

**VARIANCE ESTIMATION IN STEADY-STATE
SIMULATION, SELECTING THE BEST SYSTEM, AND
DETERMINING A SET OF FEASIBLE SYSTEMS VIA
SIMULATION**

A Thesis
Presented to
The Academic Faculty

by

Demet Batur

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Industrial and Systems Engineering
Georgia Institute of Technology
May 2006

Copyright © 2006 by Demet Batur

**VARIANCE ESTIMATION IN STEADY-STATE
SIMULATION, SELECTING THE BEST SYSTEM, AND
DETERMINING A SET OF FEASIBLE SYSTEMS VIA
SIMULATION**

Approved by:

Dr. Seong-Hee Kim, advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. David Goldsman, advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Sigrún Andradóttir
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Brani Vidakovic
School of Biomedical Engineering
Georgia Institute of Technology

Dr. Barry L. Nelson
Department of Industrial Engineering and
Management Sciences
Northwestern University

Date Approved: 24 March 2006

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere thanks to my advisors, Dr. Seong-Hee Kim and Dr. David Goldsman, for their guidance, support, understanding, and friendship. I am grateful to them for believing in me and helping me encounter all the obstacles of this hard study.

I also would like to thank Dr. Barry L. Nelson for travelling all the way from Illinois for my proposal and defense presentations and for his valuable comments and suggestions in every detail of my thesis.

I am also thankful to Dr. Sigrún Andradóttir and Dr. Brani Vidakovic for their willingness to serve in my committee and for their time and valuable input.

I also would like to thank Can Vuran for his friendship, love, support, and companionship throughout the hours-long studies.

Finally, and most importantly, I want to thank my parents and my brother, Ümit, for their love, endless support, and encouragement. Without them I would not be able to achieve my goals. I am grateful to all of them.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	viii
SUMMARY	ix
I INTRODUCTION	1
II AN IMPROVED STANDARDIZED TIME SERIES DURBIN–WATSON VARIANCE ESTIMATOR FOR STEADY-STATE SIMULATION . .	5
2.1 Background	6
2.1.1 Properties of the Output Data	6
2.1.2 The Weighted Area Estimator	6
2.1.3 The Weighted Cramér–von Mises Estimator	7
2.1.4 The Weighted Durbin–Watson Estimator	8
2.1.5 The Jackknifed Durbin–Watson Estimator	9
2.2 The Modified Jackknifed Durbin–Watson Estimator	10
2.3 Batching	12
2.3.1 Batched STS Estimators	13
2.3.2 Nonoverlapping Batch Means Estimator	15
2.3.3 Overlapping Batch Means Estimator	16
2.3.4 Recapitulation	16
2.4 Rebatching	16
2.5 Examples	20
2.5.1 Exact Example: Moving Average Process	20
2.5.2 Monte Carlo Example: Autoregressive Process	22
2.6 Conclusion	25
III FULLY SEQUENTIAL SELECTION PROCEDURES WITH PARABOLIC BOUNDARY	26
3.1 Procedures	28
3.1.1 First Procedure	28

3.1.2	Second Procedure	35
3.2	Design of the Procedures	35
3.2.1	Choice of λ	36
3.2.2	Choice of ξ	39
3.3	Analytical Comparisons of the Continuation Regions	43
3.4	Experimental Results	47
3.4.1	Independent Case	49
3.4.2	Correlated Case	51
3.5	Conclusion	53
IV	FEASIBILITY	54
4.1	Problem	56
4.2	Generic Algorithms	59
4.2.1	Multiple Feasibility Check Procedure (\mathcal{F}_B)	59
4.2.2	Accelerated Feasibility Check Procedure (\mathcal{F}_A)	62
4.2.3	Accelerated Feasibility Check Procedure with Variance Update (\mathcal{F}_A^+)	67
4.3	Example Procedures	67
4.4	Empirical Evaluation	72
4.4.1	Multivariate Normal Example	72
4.4.2	Queueing Example	77
4.5	Conclusion	79
V	CONTRIBUTIONS	80
APPENDIX A	— REBATCHED ESTIMATORS	81
APPENDIX B	— MA(1) EXAMPLE	84
APPENDIX C	— PROCEDURE \mathcal{P}_2	86
APPENDIX D	— DETERMINATION OF VECTOR \mathbf{A}	89
REFERENCES	92

LIST OF TABLES

1	Estimated Mean and Variance of the Area, CvM, DW, JDW, and MJDW Estimators (Based on One Batch) for the Variance Parameter ($\sigma^2 = 19$) of an AR(1) Process with $\phi = 0.9$	22
2	Estimated Mean and Variance of the Batched Area, CvM, JDW, MJDW, NBM, and OBM Estimators for the Variance Parameter ($\sigma^2 = 19$) of an AR(1) Process with $\phi = 0.9$ and $n = 4096$ [16384]	23
3	Estimated Mean and Variance of the Rebatched Area, CvM, JDW, and MJDW Estimators for the Variance Parameter ($\sigma^2 = 19$) of an AR(1) Process with $\phi = 0.9$ and $n = 4096$ [16384]	24
4	$\xi_{\beta_{min}}$ and ξ_{α} Values when $\lambda = 0.9268$, $n_0 = 10, 15, 20, 24$, and $\alpha = 0.01, 0.025, 0.05, 0.10$ in \mathcal{P}_1	41
5	$\xi_{\beta_{min}}$ and ξ_{α} Values when $\lambda = 0.3$, $n_0 = 10, 15, 20, 24$, and $\alpha = 0.01, 0.025, 0.05, 0.10$ in \mathcal{P}_2	41
6	Actual and Approximate ξ_{β} Values when $n_0 = 20$ and $\lambda = 0.9268$ in \mathcal{P}_1 . . .	42
7	$V_{\mathcal{KN}}/V_{\mathcal{P}_1}$ when $n_0 = 10, 24$ and $k = 2, 5, 10, 25, 100, 200, 400$	45
8	$V_{\mathcal{KN}}/V_{\mathcal{P}_2}$ and $H_{\mathcal{KN}}/H_{\mathcal{P}_2}$ when $n_0 = 10, 24$ and $k = 2, 5, 10, 25, 100, 200, 400$	46
9	$V_{\mathcal{P}_1}/V_{\mathcal{P}_2}$ when $n_0 = 10, 24$ and $k = 2, 5, 10, 25, 100, 200, 400$	46
10	The Configurations of Means and Variances Used in the Experiments	48
11	Estimated PCS in \mathcal{KN} , \mathcal{P}_1 , and \mathcal{P}_2 when $n_0 = 24, \delta = 1/\sqrt{n_0}$, and $1 - \alpha = 0.95$ as a Function of k when Systems are Simulated Independently	50
12	Sample Average of the Total Number of Observations in \mathcal{KN} , \mathcal{P}_1 , and \mathcal{P}_2 when $n_0 = 24, \delta = 1/\sqrt{n_0}$, and $1 - \alpha = 0.95$ as a Function of k	51
13	Sample Average of the Total Number of Observations in \mathcal{KN} , \mathcal{P}_1 , and \mathcal{P}_2 when $n_0 = 24, \delta = 1/\sqrt{n_0}, 1 - \alpha = 0.95$, and $k = 25$ as a Function of ρ . . .	52
14	The Mean Configurations of the Performance Measures Associated with s Stochastic Constraints	72
15	Sample Average of Total Number of Vectors (SANV) when One System ($k = 1$) is Considered with $s = 5, \rho = 0.0$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$	73
16	Estimated PCD when One System ($k = 1$) is Considered with $s = 5, \rho = 0.0$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$	73
17	SANV when One System ($k = 1$) is Considered with $s = 5, \rho = -0.15$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$	74
18	Estimated PCD when One System ($k = 1$) is Considered with $s = 5, \rho = -0.15$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$	74

19	SANV when One System ($k = 1$) is Considered with $s = 5$, $\rho = 0.3$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$	74
20	Estimated PCD when One System ($k = 1$) is Considered with $s = 5$, $\rho = 0.3$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$	75
21	SANV and Estimated PCD when Nine Systems ($k = 9$) are Considered with $s = 5$ and $\alpha = \alpha_0 = \alpha_1 = 0.05$	76
22	SANV and Estimated PCD when Nine Systems ($k = 9$) are Considered with $s = 5$, $\alpha = 0.05$, $\alpha_0 = 0.05/(k + 1)$, and $\alpha_1 = 0.05k/(k + 1)$	76
23	External Arrival and Service Rates for the Job Shop Example	78
24	Average Buffer Size and Tolerance Levels for the Job Shop Example	78

LIST OF FIGURES

1	The DW Estimator Based on Different Portions of a Sample of Size n	10
2	The Area Estimator Based on Different Portions of a Sample of Size n when the Number of Rebatching Levels is Two	17
3	Triangular and Parabolic Continuation Regions	27
4	$\lambda \xi^{3/2}$ versus λ when $n_0 = 10$ and $\beta = 0.05$ in \mathcal{P}_1	36
5	$\lambda \xi^{3/2}$ versus λ when $n_0 = 24$ and $\beta = 0.05$ in \mathcal{P}_2	37
6	$g(\xi; \lambda, n_0)$ versus ξ Graph when $n_0 = 10$ and $\lambda = 0.9268$ in \mathcal{P}_1	38
7	η^2 versus ξ Graph when $n_0 = 20$, $\lambda = 0.9268$, and $0.002050 < \beta < 0.05$ in \mathcal{P}_1	38
8	η^2 versus ξ Graph when $n_0 = 10$, $\lambda = 0.3$, and $0.00244 < \beta < 0.05$ in \mathcal{P}_2	39
9	Linearity Relationship Between η_β^2 and ξ_β	40
10	Continuation Regions of \mathcal{P}_1 and \mathcal{KN}	44
11	Continuation Regions of \mathcal{P}_2 and \mathcal{KN} for Large k	45
12	Desirable (D), Acceptable (A), and Unacceptable (U) Regions when There Exist Two Stochastic Constraints	58
13	A Triangular Continuation Region for the Constraint $y \leq Q$	59
14	Algorithmic Statement of \mathcal{F}	60
15	Algorithmic Statement of \mathcal{F}_B	61
16	D^a , A^a , and U^a Regions for the Aggregated Measure when There Exist Two Stochastic Constraints	62
17	Algorithmic Statement of \mathcal{F}_A	63
18	Algorithmic Statement of \mathcal{F}_A^+	65
19	Algorithmic Statement of \mathcal{F}_B^{AGK}	66
20	Algorithmic Statement of \mathcal{F}_A^{AGK}	69
21	Algorithmic Statement of \mathcal{F}_A^{AGK+}	70
22	Job Shop Example	77
23	A Possible Choice of ℓ_1 and ℓ_2	89
24	Determining Line ℓ_1	90

SUMMARY

In this thesis, we first present a variance estimation technique based on the standardized time series methodology for steady-state simulations. The proposed variance estimator has competitive bias and variance compared to the existing estimators in the literature. We also present the technique of rebatching to further reduce the bias and variance of our variance estimator. Second, we present two fully sequential indifference-zone procedures to select the best system from a number of competing simulated systems when best is defined by the maximum or minimum expected performance. These two procedures have parabola shaped continuation regions rather than the triangular continuation regions employed in several papers. The procedures we present accommodate unequal and unknown variances across systems and the use of common random numbers. However, we assume that basic observations are independent and identically normally distributed. Finally, we present procedures for finding a set of feasible or near-feasible systems among a finite number of simulated systems in the presence of multiple stochastic constraints, especially when the number of systems or constraints is large.

CHAPTER I

INTRODUCTION

Simulation is often used to study the behavior of a system of interest or compare a number of systems, especially when a system is too complicated to apply analytical or numerical methods.

Since random samples from probability distributions are used as input to derive output from a simulation, the output data are also random and they must be analyzed carefully by appropriate statistical techniques. There are two types of simulations with regard to output analysis: terminating and steady-state. In terminating simulations, we are interested in the performance of the system over a finite, possibly random, time period (e.g., a bank that opens at 8 am and closes after all the customers that come before 5 pm leave). In steady-state simulations, we are interested in the long-run average performance of the system as time goes to infinity assuming that the system eventually settles down (e.g., a continuously running manufacturing line).

For steady-state simulation output process, Y_1, Y_2, \dots, Y_n , we typically estimate the mean performance, μ , by \bar{Y}_n , the sample average of the first n observations. In order to give a measure of the precision of \bar{Y}_n or to build a confidence interval for μ , we also need to estimate the *variance parameter*, $\sigma^2 \equiv \lim_{n \rightarrow \infty} n \text{Var}(\bar{Y}_n)$.

There are a number of different techniques in the literature for the estimation of the variance parameter, e.g., the methods of multiple replications (MR), nonoverlapping batch means (NBM) ([33]), overlapping batch means (OBM) ([28]), replicated batch means ([1]) — a combination of MR and NBM — and standardized time series (STS) ([34]). The weighted area ([19]) and weighted Cramér–von Mises (CvM) ([16]) estimators are the most well-studied variance parameter estimators based on the STS methodology. These two estimators are combined in [15] to obtain new STS estimators — the Durbin–Watson (DW) and jackknifed DW (JDW) — with competitive bias and lower asymptotic variance than

the area, CvM, and NBM estimators. In this thesis, we improve the JDW estimator in such a way that the resulting estimator has an asymptotic variance even smaller than that of the JDW estimator while maintaining almost the same bias value. We also propose the technique of *rebatching* to further reduce the asymptotic variance and bias of this resulting estimator. Rebatching differs from batching by the way it reuses the same data with different batch sizes.

We are also concerned with ranking and selection (R&S) procedures, which typically perform comparisons among a finite number of simulated systems with some guarantee about correctness. There are four main types of comparison problems in simulation: (1) selection-of-the-best, (2) comparison with a standard, (3) multinomial selection, and (4) Bernoulli selection. For the details of these comparison problems, see [20]. Selection-of-the-best is probably the most popular type of comparison problem in simulation. In this problem, the goal is to choose the alternative with the largest or smallest expected performance measure among a finite number of simulated systems.

As a solution to the selection-of-the-best problem, a number of different approaches have been proposed. The indifference-zone approach tries to choose a system whose performance measure is at least a user-specified constant — called the indifference-zone parameter — better than all the other alternative systems with the probability of correct selection (PCS) no less than a pre-specified amount. Recent references include [29], [5], [22], and [24]. On the other hand, in [11], [10], and [9], completely different procedures from a decision-theoretic point of view, and in [7] and [8] heuristic procedures that maximize PCS under a budget constraint are proposed. Instead of providing a PCS guarantee, these Bayesian procedures attempt to allocate a finite computation budget to maximize the posterior PCS of the selected system.

In [22], a fully sequential procedure is defined as one that takes a single basic observation from each alternative that is still in play at the current stage of sampling and eliminates systems immediately when there is evidence that they are inferior. Fully sequential procedures have a boundary called a *continuation region*. For example, in [22], the authors use a triangular continuation region in their fully sequential procedure \mathcal{KN} . The \mathcal{KN} procedure

is known to be highly efficient for selecting the best system. In this thesis, we present two fully sequential indifference-zone procedures with *parabolic* continuation regions for the selection-of-the-best problem.

While comparing alternative system designs, we may face the challenge of constraints on some secondary performance measures. A significant amount of work has been performed for the selection-of-the-best problem in the simulation community (see [17] and [23] for recent literature). However, very little work has been done for finding the best system in the presence of stochastic constraints on some secondary performance measures. In [6], multiple performance measures are handled, and in [32], a two-stage procedure with a constraint on variance is proposed. However, their methods are either very hard to apply in practice or are focused only on a special case. In [2], a R&S procedure that determines the feasibility of systems in the presence of one stochastic constraint is presented and combined with a selection-of-the-best procedure to identify the best feasible system. Their feasibility check procedure can handle a general stochastic constraint on a secondary performance measure, and determines the feasibility of systems in consideration correctly with high probability. Yet, the work needs to be further extended to the case of multiple constraints.

In this thesis, we present three generic procedures for the problem of finding a set of feasible or near-feasible systems among a finite number of simulated systems in the presence of stochastic constraints. Real-life optimization problems may include a large number of stochastic constraints. Hence, a procedure that can efficiently handle a feasibility check on multiple stochastic constraints is a critical step in solving practical problems.

First, we present a generic R&S procedure that detects feasibility of one system in the presence of one constraint. Then we extend that procedure to the case of multiple systems and constraints by the use of the Bonferroni inequality. Unfortunately, the resulting procedure tends to be conservative when the number of systems or constraints is large. As a remedy, we present a procedure that accelerates the elimination of infeasible systems by re-using collected observations across stochastic constraints for each system. We also present a variance updating version of this procedure in which variance estimates are updated as more observations are obtained.

The thesis is organized as follows: In Chapter 2, we present a new asymptotic variance parameter estimator for steady-state simulation and the technique of rebatching to further reduce the asymptotic variance and bias of this estimator. In Chapter 3, we present two fully sequential indifference-zone selection procedures with parabolic boundaries for the selection-of-the-best problem. We compare the performance of these procedures with that of \mathcal{KN} by empirical studies based on independent and identically distributed (IID) normal data. In Chapter 4, we present procedures for determining a set of feasible or near-feasible systems in the presence of multiple stochastic constraints, especially when the number of systems or constraints is large. Finally, we summarize the research contributions in Chapter 5.

CHAPTER II

AN IMPROVED STANDARDIZED TIME SERIES DURBIN–WATSON VARIANCE ESTIMATOR FOR STEADY-STATE SIMULATION

For steady-state simulation output process, Y_1, Y_2, \dots, Y_n , we typically estimate the mean performance, μ , by \bar{Y}_n , the sample average of the first n observations. In order to give a measure of the precision of \bar{Y}_n or to build a confidence interval for μ , we also need to estimate the *variance parameter*, $\sigma^2 \equiv \lim_{n \rightarrow \infty} n \operatorname{Var}(\bar{Y}_n)$.

The weighted area ([19]) and weighted Cramér–von Mises (CvM) ([16]) estimators are the most well-studied variance parameter estimators based on the standardized time series (STS) methodology. These two estimators are combined in [15] to obtain new STS estimators — the Durbin–Watson (DW) and jackknifed DW (JDW) — with competitive bias and lower asymptotic variance than the area, CvM, and nonoverlapping batch means (NBM) estimators. In this chapter, we improve the JDW estimator in such a way that the resulting estimator, the modified jackknifed Durbin–Watson (MJDW) estimator, has an asymptotic variance even smaller than that of the JDW estimator while maintaining almost the same bias value. Then we apply the technique of *rebatching* to further reduce the bias and asymptotic variance of the MJDW. Rebatching differs from batching by the way it reuses the same data with different batch sizes.

The chapter is organized as follows: In Section 2.1, we give some background information about the area, CvM, DW, and JDW estimators. In Section 2.2, we define the improved estimator and present a theorem related to the expectation and asymptotic variance of this new estimator. Section 2.3 discusses the method of batching along with the NBM and OBM estimators, and gives the expected value and asymptotic variance of batched versions of all the estimators. Section 2.4 discusses the method of rebatching, and gives the expected value

and asymptotic variance of rebatched versions of all the estimators. Section 2.5 provides some exact and Monte Carlo examples that illustrate the performance of the new estimator. Section 2.6 concludes the chapter.

2.1 Background

We review some results from the literature that will be used in the rest of our chapter.

2.1.1 Properties of the Output Data

We assume that the output data, Y_1, Y_2, \dots, Y_n , are from a stationary stochastic process (e.g., a steady-state simulation process) that satisfies a Functional Central Limit theorem (FCLT) condition, given below:

Assumption FCLT *There exist μ and positive σ such that as $n \rightarrow \infty$,*

$$X_n \Rightarrow \sigma \mathcal{W},$$

where \mathcal{W} is a standard Brownian motion process, \Rightarrow denotes weak convergence as $n \rightarrow \infty$, and

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

where $\bar{Y}_j \equiv \sum_{k=1}^j Y_k / j$, $j = 1, 2, \dots, n$, and $\lfloor \cdot \rfloor$ is the greatest integer function.

The *standardized time series* of a stochastic process Y_1, Y_2, \dots, Y_n is defined as

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

Under the FCLT assumption, it can be shown that $T_n \Rightarrow \mathcal{B}$, where \mathcal{B} is a standard Brownian bridge process on $[0, 1]$. All finite-dimensional joint distributions of \mathcal{B} are normal with $E[\mathcal{B}(t)] = 0$ and $\text{Cov}(\mathcal{B}(s), \mathcal{B}(t)) = \min(s, t) - st$, $0 < s, t < 1$.

2.1.2 The Weighted Area Estimator

The square of the weighted area under the STS and its limiting functional are defined as

$$A(f; n) \equiv \left(\frac{1}{n} \sum_{k=1}^n f\left(\frac{k}{n}\right) \sigma T_n\left(\frac{k}{n}\right) \right)^2$$

and

$$A(f) \equiv \left(\int_0^1 f(t) \sigma \mathcal{B}(t) dt \right)^2,$$

respectively, where the weighting function $f(t)$ is continuous on $[0, 1]$ and chosen to satisfy $E[A(f)] = \sigma^2$. Under mild conditions, the continuous mapping theorem (CMT) ([4]) implies $A(f; n) \xrightarrow{\mathcal{D}} A(f)$ as $n \rightarrow \infty$, where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution; and so we call $A(f; n)$ the *weighted area estimator* for σ^2 .

Theorem 1 gives expressions for the expected value and asymptotic variance of $A(f; n)$. Let the covariance function $R_k \equiv \text{Cov}(Y_1, Y_{1+k})$, $k = 0, \pm 1, \pm 2, \dots$, and the constant $\gamma \equiv -2 \sum_{k=1}^{\infty} k R_k$ ([35]). In addition, let $F \equiv \int_0^1 f(s) ds$, $\bar{F} \equiv \int_0^1 F(s) ds$, and $F^* \equiv [(F - \bar{F})^2 + \bar{F}^2]/2$. Finally, the notation $p(n) = o(q(n))$ means that $p(n)/q(n) \rightarrow 0$ as $n \rightarrow \infty$.

Theorem 1 ([14] and [19]) *Suppose Y_1, Y_2, \dots, Y_n is a stationary process for which Assumption FCLT holds, $\sum_{k=1}^{\infty} k^2 |R_k| < \infty$, and $\sum_{k=-\infty}^{\infty} R_k = \sigma^2 > 0$. Then*

$$E[A(f; n)] = \sigma^2 + \frac{F^* \gamma}{n} + o(1/n). \quad (1)$$

If we also assume uniform integrability of $A^2(f; n)$, then as $n \rightarrow \infty$,

$$\text{Var}(A(f; n)) \rightarrow \text{Var}(A(f)) = 2\sigma^4. \quad (2)$$

Example 1 *Theorem 1 implies that the area estimator with constant weighting function $f_0(t) \equiv \sqrt{12}$, for all $t \in [0, 1]$, has $E[A(f_0; n)] = \sigma^2 + 3\gamma/n + o(1/n)$. The area estimator with weighting functions $f_2(t) \equiv \sqrt{840}(3t^2 - 3t + 1/2)$ or $f_{\cos, j}(t) \equiv \sqrt{8}\pi j \cos(2\pi j t)$, $j = 1, 2, \dots$, are examples of first-order unbiased estimators, i.e., estimators with $o(1/n)$ bias.*

2.1.3 The Weighted Cramér–von Mises Estimator

The area under the square of the STS and its limiting functional are defined as

$$C(g; n) \equiv \frac{1}{n} \sum_{k=1}^n g\left(\frac{k}{n}\right) \sigma^2 T_n^2\left(\frac{k}{n}\right)$$

and

$$C(g) \equiv \int_0^1 g(t) \sigma^2 \mathcal{B}^2(t) dt,$$

respectively, where the weighting function $g(t)$ is normalized so that $E[C(g)] = \sigma^2$ and $g''(t)$ is continuous and bounded on $[0, 1]$. Under mild conditions, the CMT implies $C(g; n) \xrightarrow{\mathcal{D}} C(g)$ as $n \rightarrow \infty$, and we call $C(g; n)$ the *weighted Cramér–von Mises* (CvM) estimator for σ^2 .

Theorem 2 concerns the expected value and asymptotic variance of $C(g; n)$. Let $G \equiv \int_0^1 g(t) dt$.

Theorem 2 ([16]) *Under the conditions of Theorem 1,*

$$E[C(g; n)] = \sigma^2 + \frac{\gamma}{n}(G - 1) + o(1/n). \quad (3)$$

If we also assume uniform integrability of $C^2(g; n)$, then as $n \rightarrow \infty$,

$$\text{Var}(C(g; n)) \rightarrow \text{Var}(C(g)) = 4\sigma^4 \int_0^1 g(t)(1-t)^2 \int_0^t g(s)s^2 ds dt. \quad (4)$$

Example 2 *Theorem 2 implies that the CvM estimator with constant weighting function $g_0(t) \equiv 6$, for all $t \in [0, 1]$, has $E[C(g_0; n)] = \sigma^2 + 5\gamma/n + o(1/n)$ and $\text{Var}(C(g_0)) = 0.8\sigma^4$.*

When the weighting functions are $g_2^(t) \equiv -24 + 150t - 150t^2$ and*

$$g_4^*(t) \equiv \frac{-1310}{21} + \frac{19270t}{21} - \frac{25230t^2}{7} + \frac{16120t^3}{3} - \frac{8060t^4}{3},$$

the resulting CvM estimators, $C(g_2^; n)$ and $C(g_4^*; n)$, are first-order unbiased for σ^2 and have limiting variances $\text{Var}(C(g_2^*)) = 1.729\sigma^4$ and $\text{Var}(C(g_4^*)) = 1.042\sigma^4$, respectively.*

2.1.4 The Weighted Durbin–Watson Estimator

In [15], the authors define

$$D(h; n) \equiv \frac{1}{n} \sum_{k=1}^n h\left(\frac{k}{n}\right) \left(\sigma T_n\left(\frac{k}{n}\right) - \frac{\sqrt{A(f_0; n)}}{\sqrt{12}} \right)^2$$

and the associated limiting functional

$$D(h) \equiv \sigma^2 \int_0^1 h(t)(\mathcal{B}(t) - \bar{\mathcal{B}})^2 dt,$$

where $\bar{\mathcal{B}} \equiv \int_0^1 \mathcal{B}(t) dt$, $h(t)$ is normalized so that $E[D(h)] = \sigma^2$, and $h''(t)$ is continuous and bounded on $[0, 1]$. By the CMT, $D(h; n) \xrightarrow{\mathcal{D}} D(h)$ as $n \rightarrow \infty$, and we call $D(h; n)$ the *weighted Durbin–Watson* (DW) estimator for σ^2 .

Theorem 3 gives expressions for the expected value and asymptotic variance of $D(h; n)$. Let $H(t) \equiv \int_0^t h(u) du$, $\bar{H}(t) \equiv \int_0^t H(u) du$, and $\hat{H}(t) \equiv \int_0^t \bar{H}(u) du$. In order to normalize $h(t)$, we note that for $s \leq t$,

$$q(s, t) \equiv \text{Cov}(\mathcal{B}(s) - \bar{\mathcal{B}}, \mathcal{B}(t) - \bar{\mathcal{B}}) = s(1-t) - \frac{s-s^2}{2} - \frac{t-t^2}{2} + \frac{1}{12}.$$

Theorem 3 ([15]) *Under the assumptions of Theorem 1,*

$$\mathbb{E}[D(h; n)] = \sigma^2 \left(1 - \frac{h(0) - h(1)}{24n} \right) + \frac{\gamma}{n} (9 - \bar{H}(1) + 2\hat{H}(1)) + o(1/n). \quad (5)$$

If we also assume uniform integrability of $D^2(h; n)$, then as $n \rightarrow \infty$,

$$\text{Var}(D(h; n)) \rightarrow \text{Var}(D(h)) = 4\sigma^4 \int_0^1 \int_0^t h(s)h(t)q^2(s, t) ds dt. \quad (6)$$

Example 3 *If we define $h_0(t) \equiv 12$, for all $t \in [0, 1]$, then $D(h_0; n) = 2C(g_0; n) - A(f_0; n)$. Further, $\mathbb{E}[D(h_0; n)] = \sigma^2 + 7\gamma/n + o(1/n)$ and $\text{Var}(D(h_0; n)) \rightarrow 0.4\sigma^4$ as $n \rightarrow \infty$.*

The DW estimator $D(h_0; n)$ has very small asymptotic variance but very high bias. To reduce bias while maintaining small asymptotic variance, in [15], a jackknifed version of $D(h_0; n)$ is introduced.

2.1.5 The Jackknifed Durbin–Watson Estimator

The “jackknifed” version of $D(h_0; n)$ is

$$D_{J,r}(n) \equiv \frac{D(h_0; n)}{1-r} - \frac{rD(h_0; rn)}{1-r}, \quad (7)$$

where r is fixed in $(0,1)$. In fact, we call $D_{J,r}(n)$ the *jackknifed Durbin–Watson* (JDW) estimator for σ^2 . Theorem 4 gives expressions for the expected value and asymptotic variance of $D_{J,r}(n)$.

Theorem 4 ([15]) *Under the assumptions of Theorem 1,*

$$\mathbb{E}[D_{J,r}(n)] = \sigma^2 + o(1/n). \quad (8)$$

If we also assume uniform integrability of $D_{J,r}^2(n)$, then as $n \rightarrow \infty$,

$$\text{Var}(D_{J,r}(n)) \rightarrow \frac{2(1+r+2r^2-2r^3)\sigma^4}{5(1-r)}. \quad (9)$$

Example 4 *In [15], the choice of $r = 1/2$ is recommended. If $r = 1/2$, then as $n \rightarrow \infty$, $\text{Var}(D_{J,1/2}(n)) \rightarrow 1.4\sigma^4$.*

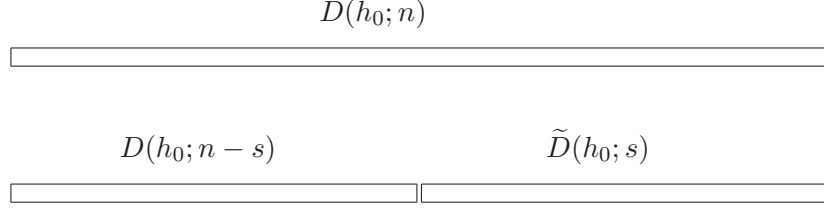


Figure 1: The DW Estimator Based on Different Portions of a Sample of Size n

2.2 The Modified Jackknifed Durbin–Watson Estimator

The main result of this chapter is a new estimator that is first-order unbiased for σ^2 but has smaller asymptotic variance than the JDW estimator.

Let $\tilde{D}(h_0; s)$, $1 \leq s \leq n$, be the DW estimator for σ^2 based on the *last* s observations out of the original n . Figure 1 illustrates three DW estimators, $D(h_0; n)$, $D(h_0; n-s)$, and $\tilde{D}(h_0; s)$, as they are applied to different portions of a sample. We can generalize the JDW estimator $D_{J,r}(n)$ from Equation (7) by including $\tilde{D}(h_0; (1-r)n)$ in the mix:

$$\tilde{D}_{J,r}(n) \equiv \beta_1 D(h_0; n) + \beta_2 D(h_0; rn) + \beta_3 \tilde{D}(h_0; (1-r)n), \quad (10)$$

where

$$\beta_1 = \frac{(-2r+1)(4r^3-6r^2-r+2)}{2r(4r^4-8r^3+2r^2+2r-1)} + \frac{1}{r}, \quad (11)$$

$$\beta_2 = \frac{r(4r^3-6r^2-r+2)}{2(4r^4-8r^3+2r^2+2r-1)}, \quad (12)$$

$$\beta_3 = \frac{(r-1)(4r^3-6r^2-r+1)}{2(4r^4-8r^3+2r^2+2r-1)}, \quad (13)$$

and r is fixed in $(0, 1)$. Note that $\beta_1 + \beta_2 + \beta_3 = 1$. We call $\tilde{D}_{J,r}(n)$ the *modified jackknifed Durbin–Watson* (MJDW) estimator for σ^2 . Note that the β coefficients are determined in such a way that the MJDW estimator is first-order unbiased.

Theorem 5 gives expressions for the expected value and asymptotic variance of $\tilde{D}_{J,r}(n)$.

Theorem 5 *Under the conditions of Theorem 1,*

$$\mathbb{E}[\tilde{D}_{J,r}(n)] = \sigma^2 + o\left(\frac{1}{n}\right). \quad (14)$$

Further, assuming that $\tilde{D}_{J,r}^2(n)$ is uniformly integrable, then as $n \rightarrow \infty$,

$$\text{Var}(\tilde{D}_{J,r}(n)) \rightarrow \left(\frac{1}{5} + \frac{r^2 - r - 1}{5(4r^4 - 8r^3 + 2r^2 + 2r - 1)} \right) \sigma^4. \quad (15)$$

Proof First, we prove the expectation result. Example 3 and symmetry imply

$$\begin{aligned} E[D(h_0; n)] &= \sigma^2 + \frac{7\gamma}{n} + o\left(\frac{1}{n}\right), \\ E[D(h_0; rn)] &= \sigma^2 + \frac{7\gamma}{rn} + o\left(\frac{1}{n}\right), \\ E[\tilde{D}(h_0; (1-r)n)] &= \sigma^2 + \frac{7\gamma}{(1-r)n} + o\left(\frac{1}{n}\right). \end{aligned}$$

Hence,

$$\begin{aligned} E[\tilde{D}_{J,r}(n)] &= \beta_1 E[D(h_0; n)] + \beta_2 E[D(h_0; rn)] + \beta_3 E[\tilde{D}(h_0; (1-r)n)] \\ &= \sigma^2 + o\left(\frac{1}{n}\right). \end{aligned} \quad (16)$$

Now, we prove the variance result. From [15], we know that as $n \rightarrow \infty$,

$$\text{Cov}(A(f_0; n), A(f_0; rn)) \rightarrow 2r^3\sigma^4, \quad (17)$$

$$\text{Cov}(A(f_0; n), C(g_0; rn)) \rightarrow (6/5)r^3\sigma^4, \quad (18)$$

$$\text{Cov}(C(g_0; n), A(f_0; rn)) \rightarrow (6/5)r^2\sigma^4, \quad (19)$$

$$\text{Cov}(C(g_0; n), C(g_0; rn)) \rightarrow (4/5)r^2\sigma^4. \quad (20)$$

From Example 3 and Equations (17)–(20), we find that as $n \rightarrow \infty$,

$$\begin{aligned} \text{Cov}(D(h_0; n), D(h_0; rn)) &= \text{Cov}(2C(g_0; n) - A(f_0; n), 2C(g_0; rn) - A(f_0; rn)) \\ &\rightarrow \frac{2}{5}r^2(2-r)\sigma^4. \end{aligned} \quad (21)$$

By symmetry,

$$\text{Cov}(D(h_0; n), \tilde{D}(h_0; (1-r)n)) \rightarrow \frac{2}{5}(1-r)^2(1+r)\sigma^4. \quad (22)$$

Notice that $D(h_0; rn)$ and $\tilde{D}(h_0; (1-r)n)$ become independent of each other as $n \rightarrow \infty$ because they are composed of disjoint sets of observations and the limiting Brownian motion has independent increments. Hence, as $n \rightarrow \infty$,

$$\text{Cov}(D(h_0; rn), \tilde{D}(h_0; (1-r)n)) \rightarrow 0. \quad (23)$$

Further, we know from Example 3 that as $n \rightarrow \infty$,

$$\text{Var}(D(h_0; n)) \rightarrow (2/5)\sigma^4. \quad (24)$$

Thus, in the same way, as $n \rightarrow \infty$,

$$\text{Var}(D(h_0; rn)) \rightarrow (2/5)\sigma^4 \quad (25)$$

and

$$\text{Var}(\tilde{D}(h_0; (1-r)n)) \rightarrow (2/5)\sigma^4. \quad (26)$$

Finally, we put everything together to obtain

$$\begin{aligned} \text{Var}(\tilde{D}_{J,r}(n)) &= \text{Var}(\beta_1 D(h_0; n) + \beta_2 D(h_0; rn) + \beta_3 \tilde{D}(h_0; (1-r)n)) \\ &= \beta_1^2 \text{Var}(D(h_0; n)) + \beta_2^2 \text{Var}(D(h_0; rn)) + \beta_3^2 \text{Var}(\tilde{D}(h_0; (1-r)n)) \\ &\quad + 2\beta_1\beta_2 \text{Cov}(D(h_0; n), D(h_0; rn)) \\ &\quad + 2\beta_2\beta_3 \text{Cov}(D(h_0; rn), \tilde{D}(h_0; (1-r)n)) \\ &\quad + 2\beta_1\beta_3 \text{Cov}(D(h_0; n), \tilde{D}(h_0; (1-r)n)), \end{aligned} \quad (27)$$

and the result follows by Equations (21)–(26). \square

Example 5 *We recommend the choice of $r = 1/2$, which leads to the minimum possible asymptotic variance. If $r = 1/2$, then as $n \rightarrow \infty$, $\text{Var}(\tilde{D}_{J,1/2}(n)) \rightarrow 1.2\sigma^4 < 1.4\sigma^4$, which is the asymptotic variance of JDW. Further, if $r = 1/2$, then $\beta_1 = 2, \beta_2 = -1/2$, and $\beta_3 = -1/2$, and the estimator takes the intuitive form*

$$\tilde{D}_{J,1/2}(n) = 2D(h_0; n) - \frac{1}{2}D(h_0; n/2) - \frac{1}{2}\tilde{D}(h_0; n/2).$$

2.3 *Batching*

In Section 2.3.1, we show how batched versions of the STS estimators can be obtained, and in Section 2.3.2 and Section 2.3.3 we review the NBM and OBM variance parameter estimators, respectively.

2.3.1 Batched STS Estimators

The technique of batching is a way to reduce the variance of a variance estimator — although at the cost of a possible increase in the bias. When we have a long run of n observations, we often divide it into b nonoverlapping batches, each of size m (such that $n = bm$). In particular, batch i consists of observations $Y_{(i-1)m+1}, \dots, Y_{im}$, $i = 1, 2, \dots, b$. Then we form an estimator from each batch of size m individually and take the sample average of the estimators to obtain a “batched” estimator for the variance parameter.

With this motivation in mind, let $A_i(f; b, m)$, $C_i(g; b, m)$, $D_i(h; b, m)$, $D_{J,r,i}(b, m)$, and $\tilde{D}_{J,r,i}(b, m)$ be the area, CvM, DW, JDW, and MJDW estimators applied to the i th batch of size m , respectively. We take the sample averages over the b batches to obtain

$$\begin{aligned} A(f; b, m) &\equiv \frac{1}{b} \sum_{i=1}^b A_i(f; b, m), \\ C(g; b, m) &\equiv \frac{1}{b} \sum_{i=1}^b C_i(g; b, m), \\ D(h; b, m) &\equiv \frac{1}{b} \sum_{i=1}^b D_i(h; b, m), \\ D_{J,r}(b, m) &\equiv \frac{1}{b} \sum_{i=1}^b D_{J,r,i}(b, m), \\ \tilde{D}_{J,r}(b, m) &\equiv \frac{1}{b} \sum_{i=1}^b \tilde{D}_{J,r,i}(b, m), \end{aligned}$$

the batched area, CvM, DW, JDW, and MJDW estimators, respectively.

The expected values of the batched area estimators with the weighting functions $f_0(t)$, $f_2(t)$, and $f_{\cos,j}(t)$, $j = 1, 2, \dots$, for all $t \in [0, 1]$ are

$$\begin{aligned} E[A(f_0; b, m)] &= \sigma^2 + \frac{3\gamma}{m} + o(1/m), \\ E[A(f_2; b, m)] &= \sigma^2 + o(1/m), \\ E[A(f_{\cos,j}; b, m)] &= \sigma^2 + o(1/m). \end{aligned}$$

The expected values of the batched CvM estimators with the weighting functions $g_0(t)$, $g_2^*(t)$, and $g_4^*(t)$ for all $t \in [0, 1]$ are

$$E[C(g_0; b, m)] = \sigma^2 + \frac{5\gamma}{m} + o(1/m),$$

$$E[C(g_2^*; b, m)] = \sigma^2 + o(1/m),$$

$$E[C(g_4^*; b, m)] = \sigma^2 + o(1/m).$$

The expected value of the batched DW estimator with the weighting function $h_0(t)$ for all $t \in [0, 1]$ is

$$E[D(h_0; b, m)] = \sigma^2 + \frac{7\gamma}{m} + o(1/m).$$

The expected value of the batched JDW estimator when $r = 1/2$ is

$$E[D_{J,1/2}(b, m)] = \sigma^2 + o(1/m).$$

The expected value of the batched MJDW estimator when $r = 1/2$ is

$$E[\tilde{D}_{J,1/2}(b, m)] = \sigma^2 + o(1/m).$$

The only difference compared to the expected values of the corresponding unbatched estimators is the use of m instead of n , resulting in an increase in the bias (see Examples 1, 2, 3 and Equations (8) and (14)).

Under suitable moment and mixing conditions, for large enough batch size m , we can make the assumption that the estimators from two different batches are approximately independent. The asymptotic variances of the batched area estimators with the weighting functions $f_0(t)$, $f_2(t)$, and $f_{\cos,j}(t)$, $j = 1, 2, \dots$, for all $t \in [0, 1]$ as $m \rightarrow \infty$ are

$$\begin{aligned} \text{Var}(A(f_0; b, m)) &\rightarrow \frac{2\sigma^4}{b}, \\ \text{Var}(A(f_2; b, m)) &\rightarrow \frac{2\sigma^4}{b}, \\ \text{Var}(A(f_{\cos,j}; b, m)) &\rightarrow \frac{2\sigma^4}{b}. \end{aligned}$$

The asymptotic variances of the batched CvM estimators with the weighting functions $g_0(t)$, $g_2^*(t)$, and $g_4^*(t)$ for all $t \in [0, 1]$ as $m \rightarrow \infty$ are

$$\begin{aligned} \text{Var}(C(g_0; b, m)) &\rightarrow \frac{0.8\sigma^4}{b}, \\ \text{Var}(C(g_2^*; b, m)) &\rightarrow \frac{1.729\sigma^4}{b}, \\ \text{Var}(C(g_4^*; b, m)) &\rightarrow \frac{1.042\sigma^4}{b}. \end{aligned}$$

The asymptotic variance of the batched DW estimator with the weighting function $h_0(t)$ for all $t \in [0, 1]$ as $m \rightarrow \infty$ is

$$\text{Var}(D(h_0; b, m)) \rightarrow \frac{0.4\sigma^4}{b}.$$

The asymptotic variance of the batched JDW estimator when $r = 1/2$ as $m \rightarrow \infty$ is

$$\text{Var}(D_{J,1/2}(b, m)) \rightarrow \frac{1.4\sigma^4}{b}.$$

The asymptotic variance of the batched MJDW estimator when $r = 1/2$ as $m \rightarrow \infty$ is

$$\text{Var}(\tilde{D}_{J,1/2}(b, m)) \rightarrow \frac{1.2\sigma^4}{b}.$$

Hence, the batched versions of the estimators have asymptotic variances approximately $(1/b)$ times the asymptotic variances of the analogous unbatched versions (see Equation (2) and Examples 2, 3, 4, and 5).

2.3.2 Nonoverlapping Batch Means Estimator

The *nonoverlapping batch means* (NBM) estimator for σ^2 is

$$\mathcal{N}(b, m) \equiv \frac{m}{b-1} \sum_{i=1}^b (\bar{Y}_{i,m} - \bar{Y}_n)^2,$$

where $\bar{Y}_{i,m} \equiv \sum_{j=1}^m Y_{(i-1)m+j}/m$, $i = 1, 2, \dots, b$, are the batch means of the process. These batch means are assumed to be IID normal random variables, at least for large enough batch size m . Under mild conditions, $\mathcal{N}(b, m) \xrightarrow{\mathcal{D}} \sigma^2 \chi_{b-1}^2 / (b-1)$ as $m \rightarrow \infty$ with b fixed, where χ_ν^2 denotes the chi-square distribution with ν degrees of freedom. In addition, by [12], [18], and [35], we have

$$\mathbb{E}[\mathcal{N}(b, m)] = \sigma^2 + \frac{\gamma(b+1)}{bm} + o(1/m)$$

and as $m \rightarrow \infty$,

$$\text{Var}(\mathcal{N}(b, m)) \rightarrow \frac{2\sigma^4}{b-1}.$$

2.3.3 Overlapping Batch Means Estimator

The *overlapping batch means* (OBM) estimator for σ^2 is

$$\mathcal{O}(b, m) \equiv \frac{nm}{(n-m+1)(n-m)} \sum_{i=1}^{n-m+1} (\bar{Y}_{i,m}^{\mathcal{O}} - \bar{Y}_n)^2,$$

where $\bar{Y}_{i,m}^{\mathcal{O}} \equiv \sum_{k=0}^{m-1} Y_{i+k}/m$, $i = 1, 2, \dots, n-m+1$, are the i th overlapping batch means ([28]). Under mild conditions, for large b ,

$$\mathbb{E}[\mathcal{O}(b, m)] \doteq \sigma^2 + \frac{\gamma}{m} + o(1/m),$$

and as $m \rightarrow \infty$,

$$\text{Var}(\mathcal{O}(b, m)) \rightarrow \frac{4\sigma^4}{3b}.$$

2.3.4 Recapitulation

All of the variance estimators examined herein are asymptotically unbiased as the batch size increases. However, the NBM and OBM estimators tend to have more bias than certain first-order unbiased STS estimators — e.g., $A(f_2; b, m)$, $C(g_2^*; b, m)$, and our new estimator $\tilde{D}_{J,1/2}(b, m)$. Further, for comparable choices of b and m , the asymptotic variance of $\tilde{D}_{J,1/2}(b, m)$ is less than that of $\mathcal{N}(b, m)$ or $\mathcal{O}(b, m)$.

2.4 Rebatching

In this section, we show how rebatched versions of the STS estimators can be obtained and present their statistical properties.

Rebatching is a technique to further reduce the bias and variance of a batched estimator of a variance parameter of a steady-state simulation process. It is a different way of batching the observations. In “standard” batching, we divide n observations into b nonoverlapping batches, each with size m . Then we form an estimator from each batch of size m individually and take the sample average of the estimators to obtain a “batched” estimator for the variance parameter. As we decrease the batch size and increase the number of batches, we obtain a more-biased but less variable variance estimator. In the rebatching technique, we combine variance estimators computed from batches of different sizes. For example, if

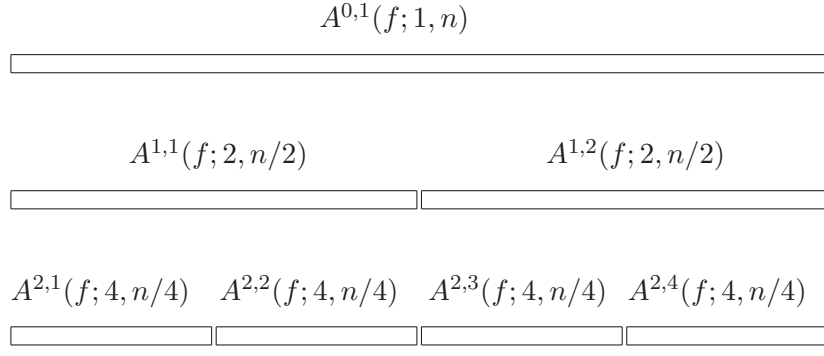


Figure 2: The Area Estimator Based on Different Portions of a Sample of Size n when the Number of Rebatching Levels is Two

$b = 4$, we take the sample average of the estimators obtained from one batch of size n , from two nonoverlapping batches of size $n/2$, and from four nonoverlapping batches of size $n/4$. The resulting rebatched variance estimator often has smaller bias and variance than the batched version based solely on $b = 4$ batches. This technique was first presented in [3] and applied to the STS area and CvM estimators. In this study, we apply the rebatching technique on the DW, JDW, and MJDW estimators, as well.

For notational convenience, we assume that we have a sample of $n = 2^k m$ observations for some integer $k \geq 0$. Let $A^{i,j}(f; 2^i, n/2^i)$, $C^{i,j}(g; 2^i, n/2^i)$, $D^{i,j}(h; 2^i, n/2^i)$, $D_{J,r}^{i,j}(2^i, n/2^i)$, and $\tilde{D}_{J,r}^{i,j}(2^i, n/2^i)$, for $i = 0, 1, \dots, k$ and $j = 1, 2, \dots, 2^i$, be the area, CvM, DW, JDW, and MJDW estimators applied to the j th nonoverlapping batch of size $n/2^i$ at the i th level of rebatching, respectively. At the i th level of rebatching, we form 2^i batches each with size $n/2^i$ from the sample of size n . For example, Figure 2 illustrates seven area estimators $A^{0,1}(f; 1, n)$, $A^{1,1}(f; 2, n/2)$, $A^{1,2}(f; 2, n/2)$, $A^{2,1}(f; 4, n/4)$, $A^{2,2}(f; 4, n/4)$, $A^{2,3}(f; 4, n/4)$, and $A^{2,4}(f; 4, n/4)$, as they are applied to different portions of a sample when the number of rebatching levels is two.

When we take the sample average of these estimators obtained from different portions of a sample of size n and the number of rebatching levels is k , we obtain

$$\begin{aligned}
 A^R(f; n, k) &\equiv \frac{\sum_{i=0}^k \sum_{j=1}^{2^i} A^{i,j}(f; 2^i, n/2^i)}{2^{k+1} - 1}, \\
 C^R(g; n, k) &\equiv \frac{\sum_{i=0}^k \sum_{j=1}^{2^i} C^{i,j}(g; 2^i, n/2^i)}{2^{k+1} - 1}, \\
 D^R(h; n, k) &\equiv \frac{\sum_{i=0}^k \sum_{j=1}^{2^i} D^{i,j}(h; 2^i, n/2^i)}{2^{k+1} - 1},
 \end{aligned}$$

$$D_{J,r}^R(n, k) \equiv \frac{\sum_{i=0}^k \sum_{j=1}^{2^i} D_{J,r}^{i,j}(2^i, n/2^i)}{2^{k+1} - 1},$$

$$\tilde{D}_{J,r}^R(n, k) \equiv \frac{\sum_{i=0}^k \sum_{j=1}^{2^i} \tilde{D}_{J,r}^{i,j}(2^i, n/2^i)}{2^{k+1} - 1},$$

the rebatched area, CvM, DW, JDW, and MJDW estimators, respectively. In [3], it is shown that the rebatched estimators, $A^R(f; n, k)$ and $C^R(g; n, k)$, have smaller bias and variance than the respective batched versions, $A(f; b, m)$ and $C(g; b, m)$, when $n = bm = 2^k m$. In this study, we compare the batched and rebatched versions of the area, CvM, JDW, and MJDW estimators. Here, we will consider the particular case of $k = 2$ in rebatching and consequently $b = 4$ in “standard” batching such that $n = 4m$. However, the results can be generalized to larger k and b values.

The expected values of the rebatched area estimators with the weighting functions $f_0(t)$, $f_2(t)$, and $f_{\cos,j}(t)$, $j = 1, 2, \dots$, for all $t \in [0, 1]$ when $b = 4$ are

$$\begin{aligned} E[A^R(f_0; n, 2)] &= \sigma^2 + \frac{3}{4} \frac{3\gamma}{m} + o(1/m), \\ E[A^R(f_2; n, 2)] &= \sigma^2 + o(1/m), \\ E[A^R(f_{\cos,j}; n, 2)] &= \sigma^2 + o(1/m). \end{aligned}$$

The expected values of the rebatched CvM estimators with the weighting functions $g_0(t)$, $g_2^*(t)$, and $g_4^*(t)$ for all $t \in [0, 1]$ when $b = 4$ are

$$\begin{aligned} E[C^R(g_0; n, 2)] &= \sigma^2 + \frac{3}{4} \frac{5\gamma}{m} + o(1/m), \\ E[C^R(g_2^*; n, 2)] &= \sigma^2 + o(1/m), \\ E[C^R(g_4^*; n, 2)] &= \sigma^2 + o(1/m). \end{aligned}$$

The expected value of the rebatched DW estimator with the weighting function $h_0(t)$ for all $t \in [0, 1]$ when $b = 4$ is

$$E[D^R(h_0; n, 2)] = \sigma^2 + \frac{3}{4} \frac{7\gamma}{m} + o(1/m). \quad (28)$$

The expected value of the rebatched JDW estimator when $r = 1/2$ and $b = 4$ is

$$E[D_{J,1/2}^R(n, 2)] = \sigma^2 + o(1/m). \quad (29)$$

The expected value of the rebatched MJDW estimator when $r = 1/2$ and $b = 4$ is

$$E[\tilde{D}_{J,1/2}^R(n, 2)] = \sigma^2 + o(1/m). \quad (30)$$

When we compare these expected values with respect to the expected values of their batched versions in Section 2.3.1, we see that the rebatched estimators, except the ones that are first-order unbiased, are less biased than their batched versions.

The asymptotic variances of the rebatched area estimators with the weighting functions $f_0(t)$, $f_2(t)$, and $f_{\cos,j}(t)$, $j = 1, 2, \dots$, for all $t \in [0, 1]$ and the corresponding asymptotic variances of the batched area estimators when $b = 4$ as $m \rightarrow \infty$ are

$$\begin{aligned} \text{Var}(A^R(f_0; n, 2)) &\rightarrow 0.3520\sigma^4 < 0.5\sigma^4, \\ \text{Var}(A^R(f_2; n, 2)) &\rightarrow 0.2983\sigma^4 < 0.5\sigma^4, \\ \text{Var}(A^R(f_{\cos,j}; n, 2)) &\rightarrow 0.2951\sigma^4 < 0.5\sigma^4. \end{aligned}$$

The asymptotic variances of the rebatched CvM estimators with the weighting functions $g_0(t)$, $g_2^*(t)$, and $g_4^*(t)$ for all $t \in [0, 1]$ and the corresponding asymptotic variances of the batched CvM estimators when $b = 4$ as $m \rightarrow \infty$ are

$$\begin{aligned} \text{Var}(C^R(g_0; n, 2)) &\rightarrow 0.1714\sigma^4 < 0.2\sigma^4, \\ \text{Var}(C^R(g_2^*; n, 2)) &\rightarrow 0.2846\sigma^4 < 0.4322\sigma^4, \\ \text{Var}(C^R(g_4^*; n, 2)) &\rightarrow 0.2290\sigma^4 < 0.2605\sigma^4. \end{aligned}$$

The asymptotic variance of the rebatched DW estimator with the weighting function $h_0(t)$ for all $t \in [0, 1]$ and the corresponding asymptotic variance of the batched DW estimator when $b = 4$ as $m \rightarrow \infty$ are

$$\text{Var}(D^R(h_0; n, 2)) \rightarrow 0.1010\sigma^4 > 0.1\sigma^4. \quad (31)$$

The asymptotic variance of the rebatched JDW estimator when $r = 1/2$ and the corresponding asymptotic variance of the batched JDW estimator when $b = 4$ as $m \rightarrow \infty$ are

$$\text{Var}(D_{J,1/2}^R(n, 2)) \rightarrow 0.2498\sigma^4 < 0.35\sigma^4. \quad (32)$$

The asymptotic variance of the rebatched MJDW estimator when $r = 1/2$ and the corresponding asymptotic variance of the batched MJDW estimator when $b = 4$ as $m \rightarrow \infty$ are

$$\text{Var}(\tilde{D}_{J,1/2}^R(n, 2)) \rightarrow 0.2212\sigma^4 < 0.3\sigma^4. \quad (33)$$

The derivations for the rebatched area and CvM estimators are given in [3] and the derivations for the rebatched DW, JDW, and MJDW are given in Appendix A.

We see that all rebatched estimators, except the rebatched DW estimator, are asymptotically less variable than their batched versions. For the rebatched DW estimator, we observe a slight increase in variance with respect to its batched version. We would not observe this increase if we were using optimal weights — minimizing the asymptotic variance of the rebatched estimator — instead of equal weights in rebatching. We prefer to use equal weights because the difference in the asymptotic variance of the rebatched estimator with equal weights and optimal weights is small. For example, the asymptotic variance of the rebatched DW estimator with optimal weights — $0.09186\sigma^4$ — is slightly less than the variance of the estimator with equal weights — $0.1010\sigma^4$.

Both rebatched and batched versions of a variance estimator require the same number of observations. However, a rebatched estimator reuses the same data by forming batches of different sizes. This requires slightly more computational time but the resulting rebatched estimator is often less biased and less variable than its batched version.

2.5 Examples

In this section, we illustrate the performance of the new estimator on exact and Monte Carlo examples. Section 2.5.1 gives analytical results on a moving average process, showing that the MJDW estimator performs in accordance with the theory. Section 2.5.2 compares the performance of the new estimator with that of its competitors on an autoregressive process.

2.5.1 Exact Example: Moving Average Process

A first-order moving average [MA(1)] process is defined as $Y_i = \theta\epsilon_{i-1} + \epsilon_i$, $i \geq 1$, where the ϵ_i 's are IID $\text{Nor}(0, 1)$ random variables. The MA(1) process has covariance function

$R_0 = 1 + \theta^2$, $R_{\pm 1} = \theta$, $R_k = 0$ elsewhere, variance parameter $\sigma^2 = (1 + \theta)^2$, and constant $\gamma = -2\theta$. We next calculate the expected value and asymptotic variance of the MJDW estimator.

We know from [15] that for the MA(1),

$$E[D(h_0; n)] = \sigma^2 + \frac{7\gamma}{n} - \frac{\sigma^2 + 12\gamma}{n^2} + \frac{5\gamma}{n^3}. \quad (34)$$

Plugging Equation (34) into Equation (16) and then inserting Equations (11)–(13), we obtain

$$\begin{aligned} E[\tilde{D}_{J,r}(n)] &= \beta_1 \left(\sigma^2 + \frac{7\gamma}{n} - \frac{\sigma^2 + 12\gamma}{n^2} + \frac{5\gamma}{n^3} \right) + \beta_2 \left(\sigma^2 + \frac{7\gamma}{rn} - \frac{\sigma^2 + 12\gamma}{r^2 n^2} + \frac{5\gamma}{r^3 n^3} \right) \\ &\quad + \beta_3 \left(\sigma^2 + \frac{7\gamma}{(1-r)n} - \frac{\sigma^2 + 12\gamma}{(1-r)^2 n^2} + \frac{5\gamma}{(1-r)^3 n^3} \right) \\ &= \sigma^2 - \left(\frac{8r^4 - 16r^3 + 3r^2 + 5r - 2}{2r(4r^4 - 8r^3 + 2r^2 + 2r - 1)(r - 1)} \right) \left(\frac{\sigma^2 + 12\gamma}{n^2} \right) \\ &\quad + \left(\frac{-9r^4 + 18r^3 - 4r^2 - 5r + 2}{2r^2(4r^4 - 8r^3 + 2r^2 + 2r - 1)(r - 1)^2} \right) \left(\frac{5\gamma}{n^3} \right) \end{aligned}$$

which shows that $\tilde{D}_{J,r}(n)$ is first-order unbiased for σ^2 , as established in Theorem 5.

We also know from [15] that for the MA(1),

$$\text{Var}(D(h_0; n)) = \frac{2\sigma^4}{5} + \frac{4\gamma\sigma^2}{5n} + O(n^{-2}), \quad (35)$$

so that

$$\text{Var}(D(h_0; rn)) = \frac{2\sigma^4}{5} + \frac{4\gamma\sigma^2}{5rn} + O(n^{-2}) \quad \text{and} \quad (36)$$

$$\text{Var}(\tilde{D}(h_0; (1-r)n)) = \frac{2\sigma^4}{5} + \frac{4\gamma\sigma^2}{5(1-r)n} + O(n^{-2}). \quad (37)$$

In Appendix B, we show that

$$\text{Cov}(D(h_0; n), D(h_0; rn)) = \frac{2r^2(2-r)\sigma^4}{5} + \frac{2r^2\gamma\sigma^2}{5n} + O(n^{-2}), \quad (38)$$

$$\text{Cov}(D(h_0; n), \tilde{D}(h_0; (1-r)n)) = \frac{2(1-r)^2(1+r)\sigma^4}{5} + \frac{2(1-r)^2\gamma\sigma^2}{5n} + O(n^{-2}), \quad (39)$$

$$\text{Cov}(D(h_0; rn), \tilde{D}(h_0; (1-r)n)) = \frac{\gamma^2}{2n^2 r(1-r)} + O(n^{-3}). \quad (40)$$

Starting with Equation (27) and plugging in Equations (35)–(40), we obtain

$$\begin{aligned} \text{Var}(\tilde{D}_{J,r}(n)) &= \frac{2}{5}\sigma^4 \left(\beta_1^2 + \beta_2^2 + \beta_3^2 + 2\beta_1\beta_2r^2(2-r) + 2\beta_1\beta_3(1-r)^2(1+r) \right) \\ &\quad + \frac{4\gamma\sigma^2}{5n} \left(\beta_1^2 + \frac{\beta_2^2}{r} + \frac{\beta_3^2}{1-r} + \beta_1\beta_2r^2 + \beta_1\beta_3(1-r)^2 \right) + O(n^{-2}). \end{aligned}$$

Table 1: Estimated Mean and Variance of the Area, CvM, DW, JDW, and MJDW Estimators (Based on One Batch) for the Variance Parameter ($\sigma^2 = 19$) of an AR(1) Process with $\phi = 0.9$

Estimator	$n = 256$		$n = 512$		$n = 1024$		$n = 2048$		True Var ($n \rightarrow \infty$)
	\hat{E}	$\widehat{\text{Var}}$	\hat{E}	$\widehat{\text{Var}}$	\hat{E}	$\widehat{\text{Var}}$	\hat{E}	$\widehat{\text{Var}}$	
$A(f_0; n)$	16.89	570	17.92	643	18.44	674	18.71	694	722
$A(f_2; n)$	18.09	650	18.84	712	18.86	709	19.00	712	722
$A(f_{\cos,1}; n)$	15.79	503	17.89	635	18.73	704	19.02	734	722
$C(g_0; n)$	15.77	239	17.28	266	18.10	275	18.56	280	289
$C(g_2^*; n)$	18.03	540	18.78	597	18.88	602	18.97	605	624
$C(g_4^*; n)$	17.10	303	18.25	337	18.77	358	18.95	368	376
$D(h_0; n)$	14.66	118	16.64	134	17.77	141	18.41	143	144
$D_{J,0.5}(n)$	17.95	400	18.69	462	18.89	487	19.01	496	505
$\tilde{D}_{J,0.5}(n)$	17.96	358	18.69	405	18.88	421	19.02	426	433

When we substitute in the β_j 's from Equations (11)–(13) and let $n \rightarrow \infty$, we obtain the same asymptotic variance result as in Theorem 5.

2.5.2 Monte Carlo Example: Autoregressive Process

A first-order autoregressive [AR(1)] process is defined as $Y_i = \phi Y_{i-1} + \epsilon_i$, $i \geq 1$, where the ϵ_i 's are IID $\text{Nor}(0, 1 - \phi^2)$ random variables, and Y_0 is a $\text{Nor}(0, 1)$ random variable initialized independently of the others. The AR(1) process has covariance function $R_k = \phi^{|k|}$ for all k , variance parameter $\sigma^2 = (1 + \phi)/(1 - \phi)$, and constant $\gamma = -2\phi/(1 - \phi)^2$.

In this example, we set $\phi = 0.9$, which gives a highly positively autocorrelated process with variance parameter $\sigma^2 = 19$. We ran 100,000 independent replications of the process for each of sample sizes $n = 256, 512, 1024$, and 2048. For each sample size, we computed the area, CvM, DW, JDW, and MJDW estimators. To estimate the expectation and asymptotic variance of each of these estimators, we recorded the sample mean and sample variance from each of the 100,000 replications. The estimated expectation (\hat{E}) and variance ($\widehat{\text{Var}}$) values for each of the estimators are given in Table 1. With 100,000 replications, the \hat{E} values are valid up to the second decimal digit. The right-most column of the table, denoted “True Var ($n \rightarrow \infty$)”, is simply the list of asymptotic variances of the various estimators obtained from Theorem 1 and Examples 2–5.

Table 2: Estimated Mean and Variance of the Batched Area, CvM, JDW, MJDW, NBM, and OBM Estimators for the Variance Parameter ($\sigma^2 = 19$) of an AR(1) Process with $\phi = 0.9$ and $n = 4096$ [16384]

Estimator	$b = 4$		$b = 8$		$b = 16$	
	\hat{E}	$\widehat{\text{Var}}$	\hat{E}	$\widehat{\text{Var}}$	\hat{E}	$\widehat{\text{Var}}$
$A(f_2; b, m)$	18.84 [18.98]	176 [180]	18.76 [18.96]	89 [90]	18.13 [18.92]	41 [44]
$C(g_2^*; b, m)$	18.86 [18.99]	150 [155]	18.73 [18.96]	75 [77]	18.06 [18.93]	34 [38]
$D(h_0; b, m)$	17.79 [18.69]	35 [36]	16.69 [18.39]	17 [18]	14.68 [17.81]	7 [9]
$D_{J,0.5}(b, m)$	18.88 [19.01]	122 [124]	18.71 [18.98]	58 [62]	17.96 [18.93]	25 [30]
$\tilde{D}_{J,0.5}(b, m)$	18.88 [18.99]	104 [107]	18.70 [18.97]	51 [53]	17.96 [18.93]	23 [26]
$\mathcal{N}(b, m)$	18.80 [19.00]	236 [240]	18.58 [18.92]	98 [102]	18.24 [18.82]	45 [47]
$\mathcal{O}(b, m)$	18.76 [18.98]	153 [152]	18.57 [18.92]	70 [69]	18.23 [18.82]	32 [32]

The table shows that, as the sample size n becomes large, the estimators — based on one large batch of size n — become less biased. Further, as n becomes large, all of the estimated variances approach their true asymptotic values from below. This makes sense because the AR(1) process with $\phi = 0.9$ is highly positively correlated, and when n is small, the variances are underestimated.

A good estimator has to have both small bias and small variance. Hence, we evaluate the performance of our estimators based on these two criteria. We did not consider MSE (mean square error) — the sum of variance and the square of bias — because MSE can sometimes be misleading. For example, when we have two estimators: (1) one with very low variance but very high bias and (2) one with larger variance but low bias, it is possible that the first one gives smaller MSE than the second one although the second estimator is better than the first in some practical sense.

In Table 1, we observe that $D(h_0; n)$ and $C(g_0; n)$ are the two least variable estimators, but unfortunately, they are also the most biased ones. Among the rest, $C(g_4^*; n)$ has the smallest variance, with $\tilde{D}_{J,0.5}(n)$ following close behind. Both $\tilde{D}_{J,0.5}(n)$ and $C(g_4^*; n)$ are first-order unbiased estimators, although the AR(1) empirical results suggest that $\tilde{D}_{J,0.5}(n)$ may be a bit less biased than $C(g_4^*; n)$ for small sample sizes.

Now, we will look at the case of batched estimators. Table 2 shows the estimated mean and variances of $A(f_2; b, m)$, $C(g_2^*; b, m)$, $D(h_0; b, m)$, $D_{J,0.5}(b, m)$, $\tilde{D}_{J,0.5}(b, m)$, $\mathcal{N}(b, m)$,

Table 3: Estimated Mean and Variance of the Rebatched Area, CvM, JDW, and MJDW Estimators for the Variance Parameter ($\sigma^2 = 19$) of an AR(1) Process with $\phi = 0.9$ and $n = 4096$ [16384]

Estimator	$b = 4$ ($k = 2$)		$b = 8$ ($k = 3$)		$b = 16$ ($k = 4$)	
	\hat{E}	$\widehat{\text{Var}}$	\hat{E}	$\widehat{\text{Var}}$	\hat{E}	$\widehat{\text{Var}}$
$A^R(f_2; n, k)$	18.95 [18.93]	107 [106]	18.82 [18.98]	50 [51]	18.47 [18.95]	24 [25]
$C^R(g_2^*; n, k)$	18.94 [18.95]	101 [101]	18.81 [18.97]	46 [48]	18.43 [18.95]	22 [24]
$D^R(h_0; n, k)$	18.09 [18.75]	36 [36]	17.34 [18.56]	18 [18]	15.97 [18.18]	8 [9]
$D_{J,0.5}^R(n, k)$	18.96 [18.96]	89 [89]	18.83 [19.00]	41 [43]	18.37 [18.95]	19 [21]
$\tilde{D}_{J,0.5}^R(n, k)$	18.94 [18.96]	79 [79]	18.82 [18.99]	37 [38]	18.37 [18.95]	17 [19]

and $\mathcal{O}(b, m)$ when the total number of observations is 4096 or 16384 and $b = n/m = 4, 8, \text{ or } 16$. All results are again based on 100,000 independent replications of the process. For example, if $n = 16384$ and $b = 4$, then the batch size is 4096, the estimated expected value of $A(f_2; b, m)$ is 18.98, and the estimated variance is 180. We only give results for the area estimator with weighting function $f_2(t)$ and the CvM estimator with weighting function $g_2^*(t)$ because the batched versions of these two performed better in terms of bias in our small experiment than the batched versions of the area and CvM estimators with other weighting functions.

The results show that as b increases for a fixed sample size n , the variance of each estimator decreases and the bias increases. This is consistent with the theoretical results. Similar to the case of unbatched estimators, $D(h_0; b, m)$ is the estimator with the smallest variance but the largest bias. Among the rest, $\tilde{D}_{J,0.5}(b, m)$ has a significantly smaller variance than the others, yet maintains a competitively small bias.

Now, we will look at the case of rebatched estimators. Table 3 shows the estimated mean and variances of $A^R(f_2; n, k)$, $C^R(g_2^*; n, k)$, $D^R(h_0; n, k)$, $D_{J,0.5}^R(n, k)$, and $\tilde{D}_{J,0.5}^R(n, k)$ when the total number of observations is 4096 or 16384 and $b = n/m = 4, 8, \text{ or } 16$. Again, we ran 100,000 independent replications of the process. When $b = 4, 8, \text{ or } 16$, the number of rebatching levels is 2, 3, or 4, respectively. When we compare these rebatched estimators with respect to their batched versions in Table 2, we observe a decrease in the variance of all estimators except the DW estimator. The results for $b = 4$ — when the number of

rebatching levels is two — match with the theoretical results, and the results for higher levels of rebatching are consistent with the $b = 4$ results as we conjectured. The rebatched estimators are also less biased than their batched versions. In particular, if they have first-order bias, the difference in the bias values is more visible. As a result of these experimental results for the rebatched estimators, we see that $\tilde{D}_{J,0.5}^R(n, k)$ is still the least variable one among the competitively low bias estimators.

2.6 Conclusion

In this chapter, we present a new variance parameter estimator for steady-state simulations — the MJDW estimator. The new estimator is based on the STS methodology and is an extension of the JDW estimator which is itself a combination of the STS unweighted area and CvM estimators. We compare the performance of the new estimator with the area, CvM, DW, and JDW estimators (the last two of which are from [15]). The comparison is done on the basis of bias and asymptotic variance. We show that the MJDW estimator has a competitive bias and lower asymptotic variance than many of the other estimators.

When the batched and rebatched versions of the area, CvM, DW, JDW, and MJDW, NBM, and OBM estimators are considered, the rebatched MJDW estimator still shows a competitive bias and lower variance than most of the other estimators.

CHAPTER III

FULLY SEQUENTIAL SELECTION PROCEDURES WITH PARABOLIC BOUNDARY

This chapter concerns selection-of-the-best problem. The aim is to choose the best system with the maximum or minimum expected performance measure from finitely many simulated systems when the variances across systems are unknown and unequal.

In this problem, we assume that basic observations are IID normal. This assumption can be satisfied when basic observations are within-replication averages in terminating or steady-state simulations, or when they are batch means in steady-state simulations that take a single-replication approach ([25]). One of the most recent statistically valid indifference-zone procedures for IID normal data is a fully sequential procedure, \mathcal{KN} , proposed by [22]. In [22], a fully sequential procedure is defined as the one that takes a single basic observation from each alternative that is still in play at the current stage of sampling and eliminates systems immediately when there is evidence that they are inferior. Fully sequential procedures have a boundary called a continuation region. Figure 3 shows two different types of continuation regions, \mathcal{T} and \mathcal{P} that have triangular and parabolic shapes, respectively. The horizontal axis in Figure 3 represents the stage number and the vertical axis represents the value of a monitoring statistic at each stage. As long as a monitoring statistic stays within the continuation region, sampling continues. When a monitoring statistic exits a continuation region, one system is eliminated depending on which direction — through the upper boundary or the lower boundary — the exit is made.

\mathcal{KN} is closely related to and derived from a hypothesis test on a drift of a Brownian motion process. In [26], it is stated that for the known drift of a Brownian motion, a triangular boundary is optimal in a sense that it minimizes a certain type of Bayes risk which is linear in the number of basic observations and quadratic in the magnitude of the drift. However, they show that for the unknown drift of a Brownian motion process,

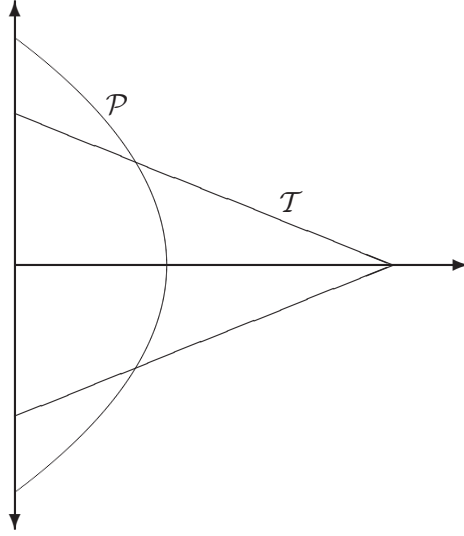


Figure 3: Triangular and Parabolic Continuation Regions

parabolic boundaries are optimal in terms of the expected value of the specific type of Bayes risk.

From this motivation, a new fully sequential procedure whose continuation region is parabolic for IID normal data, is proposed in [37]. However, the procedure assumes known variances which is unlikely to be true in practice. In this chapter, we propose two new fully sequential indifference-zone procedures for selection-of-the-best problem, one of which is a natural extension of [37]. Our procedures have parabolic continuation regions and assume unknown and unequal variances across systems. We compare the performance of the proposed procedures with that of \mathcal{KN} , which is one of the most efficient statistically valid indifference-zone procedures for IID data in terms of the number of observations required.

There exists another recent ranking and selection (R&S) procedure called $\mathcal{KN}++$ in [24]. This procedure updates variance estimates as more observations become available and is shown to be very effective in terms of the number of observations required. However, it is specifically designed for steady-state simulation that employs the single-replication design and takes individual observations — that are, at best, stationary and dependent — as basic observations. Although $\mathcal{KN}++$ is asymptotically valid for stationary and dependent data

and highly efficient, in [27], it is shown that $\mathcal{KN}++$ does not always satisfy the PCS (probability of correct selection) requirement for finite-sample size and is only heuristic even for IID normal data. Our interests in this chapter are on procedures whose validity is established for IID normal data. Hence, when we compare the performance of our procedures to other existing procedures, we do not consider $\mathcal{KN}++$ or other procedures that are designed for steady-state simulation.

The chapter is organized as follows: In Section 3.1, algorithms for the two new procedures are given and we prove that the procedures choose one of the good systems with at least a pre-specified PCS. In Section 3.2, we discuss implementation issues including the determination of the values of parameters for the new procedures. Section 3.3 provides analytical comparisons among \mathcal{KN} and the two proposed procedures. Section 3.4 compares the performance of the proposed procedures with that of \mathcal{KN} by empirical studies based on IID normal data, followed by a conclusion in Section 3.5.

3.1 *Procedures*

We design two fully sequential indifference-zone procedures. Both have parabolic continuation regions but with different parameters. We assume that there are k systems. The best system is defined as the one with the largest expected performance when the difference between the expected performances of the best and the second best is at least δ . Our procedures guarantee to select the best system with a PCS greater than or equal to $1 - \alpha$. If there are inferior systems whose means are within δ of the true best system, then those systems are called “good” systems and the procedures select one of these systems with at least $1 - \alpha$ PCS.

In the procedures, X_{ij} , $i = 1, 2, \dots, k$ and $j = 1, 2, \dots$, denotes the j th observation from the i th system and X_{ij} , $j = 1, 2, \dots$, are assumed to be IID normal with mean μ_i and variance σ_i^2 . The variances are unknown to the experimenter and may be unequal.

3.1.1 **First Procedure**

We present the first procedure that we call \mathcal{P}_1 .

Procedure \mathcal{P}_1

Setup: Choose nominal PCS $1 - \alpha$, indifference zone $\delta > 0$, and first stage sample size $n_0 \geq 2$. Then determine λ and ξ referring to **Parameters**.

Initialization: Let $I = \{1, 2, \dots, k\}$ be the initial set of systems.

Obtain observations $X_{ij}, j = 1, 2, \dots, n_0$, from each system i .

For all $i \neq \ell$ compute the sample variance of the difference between systems i and ℓ which is

$$S_{i\ell}^2 = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (X_{ij} - X_{\ell j} - [\bar{X}_i(n_0) - \bar{X}_\ell(n_0)])^2,$$

where $\bar{X}_i(n_0)$ is the sample average of the first n_0 observations from system i . Let

$$N_{i\ell} = \left\lfloor \frac{(n_0 - 1)^2 S_{i\ell}^4 \xi}{\delta^2} \right\rfloor,$$

where $\lfloor \cdot \rfloor$ indicates truncation of any fractional part. Then let

$$N_i = \max_{\ell \neq i} N_{i\ell}.$$

Here, $N_i + 1$ is the maximum number of observations that can be taken from system i .

Set the number of observations, r , equal to n_0 , and go to the next section.

Screening: Set $I^{\text{old}} = I$. Let

$$I = \left\{ i : i \in I^{\text{old}} \text{ and } \sum_{j=1}^r X_{ij} > \sum_{j=1}^r X_{\ell j} - R_{i\ell}(r), \forall \ell \in I^{\text{old}}, \ell \neq i \right\},$$

where

$$R_{i\ell}(r) = \lambda \sqrt{\max \left\{ \frac{(n_0 - 1)^2 S_{i\ell}^4 \xi}{\delta^2} - r, 0 \right\}}.$$

Stopping Rule: If $|I| = 1$, then stop and select the system $i \in I$ as the best. Otherwise, take one additional observation $X_{i,r+1}$ from each system $i \in I$, and set $r = r + 1$, and go to **Screening**.

(If the objective is to select a subset of size m containing the best system, then the stopping rule should be $|I| = m > 1$.)

Parameters: The parameter λ is any positive real number. For $1 - \alpha = 0.95$, we recommend taking $\lambda = 0.9268$ as given in [13]. Section 3.2.1 discusses the choice of λ for other values of α . For a given λ value, ξ is calculated as the solution to the equation

$$g(\xi; \lambda, n_0) = \beta, \quad (41)$$

where

$$g(\xi; \lambda, n_0) \equiv \mathbb{E} \left[\frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp \left\{ -\frac{\xi}{2} \left(\chi_{n_0-1}^2 \right)^2 \right\} \right. \\ \left. \times \sum_{n=0}^{\infty} \left(\frac{k_{2n+1}(\lambda)}{\ell_{2n+1}(\lambda)} \frac{2^n n!}{(2n+1)!} \left(\chi_{n_0-1}^2 \sqrt{\xi} \right)^{2n+1} \right) \right].$$

In this function, $k_n(\cdot)$ and $\ell_n(\cdot)$ are certain Hermite functions which are defined in Lemma 1 and $\chi_{n_0-1}^2$ is a chi-squared random variable with $n_0 - 1$ degrees of freedom. If common random numbers (CRN) are employed, we set $\beta = \alpha/(k-1)$ and if systems are simulated independently, set $\beta = 1 - (1 - \alpha)^{1/(k-1)}$. This equation does not have a closed form solution. We discuss how to determine the value of ξ in Section 3.2.2.

To prove that the procedure satisfies the PCS requirement, we need the following three lemmas:

Lemma 1 ([13]) *Let $\mathcal{W}(t, \Delta)$ be a Brownian motion process on $[0, +\infty)$ with a drift $\Delta > 0$. Consider the parabola $y = a\sqrt{s-t}$, $0 \leq t \leq s$, $a > 0$. Let R be $(-y, y)$ and T the first time that $\mathcal{W}(t, \Delta) \notin R$. That is,*

$$T = \inf\{t : 0 \leq t \leq s \text{ and } |\mathcal{W}(t, \Delta)| = a\sqrt{s-t}\}.$$

Then

$$\Pr\{\mathcal{W}(T, \Delta) < 0\} = \frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp\{-s\Delta^2/2\} \sum_{n=0}^{\infty} \left(\frac{k_{2n+1}(a)}{\ell_{2n+1}(a)} \frac{2^n n!}{(2n+1)!} (\Delta\sqrt{s})^{2n+1} \right),$$

where $k_n(\cdot)$ and $\ell_n(\cdot)$ are certain Hermite functions which are defined as

$$k_n(x) = \mathbb{E}[(Z+x)^n] \text{ and } l_n(x) = \mathbb{E}[(Z+x)^n \frac{1}{2} \text{sgn}(Z+x)]$$

for $n = 0, 1, \dots$, where Z is a standard $N(0, 1)$ random variable.

Let $\phi(x)$ and $\Phi(x)$ be the pdf and cdf of $N(0,1)$. Then one can get the recurrence relations

$$\begin{aligned} k_{n+1}(x) &= xk_n(x) + nk_{n-1}(x), \\ \ell_{n+1}(x) &= x\ell_n(x) + n\ell_{n-1}(x), \quad n \geq 1, \end{aligned}$$

where

$$k_0(x) = 1, \ell_0(x) = \Phi(x) - \frac{1}{2}, k_1(x) = x, \text{ and } \ell_1(x) = x\left(\Phi(x) - \frac{1}{2}\right) + \phi(x).$$

Lemma 2 ([21]) Suppose that a continuation region R is $(-g(t), g(t))$ given by a non-negative function $g(t), t \geq 0$. Consider two processes: a continuous process $\mathcal{W}(t, \Delta), t \geq 0$, with $\Delta > 0$ and a discrete process obtained by observing $\mathcal{W}(t, \Delta)$ at a random, increasing sequence of times $\{t_i : i = 1, 2, \dots\}$ taking values in a given countable set. Let $\tau_C = \inf\{t > 0 : \mathcal{W}(t, \Delta) \notin R\}$ and $\tau_D = \inf\{t_i : \mathcal{W}(t_i, \Delta) \notin R\}$ and assume that $\tau_D < \infty$ almost surely. Note that $\tau_D \geq \tau_C$. The error probabilities are

$$\Pr\{\mathcal{E}_C\} \equiv \Pr\{\mathcal{W}(\tau_C, \Delta) \leq -g(\tau_C)\} = \Pr\{\mathcal{W}(\tau_C, \Delta) < 0\},$$

$$\Pr\{\mathcal{E}_D\} \equiv \Pr\{\mathcal{W}(\tau_D, \Delta) \leq -g(\tau_D)\} = \Pr\{\mathcal{W}(\tau_D, \Delta) < 0\}.$$

Consider an outcome $\{(b(t); t \geq 0), \{t_i\}\}$ where $b(t)$ is the path of a Brownian motion. The conditional distribution of $\{t_i\}$ given $\mathcal{W}(t, \Delta) = b(t), t \geq 0$, is the same as that given $\mathcal{W}(t, \Delta) = -b(t), t \geq 0$. Under these conditions,

$$\Pr\{\mathcal{E}_D\} \leq \Pr\{\mathcal{E}_C\}.$$

Lemma 1 gives the probability of incorrect selection for a continuous Brownian motion process with a drift. However, in our procedure we only observe the process at integer times. When each observation is IID normal, the partial sums of the differences behave like Brownian motion process with drift at each integer point, but Brownian motion with drift is still only an approximation for our discrete process. However, Lemma 2 states that under very general conditions, the probability of incorrect selection does not increase when the Brownian motion process is observed at discrete times compared to the case where the

process is observed continuously; thus, procedures designed for continuous Brownian motion process with a drift provide an upper bound on the probability of incorrect selection for a discrete process.

Lemma 3 ([36]) *Let V_1, V_2, \dots, V_k be independent random variables, and let $g_j(v_1, v_2, \dots, v_k)$, $j = 1, 2, \dots, p$, be nonnegative, real-valued functions, each one nondecreasing in each of its arguments. Then*

$$\mathbb{E}\left[\prod_{j=1}^p g_j(V_1, V_2, \dots, V_k)\right] \geq \prod_{j=1}^p \mathbb{E}[g_j(V_1, V_2, \dots, V_k)].$$

Without loss of generality, we assume that $\mu_1 \geq \mu_2 \geq \dots \geq \mu_k$. Now, we present the main result.

Theorem 6 *Let $\mathbf{X}_j = (X_{1j}, X_{2j}, \dots, X_{kj})'$, $j = 1, 2, \dots$, be vectors of observations across all k systems. Suppose that $\mathbf{X}_1, \mathbf{X}_2, \dots$ are distributed IID multivariate normal with mean vector μ and covariance matrix Σ , where μ is an unknown vector with the property $\mu_1 \geq \mu_2 + \delta$ and Σ is an unknown, positive definite matrix. Then \mathcal{P}_1 selects system 1 with probability greater than or equal to $1 - \alpha$.*

Proof: Consider two systems, 1 and i , such that $\mu_1 \geq \mu_i + \delta$. Select a value of ξ such that $g(\xi; \lambda, n_0) = \beta$ for some $0 < \beta < 1/2$. Let

$$T = \min\left\{r : r \geq n_0 \text{ and } -R_{1i}(r) < \sum_{j=1}^r (X_{1j} - X_{ij}) < R_{1i}(r) \text{ is violated}\right\}.$$

Notice that T is the stage at which the procedure terminates. Let ICS_i be the event of incorrect selection when only two systems, 1 and i , are considered. Then

$$\begin{aligned} \Pr\{\text{ICS}_i\} &= \Pr\left\{\sum_{j=1}^T (X_{1j} - X_{ij}) < -\lambda \sqrt{\max\left\{\frac{(n_0 - 1)^2 S_{1i}^4 \xi}{\delta^2} - T, 0\right\}}\right\} \\ &= \Pr\left\{\sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}}\right) < -\lambda \sqrt{\max\left\{\frac{(n_0 - 1)^2 S_{1i}^4 \xi}{\delta^2 \sigma_{1i}^2} - \frac{T}{\sigma_{1i}^2}, 0\right\}}\right\} \\ &\leq \Pr_{\text{SC}}\left\{\sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}}\right) < -\lambda \sqrt{\max\left\{\frac{(n_0 - 1)^2 S_{1i}^4 \xi}{\delta^2 \sigma_{1i}^2} - \frac{T}{\sigma_{1i}^2}, 0\right\}}\right\} \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[\Pr_{\text{SC}} \left\{ \sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < -\lambda \sqrt{\max \left\{ \frac{(n_0 - 1)^2 S_{1i}^4 \xi}{\delta^2 \sigma_{1i}^2} - \frac{T}{\sigma_{1i}^2}, 0 \right\}} \middle| S_{1i} \right\} \right] \\
&\leq \mathbb{E} \left[\Pr_{\text{SC}} \left\{ \sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < 0 \middle| S_{1i} \right\} \right],
\end{aligned}$$

where $\sigma_{1i}^2 = \text{Var}(X_{1j} - X_{ij})$ and “SC” denotes the slippage configuration $\mu_1 = \mu_i + \delta$.

Since $(X_{1j} - X_{ij})/\sigma_{1i}, j = 1, 2, \dots$, are IID $N(\delta/\sigma_{1i}, 1)$ random variables under the SC, $\sum_{j=1}^T (X_{1j} - X_{ij})/\sigma_{1i}$ behaves like a Brownian motion process with drift δ/σ_{1i} at integer points. Therefore, if we let

$$a = \lambda, \quad s = \frac{(n_0 - 1)^2 S_{1i}^4 \xi}{\delta^2 \sigma_{1i}^2} \quad \text{and} \quad t = \frac{r}{\sigma_{1i}^2}, \quad r = n_0, \dots, N_{1i} + 1,$$

then by Lemma 1 and 2 we get

$$\begin{aligned}
&\mathbb{E} \left[\Pr_{\text{SC}} \left\{ \sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < 0 \middle| S_{1i} \right\} \right] \leq \\
&\mathbb{E} \left[\frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp \left\{ -\frac{(n_0 - 1)^2 S_{1i}^4 \xi}{\sigma_{1i}^4} \frac{\xi}{2} \right\} \sum_{n=0}^{\infty} \left(\frac{k_{2n+1}(\lambda)}{\ell_{2n+1}(\lambda)} \frac{2^n n!}{(2n+1)!} \left(\frac{(n_0 - 1) S_{1i}^2}{\sigma_{1i}^2} \sqrt{\xi} \right)^{2n+1} \right) \right].
\end{aligned}$$

Since $(n_0 - 1) S_{1i}^2 / \sigma_{1i}^2$ is chi-square distributed with $n_0 - 1$ degrees of freedom, the expectation above is equivalent to

$$\mathbb{E} \left[\frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp \left\{ -\frac{\xi}{2} (\chi_{n_0-1}^2)^2 \right\} \sum_{n=0}^{\infty} \left(\frac{k_{2n+1}(\lambda)}{\ell_{2n+1}(\lambda)} \frac{2^n n!}{(2n+1)!} (\chi_{n_0-1}^2 \sqrt{\xi})^{2n+1} \right) \right]. \quad (42)$$

Notice that (42) is the $g(\xi; \lambda, n_0)$ function, and this expectation is equal to β by the way we choose ξ . Therefore,

$$\Pr\{\text{ICS}_i\} \leq \beta. \quad (43)$$

Now, assume that we have $k \geq 2$ systems, and let ICS be the event that an incorrect selection is made by the procedure. If we set $\beta = \alpha/(k-1)$, then by the Bonferroni inequality we get

$$\Pr\{\text{ICS}\} \leq \sum_{i=2}^k \Pr\{\text{ICS}_i\} \leq \sum_{i=2}^k \frac{\alpha}{k-1} = \alpha.$$

The above inequality holds whether CRN are employed or systems are simulated independently. However, if systems are simulated independently (Σ is a diagonal matrix in that

case), a tighter boundary can be achieved, and its proof goes as follows:

$$\Pr\{\text{CS}\} \geq \Pr\left\{\bigcap_{i=2}^k \text{CS}_i\right\}$$

because the intersection event requires system 1 to eliminate each inferior system i individually, whereas in reality some system $\ell \neq 1, i$ could eliminate i . Thus,

$$\begin{aligned} \Pr\{\text{CS}\} &\geq \Pr\left\{\bigcap_{i=2}^k \text{CS}_i\right\} \\ &= \mathbb{E}\left[\Pr\left\{\bigcap_{i=2}^k \text{CS}_i \mid X_{11}, \dots, X_{1, N_1+1}\right\}\right] \\ &= \mathbb{E}\left[\prod_{i=2}^k \Pr\{\text{CS}_i \mid X_{11}, \dots, X_{1, N_1+1}\}\right], \end{aligned} \quad (44)$$

where the last equality follows because the events are conditionally independent. Clearly, (44) does not increase if we assume the slippage configuration, so we do so from here on.

Now, notice that $\Pr\{\text{CS}_i \mid X_{11}, \dots, X_{1, N_1+1}\}$ is nondecreasing in $X_{1j}, j = 1, \dots, N_1 + 1$.

Therefore, by Lemma 3,

$$\begin{aligned} (44) &\geq \prod_{i=2}^k \mathbb{E}[\Pr_{\text{SC}}\{\text{CS}_i \mid X_{11}, \dots, X_{1, N_1+1}\}] \\ &= \prod_{i=2}^k \mathbb{E}[1 - \Pr_{\text{SC}}\{\text{ICS}_i\}] \\ &\geq \prod_{i=2}^k (1 - \beta) \\ &= \left\{1 - \left(1 - (1 - \alpha)^{1/(k-1)}\right)\right\}^{k-1} = 1 - \alpha, \end{aligned}$$

where the last inequality comes from (43). \square

Corollary 1 *If $\mu_1 < \mu_2 + \delta$, then with probability $\geq 1 - \alpha$ the proposed procedure selects one of the systems whose means are less than δ from μ_1 .*

Proof: The proof is similar to that of Corollary 1 of [22]. \square

Corollary 1 guarantees that the proposed procedure chooses one of the “good” systems with at least $1 - \alpha$ guarantee when an alternative is less than δ better than the other alternatives.

3.1.2 Second Procedure

In this section, rather than presenting the second procedure that we call \mathcal{P}_2 , we only describe the differences between \mathcal{P}_1 and \mathcal{P}_2 , and give the full details in Appendix C.

Basically, \mathcal{P}_2 is a natural extension of [37] to the case with unknown variances. It is very similar to \mathcal{P}_1 but \mathcal{P}_2 takes a different continuation region $R_{i\ell}(r)$ and parameter ξ . More specifically, in **Screening** the continuation region $R_{i\ell}(r)$ for \mathcal{P}_2 is defined as

$$R_{i\ell}(r) = \lambda S_{i\ell} \sqrt{n_0 - 1} \sqrt{\max \left\{ \frac{(n_0 - 1) S_{i\ell}^2 \xi}{\delta^2} - r, 0 \right\}},$$

and **Parameters** needs to be revised as follows: $\lambda = 0.3$ is recommended for $1 - \alpha = 0.95$ — the detail of the choice of λ is discussed in Section 3.2.1 — and the constant ξ is the solution to the equation

$$h(\xi; \lambda, n_0) = \beta, \quad (45)$$

where

$$\begin{aligned} h(\xi; \lambda, n_0) \equiv & \mathbb{E} \left[\frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp \left\{ -\chi_{n_0-1}^2 \frac{\xi}{2} \right\} \right. \\ & \times \sum_{n=0}^{\infty} \left(\frac{k_{2n+1}(\lambda \sqrt{\chi_{n_0-1}^2})}{\ell_{2n+1}(\lambda \sqrt{\chi_{n_0-1}^2})} \frac{2^n n!}{(2n+1)!} \left(\sqrt{\chi_{n_0-1}^2 \xi} \right)^{2n+1} \right) \Big]. \end{aligned}$$

Theorem 7 *Under the same assumptions as in Theorem 6, \mathcal{P}_2 selects system 1 with probability greater than or equal to $1 - \alpha$.*

The theorem is proven in Appendix C.

3.2 Design of the Procedures

In this section, we discuss the implementation issues of \mathcal{P}_1 and \mathcal{P}_2 . The main issues involve the determination of λ and ξ . Besides the user-specified parameters δ , α , and n_0 that are required by all the indifference-zone procedures, \mathcal{P}_1 and \mathcal{P}_2 require two additional parameters, λ and ξ . The parameter λ is another user-specified parameter, but ξ is obtained from Equation (41) or (45). Procedure \mathcal{KN} also requires two parameters, say, c and η . The parameters c and η of \mathcal{KN} are similar to λ and ξ , respectively. The constant c in \mathcal{KN} is restricted to any positive integer, and the unique solution of η exists in a closed form with

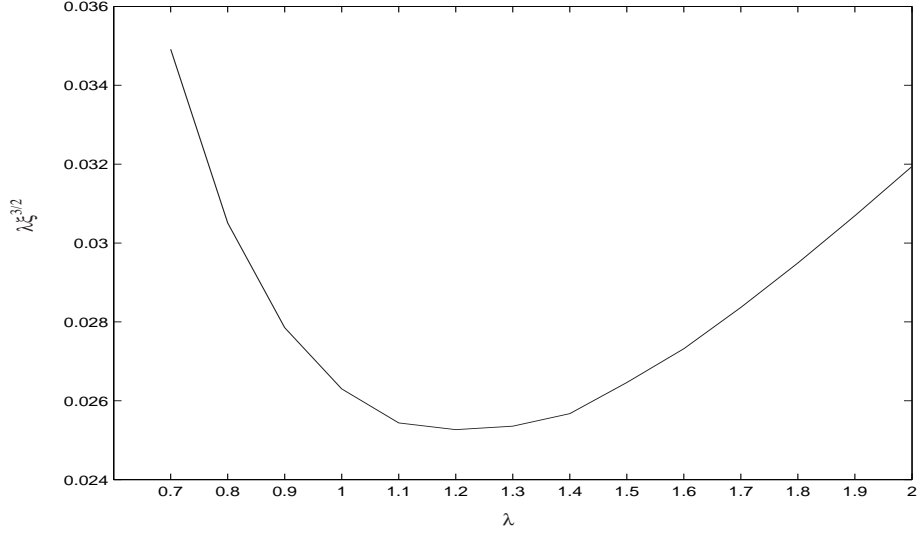


Figure 4: $\lambda\xi^{3/2}$ versus λ when $n_0 = 10$ and $\beta = 0.05$ in \mathcal{P}_1

the choice of $c = 1$ ([22]). However, in our procedures, λ can be any positive real number and the solution ξ to Equation (41) or (45) does not exist in a closed form. Therefore, in this section, we discuss which values of λ to choose and how to determine ξ quickly such that it satisfies Equation (41) or (45).

3.2.1 Choice of λ

Our work is based on [13] that presents a statistical procedure that determines whether the drift of a Brownian motion process is positive or negative when variance is known. They derive the expected first exit time of the process through a parabolic boundary, and recommend that one chooses the value of λ that minimizes the expected first exit time when the drift is zero. In [37], the value of λ is also determined in this manner.

In our case, due to unknown variances, we doubt if there exists a closed form of the expected first exit time. Therefore, we find a value λ that minimizes the area of the continuation region, which is proportional to the product of the vertical and horizontal intercepts. In [22], the area of the continuation region is also used to find the best choice of c . For example, in \mathcal{P}_1 , since the product of the horizontal and vertical intercepts is equal to $\lambda\xi^{3/2}(n_0 - 1)^3 S_{i\ell}^6 / \delta^3$, we only need to consider how $\lambda\xi^{3/2}$ changes with respect to λ since n_0 , δ , and $S_{i\ell}^2$ are not affected by the choice of λ . In \mathcal{P}_2 , we also get the area of the

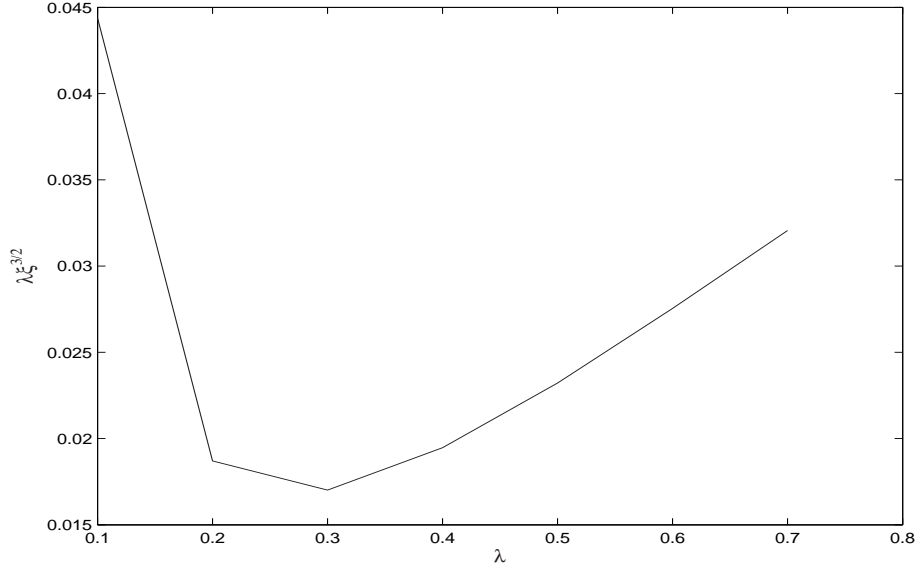


Figure 5: $\lambda\xi^{3/2}$ versus λ when $n_0 = 24$ and $\beta = 0.05$ in \mathcal{P}_2

continuation region proportional to $\lambda\xi^{3/2}$. Recall that λ can be arbitrarily chosen as any real positive number while ξ is the solution to $g(\xi; \lambda, n_0) = \beta$ or $h(\xi; \lambda, n_0) = \beta$ where β is either $\alpha/(k-1)$ or $1 - (1-\alpha)^{1/(k-1)}$ depending on the use of CRN.

Figure 4 shows the graph of $\lambda\xi^{3/2}$ versus λ when $n_0 = 10$ and $\beta = 0.05$ in \mathcal{P}_1 . As n_0 increases, the graph becomes even flatter around its minimum, especially for $0.9 < \lambda < 1.6$. When β decreases due to a decrease in α or increase in k , the graph tends to shift to the right. The range of α of general interest is $0.01 \leq \alpha \leq 0.1$ and this leads to the range of β as $0 < \beta \leq 0.1$. For this range of β , the minimum of $\lambda\xi^{3/2}$ is still achieved around $\lambda = 1.0$. Since we usually take n_0 larger than 10 — in this paper we recommend $n_0 = 24$, which will be discussed in the next section — and the graph does not dramatically shift to the right when β changes, we recommend taking $\lambda = 0.9268$ for all values of α . This is also the choice in [13] when $\beta = 0.05$.

Figure 5 shows the graph of $\lambda\xi^{3/2}$ versus λ when $n_0 = 24$ and $\beta = 0.05$ in \mathcal{P}_2 . The minimum is achieved around $\lambda = 0.3$. Unlike \mathcal{P}_1 , as n_0 decreases, the graph of \mathcal{P}_2 becomes flatter around its minimum — we conjecture that this reverse tendency is due to the difference in Equations (41) and (45). For $0 < \beta \leq 0.01$, the minimum is still achieved around $\lambda = 0.3$; therefore, we recommend taking $\lambda = 0.3$ for all values of α .

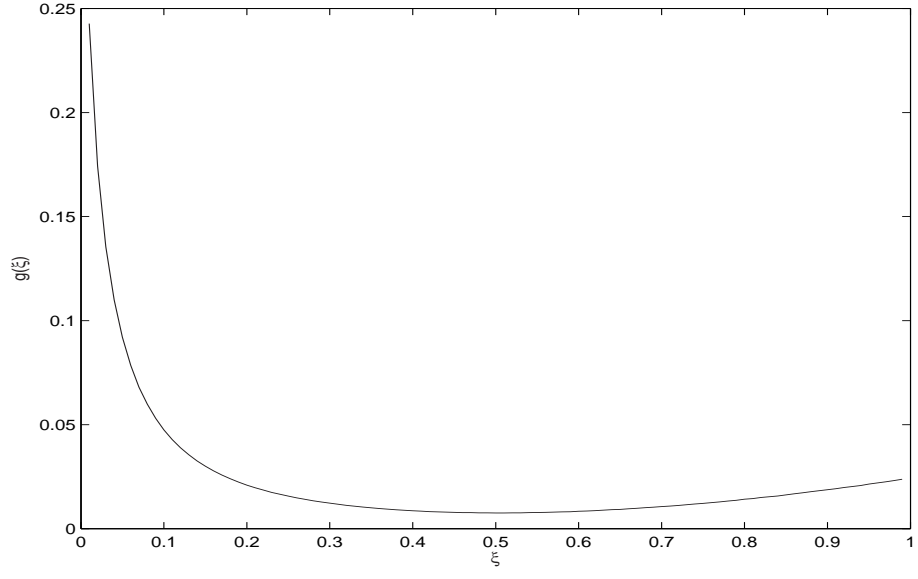


Figure 6: $g(\xi; \lambda, n_0)$ versus ξ Graph when $n_0 = 10$ and $\lambda = 0.9268$ in \mathcal{P}_1

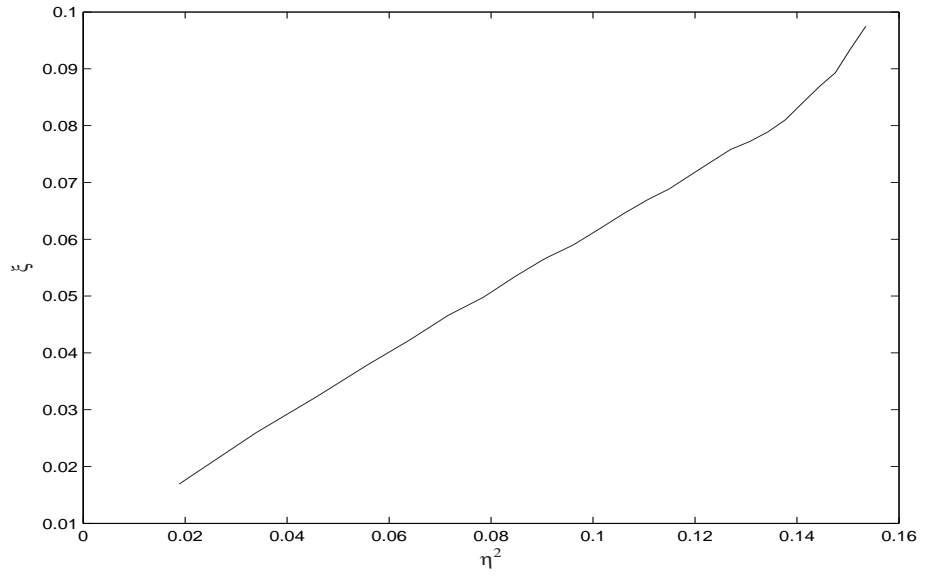


Figure 7: η^2 versus ξ Graph when $n_0 = 20$, $\lambda = 0.9268$, and $0.002050 < \beta < 0.05$ in \mathcal{P}_1

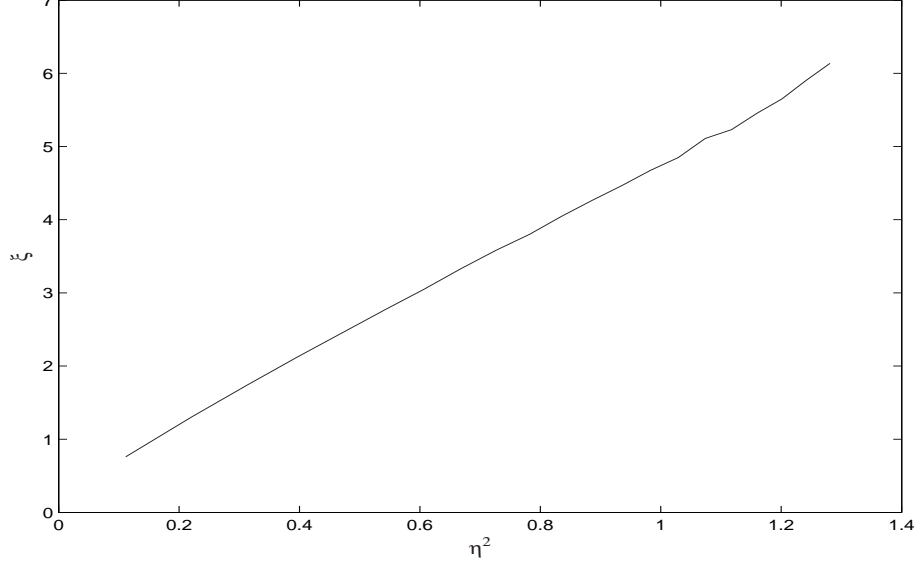


Figure 8: η^2 versus ξ Graph when $n_0 = 10$, $\lambda = 0.3$, and $0.00244 < \beta < 0.05$ in \mathcal{P}_2

3.2.2 Choice of ξ

Once we pick a value for λ , we can determine the value of ξ from Equation (41) or (45) for given n_0 and β in \mathcal{P}_1 or \mathcal{P}_2 , respectively. Since the expectations are not analytically solvable and quite complicated to be solved by numerical integration, we find them by simulation. More specifically, for given n_0 and λ , we generated four million chi-squared random variables to estimate the expectation in Equation (41) or (45) at different values of ξ and created tables of $g(\xi; \lambda, n_0)$ and $h(\xi; \lambda, n_0)$ with respect to ξ from 0.01 to 1 for \mathcal{P}_1 and from 0.01 to 7 for \mathcal{P}_2 with an increment of 0.01. Then one can find ξ at which the estimated expectation is equal to β by interpolation.

The difficulty in determination of ξ is that the existence of ξ is not guaranteed for all possible values of β , especially for small β . The parameter β , determined by α and k , gets smaller and approaches to zero as k increases. Therefore, we also need $g(\xi; \lambda, n_0)$ or $h(\xi; \lambda, n_0)$ to decrease as ξ increases in order to guarantee that ξ exists such that $g(\xi; \lambda, n_0)$ or $h(\xi; \lambda, n_0)$ is equal to β for all possible values of β . However, $g(\xi; \lambda, n_0)$ and $h(\xi; \lambda, n_0)$ for some n_0 values do not approach to zero as ξ increases. Instead, they have a non-zero positive minimum which we denote as β_{min} . This implies that if β is smaller than β_{min} , then ξ does not exist such that $g(\xi; \lambda, n_0) = \beta$ or $h(\xi; \lambda, n_0) = \beta$. Since ξ can be computed

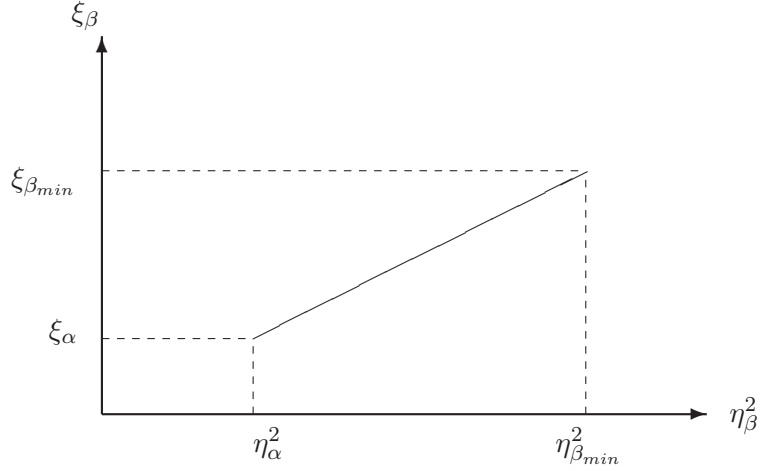


Figure 9: Linearity Relationship Between η_β^2 and ξ_β

up to a quite small β for $n_0 > 15$ in \mathcal{P}_2 , this problem is more severe in \mathcal{P}_1 compared to \mathcal{P}_2 . For example, Figure 6 shows that when $n_0 = 10$ and $\lambda = 0.9268$, β_{min} is 0.0076 for \mathcal{P}_1 . This implies that if $\alpha = 0.05$, then ξ for \mathcal{P}_1 is determined only up to $k = 7$ in which case $\beta = \alpha/(k - 1) \simeq 0.0083$.

In [22], it is shown that the parameter η when $c = 1$ is determined by

$$\eta = \frac{1}{2} \left[(2\beta)^{-2/(n_0-1)} - 1 \right], \quad (46)$$

and we observed that η^2 and ξ seem to have an approximate linear relationship for the range of β where we are able to get ξ from $g(\xi; \lambda, n_0) = \beta$ or $h(\xi; \lambda, n_0) = \beta$, that is, for the range of β greater than or equal to β_{min} . Figure 7 shows an example of the linear relationship between η^2 and ξ when $n_0 = 20$, $\lambda = 0.9268$, and $0.002050 < \beta < 0.05$ in \mathcal{P}_1 . Similarly, Figure 8 shows the linear relationship between η^2 and ξ when $n_0 = 10$, $\lambda = 0.3$, and $0.00244 < \beta < 0.05$ in \mathcal{P}_2 . We observed a similar linear relationship for all other n_0 values. We will use this linear relationship to get ξ for $\beta < \beta_{min}$.

We will add one more subscript to ξ and η and use ξ_β and η_β to represent values of ξ and η at specific β , respectively. As shown in Figure 9, we use the points $(\eta_\alpha^2, \xi_\alpha)$ and $(\eta_{\beta_{min}}^2, \xi_{\beta_{min}})$ to determine the line

$$Y - \xi_\alpha = (x - \eta_\alpha^2) \frac{\xi_{\beta_{min}} - \xi_\alpha}{\eta_{\beta_{min}}^2 - \eta_\alpha^2}.$$

Table 4: $\xi_{\beta_{min}}$ and ξ_{α} Values when $\lambda = 0.9268$, $n_0 = 10, 15, 20, 24$, and $\alpha = 0.01, 0.025, 0.05, 0.10$ in \mathcal{P}_1

n_0	$1 - \beta_{min}$	$\xi_{\beta_{min}}$	$\xi_{0.01}$	$\xi_{0.025}$	$\xi_{0.05}$	$\xi_{0.10}$
10	0.9924	0.49	0.3500	0.1744	0.09554	0.04556
15	0.9956	0.17	0.09474	0.05366	0.03194	0.01720
20	0.9980	0.10	0.04282	0.02618	0.01686	0.008229
24	0.99954	0.10	0.02742	0.01755	0.01	0.005457

Table 5: $\xi_{\beta_{min}}$ and ξ_{α} Values when $\lambda = 0.3$, $n_0 = 10, 15, 20, 24$, and $\alpha = 0.01, 0.025, 0.05, 0.10$ in \mathcal{P}_2

n_0	$1 - \beta_{min}$	$\xi_{\beta_{min}}$	$\xi_{0.01}$	$\xi_{0.025}$	$\xi_{0.05}$	$\xi_{0.10}$
10	0.9976	6.14	2.49	1.3167	0.7543	0.3809
15	0.99973	4.11	0.84	0.5062	0.3196	0.1763
20	0.99996	2.96	0.46	0.2956	0.1947	0.1116
24	0.99999	2.30	0.3329	0.2196	0.1476	0.08608

If the linear tendency continues, ξ_{β} values for $\beta < \beta_{min}$ can be approximated by

$$Y = (\eta_{\beta}^2 - \eta_{\alpha}^2) \frac{\xi_{\beta_{min}} - \xi_{\alpha}}{\eta_{\beta_{min}}^2 - \eta_{\alpha}^2} + \xi_{\alpha}.$$

Therefore, for the approximation of ξ_{β} for $\beta < \beta_{min}$, we only need the information about $(\eta_{\alpha}^2, \xi_{\alpha})$, $(\eta_{\beta_{min}}^2, \xi_{\beta_{min}})$, and η_{β}^2 . The values of η can be easily computed from Equation (46) and the values of ξ_{α} and $\xi_{\beta_{min}}$ for the popular choices of n_0 and α when $n_0 = 10, 15, 20, 24$ and $\alpha = 0.01, 0.025, 0.05, 0.10$ can be found in Table 4 for \mathcal{P}_1 and in Table 5 for \mathcal{P}_2 .

Table 6 shows the actual ξ_{β} found from $g(\xi; n_0, \lambda) = \beta$ and approximate ξ_{β} from the linear relationship when $n_0 = 20$ and $\lambda = 0.9268$ in \mathcal{P}_1 . The approximation errors are quite small for the range of $\beta_{min} < \beta < 0.05$. Also, we tested the performance of the approximation technique when $k = 3$. The most difficult case in terms of providing the PCS requirement is $k = 2$ because as k increases we expect the actual PCS to become much more large than the nominal PCS due to the Bonferroni inequality. However, to use the approximation technique k needs to be at least 3 because $\beta = \alpha$ when $k = 2$. When $k = 3$, for different α values, converging to zero, we determined the ξ values by the approximation technique. Then we checked if the PCS requirement is satisfied on the same experimental setup, which will be explained in Section 3.4. The results show that the PCS requirement

Table 6: Actual and Approximate ξ_β Values when $n_0 = 20$ and $\lambda = 0.9268$ in \mathcal{P}_1

β	Act. ξ_β	Appr. ξ_β	Error
0.05	0.0169	0.0169	0.0000
0.025321	0.0260	0.0260	0.0000
0.016952	0.0323	0.0330	0.0007
0.012741	0.0378	0.0389	0.0011
0.010206	0.0423	0.0440	0.0017
0.008512	0.0466	0.0485	0.0019
0.007301	0.0498	0.0526	0.0028
0.006391	0.0535	0.0564	0.0029
0.005683	0.0566	0.0598	0.0032
0.005116	0.0590	0.0631	0.0041
0.004652	0.0618	0.0662	0.0044
0.004265	0.0645	0.0691	0.0046
0.003938	0.0669	0.0718	0.0049
0.003657	0.0689	0.0744	0.0055
0.003414	0.0713	0.0769	0.0056
0.003201	0.0736	0.0793	0.0057
0.003013	0.0758	0.0816	0.0058
0.002846	0.0772	0.0838	0.0066
0.002696	0.0789	0.0860	0.0071
0.002561	0.0810	0.0881	0.0071
0.00244	0.0840	0.0901	0.0061
0.002329	0.0868	0.0920	0.0052
0.002228	0.0893	0.0939	0.0046
0.002135	0.0935	0.0957	0.0022
0.002050	0.0975	0.0975	0.0000

is satisfied.

Though there is an approximate linear tendency between η_β^2 and ξ_β for all n_0 and $\beta_{min} < \beta \leq \alpha$, it does not necessarily guarantee the linear tendency for $\beta < \beta_{min}$ and a deviation from the linearity assumption is possible as β gets smaller than β_{min} . When $n_0 = 24$, ξ_β is determined up to a quite small β by Equations (41) and (45). Therefore, we recommend using 24 as the initial sample size in order to make sure that the linear relationship holds up to a quite small β . However, when the number of systems is large and each replication takes long, the choice of $n_0 = 24$ may not be a good one. In that case, one can use n_0 smaller than 24 and approximate the value of ξ for large k using the linear approximation proposed in this section.

3.3 Analytical Comparisons of the Continuation Regions

In this section, we perform analytical comparisons among the parabolic regions of \mathcal{P}_1 and \mathcal{P}_2 and the triangular region of \mathcal{KN} . We will use $R(t)$ to denote the generic continuation region, and H and V to denote intercepts of $R(t)$ with the horizontal and vertical axes, respectively.

In [22], it is stated that if the experimenter has no idea whether there are a few dominant systems or a number of close competitors, the choice of $c = 1$ appears to be a good compromise solution. Hence, we will take $c = 1$ throughout this chapter. The triangular continuation region for \mathcal{KN} when $c = 1$ is defined as

$$R(t)_{\mathcal{KN}} = \left(-\frac{\eta(n_0 - 1)S_{il}^2}{\delta} + \frac{\delta}{2}t, \frac{\eta(n_0 - 1)S_{il}^2}{\delta} - \frac{\delta}{2}t \right), \quad 0 \leq t \leq \frac{2\eta(n_0 - 1)S_{il}^2}{\delta^2}.$$

Then the intercepts of \mathcal{KN} , \mathcal{P}_1 , and \mathcal{P}_2 are

$$\begin{aligned} V_{\mathcal{KN}} &= \frac{\eta(n_0 - 1)S_{il}^2}{\delta}, & H_{\mathcal{KN}} &= \frac{2\eta(n_0 - 1)S_{il}^2}{\delta^2}, \\ V_{\mathcal{P}_1} &= \frac{\lambda_1\sqrt{\xi_1}(n_0 - 1)S_{il}^2}{\delta}, & H_{\mathcal{P}_1} &= \frac{\xi_1(n_0 - 1)^2S_{il}^4}{\delta^2}, \\ V_{\mathcal{P}_2} &= \frac{\lambda_2\sqrt{\xi_2}(n_0 - 1)S_{il}^2}{\delta}, & H_{\mathcal{P}_2} &= \frac{\xi_2(n_0 - 1)S_{il}^2}{\delta^2}, \end{aligned}$$

where λ_1 and λ_2 represent the values of λ for \mathcal{P}_1 and \mathcal{P}_2 , respectively. Similarly, ξ_1 and ξ_2 represent the values of ξ for \mathcal{P}_1 and \mathcal{P}_2 . There should be no confusion between ξ_1 or ξ_2 and ξ_α or $\xi_{\beta_{\min}}$ — defined in Section 3.2.2 — since α and β_{\min} are always smaller than one.

The analysis in this section is based on ξ_1 , ξ_2 , and η values assuming that systems are simulated independently and $1 - \alpha = 0.95$. If CRN are employed, the values of ξ_1 , ξ_2 , and η differ from those without CRN, resulting slightly different continuation regions; however, this change is not significant enough to affect our analysis of the relationships among the regions.

Given that all three procedures provide actual PCS larger than the nominal PCS, we prefer the one with a smaller V and a smaller H since it provides a tighter screening step and guarantees that the procedure ends earlier if sampling continues to the end stage. More specifically, a smaller V means tighter screening at the beginning; therefore, a procedure with smaller V works efficiently if there exist a few dominant systems. However, if all

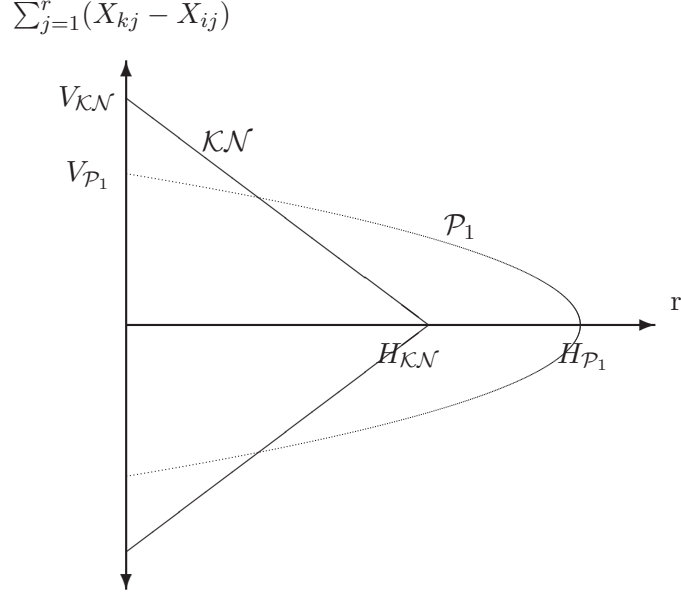


Figure 10: Continuation Regions of \mathcal{P}_1 and \mathcal{KN}

systems are close to the best system (e.g., the SC), then each system is likely to take a number of observations close to H and a procedure with a smaller H is preferred. Therefore, if the continuation region of a procedure is strictly inside of that of another procedure, then it is clear that the former dominates the latter. For the comparison, we will focus on the values of H and V and the shape of the continuation regions.

\mathcal{P}_1 vs. \mathcal{KN} : One can easily compute the ratio of vertical intercepts of \mathcal{KN} and \mathcal{P}_1 . The ratio $V_{\mathcal{KN}}/V_{\mathcal{P}_1}$ is equal to $\eta/(\lambda_1\sqrt{\xi_1})$. Table 7 shows the ratios of $V_{\mathcal{KN}}$ and $V_{\mathcal{P}_1}$ for $n_0 = 10, 24$ and $k = 2, 5, 10, 25, 100, 200, 400$. From these results, we conjecture that \mathcal{P}_1 has a tighter screening at the beginning than \mathcal{KN} . This tightness tends to be stronger as the number of systems k increases, but not always.

Now, consider the horizontal intercepts of the two procedures. The ratio $H_{\mathcal{KN}}/H_{\mathcal{P}_1}$ is equal to $2\eta/((n_0 - 1)S_{i\ell}^2\xi_1)$. Since $S_{i\ell}^2$ is random, it is hard to evaluate this ratio. However, in general, when $S_{i\ell}^2 > 1$, this ratio is less than one and $H_{\mathcal{P}_1}$ is larger than $H_{\mathcal{KN}}$.

From this analysis, one can conjecture that \mathcal{P}_1 starts with a tighter screening than \mathcal{KN} but after some point, the screening of \mathcal{KN} becomes tighter as one can see in Figure 10. It implies that when k is large and there exist a number of dominant systems, \mathcal{P}_1 is expected

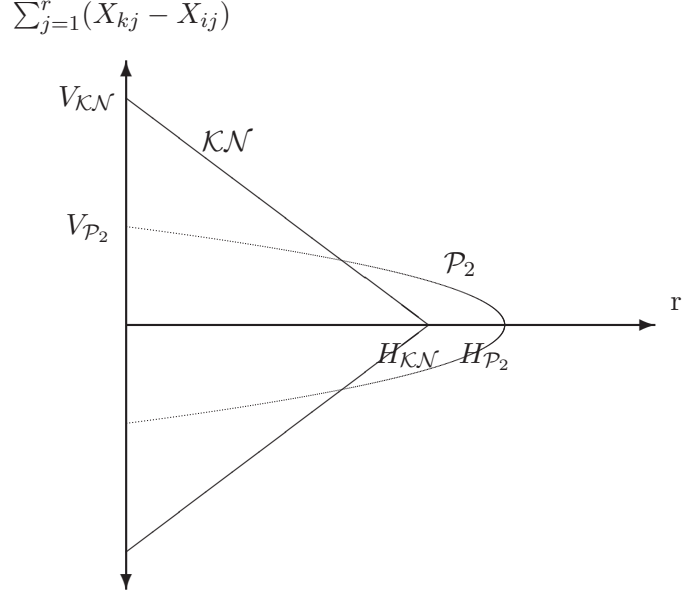


Figure 11: Continuation Regions of \mathcal{P}_2 and \mathcal{KN} for Large k

Table 7: $V_{\mathcal{KN}}/V_{\mathcal{P}_1}$ when $n_0 = 10, 24$ and $k = 2, 5, 10, 25, 100, 200, 400$

k	$n_0 = 10$	$n_0 = 24$
2	1.166	1.196
5	1.260	1.295
10	1.218	1.366
25	1.264	1.434
100	1.269	1.476
200	1.270	1.433
400	1.271	1.437

to be more efficient due to its tighter screening at the beginning. However, under a configuration close to the SC where the procedure is likely to continue to the end stage, we expect \mathcal{KN} to be more efficient than \mathcal{P}_1 .

\mathcal{P}_2 vs. \mathcal{KN} : The ratio $V_{\mathcal{KN}}/V_{\mathcal{P}_2}$ is equal to $\eta/(\lambda_2\sqrt{\xi_2})$. Table 8 shows $V_{\mathcal{KN}}/V_{\mathcal{P}_2}$ and $H_{\mathcal{KN}}/H_{\mathcal{P}_2}$ ratios for $n_0 = 10, 24$ and $k = 2, 5, 10, 25, 100, 200, 400$. The numbers in the table imply that \mathcal{P}_2 is tighter than \mathcal{KN} at the beginning of the procedure and it becomes much more tight as k increases.

The ratio $H_{\mathcal{KN}}/H_{\mathcal{P}_2}$ is equal to $2\eta/\xi_2$. $H_{\mathcal{KN}}$ is always smaller than $H_{\mathcal{P}_2}$ when $n_0 = 10$. However, when $n_0 = 24$, $H_{\mathcal{KN}}$ is larger than $H_{\mathcal{P}_2}$ for $k < 50$. Therefore, when $n_0 = 24$ and

Table 8: $V_{\mathcal{KN}}/V_{\mathcal{P}_2}$ and $H_{\mathcal{KN}}/H_{\mathcal{P}_2}$ when $n_0 = 10, 24$ and $k = 2, 5, 10, 25, 100, 200, 400$

k	$n_0 = 10$		$n_0 = 24$	
	$V_{\mathcal{KN}}/V_{\mathcal{P}_2}$	$H_{\mathcal{KN}}/H_{\mathcal{P}_2}$	$V_{\mathcal{KN}}/V_{\mathcal{P}_2}$	$H_{\mathcal{KN}}/H_{\mathcal{P}_2}$
2	1.282	0.886	0.961	1.501
5	1.441	0.593	1.143	1.250
10	1.501	0.476	1.229	1.142
25	1.526	0.355	1.329	1.047
100	1.541	0.237	1.445	0.919
200	1.506	0.187	1.487	0.855
400	1.548	0.164	1.532	0.803

Table 9: $V_{\mathcal{P}_1}/V_{\mathcal{P}_2}$ when $n_0 = 10, 24$ and $k = 2, 5, 10, 25, 100, 200, 400$

k	$n_0 = 10$	$n_0 = 24$
2	1.099	0.804
5	1.143	0.882
10	1.232	0.899
25	1.207	0.926
100	1.215	0.979
200	1.185	1.038
400	1.218	1.066

k is small (say $k < 50$), \mathcal{P}_2 is likely to outperform \mathcal{KN} because both V and H of \mathcal{P}_2 are smaller than those of \mathcal{KN} — implying that it is possible that the continuation region of \mathcal{P}_2 is inside that of \mathcal{KN} . On the other hand, as one can see in Table 8, for large k ($k \geq 50$) there does not exist a uniform superiority between \mathcal{KN} and \mathcal{P}_2 because one has a smaller V but the other has a smaller H . Figure 11 demonstrates possible continuation regions of these two procedures for large k . From this figure, we conjecture that \mathcal{P}_2 will show better performance than \mathcal{KN} in general due to a tighter initial screening. However, when k is large (say, ≥ 50) and most systems are close to the best system like the SC, it is possible that \mathcal{KN} outperforms \mathcal{P}_2 .

\mathcal{P}_1 vs. \mathcal{P}_2 : The ratio $V_{\mathcal{P}_1}/V_{\mathcal{P}_2}$ is equal to $(\lambda_1\sqrt{\xi_1})/(\lambda_2\sqrt{\xi_2})$. Table 9 shows the ratios of $V_{\mathcal{P}_1}$ and $V_{\mathcal{P}_2}$ for $n_0 = 10, 24$ and $k = 2, 5, 10, 25, 100, 200, 400$. The numbers in the table imply that \mathcal{P}_2 is slightly tighter than \mathcal{P}_1 either when $n_0 = 10$ or when $n_0 = 24$ and k is

large.

The ratio $H_{\mathcal{P}_1}/H_{\mathcal{P}_2}$ is equal to $\xi_1(n_0 - 1)S_{i\ell}^2/\xi_2$. Although $S_{i\ell}^2$ is random, it is clear that this ratio is greater than one in general. Actually, it can be quite large because of the $(n_0 - 1)$ and $S_{i\ell}^2$ factors, implying that H of \mathcal{P}_1 is likely to be a lot larger than that of \mathcal{P}_2 .

As a result, we can say that except for the case when $n_0 = 24$ and k is small, \mathcal{P}_2 is likely to outperform \mathcal{P}_1 because both V and H of \mathcal{P}_2 are smaller than those of \mathcal{P}_1 . When $n_0 = 24$ and k is small, \mathcal{P}_1 has a slightly tighter screening than \mathcal{P}_2 at the beginning. However, $H_{\mathcal{P}_1}$ is usually much larger than $H_{\mathcal{P}_2}$ because n_0 is larger than 10 and $S_{i\ell}^2$ is usually larger than one. This can overshadow the advantage of \mathcal{P}_1 having a smaller $V_{\mathcal{P}_1}$. Thus, it is not clear which procedure will do better for small k and should be tested by experiments.

3.4 *Experimental Results*

In this section, we compare the experimental results of \mathcal{KN} , \mathcal{P}_1 , and \mathcal{P}_2 based on IID normal data. In [22], the performance of \mathcal{KN} is compared to that of Rinott procedure ([31]) and a two-stage screen-and-select procedure proposed by [29], which are both statistically valid indifference-zone selection procedures. The performance is evaluated in terms of the sample average of the total number of basic observations required by each procedure when all the procedures achieve the nominal PCS. These results show that \mathcal{KN} is highly efficient for finding the best system compared to the other two procedures. Therefore, we use \mathcal{KN} as our benchmark for comparison and test the performance of our procedures on the same scenarios used in [22].

Two configurations of the true means are employed: slippage configuration (SC) and monotonically decreasing mean (MDM) configuration. For each mean configuration, three variance configurations are tested: increasing variance (IV), decreasing variance (DV), and common variance (CV). When we combine these configurations, we get a total of six different configurations. The particular mean and variances that we use for the experiments are given in Table 10. Notice that system 1 is set to be the best system.

The nominal PCS is set at $1 - \alpha = 0.95$. As suggested in Section 3.2, λ and n_0 are set to $\lambda = 0.9268$ for \mathcal{P}_1 and 0.3 for \mathcal{P}_2 , and $n_0 = 24$ for all configurations, respectively. Each

Table 10: The Configurations of Means and Variances Used in the Experiments

SC-CV	
mean	$\mu_1 = \delta, \mu_2 = \mu_3 = \dots = 0$
variance	$\sigma_1^2 = \sigma_2^2 = \dots = 1$
MDM-CV	
mean	$\mu_i = \delta - (i - 1)\delta, i \geq 1$
variance	$\sigma_1^2 = \sigma_2^2 = \dots = 1$
MDM-IV	
mean	$\mu_i = \delta - (i - 1)\delta, i \geq 1$
variance	$\sigma_i^2 = 1 + (i - 1)\delta, i \geq 1$
MDM-DV	
mean	$\mu_i = \delta - (i - 1)\delta, i \geq 1$
variance	$\sigma_i^2 = 1 / (1 + (i - 1)\delta), i \geq 1$
SC-IV	
mean	$\mu_1 = \delta, \mu_2 = \mu_3 = \dots = 0$
variance	$\sigma_i^2 = 1 + (i - 1)\delta, i \geq 1$
SC-DV	
mean	$\mu_1 = \delta, \mu_2 = \mu_3 = \dots = 0$
variance	$\sigma_i^2 = 1 / (1 + (i - 1)\delta), i \geq 1$

result is obtained from 500 macro-replications of the entire experiment. In order to check if each procedure is able to provide the nominal PCS, we record the estimated PCS for each configuration. If all procedures have estimated PCS larger than the nominal PCS, the comparison will be done based on the sample average of the total number of observations required. With 500 macro-replications, the first two digits of these sample averages are statistically meaningful.

Our experimental results support that both \mathcal{P}_1 and \mathcal{P}_2 are statistically valid since the estimated PCS values are all higher than the nominal PCS when systems are simulated independently, as shown in Table 11. For \mathcal{P}_2 , we observe that there are a few configurations where the estimated PCS values are lower than 0.95 when we make 500 macro-replications. However, when the number of macro-replications is increased to 5000, the estimated PCS values of \mathcal{P}_2 are all over 0.95. The numbers in Table 11 are from 500 macro-replications except the estimated PCS of \mathcal{P}_2 for $k = 2$ and 5. They are from 5000 macro-replications. The estimated PCS increases when CRN are employed though we did not report them in this paper. Therefore, \mathcal{KN} , \mathcal{P}_1 , and \mathcal{P}_2 seem to be all statistically valid as expected. From

now on, we will focus on the comparison among three procedures based on the sample sizes required until we reach a decision.

We categorize our experimental results into two cases: independent and correlated. In the independent case, all systems are simulated independently. In the correlated case, a constant correlation factor, ρ , is used between all pairs of systems.

3.4.1 Independent Case

We first compare the performance of the two new procedures with \mathcal{KN} when systems are simulated independently. Overall, \mathcal{P}_2 performs better than both \mathcal{P}_1 and \mathcal{KN} in most cases in terms of the total number of observations required. However, the computational time needed to search for the ξ_2 parameter for a given β is very long compared to that of ξ_1 unless a table for ξ_2 already exists. Therefore, if the gain in the simulation time when \mathcal{P}_2 is applied is not significant enough to compensate time loss in computing the ξ_2 parameter, one may prefer to use \mathcal{P}_1 , which also shows better performance than \mathcal{KN} under MDM configurations when k is large and/or CRN effect is strong.

\mathcal{P}_1 vs. \mathcal{KN} : Table 12 shows the sample average of the total number of observations taken in \mathcal{KN} , \mathcal{P}_1 , and \mathcal{P}_2 when $n_0 = 24$, $\delta = 1/\sqrt{n_0}$, and $1 - \alpha = 0.95$ as a function of k . The results of \mathcal{KN} are taken from [22]. Numbers in boldface represent the cases that \mathcal{P}_1 or \mathcal{P}_2 shows better performance — i.e., spends a smaller number of observations — than \mathcal{KN} . Numbers in a box show which procedure among those three spends the smallest observations for each configuration.

When $k = 2$ or 5 , \mathcal{KN} is superior to \mathcal{P}_1 . However, as k increases, we observe that \mathcal{P}_1 defeats \mathcal{KN} in all the MDM configurations and SC with decreasing variances. As expected in Section 3.3, this is due to the fact that \mathcal{P}_1 has a tighter screening at the beginning and this tightness becomes even stronger as k increases. Therefore, under MDM configurations with large k , \mathcal{P}_1 does a better job in detecting inferior systems early. Similarly, under SC, if we have decreasing variances as means become inferior, it is easier to detect inferior systems due to low variances of inferior systems. These low variances help \mathcal{P}_1 to eliminate inferior systems early in the experiment and, thus, to show better performance than \mathcal{KN} in SC-DV.

Table 11: Estimated PCS in \mathcal{KN} , \mathcal{P}_1 , and \mathcal{P}_2 when $n_0 = 24$, $\delta = 1/\sqrt{n_0}$, and $1 - \alpha = 0.95$ as a Function of k when Systems are Simulated Independently

k		SC-CV	MDM-CV	MDM-IV	MDM-DV	SC-IV	SC-DV
2	\mathcal{KN}	0.958	0.958	0.954	0.960	0.954	0.960
	\mathcal{P}_1	0.982	0.982	0.968	0.972	0.968	0.972
	\mathcal{P}_2	0.957	0.957	0.952	0.954	0.952	0.954
5	\mathcal{KN}	0.960	0.988	0.986	0.982	0.958	0.972
	\mathcal{P}_1	0.982	0.992	0.992	0.996	0.972	0.986
	\mathcal{P}_2	0.962	0.987	0.988	0.989	0.960	0.963
10	\mathcal{KN}	0.964	0.992	0.994	0.996	0.958	0.968
	\mathcal{P}_1	0.970	0.996	0.994	0.998	0.986	0.974
	\mathcal{P}_2	0.978	0.988	0.998	0.996	0.966	0.972
25	\mathcal{KN}	0.964	0.998	1.000	1.000	0.950	0.982
	\mathcal{P}_1	0.992	1.000	0.998	1.000	0.984	0.980
	\mathcal{P}_2	0.962	0.996	1.000	0.998	0.964	0.984
100	\mathcal{KN}	0.956	0.998	1.000	1.000	0.960	0.996
	\mathcal{P}_1	0.980	1.000	1.000	1.000	0.982	1.000
	\mathcal{P}_2	0.948	0.998	1.000	0.998	0.950	0.992

However, in the SC-CV and SC-IV, most elimination events are likely to take place at a stage close to the end of the procedure. We have seen that the screening of \mathcal{P}_1 is tighter than that of \mathcal{KN} only at the beginning of experimentation, but it soon becomes loose and its horizontal intercept is usually larger than that of \mathcal{KN} . Therefore, \mathcal{KN} tends to show better performance under these configurations.

The superiority of \mathcal{P}_1 is most noticeable under MDM-IV. Under MDM-DV, it is already easy to detect inferior systems due to their small variances. Thus, having a tighter continuation region does not help that much, and \mathcal{KN} and \mathcal{P}_1 show similar performance. However, under MDM-IV where it is more difficult to detect inferior systems due to their high variances, the tightness of \mathcal{P}_1 stands out and helps \mathcal{P}_1 to outperform \mathcal{KN} .

\mathcal{P}_2 vs. \mathcal{KN} : When $k = 2$, \mathcal{KN} is better than \mathcal{P}_2 under all configurations we tested. This is because $V_{\mathcal{KN}} < V_{\mathcal{P}_2}$ — i.e., \mathcal{KN} has a tighter screening — although $H_{\mathcal{KN}} > H_{\mathcal{P}_2}$ when k is small. However, for $k \geq 5$, \mathcal{P}_2 outperforms \mathcal{KN} in all configurations as expected in Section 3.3 because $V_{\mathcal{KN}}$ becomes larger than $V_{\mathcal{P}_2}$ as k increases although $H_{\mathcal{KN}}$ becomes slightly less than $H_{\mathcal{P}_2}$. In Table 8, we observe that when $n_0 = 10$, the ratio $V_{\mathcal{KN}}/V_{\mathcal{P}_2}$

Table 12: Sample Average of the Total Number of Observations in \mathcal{KN} , \mathcal{P}_1 , and \mathcal{P}_2 when $n_0 = 24, \delta = 1/\sqrt{n_0}$, and $1 - \alpha = 0.95$ as a Function of k

k		SC-CV	MDM-CV	MDM-IV	MDM-DV	SC-IV	SC-DV
2	\mathcal{KN}	158	158	175	147	175	147
	\mathcal{P}_1	187	187	207	167	207	167
	\mathcal{P}_2	172	172	185	154	185	154
5	\mathcal{KN}	738	456	542	403	939	594
	\mathcal{P}_1	800	485	568	409	1050	643
	\mathcal{P}_2	725	448	523	380	940	575
10	\mathcal{KN}	1727	761	981	630	2868	1149
	\mathcal{P}_1	1861	710	910	615	3049	1185
	\mathcal{P}_2	1689	695	890	612	2758	1124
25	\mathcal{KN}	5015	1333	2009	1157	13399	2312
	\mathcal{P}_1	5109	1217	1647	1091	14154	2241
	\mathcal{P}_2	4705	1185	1639	1072	12336	2252
100	\mathcal{KN}	23956	3483	6312	3215	189401	5331
	\mathcal{P}_1	24634	3303	4682	3101	192760	4990
	\mathcal{P}_2	21123	3235	4653	3052	170987	5210

increases as k increases. However, due to the sharp decrease in the ratio $H_{\mathcal{KN}}/H_{\mathcal{P}_2}$, \mathcal{KN} shows better performance than \mathcal{P}_2 under SC for some values of k , especially when k is large as expected in Section 3.3. But, in practice the SC configuration is unrealistic when k is large.

\mathcal{P}_1 vs. \mathcal{P}_2 : From Section 3.3, we know that when $n_0 = 10$, the continuation region of \mathcal{P}_2 is inside that of \mathcal{P}_1 . Hence, \mathcal{P}_2 is expected to show better performance compared to \mathcal{P}_1 under all configurations and the experimental results support this. When $n_0 = 24$ and $k > 150$, we observe a similar result. However, when $n_0 = 24$ and $k \leq 150$, since $V_{\mathcal{P}_1}$ is slightly less than $V_{\mathcal{P}_2}$, it is probable that under some configurations \mathcal{P}_1 outperforms \mathcal{P}_2 . As shown in Table 12 when $k = 25$ or $k = 100$ under SC-DV, \mathcal{P}_2 requires slightly more observations than \mathcal{P}_1 .

3.4.2 Correlated Case

We tested four different correlation factors (ρ): 0.02, 0.25, 0.50, and 0.75. In [22], it is shown that a minimum correlation required for a CRN case to outperform an independent case

Table 13: Sample Average of the Total Number of Observations in \mathcal{KN} , \mathcal{P}_1 , and \mathcal{P}_2 when $n_0 = 24, \delta = 1/\sqrt{n_0}, 1 - \alpha = 0.95$, and $k = 25$ as a Function of ρ

ρ		SC-CV	MDM-CV	MDM-IV	MDM-DV	SC-IV	SC-DV
0.02	\mathcal{KN}	4926	1313	1965	1143	13241	2304
	\mathcal{P}_1	5001	1194	1628	1074	14226	2178
	\mathcal{P}_2	4509	1177	1612	1052	12342	2181
0.25	\mathcal{KN}	3804	1093	1554	983	10736	1792
	\mathcal{P}_1	3873	1017	1286	949	11021	1634
	\mathcal{P}_2	3513	1001	1290	912	10047	1715
0.50	\mathcal{KN}	2564	870	1118	813	7750	1288
	\mathcal{P}_1	2426	823	949	786	8001	1119
	\mathcal{P}_2	2349	825	961	779	7225	1243
0.75	\mathcal{KN}	1289	682	749	667	4960	822
	\mathcal{P}_1	1125	659	692	649	4927	728
	\mathcal{P}_2	1218	663	698	655	4479	812

is approximately 0.02 for a triangular continuation region. We found that when $\rho = 0.02$, both procedures \mathcal{P}_1 and \mathcal{P}_2 also spend slightly smaller but very close number of observations on average compared to their independent cases in most of the configurations. Thus, the correlation in the amount of 0.02 also seems to be an approximate minimum amount of correlation required for the new procedures with CRN to be as efficient as independent cases. For $\rho = 0.25, 0.50$, or 0.75 , all experiment designs show a lot better performance compared to their independent cases, and the average total number of observations reduces as the correlation factor increases. The simulation results of \mathcal{KN} , \mathcal{P}_1 , and \mathcal{P}_2 for different ρ values when $n_0 = 24, \delta = 1/\sqrt{n_0}, 1 - \alpha = 0.95$, and $k = 25$ are given in Table 13.

Interestingly, when systems are simulated independently, \mathcal{P}_1 works better than \mathcal{KN} only under all the MDM configurations and some SC with decreasing variances for $k \geq 10$. However, as ρ increases, \mathcal{P}_1 shows better performance than \mathcal{KN} under more configurations. For example, when $\rho = 0.75$, \mathcal{P}_1 outperforms \mathcal{KN} in all the configurations. In Section 3.3, we show that the ratio of vertical intercepts of \mathcal{KN} and \mathcal{P}_1 is $V_{\mathcal{KN}}/V_{\mathcal{P}_1} = \eta/(\lambda_1\sqrt{\xi_1})$ and does not depend on sample variance of the difference between two systems $S_{i\ell}^2$. Therefore, this ratio does not change with the use of CRN. However, the horizontal intercepts of \mathcal{KN} and \mathcal{P}_1 ($H_{\mathcal{KN}}$ and $H_{\mathcal{P}_1}$) are a function of $S_{i\ell}^2$ and $S_{i\ell}^4$, respectively. This implies that if

there is a decrease in $S_{i\ell}^2$ due to the use of CRN, then the decrease in $H_{\mathcal{P}_1}$ is much larger than that in $H_{\mathcal{KN}}$. This explains why \mathcal{P}_1 catches up the performance of \mathcal{KN} even under SC when the effect of CRN become stronger.

The performance of \mathcal{P}_2 is better than that of \mathcal{KN} under all configurations for all values of ρ we tested. This is similar to the independent case.

For small ρ , \mathcal{P}_2 is still better than \mathcal{P}_1 like the independent case. However, as ρ increases, \mathcal{P}_1 starts to show slightly better results than \mathcal{P}_2 . This is again because of the fact that the horizontal intercepts of \mathcal{P}_1 and \mathcal{P}_2 ($H_{\mathcal{P}_1}$ and $H_{\mathcal{P}_2}$) are a function of $S_{i\ell}^4$ and $S_{i\ell}^2$, respectively.

3.5 Conclusion

In this chapter, we propose two fully sequential procedures with parabolic boundaries for IID normal data when variances are unknown and unequal. Our procedures are appropriate for the use in simulation environments unlike the procedure in [37] that assumes known variances. The procedures allow for the use of CRN. The \mathcal{P}_2 procedure performs better than both \mathcal{KN} and \mathcal{P}_1 in most cases in terms of the total number of observations. However, \mathcal{P}_1 shows competitive performance compared to \mathcal{KN} and \mathcal{P}_2 , when k is large and the effect of CRN is strong.

CHAPTER IV

FEASIBILITY

Significant amount of work has been performed for *selection-of-the-best* problem in the simulation community ([17] and [23] for recent literature). However, very little work has been done for finding the best system in the presence of stochastic constraints on some secondary performance measures. In [6], multiple performance measures are handled, and in [32], a two-stage procedure with a constraint on variance is proposed. However, their methods are either very hard to apply in practice or focused only on a special case. In [2], a R&S procedure that determines the feasibility of systems in the presence of one stochastic constraint is presented and combined with a selection-of-the-best procedure to identify the best feasible system. Their feasibility check procedure can handle a general stochastic constraint on a secondary performance measure, and determines the feasibility of systems in consideration correctly with high probability. Yet, the work needs to be further extended to the case of multiple constraints.

In this chapter, we tackle the problem of determining a set of feasible or near-feasible systems that satisfy a number of stochastic constraints, especially when the number of systems or constraints is large. Determining the best feasible system in the presence of multiple constraints is beyond the scope of this study, and it is a topic of ongoing research, for which the main results of this chapter will serve as a critical step.

Let \mathcal{F} be a R&S procedure that checks the feasibility of *one* system in the presence of *one* stochastic constraint with pre-specified probability of correct decision (PCD). Any \mathcal{F} procedure can easily be extended to the case of multiple systems and constraints by the use of the Bonferroni inequality. We call the extended procedure \mathcal{F}_B . Unfortunately, \mathcal{F}_B tends to be conservative, and this conservatism becomes more serious as the number of systems or constraints increases. To lessen this problem, we present a screening procedure that accelerates the elimination of infeasible systems. The idea is to re-use collected observations

by taking a linear combination of the observations across the stochastic constraints for each system.

How does this idea help accelerating the elimination of infeasible systems? When there are multiple constraints, a system is feasible only when all the constraints are satisfied. If at least one constraint is violated, then the system is infeasible and eliminated immediately regardless of the feasibility of the other constraints. If a system is clearly infeasible for only one or a few constraints but feasible for the other constraints, then it might be more effective to keep a feasibility check procedure for each constraint separately. However, if a system is slightly infeasible for several constraints, then an aggregation of collected observations across all the constraints would make it easier to detect the infeasibility of the system. From this motivation, we design an accelerated procedure that performs elimination based on aggregated observations across all the constraints as well as observations corresponding each constraint for each system.

Finally, to further improve the efficiency of the accelerated procedure, we consider variance updating. In a variance updating procedure, variance estimates are updated as more observations are obtained. In [24], a variance-updating R&S procedure — called $\mathcal{KN}++$ — for a steady-state simulation is presented. This procedure uses raw observations from a single replication as basic observations. They showed that $\mathcal{KN}++$ is asymptotically valid and performs significantly better than a procedure without variance update. In [27], $\mathcal{KN}++$ is applied to IID normal data. They found that (i) the procedure shows significant savings from 20% up to 80% in terms of the number of observations compared to a corresponding non-updating version; (ii) the procedure is still asymptotically valid; (iii) the procedure does not guarantee a correct selection with a pre-specified probability for a finite sample size even for IID normal data; but (iv) the degradation from the nominal probability requirement is not significant.

The chapter is organized as follows: Section 4.1 formulates our problem and gives notation and definition. Section 4.2 provides the generic algorithms of \mathcal{F} , \mathcal{F}_B , and an accelerated version of \mathcal{F}_B with and without variance updating. Example procedures are presented in Section 4.3 by extending Algorithm I of [2] to the case of multiple constraints. Finally,

we compare the performance of the proposed procedures by experimental results based on multivariate normal random variables and a simple queueing model in Section 4.4, followed by a conclusion in Section 4.5.

4.1 Problem

In this section, we define our problem and notation. Our problem is to determine a set of feasible or near-feasible systems from k simulated systems that satisfy s stochastic constraints. Let $Y_{i\ell j}$, for $i = 1, 2, \dots, k$, $\ell = 1, 2, \dots, s$, $j = 1, 2, \dots$, denote an observation from the j th replication associated with the ℓ th performance measure (or the ℓ th constraint) from the i th system, and let $\mathbf{Y}_{ij} = (Y_{i1j}, Y_{i2j}, \dots, Y_{isj})'$ be the vector of the j th observations across all s performance measures from system i . The expected performance measures of system i are defined as $\mathbf{y}_i = E[\mathbf{Y}_{ij}] = (y_{i1}, y_{i2}, \dots, y_{is})'$ where $E[Y_{i\ell j}] = y_{i\ell}$ for $\ell = 1, 2, \dots, s$ and $j = 1, 2, \dots$. Our assumption is that \mathbf{Y}_{ij} satisfies the following:

Assumption 1

$$\mathbf{Y}_{ij} = \begin{bmatrix} Y_{i1j} \\ Y_{i2j} \\ \vdots \\ Y_{isj} \end{bmatrix} \stackrel{\text{IID}}{\sim} \text{MN}(\mathbf{y}_i, \boldsymbol{\Sigma}_i),$$

where \sim and **MN** represent ‘are distributed as’ and multivariate normal, respectively, and $\boldsymbol{\Sigma}_i$ is the variance-covariance matrix of \mathbf{Y}_{ij} .

The observations $Y_{i\ell j}$ tend to be normally distributed when we take within-replication averages or batch means as basic observations. Also, observations corresponding different performance measures from a system are likely to be correlated in reality, such as total inventory cost versus the total number of back orders for an inventory system. Therefore, the assumption of multivariate normal random variables is plausible. The vectors, \mathbf{Y}_{ij} and $\mathbf{Y}_{\nu j}$ can be dependent for $i \neq \nu$ (i.e., CRN across systems are possible). However, since feasibility check does not require comparison among systems, proposed statistical procedures in this chapter will not benefit at all by the use of CRN. Nevertheless, we consider the case

of CRN as well as the independent case because the proposed procedures will eventually be embedded into a procedure for finding the best *feasible* system that requires comparison among systems.

Throughout the chapter, we write vectors in boldface, random variables in upper case, and their realizations in lower case.

A feasible system has the vector of mean performance measures smaller than or equal to a constant vector $\mathbf{Q} = (Q_1, Q_2, \dots, Q_s)'$. More specifically, system i is feasible if $\mathbf{y}_i \leq \mathbf{Q}$. Unfortunately, for stochastic systems, it is impossible to guarantee identifying all feasible systems that satisfy s stochastic constraints. Instead, in [2], *tolerance level* which is similar to the indifference-zone (IZ) parameter — minimum difference worth detecting — is introduced. We adopt the same approach in this chapter. For each constraint ℓ , $\ell = 1, 2, \dots, s$, a decision maker will be asked to give a range around Q_ℓ , say, (Q_ℓ^-, Q_ℓ^+) such that $Q_\ell^- \leq Q_\ell \leq Q_\ell^+$ and $Q_\ell^- < Q_\ell^+$. Let $\mathbf{Q}^- = (Q_1^-, \dots, Q_s^-)'$ and $\mathbf{Q}^+ = (Q_1^+, \dots, Q_s^+)'$. Then three regions are defined for the constraints we consider:

- $\mathbf{y}_i \leq \mathbf{Q}^-$: This is the *desirable region*. If a system is in this region, then it is feasible.
- $(\mathbf{y}_i < \mathbf{Q}^+) \setminus (\mathbf{y}_i \leq \mathbf{Q}^-)$: This is the *acceptable region*. If a system is in this region, it is either feasible or infeasible and can be declared feasible or infeasible regardless of its true feasibility by our procedure.
- $(y_{i1} \geq Q_1^+) \cup (y_{i2} \geq Q_2^+) \cup \dots \cup (y_{is} \geq Q_s^+)$: This is the *unacceptable region*. If a system is in this region, then it is infeasible and should be eliminated.

Furthermore, we can define the following three sets for the constraints in consideration:

$$\begin{aligned} S_D &= \text{the set of all desirable systems;} \\ S_A &= \text{the set of all acceptable systems; and} \\ S_U &= \text{the set of all unacceptable systems.} \end{aligned}$$

For given Q_ℓ^- and Q_ℓ^+ , we define q_ℓ and ϵ_ℓ as $q_\ell = (Q_\ell^- + Q_\ell^+)/2$ and $\epsilon_\ell = (Q_\ell^+ - Q_\ell^-)/2$, respectively. Then one can easily see that $Q_\ell^- = q_\ell - \epsilon_\ell$ and $Q_\ell^+ = q_\ell + \epsilon_\ell$. The parameter

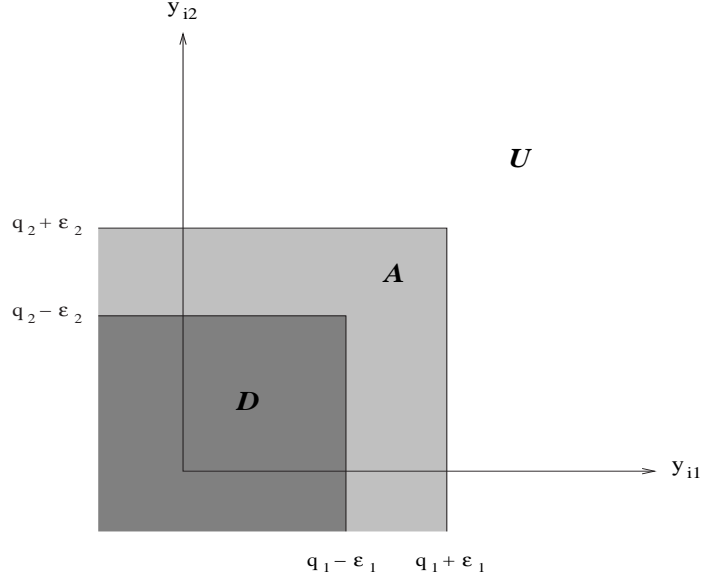


Figure 12: Desirable (D), Acceptable (A), and Unacceptable (U) Regions when There Exist Two Stochastic Constraints

q_ℓ , called the *target value* of the ℓ th constraint, behaves as a cut-off point between feasible and infeasible systems for the ℓ th constraint. The parameter ϵ_ℓ is the *tolerance level* of the ℓ th constraint, indicating how much we are willing to be off and above from q_ℓ . Then $\mathbf{q} = (q_1, q_2, \dots, q_s)'$ and $\mathbf{E} = (\epsilon_1, \epsilon_2, \dots, \epsilon_s)'$ represent the vectors of target values and tolerance levels for the s constraints, respectively.

Figure 12 shows the desirable (D), acceptable (A), and unacceptable (U) regions in terms of q_ℓ and $\epsilon_\ell, \ell = 1, 2$, when there are two stochastic constraints. Our procedures will also be presented in terms of q_ℓ and ϵ_ℓ .

Finally, a correct decision (CD) is defined as the event that a procedure selects a set F such that $S_D \subset F \subset (S_D \cup S_A)$, and a statistically valid procedure should guarantee the following probability statement:

$$\text{PCD} \equiv \Pr\{\text{CD}\} = \Pr\{S_D \subset F \subset (S_D \cup S_A)\} \geq 1 - \alpha,$$

where $1 - \alpha$ is the nominal confidence level.

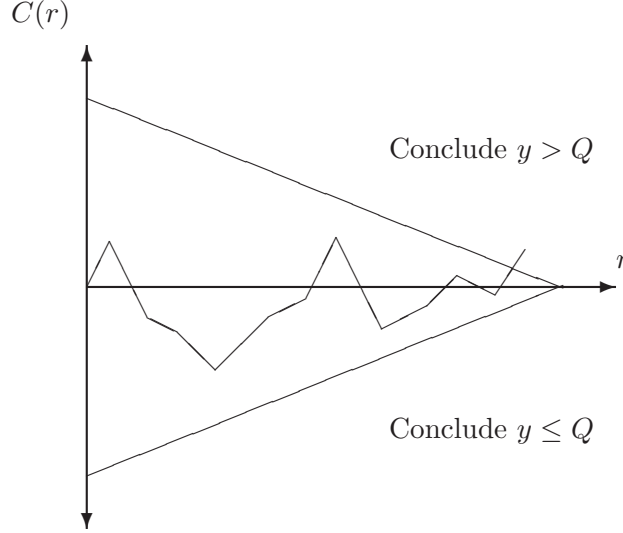


Figure 13: A Triangular Continuation Region for the Constraint $y \leq Q$

4.2 Generic Algorithms

In this section, we present generic algorithms for checking feasibility of k systems when there are s stochastic constraints. The purpose of this section is to provide a framework that helps one extend any procedure that works for one system and one constraint to more general cases.

4.2.1 Multiple Feasibility Check Procedure (\mathcal{F}_B)

In this subsection, we present the generic algorithm of \mathcal{F} for the case of one system and one constraint. Then we show how it can be extended to the case of multiple constraints and multiple systems.

The procedure \mathcal{F} requires tolerance level ϵ and target value q for the constraint in consideration, $y \leq Q$. The procedure has a monitoring statistic $C(r)$ of the observations from the system where r is the current sampling stage. The procedure also requires $R(r; \epsilon, h(\cdot), S^2(\cdot))$ which is a non-negative real-valued function of r that takes as parameters the tolerance level for the constraint ϵ , a non-negative value h that depends on other parameters, and the usual sample variance S^2 of a number of observations.

The boundary $(-R(r; \epsilon, h(\cdot), S^2(\cdot)), R(r; \epsilon, h(\cdot), S^2(\cdot)))$ defines a so called continuation region for the procedure. In our setting, \mathcal{F} is a fully sequential procedure where one basic

Generic Algorithm of \mathcal{F} (one system and one constraint)

Setup: Choose confidence level $1 - \alpha$, target value q , tolerance level ϵ , and first stage sample size $n_0 \geq 2$.

Initialization: Obtain n_0 observations from the system and set $r = n_0$. Compute $S^2(n_0)$ and $h(n_0)$ to determine $R(r; \epsilon, h(n_0), S^2(n_0))$ such that $\text{PCD} \geq 1 - \alpha$.

Feasibility Check: If $C(r) \geq +R(r; \epsilon, h(n_0), S^2(n_0))$, declare that the system is infeasible. Else if $C(r) \leq -R(r; \epsilon, h(n_0), S^2(n_0))$, declare that the system is feasible. Else set $r = r + 1$.

Stopping Rule: Continue [Feasibility Check] until a feasibility decision is made for the system.

Figure 14: Algorithmic Statement of \mathcal{F}

observation is sampled at each stage and sampling continues as long as $C(r)$ stays within the continuation region. Otherwise, the procedure stops and a decision is made depending on through which side of the boundary the exit occurs. For example, Figure 13 shows a triangular continuation region where the horizontal and vertical axes in the figure denote stage number, r , and $C(r)$, respectively. If the exit occurs through the upper boundary, we conclude that the system is infeasible. On the other hand, if the exit occurs through the lower boundary, we conclude that the system is feasible. This continuation region is set up in a way that the actual PCD is guaranteed to be at least $1 - \alpha$ through the choice of an appropriate value of $h(\cdot)$. The generic algorithm of \mathcal{F} is given in Figure 14.

To extend \mathcal{F} to general cases, we need \mathcal{F} satisfies the following:

Assumption 2 *When there is only one system with one constraint, $h(n_0)$ can be determined such that \mathcal{F} with the continuation region $R(r; \epsilon, h(n_0), S^2(n_0))$ satisfies $\text{PCD} \geq 1 - \alpha$.*

In the rest of this section, all the lemmas and corollaries will be given under the assumptions that \mathbf{Y}_{ij} and \mathcal{F} satisfy Assumptions 1 and 2, respectively.

Now, we extend \mathcal{F} to the case of k systems and s constraints. Let $C_{i\ell}(r)$ represent the monitoring statistic of observations $Y_{i\ell j}$ for $j = 1, \dots, r$ for the feasibility check of the ℓ th constraint of system i ; $S_{i\ell}^2(n_0)$ be the usual sample variance of the observations $Y_{i\ell j}$ for $j = 1, \dots, n_0$ from the ℓ th constraint of system i ; $R(r; \epsilon_\ell, h(n_0), S_{i\ell}^2(n_0))$ denote a certain

Generic Algorithm of \mathcal{F}_B (k systems and s constraints)

Setup: Choose confidence level $1 - \alpha$, target value q_ℓ , tolerance level ϵ_ℓ for $\ell = 1, \dots, s$, and first stage sample size $n_0 \geq 2$.

Initialization: Obtain n_0 observations from each system and set $r = n_0$. Compute $S_{i\ell}^2(n_0)$ for each constraint of each system and $h(n_0)$ to determine $R(r; \epsilon_\ell, h(n_0), S_{i\ell}^2(n_0))$ such that $\Pr\{\text{CD}_{i\ell}\} \geq 1 - \beta$.

Feasibility Check: For each system i , if there exist at least one ℓ such that $C_{i\ell}(r) \geq +R(r; \epsilon_\ell, h(n_0), S_{i\ell}^2(n_0))$, declare that the system is infeasible. Else if $C_{i\ell}(r) \leq -R(r; \epsilon_\ell, h(n_0), S_{i\ell}^2(n_0))$ for all ℓ , declare that the system is feasible. Else set $r = r + 1$.

Stopping Rule: Continue [Feasibility Check] until a feasibility decision is made for all systems.

Figure 15: Algorithmic Statement of \mathcal{F}_B

shaped boundary for the ℓ th constraint of system i ; and $\text{CD}_{i\ell}$ and $\text{ICD}_{i\ell}$ are the correct and incorrect decision events, respectively, when the ℓ th constraint of system i is considered in isolation. Then \mathcal{F}_B is given as in Figure 15.

Lemma 4 *If $\beta = \alpha/(ks)$, then \mathcal{F}_B satisfies $\text{PCD} \geq 1 - \alpha$.*

Proof: One can notice that \mathcal{F}_B is same as applying \mathcal{F} to each constraint of each system in isolation with $h(n_0)$ such that $\Pr\{\text{CD}_{i\ell}\} \geq 1 - \beta$. Then

$$\begin{aligned}
 \text{PCD} &= \Pr\{\cap_{i=1}^k \cap_{\ell=1}^s \text{CD}_{i\ell}\} \\
 &\geq 1 - \sum_{i=1}^k \sum_{\ell=1}^s \Pr\{\text{ICD}_{i\ell}\} \\
 &\geq 1 - \sum_{i=1}^k \sum_{\ell=1}^s \beta \\
 &= 1 - ks\beta \\
 &= 1 - ks \frac{\alpha}{ks} = 1 - \alpha,
 \end{aligned}$$

where the first inequality follows from the Bonferroni inequality. \square

Corollary 2 *If each system is simulated independently without the use of CRN across systems, then $\beta = (1 - (1 - \alpha)^{1/k})/s$ satisfies $\text{PCD} \geq 1 - \alpha$.*

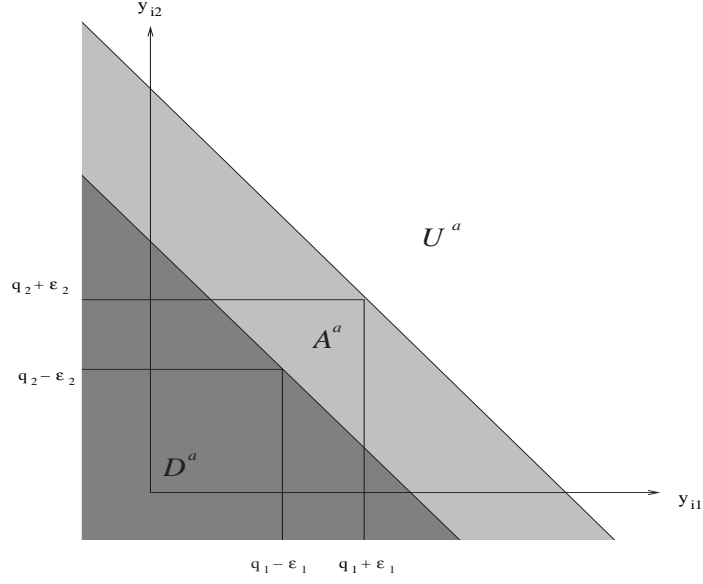


Figure 16: D^a , A^a , and U^a Regions for the Aggregated Measure when There Exist Two Stochastic Constraints

Proof:

$$\begin{aligned}
\text{PCD} &= \Pr\{\cap_{i=1}^k \cap_{\ell=1}^s \text{CD}_{i\ell}\} \\
&= \prod_{i=1}^k \Pr\{\cap_{\ell=1}^s \text{CD}_{i\ell}\} \\
&\geq \prod_{i=1}^k \left(1 - \sum_{\ell=1}^s \Pr\{\text{ICD}_{i\ell}\}\right) \\
&\geq (1 - s\beta)^k \\
&= \left(1 - (1 - (1 - \alpha)^{1/k})\right)^k = 1 - \alpha,
\end{aligned}$$

where the first inequality follows from the Bonferroni inequality. \square

4.2.2 Accelerated Feasibility Check Procedure (\mathcal{F}_A)

When the number of systems or constraints is large, \mathcal{F}_B becomes conservative in terms of the number of observations required until a decision is made and the actual PCD, mainly due to the Bonferroni inequality. To lessen this conservatism, we develop a screening procedure in which basic observations across constraints are aggregated into one observation by a linear combination. More specifically, an aggregated observation Y_{ij}^a is defined as $Y_{ij}^a = \mathbf{A}'\mathbf{Y}_{ij}$ where $\mathbf{A} = (a_1, a_2, \dots, a_s)'$ is a vector of positive constants. Since $\mathbf{Y}_{ij}, j = 1, 2, \dots$, are

Generic Algorithm of \mathcal{F}_A (k systems and s constraints)

Setup: Choose confidence level $1-\alpha$, target value q_ℓ , and tolerance level ϵ_ℓ for $\ell = 1, \dots, s$, and first stage sample size $n_0 \geq 2$. Also, calculate q^a and ϵ^a .

Initialization: Obtain n_0 observations from each system and set $r = n_0$. Compute $S_{i\ell}^2(n_0)$ and $S_i^2(n_0)$. Also, calculate $h_0(n_0)$ to determine $R^a(r; \epsilon^a, h_0(n_0), S_i^2(n_0))$ such that $\Pr\{\text{CD}_i^a\} \geq 1 - \beta_0$, and $h_1(n_0)$ to determine $R(r; \epsilon_\ell, h_1(n_0), S_{i\ell}^2(n_0))$ such that $\Pr\{\text{CD}_{i\ell}\} \geq 1 - \beta_1$.

Feasibility Check: For each system i , if there exist at least one ℓ such that $C_{i\ell}(r) \geq +R(r; \epsilon_\ell, h_1(n_0), S_{i\ell}^2(n_0))$ or $C_i^a(r) \geq +R^a(r; \epsilon^a, h_0(n_0), S_i^2(n_0))$, declare that the system is infeasible. Else if $C_{i\ell}(r) \leq -R(r; \epsilon_\ell, h_1(n_0), S_{i\ell}^2(n_0))$ for all ℓ , declare that the system is feasible. Else set $r = r + 1$.

Stopping Rule: Continue [Feasibility Check] until a feasibility decision is made for all systems.

Figure 17: Algorithmic Statement of \mathcal{F}_A

assumed to be IID multivariate normal, the aggregated observations $Y_{ij}^a, j = 1, 2, \dots$, are also IID normal. Therefore, we can apply \mathcal{F}_B directly to the aggregated observations with aggregated tolerance level $\epsilon^a = \mathbf{A}'\mathbf{E}$ and aggregated target value $q^a = \mathbf{A}'\mathbf{q}$.

The difficulty is that aggregation results in different desirable, acceptable, and unacceptable regions from those defined by the original individual constraints. More specifically, let D^a , A^a , and U^a denote desirable, acceptable, and unacceptable regions defined by aggregation with a vector \mathbf{A} . Shaded triangles of Figure 16 show the three regions for the aggregated constraint with Y_{ij}^a , ϵ^a and q^a while rectangles show those for the original individual constraints when there are two constraints. As one can see from Figure 16, systems in U^a and A^a fall into the unacceptable region U or acceptable region A in terms of the original constraints, so the screening procedure with aggregated observations is likely to eliminate systems in U and A only. However, D^a and A^a contain some unacceptable systems in terms of the original constraints. Therefore, it is possible that a system declared feasible by the screening procedure with aggregated observations is actually an unacceptable system in U . This implies that we can confidently eliminate a system if the system is

declared as infeasible by the screening procedure with aggregated observations, but a decision that a system is feasible by the screening procedure is untrustworthy because D^a and A^a may contain some unacceptable systems in terms of the original constraints. Therefore, the screening procedure with aggregated observations can not be used solely to make the feasibility decision, but can help accelerate the elimination of unacceptable systems.

Let $S_i^2(n_0)$ represent the sample variance of Y_{ij}^a for $j = 1, \dots, n_0$; $C_i^a(r)$ be the monitoring statistic of aggregated observations Y_{ij}^a for $j = 1, 2, \dots, r$ of system i ; and $R^a(r; \epsilon^a, h(n_0), S_i^2(n_0))$ be the continuation boundary for the aggregated constraint of system i . Moreover, define CD_i^a as the event that a procedure based on aggregated observations Y_{ij}^a eliminates some unacceptable and acceptable systems, but none of the desirable systems are eliminated. The accelerated feasibility check procedure called \mathcal{F}_A that combines the screening procedure with aggregated observations with \mathcal{F}_B is given in Figure 17.

Remark: The choice of the vector \mathbf{A} is discussed in Appendix D.

Lemma 5 *If one chooses β_0 and β_1 such that $k\beta_0 + ks\beta_1 = \alpha$, then \mathcal{F}_A satisfies $\text{PCD} \geq 1 - \alpha$.*

Proof:

$$\begin{aligned}
\text{PCD} &= \Pr\{(\cap_{i=1}^k \cap_{\ell=1}^s CD_{i\ell}) \cap (\cap_{i=1}^k CD_i^a)\} \\
&\geq \Pr\{\cap_{i=1}^k \cap_{\ell=1}^s CD_{i\ell}\} + \Pr\{\cap_{i=1}^k CD_i^a\} - 1 \\
&\geq (1 - ks\beta_1) + (1 - k\beta_0) - 1 \\
&= 1 - \alpha,
\end{aligned}$$

where the second inequality follows from the Bonferroni inequality. \square

Corollary 3 *If systems are simulated independently without CRN across systems, and β_0 and β_1 are such that $(1 - \beta_0)^k + (1 - s\beta_1)^k = 2 - \alpha$, then \mathcal{F}_A satisfies $\text{PCD} \geq 1 - \alpha$.*

Proof:

$$\text{PCD} = \Pr\{(\cap_{i=1}^k \cap_{\ell=1}^s CD_{i\ell}) \cap (\cap_{i=1}^k CD_i^a)\}$$

Generic Algorithm of $\mathcal{F}_{\mathcal{A}}^+$ (k systems and s constraints)

Setup: Choose confidence level $1-\alpha$, target value q_ℓ , and tolerance level ϵ_ℓ for $\ell = 1, \dots, s$, and first stage sample size $n_0 \geq 2$. Also, calculate q^a and ϵ^a , and decide when to update variance estimates (e.g., every 10 observations).

Initialization: Obtain n_0 observations from each system and set $r = n_0$.

Update: Whenever the current stage r reaches next variance update point, update $S_{i\ell}^2(r)$ and $S_i^2(r)$ based on r observations. Also, calculate $h_0(r)$ and $h_1(r)$ as one would do in $\mathcal{F}_{\mathcal{A}}$ but with r not with n_0 .

Otherwise, all the parameters remain unchanged.

Feasibility Check: For each system i , if there exist at least one ℓ such that $C_{i\ell}(r) \geq +R(\epsilon_\ell, h_1(r), S_{i\ell}^2(r))$ or $C_i^a(r) \geq +R^a(\epsilon^a, h_0(r), S_i^2(r))$, declare that the system is infeasible. Else if $C_{i\ell}(r) \leq -R(\epsilon_\ell, h_1(r), S_{i\ell}^2(r))$ for all ℓ , declare that the system is feasible. Else set $r = r + 1$.

Stopping Rule: Continue [Update] and [Feasibility Check] until a feasibility decision is made for all systems.

Figure 18: Algorithmic Statement of $\mathcal{F}_{\mathcal{A}}^+$

$$\begin{aligned}
&\geq \Pr\{\cap_{i=1}^k \cap_{\ell=1}^s \text{CD}_{i\ell}\} + \Pr\{\cap_{i=1}^k \text{CD}_i^a\} - 1 \\
&= \prod_{i=1}^k \Pr\{\cap_{\ell=1}^s \text{CD}_{i\ell}\} + \prod_{i=1}^k \Pr\{\text{CD}_i^a\} - 1 \\
&\geq \prod_{i=1}^k \left(1 - \sum_{\ell=1}^s \Pr\{\text{ICD}_{i\ell}\}\right) + \prod_{i=1}^k \Pr\{\text{CD}_i^a\} - 1 \\
&\geq (1 - s\beta_1)^k + (1 - \beta_0)^k - 1 \\
&= 1 - \alpha,
\end{aligned}$$

where the second inequality comes from the Bonferroni inequality. \square

Lemma 5 and Corollary 3 imply that one can split the overall error α between screening procedures based on original observations $Y_{i\ell j}$ and aggregated observations Y_{ij}^a . This is very similar to the decomposition lemma of [29].

Before presenting extended procedures for multiple constraints, we need some more definitions:

$$\begin{aligned}
g(\eta, d) &\equiv \frac{1}{2} (1 + 2\eta)^{-d/2}, \\
R(r; a, b, c) &\equiv \max\left\{0, \frac{bc}{a} - \frac{a}{2}r\right\} \text{ for } a, b, c > 0.
\end{aligned}$$

Procedure \mathcal{F}_B^{AGK}

Setup: Choose confidence level $1 - \alpha$, vector of tolerance levels $\mathbf{E} = (\epsilon_1, \epsilon_2, \dots, \epsilon_s)'$, and first stage sample size $n_0 \geq 2$. Find η as the solution to the equation $g(\eta, n_0 - 1) = \beta$.

Initialization: Let $I = \{1, 2, \dots, k\}$, $F = \emptyset$, and $K_i = \emptyset$, $i = 1, 2, \dots, k$, be the set of undetermined systems, the set of systems declared ‘feasible’, and the set of constraint indexes of system i that are already checked as feasible for system i , respectively. Set $h(n_0) = \eta(n_0 - 1)$.

Obtain observations \mathbf{Y}_{ij} , $j = 1, 2, \dots, n_0$, from each system i . For each system i and constraint $\ell = 1, 2, \dots, s$, compute the sample variance

$$S_{i\ell}^2(n_0) = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (Y_{i\ell j} - \bar{Y}_{i\ell}(n_0))^2,$$

where $\bar{Y}_{i\ell}(n_0)$ is the sample average of the first n_0 observations associated with constraint ℓ from system i .

Set the number of observations $r = n_0$ and go to **Feasibility Check**.

Feasibility Check: For each $i \in I$ and any $\ell \notin K_i$, $\ell = 1, 2, \dots, s$, if

$$\sum_{j=1}^r (Y_{i\ell j} - q_\ell) \geq +R(r; \epsilon_\ell, h(n_0), S_{i\ell}^2(n_0)),$$

then eliminate i from I ; else if

$$\sum_{j=1}^r (Y_{i\ell j} - q_\ell) \leq -R(r; \epsilon_\ell, h(n_0), S_{i\ell}^2(n_0)),$$

then add ℓ to K_i .

For each $i \in I$, if $|K_i| = s$, then move i from I to F .

Stopping Rule: If $|I| = 0$, then return F as a set of feasible systems.

Otherwise, take one additional observation $\mathbf{Y}_{i, r+1}$ from each system $i \in I$. Then set $r = r + 1$ and go to **Feasibility Check**.

Figure 19: Algorithmic Statement of \mathcal{F}_B^{AGK}

4.2.3 Accelerated Feasibility Check Procedure with Variance Update (\mathcal{F}_A^+)

To further improve the performance of \mathcal{F}_A , we design a variance updating version of \mathcal{F}_A , which is called \mathcal{F}_A^+ . \mathcal{F}_A^+ is basically same as \mathcal{F}_A except that we update variance estimates as more observations become available. The generic algorithm of \mathcal{F}_A^+ is given in Figure 18.

For a finite sample size, \mathcal{F}_A^+ is heuristic, but in [27], it is shown that the degradation from the nominal confidence level $1 - \alpha$ is very minor for the selection-of-the-best problem when updating is applied to a statistically valid procedure for IID normal observations. Therefore, we expect the degradation from the nominal PCD in \mathcal{F}_A^+ to be insignificant as well when \mathcal{F}_A^+ is directly taken from a statistically valid \mathcal{F}_A .

Moreover, if $\liminf_{\epsilon^a \rightarrow 0} \Pr\{\text{CD}_i^a\} \geq 1 - \beta_0$ and $\liminf_{\epsilon_\ell \rightarrow 0} \Pr\{\text{CD}_{i\ell}\} \geq 1 - \beta_1$ hold, then it can be shown that \mathcal{F}_A^+ is asymptotically valid. However, for proving the asymptotic validity one needs to assume that

$$Y_{i\ell j} = y_{i\ell} + k_{i\ell j} \quad (47)$$

represent the output process from constraint ℓ of system i , where $\{k_{i\ell j}, j = 1, 2, \dots\}$ are IID mean-zero normal random variables. Then let $y_{i\ell} = q_\ell + \epsilon_\ell$ so that as $\epsilon_\ell \rightarrow 0$ the true mean goes to q_ℓ . Under this model, as the problem becomes more and more difficult, the procedure's PCD becomes at least as large as the desired PCD.

In reality, we do not know the true differences between the means $y_{i\ell}$ and the target values q_ℓ . The most important case in a feasibility detection procedure is when the differences are small and we demand to be able to detect small differences. The asymptotic validity of the \mathcal{F}_A^+ procedure shows that if the true, unknown differences are small, then the procedure will achieve approximately the desired PCD when we also require it to detect small differences.

4.3 Example Procedures

In this section, we construct example procedures of \mathcal{F}_B , \mathcal{F}_A , and \mathcal{F}_A^+ . The key in constructing a statistically valid procedure is to know how to determine a continuation region so that the probability requirement is satisfied. The continuation region can be triangular,

parabolic or any arbitrary shaped. In [22], a triangular continuation region for selection-of-the-best problem is used. In Chapter 3, it is shown that a parabolic continuation region can be more efficient in terms of the total number of observations than a triangular continuation region. As long as one knows how to determine the continuation region that works for one system and one constraint, she should be able to extend it to more general cases using Lemmas 4 and 5.

In [2], a procedure — called Algorithm I of \mathcal{AGK} — that checks the feasibility of k systems with one constraint correctly satisfying the PCD requirement is presented. By setting $k = 1$, we take their Algorithm I as our basic procedure \mathcal{F}^{AGK} and extend it to more general cases to get \mathcal{F}_B^{AGK} , \mathcal{F}_A^{AGK} , and \mathcal{F}_A^{AGK+} .

The procedure \mathcal{F}_B^{AGK} is described in Figure 19.

Theorem 8 *If the parameter β is set to $\beta = \alpha/ks$ when CRN are employed or $\beta = (1 - (1 - \alpha)^{1/k})/s$ when systems are simulated independently, then the \mathcal{F}_B^{AGK} procedure guarantees $\text{PCD} \geq 1 - \alpha$.*

Proof: By \mathcal{AGK} , $\Pr\{\text{CD}_{il}\} \geq 1 - g(\eta, n_0 - 1)$. Since we set $g(\eta, n_0 - 1) = \beta$, $\Pr\{\text{CD}_{il}\} \geq 1 - \beta$. Then by Lemma 4 or by Corollary 2, it is straight-forward to show that $\text{PCD} \geq 1 - \alpha$.

□

The procedure \mathcal{F}_A^{AGK} is given in Figure 20.

Theorem 9 *Suppose that the parameter β_0 and β_1 are set to $\beta_0 = \alpha_0/k$ and $\beta_1 = \alpha_1/(ks)$ when CRN are used. Or β_0 and β_1 are selected such that $\beta_0 = 1 - (1 - \alpha_0)^{1/k}$ and $\beta_1 = (1 - (1 - \alpha_1)^{1/k})/s$ when systems are simulated independently. Then \mathcal{F}_A^{AGK} guarantees $\text{PCD} \geq 1 - (\alpha_0 + \alpha_1)$.*

Proof: By \mathcal{AGK} , $\Pr\{\text{CD}_{il}\} \geq 1 - g(\eta_1, n_0 - 1)$. One can notice that the screening part for aggregated observations is basically Algorithm I of \mathcal{AGK} with ϵ^a , q^a , and Y_{ij}^a except that we ignore any ‘feasible’ decision made by the procedure. Therefore, if $i \in U^a$, then

$$\Pr\{\text{CD}_i^a\} = \Pr\{\text{system } i \text{ is declared as infeasible}\} \geq 1 - g(\eta_0, n_0 - 1);$$

Procedure $\mathcal{F}_{\mathcal{A}}^{AGK}$

Setup: Choose confidence level α_0 and α_1 , vector of tolerance levels $\mathbf{E} = (\epsilon_1, \epsilon_2, \dots, \epsilon_s)'$, and first stage sample size $n_0 \geq 2$. Compute $\epsilon^a = \mathbf{A}'\mathbf{E}$ and $q^a = \mathbf{A}'\mathbf{q}$ where $\mathbf{A} = [a_\ell]_{\ell=1,2,\dots,s}$ such that $a_\ell = \prod_{j=1, j \neq \ell}^s \epsilon_j$. Find η_0 as the solution to the equation $g(\eta_0, n_0 - 1) = \beta_0$ and η_1 as the solution to the equation $g(\eta_1, n_0 - 1) = \beta_1$.

Initialization: Let $I = \{1, 2, \dots, k\}$, $F = \emptyset$, and $K_i = \emptyset, i = 1, 2, \dots, k$, be the set of undetermined systems, the set of systems declared ‘feasible’, and the set of constraint indexes of system i that are already checked as feasible for system i , respectively. Set $h_0(n_0) = \eta_0(n_0 - 1)$ and $h_1(n_0) = \eta_1(n_0 - 1)$.

Obtain observations $\mathbf{Y}_{ij}, j = 1, 2, \dots, n_0$, from each system i . Compute $Y_{ij}^a = \mathbf{A}'\mathbf{Y}_{ij}$. For each system i , compute the sample variance

$$S_i^2(n_0) = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (Y_{ij}^a - \bar{Y}_i^a(n_0))^2,$$

where $\bar{Y}_i^a(n_0)$ is the sample average of the first n_0 aggregated observations from system i .

For each system i and constraint $\ell = 1, 2, \dots, s$, compute the sample variance

$$S_{i\ell}^2(n_0) = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (Y_{i\ell j} - \bar{Y}_{i\ell}(n_0))^2,$$

where $\bar{Y}_{i\ell}(n_0)$ is the sample average of the first n_0 observations associated with constraint ℓ from system i .

Set the number of observations $r = n_0$ and go to **Feasibility Check**.

Feasibility Check: For each system $i \in I$, if $\sum_{j=1}^r (Y_{ij}^a - q^a) \geq +R(r; \epsilon^a, h_0(n_0), S_i^2(n_0))$, then eliminate i from I . Otherwise, for each $i \in I$ and any $\ell \notin K_i, \ell = 1, 2, \dots, s$, if $\sum_{j=1}^r (Y_{i\ell j} - q_\ell) \geq +R(r; \epsilon_\ell, h_1(n_0), S_{i\ell}^2(n_0))$, then eliminate i from I ; else if $\sum_{j=1}^r (Y_{i\ell j} - q_\ell) \leq -R(r; \epsilon_\ell, h_1(n_0), S_{i\ell}^2(n_0))$, then add ℓ to K_i .

For each $i \in I$, if $|K_i| = s$, then move i from I to F .

Stopping Rule: If $|I| = 0$, then return F as a set of feasible systems.

Otherwise, take one additional observation $\mathbf{Y}_{i,r+1}$ from each system $i \in I$ and let $Y_{i,r+1}^a = \mathbf{A}'\mathbf{Y}_{i,r+1}$. Set $r = r + 1$ and go to **Feasibility Check**.

Figure 20: Algorithmic Statement of $\mathcal{F}_{\mathcal{A}}^{AGK}$

and if $i \in D^a$, then

$$\Pr\{\text{CD}_i^a\} = \Pr\{\text{system } i \text{ is not declared as infeasible}\} \geq 1 - g(\eta_0, n_0 - 1).$$

Also, notice that U^a contains only some unacceptable systems. Thus, $\Pr\{\text{CD}_i^a\} \geq 1 -$

Procedure \mathcal{F}_A^{AGK+}

Setup: Choose confidence level α_0 and α_1 , vector of tolerance levels $\mathbf{E} = (\epsilon_1, \epsilon_2, \dots, \epsilon_s)'$, and first stage sample size $n_0 \geq 2$. Compute $\epsilon^a = \mathbf{A}'\mathbf{E}$ and $q^a = \mathbf{A}'\mathbf{q}$ where $\mathbf{A} = [a_\ell]_{\ell=1,2,\dots,s}$ such that $a_\ell = \prod_{j=1, j \neq \ell}^s \epsilon_j$.

Initialization: Let $I = \{1, 2, \dots, k\}$, $F = \emptyset$, and $K_i = \emptyset, i = 1, 2, \dots, k$, be the set of undetermined systems, the set of systems declared ‘feasible’, and the set of constraint indexes of system i that are already checked as feasible for system i , respectively.

Obtain observations $\mathbf{Y}_{ij}, j = 1, 2, \dots, n_0$, from each system i . Compute $Y_{ij}^a = \mathbf{A}'\mathbf{Y}_{ij}$. Set the number of observations $r = n_0$.

Update: If we have reached the next update point, then for each system $i \in I$, compute η_0 as the solution to the equation $g(\eta_0, r-1) = \beta_0$ and η_1 as the solution to the equation $g(\eta_1, r-1) = \beta_1$, and update $h_0(r) = \eta_0(r-1)$ and $h_1(r) = \eta_1(r-1)$. Also, for each system $i \in I$, compute the sample variance

$$S_i^2(r) = \frac{1}{r-1} \sum_{j=1}^r (Y_{ij}^a - \overline{Y_i^a}(r))^2,$$

where $\overline{Y_i^a}(r)$ is the sample average of the first r aggregated observations from system i . Finally, for each system i and constraint $\ell = 1, 2, \dots, s$, compute the sample variance

$$S_{i\ell}^2(r) = \frac{1}{r-1} \sum_{j=1}^r (Y_{i\ell j} - \overline{Y_{i\ell}}(r))^2,$$

where $\overline{Y_{i\ell}}(r)$ is the sample average of the first r observations associated with constraint ℓ from system i .

Feasibility Check: For each system $i \in I$, if $\sum_{j=1}^r (Y_{ij}^a - q^a) \geq +R(r; \epsilon^a, h_0(r), S_i^2(r))$, then eliminate i from I . Otherwise, for each $i \in I$ and any $\ell \notin K_i, \ell = 1, 2, \dots, s$, if $\sum_{j=1}^r (Y_{i\ell j} - q_\ell) \geq +R(r; \epsilon_\ell, h_1(r), S_{i\ell}^2(r))$, then eliminate i from I ; else if $\sum_{j=1}^r (Y_{i\ell j} - q_\ell) \leq -R(r; \epsilon_\ell, h_1(r), S_{i\ell}^2(r))$, then add ℓ to K_i .

For each $i \in I$, if $|K_i| = s$, then move i from I to F .

Stopping Rule: If $|I| = 0$, then return F as a set of feasible systems.

Otherwise, take one additional observation $\mathbf{Y}_{i,r+1}$ from each system $i \in I$ and let $Y_{i,r+1}^a = \mathbf{A}'\mathbf{Y}_{i,r+1}$. Set $r = r + 1$ and go to **Update**.

Figure 21: Algorithmic Statement of \mathcal{F}_A^{AGK+}

$g(\eta_0, n_0 - 1)$. By Lemma 5 or Corollary 3 and the way we choose the values of η_0 and η_1 , the theorem follows. \square

As shown in Theorem 9, \mathcal{F}_A^{AGK} guarantees CD with probability at least $1 - (\alpha_0 + \alpha_1)$. When the overall nominal confidence level is $1 - \alpha$, choosing α_0 and α_1 such that $\alpha_0 + \alpha_1 = \alpha$ guarantees CD with probability at least $1 - \alpha$. However, with this choice of α_0 and α_1 , \mathcal{F}_A^{AGK} might not always perform better than \mathcal{F}_B^{AGK} . For example, suppose that the overall nominal confidence level is 95% and we choose $\alpha_0 = \alpha_1 = 0.025$. If all the systems are in the desirable region D , elimination based on aggregated observations is not likely to be utilized and \mathcal{F}_A^{AGK} becomes very similar to \mathcal{F}_B^{AGK} except that \mathcal{F}_A^{AGK} uses a larger confidence level $1 - \alpha_1 = 0.975$ instead of $1 - \alpha = 0.95$. Thus, the performance of \mathcal{F}_A^{AGK} is likely to be worse than \mathcal{F}_B^{AGK} in this situation. However, it is possible that \mathcal{F}_A^{AGK} performs better than \mathcal{F}_B^{AGK} if there are a number of acceptable or unacceptable systems in terms of the original constraints: the saving from eliminating acceptable or unacceptable systems earlier by the screening part with aggregated observations might be large enough to compensate using a larger probability $1 - \alpha_1$ than $1 - \alpha$.

On the other hand, if we choose $\alpha_1 = \alpha$ and $\alpha_0 > 0$, then \mathcal{F}_A^{AGK} is guaranteed to perform better than \mathcal{F}_B^{AGK} in terms of the number of observations required until we reach a decision. However, the efficiency is achieved at the cost of PCD: the actual PCD for \mathcal{F}_A^{AGK} is now only guaranteed to be $\geq 1 - (\alpha_0 + \alpha_1)$. However, we know that \mathcal{F}_B^{AGK} and \mathcal{F}_A^{AGK} are already quite conservative and the actual PCD is usually larger than the nominal level $1 - (\alpha_0 + \alpha_1)$. So, if one chooses $0 < \alpha_0 \leq \alpha$ and $\alpha_1 = \alpha$, it will certainly help \mathcal{F}_A^{AGK} perform better than \mathcal{F}_B^{AGK} while the actual PCD is mostly $\geq 1 - \alpha$.

Finally, the variance updating version \mathcal{F}_A^{AGK+} is given in Figure 21. As discussed in Section 4.2.3, \mathcal{F}_A^{AGK+} is heuristic, but actually one can show that $\liminf_{\epsilon^a \rightarrow 0} \Pr\{CD_i^a\} \geq 1 - \beta_0$ and $\liminf_{\epsilon_\ell \rightarrow 0} \Pr\{CD_{i\ell}\} \geq 1 - \beta_1$ by arguments similar to the proofs of Algorithm I of \mathcal{AGK} and Theorem 2 of [24] when $\mathbf{Y}_{ij}, j = 1, 2, \dots$ satisfy Equation (47). Therefore, \mathcal{F}_A^{AGK+} satisfies the PCD requirement as tolerance levels approach to zero.

Table 14: The Mean Configurations of the Performance Measures Associated with s Stochastic Constraints

desirable	
D1	$y_\ell = -\epsilon, \ell = 1, 2, \dots, s$
D2	$y_\ell = -\ell\epsilon, \ell = 1, 2, \dots, s$
D3	$y_\ell = -10\epsilon, \ell = 1, 2, \dots, s$
acceptable	
A1	$y_1 = y_2 = -2\epsilon, y_\ell = -\epsilon/2, \ell = 3, 4, \dots, s$
A2	$y_\ell = 0, \ell = 1, 2, \dots, s$
A3	$y_\ell = \epsilon/2, \ell = 1, 2, \dots, s$
unacceptable	
U1	$y_1 = y_2 = -2\epsilon, y_\ell = \epsilon, \ell = 3, 4, \dots, s$
U2	$y_\ell = \epsilon, \ell = 1, 2, \dots, s$
U3	$y_\ell = \ell\epsilon, \ell = 1, 2, \dots, s$

4.4 Empirical Evaluation

In this section, we compare the performance of \mathcal{F}_B^{AGK} , \mathcal{F}_A^{AGK} , and \mathcal{F}_A^{AGK+} .

4.4.1 Multivariate Normal Example

Without loss of generality, we assume that $\mathbf{q} = (0, 0, \dots, 0)'$, and the tolerance level for each constraint is set to $\epsilon_\ell = 1/\sqrt{n_0}$.

Output vectors corresponding s constraints for system i are directly generated from a multivariate normal distribution with a mean vector \mathbf{y}_i and a positive definite variance-covariance matrix Σ_i .

We consider three mean configurations \mathbf{y}_i from each of desirable, acceptable, and unacceptable regions. The mean configuration of a system in the desirable region will take one of D1, D2, and D3 configurations. Similarly, A1 to A3 are for systems in the acceptable region, and U1 to U3 are for systems in the unacceptable region. These nine mean configurations are shown in Table 14. The D1, A1, and U1 configurations are respectively more difficult than the D2, A2, and U2 configurations, and D2, A2, and U2 are more difficult than D3, A3, and U3. For example, a system with the D1 configuration barely falls into the desirable region with \mathbf{y}_i whose elements are exactly equal to $-\epsilon$. On the other hand, the D2 or D3 configuration has much smaller expected performance measures than $-\epsilon$ for all constraints;

Table 15: Sample Average of Total Number of Vectors (SANV) when One System ($k = 1$) is Considered with $s = 5$, $\rho = 0.0$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$

	CV		IV		DV	
	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}
D1	72	71	125	122	124	121
D2	47	47	56	56	102	102
D3	11	11	18	18	17	18
A1	86	85	165	164	118	117
A2	52	26	81	37	81	37
A3	28	11	40	14	40	14
U1	25	20	47	33	31	24
U2	20	10	28	11	28	11
U3	11	10	16	10	12	10

Table 16: Estimated PCD when One System ($k = 1$) is Considered with $s = 5$, $\rho = 0.0$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$

	CV		IV		DV	
	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}
D1	0.963	0.955	0.960	0.935	0.960	0.935
D2	0.993	0.993	0.993	0.993	0.993	0.993
D3	1.000	1.000	1.000	1.000	1.000	1.000
U1	1.000	1.000	1.000	1.000	1.000	1.000
U2	1.000	1.000	1.000	1.000	1.000	1.000
U3	1.000	1.000	1.000	1.000	1.000	1.000

thus, it should be easier to detect that the system is a desirable system.

The s by s variance-covariance matrix of system i is Σ_i and it is assumed to have diagonal elements σ_ℓ^2 for $\ell = 1, 2, \dots, s$ and non-diagonal elements $\rho\sigma_\ell\sigma_{\ell'}$ for $\ell \neq \ell'$ and $\ell, \ell' = 1, \dots, s$. That is, system i generates \mathbf{Y}_{ij} , $j = 1, 2, \dots$, that are multivariate normally distributed with mean vectors \mathbf{y}_i and marginal variances σ_ℓ^2 for $\ell = 1, \dots, s$ and equal correlation ρ between each pair of constraints. The correlation ρ varies over $\rho = \{-0.15, 0.0, 0.3\}$, where the numbers are chosen to ensure that the variance-covariance matrix Σ_i is positive definite. The marginal variances σ_ℓ^2 take one of three configurations: constant variances (CV), increasing variances (IV), and decreasing variances (DV). The variances in the CV configuration are all set to 1. In IV and DV, the variance of each constraint $\ell = 1, 2, \dots, s$, is set to $1 + (\ell - 1)\epsilon$ and $1 + (s - \ell)\epsilon$, respectively.

Table 17: SANV when One System ($k = 1$) is Considered with $s = 5$, $\rho = -0.15$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$

	CV		IV		DV	
	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}
D1	72	72	126	125	126	125
D2	47	47	55	55	102	102
D3	11	11	18	18	18	18
A1	87	87	169	169	120	120
A2	49	22	74	27	75	27
A3	27	10	39	11	39	11
U1	25	20	46	32	30	23
U2	20	10	28	10	28	10
U3	11	10	16	10	12	10

Table 18: Estimated PCD when One System ($k = 1$) is Considered with $s = 5$, $\rho = -0.15$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$

	CV		IV		DV	
	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}
D1	0.959	0.959	0.956	0.953	0.959	0.956
D2	0.993	0.993	0.993	0.993	0.991	0.991
D3	1.000	1.000	1.000	1.000	1.000	1.000
U1	1.000	1.000	1.000	1.000	1.000	1.000
U2	1.000	1.000	1.000	1.000	1.000	1.000
U3	1.000	1.000	1.000	1.000	1.000	1.000

Table 19: SANV when One System ($k = 1$) is Considered with $s = 5$, $\rho = 0.3$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$

	CV		IV		DV	
	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}
D1	67	67	120	116	120	116
D2	46	46	54	54	100	100
D3	11	11	17	17	17	17
A1	83	82	161	158	114	115
A2	60	38	94	58	94	58
A3	31	17	44	24	44	24
U1	27	24	49	42	32	30
U2	22	12	30	17	30	17
U3	11	10	17	11	12	10

Table 20: Estimated PCD when One System ($k = 1$) is Considered with $s = 5$, $\rho = 0.3$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$

	CV		IV		DV	
	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}
D1	0.962	0.932	0.960	0.924	0.961	0.924
D2	0.991	0.991	0.994	0.994	0.992	0.992
D3	1.000	1.000	1.000	1.000	1.000	1.000
U1	1.000	1.000	1.000	1.000	1.000	1.000
U2	1.000	1.000	1.000	1.000	1.000	1.000
U3	1.000	1.000	1.000	1.000	1.000	1.000

The number of systems is $k = 1$ or 9, and the number of constraints is $s = 5$. The first-stage sample size is set to $n_0 = 10$ and the overall nominal confidence is $1 - \alpha = 0.95$. In \mathcal{F}_A^{AGK} , we first set $\alpha_0 = \alpha_1 = 0.05$ which means that \mathcal{F}_A^{AGK} only guarantees the actual PCD ≥ 0.90 and based on the discussion at the end of Section 4.3 this is the way we recommend setting the α values. In the second case, we take $\alpha_0 = 0.05/(k+1)$ and $\alpha_1 = 0.05k/(k+1)$ so that \mathcal{F}_A^{AGK} also guarantees the actual PCD ≥ 0.95 . We make 10,000 experiments (complete repetitions) and report estimated PCD and sample average of the total number of vectors \mathbf{Y}_{ij} (SANV), assuming that \mathbf{Y}_{ij} across s constraints are simultaneously obtained from a replication.

Tables 15 and 16 show SANV and the corresponding estimated PCD of \mathcal{F}_B^{AGK} and \mathcal{F}_A^{AGK} when $k = 1$, $s = 5$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$ with correlation across constraints $\rho = 0$. For a system with the D1, D2, or D3 configuration, both procedures require almost the same total number of vectors (replications) as expected, because (i) the value of α of \mathcal{F}_B^{AGK} is same as the value of α_1 of \mathcal{F}_A^{AGK} ; and (ii) the screening with aggregated observations of \mathcal{F}_A^{AGK} is unlikely to be utilized. For the A1, A2, A3, U1, U2, and U3 configurations, \mathcal{F}_A^{AGK} shows meaningful savings in the total number of vectors compared with \mathcal{F}_B^{AGK} in many cases. However, some — but not significant — degradation in PCD is possible as shown in Table 16. For example, for the D1 configuration with IV and DV, the estimated PCD is around 0.935 which is slightly smaller than the nominal value 0.95. However, for the other configurations, the estimated PCD are all higher than 0.95.

Table 21: SANV and Estimated PCD when Nine Systems ($k = 9$) are Considered with $s = 5$ and $\alpha = \alpha_0 = \alpha_1 = 0.05$

	$\rho = -0.15$			$\rho = 0.0$			$\rho = 0.3$		
	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_A^{AGK+}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_A^{AGK+}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_A^{AGK+}
CV	758	556	296	764	586	310	776	655	352
	0.994	0.994	0.991	0.994	0.992	0.989	0.994	0.991	0.989
IV	1221	896	439	1238	951	475	1257	1069	555
	0.993	0.993	0.992	0.993	0.990	0.989	0.995	0.989	0.989
DV	1277	972	472	1282	1022	507	1305	1131	580
	0.994	0.993	0.991	0.994	0.988	0.989	0.995	0.988	0.988

Table 22: SANV and Estimated PCD when Nine Systems ($k = 9$) are Considered with $s = 5$, $\alpha = 0.05$, $\alpha_0 = 0.05/(k + 1)$, and $\alpha_1 = 0.05k/(k + 1)$

	$\rho = -0.15$		$\rho = 0.0$		$\rho = 0.3$	
	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}	\mathcal{F}_B^{AGK}	\mathcal{F}_A^{AGK}
CV	758	612	764	676	776	763
	0.994	0.994	0.994	0.995	0.994	0.994
IV	1221	993	1238	1106	1257	1243
	0.993	0.994	0.993	0.994	0.995	0.995
DV	1277	1063	1282	1170	1305	1297
	0.994	0.995	0.994	0.995	0.995	0.994

Tables 17 and 18 show SANV and corresponding estimated PCD values when $k = 1$, $s = 5$, $\rho = -0.15$, and $\alpha = \alpha_0 = \alpha_1 = 0.05$. The results for $\rho = 0.3$ are given in Tables 19 and 20. The tables show that regardless of negative or positive correlation across constraints, \mathcal{F}_A^{AGK} performs as well as or better than \mathcal{F}_B^{AGK} at the cost of slight degradation in PCD in some cases.

Now, we consider $k = 9$ systems, one from each mean configuration D1 through U3. Variances of the nine systems are assumed to all follow one of CV, IV, and DV configurations. We apply \mathcal{F}_B^{AGK} , \mathcal{F}_A^{AGK} , and \mathcal{F}_A^{AGK+} to compare the performance of the three procedures.

In Table 21, the SANV and estimated PCD when $\alpha = \alpha_0 = \alpha_1 = 0.05$ are given. As it is seen, estimated PCD are all larger than the nominal value 0.95 and \mathcal{F}_A^{AGK} spends fewer vectors for all configurations tested. This is expected since we have three systems from the acceptable and unacceptable systems. If all systems are in the desirable region, then

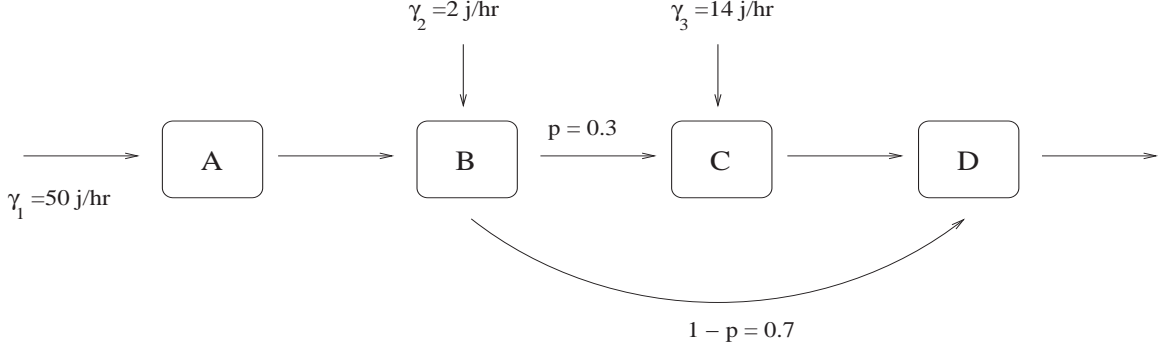


Figure 22: Job Shop Example

\mathcal{F}_A^{AGK} spends about the same number of replications as \mathcal{F}_B^{AGK} , which we do not include in this chapter. In reality, if k is large, it is very unlikely that all the systems are in the desirable region. Therefore, we expect \mathcal{F}_A^{AGK} to be more efficient than \mathcal{F}_B^{AGK} with very little or no degradation in PCD when k is large. The variance updating version, \mathcal{F}_A^{AGK+} , can further increase the efficiency of the procedure. For example, when $\rho = 0$ with the CV configuration, \mathcal{F}_A^{AGK+} spends only 310 replications with the estimated PCD equal to 0.989 while \mathcal{F}_B^{AGK} and \mathcal{F}_A^{AGK} spend 764 and 586 replications, respectively.

In Table 22, the SANV and estimated PCD when $\alpha = 0.05$, $\alpha_0 = 0.05/(k+1)$, and $\alpha_1 = 0.05k/(k+1)$ are given. As it is seen, \mathcal{F}_A^{AGK} again spends fewer vectors than \mathcal{F}_B^{AGK} for all configurations tested, but the difference is less compared to the previous case because of the way we set α_0 and α_1 .

4.4.2 Queueing Example

In this section, we consider a small queueing example to evaluate the performance of \mathcal{F}_A^{AGK} , \mathcal{F}_B^{AGK} , and \mathcal{F}_A^{AGK+} . There is a small job shop with nine agile workers and four stations named Station A, B, C, and D. The 30% of items arriving at Station A move along Station B \rightarrow Station C \rightarrow Station D. The rest of items (70% of them) move along Station B \rightarrow Station D. Moreover, Stations B, C, and D have external arrivals. The diagram of this job shop is shown in Figure 22.

We want to determine feasible allocations of the nine agile workers over four stations when there are constraints that the expected average number of jobs waiting in queue at

Table 23: External Arrival and Service Rates for the Job Shop Example

Station	γ_ℓ (jobs/hr)	μ_ℓ (jobs/worker/hr)
1	50	27
2	2	55
3	14	31
4	0	80

Table 24: Average Buffer Size and Tolerance Levels for the Job Shop Example

Station	q_ℓ (jobs)	ϵ_ℓ (jobs)
1	12	0.5
2	1	0.1
3	1	0.1
4	5	0.1

each station should be smaller than or equal to a pre-determined average buffer space for the station. We assume that the number of external arrivals to each station follows a Poisson distribution with rate γ_ℓ and service times of each worker at each workstation is exponentially distributed with rate μ_ℓ for $\ell = 1, 2, 3, 4$. The values of γ_ℓ and μ_ℓ for each station ℓ are given in Table 23. The total service rate at a station is the service rate per worker times the number of assigned workers. For example, the service rate of a worker at Station A is 27 jobs/hour. If two workers are assigned, then the total service rate at Station A becomes 54 jobs/hr.

We assume that at least one worker should be assigned to each station. Then there are 56 different ways to allocate the nine agile workers over the four stations. Among those configurations, 35 of them are stable — i.e., the expected number of jobs waiting in queue does not blow up in steady-state. In this experiment, we consider only stable configurations (thus, $k = 35$). Among those configurations, 25 are unacceptable, 6 are acceptable, and 4 are desirable. The pre-determined average buffer size (q_ℓ) and the tolerance level (ϵ_ℓ) for each station are given in Table 24.

Since all service times and external inter-arrival times are assumed to be exponentially distributed, this job shop is a Jackson network. Therefore, the expected average number of

jobs waiting in the queue of each station can be analytically obtained. This enables us to check if a tested procedure makes a correct decision about the feasibility of a system.

The nominal confidence level $1 - \alpha$ is set to 0.95 and we set $1 - \alpha_0 = 1 - \alpha_1 = 0.95$ for \mathcal{F}_A^{AGK} . The first-stage sample size is $n_0 = 10$. For $k = 35$ systems, all three procedures were able to make a correct decision by returning a set F containing all the desirable systems and some acceptable systems only. However, to reach the decision, \mathcal{F}_B^{AGK} required 1980 replications while \mathcal{F}_A^{AGK} and \mathcal{F}_A^{AGK+} spent 899 and 624 replications, respectively, giving more than 50% savings in the number of replications compared with \mathcal{F}_B^{AGK} .

4.5 Conclusion

In discrete optimization with several stochastic constraints for a complicated model, the first step is to identify a set of feasible or near-feasible systems. We present three generic procedures that are useful for the purpose. The proposed generic procedures can incorporate a continuation region of any shape, provided that one knows how to set up a continuation region of the shape to deliver the PCD requirement when there is only one system with one stochastic constraint. Then the procedure can be extended to \mathcal{F}_B that can handle multiple systems and constraints. The performance of \mathcal{F}_B can be accelerated by a procedure called \mathcal{F}_A that reuses observations across constraints by applying screening to aggregated observations. Variance updating further helps improving the efficiency of \mathcal{F}_A quite a bit.

The next step is to solve a problem of finding the best feasible system. Or one may be interested in investigating feasibility check procedures with different continuation regions other than a triangular one. These are the subjects of ongoing research.

CHAPTER V

CONTRIBUTIONS

We present a new variance parameter estimator for steady-state simulations with competitive or better statistical properties compared to the existing estimators in the literature. By using a rebatching technique, we further lower the bias and variance of this estimator compared to its batched version.

We present two fully sequential indifference-zone procedures for the selection-of-the-best system among a finite number simulated systems when the variances are unknown and unequal and observations within systems are IID normal. These procedures are shown to be efficient in terms of the sampling cost when the number of systems is large or CRN are employed.

We present three generic procedures for determining a set of feasible systems in the presence of multiple stochastic constraints, especially when the number of systems or constraints is large. The proposed generic procedures can incorporate with a continuation region of any shape, provided that one knows how to set up a continuation region of the shape to deliver the PCD requirement when there is only one system with one stochastic constraint. Currently available selection procedures can handle at most one stochastic constraint. However, practical optimization problems may include a large number of stochastic constraints. In this context, our proposed procedure serves as a critical step to move from optimization with one stochastic constraint to the problem with multiple constraints. Consequently, simulation can be considered as a solution technique for more practical and meaningful optimization problems in the world.

APPENDIX A

REBATCHED ESTIMATORS

Derivation of Equations (28)–(33). First we prove the expectation results. Examples 3, 4, and 5 imply that

$$\begin{aligned} E[D^{i,j}(h_0; 2^i, n/2^i)] &= \sigma^2 + 7\gamma 2^i/n + o(1/n), \\ E[D_{J,0.5}^{i,j}(2^i, n/2^i)] &= \sigma^2 + o(1/n), \\ E[\tilde{D}_{J,0.5}^{i,j}(2^i, n/2^i)] &= \sigma^2 + o(1/n), \end{aligned} \tag{48}$$

for $i = 0, 1, 2$ and $j = 1, 2, \dots, 2^i$. Hence,

$$\begin{aligned} E[D^R(h_0; n, 2)] &= \frac{1}{7}(E[D^{01}(h_0; 1, n)] + E[D^{11}(h_0; 2, n/2)] + E[D^{12}(h_0; 2, n/2)] + E[D^{21}(h_0; 4, n/4)] \\ &\quad + E[D^{22}(h_0; 4, n/4)] + E[D^{23}(h_0; 4, n/4)] + E[D^{24}(h_0; 4, n/4)]) \\ &= \sigma^2 + 21\gamma/n + o(1/n). \end{aligned}$$

By similar calculations, we get

$$\begin{aligned} E[D_{J,0.5}^R(n, 2)] &= \sigma^2 + o(1/n), \\ E[\tilde{D}_{J,0.5}^R(n, 2)] &= \sigma^2 + o(1/n). \end{aligned}$$

Now, we prove the variance results. From [3], we know that as $m \rightarrow \infty$,

$$\text{Cov}(A_i(f_0; \delta, m), A_1(f_0; 1, \delta m)) \rightarrow 2 \frac{\sigma^4}{\delta^3}, \tag{49}$$

$$\text{Cov}(C_i(g_0; \delta, m), C_1(g_0; 1, \delta m)) \rightarrow \frac{4}{5} \frac{\sigma^4}{\delta^2} \tag{50}$$

for $i = 1, 2, \dots, \delta$ for any positive integer δ .

By similar calculations, we get as $m \rightarrow \infty$,

$$\text{Cov}(C_i(g_0; \delta, m), A_1(f_0; 1, \delta m)) \rightarrow \frac{6}{5} \frac{\sigma^4}{\delta^3}, \tag{51}$$

$$\text{Cov}(A_i(f_0; \delta, m), C_1(g_0; 1, \delta m)) \rightarrow \frac{6}{5} \frac{\sigma^4}{\delta^2} \tag{52}$$

for $i = 1, 2, \dots, \delta$ for any positive integer δ .

From Equations (49)–(52), we get as $m \rightarrow \infty$,

$$\begin{aligned}
& \text{Cov}(D_i(h_0; \delta, m), D_1(h_0; 1, \delta m)) \\
&= 4\text{Cov}(C_i(g_0; \delta, m), C_1(g_0; 1, \delta m)) - 2\text{Cov}(C_i(g_0; \delta, m), A_1(f_0; 1, \delta m)) \\
&\quad - 2\text{Cov}(A_i(f_0; \delta, m), C_1(g_0; 1, \delta m)) + \text{Cov}(A_i(f_0; \delta, m), A_1(f_0; 1, \delta m)) \\
&= \frac{4}{5} \frac{\sigma^4}{\delta^2} - \frac{2}{5} \frac{\sigma^4}{\delta^3}.
\end{aligned} \tag{53}$$

By using Equation (53) and Example 3, we get as $n \rightarrow \infty$,

$$\begin{aligned}
& \text{Var}(D^R(h_0; n, 2)) \\
&= \frac{1}{49} \text{Var} \left(D^{01}(h_0; 1, n) + D^{11}(h_0; 2, n/2) + D^{12}(h_0; 2, n/2) + D^{21}(h_0; 4, n/4) \right. \\
&\quad \left. + D^{22}(h_0; 4, n/4) + D^{23}(h_0; 4, n/4) + D^{24}(h_0; 4, n/4) \right) \\
&= 0.1010\sigma^4.
\end{aligned} \tag{54}$$

Similarly, as $n \rightarrow \infty$,

$$\begin{aligned}
& \text{Var}(D_{J,0.5}^R(n, 2)) \\
&= \frac{1}{49} \text{Var} \left(D_{J,0.5}^{01}(1, n) + D_{J,0.5}^{11}(2, n/2) + D_{J,0.5}^{12}(2, n/2) + D_{J,0.5}^{21}(4, n/4) \right. \\
&\quad \left. + D_{J,0.5}^{22}(4, n/4) + D_{J,0.5}^{23}(4, n/4) + D_{J,0.5}^{24}(h_0; 4, n/4) \right) \\
&= \frac{1}{49} \text{Var} \left(2D^{01}(h_0; 1, n) - D^{11}(h_0; 2, n/2) + 2D^{11}(h_0; 2, n/2) - D^{21}(h_0; 4, n/4) \right. \\
&\quad \left. + 2D^{12}(h_0; 2, n/2) - D^{23}(h_0; 4, n/4) + 2D^{21}(h_0; 4, n/4) - D^{31}(h_0; 8, n/8) \right. \\
&\quad \left. + 2D^{22}(h_0; 4, n/4) - D^{33}(h_0; 8, n/8) + 2D^{23}(h_0; 4, n/4) - D^{35}(h_0; 8, n/8) \right. \\
&\quad \left. + 2D^{24}(h_0; 4, n/4) - D^{37}(h_0; 8, n/8) \right) \\
&= 0.2498\sigma^4.
\end{aligned} \tag{55}$$

Again by similar calculations, as $n \rightarrow \infty$,

$$\begin{aligned}
& \text{Var}(\tilde{D}_{J,0.5}^R(n, 2)) \\
&= \frac{1}{49} \text{Var} \left(\tilde{D}_{J,0.5}^{01}(1, n) + \tilde{D}_{J,0.5}^{11}(2, n/2) + \tilde{D}_{J,0.5}^{12}(2, n/2) + \tilde{D}_{J,0.5}^{21}(4, n/4) \right. \\
&\quad \left. + \tilde{D}_{J,0.5}^{22}(4, n/4) + \tilde{D}_{J,0.5}^{23}(4, n/4) + \tilde{D}_{J,0.5}^{24}(4, n/4) \right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{49} \text{Var} \left(D^{01}(h_0; 1, n) - \frac{1}{2} D^{11}(h_0; 2, n/2) - \frac{1}{2} D^{12}(h_0; 2, n/2) \right. \\
&\quad + D^{11}(h_0; 2, n/2) - \frac{1}{2} D^{21}(h_0; 4, n/4) - \frac{1}{2} D^{22}(h_0; 4, n/4) \\
&\quad + D^{12}(h_0; 2, n/2) - \frac{1}{2} D^{23}(h_0; 4, n/4) - \frac{1}{2} D^{24}(h_0; 4, n/4) \\
&\quad + D^{21}(h_0; 4, n/4) - \frac{1}{2} D^{31}(h_0; 8, n/8) - \frac{1}{2} D^{32}(h_0; 8, n/8) \\
&\quad + D^{22}(h_0; 4, n/4) - \frac{1}{2} D^{33}(h_0; 8, n/8) - \frac{1}{2} D^{34}(h_0; 8, n/8) \\
&\quad + D^{23}(h_0; 4, n/4) - \frac{1}{2} D^{35}(h_0; 8, n/8) - \frac{1}{2} D^{36}(h_0; 8, n/8) \\
&\quad \left. + D^{24}(h_0; 4, n/4) - \frac{1}{2} D^{37}(h_0; 8, n/8) - \frac{1}{2} D^{38}(h_0; 8, n/8) \right) \\
&= 0.2212\sigma^4.
\end{aligned} \tag{56}$$

APPENDIX B

MA(1) EXAMPLE

Derivation of Equations (38)–(40) From [15], we know that

$$\text{Cov}(A(f_0; n), A(f_0; rn)) = 2r^3\sigma^4 + \frac{6r\gamma\sigma^2}{n} + O(n^{-2}), \quad (57)$$

$$\text{Cov}(A(f_0; n), C(g_0; rn)) = \frac{6r^3\sigma^4}{5} + \frac{3r\gamma\sigma^2}{n} + O(n^{-2}), \quad (58)$$

$$\text{Cov}(C(g_0; n), A(f_0; rn)) = \frac{6r^2\sigma^4}{5} + \frac{3r(2-r)\gamma\sigma^2}{n} + O(n^{-2}), \quad (59)$$

$$\text{Cov}(C(g_0; n), C(g_0; rn)) = \frac{4r^2\sigma^4}{5} + \frac{r(15-7r)\gamma\sigma^2}{5n} + O(n^{-2}). \quad (60)$$

From Example 3, we have

$$\text{Cov}(D(h_0; n), D(h_0; rn)) = \text{Cov}(2C(g_0; n) - A(f_0; n), 2C(g_0; rn) - A(f_0; rn)).$$

Invocation of Equations (57)–(60) immediately yields Equation (38); and then Equation (39) follows from symmetry.

Before deriving Equation (40), let $\tilde{Y}_s \equiv \sum_{i=n-s+1}^n Y_i/s$, $s < n$, denote the average of the last s observations out of Y_1, Y_2, \dots, Y_n . For the MA(1) process under study,

$$\text{Cov}(\bar{Y}_m - \bar{Y}_j, \tilde{Y}_{n-m} - \tilde{Y}_k) = \begin{cases} -\gamma/[2m(n-m)], & j < m; k < n-m \\ 0, & \text{otherwise.} \end{cases} \quad (61)$$

Further, we define \tilde{T}_s , $s < n$, as the standardized time series formed from the last s observations, and $\tilde{A}(f; s)$ and $\tilde{C}(g; s)$ as the area and CvM estimators, respectively, calculated from the last s observations. We will also make use of:

Lemma 6 ([30]). *If X_1 and X_2 are jointly normal with mean zero, then $\text{Cov}(X_1^2, X_2^2) = 2\text{Cov}^2(X_1, X_2)$.*

From Equation (61) and Lemma 6, we obtain

$$\begin{aligned}
& \text{Cov}(C(g_0; m), \tilde{C}(g_0, n - m)) \\
&= \text{Cov}\left(\frac{1}{m} \sum_{j=1}^m g_0\left(\frac{j}{m}\right) \sigma^2 T_m^2\left(\frac{j}{m}\right), \frac{1}{n - m} \sum_{k=1}^{n-m} g_0\left(\frac{k}{n - m}\right) \sigma^2 \tilde{T}_{n-m}^2\left(\frac{k}{n - m}\right)\right) \\
&= \frac{36}{m(n - m)} \sum_{j=1}^m \sum_{k=1}^{n-m} \text{Cov}\left(\sigma^2 T_m^2\left(\frac{j}{m}\right), \sigma^2 \tilde{T}_{n-m}^2\left(\frac{k}{n - m}\right)\right) \\
&= \frac{72}{m(n - m)} \sum_{j=1}^m \sum_{k=1}^{n-m} \text{Cov}^2\left(\sigma T_m\left(\frac{j}{m}\right), \sigma \tilde{T}_{n-m}\left(\frac{k}{n - m}\right)\right) \\
&= \frac{72}{m(n - m)} \sum_{j=1}^m \sum_{k=1}^{n-m} \text{Cov}^2\left(\frac{j}{\sqrt{m}}(\bar{Y}_m - \bar{Y}_j), \frac{k}{\sqrt{n - m}}(\tilde{Y}_{n-m} - \tilde{Y}_k)\right) \\
&= \frac{72}{m^2(n - m)^2} \sum_{j=1}^m \sum_{k=1}^{n-m} j^2 k^2 \text{Cov}^2(\bar{Y}_m - \bar{Y}_j, \tilde{Y}_{n-m} - \tilde{Y}_k) \\
&= \frac{72}{m^2(n - m)^2} \sum_{j=1}^{m-1} \sum_{k=1}^{n-m-1} j^2 k^2 \frac{\gamma^2}{4m^2(n - m)^2} \\
&= \frac{\gamma^2(m - 1)(2m - 1)(n - m - 1)(2n - 2m - 1)}{2m^3(n - m)^3}.
\end{aligned}$$

Substitution of $m = rn$ yields

$$\text{Cov}(C(g_0; rn), \tilde{C}(g_0, (1 - r)n)) = \frac{2\gamma^2}{n^2(1 - r)r} + O(n^{-3}). \quad (62)$$

By similar calculations, we obtain

$$\text{Cov}(C(g_0; rn), \tilde{A}(f_0; (1 - r)n)) = \frac{3\gamma^2}{n^2 r(1 - r)} + O(n^{-3}), \quad (63)$$

$$\text{Cov}(A(f_0; rn), \tilde{C}(g_0; (1 - r)n)) = \frac{3\gamma^2}{n^2 r(1 - r)} + O(n^{-3}), \quad (64)$$

$$\text{Cov}(A(f_0; rn), \tilde{A}(f_0, (1 - r)n)) = \frac{9\gamma^2}{2n^2 r(1 - r)} + O(n^{-3}). \quad (65)$$

Finally, Equations (62)–(65) imply

$$\begin{aligned}
& \text{Cov}(D(h_0; rn), \tilde{D}(h_0; (1 - r)n)) \\
&= \text{Cov}(2C(g_0; rn) - A(f_0; rn), 2\tilde{C}(g_0; (1 - r)n) - \tilde{A}(f_0; (1 - r)n)) \\
&= \frac{\gamma^2}{2n^2 r(1 - r)} + O(n^{-3}).
\end{aligned}$$

APPENDIX C

PROCEDURE \mathcal{P}_2

Setup: Choose nominal level of PCS $1 - \alpha$, indifference zone δ , and first stage sample size $n_0 \geq 2$. Then determine λ and ξ referring to **Parameters**.

Initialization: Let $I = \{1, 2, \dots, k\}$ be the initial set of systems.

Obtain observations $X_{ij}, j = 1, 2, \dots, n_0$, from each system i .

For all $i \neq \ell$ compute the sample variance of the difference between systems i and ℓ which is

$$S_{i\ell}^2 = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (X_{ij} - X_{\ell j} - [\bar{X}_i(n_0) - \bar{X}_\ell(n_0)])^2.$$

Let

$$N_{i\ell} = \left\lfloor \frac{(n_0 - 1)S_{i\ell}^2 \xi}{\delta^2} \right\rfloor,$$

where $\lfloor \cdot \rfloor$ indicates truncation of any fractional part. Then let

$$N_i = \max_{\ell \neq i} N_{i\ell}.$$

Here, $N_i + 1$ is the maximum number of observations that can be taken from system i .

Set the number of observations, r , equal to n_0 , and go to the next section.

Screening: Set $I^{\text{old}} = I$. Let

$$I = \left\{ i : i \in I^{\text{old}} \text{ and } \sum_{j=1}^r X_{ij} > \sum_{j=1}^r X_{\ell j} - R_{i\ell}(r), \forall \ell \in I^{\text{old}}, \ell \neq i \right\},$$

where

$$R_{i\ell}(r) = \lambda S_{i\ell} \sqrt{n_0 - 1} \sqrt{\max \left\{ \frac{(n_0 - 1)S_{i\ell}^2 \xi}{\delta^2} - r, 0 \right\}}.$$

Stopping Rule: If $|I| = 1$, then stop and select the system $i \in I$ as the best. Otherwise, take one additional observation $X_{i,r+1}$ from each system $i \in I$, and set $r = r + 1$, and go to **Screening**.

(If the objective is to select a subset of size of m containing the best system, then the stopping rule should be $|I| = m > 1$.)

Parameters: The parameter λ is any positive real number. For $1 - \alpha = 0.95$, we recommend taking $\lambda = 0.3$. Section 3.2.1 discusses the choice of λ for other values of α . For a given λ value, ξ is calculated as the solution to the equation

$$h(\xi; \lambda, n_0) = \beta, \quad (66)$$

where

$$\begin{aligned} h(\xi; \lambda, n_0) &\equiv \mathbb{E} \left[\frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp \left\{ -\chi_{n_0-1}^2 \frac{\xi}{2} \right\} \right. \\ &\quad \times \sum_{n=0}^{\infty} \left(\frac{k_{2n+1}(\lambda \sqrt{\chi_{n_0-1}^2})}{\ell_{2n+1}(\lambda \sqrt{\chi_{n_0-1}^2})} \frac{2^n n!}{(2n+1)!} \left(\sqrt{\chi_{n_0-1}^2} \xi \right)^{2n+1} \right) \Big]. \end{aligned}$$

Now we prove that \mathcal{P}_2 also provides the predetermined probability of correct selection.

Proof of Theorem 7: Consider two systems 1 and i such that $\mu_1 \geq \mu_i + \delta$. Select a value of ξ such that $h(\xi; \lambda, n_0) = \beta$ for some $0 < \beta < 1/2$. Let

$$T = \min \{ r : r \geq n_0 \text{ and } -R_{1i}(r) < \sum_{j=1}^r (X_{1j} - X_{ij}) < R_{1i}(r) \text{ is violated} \}.$$

Notice that T is the stage at which the procedure terminates. Let ICS_i be the event of incorrect selection when only two systems, 1 and i , are considered. Then

$$\begin{aligned} \Pr \{ \text{ICS}_i \} &= \Pr \left\{ \sum_{j=1}^T (X_{1j} - X_{ij}) < -\lambda S_{1i} \sqrt{n_0 - 1} \sqrt{\max \left\{ \frac{(n_0 - 1) S_{1i}^2 \xi}{\delta^2} - T, 0 \right\}} \right\} \\ &= \Pr \left\{ \sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < -\lambda \frac{S_{1i}}{\sigma_{1i}} \sqrt{n_0 - 1} \sqrt{\max \left\{ \frac{(n_0 - 1) S_{1i}^2 \xi}{\delta^2} - T, 0 \right\}} \right\} \\ &\leq \Pr_{\text{SC}} \left\{ \sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < -\lambda \frac{S_{1i}}{\sigma_{1i}} \sqrt{n_0 - 1} \sqrt{\max \left\{ \frac{(n_0 - 1) S_{1i}^2 \xi}{\delta^2} - T, 0 \right\}} \right\} \\ &= \mathbb{E} \left[\Pr_{\text{SC}} \left\{ \sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < -\lambda \frac{S_{1i}}{\sigma_{1i}} \sqrt{n_0 - 1} \sqrt{\max \left\{ \frac{(n_0 - 1) S_{1i}^2 \xi}{\delta^2} - T, 0 \right\}} \right\} \middle| S_{1i} \right] \end{aligned}$$

$$\leq \mathbb{E} \left[\Pr_{\text{SC}} \left\{ \sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < 0 \middle| S_{1i} \right\} \right]. \quad (67)$$

If we let

$$a = \lambda \sqrt{n_0 - 1} \frac{S_{1i}}{\sigma_{1i}}, \quad s = \frac{(n_0 - 1)S_{1i}^2}{\delta^2} \xi, \quad t = n_0, \dots, N_{1i} + 1,$$

then by using Lemma 1 and 2 we get

$$\begin{aligned} \mathbb{E} \left[\Pr_{\text{SC}} \left\{ \sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < 0 \middle| S_{1i} \right\} \right] &\leq \\ E \left[\frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp \left\{ -\frac{(n_0 - 1)S_{1i}^2}{\sigma_{1i}^2} \frac{\xi}{2} \right\} \sum_{n=0}^{\infty} \left(\frac{k_{2n+1} \left(\lambda \frac{S_{1i}\sqrt{n_0-1}}{\sigma_{1i}} \right)}{\ell_{2n+1} \left(\lambda \frac{S_{1i}\sqrt{n_0-1}}{\sigma_{1i}} \right)} \frac{2^n n!}{(2n+1)!} \left(\sqrt{\frac{(n_0 - 1)S_{1i}^2}{\sigma_{1i}^2}} \xi \right)^{2n+1} \right) \right]. \end{aligned}$$

Since $(n_0 - 1)S_{1i}^2/\sigma_{1i}^2$ is chi-square distributed with $n_0 - 1$ degrees of freedom, the expectation above is equivalent to

$$\mathbb{E} \left[\frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp \left\{ -\chi_{n_0-1}^2 \frac{\xi}{2} \right\} \sum_{n=0}^{\infty} \left(\frac{k_{2n+1} \left(\lambda \sqrt{\chi_{n_0-1}^2} \right)}{\ell_{2n+1} \left(\lambda \sqrt{\chi_{n_0-1}^2} \right)} \frac{2^n n!}{(2n+1)!} \left(\sqrt{\chi_{n_0-1}^2} \xi \right)^{2n+1} \right) \right]. \quad (68)$$

Notice that (68) is $h(\xi; \lambda, n_0)$. So, this expectation is equal to β .

The proof beyond this point is same as that of Theorem 6. \square

APPENDIX D

DETERMINATION OF VECTOR A

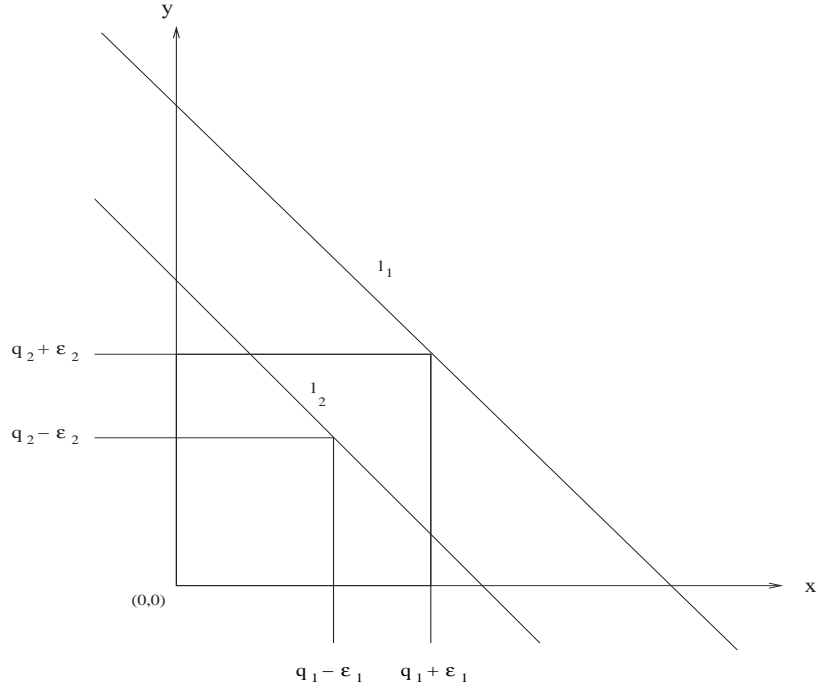


Figure 23: A Possible Choice of ℓ_1 and ℓ_2

The $\mathbf{A} = (a_1, a_2, \dots, a_s)'$ vector determines the boundaries among the D^a , A^a , and U^a regions. When we have two stochastic constraints, we set lines ℓ_1 and ℓ_2 that determine the boundaries among these regions. The lines are parallel to each other and pass through the corners of the original rectangular regions. Hence, the U^a region does not contain any desirable or acceptable systems and the A^a region does not contain any desirable systems in terms of the original constraints. A possible choice of ℓ_1 and ℓ_2 for some target values and tolerance levels when $s = 2$ is shown in Figure 23. Since the screening procedure with aggregated observations makes infeasibility decision only, we focus on determining ℓ_1 .

We start with $s = 2$ and then generalize the results to any s . While determining ℓ_1 , we can assume that $q_1 = 0$ and $q_2 = 0$ without loss of generality. Then from Figure 24, we see

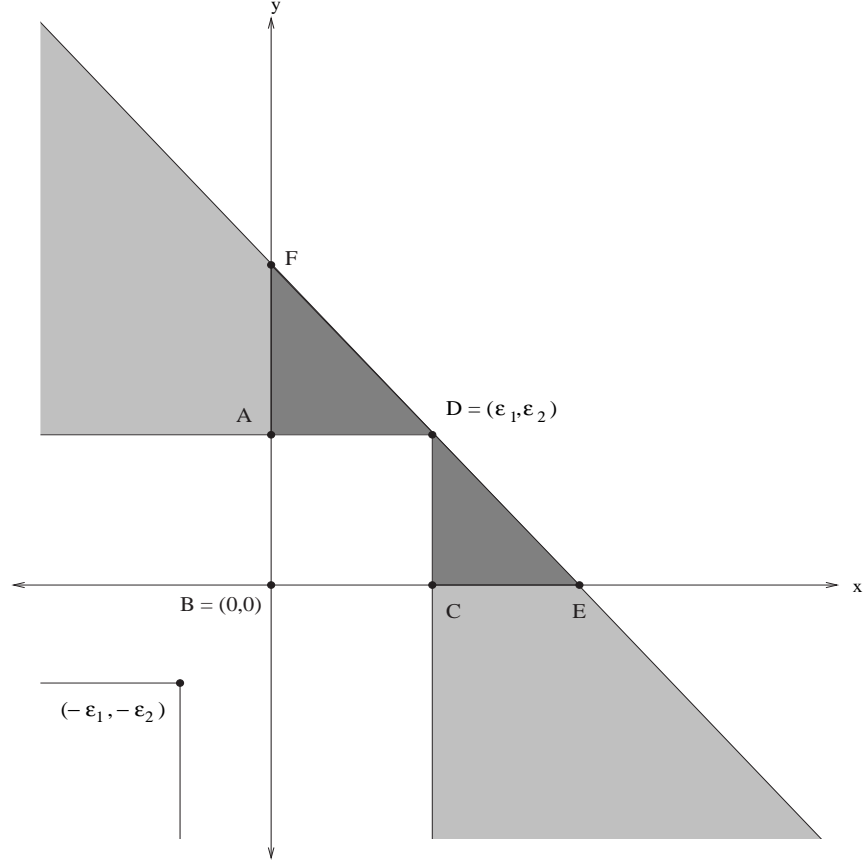


Figure 24: Determining Line ℓ_1 .

that for a vector \mathbf{A} , ℓ_1 has the following form:

$$a_1x + a_2y + c = 0,$$

where $a_1 > 0$, $a_2 > 0$, and $a_1\epsilon_1 + a_2\epsilon_2 + c = 0$.

While determining the values of a_1 and a_2 , we try to find a_1 and a_2 that minimize the dark shaded regions. One may want to find the values of a_1 and a_2 that minimize the total area of the light and dark shaded regions, hoping that she can minimize the possible number of infeasible systems that may fall in the A^a or D^a region. Notice that if it happens, the infeasible system is likely to be survived from the screening procedure with aggregated observations. However, the light shaded regions contain systems which are feasible in terms of one constraint and infeasible in terms of the other constraint. In this case, it is not clear if aggregated observations would make it easy to detect infeasibility. On the other hand, the

dark shaded regions contain systems that are infeasible for both constraints and aggregation likely makes infeasibility more noticeable. Therefore, the regions that actually matter for the efficiency of the procedure are the dark shaded regions and we want to minimize its area.

Our optimization problem becomes:

$$\begin{aligned}
\min_{a_1, a_2} \quad & \frac{c^2}{a_1 a_2} \\
s.t. \quad & a_1 \epsilon_1 + a_2 \epsilon_2 + c = 0 \\
& a_1, a_2 > 0
\end{aligned} \tag{69}$$

The objective function is basically the area of Triangle FBE. However, this is equivalent to minimizing the area of the dark shaded region because the area of Rectangle ABCD does not depend on the choice of a_1 and a_2 . The solution to problem (69) is $a_1 = \epsilon_2$ and $a_2 = \epsilon_1$, and \mathbf{A} vector is $(\epsilon_2, \epsilon_1)'$.

For general s , we solve

$$\begin{aligned}
\min \quad & \frac{c^s}{a_1 a_2 \dots a_s} \\
s.t. \quad & a_1 \epsilon_1 + a_2 \epsilon_2 + \dots + a_s \epsilon_s + c = 0 \\
& a_i > 0, i = 1, 2, \dots, s,
\end{aligned}$$

and we get $\mathbf{A} = [a_\ell]_{\ell=1,2,\dots,s}$ such that $a_\ell = \prod_{j=1, j \neq \ell}^s \epsilon_j$.

REFERENCES

- [1] ANDRADÓTTIR, S. and ARGON, N. T., “Variance estimation using replicated batch means,” in *Proceedings of the 2001 Winter Simulation Conference*, ed. B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 338–343, 2001.
- [2] ANDRADÓTTIR, S., GOLDSMAN, D., and KIM, S.-H., “Fully sequential procedures for comparing constrained systems via simulation,” in *Proceedings of the 2005 Winter Simulation Conference*, ed. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 2005.
- [3] BATUR, D., GOLDSMAN, D., KIM, S.-H., and NELSON, B. L., “Rebatched variance estimators for simulations,” *working paper*, 2006.
- [4] BILLINGSLEY, P., *Convergence of Probability Measures*. John Wiley & Sons, New York, 1968.
- [5] BOESEL, J., NELSON, B. L., and KIM, S.-H., “Using ranking and selection to ‘clean up’ after simulation optimization,” *Operations Research*, vol. 51, pp. 814–825, 2003.
- [6] BUTLER, J., MORRICE, D. J., and MULLARKEY, P. W., “A multiple attribute utility theory approach to ranking and selection,” *Management Science*, vol. 47, pp. 800–816, 2001.
- [7] CHEN, H.-C., CHEN, C.-H., DAI, L., and YÜCESAN, E., “New development of optimal computing budget allocation for discrete event simulation,” in *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 334–341, 1997.
- [8] CHEN, H.-C., CHEN, C.-H., and YÜCESAN, E., “Computing efforts allocation for ordinal optimization and discrete event simulation,” *IEEE Transactions on Automatic Control*, vol. 45, pp. 960–964, 2000.
- [9] CHICK, S. and INOUE, K., “New procedures for identifying the best simulated system using common random numbers,” *Management Science*, vol. 47, pp. 1133–1149, 2001.
- [10] CHICK, S. and INOUE, K., “New two-stage and sequential procedures for selecting the best simulated system,” *Operations Research*, vol. 49, pp. 1609–1624, 2001.
- [11] CHICK, S. E., “Selecting the best system: A decision-theoretic approach,” in *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 326–333, 1997.
- [12] CHIEN, C., GOLDSMAN, D., and MELAMED, B., “Large-sample results for batch means,” *Management Science*, vol. 43, pp. 1288–1295, 1997.

- [13] FEREBEE, B., “Tests with parabolic boundary for the drift of a Wiener process,” *Annals of Statistics*, vol. 10, pp. 882–894, 1982.
- [14] FOLEY, R. D. and GOLDSMAN, D., “Confidence intervals using orthonormally weighted standardized time series,” *ACM Transactions on Modeling and Computer Simulation*, vol. 9, pp. 297–325, 1999.
- [15] GOLDSMAN, D., KANG, K., KIM, S.-H., SEILA, A. F., and TOKOL, G., “Combining standardized time series area and Cramér–von Mises variance estimators,” tech. rep., School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, 2005.
- [16] GOLDSMAN, D., KANG, K., and SEILA, A. F., “Cramér–von Mises variance estimators for simulations,” *Operations Research*, vol. 47, pp. 299–309, 1999.
- [17] GOLDSMAN, D., KIM, S.-H., and NELSON, B. L., “Statistical selection of the best system,” in *Proceedings of the 2005 Winter Simulation Conference*, ed. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 178–187, 2005.
- [18] GOLDSMAN, D. and MEKETON, M. S., “A comparison of several variance estimators,” tech. rep., School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 1986.
- [19] GOLDSMAN, D., MEKETON, M. S., and SCHRUBEN, L. W., “Properties of standardized time series weighted area variance estimators,” *Management Science*, vol. 36, pp. 602–612, 1990.
- [20] GOLDSMAN, D. and NELSON, B. L., *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, ch. 8 Comparing Systems via Simulation, pp. 273–306. New York: John Wiley & Sons, 1998.
- [21] JENNISON, C., JOHNSTONE, I. M., and TURNBULL, B. W., “Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means,” tech. rep., Department of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853, 1980.
- [22] KIM, S.-H. and NELSON, B. L., “A fully sequential procedure for indifference-zone selection in simulation,” *ACM TOMACS*, vol. 11, pp. 251–273, 2001.
- [23] KIM, S.-H. and NELSON, B. L., *Handbooks in Operations Research and Management Science*, ch. 17 Selecting the Best: Theory and Method. Elsevier, Oxford, UK, 2006.
- [24] KIM, S.-H. and NELSON, B. L., “On the asymptotic validity of fully sequential selection procedures for steady-state simulation,” *to appear in Operations Research*, 2006.
- [25] LAW, A. M. and KELTON, D., *Simulation Modeling and Analysis, 3rd ed.* New York: McGraw-Hill, 2000.
- [26] LERCHE, H. R., *Lecture Notes in Statistics*, ch. Boundary Crossing of Brownian Motion, pp. 98–139. New York: Springer-Verlag, 1986.

- [27] MALONE, G. J., KIM, S.-H., GOLDSMAN, D., and BATUR, D., "Performance of variance updating procedures on various data," in *Proceedings of the 2005 Winter Simulation Conference*, ed. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 2005.
- [28] MEKETON, M. S. and SCHMEISER, B. W., "Overlapping batch means: Something for nothing?," in *Proceedings of the 1984 Winter Simulation Conference*, ed. S. Sheppard, U. W. Pooch, and C. D. Pegden. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 227–230, 1984.
- [29] NELSON, B. L., SWANN, J., GOLDSMAN, D., and SONG, W., "Simple procedures for selecting the best simulated system when the number of alternatives is large," *Operations Research*, vol. 49, pp. 950–963, 2001.
- [30] PATEL, J. K. and READ, C. B., *Handbook of the Normal Distribution, 2nd Ed.* Marcel Dekker, New York, 1996.
- [31] RINOTT, Y., "On two-stage selection procedures and related probability-inequalities," *Comm. Stat.-Thy. and Meth*, vol. A7, pp. 799–811, 1978.
- [32] SANTNER, T. J. and TAMHANE, A. C., *Design of Experiments – Ranking and Selection: Essays in Honor of Robert E. Bechhofer*, ch. Designing Experiments for Selecting a Normal Population with a Large Mean and a Small Variance, pp. 179–198. No. 179–198, New York: Marcel-Dekker, 1984.
- [33] SCHMEISER, B. W., "Batch size effects in the analysis of simulation output," *Operations Research*, vol. 30, pp. 556–568, 1982.
- [34] SCHRUBEN, L. W., "Confidence interval estimation using standardized time series," *Operations Research*, vol. 31, pp. 1090–1108, 1983.
- [35] SONG, W.-M. T. and SCHMEISER, B. W., "Optimal mean-squared-error batch sizes," *Management Science*, vol. 41, pp. 110–123, 1995.
- [36] TAMHANE, A. C., "Multiple comparisons in model I one-way ANOVA with unequal variances," *Communications in Statistics*, vol. A6, pp. 15–32, 1977.
- [37] ZHU, L., GOLDSMAN, D., KIM, S.-H., and BATUR, D., "A sequential procedure with parabolic boundary for known-variance normal population selection.," tech. rep., School of Industrial Systems and Engineering, Georgia Institute of Technology, Atlanta, GA 30332, 2005.