $\widehat{\Sigma}_{i,n}^{(p,q)} = (1/(n-1))\sum_{j=1}^{n}[Z_{i,j}^{(p)} - \bar{Z}_{i,n}^{(p)}][Z_{i,j}^{(q)} - \bar{Z}_{i,n}^{(q)}]$. It is straightforward to show that $V_{i,n} = (\nabla g(\bar{Z}_{i,n}))^\top \widehat{\Sigma}_{i,n} (\nabla g(\bar{Z}_{i,n}))$ satisfies Assumption 2.

Example 3 (Regenerative processes on the cycle time scale) For each alternative i, let $X_i = [X_i(t) : t \geq 0]$ be a continuous-time stochastic process evolving on a state space S_i under a probability measure P_i having expectation E_i. Let $f_i : S_i \to \Re$ be a reward function on S_i, and we want to compare the alternatives in terms of their long-

run average rewards. Assume that for each alternative i, there exists a sequence of times $A_{i,-1} = 0 \leq A_{i,0} < A_{i,1} < \cdots$ such that X_i is regenerative with respect to the sequence $(A_{i,j} : j \geq 0)$ under measure P_i. For each $j \geq 1$, $[X_i(s) : A_{i,j-1} \leq s < A_{i,j}]$ is the jth regenerative cycle of alternative i. Let $\tau_{i,j} = A_{i,j} - A_{i,j-1}$ be the length of the jth regenerative cycle for process i, and define $Y_{i,j} = \int_{A_{i,j-1}}^{A_{i,j}} f_i(X_i(t))\, dt$, which is the cumulative reward over the jth regenerative cycle of the ith process. It is well known that $(Y_{i,j}, \tau_{i,j})$, $j = 1, 2, \ldots$ are i.i.d. Thus, assuming that $E_i[Y_{i,j}^2] < \infty$ and $E_i[\tau_{i,j}^2] < \infty$, we then can use the setting of Example 2 to handle this case. Specifically, let $Z_{i,j} = (Z_{i,j}^{(1)}, Z_{i,j}^{(2)})$, where $Z_{i,j}^{(1)} = Y_{i,j}$ and $Z_{i,j}^{(2)} = \tau_{i,j}$, and let $g(z_1, z_2) = z_1/z_2$. In this case, we want to estimate $\theta = E_i[Y_{i,1}]/E_i[\tau_{i,1}]$, and its estimator $\widehat{\theta}_{i,n} = g(\bar{Z}_{i,n}^{(1)}, \bar{Z}_{i,n}^{(2)}) = (\sum_{j=1}^n Y_{i,j})/(\sum_{j=1}^n \tau_{i,j})$ satisfies Assumption 1 by Example 2, with $\sigma_i^2 = E_i[(Y_{i,1} - \theta_i \tau_{i,1})^2]/(E_i[\tau_{i,1}])^2$. Note that $\widehat{\theta}_{i,n}$ is the estimator of θ_i based on a fixed number n of cycles simulated. If we let $\bar{\tau}_{i,n} = (1/n)\sum_{j=1}^n \tau_{i,j}$, then $V_{i,n} = (1/(n\bar{\tau}_{i,n}^2))\sum_{j=1}^n \left[Y_{i,j} - \widehat{\theta}_{i,n}\tau_{i,j}\right]^2$ satisfies Assumption 2 ([30]).

Example 4 (Quantiles) Suppose there are k independent (not necessarily normally distributed) populations and we perform i.i.d. sampling within each population. Let H_i denote the distribution function of population i. For $0 < y < 1$, let $\xi_{i,y}$ denote the yth quantile of H_i; i.e., $\xi_{i,y} = \inf\{x : H_i(x) \geq y\} \equiv H_i^{-1}(y)$. The goal is to determine for fixed $0 < p < 1$ the population i with the largest pth quantile $\theta_i = \xi_{i,p}$. Assume that for each i, H_i is differentiable at $\xi_{i,p}$, with $H_i'(\xi_{i,p}) = h_i(\xi_{i,p}) > 0$, where H_i' denotes the derivative of H_i and h_i is the density of H_i (with respect to Lebesgue measure). Let $Z_{i,1}, Z_{i,2}, \ldots$ be i.i.d. samples from distribution H_i, and let $H_{i,n}$ denote the empirical distribution function based on the first n samples; i.e., $H_{i,n}(x) = (1/n)\sum_{j=1}^n 1\{Z_{i,j} \leq x\}$, where $1\{C\}$ denotes the indicator function of the event C. The sample pth quantile is then $H_{i,n}^{-1}(p)$. Consequently, the discrete-time estimator $\widehat{\theta}_{i,n} = H_{i,n}^{-1}(p)$ satisfies the CLT in (2.1) with $\eta = 1/2$ and $\sigma_i^2 = p(1-p)/h^2(\xi_{i,p})$; e.g., see p. 77 of [65]. Assumption 1 holds by Theorem 5 of [4]. To construct a consistent estimator of σ_i^2, let $V_{i,n} = p(1-p)[(H_{i,n}^{-1}(p+\epsilon_n) - H_{i,n}^{-1}(p-\epsilon_n))/(2\epsilon_n)]^2$. Then Bloch and Gastwirth (1968) show that

$V_{i,n}$ satisfies Assumption 2 when $\epsilon_n \to 0$ and $n\epsilon_n \to \infty$ as $n \to \infty$. The reference [11] shows that $\epsilon_n = O(n^{-1/5})$ minimizes the MSE[1] of the estimator of $1/h_i(\xi_{i,p})$. Hall and Sheather (1988) show that $\epsilon_n = O(n^{-1/3})$ asymptotically minimizes coverage error for a confidence interval for a single quantile when using a single-stage procedure. Although this is different than our context, since we consider simultaneous confidence intervals for differences of quantiles and a two-stage procedure, using $\epsilon_n = O(n^{-1/3})$ for our MCB intervals may be a reasonable choice in terms of coverage. Section 2.8.2 explores this issue in some numerical experiments on a small example.

Example 5 (Steady-state simulations) Consider again the model in Example 3, but we no longer assume that X_i is necessarily regenerative. Let

$$\widehat{\theta}_i(t) = \frac{1}{t}\int_0^t f_i(X_i(s))\,ds, \tag{2.9}$$

which is the time-average reward of the process X_i up to time t. Section 4.4 of [75] discusses stochastic processes (including Markov chains, mixing processes, associated processes and martingales) for which $\widehat{\theta}_i(t)$ satisfies the FCLT in Condition 2 with $\omega = \upsilon = 1$ and variance parameter σ_i^2, so Assumption 1 holds by Proposition 2. Examples of estimators of σ_i^2 satisfying Assumption 2 include regenerative estimators ([30]), spectral estimators ([18]), autoregressive estimators ([24], p. 252), and various batch means and batched area estimators in which the number of batches $m \to \infty$ at an appropriate rate as the run length increases ([19]). The references [49] and [12] provide overviews of these techniques.

For example, consider the regenerative case, in which we form our estimator $\widehat{\theta}_i(t)$ in (2.9) based on simulating the process X_i up to a fixed time t, resulting in a random number of cycles completed. This differs from Example 3, which simulated for a fixed number n of cycles. Define $N_i(t) = \sup\{j \geq 0 : A_{i,j} \leq t\}$, which is the number of

[1]If $\widehat{\sigma}^2$ is a candidate estimator of σ^2, then we may evaluate the performance of $\widehat{\sigma}^2$ by determining its bias, $\text{Bias}[\widehat{\sigma}^2] \equiv E[\widehat{\sigma}^2] - \sigma^2$; its variance, $\text{Var}[\widehat{\sigma}^2] \equiv E[(\widehat{\sigma}^2 - E[\widehat{\sigma}^2])^2]$; and finally its mean-squared error, $MSE[\widehat{\sigma}^2] \equiv E[(\widehat{\sigma}^2 - \sigma^2)^2] = \text{Bias}^2[\widehat{\sigma}^2] + \text{Var}[\widehat{\sigma}^2]$.

regenerative cycles that process i completes by time t. Under the same assumptions as in Example 3, the estimator $\widehat{\theta}_i(t)$ in (2.9) satisfies Assumption 1 with $\sigma_i^2 = E_i[(Y_{i,1} - \theta_i\tau_{i,1})^2]/E_i[\tau_{i,1}]$. The variance estimator

$$V_i(t) = \frac{1}{t}\sum_{j=1}^{N_i(t)}\left[Y_{i,j} - \widehat{\theta}_i(t)\tau_{i,j}\right]^2 \tag{2.10}$$

satisfies Assumption 2 ([30]).

Example 6 (Kiefer-Wolfowitz stochastic approximation) For each alternative i, suppose $\beta_i(\rho_i)$ is a real-valued function that can be represented as $\beta_i(\rho_i) = E[Z_i(\rho_i)]$, where $\rho_i \in \Re$ and β_i is three-times differentiable on $\Re$. Suppose β_i is minimized at $\theta_i = \rho_i^*$, and ρ_i^* is the unique solution to $\beta_i'(\rho_i) = 0$, where β_i' denotes the derivative of β_i. We want to compare the alternatives in terms of the $\theta_1, \ldots, \theta_k$. One approach to finding ρ_i^* is to apply to each alternative i the Kiefer-Wolfowitz stochastic approximation algorithm ([44]), which generates a sequence $\rho_{i,1}, \rho_{i,2}, \ldots$ that converges a.s. to ρ_i^*. Specifically, for each alternative i, let $(c_{i,n} : n \geq 0)$ be a deterministic sequence of nonnegative constants, and take $\rho_{i,n+1} = \rho_{i,n} - c_{i,n}X_{i,n+1}$, where $X_{i,n+1}$ is generated independently conditional on $\rho_{i,n}$, i.e.,

$$P\{X_{i,n+1} \in A \mid (\rho_{i,j}, X_{i,j}), j \leq n\} = P\left\{\frac{Z_i(\rho_{i,n} + h_{i,n+1}) - Z_i(\rho_{i,n} - h_{i,n+1})}{2h_{i,n+1}} \in A\right\},$$

with $Z_i(\rho_{i,n} + h_{i,n+1})$ and $Z_i(\rho_{i,n} - h_{i,n+1})$ generated independently. We then define the discrete-time estimation process $\widehat{\theta}_{i,n} = \rho_{i,n}$ for $n \geq 0$. Suppose we choose the constants $c_{i,n}$ and $h_{i,n}$ as $c_{i,n} = c_i n^{-1}$ and $h_{i,n} = h_i n^{-1/3}$ for constants c_i and h_i. Assume that $c_i\beta_i''(\rho_i^*) > 1/3$, where β_i'' denotes the second derivative of β_i. Then under additional mild regularity conditions, [60] establishes the FCLT in Condition 2 with noncanonical $\eta = 1/3$, $\upsilon = 2b_i - 2/3$, $\omega = b_i$, $b_i = c_i\beta_i''(\rho_i^*)$, $\sigma_i^2 = c_i^2\kappa_i^2/((2b_i - 2/3)(4h_i^2))$, and $\kappa_i^2 = 2\mathrm{Var}[Z_i(\rho_i^*)]$, so Assumption 1 holds by Proposition 2.

The reference [73], p. 189, provides some directions on how to construct an estimator $V_{i,n}$ for σ_i^2 so that Assumption 2 holds.

2.7 Asymptotic Comparison of WCVE and STS MCB Methods

We now compare our WCVE MCB methods and the STS MCB procedures of [21], when both are applied in the setting of steady-state simulations from Example 5 of Section 2.6. Hence, each alternative i has a corresponding simulated process $X_i = [X_i(t) : t \geq 0]$ such that $f_i(X_i(t)) \Rightarrow f_i(X_i(\infty))$ as $t \to \infty$, $\theta_i = E[f_i(X_i(\infty))]$ is the steady-state mean reward, and Assumption 1 holds with $\eta = 1/2$. While Damerdji and Nakayama only develop MCB procedures, their methods can also be extended to become asymptotically valid selection procedures by assuming a framework similar to that used in Section 2.4 and Assumption 3.

Before doing the comparison, we need some background on STS methods. Assume each simulated process $X_i \in D[0, \infty)$, and define the estimator $\widehat{\theta}_i(t)$ as in (2.9). The validity of STS methods requires the FCLT in Condition 2 to hold with $\eta = 1/2$ and $\upsilon = \omega = 1$, so $U_i(t) = \sigma_i B_i(t)$ and $U_{i,n}(t) = n^{1/2}\left(\bar{X}_{i,n}(t) - \theta_i t\right)$ with $\bar{X}_{i,n}(t) = \frac{1}{n}\int_0^{nt} f_i(X_i(s))\, ds$. Both $U_{i,n}(t)$ and $U_i(t)$ are continuous at all $t \geq 0$ almost surely, and to simplify the discussion, we modify Condition 2 to hold in $C[0, \infty)$ rather than $C[0, 1]$. We assume that the processes X_i, $i = 1, \ldots, k$, are simulated independently, so the limiting Brownian motions B_i, $i = 1, \ldots, k$, in Condition 2 are independent.

Each STS method has a corresponding function g, whose square, when scaled and applied to an integrated version of the original process (i.e., $\bar{X}_{i,n}$ or $U_{i,n}$), provides an estimate of the process's variance parameter. To define the function g, first define the restriction mapping $r_{[0,1]} : C[0, \infty) \to C[0, 1]$ with $r_{[0,1]}(x)(t) = x(t)$ for $0 \leq t \leq 1$ and $x \in C[0, \infty)$. Let B be a standard one-dimensional Brownian motion on $[0, \infty)$. Assume $g : C[0, \infty) \to \Re$ satisfies the following conditions:

C1. The value of $g(x)$ for $x \in C[0, \infty)$ depends only on $[x(t) : 0 \leq t \leq 1]$; i.e., there exists a function $g_{[0,1]} : C[0, 1] \to \Re$ such that $g(x) = g_{[0,1]}(r_{[0,1]}(x))$ for all $x \in C[0, \infty)$.

C2. $g(\beta x) = \beta g(x)$ for all $\beta \in \Re$ with $\beta > 0$ and $x \in C[0, \infty)$;

C3. $g(x - \beta e) = g(x)$ for all $\beta \in \Re$ and $x \in C[0, \infty)$, where $e \in C[0, \infty)$ with $e(t) = t$;

C4. $P\{g(B) > 0\} = 1$;

C5. $P\{B \in D(g)\} = 0$, where $D(g)$ denotes the set of discontinuities of g;

C6. $g(B)$ has a density function f_g with respect to Lebesgue measure and $f_g(\beta) > 0$ for all $\beta \in \Re$ with $\beta > 0$.

Conditions C2–C5 are from [29], and [55] and [21] added C1 and a weaker version of C6 for two-stage STS procedures. For two-stage procedures, we rescale time so that the first stage is the unit interval, so C1 ensures that g only depends on the process's evolution in the first stage. Instead of C6, [21] assume that $P\{g(B) \in A\} = 0$ for countable sets A. However, all known STS functions g satisfy both conditions, and we believe (although we have no formal proof) that conditions C1–C5 imply C6. Indeed, many (but not all) STS functions g result in $g^2(B)$ having a χ^2 distribution with some number of degrees of freedom, so C6 holds. One exception is Calvin and Nakayama's maximum estimator ([15]), for which $g(B)$ has a Weibull distribution, but this also satisfies C6.

The reference [29] provide functions g for several STS methods, including Schruben 's area estimator ([64]), and standardized maximum estimator. Also, they show that batch means (BM) with a fixed number m of batches has STS function $g_{\mathrm{bm},m} : C[0, \infty) \to \Re$ with

$$g_{\mathrm{bm},m}(x) = \left[\left(\frac{m}{m-1} \right) \sum_{i=1}^{m} \left(\Delta_m x(i/m) - x(1)/m \right)^2 \right]^{1/2}, \tag{2.11}$$

where $\Delta_h x(t) = x(t) - x(t - 1/h)$. Note that $\Delta_m B(i/m)$, $i = 1, \ldots, m$, are increments of B, so are i.i.d. $N(0, 1/m)$. Since $B(1)/m$ is the sample average of these, $(m - 1)g^2_{\mathrm{bm},m}(B)$ is distributed as $\chi^2(m-1)$, a χ^2 distribution with $m-1$ degrees of freedom, so $E[g^2_{\mathrm{bm},m}(B)] = 1$.

We now describe how to apply batching with $m \geq 1$ batches to other STS methods. For each $j = 1, \ldots, m$, define the mapping $\Lambda_{j,m} : C[0,\infty) \to C[0,\infty)$ as

$$\Lambda_{j,m}(x)(t) = \sqrt{m}\left[x\left(\frac{j-1+t}{m}\right) - x\left(\frac{j-1}{m}\right)\right], \quad t \geq 0,$$

for $x \in C[0,\infty)$. Note that if B is a standard Brownian motion, then $\Lambda_{j,m}(B)$ is also a standard Brownian motion. Moreover, the processes $[\Lambda_{j,m}(B)(t) : 0 \leq t \leq 1]$, $j = 1, \ldots, m$, are independent standard Brownian motions on the unit interval by the independent-increments property of B. Then for a STS function g and $p > 0$ such that $E[g^p(B)] < \infty$, we define the batched STS function $\bar{g}_m$ with $m \geq 1$ batches and using the p-norm as

$$\bar{g}_m(x) = \left(\frac{1}{m}\sum_{j=1}^{m} \frac{g^p \circ \Lambda_{j,m}(x)}{E[g^p(B)]}\right)^{1/p}, \tag{2.12}$$

where $f \circ h(x) = f(h(x))$; see [15] and [29].

2.7.1 Overlapping area estimator

Let T be the first-stage sample size, b be the batch size and $m = T/b$ be the ratio of the sample size to the batch size (for more details, see Section 2.8). The bridging map $\Theta : (x,s) \in C[0,\infty) \times [0, \frac{m-1}{m}] \to \Theta(x,s) \in C[0,1]$ is defined by

$$\Theta(x,s)(t) \equiv \sqrt{m}\left[\left[x\left(s+\frac{t}{m}\right) - x\left(s\right)\right] - t\left[x\left(s+\frac{1}{m}\right) - x\left(s\right)\right]\right] \text{ for } t \in [0,1].$$

We may note that $\Theta(B,s)(\cdot)$ is a standard Brownian bridge process on $[0,1]$.

Define a weighting function $w(t)$ that satisfies the following conditions: $w(t)$ is continuous at every $t \in [0,1]$ and

$$\text{Var}\left[\int_0^1 w(t)\overline{Z}(t)dt\right] = 1,$$

where $\overline{Z}(\cdot)$ is a standard Brownian bridge process on $[0,1]$. The weighting function $w(\cdot)$ is normalized so that $\int_0^1\int_0^1 w(s)w(t)(\min(s,t) - st)\,ds\,dt = 1$. Popular choices of

weighting functions are $w_0(t) = \sqrt{12}$, $w_2(t) = \sqrt{840}(3t^2 - 3t + 1/2)$ and $w_{\cos,j}(t) = \sqrt{8}\pi j \cos(2\pi jt)$ with $j = 1, 2, \ldots$ for $t \in [0, 1]$. The last function is known as the trigonometric weighting function and is used in our experiments.

Then the overlapping area map $g : x \in C[0, \infty) \to g(x) \in \Re$ is defined by

$$g^2(x) = \frac{m}{m-1} \int_0^{1-\frac{1}{m}} \left[\int_0^1 w(u)\Theta(x, s)(u)\, du \right]^2 ds.$$

It may be verified that $g : C[0, \infty) \to \Re$ satisfies the conditions C1-C6 of Section 2.7. We shall show some of them. C1 follows from definition of g. We observe that

$$g^2(\beta x) = \frac{m}{m-1} \int_0^{1-\frac{1}{m}} \left[\int_0^1 w(u)\Theta(\beta x, s)(u)\, du \right]^2 ds.$$

Now $\Theta(\beta x, s)(t) = \beta\Theta(x, s)(t)$. Hence

$$\begin{aligned} g^2(\beta x) &= \frac{m}{m-1} \int_0^{1-\frac{1}{m}} \left[\int_0^1 w(u)\beta\Theta(x, s)(u)\, du \right]^2 ds \\ &= \beta^2 \frac{m}{m-1} \int_0^{1-\frac{1}{m}} \left[\int_0^1 w(u)\Theta(x, s)(u)\, du \right]^2 ds \\ &= \beta^2 g^2(x). \end{aligned}$$

So $g(\beta x) = \beta g(x)$ and hence C2 is satisfied. Next, we observe that for all $\beta \in \Re$ and $e \in C[0, \infty)$ with $e(t) = t$, we have

$$\begin{aligned} &\Theta(x - \beta e, s)(t) \\ &= \sqrt{m} \left[\left[(x - \beta e) \left(s + \frac{t}{m} \right) - (x - \beta e)(s) \right] - t \left[(x - \beta e) \left(s + \frac{1}{m} \right) - (x - \beta e)(s) \right] \right] \\ &= \sqrt{m} \left[\left[x \left(s + \frac{t}{m} \right) - x(s) - \beta \left(s + \frac{t}{m} \right) + \beta s \right] - t \left[x \left(s + \frac{1}{m} \right) - x(s) \right. \right. \\ &\left. \left. - \beta \left(s + \frac{1}{m} \right) + \beta s \right] \right] \\ &= \sqrt{m} \left[\left[x \left(s + \frac{t}{m} \right) - x(s) - \beta \frac{t}{m} \right] - t \left[x \left(s + \frac{1}{m} \right) - x(s) - \beta/m \right] \right] \\ &= \sqrt{m} \left[\left[x \left(s + \frac{t}{m} \right) - x(s) \right] - t \left[x \left(s + \frac{1}{m} \right) - x(s) \right] \right] \\ &= \Theta(x, s)(t). \end{aligned}$$

So $g(x - \beta e) = g(x)$. Hence C3 is also satisfied.

The proof of C5 follows from a modification of the proof of Proposition A-1 of [2].

2.7.2 Overlapping CvM estimator

For the details of this estimator, see [2]. We maintain the same notations from the previous subsection. We also define $C(w_N) \equiv \int_0^1 w_N(t)(\overline{Z}(t))^2\,dt$, where $w_N(\cdot)$ is the normalized weighting function (the subscript N is used to emphasize the fact that the weighting function is normalized) so that $E[C(w_N)] = 1$ and $d^2 w_N(t)/dt^2$ is continuous at every $t \in [0, 1]$. Here, $C(w_N)$ is the limiting functional of the weighted CvM estimator for σ^2.

Then we define the overlapping CvM map $g : x \in C[0, \infty) \to g(x) \in \Re$ as

$$g^2(x) = \frac{m}{m-1} \int_0^{1-\frac{1}{m}} \int_0^1 w_N(u) \left[\Theta(x, s)(u)\right]^2 \, du\, ds.$$

2.7.3 Comparison of WCVE and STS MCB procedures

We now describe the two-stage MCB procedures of [21]. Let $T_{0,i}$ be the first-stage run length, and for g_m defined as either $g_{\mathrm{bm},m}$ in (2.11) or $\bar{g}_m$ in (2.12), let

$$V'_{i,m}(T_{0,i}) = T_{0,i}\, g_m^2(\bar{X}_{i,T_{0,i}}) = g_m^2(U_{i,T_{0,i}}) \tag{2.13}$$

as the first-stage estimate of σ_i^2, where the second equality in (2.13) follows from C2 and C3. Then the STS MCB method for constructing intervals with absolute-width parameter δ is the same as Procedure A in Section 2.3 except (2.4) is changed to

$$T'_{i,m}(\delta) = \max\left(T_{0,i}, \frac{\gamma'^2 V'_{i,m}(T_{0,i})}{\delta^2}\right), \tag{2.14}$$

where γ' is a constant to be discussed shortly, and we replace $\widetilde{\theta}_i(\delta)$ in steps 4 and 5 with $\widetilde{\theta}'_i(\delta) = \widehat{\theta}_i(T'_{i,m}(\delta))$. Similar modifications can be made to Procedure R of Section 2.5 to construct STS MCB intervals with relative-width parameter δ. ([21] actually define their two-stage STS MCB procedures to determine the total number of batches to

simulate for each system, but these can easily be modified to determine instead the total run length required, as in (2.14).)

We now provide more details on the constant γ' in (2.14). Suppose that we want to construct STS MCB intervals having asymptotic joint confidence level at least $1-\alpha$. Then $\gamma' = \gamma'(k, 1-\alpha, g_m)$ in (2.14) is chosen to satisfy

$$E\left[\prod_{i=1}^{k-1} \Phi\left(\frac{\gamma'}{[(1/g_m^2(B_i)) + (1/g_m^2(B_k))]^{1/2}}\right)\right] = 1-\alpha. \tag{2.15}$$

In contrast our WCVE MCB method requires the constant $\gamma = \gamma(k, 1-\alpha)$ in (2.4), which satisfies

$$\left[\Phi(\gamma/\sqrt{2})\right]^{k-1} = 1-\alpha, \tag{2.16}$$

so $\gamma = \sqrt{2}\, z_{(1-\alpha)^{1/(k-1)}}$.

When g_m is the BM function $g_{\mathrm{bm},m}$ in (2.11) with $m \geq 2$ batches, $[(1/g_m^2(B_i)) + (1/g_m^2(B_k))]$ in (2.15) is distributed as $(m-1)[(1/\chi_i^2) + (1/\chi_k^2)]$, where $\chi_1^2, \ldots, \chi_k^2$ are independent χ^2 random variables, each with $m-1$ degrees of freedom. In this case the parameter $\gamma'(k, 1-\alpha, g_{\mathrm{bm},m})$ in (2.15) is exactly Rinott's constant ([59]) in his two-stage selection procedure for comparing independent normal populations when the first-stage sample size for each population is m. The reference [6] provide tables of values for γ'.

We are now in a position to compare WCVE and STS MCB procedures in the steady-state simulation setting, where $\eta = 1/2$. We first state our assumptions used in the comparison:

Assumption 4 *The k alternatives are simulated independently, and satisfy the FCLT in Condition 2 in $C[0,\infty)$ with $\eta = 1/2$. The STS function g_m is either the batch means function $g_{bm,m}$ in (2.11) with $m \geq 2$ batches, or the g functions defined in Subsections 2.7.1 and 2.7.2, or the batched STS function $\bar{g}_m$ in (2.12) with $m \geq 1$ batches and $p > 0$ such that $E[g^p(B)] < \infty$ and $E[\bar{g}_m^2(B)] = 1$, where B is a standard Brownian motion. The first-stage run length for each alternative i for both the WCVE and STS MCB methods is $T_{0,i} = \zeta_i \delta^{-2}$ for any constant $\zeta_i > 0$.*

The assumption that $E[\bar{g}_m^2(B)] = 1$, which we earlier showed holds for BM, is not restrictive as long as $E[\bar{g}_m^2(B)] < \infty$. It ensures that the limiting variance estimator is unbiased since (2.13) and the FCLT imply $V'_{i,m}(T_{0,i}) \Rightarrow g_m^2(\sigma_i B_i) = \sigma_i^2 g_m^2(B_i)$ as $\delta \to 0$ by C2.

We now compare the STS and WCVE MCB methods in terms of their total run lengths $T'_{i,m}(\delta)$ and $T_i(\delta)$. We also compare the methods in terms of the *potential* total run lengths, which we define as the second terms in the maximums in (2.14) and (2.4). Specifically, these are

$$\bar{T}'_{i,m}(\delta) = \frac{\gamma'^2 V'_{i,m}(T_{0,i})}{\delta^2} \quad \text{and} \quad \bar{T}_i(\delta) = \frac{\gamma^2 V_i(T_{0,i})}{\delta^2}, \tag{2.17}$$

for the STS and WCVE MCB methods, respectively. For the comparisons, we define the ratios $R_{i,m}(\delta) = T'_{i,m}(\delta)/T_i(\delta)$ and $\bar{R}_{i,m}(\delta) = \bar{T}'_{i,m}(\delta)/\bar{T}_i(\delta)$, and recall the definition of τ_i in (2.5).

Theorem 4 *Under Assumptions 2 and 4, the following hold for each number $m \geq 1$ of batches and for each alternative i:*

(i) $\delta^2 T'_{i,m}(\delta) \Rightarrow \tau'_{i,m}$ *as* $\delta \to 0$, *where* $\tau'_{i,m} = \max[\zeta_i, \gamma'^2 \sigma_i^2 g_m^2(B_i)] > 0$ *a.s. is nondegenerate.*

(ii) $\gamma'(k, 1-\alpha, g_m) > \gamma(k, 1-\alpha)$.

(iii) $R_{i,m}(\delta) \Rightarrow R_{i,m}$ *and* $\bar{R}_{i,m}(\delta) \Rightarrow \bar{R}_{i,m}$ *as* $\delta \to 0$, *where* $R_{i,m} = \frac{\tau'_{i,m}}{\tau_i}$ *and* $\bar{R}_{i,m} = \frac{\gamma'^2 g_m^2(B_i)}{\gamma^2}$.

(iv) If $\{V'_{i,m}(t) : t > 0\}$ *is uniformly integrable, then* $\lim_{\delta \to 0} E[R_{i,m}(\delta)] > 1$. *If* $\{V'_{i,m}(t)/V_i(t) : t > 0\}$ *is uniformly integrable, then* $\lim_{\delta \to 0} E\left[\bar{R}_{i,m}(\delta)\right] = \left[\frac{\gamma'(k,1-\alpha,g_m)}{\gamma(k,1-\alpha)}\right]^2 > 1$.

Theorem 4(iv) shows that on average, the total run length for STS MCB methods is asymptotically strictly larger than that for WCVE MCB methods. The reason is that

for a fixed number m of batches, the STS variance estimator is not consistent as the first-stage run length grows, whereas the estimator $V_i(t)$ is consistent by Assumption 2.

We now examine the variability of the total run lengths $T_i(\delta)$ and $T'_{i,m}(\delta)$ (defined in (2.4) and (2.14), respectively) and the potential total run lengths $\bar{T}_i(\delta)$ and $\bar{T}'_{i,m}(\delta)$ (defined in (2.17)) of the WCVE and STS MCB methods. Theorem 1(ii) shows $\delta^2 T_i(\delta) \Rightarrow \tau_i$ as $\delta \to 0$ for WCVE MCB procedures. For STS MCB methods with a fixed number $m \geq 1$ of batches, Theorem 4(i) gives $\delta^2 T'_{i,m}(\delta) \Rightarrow \tau'_{i,m}$ as $\delta \to 0$. In the case of WCVE MCB, $\tau_i > 0$ is deterministic, whereas for STS MCB, $\tau'_{i,m}$ is a nondegenerate positive random variable. Hence, we see that STS MCB methods have asymptotically more variable total run lengths than WCVE MCB procedures. The reference [57] quantifies this observation another way by comparing the limiting variances (appropriately normalized) of $T_i(\delta)$, $T'_{i,m}(\delta)$, $\bar{T}_i(\delta)$ and $\bar{T}'_{i,m}(\delta)$.

We now establish what happens to STS MCB methods as the number of batches grows large.

Theorem 5 *Under Assumptions 2 and 4, the following hold:*

(i) $\gamma'(k, 1-\alpha, g_m) \to \gamma(k, 1-\alpha)$ *as* $m \to \infty$.

(ii) $\lim_{m\to\infty} \lim_{\delta\to 0} \delta^2 T'_{i,m}(\delta) \Rightarrow \tau_i$, *where* τ_i *is defined in (2.5).*

(iii) $R_{i,m} \Rightarrow 1$ *and* $\bar{R}_{i,m} \Rightarrow 1$ *as* $m \to \infty$.

(iv) If $\{g_m^2(B) : m \geq 1\}$ *is uniformly integrable, then* $E[R_{i,m}] \to 1$ *and* $E\left[\bar{R}_{i,m}\right] \to 1$ *as* $m \to \infty$.

From this we now can make the following comparisons of the STS MCB procedures applied to the limiting Brownian motions and our WCVE MCB methods. As the number of batches grows large, STS MCB methods become comparable to WCVE MCB methods in terms of overall run lengths by Theorem 5(iii)–(iv). But for any fixed number of batches, STS MCB methods have asymptotically longer run lengths on average than WCVE MCB methods by Theorem 4(iv).

As an illustration, we list values of $\gamma(k, 1-\alpha)$ for the WCVE MCB methods and $\gamma'(k, 1-\alpha, g_{\mathrm{bm},m})$ for the BM MCB methods for $k = 4$ systems and nominal confidence level $1-\alpha = 0.9$ for different numbers m of batches:

$$\begin{aligned}\gamma'(4, 0.9, g_{\mathrm{bm},10}) &= 2.912,\\ \gamma'(4, 0.9, g_{\mathrm{bm},30}) &= 2.666,\\ \gamma'(4, 0.9, g_{\mathrm{bm},50}) &= 2.627,\\ \gamma(4, 0.9) &= 2.571.\end{aligned}$$

The values of $\gamma'(4, 0.9, g_{\mathrm{bm},m})$ are always strictly greater than $\gamma(4, 0.9)$, as noted by Theorem 4(ii), but $\gamma'(4, 0.9, g_{\mathrm{bm},m})$ approaches $\gamma(4, 0.9)$ as m gets large, which is consistent with Theorem 5(i). For comparison, $\bar{\gamma}(4, 0.9) = 2.452$; see Theorem 1(iii).

2.8 Experimental Results

In this section, we demonstrate the asymptotic validity of our procedures as the prescribed width of the confidence intervals shrinks to zero. Also, for the steady-state simulation context, we compare our procedures based on WCVEs (see Sections 2.3 and 2.5) with techniques that instead use standardized time series methods (see Section 2.7). The simulation experiments are conducted for comparison of the mean waiting time of M/M/1 queues for both medium and heavy traffic intensities. We have also considered the comparison of a fixed quantile of the longest path of different stochastic activity networks (SANs).

2.8.1 M/M/1 queues

We ran simulation experiments comparing the WCVE and STS MCB procedures with relative-width parameter δ. We compared $k = 4$ systems, where each system corresponds to an M/M/1 queue. For each system i, we simulated the discrete-time process

of waiting times $[W_{i,j} : j = 0, 1, 2, \ldots]$. Each system's simulation started with $W_{i,0} = 0$, i.e., the initial customer has no wait, and we used Lindley's equation to simulate the successive waiting times as $W_{i,j+1} = \max(W_{i,j} + S_{i,j} - A_{i,j}, 0)$, where $S_{i,j}$ is the service time of the jth customer and $A_{i,j}$ is the interarrival time between the jth and $(j+1)$th customers in the ith system. We conducted two sets of experiments. For both sets, the service rate is 1. In the first set, the 4 systems have interarrival rates 0.50, 0.51, 0.51, 0.51, respectively (see Tables 2.1 and 2.2). In the second set, we deal with higher traffic intensities and consider the following interarrival rates: 0.80, 0.81, 0.81 and 0.81 (see Tables 2.3 and 2.4).

Let θ_i be the steady-state expected waiting time for system i. We know that $\theta_i = E^2[S_{j,1}]/(E[A_{i,1}] - E[S_{j,1}])$. Each system i is regenerative with regenerations occurring when $W_{i,j} = 0$, so Assumption 1 holds with $\eta = 1/2$ (see Example 5 in Section 2.6). We compare the systems in terms of the θ_i, where we assume smaller θ_i is better.

We use the regenerative method's (RM) variance estimator in (2.10) in the WCVE MCB method. We applied Procedure R using γ rather than $\bar{\gamma}$; see Theorem 3(iii). For the STS MCB procedure, we employed batch means with either $m = 10$ batches (BM10) or 20 batches (BM20).

For the STS MCB procedure, we also used the overlapping area estimator with trigonometric weighting functions, which we briefly described in Section 2.7.1. For the details of this algorithm, please refer to [1] and its online supplement. In Section 2.7.1, we defined the overalapping area estimator as applied to a continuous-time process, and we now explain how to compute it for a discrete-time process.

Suppose we have a discrete-time simulation output process $\{X_i : i = 1, 2, \ldots\}$. Let T be the first-stage run length. Consider a sample $\{X_i : i = 1, \ldots, T\}$, which is divided into $(T - b + 1)$ overlapping batches, each of size b so that the observations $\{X_{i+k} : k = 0, \ldots, b-1\}$ constitute batch i for $i = 1, \ldots, T-b+1$. Let $m \equiv T/b$ denote the ratio of the sample size to the batch size. The standardized time series computed

from overlapping batch i is

$$Y^o_{i,b}(t) \equiv \lfloor bt \rfloor (\overline{X}^o_{i,b} - \overline{X}^o_{i,\lfloor bt \rfloor})/\sqrt{b}$$

for $t \in [0,1]$ and $i = 1, \ldots, T-b+1$, where

$$\overline{X}^o_{i,j} \equiv \frac{1}{j} \sum_{k=0}^{j-1} f(X_{i+k}),$$

for $i = 1, \ldots, T-b+1$ and $j = 1, \ldots, b$. We define the above function f as the identity function $f(x) = x$ in the experiments.

We define the overlapping area estimator computed from overlapping batch i by

$$A^o_i(w;b) \equiv \left[\frac{1}{b} \sum_{k=1}^{b} w\left(\frac{k}{b}\right) Y^o_{i,b}\left(\frac{k}{b}\right) \right]^2, i = 1, \ldots, T-b+1.$$

The overlapping area estimator for σ^2 is

$$A^o(w;m,b) \equiv \frac{1}{T-b+1} \sum_{i=1}^{T-b+1} A^o_i(w;b).$$

In particular, we have used the following estimator:

$$\overline{A}^o_l(w_{\cos,j};m,b) \equiv (1/l) \sum_{j=1}^{l} A^o(w_{\cos,j};m,b),$$

where $w_{\cos,j}(t) = \sqrt{8}\pi j \cos(2\pi j t)$. This family of estimators is first-order unbiased, since each of them is the average of l first-order unbiased estimators. We have used this estimator with $l = 1$ (though it is possible to choose $l = 2, 3, \ldots$), batch size $T/10$ (which we denote as Area10) or $T/20$ (denoted as Area20), where T is the first-stage run length. In the initialization step, we arbitrarily choose the user-selected positive integer $j = 1$. Other possible choices could be $2, 3, \ldots$.

From [1], we find that the area estimator $\overline{A}^o_1(w_{\cos,1};m,b)$ that we used here approximately follows a Chi-square distribution:

$$\overline{A}^o_1(w_{\cos,1};m,b) \sim \frac{E(\overline{A}^o_1(w_{\cos,1};m,b))\chi^2(v_{\text{eff}})}{v_{\text{eff}}},$$

Here, $\chi^2(k)$ is a Chi-square distribution with k degrees of freedom and $v_{\text{eff}} = \text{Round}\left(\frac{2E^2(A)}{\text{Var}(A)}\right)$ and Round(z) denotes rounding of z towards the nearest integer. The variance of the estimator is given by

$$\text{Var}(\overline{A}_1^o(w_{\cos,1}; m, b)) = \frac{((16\pi^2 + 30)m - (20\pi^2 + 33))\sigma^4}{(24\pi^2(m-1)^2}.$$

For the overlapping area estimator, $E(\overline{A}_1^o(w_{\cos,1}; m, b)) \approx \sigma^2$. See Theorem 3 of [2] for details. The values of v_{eff} are 23 and 48 for $m = 10$ and $m = 20$ respectively. The values of γ' are available from Table 2.13, p. 62 of [6] and also from the FORTRAN program available from [7].

Recall we defined the first-stage run length for system i as $T_{0,i} = \zeta_i \delta^{-\lambda}$, and in our experiments we took $\zeta_i = \zeta$ for all i. We varied the values of ζ, δ, and λ, which gave rise to different first-stage run lengths, where the run length denotes the number of customers simulated. We used $\zeta = 256$, 1024, 4096; $\delta = 0.2$, 0.1, 0.05; and $\lambda = 1$, 2 for the first set of experiments (moderate traffic intensity) and $\zeta = 8320$, 16640, 33280; $\delta = 0.2$, 0.1, 0.05; and $\lambda = 1$, 2 for the second set of experiments (heavy traffic intensity).

We ran 2000 independent replications in each experiment to estimate the MCB intervals' joint coverage, which had a nominal lower bound of $1 - \alpha = 0.9$. In each replication we computed the sum of the scaled total run lengths across all systems as $\sum_{i=1}^4 \delta^2 T_i(\delta)$ and $\sum_{i=1}^4 \delta^2 T'_{i,m}(\delta)$. We then calculated the sample mean and standard deviation of these quantities across the replications.

Table 2.1 and Table 2.3 contain the coverage results for the first set of experiments and the second set of experiments respectively. For each method, coverage increases as δ decreases for fixed ζ, with coverages larger than or just slightly below 90% for at least the larger values of ζ when δ is small. Thus, we see the asymptotic validity of Theorem 3(i) and procedure R.1 of [21] taking effect. Coverage also increases as ζ increases for fixed δ for each method. In addition, the coverages are higher for $\lambda = 2$ than for $\lambda = 1$, so it appears that the shorter first-stage lengths ($\lambda = 1$) require smaller

Table 2.1 Coverage results (in percent) for moderate traffic intensity cases

		$\lambda = 1$			$\lambda = 2$		
		ζ			ζ		
δ	Method	256	1024	4096	256	1024	4096
0.05	BM10	61.30	70.50	92.20	78.00	96.00	100.00
	Area10	61.35	69.40	92.25	77.70	95.80	99.85
	BM20	60.70	69.55	92.30	77.85	95.85	99.90
	Area20	60.20	68.90	92.30	77.30	95.55	99.55
	RM	60.10	69.20	92.15	77.20	95.50	99.45
0.1	BM10	60.30	66.30	84.85	77.80	95.15	99.80
	Area10	57.90	65.25	84.80	77.70	94.80	98.85
	BM20	57.65	65.85	84.70	77.40	94.65	99.30
	Area20	57.15	66.00	84.55	77.05	94.70	98.75
	RM	56.05	65.75	84.60	77.00	94.50	98.20
0.2	BM10	55.90	62.20	74.45	73.70	94.35	97.75
	Area10	54.20	61.05	74.25	73.05	93.65	97.40
	BM20	53.75	61.10	74.40	73.20	93.90	97.55
	Area20	52.40	60.05	74.05	72.85	93.55	97.35
	RM	52.05	59.75	74.10	72.80	92.50	96.80

δ for the asymptotics to take effect than for the longer first-stage lengths ($\lambda = 2$). All of the coverages for $\lambda = 1$ are below the nominal lower bound of 90%, so we see the asymptotic validity of the MCB intervals now appears to hold. For almost all the experiments, BM10 has higher coverage than RM, which seems to indicate that batch means with a small number of batches is more robust than the regenerative method. In contrast, when the coverages are below the nominal lower bound, the values for BM20 are lower than both RM and BM10. This appears to be because for a fixed run length, the Brownian approximation on which BM is based breaks down as the number of batches increases; this leads to a degradation in the BM estimate of the variance as m increases, which leads to poorer coverage. We also observe that when the coverages are above the nominal level, coverage of Area10 is greater than or equal to the coverage of Area20, which in turn is greater than RM.

Tables 2.2 and 2.4 contain the results of the sum of the scaled total run lengths across all systems for the first and the second set of experiments respectively. We begin by making the following observations for the combinations of method, δ, λ and ζ that have coverage levels in Table 2.1 at least or just slightly below 90%. First, for

Table 2.2 Scaled total run lengths ($\times 10^3$) for moderate traffic intensity cases

		$\lambda = 1$						$\lambda = 2$					
		Mean			Std. Dev.			Mean			Std. Dev.		
		ζ			ζ			ζ			ζ		
δ	Method	256	1024	4096	256	1024	4096	256	1024	4096	256	1024	4096
0.05	BM10	9.01	11.57	12.92	1.41	0.94	0.72	11.92	12.75	17.26	0.73	0.73	0.71
	Area10	8.80	11.16	12.21	1.13	0.53	0.44	11.44	12.41	16.95	0.54	0.43	0.42
	BM20	8.50	10.80	11.82	1.09	0.48	0.42	11.73	12.40	16.68	0.44	0.39	0.38
	Area20	8.37	10.79	11.77	0.93	0.42	0.33	11.30	12.20	16.31	0.32	0.29	0.24
	RM	8.11	10.28	11.57	0.90	0.40	0.24	11.35	12.16	16.31	0.25	0.23	0.17
0.1	BM10	8.39	11.04	12.40	2.34	0.88	0.78	11.59	12.26	16.81	0.83	0.68	0.72
	Area10	8.23	10.81	11.72	2.09	0.67	0.54	11.39	11.72	16.59	0.51	0.36	0.43
	BM20	8.19	10.29	11.51	1.93	0.74	0.48	10.98	11.72	16.51	0.54	0.37	0.41
	Area20	8.12	10.26	11.48	1.14	0.52	0.36	10.98	11.34	16.38	0.44	0.25	0.28
	RM	8.04	10.17	11.25	0.94	0.37	0.23	10.73	11.15	16.37	0.25	0.18	0.21
0.2	BM10	8.29	10.68	11.82	1.44	1.45	0.67	11.16	11.58	16.48	1.14	0.75	0.78
	Area10	8.03	10.34	11.28	1.09	1.31	0.58	10.73	11.35	16.49	0.82	0.55	0.54
	BM20	7.90	10.17	11.22	1.02	1.15	0.64	10.61	11.27	16.33	0.77	0.47	0.44
	Area20	7.57	10.19	11.18	1.00	1.14	0.53	10.59	11.13	16.39	0.71	0.35	0.33
	RM	7.51	9.95	10.86	0.88	0.34	0.28	10.40	11.17	16.35	0.62	0.28	0.24

Table 2.3 Coverage results (in percent) for high traffic intensity cases

		$\lambda = 1$			$\lambda = 2$		
		ζ			ζ		
δ	Method	8320	16640	33280	8320	16640	33280
0.05	BM10	78.45	82.85	87.45	93.90	97.70	98.80
	Area10	78.35	82.80	85.95	93.60	97.20	98.75
	BM20	77.55	83.20	86.70	93.70	97.10	98.70
	Area20	78.35	82.75	85.60	93.60	97.20	98.45
	RM	77.50	82.85	85.90	93.80	97.20	98.60
0.1	BM10	68.40	83.65	89.35	88.50	93.95	98.50
	Area10	67.25	82.90	87.80	88.25	93.50	97.50
	BM20	68.80	84.40	87.80	88.10	93.75	97.90
	Area20	67.65	81.95	86.85	88.00	93.10	97.50
	RM	66.70	82.30	86.95	88.05	93.95	97.75
0.2	BM10	58.80	78.75	82.40	79.75	89.60	96.55
	Area10	59.65	77.80	82.05	79.85	88.30	95.95
	BM20	59.60	78.25	81.55	81.00	88.20	96.00
	Area20	59.95	77.35	82.55	79.65	89.80	95.40
	RM	59.40	77.30	81.15	79.40	89.20	94.75

these cases, RM has smaller or equal average total run length than BM10 and BM20, and also Area10 and Area20, which agrees with Theorem 4. The same is true for the standard deviation of the total run lengths; see the discussion about run-length variability following the proof of Theorem 4. Also, BM20 has average total run length close to that of RM, which is consistent with Theorem 5. However, the variance for BM20 is larger than the variance for RM, so it appears the asymptotics as the number m of batches grows large hold sooner for the mean run length than for the variance. Finally, the asymptotic (as δ gets small) orderings for the mean and variance take effect sooner for $\lambda = 2$ than for $\lambda = 1$. The mean run length of BM20, Area20 and RM are very close to each other.

2.8.2 Stochastic Activity Networks (SAN)

SANs are useful in modeling, planning and management of large projects. We consider the same SAN used by [40]. There are 3 paths and 5 activities in the network. The duration of the ith activity is $A_i, 1 \leq i \leq 5$, which are i.i.d. exponentials with rate 1. Let us denote the density function of A_i by f_i for $1 \leq i \leq 5$. Hence $f_i(t) = \exp(-t)$

Table 2.4 Scaled total run lengths ($\times 10^6$) for high traffic intensity cases

		$\lambda = 1$						$\lambda = 2$					
		Mean			Std. Dev.			Mean			Std. Dev.		
		ζ			ζ			ζ			ζ		
δ	Method	8320	16640	33280	8320	16640	33280	8320	16640	33280	8320	16640	33280
0.05	BM10	0.90	1.11	1.37	0.77	0.85	0.97	2.16	2.53	2.83	1.07	1.05	0.80
	Area10	0.77	0.97	1.19	0.67	0.74	0.84	1.95	2.24	3.11	0.97	0.92	0.86
	BM20	0.78	0.97	1.20	0.67	0.72	0.83	1.93	2.26	2.82	0.95	0.93	1.02
	Area20	0.73	0.92	1.13	0.62	0.70	0.79	1.86	2.17	2.88	0.93	0.89	0.85
	RM	0.71	0.88	1.09	0.59	0.67	0.76	1.79	2.09	2.77	0.89	0.85	0.61
0.1	BM10	1.05	1.34	1.60	1.63	1.91	2.08	2.20	2.96	3.75	2.51	2.91	3.24
	Area10	0.93	1.16	1.38	1.49	1.70	1.79	1.92	2.51	3.26	2.18	2.37	2.84
	BM20	0.92	1.16	1.38	1.44	1.64	1.85	1.94	2.56	3.32	2.20	2.39	2.82
	Area20	0.88	1.11	1.33	1.42	1.60	1.74	1.84	2.43	3.11	2.08	2.34	2.65
	RM	0.84	1.07	1.28	1.33	1.55	1.70	1.75	2.33	2.99	1.94	2.23	2.53
0.2	BM10	1.15	1.26	1.75	3.63	3.03	4.33	1.93	2.69	3.22	4.65	5.60	6.31
	Area10	0.99	1.08	1.52	2.95	2.57	3.85	1.69	2.35	2.71	4.21	4.97	5.21
	BM20	0.98	1.07	1.53	2.95	2.46	3.77	1.69	2.38	2.79	4.10	4.95	5.37
	Area20	0.93	1.02	1.43	2.83	2.40	3.51	1.62	2.24	2.59	4.07	4.78	5.04
	RM	0.89	0.98	1.37	2.72	2.30	3.38	1.56	2.16	2.46	3.99	4.70	4.74

for $t \geq 0$. We specify the jth path by the set of activities on path j and denote the set by B_j. Let $B_1 = \{1, 2\}$, $B_2 = \{1, 3, 5\}$ and $B_3 = \{4, 5\}$. The length of path j is $T_j = \sum_{i \in B_j} A_i$. The length of the longest path is $X = \max(T_1, T_2, T_3)$.

Our objective is to compare $k = 4$ versions of the SAN described above. Since we shall work with more than one SAN, we shall use i to index the ith SAN, for $i = 1, 2, 3, 4$. In particular, X_i will denote the length of the longest path of the ith SAN and $\xi_{p,i}$ will denote the pth quantile of X_i. The four SANs are as in the previous paragraph but three of them have an additional "delay" element introduced at the end, while one SAN does not have the delay element. For our experiments, we have chosen the value of the delay element as 0.30. The SANs will be compared in terms of $\xi_{p,i}$ for $i = 1, 2, 3, 4$ and $p = 0.95$. Here, we assume that smaller values are better.

We used the variance estimator given in Example 4. We have considered two different choices of ϵ_n viz., $\epsilon_n = cn^{-1/5}$ and also $\epsilon_n = cn^{-1/3}$, where n is the first-stage run length and the constant c is chosen as 0.1. In this case also, the first-stage run length for system i is taken as $n = \zeta\delta^{-\lambda}$. Table 2.5 contains the coverage results and Table 2.6 contains the results of the sum of the scaled total run lengths across all systems. We observe that in this case, the coverage values are usually higher when we choose $\epsilon_n = 0.1n^{-1/5}$ instead of choosing $\epsilon_n = 0.1n^{-1/3}$.

Since the run lengths are higher and less variable in the case of the estimator given by [11] (i.e., the estimator corresponding to $\epsilon_n = 0.1n^{-1/5}$) at least when $\lambda = 1$, this might lead to higher coverage. The theory of asymptotic coverage error minimization in the case of a single quantile is probably not applicable for the comparison of multiple quantiles and that could be the possible reason behind the poorer coverage of the estimator given by [34] (i.e., the estimator corresponding to $\epsilon_n = 0.1n^{-1/3}$).

Table 2.5 Coverage results (in percent) for the 0.95 quantile of the longest path of a SAN

δ	$\frac{\epsilon_n}{0.1}$	$\lambda = 1$			$\lambda = 2$		
		ζ			ζ		
		1024	2048	4096	1024	2048	4096
0.05	$n^{-1/3}$	45.40	51.20	62.95	72.65	87.40	96.95
	$n^{-1/5}$	49.55	54.05	64.20	73.30	88.15	97.30
0.1	$n^{-1/3}$	44.20	47.70	53.30	71.20	86.80	95.75
	$n^{-1/5}$	46.30	48.50	54.65	72.65	87.65	96.20
0.2	$n^{-1/3}$	42.60	44.25	45.45	68.90	83.75	94.30
	$n^{-1/5}$	45.00	46.55	47.55	69.20	83.95	94.35

2.9 Proofs

2.9.1 Proof of Theorem 1

We start by proving part (i). Since $T_{0,i} = \zeta_i \delta^{-\lambda}$ for $\zeta_i > 0$ and $0 < \lambda \leq 1/\eta$, (2.4) implies

$$\tau_i(\delta) \equiv \delta^{1/\eta} T_i(\delta) = \max\left(\zeta_i \delta^{(1/\eta)-\lambda}, \left(\gamma\sqrt{V_i(T_{0,i})}\right)^{1/\eta}\right). \tag{2.18}$$

Hence, Assumption 2 and the generalized continuous-mapping theorem (e.g., Theorem 3.4.4 of [75]) imply

$$\delta^{1/\eta} T_i(\delta) = \tau_i(\delta) \Rightarrow \tau_i \tag{2.19}$$

as $\delta \to 0$, where τ_i is defined in (2.5) and is deterministic, which proves the first part of (i). The uniform-integrability assumption and Theorem 3.5 of Billingsley (1999) establish the rest of (i).

Now we establish (ii). Define the event

$$G(\delta) = \left\{\left(\widetilde{\theta}_i(\delta) - \widetilde{\theta}_{(k)}(\delta)\right) - \left(\theta_i - \theta_{(k)}\right) < \delta, \forall i \neq (k)\right\}, \tag{2.20}$$

and we can establish

$$G(\delta) \subseteq \left\{\theta_i - \max_{j \neq i} \theta_j \in I_i(\delta),\ i = 1, 2, \ldots, k\right\} \tag{2.21}$$

Table 2.6 Scaled total run lengths (×100) for the 0.95 quantile of the longest path of a SAN

		$\lambda = 1$						$\lambda = 2$					
		Mean			Std. Dev.			Mean			Std. Dev.		
		ζ			ζ			ζ			ζ		
δ	$\frac{\epsilon_n}{0.1}$	1024	2048	4096	1024	2048	4096	1024	2048	4096	1024	2048	4096
0.05	$n^{-1/3}$	8.26	8.37	8.61	0.76	0.54	0.33	41.01	81.92	163.86	0.084	0.049	0.051
	$n^{-1/5}$	8.67	8.78	8.70	0.43	0.39	0.21	41.32	82.55	164.50	0.065	0.044	0.038
0.1	$n^{-1/3}$	8.55	8.83	16.33	0.88	0.48	0.09	41.49	82.77	163.91	0.069	0.072	0.045
	$n^{-1/5}$	8.87	8.99	16.47	0.55	0.36	0.07	41.64	82.80	164.16	0.056	0.037	0.028
0.2	$n^{-1/3}$	9.16	16.42	32.86	0.77	0.23	0.07	41.85	82.87	164.62	0.081	0.033	0.067
	$n^{-1/5}$	9.26	16.70	33.21	0.55	0.04	0.03	41.89	82.98	164.82	0.053	0.022	0.019

by following a line of reasoning developed by Hsu (1984) (also see pages 150–151 of [38] and [21]). We now show that

$$\lim_{\delta \to 0} P\{G(\delta)\} > 1 - \alpha. \tag{2.22}$$

Define

$$W_i(\delta) = \frac{1}{\delta}\left[\widetilde{\theta}_i(\delta) - \theta_i\right] = \frac{1}{\tau_i^{\eta}}(\tau_i \delta^{-1/\eta})^{\eta}\left[\widehat{\theta}_i(T_i(\delta)) - \theta_i\right], \tag{2.23}$$

and note that

$$G(\delta) = \left\{ \frac{W_i(\delta) - W_{(k)}(\delta)}{\left[\left(\sigma_i^2/\tau_i^{2\eta}(\delta)\right) + \left(\sigma_{(k)}^2/\tau_{(k)}^{2\eta}(\delta)\right)\right]^{1/2}} < \frac{1}{\left[\left(\sigma_i^2/\tau_i^{2\eta}(\delta)\right) + \left(\sigma_{(k)}^2/\tau_{(k)}^{2\eta}(\delta)\right)\right]^{1/2}}, \forall i \neq (k) \right\}.$$

Observe that each $\tau_j(\delta) \geq \left(\gamma\sqrt{V_j(T_{0,j})}\right)^{1/\eta} \geq 0$ by (2.18), so $\tau_j^{2\eta}(\delta) \geq \gamma^2 V_j(T_{0,j})$, and

$$H(\delta) \subseteq G(\delta), \tag{2.24}$$

where

$$\begin{aligned} H(\delta) \;=\; & \left\{ \frac{W_i(\delta) - W_{(k)}(\delta)}{\left[\left(\sigma_i^2/\tau_i^{2\eta}(\delta)\right) + \left(\sigma_{(k)}^2/\tau_{(k)}^{2\eta}(\delta)\right)\right]^{1/2}} \right. \\ & \left. - \; \frac{\gamma}{\left[\left(\sigma_i^2/V_i(T_{0,i})\right) + \left(\sigma_{(k)}^2/V_{(k)}(T_{0,(k)})\right)\right]^{1/2}} < 0, \forall i \neq (k) \right\}. \end{aligned} \tag{2.25}$$

We now want to show that

$$\lim_{\delta \to 0} P\{H(\delta)\} > 1 - \alpha, \tag{2.26}$$

which will then establish (2.22).

To prove (2.26), first recall (2.19), where $0 < \tau_i < \infty$ is deterministic. Hence, (2.23) and Assumption 1 ensure $(W_i(\delta), i = 1, \ldots, k) \Rightarrow (W_i, i = 1, \ldots, k)$ as $\delta \to 0$, where each $W_i \sim N(0, \sigma_i^2/\tau_i^{2\eta})$ and W_i, $i = 1, \ldots, k$, are independent. Since $V_i(T_{0,i}) = V_i(\zeta_i \delta^{-\lambda}) \Rightarrow \sigma_i^2$ by Assumption 2 with σ_i^2 deterministic, $((W_i(\delta), \tau_i(\delta), V_i(\zeta_i \delta^{-\lambda}))$, $i =$

$1,\ldots,k) \Rightarrow ((W_i,\tau_i,\sigma_i^2),\ i=1,\ldots,k)$ as $\delta \to 0$ by (2.19) and Theorem 3.9 of [10]. The continuous-mapping theorem then implies

$$\left(\frac{W_i(\delta) - W_{(k)}(\delta)}{\left[\left(\sigma_i^2/\tau_i^{2\eta}(\delta)\right) + \left(\sigma_{(k)}^2/\tau_{(k)}^{2\eta}(\delta)\right)\right]^{1/2}} - \frac{\gamma}{\left[\left(\sigma_i^2/V_i(\zeta_i\delta^{-\lambda})\right) + \left(\sigma_{(k)}^2/V_{(k)}(\zeta_i\delta^{-\lambda})\right)\right]^{1/2}}, \forall i \neq (k) \right)$$
$$\Rightarrow \quad \left(Z_i - \frac{\gamma}{\sqrt{2}}, \forall i \neq (k) \right) \tag{2.27}$$

as $\delta \to 0$, where $Z_i = [W_i - W_{(k)}]/[(\sigma_i^2/\tau_i^{2\eta}) + (\sigma_{(k)}^2/\tau_{(k)}^{2\eta})]^{1/2}$. Because the limiting random vector in (2.27) is multivariate normal and therefore has a continuous distribution, $P\{Z_i - \gamma/\sqrt{2} = 0, \forall i \neq (k)\} = 0$, so

$$\lim_{\delta\to 0} P\{H(\delta)\} = P\left\{ Z_i - \frac{\gamma}{\sqrt{2}} < 0, \forall i \neq (k) \right\} \tag{2.28}$$

by the Portmanteau theorem (e.g., Theorem 2.1 of [10]).

Each $Z_i \sim N(0,1)$, and for $i \neq j$ with $i, j \neq (k)$,

$$\mathrm{Cov}(Z_i, Z_j) = \frac{\sigma_{(k)}^2/\tau_{(k)}^{2\eta}}{\left(\left[\left(\sigma_i^2/\tau_i^{2\eta}\right) + \left(\sigma_{(k)}^2/\tau_{(k)}^{2\eta}\right)\right]\left[\left(\sigma_j^2/\tau_j^{2\eta}\right) + \left(\sigma_{(k)}^2/\tau_{(k)}^{2\eta}\right)\right]\right)^{1/2}}, \tag{2.29}$$

which is positive. Hence, Slepian's [68] inequality implies

$$P\left\{ Z_i < \frac{\gamma}{\sqrt{2}}, \forall i \neq (k) \right\} > \prod_{i\neq(k)} P\left\{ Z_i < \frac{\gamma}{\sqrt{2}} \right\} = \prod_{i\neq(k)} \Phi(\gamma/\sqrt{2}) = 1 - \alpha \tag{2.30}$$

since $\gamma = \sqrt{2} z_{(1-\alpha)^{1/(k-1)}}$. Thus, (2.26) follows from (2.28), so (2.22) holds, thereby proving (ii).

To show (iii), note that if we replace γ with $\bar{\gamma}$, all of the arguments up to (2.30) are still valid. Hence, when $\eta = 1/2$ and $0 < \lambda < 2$, part (ii) now shows that $\tau_i = \bar{\gamma}^2\sigma_i^2$, so the covariance in (2.29) reduces to $1/2$. Consequently, we no longer need to use Slepian's inequality in (2.30) to bound $P\{Z_i < \bar{\gamma}/\sqrt{2}, \forall i \neq (k)\}$, which equals $1-\alpha$ by the definition of $\bar{\gamma}$. Also, the inequality in (2.30) holds if γ is replaced with any $x > 0$, so $\bar{\gamma} < \gamma$. Part (ii) implies $T_i(\delta)/(\bar{\gamma}\sigma_i/\delta)^2 \Rightarrow 1$ as $\delta \to 0$, and the final result follows immediately by the uniform integrability.

2.9.2 Proof of Theorem 2

We write $G_\theta(\delta)$ instead of $G(\delta)$ in (2.20) to emphasize the dependence on the parameters $\theta = (\theta_1, \ldots, \theta_k)$. Then for any $\theta \in \Omega(\delta)$ and $\delta > 0$,

$$G_\theta(\delta) \;=\; \left\{\widetilde{\theta}_{(k)}(\delta) > \widetilde{\theta}_i(\delta) + (\theta_{(k)} - \theta_i - \delta), \forall i \neq (k)\right\} \;\subseteq\; \{\widetilde{\theta}_{(k)}(\delta) > \widetilde{\theta}_i(\delta), \forall i \neq (k)\} \;=\; \mathrm{CS}_\theta(\delta)$$

since $\theta_{(k)} - \theta_i - \delta > 0$ for $(\theta_1, \ldots, \theta_k) \in \Omega(\delta)$. Also, the proof of (2.21) holds for all $\theta \in \Omega(\delta)$, so

$$G_\theta(\delta) \;\subseteq\; \mathrm{CS}_\theta(\delta) \cap \left\{\theta_i - \max_{j \neq i} \theta_j \in I_i(\delta),\; i = 1, 2, \ldots, k\right\} \text{ for all } \theta \in \Omega(\delta) \text{ and } \delta > 0. \quad (2.31)$$

Because of Assumption 3, we can write $\widetilde{\theta}_i(\delta) = \theta_i + \widetilde{Y}_i(\delta)$, where $\widetilde{Y}_i(\delta) = Y_i(T_i(\delta))$, and the distribution of $\widetilde{Y}_i(\delta)$ does not depend on θ_i. Consequently, for any $\theta \in \Omega(\delta)$,

$$G_\theta(\delta) \;=\; \left\{\widetilde{\theta}_{(k)}(\delta) - \theta_{(k)} > \widetilde{\theta}_i(\delta) - \theta_i - \delta, \forall i \neq (k)\right\} \;=\; \left\{\widetilde{Y}_{(k)}(\delta) > \widetilde{Y}_i(\delta) - \delta, \forall i \neq (k)\right\},$$

so for all $\delta > 0$, the probability of $G_\theta(\delta)$ does not depend on θ because of Assumption 3. Hence, we again write $G(\delta)$ rather than $G_\theta(\delta)$, and (2.31) implies

$$P\{G(\delta)\} \leq \inf_{\theta \in \Omega(\delta)} P\left\{\mathrm{CS}_\theta(\delta),\; \theta_i - \max_{\ell \neq i} \theta_\ell \in I_i(\delta),\; \forall i = 1, \ldots, k\right\}$$

for all $\delta > 0$. Similarly, the probability of $H(\delta)$ in (2.25) does not depend on θ since $W_i(\delta) = \widetilde{Y}_i(\delta)/\delta$ by (2.23) and Assumption 3, so

$$P\{H(\delta)\} \leq P\{G(\delta)\} \leq \inf_{\theta \in \Omega(\delta)} P\left\{\mathrm{CS}_\theta(\delta),\; \theta_i - \max_{\ell \neq i} \theta_\ell \in I_i(\delta),\; \forall i = 1, \ldots, k\right\} \quad (2.32)$$

for all $\delta > 0$ by (2.24). Replacing $\widehat{\theta}_i(T_i(\delta)) - \theta_i$ with $Y_i(T_i(\delta))$ throughout the proof of (2.26), we see that (2.26) holds regardless of θ. Thus, taking limits as $\delta \to 0$ in (2.32) completes the proof by (2.26).

2.10 Proof of Theorem 3

We start by proving part (i). Since $T_{0,i} = \zeta_i \delta^{-\lambda}$ for $\zeta_i > 0$ and $0 < \lambda \leq 1/\eta$, (2.6) implies

$$\tau_{i,r}(\delta) \equiv \delta^{1/\eta}\, T_{i,r}(\delta) = \max\left(\zeta_i \delta^{(1/\eta)-\lambda}, \left(\frac{\gamma\sqrt{V_i(T_{0,i})}}{\epsilon_i(T_0)}\right)^{1/\eta}\right). \quad (2.33)$$

Note that (2.1) implies

$$\widehat{\theta}_i(T_{0,i}) \;=\; T_{0,i}^{-\eta}\, T_{0,i}^{\eta}[\widehat{\theta}_i(T_{0,i}) - \theta_i] + \theta_i \;\Rightarrow\; 0 \cdot N(0, \sigma_i^2) + \theta_i \;=\; \theta_i \tag{2.34}$$

as $\delta \to 0$. Since each θ_i is deterministic, Theorem 3.9 of [10] yields $(\widehat{\theta}_i(T_{0,i}) : i = 1, \ldots, k) \Rightarrow (\theta_i : i = 1, \ldots, k)$. Therefore, we have that $\epsilon_i(T_0)$ defined in (2.7) satisfies $\epsilon_i(T_0) \Rightarrow \epsilon_i$ as $\delta \to 0$ by the continuous-mapping theorem, where ϵ_i is defined immediately after (2.8) and is deterministic. Assumption 2 then implies $(\epsilon_i(T_0), V_i(T_{0,i})) \Rightarrow (\epsilon_i, \sigma_i^2)$ as $\delta \to 0$ by Theorem 3.9 of [10]. Hence, it follows from the generalized continuous-mapping theorem that

$$\delta^{1/\eta}\, T_{i,r}(\delta) = \tau_{i,r}(\delta) \Rightarrow \tau_{i,r} \tag{2.35}$$

as $\delta \to 0$, where $\tau_{i,r}$ is defined in (2.8) and is deterministic, which proves the first part of (i). The uniform-integrability assumption and Theorem 3.5 of Billingsley (1999) establish the second part of (i).

Now we establish (i). Define the event $G_r(\delta) = A_1(\delta) \cap A_2(\delta)$, where

$$\begin{aligned} A_1(\delta) &= \left\{ \left(\widetilde{\theta}_{(k),r}(\delta) - \widetilde{\theta}_{i,r}(\delta) \right) - \left(\theta_{(k)} - \theta_i \right) \geq -\delta \left(\widetilde{\theta}_{(k),r}(\delta) - \widetilde{\theta}_{i,r}(\delta) \right), \forall i \neq (k) \right\}, \\ A_2(\delta) &= \left\{ \widetilde{\theta}_{(k),r}(\delta) > \widetilde{\theta}_{(k-1),r}(\delta), \forall i \neq (k) \right\}, \end{aligned}$$

and we can establish

$$G_r(\delta) \subseteq \left\{ \theta_i - \max_{j \neq i} \theta_j \in I_{i,r}(\delta),\ i = 1, 2, \ldots, k \right\} \tag{2.36}$$

by following a line of reasoning developed by [21]. We now show that

$$\lim_{\delta \to 0} P\{G_r(\delta)\} > 1 - \alpha. \tag{2.37}$$

Define

$$W_{i,r}(\delta) = \frac{1}{\delta} \left[\widetilde{\theta}_{i,r}(\delta) - \theta_i \right] = \frac{1}{\tau_{i,r}^{\eta}} (\tau_{i,r} \delta^{-1/\eta})^{\eta} \left[\widehat{\theta}_i(T_{i,r}(\delta)) - \theta_i \right] \tag{2.38}$$

and note that

$$A_1(\delta) = \left\{ \left[W_{i,r}(\delta) - W_{(k),r}(\delta) \right] - \left[\widetilde{\theta}_{(k),r}(\delta) - \widetilde{\theta}_{i,r}(\delta) \right] \leq 0, \forall i \neq (k) \right\}.$$

Since $\tau_{i,r} > 0$ in (2.35) is deterministic, Assumption 1 ensures

$$(W_{i,r}(\delta) : i = 1, \ldots, r) \Rightarrow (W_{i,r} : i = 1, \ldots, k)$$

as $\delta \to 0$, where $W_{1,r}, \ldots, W_{k,r}$ are independent and $W_{i,r} \sim N(0, \sigma_i^2/\tau_{i,r}^{2\eta})$. Also, arguing as in (2.34), we can show that $\widetilde{\theta}_i(\delta) \Rightarrow \theta_i$ as $\delta \to 0$, where θ_i is deterministic. Hence,

$$(W_{i,r}(\delta), \widetilde{\theta}_i(\delta) : i = 1, \ldots, k) \Rightarrow (W_{i,r}, \theta_i : i = 1, \ldots, k) \tag{2.39}$$

as $\delta \to 0$ by Theorem 3.9 of [10].

Now define the mapping $g = (g_1, g_2)$ with $g_1 : \Re^{2k} \to \Re^{k-1}$ and $g_2 : \Re^{2k} \to \Re$, as

$$\begin{aligned} g_1(x_i, y_i : i = 1, \ldots, k) &= ((x_i - x_{(k)}) - (y_{(k)} - y_i) : \forall\ i \neq (k)), \\ g_2(x_i, y_i : i = 1, \ldots, k) &= y_{(k)} - y_{(k-1)}. \end{aligned}$$

Observe that $A_1(\delta) = \{g_1(W_{i,r}(\delta), \widetilde{\theta}_i(\delta) : i = 1, \ldots, k) \leq 0\}$ and $A_2(\delta) = \{g_2(W_{i,r}(\delta), \widetilde{\theta}_i(\delta) : i = 1, \ldots, k) > 0\}$. Since g_1 and g_2 are continuous functions, the continuous-mapping theorem and (2.39) yield $g(W_{i,r}(\delta), \widetilde{\theta}_i(\delta) : i = 1, \ldots, k) \Rightarrow g(W_{i,r}, \theta_i : i = 1, \ldots, k)$ as $\delta \to 0$. Since $W_{i,r} : i = 1, \ldots, k$, are independent normals, $P\{g_1(W_{i,r}, \theta_i : i = 1, \ldots, k) = 0\} = 0$, so the Portmanteau theorem gives

$$\begin{aligned} P\{G_r(\delta)\} &= P\{\, g_1(W_{i,r}(\delta), \widetilde{\theta}_i(\delta) : i = 1, \ldots, k) \leq 0;\ g_2(W_{i,r}(\delta), \widetilde{\theta}_i(\delta) : i = 1, \ldots, k) > 0\,\} \\ &\to P\{\, g_1(W_{i,r}, \theta_i : i = 1, \ldots, k) \leq 0;\ g_2(W_{i,r}, \theta_i : i = 1, \ldots, k) > 0\,\} \ \equiv\ p_1 \end{aligned}$$

as $\delta \to 0$. Note that $g_2(W_{i,r}, \theta_i : i = 1, \ldots, k) = \theta_{(k)} - \theta_{(k-1)}$, so $P\{g_2(W_{i,r}, \theta_i : i = 1, \ldots, k) > 0\} = 1$ by our assumption that $\theta_{(k)} > \theta_{(k-1)}$, and consequently, $p_1 = P\{g_1(W_{i,r}, \theta_i : i = 1, \ldots, k) \leq 0\}$. Hence, to establish (2.37), we need to show that $p_1 \geq 1 - \alpha$.

To prove this, note that

$$\begin{aligned} p_1 &= P\left\{W_{i,r} - W_{(k),r} \leq \theta_{(k)} - \theta_i,\ \forall\ i \neq (k)\right\} \\ &= P\left\{ \frac{W_{i,r} - W_{(k),r}}{\left[\left(\sigma_i^2/\tau_{i,r}^{2\eta}\right) + \left(\sigma_{(k)}^2/\tau_{(k),r}^{2\eta}\right)\right]^{1/2}} \leq \frac{\theta_{(k)} - \theta_i}{\left[\left(\sigma_i^2/\tau_{i,r}^{2\eta}\right) + \left(\sigma_{(k)}^2/\tau_{(k),r}^{2\eta}\right)\right]^{1/2}},\ \forall\ i \neq (k) \right\}. \end{aligned}$$

By (2.8), each $\tau_{j,r} \geq (\gamma\sigma_j/\epsilon_j)^{1/\eta} > 0$, so $\tau_{j,r}^{2\eta} \geq (\gamma\sigma_j/\epsilon_j)^2$. Moreover, $\epsilon_j \geq \epsilon_{(k)} > 0$ for all j, so for $i \neq (k)$,

$$\begin{aligned}\left[\frac{\sigma_i^2}{\tau_{i,r}^{2\eta}} + \frac{\sigma_{(k)}^2}{\tau_{(k),r}^{2\eta}}\right]^{1/2} &\leq \left[\frac{\sigma_i^2}{(\gamma\sigma_i/\epsilon_i)^2} + \frac{\sigma_{(k)}^2}{(\gamma\sigma_{(k)}/\epsilon_{(k)})^2}\right]^{1/2} \\ &\leq \left[\frac{\sigma_i^2}{(\gamma\sigma_i/\epsilon_i)^2} + \frac{\sigma_{(k)}^2}{(\gamma\sigma_{(k)}/\epsilon_i)^2}\right]^{1/2} \\ &= \sqrt{2}\left(\frac{\theta_{(k)} - \theta_i}{\gamma}\right).\end{aligned}$$

Consequently,

$$p_1 \geq P\left\{Z_{i,r} \leq \frac{\gamma}{\sqrt{2}},\ \forall\ i \neq (k)\right\}$$

where $Z_{i,r} = [W_{i,r} - W_{(k),r}]/[(\sigma_i^2/\tau_{i,r}^{2\eta}) + (\sigma_{(k)}^2/\tau_{(k),r}^{2\eta})]^{1/2} \sim N(0,1)$. The rest of the proof follows the same line of reasoning in the proof of Theorem 1 from (2.29) onward with a subscript r added to the variables Z_i, τ_i, and T_i.

2.11 Proof of Proposition 1

For each $i = 1, \ldots, k$, note that

$$\widetilde{\tau}_{i,\delta}^{\eta}\left[\widehat{\theta}_i(\widetilde{T}_i(\delta)) - \theta_i\right] = \widetilde{\tau}_{i,\delta}^{\eta}\left[\widehat{\theta}_i(\widetilde{\tau}_{i,\delta}) - \theta_i\right] + \widetilde{\tau}_{i,\delta}^{\eta}\left[\widehat{\theta}_i(\widetilde{T}_i(\delta)) - \widehat{\theta}_i(\widetilde{\tau}_{i,\delta})\right] \equiv X_{i,\delta} + Y_{i,\delta}.$$

The fact that each $\widetilde{\tau}_{i,\delta}$ is deterministic implies $X_{i,\delta} \Rightarrow Z_i \sim N(0, \sigma_i^2)$ as $\delta \to 0$ by (2.1). It also follows that $\widehat{\theta}_i(\widetilde{\tau}_{i,\delta})$, $i = 1, \ldots, k$, are independent by the independence of $\widehat{\theta}_i$, $i = 1, \ldots, k$, since $\widetilde{\tau}_{i,\delta}$, $i = 1, \ldots, k$, are deterministic. Thus, Theorem 11.4.4 of [75] guarantees $(X_{i,\delta},\ i = 1, \ldots, k) \Rightarrow (Z_i,\ i = 1, \ldots, k)$ as $\delta \to 0$, where Z_i, $i = 1, \ldots, k$, are independent. Theorem 3.1 of [10] then ensures that it is sufficient to prove $(Y_{i,\delta},\ i = 1, \ldots, k) \Rightarrow (0, \ldots, 0)$ as $\delta \to 0$ to establish the result. But [4] shows that $Y_{i,\delta} \Rightarrow 0$ for each i, and since the limit is deterministic, they converge jointly by Theorem 3.9 of [10], completing the proof.

2.11.1 Proof of Theorem 4

First note that the FCLT in Condition 2 and C5 imply

$$g_m^2(U_{i,T_{0,i}}) = g_m^2(U_{i,\zeta_i\delta^{-\lambda}}) \Rightarrow g_m^2(\sigma_i B_i) = \sigma_i^2 g_m^2(B_i) \tag{2.40}$$

as $\delta \to 0$ by the continuous-mapping theorem and C2. Hence, $\delta^2\, T'_{i,m}(\delta) = \max\left(\zeta_i, \gamma'^2 g_m^2(U_{i,T_{0,i}})\right) \Rightarrow \tau'_{i,m}$ as $\delta \to 0$ by C5 and the continuous-mapping theorem, so (i) holds by C6.

Section 2.12 contains the proof of (ii). For (iii), note that

$$R_{i,m}(\delta) = \frac{\max\left[\zeta_i, \gamma'^2 g_m^2(U_{i,T_{0,i}})\right]}{\max\left[\zeta_i, \gamma^2 V_i(T_{0,i})\right]} \quad \text{and} \quad \bar{R}_{i,m}(\delta) = \frac{\gamma'^2 g_m^2(U_{i,T_{0,i}})}{\gamma^2 V_i(T_{0,i})} \tag{2.41}$$

for each i. Since $V_i(T_{0,i}) \Rightarrow \sigma_i^2$ as $T_{0,i} \to \infty$ by Assumption 2 and σ_i^2 is deterministic, (2.40) implies

$$(g_m^2(U_{i,T_{0,i}}), V_i(T_{0,i})) \Rightarrow (\sigma_i^2 g_m^2(B_i), \sigma_i^2) \tag{2.42}$$

as $\delta \to 0$ by Theorem 3.9 of [10], so the continuous-mapping theorem yields (iii).

Note that $0 \le R_{i,m}(\delta) \le \max\left[\zeta_i, \gamma'^2 V'_{i,m}(\zeta_i/\delta^2)\right]/\zeta_i$, so the uniform integrability of $\{V'_{i,m}(t) : t > 0\}$ implies $\{R_{i,m}(\delta) : \delta > 0\}$ also is; hence, (iii) ensures $E[R_{i,m}(\delta)] \to E[R_{i,m}]$ as $\delta \to 0$ by Theorem 3.5 of [10]. Applying Jensen's inequality results in

$$E[R_{i,m}] = E\left[\frac{\max\left[\zeta_i, \gamma'^2 \sigma_i^2 g_m^2(B_i)\right]}{\max\left[\zeta_i, \gamma^2\sigma_i^2\right]}\right] > \frac{\max\left[\zeta_i, \gamma'^2 \sigma_i^2 E[g_m^2(B_i)]\right]}{\max\left[\zeta_i, \gamma^2\sigma_i^2\right]} = \frac{\max\left[\zeta_i, \gamma'^2\sigma_i^2\right]}{\max\left[\zeta_i, \gamma^2\sigma_i^2\right]},$$

where the strict inequality holds because $P\{\gamma'^2\sigma_i^2 g_m^2(B_i) < \zeta_i\} > 0$ and $P\{\gamma'^2\sigma_i^2 g_m^2(B_i) > \zeta_i\} > 0$ by C6, and the second equality holds because we assumed $E[g_m^2(B_i)] = 1$. Thus, the first part of (iv) follows from (ii). The second part of (iv) similarly holds.

2.12 Proof of Theorem 4(ii)

First recall our definition of F_m in (2.55), and it suffices to show that $F_m(a) < [\Phi(a/\sqrt{2})]^{k-1}$ for all $a > 0$. We first show that

$$F_m(a) \le \left[\Phi\left(\frac{a}{\sqrt{2}}\right)\right]^{k-1} \tag{2.43}$$

for all $a > 0$.

Let $G(x) = P\{g_m^2(B) \le x\}$, where B is a standard Brownian motion. Note that

$$F_m(a) = \int_0^\infty \left[\int_0^\infty \Phi(aC(x,y))\, dG(x)\right]^{k-1} dG(y), \tag{2.44}$$

where $C(x,y) = \left[\frac{1}{x} + \frac{1}{y}\right]^{-1/2}$. We first show that for each $a > 0$ and $y > 0$, the inner integral in (2.44) satisfies

$$\int_0^\infty \Phi(aC(x,y))\, dG(x) \le \Phi(aC(1,y)). \tag{2.45}$$

Let K be any distribution function with $K(0) = 0$ and $K(\infty) = 1$, and let $\overline{K}(x) = 1 - K(x)$. For each $a > 0$ and $y > 0$, note that

$$\begin{aligned}
\int_0^\infty \Phi(aC(x,y))\, dK(x) &= -\int_0^\infty \Phi(aC(x,y))\, d\overline{K}(x) \\
&= -\left[\Phi(aC(x,y))\, \overline{K}(x)\right]_{x=0}^{\infty} + \int_0^\infty \overline{K}(x)\, \phi(aC(x,y))\, a\, c(x,y)\, dx \\
&= \frac{1}{2} + \int_0^\infty \overline{K}(x)\, \phi(aC(x,y))\, a\, c(x,y)\, dx,
\end{aligned}$$

where the penultimate step follows by applying an integration by parts, ϕ denotes the density function of a standard normal, and $c(x,y) = \frac{\partial}{\partial x} C(x,y)$. To simplify notation, let $\psi(x) = \phi(aC(x,y))\, a\, c(x,y)$, so

$$\int_0^\infty \Phi(aC(x,y))\, dK(x) = \frac{1}{2} + \int_0^\infty \overline{K}(x)\, \psi(x)\, dx. \tag{2.46}$$

Now define the distribution function H as

$$H(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x \ge 1 \end{cases},$$

and we can write the right-hand side of (2.45) as

$$\Phi(aC(1,y)) = \int_0^\infty \Phi(aC(x,y))\, dH(x) = \frac{1}{2} + \int_0^\infty \overline{H}(x)\, \psi(x)\, dx$$

by (2.46). Similarly, we can use (2.46) to express the left-hand side of (2.45) as

$$\int_0^\infty \Phi(aC(x,y))\, dG(x) = \frac{1}{2} + \int_0^\infty \overline{G}(x)\, \psi(x)\, dx.$$

So to establish (2.45), it suffices to show $\int_0^\infty \overline{G}(x)\,\psi(x)\,dx \le \int_0^\infty \overline{H}(x)\,\psi(x)\,dx$, which is equivalent to proving

$$\int_0^\infty \left[\overline{H}(x) - \overline{G}(x)\right]\,\psi(x)\,dx \ge 0. \tag{2.47}$$

We now show that $\psi(x)$ is strictly decreasing in $x > 0$ by showing its derivative $\psi'(x)$ is negative for all $x > 0$. Observe that

$$\psi'(x) = \phi'(aC(x,y))\,\left(a\,c(x,y)\right)^2 + \phi(aC(x,y))\,a\,c'(x,y), \tag{2.48}$$

where $\phi'(x) = \frac{d}{dx}\phi(x)$ and $c'(x,y) = \frac{\partial}{\partial x}c(x,y)$. Note that

$$\begin{aligned} c(x,y) &= \frac{1}{2x^2}\left[\frac{1}{x}+\frac{1}{y}\right]^{-3/2}, \\ c'(x,y) &= \frac{-1}{x^3}\left[\frac{1}{x}+\frac{1}{y}\right]^{-3/2} + \frac{3}{4x^4}\left[\frac{1}{x}+\frac{1}{y}\right]^{-5/2}. \end{aligned}$$

Since $\phi(z)$ is strictly decreasing in $z > 0$, $\phi'(z) < 0$ for $z > 0$, so $\phi'(aC(x,y)) < 0$ because $C(x,y) > 0$ for $x > 0$ and $y > 0$. Also, $c(x,y) > 0$ for all $x > 0$ and $y > 0$, so the first term in (2.48) is negative. We now show the second term in (2.48) is also negative. First note that $\phi(z) > 0$ for all z, and we now show $c'(x,y) < 0$ for all $x > 0$ and $y > 0$. This holds if and only if $\frac{3}{4x}\left[\frac{1}{x}+\frac{1}{y}\right]^{-1} < 1$, which clearly is true since $\frac{3}{4x} < \frac{1}{x} < \frac{1}{x}+\frac{1}{y}$ for $x > 0$ and $y > 0$. Therefore, (2.48) is negative, so $\psi(x)$ in (2.47) is strictly decreasing in $x > 0$.

Now note that

$$\overline{H}(x) - \overline{G}(x) \ge 0 \qquad \text{for } x < 1, \tag{2.49}$$

$$\overline{H}(x) - \overline{G}(x) \le 0 \qquad \text{for } x \ge 1. \tag{2.50}$$

Hence,

$$\int_0^1 \left[\overline{H}(x) - \overline{G}(x)\right] dx\,\psi(1) \le \int_0^1 \left[\overline{H}(x) - \overline{G}(x)\right]\psi(x)\,dx \tag{2.51}$$

and

$$\int_1^\infty \left[\overline{H}(x) - \overline{G}(x)\right] dx\,\psi(1) \le \int_1^\infty \left[\overline{H}(x) - \overline{G}(x)\right]\psi(x)\,dx. \tag{2.52}$$

Adding (2.51) and (2.52) then yields

$$\int_0^\infty \left[\overline{H}(x) - \overline{G}(x)\right] dx\, \psi(1) \leq \int_0^\infty \left[\overline{H}(x) - \overline{G}(x)\right] \psi(x)\, dx. \tag{2.53}$$

But the left-hand side of (2.53) is zero because $\int_0^\infty \overline{G}(x)\, dx = E[g^2(B)] = 1$ by Assumption 4 and $\int_0^\infty \overline{H}(x)\, dx = 1$, which establishes (2.47) and therefore (2.45).

Because of (2.44) and (2.45), if we now want to show (2.43), it suffices to prove that

$$\int_0^\infty \left[\Phi(aC(1,y))\right]^{k-1} dG(y) \leq \left[\Phi\left(\frac{a}{\sqrt{2}}\right)\right]^{k-1}. \tag{2.54}$$

Because $\Phi(a/\sqrt{2}) = \Phi(aC(1,1))$, this can be established using arguments similar to those applied to show (2.45), which proves (2.43). Because of our arguments in the proof, the only way equality can hold in (2.43) is if $P\{g_m^2(B) = 1\} = 1$, but [29] show this is impossible for any fixed $m \geq 1$. (Alternatively, C6 implies the inequalities in (2.49) and (2.50) are strict, which implies (2.51)–(2.54) also have strict inequalities.) Thus, the proof is complete.

2.12.1 Proof of Theorem 5

For $i = 1, \ldots, k$, note that $B_i(1)$ has a standard normal distribution. Define the vector $(B_1(1), \ldots, B_{k-1}(1), g_m^2(B_1), \ldots, g_m^2(B_k))$, and define the function $h : \Re^{2k-1} \to \Re$ as

$$h(z_1, \ldots, z_{k-1}, v_1, \ldots, v_k) = \max_{i<k} \left[z_i \left(\frac{1}{v_i} + \frac{1}{v_k} \right)^{1/2} \right].$$

Then evaluating the left-hand side of (2.15) at any point a rather than γ', we have

$$\begin{aligned} F_m(a) &\equiv E\left[\prod_{i=1}^{k-1} \Phi\left(\frac{a}{\left[(1/g_m^2(B_i)) + (1/g_m^2(B_k))\right]^{1/2}}\right)\right] \\ &= P\left\{h(B_1(1), \ldots, B_{k-1}(1), g_m^2(B_1), \ldots, g_m^2(B_k)) \leq a\right\} \end{aligned} \tag{2.55}$$

since $B_i(1)$ and $g(B_i)$ are independent ([29]) and $B_1, \ldots, B_k$ are independent.

If $g_m = g_{\mathrm{bm},m}$, then $g_m^2(B_i) \sim \frac{\chi^2(m-1)}{(m-1)} \Rightarrow 1$ as $m \to \infty$, where $\chi^2(t)$ denotes a χ^2 random variable with t degrees of freedom. Now suppose $g_m = \bar{g}_m$ in (2.12). For each i, the processes $[\Lambda_{j,m} B_i(t) : 0 \leq t \leq 1]$, $j = 1, \ldots, m$, are independent

standard Brownian motions on the unit interval, so it follows from C1 that $g \circ \Lambda_{j,m}(B_i)$, $j = 1, \ldots, m$, are i.i.d. Since $E[g^p(B)] < \infty$, the weak law of large numbers implies $(1/m) \sum_{j=1}^{m} g^p \circ \Lambda_{j,m}(B_i) \Rightarrow E[g^p(B)]$ as $m \to \infty$, so (2.12) yields that again

$$g_m^2(B_i) \Rightarrow 1 \tag{2.56}$$

as $m \to \infty$ by the continuous-mapping theorem.

Equation (2.56) also holds when g_m is the overlapping area estimator in Section 2.7.1 or the overlapping CvM estimator in Section 2.7.2, as we now explain. The reference [2] shows that in both cases, $g_m(B)$ converges to 1 in quadratic mean as $m \to \infty$; see Theorems 5 and 8 of [2]. Since convergence in quadratic mean implies convergence in distribution (e.g., see pp. 140–141 of [43]), we see that (2.56) then holds for these two estimators; It also can be shown that both overlapping estimators satisfy properties C1–C6.

Since $B_1(1), \ldots, B_{k-1}(1), g_m^2(B_1), \ldots, g_m^2(B_k)$, are mutually independent, (2.56) and Theorem 3.9 of [10] guarantee

$$(B_1(1), \ldots, B_{k-1}(1), g_m^2(B_1), \ldots, g_m^2(B_k)) \Rightarrow (B_1(1), \ldots, B_{k-1}(1), 1, \ldots, 1) \tag{2.57}$$

as $m \to \infty$. Each $B_i(1) \sim N(0,1)$, so $P\{h(B_1(1), \ldots, B_{k-1}(1), 1, \ldots, 1) = a\} = P\{\max_{i<k} \left[B_i(1)\sqrt{2}\right] = a\} = 0$. Thus, the a.s. continuity of h at the limit in (2.57) implies that as $m \to \infty$,

$$\begin{aligned} & P\left\{h(B_1(1), \ldots, B_{k-1}(1), g_m^2(B_1), \ldots, g_m^2(B_k)) \le a\right\} \\ & \quad \to \quad P\left\{h(B_1(1), \ldots, B_{k-1}(1), 1, \ldots, 1) \le a\right\} = \left[\Phi(a/\sqrt{2})\right]^{k-1} \equiv F(a), \end{aligned}$$

by the Portmanteau theorem. Since a was arbitrary, we have now proven $F_m \Rightarrow F$ as $m \to \infty$. The proof of Lemma 1.5.6 of [65] shows that there is a one-to-one correspondence between the elements of the set $\{t : 0 < t < 1, F_m^{-1}(t) \not\to F^{-1}(t) \text{ as } m \to \infty\}$ and the flat portions of F. However, since F is strictly increasing, $F_m^{-1}(t) \to F^{-1}(t)$ for all $0 < t < 1$, so we get (i).

Parts (ii) and (iii) hold by applying part (i), (2.56) and the continuous-mapping theorem to Theorem 4(i) and (iii). Part (iv) then follows from the uniform integrability.

CHAPTER 3

ESTIMATING THE TIME AVERAGE VARIANCE CONSTANT

3.1 Introduction

Let a real sequence $\{X_n : n \geq 1\}$ represent the output of a steady-state simulation and assume that the sequence is ergodic, i.e., we assume the existence of a finite constant μ such that $(1/n)\sum_{i=1}^{n} X_i \xrightarrow{d} \mu$, as $n \to \infty$. Here "$\xrightarrow{d}$" means convergence in distribution of random variables (see [9]). The sample mean $\overline{X}(n) \equiv (1/n)\sum_{i=1}^{n} X_i$ is the most common point estimator for the parameter μ. People are more interested in constructing an interval estimator for μ. Several interval estimation techniques are available in the literature. These include the methods of replication ([50]), batched means ([62]), autoregression ([25]), spectrum estimation ([35]), regeneration ([17]) and ARMA modeling ([63]). Several comparative studies of simulation confidence interval estimators have been reported ([22]), [50], [61]). In order to give a measure of the precision of the estimate, the *variance parameter*, $\sigma^2 \equiv \lim_{n\to\infty} nVar(\overline{X}(n))$ is also estimated. There are several different techniques in the literature for the estimation of σ^2. For example the methods of nonoverlapping batch means(NBM) ([62]), overlapping batch means (OBM) ([53]) and other standardized time series (STS) ([64]). The consensus of these studies is that the various interval estimation techniques work well only in situations in which their supporting assumptions are valid. Unfortunately, it is not always easy to verify that a particular simulated process has the properties assumed. We try to develop a general approach to interval estimation that has very mild theoretical requirements for validity, but not less than the method of batch means. This approach is an STS approach, though the focus is mainly on the estimation of the variance parameter. One may use the variance parameter to construct an interval estimator for the mean of the stochastic process under consideration.

3.2 Background and theory

Let $\overline{X}_j = \sum_{k=1}^{j} X_k/j,\ j = 1, 2, \ldots, n.$

Assumption 5 *Suppose that the series* $\sigma^2 = Var(X_1) + 2\sum_{k=1}^{\infty} Cov(X_1, X_{1+k})$ *converges absolutely and* $\sigma^2 > 0$. *For each positive integer* n, *let*

$$Y_n\left(\frac{i}{n}\right) \equiv \frac{i(\overline{X}_i - \mu)}{\sigma\sqrt{n}} \quad for \quad 0 \le i \le n. \tag{3.1}$$

We define $Y_n(t)$ *by linear interpolation elsewhere on* $[0, 1]$ *so that it is a continuous function. Then* $Y_n \xrightarrow{d} B(\cdot)$ *as* $n \to \infty$, *where* $B(\cdot)$ *is a standard Brownian motion process on* $[0, 1]$.

Assumption 5 is satisfied by almost all the stochastic processes relevant to the steady-state simulations. For example, consider the mixed autoregressive-moving average time series (stationary and invertible), finite-state discrete-time Markov chains (irreducible and aperiodic), finite state continuous-time Markov processes, waiting times in a stable $M/M/1$ queueing system etc. This assumption has also been used in [2]. The class of processes that satisfy this assumption (also known as *Assumption FCLT*) is given in [28]. The related references are also given in [28]. Assumption 5 is necessary, since it has been directly used in the proof of Theorem 9 and hence also is necessary for Theorem 10, Theorem 11 and Theorem 12.

Next we discuss a few variance estimators. Without loss of generality, we may assume that $E[X_1] = 0$. We also assume that $n = mb$, where m is the batch size and b is sometimes interpreted as the number of nonoverlapping batches.

3.2.1 Nonoverlapping Batch Means

Nonoverlapping Batch Means (NBM) is a very popular estimator having practical use. It is defined as

$$N(b, m) = \frac{m}{b-1}\sum_{i=1}^{b}(\overline{X}_{i,m} - \overline{X}_n)^2 = \frac{m}{b-1}\left(\sum_{i=1}^{b}\overline{X}_{i,m}^2 - b\overline{X}_n^2\right). \tag{3.2}$$

Here, $\overline{X}_{i,m} = \sum_{j=1}^{m} X_{(i-1)m+j}/m$ is the ith nonoverlapping batch mean, $i = 1, 2, \ldots, b$, i.e., the sample mean of the observations $X_{(i-1)m+1}, X_{(i-1)m+2}, \ldots, X_{im}$, which form the ith nonoverlapping batch. For details, see [16], [33] and [70].

3.2.2 Overlapping Batch Means

This estimator is defined as

$$O(b, m) = \frac{nm}{(n-m+1)(n-m)} \sum_{i=1}^{n-m+1} (\overline{X}^{o}_{i,m} - \overline{X}_n)^2. \tag{3.3}$$

Here, $\overline{X}^{o}_{i,m} = \sum_{j=0}^{m-1} X_{i+j}/m$ is the ith overlapping batch mean, $i = 1, 2, \ldots, n-m+1$, i.e., the sample mean of the observations $X_i, X_{i+1}, \ldots, X_{i+m-1}$, which form the ith overlapping batch. For details, see [53].

3.2.3 Standardized Time Series

The Standardized Time Series (STS) of $\mathbf{X} = X_1, X_2, \ldots$ is defined as $T_n(t) = \frac{\lfloor nt \rfloor (\overline{X}_n - \overline{X}_{\lfloor nt \rfloor})}{\sigma\sqrt{n}}$ for $t \in [0, 1]$. Under assumption 5, [29] and [64] show that $T_n \xrightarrow{d} B$, a Brownian bridge on $[0, 1]$, which is a Gaussian process with $E[B(t)] = 0$ and $Cov(B(s), B(s)) = \min(s, t) - st$.

3.3 Some results from the empirical process literature

3.3.1 The p-fold integrated empirical process

Let $X_1, X_2, \ldots, X_n$ be independent observations on a random variable X with continuous distribution function F and write

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i < x}$$

for the usual empirical distribution function of the sample $(X_1, X_2, \ldots, X_n)$. The random variables $U_i = F(X_i), i \geq 1$, are independent and uniformly distributed on $(0, 1)$. Let

$$G_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{U_i < t}$$

be the corresponding empirical distribution function; we have of course a.s. $F_n = G_n \circ F$. See [48] for details. For testing the null hypothesis $H_0 : F = F_0$, where F_0 is an apriori specified continuous distribution function, many celebrated statistics are based on comparing F_n and F_0; see for instance [58, 66]. In [37, 36], the authors propose new goodness-of-fit tests which are based on the integrated empirical process. Actually the underlying limiting process is the integrated Brownian bridge which has been studied in [48].

Let $p \geq 1$ be a fixed integer. Let us introduce the p-fold integrated empirical process as follows:

$$\begin{aligned}
F_{p,n}(x) &\equiv \int_{-\infty}^{x} \frac{[F_0(x) - F_0(y)]^{p-1}}{(p-1)!} F_n(y) dF_0(y). \\
F_{p,0}(x) &\equiv \int_{-\infty}^{x} \frac{[F_0(x) - F_0(y)]^{p-1}}{(p-1)!} F_0(y) dF_0(y) \\
&= \int_{-\infty}^{x} \frac{[F_0(x) - F_0(y)]^{p}}{p!} dF_0(y) \\
&= \frac{F_0(x)^{p+1}}{(p+1)!}. \\
\mathcal{F}_{p,n}(x) &\equiv \sqrt{n}[F_{p,n}(x) - F_{p,0}(x)].
\end{aligned}$$

We also introduce the similar quantities related to the uniform distribution. For $t \in [0, 1]$,

$$\begin{aligned}
G_{p,n}(t) &\equiv \int_0^t \frac{(t-s)^{p-1}}{(p-1)!} G_n(s) ds = \frac{1}{n} \sum_{i=1}^{n} \frac{[(t-U_i)^+]^p}{p!} \\
G_{p,0}(t) &\equiv \int_0^t \frac{(t-s)^{p-1}}{(p-1)!} s ds = \int_0^t \frac{(t-s)^p}{p!} ds = \frac{t^{p+1}}{(p+1)!} \\
\mathcal{G}_{p,n}(t) &\equiv \sqrt{n}[G_{p,n}(t) - G_{p,0}(t)].
\end{aligned}$$

The foregoing notations can be extended to the non-integrated case $p = 0$ by setting

$$F_{0,n}(x) \equiv F_n(x), \qquad F_{0,0}(x) \equiv F_0(x), \qquad \mathcal{F}_{0,n}(x) \equiv \sqrt{n}[F_n(x) - F_0(x)]$$

$$G_{0,n}(t) \equiv G_n(t), \qquad G_{0,0}(t) \equiv t, \qquad \mathcal{G}_{0,n}(t) \equiv \sqrt{n}[G_n(t) - t]$$

3.3.2 Relationships between integrated statistics and integrated Brownian bridge

Let $\beta \equiv (\beta(t))_{0\leq t\leq 1}$ be the standard Brownian bridge. We introduce the underlying process $\beta_p = (\beta_p(t))_{0\leq t\leq 1}$ associated with the statistics which occur in our work:

$$\beta_p(t) \equiv \int_0^t \frac{(t-s)^{p-1}}{(p-1)!}\beta(s)ds = \int_0^t \frac{(t-s)^p}{p!}d\beta(s)$$

Actually β_p is nothing but the p-fold primitive of β. It is a well-known fact that the process $\mathcal{G}_{0,n}$ converges weakly towards the Brownian bridge β. See [48] for details.

Theorem 6 *The process $\mathcal{G}_{p,n}$ converges weakly towards β_p as $n \to +\infty, p \neq 0$.*

3.3.3 Covariance function

We shall need the following result in calculating the theoretical covariance matrix of the integrated paths. We derive the following expression for the covariance function of the process β_p by using the rule:

$$\mathbb{E}\left[\int_0^s f(u)d\beta(u)\int_0^t g(u)d\beta(u)\right] = \int_0^{s\wedge t} f(u)g(u)du - \int_0^s f(u)du\int_0^t g(u)du$$

Theorem 7 *The process β_p is a centered Gaussian process with the covariance function:*

$$G_{\beta_p}(s,t) \equiv \mathbb{E}[\beta_p(s)\beta_p(t)] = \int_0^{s\wedge t} \frac{(s-u)^p}{p!}\frac{(t-u)^p}{p!}du - \frac{(st)^{p+1}}{(p+1)!^2}.$$

3.4 The context

Recall that if a simulation produces output $X = \{X_1, X_2, \ldots, X_n\}$, then the time average variance constant is given by

$$\sigma^2 \equiv \lim_{n\to\infty} n \text{ Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right).$$

in cases where this limit exists. A quadratic function (see [69]) may be applied to the output vector X in order to construct an estimator for σ^2; this approach is given in [69]. The computational complexity of this estimator is proportional to n^2 where n is the simulation run length. Batch means estimator is a typical example of this approach.

However, due to the special form of the batch means estimator, the computational complexity is linear in n. See [13] for details.

If $\widehat{\sigma}^2$ is a candidate estimator of σ^2, then we may evaluate the performance of $\widehat{\sigma}^2$ by determining its bias, $\text{Bias}[\widehat{\sigma}^2] \equiv E[\widehat{\sigma}^2] - \sigma^2$; its variance, $\text{Var}[\widehat{\sigma}^2] \equiv E[(\widehat{\sigma}^2 - E[\widehat{\sigma}^2])^2]$; and finally its mean-squared error, $\text{MSE}[\widehat{\sigma}^2] \equiv E[(\widehat{\sigma}^2 - \sigma^2)^2] = \text{Bias}^2[\widehat{\sigma}^2] + \text{Var}[\widehat{\sigma}^2]$. Usually, the lower the bias and variance, the better.

Efficiency of an estimator is defined as the reciprocal of the product of work and mean-squared error. See [13] for details. Suppose an estimator has bias B and variance V. If the estimator takes time T to compute, the efficiency of the estimator η is given by $\eta = \frac{1}{T(B^2+V)}$. In [13], the author has shown that the efficiency of the integrated path estimators is better compared to batch means.

In our approach, we take into account the covariance structure of the underlying limiting stochastic process and suggest an estimator for the TAVC. As far as we know, this has not been done earlier. Usually it is found that the estimators of the TAVC are either low in bias and high in variance or vice versa. The interesting part of our experimental result is that at least for the $AR(1)$ process (introduced later), our estimator is both low in bias and variance in comparison with the standard batch means estimator. We have given the main result in Theorem 10.

3.5 Some relevant results from the existing literature

The results mentioned in this subsection are given in [14]. Still we mention those here for the sake of completeness.

Set $\widetilde{W}_0^j = 0$ for $0 \leq j \leq k$ and for $i > 0$, set $\widetilde{W}_i^0 = \sum_{l=1}^{i} X_l$, and for $1 \leq j \leq k$, set $\widetilde{W}_i^j = j \sum_{l=1}^{i} \widetilde{W}_l^{j-1}$.

Define $\widetilde{\mathbf{W}}_n = (\widetilde{W}_n^0, \widetilde{W}_n^1, \ldots, \widetilde{W}_n^k)$, where k is the integration count parameter.

Define $W_n^0(j/n) = n^{-1/2}\sum_{i=1}^{j} X_i$, $0 \le j \le n$, and, for $0 \le t \le 1$, define $W_n^0(t)$ by linear interpolation of $W_n^0(j/n), 0 \le j \le n$. For $r \ge 1$ and $0 \le j \le n$, set

$$W_n^r(j/n) = n^{-1/2}\sum_{i=1}^{j} X_i \left(\frac{j-i+1}{n}\right)^r, 0 \le j \le n.$$

Define continuous functions: $\widetilde{W}_n^r(j/n) = \widetilde{W}_j^r$ for $0 \le j \le n$ and by linear interpolation elsewhere. Set $\mathbf{W}_n(t) = (W_n^0(t), \ldots, W_n^k(t))$ and $\widetilde{\mathbf{W}}_n(t) = (\widetilde{W}_n^0(t), \ldots, \widetilde{W}_n^k(t))$ for $0 \le t \le 1$.

Define the $(k+1) \times (k+1)$ matrix $\mathbf{A}$ by:

$$A(r,q) \equiv \begin{cases} (-1)^{r-q}\begin{bmatrix} r \\ q \end{bmatrix}, & q \le r \\ 0 & q > r \end{cases}$$

for $0 \le r, q \le k$, where $\begin{bmatrix} r \\ q \end{bmatrix}$ are the Stirling numbers of the first kind [47], satisfying the recurrence: $\begin{bmatrix} n+1 \\ m \end{bmatrix} = \begin{bmatrix} n \\ m-1 \end{bmatrix} + n\begin{bmatrix} n \\ m \end{bmatrix}$, $1 \le m \le n$, $\begin{bmatrix} n \\ n \end{bmatrix} = 1$, $\begin{bmatrix} 0 \\ 0 \end{bmatrix} = 1$, and $\begin{bmatrix} n \\ 0 \end{bmatrix} = 0$ for $n \ge 1$. The $(k+1) \times (k+1)$ diagonal matrix $\mathbf{N}_n$ is defined by:

$$N_n(r,q) \equiv \begin{cases} n^{q+1/2}, & q = r \\ 0 & q \ne r \end{cases}$$

The simulation produces the discrete iterated sums, $\widetilde{W}_n$, which may be transformed to iterated integrals using the following theorem.

Theorem 8 $\mathbf{W}_n(1) = (\mathbf{A}\mathbf{N}_n)^{-1}\widetilde{\mathbf{W}}_n$.

In order to center the output, define the map $\psi : C[0,1] \to C[0,1]$ by $\psi(f)(t) \equiv f(t) - tf(1)$, $f \in C[0,1]$, $0 \le t \le 1$, and set $\overline{W}_n^0 = \psi(W_n^0)$; i.e.,
$\overline{W}_n^0(s) = W_n^0(s) - sW_n^0(1)$, $0 \le s \le 1$.

For $r \ge 1$ and $0 \le t \le 1$, set $\overline{W}_n^r(t) \equiv r\int_{s=0}^{t} \overline{W}_n^0(s)(t-s)^{r-1}ds$.
Set $\overline{\mathbf{W}}_n(t) \equiv (\overline{W}_n^0(t), \ldots, \overline{W}_n^k(t))$, $0 \le t \le 1$.

There is a stochastic process to which $\overline{W}_n$ converges. If B denotes a standard Brownian motion, then ψB is a Brownian bridge. For $r \ge 1$, let $T^r : C[0,1] \to C[0,1]$ denote the operator of r-fold integration. For a continuous function f,

$T^r f(t) \equiv \int_{s_1=0}^{t} \dots \int_{s_r=0}^{s_{r-1}} f(s_r) ds_r \dots ds_1 = \int_{s=0}^{t} f(s) \frac{(t-s)^{r-1}}{(r-1)!} ds$. The last part follows by changing the order of integration. The next theorem provides the basis for the simulation output analysis method:

Theorem 9 *As* $n \to \infty$,
$\left(\overline{W}_n^0, \overline{W}_n^1, \dots, \overline{W}_n^k\right) \xrightarrow{d} \sigma(\psi B, T\psi B, 2T^2\psi B, \dots, k!T^k\psi B)$.

3.6 Our approach

3.6.1 Construction of the integrated paths of the simulation output process

Let the i-fold integrated Brownian bridge be denoted by $W_i(\cdot)$ for $i = 0, 1, 2, \dots$. We can calculate the theoretical covariance between $W_i(s)$ and $W_j(t)$ for all values of $i,\ j,\ s$ and t. Now we shall show how we can construct the integrated paths from the simulation output.

Let the simulation output be $X_1, X_2, \dots, X_n$. Consider the *integrated paths* (IP) $W_0, W_1, \dots, W_k$. The construction of the j-th path W_j is shown below: For $2 \le i \le n$, define the following:

$$\begin{aligned} W_0(1) &= 0 \\ W_0(i) &= \frac{1}{\sqrt{n}}\left(\sum_{k=0}^{i} X(k) - \frac{i}{n}\sum_{k=0}^{n} X(k)\right) \end{aligned}$$

For $j = 1, 2, \dots, k$, define the following:

$$\begin{aligned} W_j(1) &= 0 \\ W_j(i) &= \frac{1}{n}(W_j(i-1) + W_{j-1}(i)) \end{aligned}$$

3.6.2 Construction of the theoretical covariance matrix

Suppose we want to construct the covariance matrix for $W_a, W_{a+1}, W_{a+2}, \ldots, W_b$. Using Theorem 7, we can calculate the covariance between $W_b(s)$ and $W_b(t)$ for both the cases, viz., $s \leq t$ and $s > t$. Denote these two expressions by F_1 and F_2 respectively.

Suppose we want to calculate the covariance between $W_i(s)$ and $W_j(t)$ for $i, j \leq b$. Let $\hat{i} = b - i$ and $\hat{j} = b - j$. The expression for covariance is simply given by $\frac{\partial^{\hat{j}+\hat{i}} F}{\partial t^{\hat{j}} \partial s^{\hat{i}}}$. Note that $F = F_1$ or $F = F_2$ depending on whether $s \leq t$ or $s > t$.

3.6.3 An example

Consider the covariance between $W_1(s)$ and $W_2(t)$:

1. If $s \leq t$, then $C_{st} = \frac{s^2}{24}(s^2 - 4st + 6t^2 - 2t^3)$.
2. If $s > t$, then $C_{st} = \frac{t^3}{24}(-t + 4s - 2s^2)$.
3. If we put $t = s$, we find that both of the above formulas yield the same expression $C_{st} = \frac{s^4}{24}(3 - 2s)$, which shows the continuity of the covariance function.

Similarly, consider the covariance between $W_2(s)$ and $W_1(t)$:

1. If $s \leq t$, then $C_{st} = \frac{s^3}{24}(-s + 4t - 2t^2)$.
2. If $s > t$, then $C_{st} = \frac{t^2}{24}(t^2 - 4st + 6s^2 - 2s^3)$.
3. If we put $t = s$, we find that both of the above formulas yield the same expression $C_{st} = \frac{s^4}{24}(3 - 2s)$, which shows the continuity of the covariance function.

3.7 The estimator of the TAVC

Let C be the non-singular covariance matrix of a p-dimensional normal random variable U, i.e., $U \sim N_p(0, C)$. Let the rank-factorization of C be $C = FF^T$, where F is a $p \times m$ matrix,

Theorem 10 *If* $U \sim kN_p(0, C) = N_p(0, k^2C)$ *then* $Y = F^{-1}U \sim N_p(0, k^2)$. *The estimator of* k^2 *is given by* $\frac{Y^TY}{p} = \frac{U^TC^{-1}U}{p}$.

3.7.1 The proof of a more general result

Let U, Y be random variables and $U \sim N_p(0, k^2C)$. C is a $p \times p$ matrix of rank m. Let the rank-factorization of C be $C = FF^T$, where F is a $p \times m$ matrix, so F^T is a $m \times p$ matrix. Let F^- be a left inverse of F, which is a $m \times p$ matrix. Note that $F^-F = I_m$ and $F^T(F^-)^T = I_m$.

Define $Y = F^-U$.

$Y = kF^-U \sim N_m(0, (kF^-\ C\ (kF^-)^T)$

Or, $Y \sim N_m(0, kF^-FF^T(F^-)^Tk)$

Or, $Y \sim N_m(0, k^2)$

Hence the estimator of k is given by $\frac{Y^TY}{m}$. If C is non-singular then $m = p$ and the above expression may be simplified to

$$\frac{U^TC^{-1}U}{p} = \frac{U^TC^{-1}U}{rank(C)}$$

3.7.2 Relationship between Theorem 10 and our estimator

We may note that the above theorem may be used to construct an estimator for the TAVC using the integrated paths. If we construct several integrated paths $W_a, W_{a+1}, \ldots, W_{b-1}, W_b$ and concatenate them to form the vector U (see Theorem 10), it is clear that the mean of the vector is 0 and the covariance matrix may be computed. The constant k of Theorem 10 may be identified as the TAVC, the scaling constant for the process U.

Theorem 11 *Suppose that $X \sim N_p(\mu, \psi)$, where ψ is positive definite. Let $\Delta^2(X) = (X - \mu)'\psi^{-1}(X - \mu)$. Then $\Delta^2(X)$ follows a chi-square distribution with p degrees of freedom.*

See Theorem 3.3.2 of [26] for the proof of the above theorem.

Specializing Theorem 11 to our case, we observe that we have $\mu = 0$ and $\psi = \sigma^2C$, where σ^2 is the TAVC. So $U'(\sigma^2C)^{-1}U \sim \chi^2_k$, where k is the dimension of the covariance matrix C.

Theorem 12 *$E(U'(\sigma^2 C)^{-1}U) = k$ and $var(U'(\sigma^2 C)^{-1}U) = 2k$.*

The proof of the theorem follows from the properties of the χ^2 distribution. It may be noted that from the above theorem also, we have $\sigma^2 = \frac{E(U'C^{-1}U)}{k}$. In fact, this is the key step in the derivation of our estimator.

3.7.3 The algorithm

Let a and b be the indices of the already computed IPs, n be the number of data points generated, g be the size of an equispaced grid of points. Assume that n is a multiple of g, i.e., $n = mg$ where m is an integer. Let the g grid points be denoted by $t_1, \ldots, t_g$, where $t_k = (k-1) * m, k = 1, \ldots, g$. We sample the IPs at the grid points and construct the following vectors $\overline{W}_i$ from the IP W_i for $i = a, a+1, \ldots, b$. Note that

$$\begin{array}{rcccc} \overline{W}_a & = & W_a(t_1), & \ldots, & W_a(t_g) \\ \overline{W}_{a+1} & = & W_{a+1}(t_1), & \ldots, & W_{a+1}(t_g) \\ \vdots & & & & \\ \overline{W}_b & = & W_b(t_1), & \ldots, & W_b(t_g) \end{array}$$

We can calculate the bias and the variance of the proposed estimator for a process for which the true TAVC is known. Let B_L be the bias and V_L be the variance of our proposed estimator L. Let $\gamma = b - a + 1$ and $REPS$ be the number of replications.

Algorithm: **Input**: a, b, n, g.

Output: L, B_L and V_L.

1. Calculate the covariance matrix C for the vectors $\overline{W}_a, \overline{W}_{a+1}, \overline{W}_{a+2}, \ldots, \overline{W}_b$. This is a $\gamma g \times \gamma g$ matrix. Also calculate the inverse of C.

2. for each replication $r, 1 \leq r \leq REPS$, do the following:

   (a) Generate the vectors $\overline{W}_a, \overline{W}_{a+1}, \overline{W}_{a+2}, \ldots, \overline{W}_b$.

   (b) Concatenating the above γ vectors, we obtain a vector of size γg. Denote it by W.

   (c) The value of the estimator L in the rth replication is given by $L(r) = \frac{WC^{-1}W^T}{\gamma * g}$.

3. Calculate the sample mean and variance of L. Hence calculate the bias of L.

3.8 Experimental results

We conducted computer experiments on AR(1) process and M/M/1 queuing process and calculated the bias, variance, mean-squared error (MSE) and efficiency for the estimation of the TAVC of these two processes under different conditions. The results are shown in Tables 3.1-3.5. The "start index" a and the "end index" b of each of the tables denote the fact that the corresponding estimator is based on all the $b - a + 1$ integrated paths, starting with the ath integrated path and ending with the bth integrated path.

While computing the estimators, a frequently encountered problem is that the covariance matrix becomes ill-conditioned and consequently we get meaningless estimates. That is why many of the possible estimators could not be used at all and we had to remain content with only three or four estimators. The pairwise estimators also suffer a lot from this problem. We have observed that the estimators based on the (i, j)-th pair-wise paths give meaningful results only when $i = 1$. We have shown the results for the pair-wise estimators for the AR(1) process in Table 3.5.

3.8.1 AR(1) process

The AR(1) process is a stationary Gaussian process defined as follows: $Y_0 \sim N(0, 1)$ and $Y_i = \psi Y_{i-1} + \epsilon_i$, $i \geq 1$. Here, $-1 < \psi < 1$ and the $\{\epsilon_i\}$ are independent and $\epsilon_i \sim N(0, 1 - \psi^2)$. For this particular process, it is possible to compute the value of the theoretical TAVC. The value is given by $\sigma^2 = \frac{1+\psi}{1-\psi}$. For our experiments, we had chosen $\psi = 0.9$, resulting in $\sigma^2 = 19$.

We ran $1,000$ independent replications for each of the estimators of TAVC with run length $n = 20,000$. The best values obtained are compared with the values obtained by (i) the non-overlapping batch means estimator (NBM) and (ii) the overlapping batch means estimator (OBM) as given in [1]. We find that the best bias and

variance combination is obtained from the second row of Table 3.2 and is given by $(-0.101, 16.910)$. This compares favorably (both in terms of bias and variance) with the NBM and OBM estimators given in Tables 2 and 3 of [1], which are respectively $(0.18, 38.00)$ and $(0.15, 25.56)$.

We also ran another set of experiments for $10,000$ replications which gave slightly better results, as shown in Tables 3.3 and 3.4. Finally, we have also considered some pair-wise estimators that gave meaningful results and reported the results in Table 3.5.

3.8.2 M/M/1 waiting time process

We also considered the M/M/1 waiting time sequence, where arrivals and service times are independent and exponentially distributed with mean 0.8. For traffic intensity ρ, the theoretical value of the TAVC is known in this case also. The value is given by $\sigma^2 = \frac{\rho(2+5\rho-4\rho^2+\rho^3)}{(1-\rho)^4}$. The derivation of this result is given in [74]. We had chosen $\rho = 0.8$, which gives $\sigma^2 = 1976$.

We ran $1,000$ independent replications for each of the estimators of TAVC with run length $n = 20,000$. The values obtained are compared with the values obtained by (i) the non-overlapping batch means estimator (NBM) and (ii) the overlapping batch means estimator (OBM) as given in [1]. We find that the best bias and variance combination is obtained from the row 2 of Table 3.6 and is given by $(-1.63, 4.08 \times 10^6)$. This compares favorably in terms of bias, but not in terms of variance with the NBM and OBM estimators given in Tables 4 and 5 of [1], which are respectively $(116, 1.866 \times 10^6)$ and $(125, 1.762 \times 10^6)$. We also ran another set of experiments with $10,000$ replications and obtained slightly different values as shown in Tables 3.8 and 3.9.

3.9 Discussion

In many cases, we find that the covariance matrix becomes close to singular and consequently we get unreliable estimates. This is the reason why we practically get few useful estimators out of several possible combinatorial combinations of integrated path

Table 3.1 AR(1) process experiments with $n = 20,000$, grid size = 4, REPS = $1,000$.

Start index	End index	Bias	Variance	MSE
1	1	0.377	195.329	195.471
1	2	−0.272	85.820	85.746
1	3	−0.033	56.366	56.365

Table 3.2 AR(1) process experiments with $n = 20,000$, grid size = 20, REPS = $1,000$.

Start index	End index	Bias	Variance	MSE
1	1	0.057	38.369	38.372
1	2	−0.101	16.910	16.921
1	3	0.491	62.803	63.044

Table 3.3 AR(1) process experiments with $n = 20,000$, grid size = 4, REPS = $10,000$.

Start index	End index	Bias	Variance	MSE
1	1	−0.19	170.60	170.64
1	2	−0.12	89.31	89.33
1	3	−0.69	58.37	58.84

Table 3.4 AR(1) process experiments with $n = 20,000$, grid size = 20, REPS = $10,000$.

Start index	End index	Bias	Variance	MSE
1	1	−0.03	34.96	34.96
1	2	−0.06	17.61	17.62
1	3	−0.002	59.77	59.77

Table 3.5 Pairwise paths: AR(1) process experiments with $n = 20,000$, grid size = 20, REPS = 1000.

First path	Second path	Bias	Variance	MSE
1	2	0.040	16.939	16.941
1	3	−0.098	15.696	15.706
1	4	−0.011	19.005	19.005
1	5	0.011	17.736	17.736

Table 3.6 M/M/1 queuing process experiments with grid size 4, $n = 20,000$, REPS $= 1,000$.

Start index	End index	Bias	Variance ($\times 10^6$)	MSE ($\times 10^6$)
1	1	-111.51	3.21	3.22
1	2	-1.63	4.08	4.08
1	3	-120.73	1.80	1.82

Table 3.7 M/M/1 queuing process experiments with grid size 20, $n = 20,000$, REPS $= 1,000$.

Start index	End index	Bias	Variance ($\times 10^6$)	MSE ($\times 10^6$)
1	1	-30.13	2.66	2.67
1	2	-24.55	3.00	3.01
1	3	-23.24	4.77	4.77

Table 3.8 M/M/1 queuing process experiments with grid size 4, $n = 20,000$, REPS $= 10,000$.

Start index	End index	Bias	Variance ($\times 10^6$)	MSE ($\times 10^6$)
1	1	-36.64	2.1	2.1
1	2	-125.67	1.5	1.5
1	3	11.36	3.1	3.1

Table 3.9 M/M/1 queuing process experiments with grid size 20, $n = 20,000$, REPS $= 10,000$.

Start index	End index	Bias	Variance ($\times 10^6$)	MSE ($\times 10^6$)
1	1	-50.7	4.4	4.4
1	2	-2.17	2.7	2.7
1	3	-52.3	2.5	2.5

estimators. The general observation about the experimental results is that usually we get very low bias values when we consider more than one integrated path. Inclusion of higher order integrated paths sometimes leads to numerical instabilities and gives us counter-intuitive results. Empirically we find that the combination of the first and the second integrated paths may give promising values in many cases, though inclusion of the third integrated path may further reduce bias (provided it does not result in numerical problems). For the AR(1) experiments, the pairwise paths also gave us very low values of bias and variance, though the results become instable if we exclude the first order integrated paths. The bias and variance values are usually much lower in the case of AR(1) experiments compared to the M/M/1 experiments. It may be possible to reduce the bias further by computing a bias correction of first order/possibly higher order.

CHAPTER 4

CONCLUSIONS

We established the asymptotic validity of two-stage selection and MCB procedures under the general setting of a parameter estimator satisfying a random-time-change CLT (Assumption 1) and the existence of a weakly convergent variance estimator for the variance parameter appearing in the CLT (Assumption 2). For the case when η in the CLT takes on the canonical value of $1/2$ and the first-stage run length is asymptotically negligible compared to the total run length (i.e., when $\lambda < 2$), Theorems 1(iii) and 3(iii) established the asymptotic efficiency of our two-stage WCVE MCB methods in the sense that the total run length is asymptotically equivalent to what it would be when the variances σ_i^2 are known. However, our empirical results seem to indicate that it takes smaller values of δ for the asymptotics to take effect when $\lambda < 2$ compared to when $\lambda = 2$. In the setting of steady-state simulations, we also provided a comparison of our WCVE MCB methods with those based on STS, and showed that WCVE methods are asymptotically strictly better (in terms of the mean and variance of total run lengths) than any STS method with a fixed number of batches. As the number of batches grows to infinity, the methods become equivalent.

We have also given an estimator for TAVC based on the integrated paths. Due to ill-conditioning of the resulting covariance matrix, many of the possible estimators could not be used and we had to remain content with only three or four estimators. Though we compared the estimators with NBM and OBM estimators only, it may be mentioned that our estimator outperforms many of the estimators presented in Tables 2-5 of [1]. The novel feature of this estimator is significantly less bias, at the cost of not so significant increase in variance. This estimator is really a counter example of the so-called bias variance trade-off among the class of estimators for TAVC. In the future, we

plan to incorporate a first order (possibly higher order) bias correction to the proposed estimator, which is expected to reduce the bias further, without affecting the variance.

REFERENCES

[1] C. Alexopoulos, N. T. Argon, D. Goldsman, N. M. Steiger, G. Tokol, and J. R. Wilson. Efficient computation of overlapping variance estimators for simulation. *INFORMS Journal on Computing*, 19(3):314–327, 2007.

[2] Christos Alexopoulos, Nilay Tanik Argon, David Goldsman, Gamze Tokol, and James R. Wilson. Overlapping variance estimators for simulation: I. theory. Website, 2005. `ftp://ftp.ncsu.edu/pub/eos/pub/jwilson/ovestv50.pdf`.

[3] T. Altiok and B. Melamed. *Simulation Modeling and analysis with Arena*. Academic Pr, 2007.

[4] F. J. Anscombe. Large-sample theory of sequential estimation. *Proceedings of the Cambridge Philosophical Society*, 48:600–607, 1952.

[5] R. E. Bechhofer. A single-sample multiple-decision procedure for ranking means of normal populations with known variances. *Annals of Mathematical Statistics*, 25:16–39, 1954.

[6] R. E. Bechhofer, T. J. Santner, and D. M. Goldsman. *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. John Wiley & Sons, New York, 1995.

[7] R. E. Bechhofer, T. J. Santner, and D. M. Goldsman. Design and analysis of experiments for statistical selection, screening, and multiple comparisons. Website, 1995. `http://www.stat.ohio-state.edu/~tjs/REB-TJS-DMG/reb-tjs-dmg.zip` (accessed April, 2011).

[8] P. Billingsley. Limit theorems for randomly selected partial sums. *Annals of Mathematical Statistics*, 33:85–92, 1962.

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

[10] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, second edition, 1999.

[11] E. Bofinger. Estimation of a density function using order statistics. *Australian Journal of Statistics*, 17:1–7, 1975.

[12] P. Bratley, B. L. Fox, and L. E. Schrage. *A Guide to Simulation*. Springer-Verlag, New York, second edition, 1987.

[13] J. M. Calvin. Simulation output analysis using integrated paths. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 17(3):1–24, 2007.

[14] J. M. Calvin. Simulation output analysis using integrated paths II: Low bias estimators. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 19(3):1–28, 2009.

[15] J. M. Calvin and M. K. Nakayama. Permuted standardized time series for steady-state simulations. *Mathematics of Operations Research*, 31:351–368, 2006.

[16] C. Chien, D. Goldsman, and B. Melamed. Large-sample results for batch means. *Management Science*, pages 1288–1295, 1997.

[17] M. A. Crane and D. L. Iglehart. Simulating stable stochastic systems: III. Regenerative processes and discrete-event simulations. *Operations Research*, 23(1):33–45, 1975.

[18] H. Damerdji. Strong consistency and other properties of the spectral variance estimator. *Management Science*, 37:1424–1440, 1991.

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

[20] H. Damerdji, P. W. Glynn, M. K. Nakayama, and J. R. Wilson. Selecting the best system in transient simulations with variances known. In J. Charnes, D. Morrice, D. Brunner, and J. Swain, editors, *Proceedings of the 1996 Winter Simulation Conference*, pages 281–286, Piscataway, New Jersey, 1996. Institute of Electrical and Electronics Engineers.

[21] H. Damerdji and M. K. Nakayama. Two-stage multiple-comparison procedures for steady-state simulations. *ACM Transactions on Modeling and Computer Simulations*, 9:1–30, 1999.

[22] S. D. Duket. Simulation output analysis. Master's thesis, Purdue University., 1974.

[23] P. Eykhoff. *System identification: parameter and state estimation.* Wiley & Sons, 1974.

[24] G. S. Fishman. *Principles of Discrete Event Simulation.* Wiley, New York, 1978.

[25] G. S. Fishman. *Principles of discrete event simulation.[Book review].* John Wiley & Sons, New York, NY, 1978.

[26] B. Flury. *A first course in multivariate statistics.* Springer Verlag, 1997.

[27] B. L. Fox and P. W. Glynn. Replication schemes for limiting expectations. *Probability in the Engineering and Informational Sciences*, 3:299–318, 1989.

[28] P. W. Glynn and D. L. Iglehart. Importance sampling for stochastic systems. *Management Science*, 35:1367–1393, 1989.

[29] P. W. Glynn and D. L. Iglehart. Simulation output analysis using standardized time series. *Mathematics of Operations Research*, 15:1–16, 1990.

[30] P. W. Glynn and D. L. Iglehart. Conditions for the applicability of the regenerative method. *Management Science*, 39:1108–1111, 1993.

[31] P. W. Glynn and W. Whitt. The asymptotic efficiency of simulation estimators. *Operations Research*, 40:505–520, 1992.

[32] P. W. Glynn and W. Whitt. The asymptotic validity of sequential stopping rules in stochastic simulations. *Annals of Applied Probability*, 2:180–198, 1992.

[33] D. Goldsman and M. Meketon. *A comparison of several variance estimators.* Georgia Institute of Technology, 1986.

[34] P. Hall and S. J. Sheather. On the distribution of a Studentized quantile. *Journal of the Royal Statistical Society B*, 50:381–391, 1988.

[35] P. Heidelberger and P. D. Welch. A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM*, 24(4):233–245, 1981.

[36] N. Henze and Y.Y. Nikitin. A new approach to goodness-of-fit testing based on the integrated empirical process. *Journal of Nonparametric Statistics*, 12(3):391–416, 2000.

[37] N. Henze and Y.Y. Nikitin. Watson-type goodness-of-fit tests based on the integrated empirical process. *Mathematical Methods of Statistics*, 11(2):183–202, 2002.

[38] Y. Hochberg and A. C. Tamhane. *Multiple Comparison Procedures*. Wiley, New York, 1987.

[39] J. C. Hsu. Constrained simultaneous confidence intervals for multiple comparisons with the best. *Annals of Statistics*, 12:1136–1144, 1984.

[40] J. C. Hsu and Barry L. Nelson. Control variates for quantile estimation. *Management Science*, 36:835–851, 1990.

[41] D. L. Iglehart. *Simulating stable stochastic systems, VII: Selecting the best system*. Elsevier North-Holland, Amsterdam, 1977.

[42] D. L. Iglehart. *The regenerative method for simulation analysis*. Prentice Hall, Englewood Cliffs, New Jersey, 1978.

[43] Alan F. Karr. *Probability (Springer Texts in Statistics)*. Springer, Berlin/DE, 1993.

[44] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics*, 23:462–466, 1952.

[45] S. H. Kim and B. L. Nelson. On the asymptotic validity of fully sequential selection procedures for steady-state simulation. *Operations Research*, 54:475–488, 2006.

[46] S.-H. Kim and B. L. Nelson. *Selecting the Best System*. Elsevier, Amsterdam, 2006.

[47] D. E. Knuth. *The Art of Computer Programming: Fundamental Algorithms*. Addison-Wesley, Reading, Massachusetts, third edition, 1997.

[48] A. Lachal. Study of some new integrated statistics: computation of Bahadur efficiency, relation with non-standard boundary value problems. *Mathematical Methods of Statistics*, 10(1):73–104, 2001.

[49] A. M. Law. *Simulation Modeling and Analysis*. McGraw-Hill, New York, fourth edition, 2006.

[50] A. M. Law and W. D. Kelton. Confidence intervals for steady-state simulations: I. A survey of fixed sample size procedures. *Operations Research*, 32(6):1221–1239, 1984.

[51] E. L. Lehmann. *Elements of Large-Sample Theory*. Springer, New York, 1999.

[52] F. J. Matejcik and B. L. Nelson. Two-stage multiple comparisons with the best for computer simulation. *Operations Research*, 43:633–640, 1995.

[53] M. S. Meketon and B. Schmeiser. Overlapping batch means: Something for nothing? In *Proceedings of the 16th conference on Winter simulation*, pages 226–230. IEEE Press, 1984.

[54] N. Mukhopadhyay. Some comments on two-stage selection procedures. *Communications in Statistics, Series A*, 8:671–683, 1979.

[55] M. K. Nakayama. Two-stage stopping procedures based on standardized time series. *Management Science*, 40:1189–1206, 1994.

[56] M. K. Nakayama. Selection and multiple comparisons for regenerative systems. In L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, editors, *Proceedings of the 2006 Winter Simulation Conference*, pages 287–293, Piscataway, New Jersey, 2006. Institute of Electrical and Electronics Engineers.

[57] M. K. Nakayama. Two-stage multiple comparison procedures for steady-state simulations: Run-length variability and choosing first-stage run lengths. In preparation, 2010.

[58] Y. Nikitin. *Asymptotic efficiency of nonparametric tests*. Cambridge University Press, 1995.

[59] Y. Rinott. On two-stage selection procedures and related probability-inequalities. *Communications in Statistics—Theory and Methods*, A7:799–811, 1978.

[60] D. Ruppert. Almost sure approximations to the Robbins-Munro and Kiefer-Wolfowitz processes with dependent noise. *Annals of Probability*, 10:178–187, 1982.

[61] R. G. Sargent. An introduction to statistical analysis of simulation output data. *AGARD Modeling and Simulation of Avionics Systems and Command, Control and Commun. 13 p(SEE N 80-19809 10-59)*, 1980.

[62] B. Schmeiser. Batch size effects in the analysis of simulation output. *Operations Research*, 30(3):556–568, 1982.

[63] T. J. Schriber and R. W. Andrews. ARMA-based confidence intervals for simulation output analysis. *American Journal of Mathematical and Management Sciences*, 4(3):345–373, 1984.

[64] L. W. Schruben. Confidence interval estimation using standardized time series. *Operations Research*, 31(6):1090–1108, 1983.

[65] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York, 1980.

[66] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*. Society for Industrial Mathematics, 2009.

[67] D. S. Silvestrov. *Limit Theorems for Randomly Stopped Stochastic Processes*. Springer, London, 2004.

[68] D. Slepian. The one-sided barrier problem for Gaussian noise. *Bell System Technical Journal*, 41:463–501, 1962.

[69] W. T. Song and B. W. Schmeiser. Variance of the sample mean: Properties and graphs of quadratic-form estimators. *Operations Research*, pages 501–517, 1993.

[70] W. T. Song and B. W. Schmeiser. Optimal mean-squared-error batch sizes. *Management Science*, pages 110–123, 1995.

[71] D. W. Sullivan and J. R. Wilson. Restricted subset selection procedures for simulation. *Operations Research*, 37:52–71, 1989.

[72] J. R. Swisher, S. H. Jacobson, and E. Yucesan. Discrete-event simulation optimization using ranking, selection, and multiple comparison procedures: A survey. *ACM Transactions on Modeling and Computer Simulation*, 13:134–154, 2003.

[73] J. H. Ventner. An extension of the Robbins-Munro procedure. *Annals of Mathematical Statistics*, 38:181–190, 1967.

[74] W. Whitt. Planning queueing simulations. *Management Science*, 35(11):1341–1366, 1989.

[75] W. Whitt. *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer-Verlag, New York, 2002.