University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Industrial and Management Systems Engineering -- Dissertations and Student Research | Industrial and Management Systems Engineering

5-1-2012

System Selection Based on Controlled Risk

Lina Wang
University of Nebraska – Lincoln, lina_wang0213@yahoo.com.cn

Follow this and additional works at: http://digitalcommons.unl.edu/imsediss

Part of the Operations Research, Systems Engineering and Industrial Engineering Commons

Wang, Lina, "System Selection Based on Controlled Risk" (2012). *Industrial and Management Systems Engineering -- Dissertations and Student Research.* Paper 27.
http://digitalcommons.unl.edu/imsediss/27

This Article is brought to you for free and open access by the Industrial and Management Systems Engineering at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Industrial and Management Systems Engineering -- Dissertations and Student Research by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

SYSTEM SELECTION BASED ON CONTROLLED RISK

by

Lina Wang

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Master of Science

Major: Industrial & Management System Engineering

Under the Supervision of Professors Fred Choobineh & Demet Batur

Lincoln, Nebraska

May, 2012

SYSTEM SELECTION BASED ON CONTROLLED RISK

Lina Wang, M. S.

University of Nebraska, 2012

Adviser: Fred Choobineh & Demet Batur

The studies of using computer simulation to address stochastic systems selection problem prevail currently. Ranking and selection procedures are statistical procedures that can be used in the field of computer simulation to select the best system or the subset that includes the best system based on a performance metric of the interest with a guaranteed correct selection. The best system is defined in the majority of selection problems as a system with the largest (or smallest) mean value. However, mean is only a measure of the average behavior. Risk, which influences the stability of a system to a large extend, may not be controlled by controlling the mean.

In this study, we focus on developing stochastic system selection procedures that consider the mean and the risk of systems simultaneously. Variance is used as a measure of risk. Three different ranking and selection procedures are designed to satisfy the decision maker's preference via i) mean-variance measure, ii) mean with variance constraint measure, and iii) risk-adjusted mean measure. The validation of each procedure is proved, and the experiments are illustrated at the end of this thesis.

Contents

Contents **iii**

List of Figures **v**

List of Tables **vi**

1 Introduction **1**

2 Background **6**

2.1 Risk 6

2.1.1 Definition of Risk 6

2.1.2 Measures of Risk 7

2.2 Ranking and Selection Procedures 9

2.3 Related Work 16

2.4 Objectives of our Study 17

3 Measures for comparison **18**

3.1 Mean and Variance 18

3.2 Mean Comparison with Variance Constraint 19

3.3 Baumol's L Criterion 19

4 Methodology **23**

4.1 Mean-Variance Selection (MVS) Procedure 24

4.2 Constrained Mean-Variance Selection (CMVS) Procedure 32

4.3 Risk-Adjusted Mean Selection (RAMS) Procedure 38

4.3.1 Quantile Estimation 39

4.3.2 Quantile-Window Approach 40

4.3.3 Procedure RAMS 42

5 Experiments **47**

5.1 Experiments for MVS Procedure 47

5.2 Experiments for CMVS Procedure 50

5.3 Experiments for RAMS Procedure 51

5.4 Experiments for More than two Systems 56

5.5 Experiments for Robustness 57

6 Conclusion **61**

A Results for Experiments of Section 5.5 **63**

B Matlab Code for Procedure MVS **66**

C Matlab Code for Procedure CMVS **75**

D Matlab Code for Procedure RAMS **85**

Bibliography **94**

List of Figures

5.1 The relationship between the cross-over point and the boundaries for Configurations 9 53

5.2 The relationship between the cross-over point and the boundaries for Configurations 10 53

5.3 The relationship between the cross-over point and different G's. 54

5.4 Results from group 1 58

(a) Results from sub-group 1 for procedure MVS. 58

(b) Results from sub-group 2 for procedure MVS. 58

(c) Results from sub-group 1 for procedure CMVS. 58

(d) Results from sub-group 2 for procedure CMVS. 58

(e) Results from sub-group 1 for procedure RAMS. 58

(f) Results from sub-group 2 for procedure RAMS. 58

5.5 Results from group 2 59

(a) Results from group 2 for procedure MVS 59

(b) Results from group 2 for procedure CMVS 59

(c) Results from group 2 for procedure RAMS 59

5.6 RAMS procedure when systems are exponential and the differences between L's are all equal to δ_x. 60

List of Tables

4.1 Quantile Window Q sets at $G = 1$ 41

5.1 Configurations 1-3 of MVS procedure, $\delta = 0.05$. 48

5.2 Results for Configurations 1-3 of MVS procedure 48

5.3 Configurations 4-5 of MVS procedure, $\delta = 0.05$. 49

5.4 Results for Configurations 4-5 of MVS procedure 49

5.5 Configurations 6-8 of CMVS procedure, $(\delta = 0.05, \sigma_0 = R = 1.1)$. 50

5.6 Results for Configurations 6-8 of CMVS procedure 51

5.7 Configurations 9-10 of RAMS procedure $(\delta_x = 0.5, G = 1)$ 51

5.8 Results for Configurations 9-10 of RAMS procedure 52

5.9 Configurations of the RAMS procedure with different G values for the system pair $(\mu_1 = 100\delta_x, \sigma_1 = 60\delta_x)$ and $(\mu_2 = 50\delta_x, \sigma_2 = 10\delta_x)$, $(\delta_x = 0.5)$ 54

5.10 Results of the RAMS procedure with different G values for the system pair $(\mu_1 = 100\delta_x, \sigma_1 = 60\delta_x)$ and $(\mu_2 = 50\delta_x, \sigma_2 = 10\delta_x)$, $(\delta_x = 0.5)$ . . . 55

5.11 Configurations of the RAMS procedure when the cross over point is at q_T quantile, $(\delta_x = 0.5)$ 55

5.12 Results for the RAMS procedure when the cross over point is at q_T quantile, $(\delta_x = 0.5)$ 55

5.13 Configuration 15 for three procedures, ($\delta = 0.05, \delta_x = 0.05$, and $\sigma_0 = R = 1.1$). 56

5.14 Results of Configuration 15 for three procedures, ($\delta = 0.05, \delta_x = 0.05$, and $\sigma_0 = R = 1.1$). 56

A.1 Configurations of sub-group 1, ($\delta = 0.05, \delta_x = 0.05, R = 1.1, \sigma_0 = 10, G = 1$). 63

A.2 Configurations of sub-group 2, ($\delta = 0.05, \delta_x = 0.05, R = 1.1, \sigma_0 = 10, G = 1$). 64

A.3 Configurations of Group 2, ($\delta = 0.05, \delta_x = 0.05, R = 1.1, \sigma_0 = 10, G = 0.5$). 64

A.4 Configurations of the differences of Ls between two systems are equal to δ_x, ($G = 0.5$). 65

ACKNOWLEDGMENTS

I am very glad to say that my research for this thesis has been a wonderful learning experience. Acknowledgments for my research and thesis are given to many people.

I would like to express my sincere gratitude to my advisors Dr. Demet Batur and Dr. Fred Choobineh. First, I am grateful to them for providing me the opportunity and resources to work on thesis research. It has been great to work with them, very encouraging and enthusiastic professors. Their broad knowledge and constructive ideas have been of great value for me, and have always greatly motivated me in my research.

I 'd like to thank all the professors on my committee for their insightful comments and suggestions throughout my research.

I have gratefulness to all my friends in the lab. They are cute and friendly. They always create a happy and enjoyable environment for my research.

Finally, I want to thank my family and all my friends for their constant encouragement and support.

Chapter 1

Introduction

During this decade, the studies on solving stochastic systems selection problem via simulation prevail. The majority of papers are focused on finding the most effective method to select the best system, or subset that includes the best system based on a performance metric of the interest. The best system is often defined as the system with the smallest (or largest) expected value. However, for different selection problem, the goal of the selection may be different. It is necessary to establish the dominant set of systems based on the decision maker's preference.

Generally, in finance and economics the decision makers are classified as risk-seeking, risk-neutral, and risk-averse decision makers. We assume this classification can be applied to the field of selection problem. In this paper, our objective is to construct stochastic systems selection procedures via simulation which are based on the decision maker's preference, especially for risk-averse decision maker.

It is well known that the legacy mean dominance is the most prevalent selection criterion in the literature of stochastic system selection via simulation. Nonetheless, mean is only a measure of the average behavior. A common misconception is to confuse *estimation error* with *risk* (Henderson and Nelson (2006) [18]). The confidence

interval of a mean performance metric is only a measure of the estimation error. It is not related to the risk associated with the system's performance. Usually, decreasing the estimation error is through increasing the simulation run length or the number of simulation replications; however, the risk associated with the system's performance remains the same. Hence, the way of selecting the best system with the smallest (or largest) expected value is just suitable for risk-neutral decision makers. In the mean-based selection criterion, risk is not considered. Any system with the largest expected value can be selected as the best system, no matter what the variance value is.

However, for the risk-averse decision maker, risk is an important element in the decision making process. A system with large risk is unpredictable and hard to control. Batur and Choobineh (2009) [2] discuss different ways of measuring risk in ranking and selection. One way of measuring risk is through the variance. In this thesis, we focus on three different problems for comparing simulated systems based on the mean and variance measures of a performance metric of interest. We propose a sequential selection procedure for each problem. Without loss of generality, we assume that larger is better; similar results hold in the case of smaller is better.

The first problem is the mean-variance (MVS) selection problem which compares systems based on the mean and variance measures simultaneously. The mean and variance selection criterion was first explored by Markowitz (1959) [26]. The analysis of the relationship is described as efficient frontier. In the mean-variance based comparison approach, the system with the best mean and the smallest variance is defined as the best system. Two systems are seen as nondominant systems if one system is detected to be better based on the mean but worse based on the variance. If there are pairs with the same mean and variance values, the systems are equal.

In the MVS selection problem, the variability is measured by the variance, and

it is a breakthrough approach compared to the mean-based selection. However, in the mean-variance comparison approach, it is possible that the best system is the one with the largest variance because it also has the best mean value. The system with larger mean and variance value may be the best system for the risk-seeking and risk-neutral decision-maker but absolutely not for the risk-averse decision maker. Moreover, the mean-variance comparison approach often results in a MV-efficient set with more than one system. If the user is interested in selecting only one system, other selection criteria need to be considered to eliminate more systems.

In most problem settings, the risk-averse decision makers are interested in selecting the best system from among the systems that have acceptable levels of risk. In this study, the second problem is a constrained mean-variance (CMVS) selection problem. In the CMVS problem, a target variance is specified by decision maker. The systems that have larger variance than the target are considered infeasible. This leads to the elimination of infeasible systems which have unacceptable levels of risk. The remaining feasible systems are compared based on the mean measure to select the best system. The purpose in this problem is to set an upper bound for variance to reduce the probability of selecting a high risk system.

Baumol (1963) [4] states that a system with relatively high standard deviation σ will be relatively safe if its expected value μ is sufficiently high. He explains that by assuming $\mu - \sigma$ and $\mu + \sigma$ as the lowest and highest confidence limit of the expected value. For example, there are two types of new machines which are considered for producing emergency goods. According to the trial test, the machine A can produce 18 goods per hour, which has the variability of 5^2 goods per hour, and the machine B can produce 10 goods per hour, which has the variability of 2^2 goods per hour. Based on the mean-variance criterion, machine A and B are nondominant because A has better mean value and poorer variance. Based on the constrained mean-variance

selection criterion, machine B is the best system if the variance of the standard system is smaller than 5^2.

However, in Baumol's statement, the lowest expected production of Machine A, $\mu_A - \sigma_A = 18 - 5 = 13$, is higher than the highest expected production of Machine B, $\mu_B + \sigma_B = 10 + 2 = 12$. Also, if the variables are considered as normally distributed, the probability of obtaining the lowest production of Machine A is larger than obtaining the highest production of Machine B. That means it is more likely to get a higher production rate with Machine A; choosing machine A is safer than machine B.

In this thesis, another problem named risk-adjusted mean based selection (RAMS) problem is presented. In the RAMS problem, the comparison of systems is based on the performance measure $\mu - G\sigma$, where μ is the expected value, σ is the standard deviation, and G is a real number specified by the decision maker to express the extent of feeling about the risk. When G is zero, $\mu - G\sigma = \mu$. Hence, it is the same as comparison of only the mean values. That works for risk-neutral decision makers because risk does not have any effect on the performance measure. If G is positive, the mean is adjusted down based on the risk. As G becomes larger, the negative effects of risk on the performance measure increases. The decision makers, who love to take higher risk, may specify a negative G value. In this paper, we focus on finding the measures for risk-averse decision maker. Hence, G is only considered to have a positive value.

Baumol (1963) [4] presents the L selection criterion which compares system based on the $L = \mu - G\sigma$ measure. In the L criterion, the system with the largest L value is the best. Baumol also points out a relationship between the L criterion and Markowitz's mean-variance selection criterion. The L criterion is a derived measure from Markowitz criterion. In addition, Robison and Barry (1978) [35] explains the

relationship between the first-degree stochastic dominance and the L criterion. They show that the first-degree stochastic dominance guarantees the L criterion, but it can not be ensured backwards. However, if a system is better based on the L criterion for any G value, it also first-order stochastically dominates other systems. These relationships between the L criterion and Markowitz's mean-variance selection criterion or the first-degree stochastic dominance will be discussed in detail in Chapter 3.

We designed three sequential ranking and selection procedures to solve the selection problems introduced above. In Chapter 4, three selection procedures are presented with detail, followed by the experiments in Chapter 5. The conclusion is presented in Chapter 6.

Chapter 2

Background

2.1 Risk

The risk comes from the uncertainty of the future. Any event for which enough information can not be obtained to infer the results includes a risk.

2.1.1 Definition of Risk

For any risk there are two characters: i) uncertainty and ii) potential undesired loss. For example,

- The change of hazard or danger;
- Exposure to the chance of injury or loss;
- The chance of loss;
- The amount of possible loss;

Compared with risk, riskless is no loss or without uncertainty. For example, depositing $2000 into any bank with a certainty return rate 0.03 in a year is riskless. Buying a

U.S. Treasury bill, which mature in half year at 500, assuming that the market price is 470 today, and has a return rate is 0.064, is also riskless.

The riskless investment has only one certain outcome. On the contrary, the risky investment has more than one outcome. For instance, play the game of tossing a coin. The rule of the game is gaining 1 dollar if it is a head and losing 1 dollar if it is a tail. Hence, the probability of gaining one dollar is 0.5, as well as losing. This is a risk action since we do not know the actual outcome (more than one outcome), and it may be a loss. In addition, all of the gambling games, stocks, and almost all investments have risk. Hence, finding a way to measure the risk and rank the risky investments is very helpful and necessary for investor to make a decision before investing in a risky project.

2.1.2 Measures of Risk

In the following paragraph, several risk measures, which have appeared in the literature of Finance and Economics, are discussed (Levi [24]).

Domar and Musgrave (1944) [24] calculate the risk by considering all of the possible losses. They point out that the essential of the risk is the probability of actual return being less than zero. The proposed risk index is presented as $RI_{DM} = -\sum_{x_i \leq 0} p_i x_i$ for discrete variables and $RI_{DM} = -\int_{-\infty}^{0} f(x)x \, \mathrm{d}x$ for continuous variables. Moreover, for some of the investors who think the risk occurs when the return rate is less than riskless interest rate d, the above formula needs to be changed to $RI_{DM'} = -\sum_{x_i \leq d} p_i(x_i - d)$ for discrete variables, and $RI_{DM'} = -\int_{-\infty}^{d} f(x)(x-d) \, \mathrm{d}x$ for continuous variables. Actually, d can be any number that expresses the lowest acceptable return rate in investor's opinion. In both formulas, the RI values are positive. Thus, the larger the value of RI, the higher the risk is.

Roy (1952) [24] proposes a principle named "safety first" and calculates the risk by just taking the probability of loss into account. Roy's risk index is $RI_R = P(X \leq d)$, where d is the lowest acceptable return rate of the investor. Using the Chebycheff's inequality and selecting $k\sigma = (\mu - d)$, the probability of $P(X \leq d) \leq \frac{\sigma^2}{2(\mu-d)^2}$ can be obtained, where μ is the expected value and σ is the standard deviation.

The first two measures RI_{DM} and RI_R are attractive since they directly express the investor's intuition. It calculates only the loss of the investment. However, they do not consider the differential damage of the amount of loss. For instance, in the first method RI_{DM}, two investments are considered equal if one has 0.1 probability of losing \$100 and the other has 0.5 probability of losing \$20. However, different investors may have different opinions on the ranking of risk. Some investors may more concern about the amount of loss, and the others may think the probability of loss is more important. The same problem also exists in the second method.

Markowitz (1952) [25] develops another way of measuring risk through the variance. Variance is a measure of dispersion about the mean value. The risk action always has more than one outcome. Obtaining variance can measure the dispersion between the outcomes and infer the range of possible outcomes. Generally, variance is obtained by $\sigma^2 = \sum_{\text{all}_{x_i}} p_i(x_i - \mu)^2$ for discrete distribution, and $\sigma^2 = \int_{-\infty}^{\infty} f(x)(x - \mu)^2 \, \mathrm{d}x$ for continuous distribution. Based on the variance risk index, a mean-variance analysis is constructed.

However, the variance risk index includes all of the positive and negative return, which means the gain and the loss have been considered together. Hence, based on the larger the better, the investor prefers to maximize the gain and minimize the loss. Another measure which only considers the deviations of negative return is presented according to Markowitz's variance risk index. It is named as semi-variance risk index. In this risk index, $SV = \sum p_i(x_i - A)^2$, where A is always selected as the expected

value μ of an investment.

Baumol (1963) [4] claims that, an investment with relatively high standard deviation will be relatively safe if its expected value is sufficiently high. He points out that variability is a source of risk, nevertheless, depending on the Chebycheff's inequality, the probability of loss $\mu - G\sigma$ is lower than $\frac{1}{2G^2}$. For example, System A has a small variance and a small expected value. System B has a very large expected value but a relatively larger variance. For both systems, the relationship between $\mu_A - G\sigma_A < \mu_B - G\sigma_B$ is verified. We think system B is safer than system A. He proposed the risk index as $L = \mu - G\sigma$, where G is decided according to the safety requirement of investor.

Besides these, some of the researchers attempt to specify different utility functions to present the preference of risk. Nonetheless, all of these measures have pros and cons.

2.2 Ranking and Selection Procedures

Computer simulation is very helpful and popular for solving stochastic systems selection problem. Nevertheless, it is a challenge to determine the number of simulation replications needed for each alternative system to make a statistically valid selection. Ranking and selection (R&S) procedures are statistical procedures developed to simulate each alternative system to find the best system or subset that includes the best system based on a performance metric of the interest with a probability of correct selection guarantee. In the literature review of R&S procedures, there are four main kinds of comparison problems:

- select the system with the largest (or smallest) expected performance measure;

- compare systems with a standard system;
- select the system that has the largest probability of being the best performer (multinomial selection);
- select the system with the largest probability of success (Bernoulli selection).

In addition, R&S procedures generally can be classified into two different types, subset selection procedure and indifference-zone selection procedure [37] .

The origin of subset selection R&S procedure is the method of Gupta (1956) [16]. He presents a single-stage procedure to solve the selection problem of finding the system with the largest expected mean value. In this procedure, the samples from the competing alternatives are assumed independent, equal-sized, and normally distributed with common known variance. In Gupta's solution, I is the initial set. The systems $\ell \in I$, such that

$$\bar{X}_\ell(n) \geq \max_{i \neq \ell} \bar{X}_i(n) - h\sigma\sqrt{\frac{2}{n}},$$

compose the subset that includes all best systems, where $\bar{X}_\ell(n)$ is the sample mean value of the n outputs of system ℓ, and h is a constant that depends on $|I|$ and α. $|I|$ is the number of systems, and $1-\alpha$ is the nominal confidence level. However, for the subset selection, there are always more than one system in the best subset, and no procedure can guarantee a best subset of size 1 and meet the requirement of the probability of correct selection for an arbitrary n value.

Bechhofer (1954) [5] designs a procedure that guarantees to select the best system among K systems, whenever the difference between the mean measures of the best system and all the other systems is larger than δ, and ensures that the probability

of correct selection is above $1-\alpha$. This is the origin of the indifference-zone R&S procedures. It is a typical single stage R&S procedure, and the number of observations N for each alternative is determined via the following formula:

$$N = \lceil 2(\frac{\sigma Z_{K-1,0.5}^{(1-\alpha)}}{\delta})^2 \rceil \ , \tag{2.1}$$

where $Z_{K-1,0.5}^{(1-\alpha)}$ is a constant that can be looked up from the table in Bechhofer et. al. (1955) [6], $1-\alpha$ is the nominal confidence level, K is the number of systems, σ is the known common variance for all systems, and δ is the mean indifference zone parameter which is the smallest difference the decision maker feels is worth detecting. The best system is selected based on the largest mean value.

In the above procedures, the number of observations N for both procedures are determined by the specified parameters. The Bechhofer's procedure is powerful, it gives the required number of observations N for detecting the difference of at least δ. However, when the actual differences in the procedure are larger than δ, the N value maybe larger than needed.

Compared with the single stage procedures, the sequential selection procedures can reach a decision earlier. The first sequential selection procedure can be traced back to Wald (1947). Paulson (1964) [32] proposes a fully sequential procedure that increases one observation each time and eliminates a system from continued sampling when it is clearly figured out as an inferior system. In this procedure, a system $\ell \in I$ can survive at current stage r only if

$$\bar{X}_\ell(r) \geq \max_{i \in I} \bar{X}_i(r) - \max\{0, \frac{a}{r} - \lambda\}, \tag{2.2}$$

where $a > 0$, $0 < \lambda < \delta$, and $n = \lfloor a/\lambda \rfloor$. n is the number of needed observation for this procedure. Thus, any system that cannot satisfy Equation (2.2) is eliminated

at current stage r. With r value becoming larger at each stage, a subset selection is performed, and the hedging factor $(a/r-\lambda)$ gets smaller. At the end of the procedure, the system with the largest mean value is selected the best system.

The previous procedures are only applicable when the samples are independent and normal, and the variances are equal and known. But this situation rarely happens in the real world. Although some of single stage procedures can permit an unknown variance, Dudewicz (1996) [12] states that it is sometimes not possible to guarantee a specified probability of correct selection. In the following paragraphs, more efficient R& S procedures are illustrated for the unknown variance case.

Rinott (1978) [34] provides a two-stage R&S procedure for selecting the best system with largest mean value and unknown variance. The assumption for this method is that all simulated data are independent, identical, and from a normal distribution. In this procedure, the total sample size is determined by

$$N_i = \max\{n_0, \lceil (\frac{hS_i}{\delta})^2 \rceil\},$$

where n_0 is the initial number of observations in the first stage, S_i is the sample standard deviation calculated by the observations from the first stage, h is a constant determined based on the number of systems, $1-\alpha$, and n_0. Goldsman (2002) [14] introduces an extended Rinott's procedure (R^+). In the R^+ procedure, the observations are divided into m batches. The mean and variance value of each system comes from the batched samples.

Koening and Law (1985) [23] propose a two or more stage sampling procedure, but the procedure is only applicable when the samples are generated from independent, normally distributed competing alternatives with unknown and unequal variance. Goldsman and Nelson (1998) [15] propose a subset selection procedure which can

permit an unknown and common variance. Nelson et al. (2001) [29] proposes a two-stage procedure $NGSG$ that combines the Gupta's subset selection procedure and Bechhofer's indifference-zone selection procedure. In this procedure, the Gupta like subset selection method is used to eliminate systems that are clearly inferior after the first stage, and the Bechhofer like indifference-zone selection method is used to allocate the number of observations at the second stage. Thus, this procedure can be applied to the case of large number of alternatives.

Kim and Nelson (2001) [21] construct a fully sequential R&S procedure called KN for terminating simulation or steady-state simulation in the case of employing multiple replications. This is a more efficient procedure. In this procedure, at each stage one observation is generated for each surviving system and a system can be eliminated when it is deemed as a poorer system. However, the procedure requires all of the systems to be simulated simultaneously, and hence there is a lot of switching among the systems. In addition, the KN procedure requires that the simulated data are normal.

Kim and Nelson (2006) [22] propose another fully sequential procedure $KN++$ which is a extension of KN. It is appropriate for steady-state simulation when only one replication is employed. In this procedure, the simulated output from each alternative is assumed to follow a Function Central Limit Theorem. The variance estimators in the first stage of the KN procedure are replaced by an appropriate asymptotic variance constant, and the variance estimators are updated as the stage r increases based on the batch size m_r. This procedure has been shown to be highly efficient.

All of the procedures in the above paragraphs belong to the first type of comparison problem: select the system with the largest (or smallest) expected performance measure. It is the most common system selection problem in the literature. In the

comparison with a standard problem, the purpose is to find a system that has larger (or smaller) performance measure than the standard system. Nelson and Goldsman (2001) [27] design a two-stage procedure for this problem. Assuming system 0 is the standard system, the standard system is retained if $\bar{X}_0 + \delta$ is larger than any other systems. If not, the system with the largest mean value is selected as the best system. Kim (2005) [20] proposes a fully sequential procedure for comparison with the standard system. In this procedure, the standard system is retained if $\mu_i + \delta < \mu_0$, where μ_0 is the mean value of the standard system, μ_i is the mean value of i-th system and δ is the indifference-zone parameter.

In the multinomial selection problem, the goal is to select the best system which is most likely to be the best system in a single trial. Goldsman (1984) [13] first suggests a procedure to consider a general use of this type of problem. For example, the objective may be to determine the bulb with the longest lifetime. Assume the lifetimes of the j-th trial are $(X_{1j}, X_{2j}, \ldots, X_{kj})$. The objective is to identify the bulb with largest value of P_i, where $P_i = Pr(X_{ij} > X_{\ell j}), \ell \neq i$.

In the procedure of selecting the system with the largest probability of success, the simulated output from each system on each independent replication is denoted by X_{ij}. The value of X_{ij} will either be 0 (failure) or 1 (success). A system with the largest probability of success, $Pr\{X_{ij} = 1\}$, is deemed as the best system. There is a little research on this type of problem.

In addition, there are R&S procedures that are established using other methods. Bayesian procedure is one of these methods. The objective of the Bayesian method is to maximize the probability of correct selection by allocating no more than a specific number of observations. Chen (1995) [7] and Chen et al.(1996) [8] construct a sequential procedure based on Bayesian method using Chernoff bounds to estimate the gradient information for discrete-event system. Chick (1997) [11] presents a R&S

procedure through Bayesian decision theoretic perspective. Instead of finding the system with the largest mean value, selecting the system with the largest expected utility is the purpose in this procedure. Also, this procedure can be used when the mean and variance are unknown with the Gaussian or non-Gaussian simulation output. Chen (1997) [10] extends his previous work in 1995 and 1996 with another method to estimate the performance metric of interest and provides two parameters that crucially affect the performance of the procedures, the initial sample size and the one-time incremental observation. He recommends a good choice of the initial sample is 10 to 20, and the good selection of one-time incremental observation is 15 to 30 for the ten designs problem. Inoue and Chick (1998) [19] illustrate the Bayesian and Frequential method with different kinds of simulation output, and make a comparison between the two methods. They also improve Chick (1997)'s work in sampling costs and information value to select the best system. Based on Bayesian setting of Inoue and Chick 1998 [19], Chen (2010) [9]proposes an efficient procedure to select the optimal subset which concludes the top m systems among k systems under the framework of optimal computing budget allocation.

Nelson and Staum (2006) [28] propose a control-variate procedure which uses a control-variate estimator instead of the sample means. This procedure can provide statistically more accurate but requires more computational effort. Tsai (2009) [39] proposes a sequential procedure that utilizes a more effective sum of differences (controlled sum) than the sum of differences between two systems of KN procedure. Tsai 2009 [40] introduces a new control-variate to decrease the computation cost and remain the statistical efficiency, and proposes a two-stage control-variate combined procedure. Healey, Goldsman and Kim (2007) [17] propose a procedure that employs two new variance estimators, overlapping area and overlapping Cramer-von Mises estimator, to guarantee a certain probability of correct selection. Teng (2010) [38]

proposes a procedure to address the indifference-zone issue under multi-objective optimal computing budget allocation framework. In this paper, a Pareto set is defined by non-dominant systems. The allocation rules are set by Lagrange relaxation method. It assumes that all simulated observations from each system are sampled from independently normally distributed populations.

2.3 Related Work

Although there are a lot of studies focusing on solving stochastic systems selection via computer simulation, the main objective is to improve the efficiency. The work in the literatures which considers both average behaviors and variability of any systems are limited.

Santner and Tamhane (1984) [36] present a selection procedure which selects the best system with the largest mean value and smallest variance value from a non-empty set of good systems. The good system is decided via specifying two integers a and b, where $1 \leq a, b \leq p$, and p is the number of alternative systems. A system with a mean value among the a largest ones and variance among the b smallest ones is defined as a *good* system.

Andradottir and Kim (2010) [1] solve a constrained system selection problem where both the objective function and the constraints are based on the mean measures.

In Batur and Choobineh (2010) [3], the best system is defined as the one with the best (largest or smallest) mean and smallest variance. If there are systems with better mean but poorer variance compared to the others, there is not a single best system, but a number of nondominant systems.

2.4 Objectives of our Study

Risk is unavoidable in real world. The ability of enduring risk for different people varies. In this thesis, we want to establish the dominant set of systems based on the decision maker's preference, especially for risk-averse decision makers. Three comprehensive selection criteria are presented in following chapter.

Chapter 3

Measures for comparison

In the following subsections, we introduce the three selection criteria we consider in this study.

3.1 Mean and Variance

In system selection literature, mean is the only measure that the authors take into account, where, the risk can not be embodied. Another significant element, variance, as a way of measuring risk also needs to be considered. Hence, a selection criterion more comprehensive than the mean measure is the mean-variance dominance selection criterion. In the mean-variance dominance criterion, systems are compared based on both mean and variance measures. Refer to Batur and Choobineh (2009) [2] for a discussion of different selection criteria in ranking and selection.

In the mean-variance based comparison approach, the system with the best (largest or smallest) mean and the smallest variance is deemed as the best system. Suppose $E_A(X)$ and $E_B(X)$ are the mean values and $Var_A(X)$ and $Var_B(X)$ are the variance values of the performance metric of interest for Systems A and B, respectively.

System A has *mean-variance* dominance over System B if $E_A(X) \geq$ (or $\leq$)$E_B(X)$ and $Var_A(X) \leq Var_B(X)$ in the case of larger (or smaller) is better and at least one inequality holds strictly. If there is not a unique best system, a set of nondominant systems exists.

3.2 Mean Comparison with Variance Constraint

Based on mean-variance criterion, a system with the largest mean and variance can be selected as the best system. However, that is not an optimal decision for the person who is a risk-averse decision maker. Here, variance constraint is set for the risk-averse decision maker to control the risk of systems. The systems whose variances are smaller than that of the standard system are considered as feasible systems. The best system is selected based on the mean comparisons from all the feasible systems. That is a way to reduce the probability of choosing a system with high risk for the risk-averse decision maker.

3.3 Baumol's L Criterion

In this paper, the third measure that is considered for selection of the best system is $L = \mu - G\sigma$ [4]. This comparison works better for risk-averse decision maker. In $\mu - G\sigma$, μ is the expected value of the system, σ is the standard deviation (denotes the variability) of the system, and G is a constant which is specified by the decision maker to express the extent of feeling about the risk. When G is a positive number, $\mu - G\sigma$ is less than μ. That means the decision maker would like to trade off some of the expected value to consider the system's variability. The larger the G value the decision maker chooses, the more conservative the decision maker is. That is because

a large G value reduces the probability of the decision maker encountering a loss. For example, when G is equal to 1, for a normally distributed system, the probability of the return being less than $\mu - \sigma$ is only 0.16. If G increases to 1.5, the probability decreases to 0.07. A special case for this problem occurs when G is considered to be 0. In that case, the problem becomes the same as selecting a system based on only the mean measure. The decision maker does not consider the risk (variability) of the systems, and can be called a risk-neutral decision maker. When G is a negative number, $\mu - G\sigma$ is larger than μ. The person who always prefers a negative G is a risk seeking decision maker. This will not be included in this study.

Baumol points out that the L criterion is a derived measure from Markowitz criterion which is based on mean and variance. Therefore, systems selected based on the L criterion are a subset of systems that are selected based on the Markowitz criterion. If a system is not the best system based on the Markowitz criterion, it can not be the best system based on the L criterion from the same set of systems.

Theorem (Baumol (1963)): The Bamoul's L criterion is a derived measure from Markowitz criterion.

Proof: Suppose there are two systems A and B. A dominates B based on the Markowitz criterion. Thus, $\mu_A \geq \mu_B$ and $Var_A \leq Var_B$, where only one equality can exist at the same time. If an equal amount $G\sigma_A$ is subtracted from μ_A and μ_B, $\mu_A - G\sigma_A \geq \mu_B - G\sigma_A$ is obtained. Then $\mu_A - G\sigma_A \geq \mu_B - G\sigma_B$ is assured because $\sigma_A \leq \sigma_B$ and $\sigma_A, \sigma_B \geq 0$. Thus, $L_A = \mu_A - G\sigma_A \geq \mu_B - G\sigma_B = L_B$ is shown when A dominates B based on Markowitz criterion. $\diamond$

Lindon and Peter(1978) [35] clarify L criterion and show another relationship between the first-degree stochastic dominance and the L criterion. The first-degree stochastic dominance implies the L efficiency for any G value. That means if System A first-degree stochastically dominates System B, it also dominates System B based

on L criterion for any G value. On the contrary, if System A dominates Systems B based on the L criterion for any G value, System A also first-degree stochastically dominates Systems B.

Theorem (Lindon and Peter(1978)) [35]: The first-degree stochastic dominance implies the L efficiency for any G value. The reverse also holds.

Proof: Suppose system A first-degree stochastically dominates system B. Then the following equation must be satisfied for all y values,

$$\int_{-\infty}^{y} f_B(x)\ dx \geq \int_{-\infty}^{y} f_A(x)\ dx \tag{3.1}$$

where $y \in [-\infty, \infty]$, and $f_A(x)$ and $f_B(x)$ are the probability density functions of systems A and B, respectively.

It is well known that, for any normally distributed system, the $\mu_i - G\sigma_i$ value is equal to a corresponding quantile value of that system. The way to find the corresponding quantile value is using the Central Limit Theorem. For instance, when G is specified to be 0, the value of $\mu_i - G\sigma_i$ is equal to 0.5th quantile value of the i-th system. Depending on that, for any G value, a quantile q_T that corresponds to $\mu - G\sigma$ can be obtained by

$$\int_{-\infty}^{\mu_i - G\sigma_i} f_i(x)\ dx = \Phi(G) = q_T. \tag{3.2}$$

According to Equations (3.1) and (3.2), for a defined G value, the following can

be obtained:

$$\int_{-\infty}^{\mu_b - G\sigma_b} f_B(x)\ dx \geq \int_{-\infty}^{\mu_b - G\sigma_b} f_A(x)\ dx \tag{3.3}$$

$$\text{and} \int_{-\infty}^{\mu_a - G\sigma_a} f_A(x)\ dx = \int_{-\infty}^{\mu_b - G\sigma_b} f_B(x)\ dx \tag{3.4}$$

$$\Rightarrow \int_{-\infty}^{\mu_a - G\sigma_a} f_A(x)\ dx \geq \int_{-\infty}^{\mu_b - G\sigma_b} f_A(x)\ dx$$

$$\Rightarrow \mu_a - G\sigma_a \geq \mu_b - G\sigma_b$$

Hence $\mu_a - G\sigma_a \geq \mu_b - G\sigma_b$ is proved. In addition, the theorem can be proved backwards for any $G \in [-\infty, \infty]$. Assume that system A dominates system B based on the L criterion. Thus, the corresponding quantile value $\mu_A - G\sigma_A$ is larger than $\mu_B - G\sigma_B$. From Equations (3.3) and (3.4), the conclusion that system A first-degree stochastically dominates system B at that point q_T can be obtained. If this condition is satisfied at any G from negative infinity to positive infinity, system A first-degree stochastically dominates system B is assured.$\diamond$

Chapter 4

Methodology

In this section, three procedures are presented. These procedures compare the performance of the simulated systems based on different measures. Let $\mu_1, \mu_2, \ldots, \mu_K$, and $\sigma_1, \sigma_2, \ldots, \sigma_K$ be the unknown mean and unknown variance of the performance metric of interest of the K systems, respectively. The three problems can be formulated as follows:

1. Max μ_i and Min σ_i^2.

2. Max μ_i
   st. $\sigma_i^2 \leq \sigma_0^2$, where σ_0 is unknown.

3. Max $(\mu_i - G\sigma_i)$, where G is a prespecified parameter.

Let $X_{i\ell}$, $i = 1, 2, \ldots, K$, $\ell = 1, 2, \ldots$ be the ℓ th observation from the simulated system i. We assume that the simulation output data of each system are independent and identically distributed (IID) normal with mean μ_i and variance σ_i^2, for $i = 1, 2, \ldots, K$,

$$\{X_{i\ell} : \ell = 1, 2, \ldots\} \overset{\textbf{IID}}{\sim} \textbf{Normal}(\mu_i, \sigma_i^2).$$

Also, it is assumed that systems are simulated independently.

4.1 Mean-Variance Selection (MVS) Procedure

In the mean - variance based selection (MVS) procedure, the system with the largest μ_i value and smallest σ_i value is called the best system. If two systems have the same mean values but different variances, the system with the smaller variance is better than the other. Similarly if two systems have the same variance but different mean values, the system with the larger mean value is better than the other. Two systems are non-dominant if one system has a larger mean and variance compared to the other.

Considering the uncertainty of establishing dominance with a finite number of simulated observations, the MVS procedure is constructed using the indifference-zone parameters for both the mean and variance measures. The indifference-zone parameter for the mean value is denoted by δ. It is the smallest difference worth detecting between the mean values of two systems at a reasonable sampling cost. The mean indifference-zone parameter δ used here has a slightly different functionality from the indifference-zone parameter used in the mean-based sequential ranking and selection R&S procedures in the literature. In the mean based R&S procedures, the objective is detecting if the mean measure of one system is larger than that of another system, i.e., $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$. When mean values of two systems are within the indifference-zone range, i.e., $|\mu_1 - \mu_2| < \delta$, the experimenter is indifferent as to which system is selected. However, in the MVS procedure, the experimenter wants to determine if

$$\mu_1 > \mu_2,\ \mu_1 = \mu_2,\ \text{or}\ \mu_1 < \mu_2.$$

This way if two systems are indifferent based on the mean measure, the dominant

system is determined based on the variance measure.

By utilizing the indifference-zone parameter if $0 < \mu_1 - \mu_2 < \delta$, then both $\mu_1 > \mu_2$ and $\mu_1 = \mu_2$ decisions are considered correct. Similarly, if $-\delta < \mu_1 - \mu_2 < 0$, then $\mu_1 = \mu_2$ and $\mu_1 < \mu_2$ decisions are considered correct. On the other hand, if $\mu_1 - \mu_2 \geq \delta$, then $\mu_1 > \mu_2$ is the only correct decision. Similarly, if $\mu_1 - \mu_2 \leq -\delta$, then $\mu_1 < \mu_2$ is the only correct decision.

For the variance measure, the decision maker wants to decide if

$$\sigma_1^2/\sigma_2^2 > 1,\ \sigma_1^2/\sigma_2^2 = 1,\ \text{or}\ \sigma_2^2/\sigma_1^2 > 1.$$

The indifference-zone parameter for the ratio of variances is denoted by R^2 where $R > 1$. It is the smallest practical ratio worth detecting in the ratio of variances of two systems.

If $1 < \sigma_1^2/\sigma_2^2 < R^2$, then both $\sigma_1^2/\sigma_2^2 > 1$ and $\sigma_1^2/\sigma_2^2 = 1$ decisions are considered correct. However, if $1 < \sigma_2^2/\sigma_1^2 < R^2$, then $\sigma_2^2/\sigma_1^2 > 1$ and $\sigma_1^2/\sigma_2^2 = 1$ decisions are considered correct. If $\sigma_1^2/\sigma_2^2 \geq R^2$, then $\sigma_1^2/\sigma_2^2 > 1$ is the only correct decision. The same holds for the σ_2^2/σ_1^2 ratio.

In the proposed selection procedure, comparison of simulated systems based on the mean and variance measures are performed simultaneously. For each pairwise comparison based on the mean measure, a three-way hypothesis test is performed. The null hypothesis is the equivalence of the mean measures, i.e., $H_0^\mu : \mu_1 - \mu_2 = 0$. The alternative hypotheses are one mean measure being strictly greater than the other mean measure, i.e., $H_1^\mu : \mu_1 - \mu_2 > 0$ and $H_2^\mu : \mu_1 - \mu_2 < 0$. A similar three-way hypothesis test is performed for the ratio of the variances, i.e., $H_0^\sigma : \sigma_1^2/\sigma_2^2 = 1$, $H_1^\sigma : \sigma_1^2/\sigma_2^2 > 1$, and $H_2^\sigma : \sigma_2^2/\sigma_1^2 > 1$.

Rejection of the null hypothesis when it is true is called a Type I error. The

three-way hypothesis test is developed such that the probability of committing a Type I error is at most α, and α is distributed between the two alternatives, i.e., Pr(Accepting $H_1|H_0$ is true) $\leq \alpha/2$ and Pr(Accepting $H_2|H_0$ is true) $\leq \alpha/2$. Non-rejection of the null hypothesis when it is false is called a Type II error. The three-way hypothesis test is developed such that the probability of committing a Type II error is at most β, i.e., Pr(Do not reject $H_0|H_0$ is false) $\leq \beta$.

In general, in hypothesis testing the user has control over the Type I error; however, the Type II error is uncontrollable. The user can decrease both errors by increasing the number of observations, but it is not possible to control the probability of committing a Type II error unless a specific alternative hypothesis exists. Here in order to be able to control the probability of committing a type II error, the alternative mean hypotheses are transformed into $H_1^\mu : \mu_1 - \mu_2 \geq \delta$ and $H_2^\mu : \mu_1 - \mu_2 \leq -\delta$, where δ is the mean indifference-zone parameter.

According to the transformed three-way test, if the difference between the mean measures is larger than the indifference-zone value δ, the test detects the true difference selecting the correct alternative hypothesis with a probability of at least $1 - \beta$. For example, if $\mu_1 - \mu_2 \geq \delta$, then the decision is $H_1^\mu : \mu_1 - \mu_2 > 0$ with a probability of at least $1 - \beta$. However, if the difference between the mean measures is in the range of 0 to δ, the test detects the true difference selecting the correct alternative hypothesis or fails to reject the null hypothesis with a total probability of at least $1 - \alpha/2$. For example, if $0 < \mu_1 - \mu_2 < \delta$, then the decision is $H_0^\mu : \mu_1 = \mu_2$ or $H_1^\mu : \mu_1 - \mu_2 > 0$ with a total probability of at least $1 - \alpha/2$.

Similarly, the alternative variance hypotheses are transformed into $H_1^\sigma : \sigma_1^2/\sigma_2^2 \geq R^2$ and $H_2^\sigma : \sigma_2^2/\sigma_1^2 \geq R^2$ where $R^2 > 1$ is the variance indifference-zone parameter. Similar to the mean measure if the ratio of the variances is larger than R^2, the test detects the correct alternative hypothesis with a probability of at least $1 - \beta$. For

example if $\sigma_1^2/\sigma_2^2 \geq R^2$, then the decision is $H_1^\sigma : \sigma_1^2/\sigma_2^2 > 1$ with a probability of at least $1-\beta$. However, if the ratio of the variances is in the range of 1 to R^2, then the test detects the true ratio selecting the true alternative hypothesis or fails to reject the null hypothesis with a total probability of at least $1-\alpha/2$. For example, if $1 < \sigma_1^2/\sigma_2^2 < R^2$, then the decision is $H_0^\sigma : \sigma_1^2/\sigma_2^2 = 1$ or $H_1^\sigma : \sigma_1^2/\sigma_2^2 > 1$ with a total probability of at least $1-\alpha/2$.

Paulson (1964) [33] presents a sequential procedure for testing the three-way mean test presented above. The test is shown to satisfy the prespecified probability of committing Type I and II errors, i.e.,

$$\begin{aligned}
\Pr(\text{Do not reject } H_0^\mu \mid |\mu_1-\mu_2| \geq \delta) &\leq \beta, \\
\Pr(\text{Accept } H_1^\mu \mid \mu_1-\mu_2 = 0) &\leq \alpha/2, \\
\Pr(\text{Accept } H_2^\mu \mid \mu_1-\mu_2 = 0) &\leq \alpha/2.
\end{aligned}$$

In the procedure, the parameter d is denoted by a function of the indifference-zone parameter δ. It is used to control the sequential structure of the procedure. Paulson(1964) [33] recommends that the best efficient value of d is equal to $3\delta/8$ after many experiments.

The proposed procedure is presented as follows:

PROCEDURE MVS:

Setup: Specify the probability of correctly determining the MV-efficient set of systems $1-\alpha$. Specify the mean indifference-zone parameter δ and the variance indifference-zone parameter R^2, where $R > 1$. Set parameter d for the mean test to $3\delta/8$ and parameter λ for the variance test to $1+(0.7)(R-1)$. Set the probability

of committing Type I and II errors for each three-way hypothesis test to $\alpha_K = \beta_K = \alpha/[K(K-1)]$. Define S_I as the set of initial K systems. Define S_N as the set of pairs of nondominant systems, and set $S_N = \emptyset$. Define S_E as the set of pairs of equal systems, and set $S_E = \emptyset$. Define S_D as the set of decided systems, and set $S_D = \emptyset$. Set the initial sample size n_0 to $\lfloor 1 - ln(\alpha_K)/ln(\lambda) \rfloor + 1$.

Initialization: Obtain n_0 observations from each system $i \in S_I$. Set the number of stages r to n_0.

Calculation: For every system $i \in S_I \backslash S_D$, compute the sample variance

$$S_i^2(r) = \frac{1}{r-1} \sum_{\ell=1}^{r} (X_{i\ell} - \bar{X}_i(r))^2, \text{ where } \bar{X}_i(r) = \frac{1}{r} \sum_{\ell=1}^{r} X_{i\ell}.$$

For the comparison of the means of system pair (i,j), $i,j \in S_I$, $(i,j) \notin S_N$, $(i,j) \notin S_E$, $i \neq j$, compute the following inner and outer bounds

$$\mathrm{IB}_{ij}^{\mu}(r) = r(\delta - d) - \frac{\Big(S_i^2(r) + S_j^2(r)\Big)\Big[(1/\beta_K)^{\frac{2}{r-1}} - 1\Big](r-1)}{4d},$$

$$\mathrm{OB}_{ij}^{\mu}(r) = rd + \frac{\Big(S_i^2(r) + S_j^2(r)\Big)\Big[(2/\alpha_K)^{\frac{2}{r-1}} - 1\Big](r-1)}{4d}.$$

For the comparison of the variances of system pair (i,j), $i,j \in S_I$, $(i,j) \notin S_N$, $(i,j) \notin S_E$, $i \neq j$, compute the following inner and outer bounds

$$\mathrm{IB}_{ij}^{\sigma}(r) = \frac{\lambda\Big(S_i^2(r)/S_j^2(r)\Big)\Big(\lambda - \beta_K^{\frac{1}{r-1}}\Big)}{\Big(\lambda\beta_K^{\frac{1}{r-1}} - 1\Big)},$$

$$\mathrm{OB}_{ij}^{\sigma}(r) = \frac{\Big(S_i^2(r)/S_j^2(r)\Big)\Big[\lambda - (2/\alpha_K)^{\frac{1}{r-1}}\Big]}{\lambda\Big[\lambda(2/\alpha_K)^{\frac{1}{r-1}} - 1\Big]}.$$

Decision: For every system pair (i,j), $i,j \in S_I$, $(i,j) \notin S_N$, $(i,j) \notin S_E$, $i \neq j$:

If $-\mathrm{IB}_{ij}^{\mu}(r) < \sum_{\ell=1}^{r}(X_{i\ell} - X_{j\ell}) < \mathrm{IB}_{ij}^{\mu}(r)$, decide $\mu_i = \mu_j$.

If $\sum_{\ell=1}^{r}(X_{i\ell} - X_{j\ell}) > \mathrm{OB}_{ij}^{\mu}(r)$, decide $\mu_i > \mu_j$.

If $\sum_{\ell=1}^{r}(X_{i\ell} - X_{j\ell}) < -\mathrm{OB}_{ij}^{\mu}(r)$, decide $\mu_i < \mu_j$.

If $\mathrm{IB}_{ij}^{\sigma}(r) < R^2$ and $\mathrm{IB}_{ji}^{\sigma}(r) < R^2$, decide $\sigma_i^2 = \sigma_j^2$.

If $r > \lfloor 1 - ln(\alpha_K/2)/ln(\lambda) \rfloor$ and $\mathrm{OB}_{ij}^{\sigma}(r) > 1$, decide $\sigma_i^2 > \sigma_j^2$.

If $r > \lfloor 1 - ln(\alpha_K/2)/ln(\lambda) \rfloor$ and $\mathrm{OB}_{ji}^{\sigma}(r) > 1$, decide $\sigma_i^2 < \sigma_j^2$.

Elimination: For every system pair (i,j), $i,j \in S_I$, $(i,j) \notin S_N$, $(i,j) \notin S_E$, $i \neq j$:

If i) $\sigma_i^2 > \sigma_j^2$ and $\mu_i = \mu_j$, ii) $\sigma_i^2 = \sigma_j^2$ and $\mu_i < \mu_j$, or iii) $\sigma_i^2 > \sigma_j^2$ and $\mu_i < \mu_j$ decisions are made, eliminate system i, and set $S_I = S_I/\{i\}$. Delete every system pair in S_N or S_E which includes system i.

If i) $\sigma_i^2 < \sigma_j^2$ and $\mu_i = \mu_j$, ii) $\sigma_i^2 = \sigma_j^2$ and $\mu_i > \mu_j$, or iii) $\sigma_i^2 < \sigma_j^2$ and $\mu_i > \mu_j$ decisions are made, eliminate system j, and set $S_I = S_I/\{j\}$. Delete every system pair in S_N or S_E which includes system j.

If i) $\sigma_i^2 > \sigma_j^2$ and $\mu_i > \mu_j$ and ii) $\sigma_i^2 < \sigma_j^2$ and $\mu_i < \mu_j$ decisions are made, declare systems i and j nondominant, and set $S_N = S_N \cup \{(i,j)\}$.

If $\sigma_i^2 = \sigma_j^2$ and $\mu_i = \mu_j$ decisions are made, declare systems i and j equal, and set $S_E = S_E \cup \{(i,j)\}$.

Termination: If $|S_I| = 1$, stop and declare system $i \in S_I$ as the best system. If the S_N and S_E sets include every (i,j) pair where $i,j \in S_I, i \neq j$, stop and declare S_I as the MV-efficient set of systems; the system pairs in S_N are declared nondominant and the system pairs in S_E are declared equal.

Otherwise, for $i \in S_I$, if every (i,j) pair, $j \in S_I, i \neq j$, is in S_N or S_E, add system i to S_D. Obtain one more observation from every system $i \in S_I \backslash S_D$. Set $r = r + 1$.

While going back to the Calculation step, clear all test results for system pairs (i, j), $i, j \in S_I$, $(i, j) \notin S_N \cup S_E$, $i \neq j$. $\diamond$

In the Setup step, the following parameters are specified: i) the probability of correctly determining the MV-efficient set of systems, $1-\alpha$; ii) the mean indifference-zone parameter, δ; and iii) the variance indifference-zone parameter, R^2. The constant parameter d for the mean test and the constant parameter λ for the variance test are determined based on δ and R, respectively. The probability of committing Type I α_K or II error β_K for each test is set to $\alpha/[K(K-1)]$. Set S_I is the set of systems still under consideration; it initially includes all K systems. Set S_N is defined as the set of system pairs that are declared nondominant. Set S_E is defined as the set of system pairs that are declared equal. Set S_D is defined as the set of systems that are declared decided; no more sampling is needed from a decided system. All three sets are initially empty sets. Finally, the initial sample size n_0 is determined by α_K and λ.

In the Initialization step, n_0 observations obtained from every system $i \in S_I$, and the number of replications r is set to n_0. In the Calculation step, the sample variances are computed for every system $i \in S_I \backslash S_D$. The bounds of the continuation region $\mathrm{IB}^{\mu}_{ij}(r)$ and $\mathrm{OB}^{\mu}_{ij}(r)$ are computed for the mean hypothesis test, and $\mathrm{IB}^{\sigma}_{ij}(r)$ and $\mathrm{OB}^{\sigma}_{ij}(r)$ are computed for the variance hypothesis test. In the Decision step, decisions of the hypothesis tests for each pair of systems are made by using these bounds and the indifference-zone parameters. For a pair of systems, if none of the conditions are satisfied, the hypothesis test is inconclusive at stage r.

In the Elimination step, elimination, nondominance, or equality decisions are made according to the results of the hypothesis tests. If a system is declared to be worse than another system in S_I based on both the mean and variance measures, that system is eliminated and deleted from S_I. On the other hand, if two systems are

declared nondominant, the pair is added to the S_N set. If two systems are declared equal, the pair is added to the S_E set.

In the Termination step, there are two stopping rules. The first rule is to stop the procedure if the S_I set includes only one system. In this case, the system in the S_I set is declared the best system with a $1-\alpha$ probability of correct selection. The second rule is to stop the procedure if all pairs of the systems in the S_I set is included in the S_N or S_E set. This means that for every pair of systems that have not been eliminated, a nondominance or an equality decision is made. In this case, the S_I set is declared the MV-efficient set of systems with a $1-\alpha$ probability of correct selection; the systems in S_N and S_E are declared nondominant and equal, respectively, based on the MV selection criterion.

If none of the stopping conditions are satisfied, the S_D set is updated. For system $i \in S_I$, if every pair (i,j), $j \in S_I, i \neq j$ is included in S_N or S_E, system i is added to S_D. For example, suppose that $S_I = \{2,5,6\}$, $S_N = \{(2,6)\}$, and $S_E = \{(2,5)\}$, then System 2 is added to S_D. One more observation is obtained from every system in the $S_I \backslash S_D$ set. Stage number r is increased by one, and the procedure goes back to the Calculation step. The detailed matlab code for procedure MVS is in Appendix B.

In MVS procedure, $K(K-1)/2$ pairwise comparisons are performed in the comparison of K systems for each measure. The probability of committing Type I or II errors for each pairwise comparison and measure is set to $\alpha/[K(K-1)]$. This assures that the overall probability of selecting the correct set of nondominant systems is at least $1-\alpha$.

Theorem 1 : The MVS procedure correctly determines the MVS-efficient set of systems among K systems with at least a probability of $1-\alpha$.

Proof: Let CS denote the overall correct selection, MCS and $IMCS$ denote the correct and incorrect mean comparison selection decisions, respectively. Let VCS

and $IVCS$ denote the correct and incorrect variance comparison selection decisions, respectively, $IMCS_\ell$ denote the incorrect selection for the mean comparison of the ℓ-th pair, and $IMVS_\ell$ denote the incorrect selection for the variance comparison of the ℓ-th pair.

$$
\begin{aligned}
Pr(CS) &= Pr(MCS \cap VCS) \\
&\geq Pr(MCS) + Pr(VCS) - 1 \qquad (4.1) \\
&= (1 - Pr(IMCS)) + (1 - Pr(IVCS)) - 1 \\
&= (1 - \cup_{\ell=1}^{K(K-1)/2} Pr(IMCS_\ell)) + (1 - \cup_{\ell=1}^{K(K-1)/2} Pr(IVCS_\ell)) - 1 \\
&\geq (1 - \sum_{\ell=1}^{K(K-1)/2} \frac{\alpha}{K(K-1)}) + (1 - \sum_{\ell=1}^{K(K-1)/2} \frac{\alpha}{K(K-1)}) - 1 \qquad (4.2) \\
&= 1 - (\alpha/2) - (\alpha/2) \\
&= 1 - \alpha
\end{aligned}
$$

The two Equations (4.1) and (4.2) are correct because of the Bonferroni inequality.

4.2 Constrained Mean-Variance Selection (CMVS) Procedure

In the constrained mean variance selection (CMVS) procedure, the objective is to select a system with the largest mean value; meanwhile, its variance must be less than that of a standard system. The formulation of the problem is Max μ_i, st. $\sigma_i^2 \leq \sigma_0^2$, where σ_0 is the unknown variance of the standard system.

There are two phases in the CMVS procedure. First, the feasible systems whose

variance is smaller than that of standard system need to be found. Then, the system with the largest mean value is selected from the feasible systems. The selection behavior under this criterion tells us that: i) a system with the largest mean value can not become the best system if it has a larger variance than that of the standard system, ii) a system with a smaller mean value is the best system if all the other systems with larger mean have larger variance than the standard system. Here, nondominant set of systems can also exist, i.e., two systems pass the feasibility check and they have the same mean values.

PROCEDURE CMVS:

Setup: Specify the probability of correctly determining the CMV-efficient set of systems $1-\alpha$. Specify the mean indifference-zone parameter δ and the variance indifference-zone parameter R^2. Set parameter d for the mean test to $3\delta/8$ and parameter λ for the feasibility check to $1+(0.7)(R-1)$. Set the probability of Type I error α_{K_1} and Type II error β_{K_1} of each hypothesis test in the feasibility check phase to $\alpha/(2K)$. Set α_{K_2} and β_{K_2} of each three-way hypothesis in the selection-of-the-best phase to $\alpha/[K(K-1)]$. Define S_I as the set of initial K systems, and system 0 as the standard system with the unknown variance measure σ_0^2. Define S_F as the set of feasible systems, and set $S_F = \emptyset$. Define S_N as the set of pairs of nondominant systems, and set $S_N = \emptyset$. Define S_D as the set of decided systems, and set $S_D = \emptyset$. Set the initial sample size n_0 to $\lfloor 1 - ln(\alpha_{K_1})/ln(\lambda) \rfloor + 1$.

Initialization: Obtain n_0 observations from each system $i \in S_I$ and the standard system. Set the number of stages r to n_0.

Calculation: For every system $i \in S_I \cup S_F \backslash S_D$ and the standard system 0 if

$|S_I| \neq 0$, compute the sample variance

$$S_i^2(r) = \frac{1}{r-1}\sum_{\ell=1}^{r}(X_{i\ell} - \bar{X}_i(r))^2, \text{ where } \bar{X}_i(r) = \frac{1}{r}\sum_{\ell=1}^{r} X_{i\ell}.$$

For the feasibility check of system $i \in S_I$, compute the following lower and upper bounds

$$\mathrm{LB}_i^{\sigma}(r) = \frac{\lambda\Big(S_i^2(r)/S_0^2(r)\Big)\Big(\lambda - \beta_{K_1}^{\frac{1}{r-1}}\Big)}{\Big(\lambda\beta_{K_1}^{\frac{1}{r-1}} - 1\Big)},$$

$$\mathrm{UB}_i^{\sigma}(r) = \frac{\Big(S_i^2(r)/S_0^2(r)\Big)\Big[\lambda - (1/\alpha_{K_1})^{\frac{1}{r-1}}\Big]}{\lambda\Big[\lambda(1/\alpha_{K_1})^{\frac{1}{r-1}} - 1\Big]}.$$

For the comparison of the means of system pair (i,j), $i,j \in S_I \cup S_F$, $(i,j) \notin S_N$, $i \neq j$, compute the following inner and outer bounds

$$\mathrm{IB}_{ij}^{\mu}(r) = r(\delta - d) - \frac{\Big(S_i^2(r) + S_j^2(r)\Big)\Big[(1/\beta_{K_2})^{\frac{2}{r-1}} - 1\Big](r-1)}{4d},$$

$$\mathrm{OB}_{ij}^{\mu}(r) = rd + \frac{\Big(S_i^2(r) + S_j^2(r)\Big)\Big[(2/\alpha_{K_2})^{\frac{2}{r-1}} - 1\Big](r-1)}{4d}.$$

Feasibility check:For every system $i \in S_I$:

If $\mathrm{LB}_i^{\sigma}(r) < R^2$, then decide $H_0^{\sigma} : \sigma_i^2 \leq \sigma_0^2$. Move system i from S_I to S_F. Set $S_I = S_I/\{i\}$ and $S_F = S_F \cup \{i\}$.

If $\mathrm{UB}_i^{\sigma}(r) > 1$, then decide $H_1^{\sigma} : \sigma_i^2 > \sigma_0^2$. Eliminate system i and set $S_I = S_I/\{i\}$.

Decision: For every system pair (i,j), $i,j \in S_F$, $(i,j) \notin S_N$, $i \neq j$:

If $-\mathrm{IB}_{ij}^{\mu}(r) < \sum_{\ell=1}^{r}(X_{i\ell} - X_{j\ell}) < \mathrm{IB}_{ij}^{\mu}(r)$, decide $\mu_i = \mu_j$.

If $\sum_{\ell=1}^{r}(X_{i\ell} - X_{j\ell}) > \mathrm{OB}_{ij}^{\mu}(r)$, decide $\mu_i > \mu_j$.

If $\sum_{\ell=1}^{r}(X_{i\ell} - X_{j\ell}) < -\mathrm{OB}_{ij}^{\mu}(r)$, decide $\mu_i < \mu_j$.

For every system pair $(i,j), i \in S_F, j \in S_I, i \neq j$:

If $\sum_{\ell=1}^{r}(X_{i\ell} - X_{j\ell}) > \mathrm{OB}_{ij}^{\mu}(r)$, decide $\mu_i > \mu_j$.

Elimination: For every system pair (i,j), $i,j \in S_F$, $(i,j) \notin S_N$, $i \neq j$:

If $\mu_i < \mu_j$ decision is made, eliminate system i, and set $S_F = S_F/\{i\}$. Delete every system pair in S_N which includes system i.

If $\mu_i > \mu_j$ decision is made, eliminate system j, and set $S_F = S_F/\{j\}$. Delete every system pair in S_N which includes system j.

If $\mu_i = \mu_j$ decision is made, declare systems i and j nondominant, and set $S_N = S_N \cup \{(i,j)\}$.

For every system pair $(i,j), i \in S_F, j \in S_I, i \neq j$:

If $\mu_i > \mu_j$ decision is made, eliminate system j, and set $S_I = S_I/\{j\}$.

Termination: If $|S_I| = 0$ and $|S_F| = 1$, stop and declare system $i \in S_F$ as the best system. If $|S_I| = 0$ and $|S_F| = 0$, stop and declare that there is no feasible system. If $|S_I| = 0$ and the S_N set includes every (i,j) pair where $i,j \in S_F, i \neq j$, stop and declare S_F as the CMV-efficient set of systems.

Otherwise, when $|S_I| = 0$, for $i \in S_F$, if every (i,j) pair, $j \in S_F, i \neq j$, is in S_N, add system i to S_D. Obtain one more observation from every system $i \in S_I \cup S_F \backslash S_D$ and the standard system 0 if $|S_I| \neq 0$, set $r = r + 1$, and go back to the Calculation step.$\diamond$

In the Setup step, the following parameters are specified: i) the probability of correctly determining the CMV-efficient set of systems, $1-\alpha$; ii) the mean indifference-zone parameter, δ; and iii) the variance indifference-zone parameter, R^2. The constant parameters d and λ are determined based on δ and R, respectively. The probability of Type I or II error is set to $\alpha_{K_1} = \beta_{K_1} = \alpha/(2K)$ for the feasibility check phase

and $\alpha_{K_2} = \beta_{K_2} = \alpha/[K(K-1)]$ for the mean selection phase. Set S_I is the set of all systems still under consideration; it initially includes all systems. Set S_F is defined as the set of systems that are declared feasible. Set S_N is defined as the set of feasible system pairs that are declared nondominant. Set S_D is defined as the set of decided systems which are declared nondominant with all the feasible systems when the S_I set is empty. No more sampling is needed from a decided system. All three sets are initially empty sets. Finally, the initial sample size n_0 is determined by α_{K_1} and λ.

In the Initialization step, n_0 observations are obtained from every system in S_I and the standard system. In the Calculation step, the sample means and variances are computed for systems in the $S_I \cup S_F \backslash S_D$ set and the standard system 0 if feasibility check phase is still needed. The bounds of the continuation region LB_i^σ and UB_i^σ are computed to check the feasibility of system $i \in S_I$. The bounds of the continuation region IB_i^μ and OB_i^μ are computed for the mean comparison of system pair (i, j), $i, j \in S_I \cup S_F, (i, j) \notin S_N, i \neq j$, to find the feasible system with the largest mean.

In the Feasibility Check step, system i is moved to the S_F set if the decision that the variance of system $i \in S_I$ is smaller than or equal to the variance of the standard system is made. In the Decision step, pairwise mean comparison hypothesis tests are performed.

In the Elimination step, system $i \in S_F$ can eliminate any system which has smaller mean value. Any two systems in S_F, which are declared nondominant, are moved to S_N. It is important to notice that a system in S_I can not eliminate any system before it passes the feasibility check.

In the Termination step, the procedure stops in three cases: i) If no system is in the S_I set and only one system is in the S_F set, the system in S_F is declared the best system with a $1 - \alpha$ probability of correct selection. ii) If no system is left in the S_I and S_F sets, the problem is declared infeasible with a $1 - \alpha$ probability of correct

selection. iii) If no system is in the S_I set and the S_N set includes every (i, j) pair, $i, j \in S_F, i \neq j$, then S_F is declared the CMV-efficient set of systems with a $1 - \alpha$ probability of correct selection.

If the procedure does not terminate, the S_D set is updated. One more observation is obtained from the systems still under consideration, the stage number is increased by one, and the procedure goes back to the Calculation step.

In the comparison of K systems, K pairwise variance comparisons between the standard system and the other K systems for the feasibility check are performed. $K(K-1)/2$ pairwise mean value comparisons among all of K systems are performed.

Theorem 2: The CMVS procedure correctly determines CMVS-efficient set of systems among K systems with at least a probability of $1 - \alpha$.

Proof: Let CS denote the overall correct selection, FCS and $IFCS$ denote the correct and incorrect feasibility check selection decisions, MCS and $IMCS$ denote the correct and incorrect mean comparison selection decisions, respectively. Let $IFCS_\ell$ denote the incorrect selection for the feasibility check of the ℓ-th pair, and $IMVS_\ell$ denote the incorrect selection for the mean comparison of the ℓ-th pair.

$$
\begin{aligned}
Pr(CS) &= Pr(FCS \cap MCS) \\
&\geq Pr(FCS) + Pr(MCS) - 1 \qquad (4.3) \\
&= (1 - Pr(IFCS)) + (1 - Pr(IMCS)) - 1 \\
&= (1 - \cup_{\ell=1}^{K} Pr(IFCS_\ell)) + (1 - \cup_{\ell=1}^{K(K-1)/2} Pr(IMCS_\ell)) - 1 \\
&\geq (1 - \sum_{i=1}^{K} \frac{\alpha}{2K}) + (1 - \sum_{i=1}^{K(K-1)/2} \frac{\alpha}{K(K-1)}) - 1 \qquad (4.4) \\
&= 1 - (\alpha/2) - (\alpha/2) \\
&= 1 - \alpha.
\end{aligned}
$$

The two Equations (4.3) and (4.4) hold by the Bonferroni inequality.

4.3 Risk-Adjusted Mean Selection (RAMS) Procedure

In the risk-adjusted mean based selection (RAMS) procedure, the objective is to select the system with the maximum $L_i = \mu_i - G\sigma_i$ value. A three way hypothesis test is conducted for every system pair (i, j) based on L:

$$
\begin{aligned}
H_0^L &: L_i - L_j = 0, \qquad (4.5) \\
H_1^L &: L_i - L_j > 0, \\
H_2^L &: L_i - L_j < 0.
\end{aligned}
$$

The system with the largest value of L is selected as the best system. Two systems

are equal if their L values are the same. In the RAMS procedure, the method of calculating L value for each system is through finding the corresponding quantile value of that system. According to Section 2, we have already known, for a defined G, the quantile is equal to $\Phi(G)$ according to

$$\Pr(X_i \leq \mu_i - G\sigma_i) = \Phi(-G) = q. \tag{4.6}$$

Thus, our problem is turned to the comparison of q quantile values of each system pair.

4.3.1 Quantile Estimation

Let $X_1, X_2, \ldots, X_N$ be a random sample of size N from an absolutely continuous distribution function $F(x)$ with a probability density function $f(x)$, and $X_{[1]} \leq X_{[2]} \leq \ldots \leq X_{[N]}$ be the corresponding order statistics. The qth quantile, say $x^{[q]}$, $0 < q < 1$, of the X random variable with distribution F, is defined as $x^{[q]} = inf\{x : F(x) \geq q\}$, where *inf* indicates the infimum or the greatest lower bound.

A nonparametric point estimator of $x^{[q]}$ is the qth sample quantile $\hat{x}^{[q]} = X_{[\lfloor Nq+1 \rfloor]}$, where $\lfloor z \rfloor$ denotes the integral part of the real number z. If $f(x)$ is differentiable in the neighborhood of $x^{[q]}$ and $f(x^{[q]}) \neq 0$, the $\hat{x}^{[q]}$ is asymptotically normally distributed as N becomes larger.

The quantile estimator $\hat{x}^{[q]} = X_{[\lfloor Nq+1 \rfloor]}$ is often an inefficient estimator since it has large variance and bias for a finite N. A class of estimators that remedies these shortcomings takes a convex combination of two or more order statistics using an appropriate weight function. Among these estimators, the ones that are based on two order statistics are known to provide better results in terms of bias although they may not perform well in terms of the mean squared error (MSE). Hence, the N value

needs to be large enough to obtain an unbiased estimate even for the tail quantiles.

Parish (1990) [30] studied the behavior of a number of quantile estimators that are linear functions of order statistics under normal sampling. He estimated the bias and MSE of the estimators and concluded that the Paren (1979) [31] quantile estimator has the smallest average bias and low MSE. The estimator [31] is

$$\hat{x}^{[q]} = X_{[w]}(0.5 + w - Nq) + X_{[w+1]}(0.5 - w + Nq),$$

where $w = \lfloor Nq + 0.5 \rfloor$.

4.3.2 Quantile-Window Approach

Let q_T be the target quantile that corresponds to G by $q_T = \Phi(-G)$. The target quantile q_T happens to be a real number with infinite number of decimal digits sometimes. For example, when G is specified to be 1, q_T is 0.158655.... Estimating $x^{[q]}$ accurately using a finite sample of N observations when q_T has infinite number of decimal digits is quite challenging. Achieving high accuracy is critically important and difficult especially when H_0^L is true.

A quantile-window approach is developed to help compare the systems at a q_T quantile accurately when q_T has infinite number of decimal digits. In the quantile-window approach, a $Q = \{q_\ell, q_u\}$ set is determined such that i) q_ℓ and q_u have finite numbers of decimal digits, ii) $q_\ell N$ is the largest integer less than $q_T N$, and iii) $q_u N$ is the smallest integer larger than $q_T N$. In the RAMS procedure, instead of comparing the q_T quantile values for every pair of systems directly, the systems are compared based on the q_ℓ and q_u quantile values assuming that the quantile window is chosen narrow enough for the distributions not to possibly cross more than once in the interval defined by Q. If the comparison of the systems based on the q_ℓ and

q_u quantiles are consistent, it means there is not a crossover between the quantiles and the same comparison result holds for the q_T quantile as well. However, if the comparison results are inconsistent, it means there is a crossover between the two quantiles and a tighter window around q_T with a larger N value is needed.

Table 4.1 shows how quantile boundaries and the N values are changed based on the results of the comparisons when $G = 1$. First, N is set to 100, the two quantile boundaries are $q_\ell = 0.15$ and $q_u = 0.16$. When the decisions on both quantiles are made, a consistency check is performed. If the decisions are different at the current N value, N is increased by a multiple of 10 and the Q set is updated. Two new quantile boundaries q_ℓ and q_u are generated with more decimal digits. The new boundaries are closer to the q_T value. Hence, the quantile values can be estimated with more accuracy. As N increases to infinity, the two boundary values are changed such that they both converge to q_T. The final decision based on L_i is made when the comparison decisions of q_ℓ and q_u are consistent.

Table 4.1: Quantile Window Q sets at $G = 1$

N	100	1000	10,000	100,000	$\cdots$
q_ℓ	0.15	0.158	0.1586	0.15865	$\cdots$
q_T	$0.158655\cdots$	$0.158655\cdots$	$0.158655\cdots$	$0.158655\cdots$	$0.158655\cdots$
q_u	0.16	0.159	0.1587	0.15866	$\cdots$

Generally, if the crossover point of a system pair is far away from the q_T quantile, the initial N value of 100 is enough to make a consistent decision for both quantile boundaries. If the crossover point of a system pair is in the range of current quantile boundaries, N should be increased. The quantile boundaries are narrowed down to the point where a consistent decision can be reached at both quantiles.

4.3.3 Procedure RAMS

PROCEDURE RAMS:

Setup: Specify the probability of correctly determining the RAM-efficient set of systems $1-\alpha$, and the quantile indifference-zone parameter δ_x. Set parameter d for the mean test to $3\delta_x/8$. Set the probability of committing Type I error α_K or II error β_K for the L comparison three-way hypothesis test to $\alpha/(K(K-1))$. Define S_I as the set of initial K systems. Define S_E as the set of pairs of equal systems, and set $S_E = \emptyset$. Define S_D as the set of decided systems, and set $S_D = \emptyset$. Find the target quantile q_T value via $\Phi(-G) = q_T$. Set the initial N value to 100. Set the initial number of samples n_0 to 10.

Initialization: Determine the $Q = \{q_\ell, q_u\}$ set such that $q_\ell = \lfloor q_T N \rfloor / N$ and $q_u = \lceil q_T N \rceil / N$. Obtain n_0 independent samples of N IID observations from every system $i \in S_I \backslash S_D$. For every $i \in S_I \backslash S_D$ and quantile $q \in Q$, compute the quantile estimates $\hat{x}_{i,1}^{[q]}, \hat{x}_{i,2}^{[q]}, \ldots, \hat{x}_{i,n_0}^{[q]}$. Set the number of stages r to n_0.

Calculation:For every system $i \in S_I \backslash S_D$ and quantile $q \in Q$, compute the sample variance of the quantile estimates

$$S_i^2(r)^{[q]} = \frac{1}{r-1} \sum_{\ell=1}^{r} (\hat{x}_{i\ell}^{[q]} - \hat{x}_i^{[q]}(r))^2, \text{ where } \hat{x}_i^{[q]}(r) = \frac{1}{r} \sum_{\ell=1}^{r} \hat{x}_{i\ell}^{[q]}.$$

For every system pair (i,j), $i,j \in S_I$, $(i,j) \notin S_E$, $i \neq j$, and quantile $q \in Q$, compute the following inner and outer bounds

$$\mathrm{IB}_{ij}^{[q]}(r) = r(\delta - d) - \frac{\left(S_i^2(r)^{[q]} + S_j^2(r)^{[q]}\right)\left[(1/\beta_K)^{\frac{2}{r-1}} - 1\right](r-1)}{4d},$$

$$\mathrm{OB}_{ij}^{[q]}(r) = rd + \frac{\left(S_i^2(r)^{[q]} + S_j^2(r)^{[q]}\right)\left[(2/\alpha_K)^{\frac{2}{r-1}} - 1\right](r-1)}{4d}.$$

Decision: For every system pair (i,j), $i,j \in S_I$, $(i,j) \notin S_E$, $i \neq j$, and quantile $q \in Q$:

If $-\mathrm{IB}_{ij}^{[q]}(r) < \sum_{\ell=1}^{r}(\hat{x}_{i\ell}^{[q]} - \hat{x}_{j\ell}^{[q]}) < \mathrm{IB}_{ij}^{[q]}(r)$, decide $x_i^{[q]} = x_j^{[q]}$.

If $\sum_{\ell=1}^{r}(\hat{x}_{i\ell}^{[q]} - \hat{x}_{j\ell}^{[q]}) > \mathrm{OB}_{ij}^{[q]}(r)$, decide $x_i^{[q]} > x_j^{[q]}$.

If $\sum_{\ell=1}^{r}(\hat{x}_{i\ell}^{[q]} - \hat{x}_{j\ell}^{[q]}) < -\mathrm{OB}_{ij}^{[q]}(r)$, decide $x_i^{[q]} < x_j^{[q]}$.

Consistency Check: For every system pair (i,j), $i,j \in S_I$, $(i,j) \notin S_E$, $i \neq j$:

If $x_i^{[q_\ell]} = x_j^{[q_\ell]}$ and $x_i^{[q_u]} = x_j^{[q_u]}$ decisions are made, decide $x_i^{[q_T]} = x_j^{[q_T]}$.

If $x_i^{[q_\ell]} > x_j^{[q_\ell]}$ and $x_i^{[q_u]} > x_j^{[q_u]}$ decisions are made, decide $x_i^{[q_T]} > x_j^{[q_T]}$.

If $x_i^{[q_\ell]} < x_j^{[q_\ell]}$ and $x_i^{[q_u]} < x_j^{[q_u]}$ decisions are made, decide $x_i^{[q_T]} < x_j^{[q_T]}$.

Elimination: For every system pair (i,j), $i,j \in S_I$, $(i,j) \notin S_E$, $i \neq j$:

If $x_i^{[q_T]} = x_j^{[q_T]}$ decision is made, declare systems i and j equal, and set $S_E = S_E \cup \{(i,j)\}$.

If $x_i^{[q_T]} > x_j^{[q_T]}$ decision is made, eliminate system j, and set $S_I = S_I/\{j\}$. Delete every system pair in S_E which includes system j.

If $x_i^{[q_T]} < x_j^{[q_T]}$ decision is made, eliminate system i, and set $S_I = S_I/\{i\}$. Delete every system pair in S_E which includes system i.

Termination: If $|S_I| = 1$, stop and declare system $i \in S_I$ as the best system. If the S_E set includes every (i,j) pair where $i,j \in S_I, i \neq j$, stop and declare S_I as the RAM-efficient set of systems; systems in S_E are declared equal.

Otherwise, for $i \in S_I$, if every (i,j) pair, $j \in S_I, i \neq j$, is in S_E, then add system i to S_D. If all (i,j), $i,j \in S_I$, $i \neq j$, pairwise comparisons reached to a decision at the current N value for both q_ℓ and q_u quantiles, go back to the Initialization step with

$N = N * 10$. Otherwise, obtain one more sample of N IID observations from every system $i \in S_I \backslash S_D$, set $r = r + 1$, and go back to the Calculation step. While going back, clear all q_l and q_u test results for system pairs (i, j), $i, j \in S_I$, $(i, j) \notin S_E$, $i \neq j$.
$\diamond$

In the Setup step, a number of parameters are specified: i) the probability of correctly determining the RAM-efficient set of systems, $1 - \alpha$ and ii) the quantile Indifference-Zone parameter, δ. The test parameter d is set to $3\delta/8$. The probability of committing Type I and II errors for the L comparison three-way hypothesis test is set to $\alpha_K = \beta_K = 2\alpha/(K(K-1))$. Set S_I is defined as the set of systems still under consideration; it initially includes all K systems. Set S_E is the set of system pairs that are declared equal. Set S_D is defined as the set of systems that are declared decided. No more sampling is needed from a decided system. The target quantile q_t is determined from G by the $q_t = \Phi(-G)$ relationship. The initial value of the sample size N is 100. $Q = \{q_\ell, q_u\}$ set is determined such that $q_\ell N$ is the largest integer less than $q_T N$ and $q_u N$ is the smallest integer larger than $q_T N$. The initial number of samples n_0 is set to 10.

In the Initialization step, n_0 Independent samples of N IID observations are obtained from every system $i \in S_I$. In the Calculation step, the bounds of the continuation regions $\mathrm{IB}_{ij}^{[q]}$ and $\mathrm{OB}_{i,j}^{[q]}$ are computed.

In the Decision step, three different decisions can be made for each pair of systems and each $q \in Q$. In the Consistency Check step, if decisions for both q_ℓ and q_u quantiles of a system pair are consistent, the decision for the q_T quantile is made according to the result of each hypothesis test. In the Elimination step, systems that are found to be inferior based on the L measure are eliminated. A system pair is added to the S_E set if they are declared equal.

In the Termination step, the procedure stops in two cases: i) There is only one

system left in the S_I set. In this case, the system in the S_I set is declared the best system with a $1-\alpha$ probability of correct selection. ii) If the S_E set includes every (i,j) pair, where $i,j \in S_I$, the S_I set is declared the RAM-efficient set of systems with a $1-\alpha$ probability of correct selection.

Otherwise, if a decision is made for all pair of systems based on the current N value, we clear all of the previous comparison results for $i \in S_I \backslash S_D$ and go back to Initialization step with $N = N * 10$ changing quantile boundaries. Otherwise, one more sample is obtained from each system $i \in S_I \backslash S_D$, stage number r is increased by one, and the procedure goes back to Calculation step.

In the comparisons of L for K systems, $K(K-1)/2$ pairwise comparisons are performed. The probability of Type I and II errors for each pairwise comparison of L for each quantile value is $\alpha_K = \beta_K = \alpha/K(K-1)$. Hence, the overall probability of correctly selecting the best system among K systems based on L can be assured to be above $1-\alpha$ by Bonferroni inequality.

Theorem 3: The RAMS procedure correctly determines the RAMS-efficient sets of systems among K systems with at least a probability of $1-\alpha$.

Proof: Let CS and ICS denote the overall correct and incorrect L comparison selection,denote the overall incorrect selection decisions, respectively. ICS_{q_ℓ} denote the incorrect selection for the L comparison of the ℓ-th system pair at q_ℓ, and ICS_{q_u} denote the incorrect selection for the L comparison of the ℓ-th system pair at q_u.

$$
\begin{aligned}
Pr(CS) &= Pr(CS_{q_\ell} \cap CS_{q_u}) \\
&\geq Pr(CS_{q_\ell}) + Pr(CS_{q_u}) - 1 \qquad (4.7) \\
&= (1 - Pr(ICS_{q_\ell})) + (1 - Pr(ICS_{q_u})) - 1 \\
&= (1 - \cup_{\ell=1}^{K(K-1)/2}(Pr(ICS_{q_\ell}))) + (1 - \cup_{\ell=1}^{K(K-1)/2}(Pr(ICS_{q_u}))) - 1 \\
&\geq 2(1 - \sum_{i=1}^{K(K-1)/2} \frac{\alpha}{K(K-1)}) - 1 \qquad (4.8) \\
&= 1 - \alpha
\end{aligned}
$$

The two Equations (4.7) and (4.8) in the above are correct because of the Bonferroni inequality.

Chapter 5

Experiments

In this part, the performance of MVS, CMVS, RAMS procedures are shown. In the experiments, all system distributions are normal distribution. The nominal confidence level is set to $1 - \alpha = 0.95$. The mean indifference-zone parameter δ is shown at the table title for each configuration. The variance indifference value R^2 $(R > 1)$ is set to 1.1 for the MVS and CMVS procedures. For the RAMS procedure, the quantile indifference-zone parameter δ_x is set to 0.05. We replicate each experiment 10,000 times. The percentage of correct selection (PCS) and the sample average of total number of observations $(SATO)$ are two main statistics reported in the results table.

5.1 Experiments for MVS Procedure

In Table (5.1), five configurations are presented for the MVS procedures. Configurations 1-3 are taken from Batur and Choobineh 2010 [3]. They are utilized to test the effectiveness of the MVS procedure in detecting the smallest difference on the mean and variance measures. In Configuration 1, the mean and variance values of System 1 are equal to 1. The mean and variance values of System 2 are $1-\delta$ and R, respectively.

Table 5.1: Configurations 1-3 of MVS procedure, $\delta = 0.05$.

		Sys. 1	Sys. 2	Correct Selection
Configuration 1	μ	1	$1-\delta$	$S_I = \{1\}$
	σ	1	R	
Configuration 2	μ	1	$1-\delta$	$S_I = \{1\}$
	σ	1	1	
Configuration 3	μ	1	1	$S_I = \{1\}$
	σ	1	R	

These mean and variance configurations are called *slippage* configurations because the difference of the mean values between the first system and the second system is equal to the mean indifference-zone parameter δ. Similarly, the ratio of the variances of the second system to the first system is equal to the variance indifference-zone parameter R^2. In Configuration 2, the mean measures have the slippage configuration and the variances are all one. In Configuration 3, the mean values are equal to 1, but the variances have the slippage configuration. The best system is System 1 for these three configurations because of the larger mean value and the smaller variance value of System 1. The nondominant and the equality system sets are empty for the three configurations.

Table 5.2: Results for Configurations 1-3 of MVS procedure

	PCS	$SATO$
Configuration 1	1.000	16,006
Configuration 2	0.980	14,724
Configuration 3	0.990	19,790

The results of MVS procedure for Configurations 1-3 are shown in Table (5.2). All PCS results are above the prespecified confidence level 0.95. Configuration 1 needs more observations than Configuration 2 because the larger variance of System 2 in

Configuration 1 makes the mean hypothesis test in Configuration 1 more difficult than the mean hypothesis test in Configuration 2. Configuration 3 needs more observations than Configuration 1 since making a correct equality decision in a hypothesis test is more difficult than making a correct non-equality decision.

In Table (5.3), we test the effectiveness of the MVS procedure in identifying two equivalent systems and two nondominant systems in Configurations 4 and 5, respectively. In Configuration 4, the mean values of Systems 1 and 2 are the same, as well as the variances. They are expected to be selected as an equal system pair in the S_E set. In Configuration 5, System 2 has a larger mean value than System 1, but also has a larger variance. Thus, they are nondominant systems, and are expected to be selected as a nondominant system pair in the S_N set.

Table 5.3: Configurations 4-5 of MVS procedure, $\delta = 0.05$.

		Sys. 1	Sys. 2	Correct Selection
Configuration 4	μ	1	1	$S_I = \{1, 2\}, S_E = \{(1, 2)\}$
	σ	1	1	
Configuration 5	μ	1	$1 + \delta$	$S_I = \{1, 2\}, S_N = \{(1, 2)\}$
	σ	1	R	

Table 5.4: Results for Configurations 4-5 of MVS procedure

	PCS	$SATO$
Configuration 4	0.980	18,008
Configuration 5	0.980	16,190

The results of the MVS procedure for Configurations 4-5 are shown in Table (5.4). As seen in Configuration 4, 98 percent of selections make correct selection decisions choosing Systems 1 and 2 as an equal system pair. In Configuration 5, 98

percent of decisions conclude that Systems 1 and 2 are a nondominant system pair. It is reasonable that Configuration 4 needs more observations than Configuration 5 because detecting equal case for either mean measures or variance measures is the most difficult decision case.

5.2 Experiments for CMVS Procedure

In Table (5.5), Configuration 6 is designed to test the effectiveness of the feasibility check phase of the CMVS procedure. In Configuration 6, System 2 has a larger mean value, but it is an infeasible system because of the larger variance compared with the standard system 0 which has a standard deviation of R. Hence, System 1 is the best system. Configurations 7-8 are designed to test the effectiveness of the mean comparison of the CMVS procedure. In Configuration 7, both systems are feasible. System 1 is the best system because of the larger mean value. In Configuration 8, the two systems are nondominant, since they are less variance than the standard system and their mean values are equal.

Table 5.5: Configurations 6-8 of CMVS procedure, ($\delta = 0.05, \sigma_0 = R = 1.1$).

		Sys. 1	Sys. 2	Correct Selection
Configuration 6	μ	1	$1+\delta$	$S_I = \{1\}$
	σ	R	R^2	
Configuration 7	μ	1	$1-\delta$	$S_I = \{1\}$
	σ	R	R	
Configuration 8	μ	1	1	$S_I = \{1,2\}, S_N = \{(1,2)\}$
	σ	1	R	

The results of the CMVS procedure for Configurations 6-8 are listed in Table (5.6). In all configurations, the CMVS procedure made the correct selection decision with more than 0.95 probability. For Configuration 8, the $SATO$ is larger than the

Table 5.6: Results for Configurations 6-8 of CMVS procedure

	PCS	$SATO$
Configuration 6	0.982	3,957
Configuration 7	0.969	18,644
Configuration 8	0.960	22,645

other two configurations since both systems are feasible and they have the same mean value.

5.3 Experiments for RAMS Procedure

Configurations 9 and 10 in Table (5.7) are used to test the effectiveness of the RAMS procedure. In both configurations, the differences of L's are equal to δ_x which is the smallest difference worth detecting. In Configuration 9, the best system is System 2, because $L_1 = 100\delta_x - 50\delta_x < L_2 = 61\delta_x - 10\delta_x$. In Configuration 10, the best system is System 1.

Table 5.7: Configurations 9-10 of RAMS procedure ($\delta_x = 0.5, G = 1$)

		Sys. 1	Sys. 2	L	corss over point	Correct Selection
Configuration 9	μ σ	$100\delta_x$ $50\delta_x$	$61\delta_x$ $10\delta_x$	$-\delta_x$	0.1648	$S_I = \{2\}$
Configuration 10	μ σ	$100\delta_x$ $50\delta_x$	$60\delta_x$ $11\delta_x$	δ_x	0.1525	$S_I = \{1\}$

In Table (5.8), it is seen that the PCSs of both configurations are above the nominal confidence level 0.95, but the $SATO$s have great difference. This is due to the fact that N changes based on the relationship between the cross-over point and

Table 5.8: Results for Configurations 9-10 of RAMS procedure

	PCS	$SATO$
Configuration 9	0.999	192,504
Configuration 10	0.990	308,513

the initial quantile window of each configuration. The relationships between the cross-over point and the quantile window when $N = 100$ for each configuration are shown in Figures (5.1) and (5.2). For both configurations, the q_T is $0.1587\ldots$ according to $G = 1$, and the the range of boundaries for $N = 100$ is (0.15-0.16). In Figure (5.1), it is seen that the cross-over point of Configuration 9 is not only out of the boundaries, but also far away from the q_T line. That means the quantile window at current N value is narrow enough for the procedure to make a consistent decision for both quantiles. However, for Configuration 10, as seen in Figure (5.2), the cross-over point is 0.1525 which is very close to the q_T point we focus on and in the range of two boundaries. A consistent decision can not be obtained at the current N value of 100. Hence, N is increased and the boundaries are narrowed until a consistent decision can be made. For this reason the $SATO$ of Configuration 10 is larger than Configuration 9.

Generally, a system pair whose cross-over point is in the range of boundaries needs more observations to make a correct selection decision. When one of the boundaries and the cross-over point are within the indifference zone, the procedure may need a very large N value to narrow down the boundaries, such as 10,000. The final decision is made when both boundaries and the cross-over point are within the indifference zone.

The experiments in Table (5.9) are designed to further verify the relationship between the cross-over point and the q_T value. In Table (5.9), a system pair is

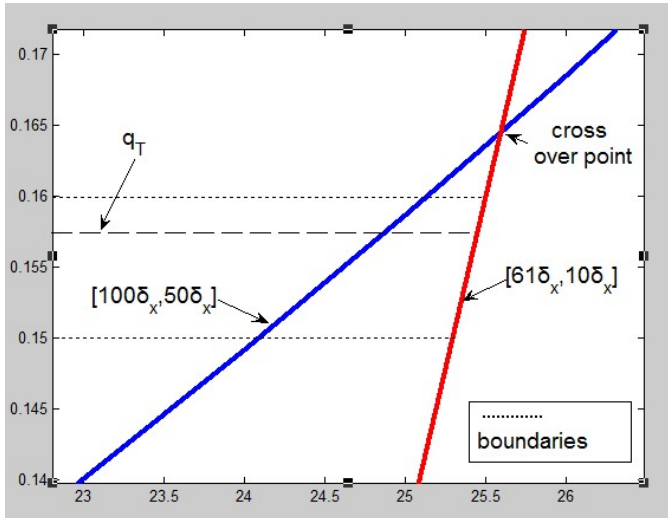


Figure 5.1: The relationship between the cross-over point and the boundaries for Configurations 9 .

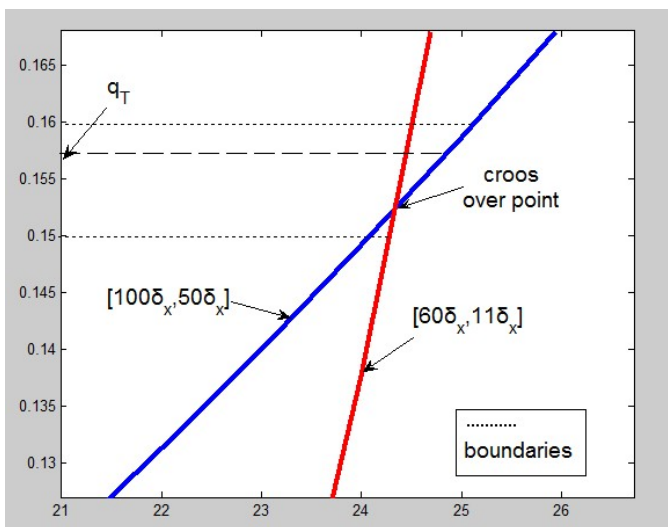


Figure 5.2: The relationship between the cross-over point and the boundaries for Configurations 10 .

performed with different G, or q_T, values. System 1 has a mean of $100\delta_x$ and standard deviation of $60\delta_x$. System 2 has a mean of $50\delta_x$ and a standard deviation of $10\delta_x$. The system pair has the cross-over point which is the same as the q_T value of $G = 1$. The comparison for $G = 1$ is expected to have the lower PCS and largest $SATO$

value because when the cross-over point is close to q_T, the $SATO$ increases rapidly. The relationships between the cross-over point and different G's are shown in Figure (5.3). The results of these experiments are shown in Table (5.10). As it's seen in the results, the largest $SATO$ value is observed when $G = 1$. As G becomes apart from $G = 1$, the $SATO$ value decrease.

Table 5.9: Configurations of the RAMS procedure with different G values for the system pair $(\mu_1 = 100\delta_x, \sigma_1 = 60\delta_x)$ and $(\mu_2 = 50\delta_x, \sigma_2 = 10\delta_x)$, $(\delta_x = 0.5)$

	q_T	cross over point	Correct Selection
$G = 0.5$	$0.3085\ldots$	0.1587	$S_I = \{1\}$
$G = 1$	$0.1587\ldots$	0.1587	$S_I = \{1, 2\}, S_E = \{(1, 2)\}$
$G = 1.5$	$0.0668\ldots$	0.1587	$S_I = \{2\}$
$G = 2$	$0.0228\ldots$	0.1587	$S_I = \{2\}$

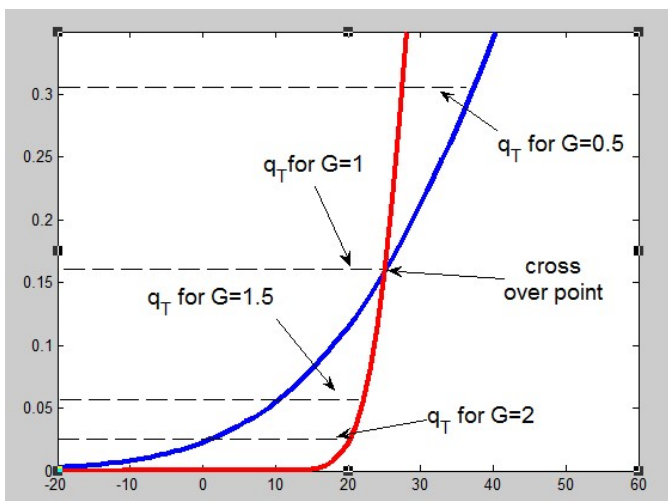


Figure 5.3: The relationship between the cross-over point and different G's.

The experiments in Table (5.11) are used to observe the change of $SATO$ as the q_T goes to tail, when q_T is equal to the cross-over point of a system pair. In Table (5.11), four different pairs of systems with different G values are presented. All of the

Table 5.10: Results of the RAMS procedure with different G values for the system pair $(\mu_1 = 100\delta_x, \sigma_1 = 60\delta_x)$ and $(\mu_2 = 50\delta_x, \sigma_2 = 10\delta_x)$, $(\delta_x = 0.5)$

	PCS	SATO
$G = 0.5$	1.000	3,888
$G = 1$	0.993	491,876
$G = 1.5$	1.000	7,181
$G = 2$	1.000	6,942

Table 5.11: Configurations of the RAMS procedure when the cross over point is at q_T quantile, $(\delta_x = 0.5)$

	q_T	Systems	L	cross over point	Correct Selection
$G = 0.5$	0.3085	$[100\delta_x, 110\delta_x]$ $[50\delta_x, 10\delta_x]$	0	0.3085	$S_I = \{1, 2\}, S_E = \{(1, 2)\}$
$G = 1$	0.1587	$[100\delta_x, 60\delta_x]$ $[50\delta_x, 10\delta_x]$	0	0.1587	$S_I = \{1, 2\}, S_E = \{(1, 2)\}$
$G = 2$	0.0228	$[100\delta_x, 35\delta_x]$ $[50\delta_x, 10\delta_x]$	0	0.0228	$S_I = \{1, 2\}, S_E = \{(1, 2)\}$

Table 5.12: Results for the RAMS procedure when the cross over point is at q_T quantile, $(\delta_x = 0.5)$

	PCS	SATO
$G = 0.5$	0.994	253,348
$G = 1$	0.996	364,930
$G = 2$	0.996	568,254

systems have the cross-over points at the q_T quantile point. For example, when G is equal to 0.5, the system pair ($[100\delta, 110\delta]$ and $[50\delta, 10\delta]$) cross at the 0.3085 quantile which is equal to $\Phi(G = 0.5)$. As we see in the Table (5.12), the $SATO$ increases as the G value increases (or q_T decreases) because it is statistically more difficult to estimate a tail quantile.

5.4 Experiments for More than two Systems

Configuration 15 is designed to test the effectiveness of the three procedures when K is larger than 2, and to verify the relationship among the three procedures. In Table (5.13), $K = 5$ systems are tested by the three procedures when the mean indifference-zone parameter $\delta = 0.05$, the quantile indifference-zone parameter $\delta_x = 0.05$, and the standard system is $\mu = 1, \sigma = R$.

Table 5.13: Configuration 15 for three procedures, ($\delta = 0.05, \delta_x = 0.05$, and $\sigma_0 = R = 1.1$).

		Sys. 1	Sys. 2	Sys. 3	Sys. 4	Sys. 5
Configuration 15	μ	1	1	$1+\delta$	$1+\delta$	1
	σ	1	1	R	R^2	R

As seen in Table (5.14), the PCS results of Configuration 15 are above the confidence level 0.95. The RAMS efficient set of systems is a subset of the MVS efficient set of systems. The CMVS procedure is capable of selecting a single best system (System 3 in this case) when a standard system for comparison is given.

Table 5.14: Results of Configuration 15 for three procedures, ($\delta = 0.05, \delta_x = 0.05$, and $\sigma_0 = R = 1.1$).

Procedures	Correct Selection	PCS	$SATO$
MVS	$S_I = \{1,2,3\}, S_N = \{(1,3)\}, \{(2,3)\}$ and $S_E = \{(1,2)\}$	0.996	83,157
CMVS	$S_I = \{3\}$	0.989	72,865
RAMS	$S_I = \{1,2\}$ and $S_E = \{(1,2)\}$	0.998	179,082

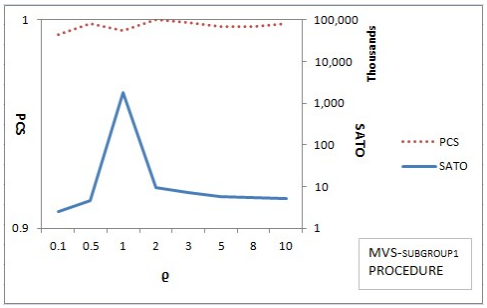
5.5 Experiments for Robustness

In this section, experiments are performed for detecting the robustness of the proposed procedures. In the experiments, two different configuration groups are used. In group 1, the system distributions are IID normal. In group 2, the system distributions are non-normal, and specifically exponential distribution is utilized. In each group, the configurations change according to different ρ values. ρ is $(\sigma_1/\mu_1)/(\sigma_2/\mu_2)$ for normal distribution sampling and μ_1/μ_2 for exponential distribution sampling. In this case, the experiments only consider two systems comparison. In all figures in this section, the x axis denotes the ρ value, the left y axis denotes PCS, and the right y axis denotes the SATO.

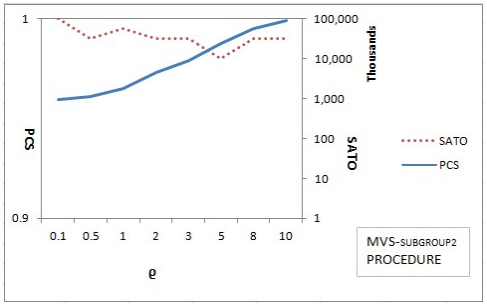
In group 1's systems comparison, two different sub-groups are designed. For both sub-groups, the second system is set such that $\sigma_2/\mu_2 = 1$. In Subgroup 1 experiments, $\sigma_1 = \sigma_2$ and μ_1 is gradually decreased so that ρ increases. In Subgroup 2 experiments, $\mu_1 = \mu_2$ and σ_1 is gradually increased so that ρ increases.

In figure (5.4), the results from group 1 are shown. The figures (5.4a) (5.4c) (5.4e) on the left side are from sub-group1, and the figures (5.4b) (5.4d) (5.4f) on the right side are from sub-group 2. It is seen that when $\rho = 1$, all experiments in sub-group1 need a larger SATO but have a lower PCS. In other words, generally, the most difficult situation for our procedures is the equality case comparison.

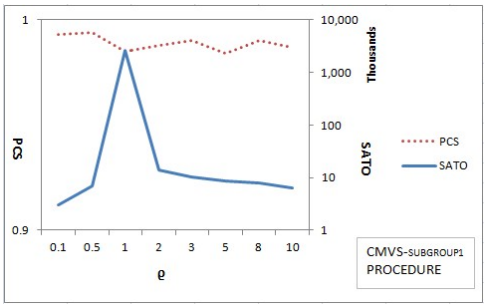
Furthermore, the ρ can be changed by changing any variables in $(\sigma_1/\mu_1)/(\sigma_2/\mu_2)$. In this experiment, when we see the configurations in sub-group 1 and sub-group 2, it can be seen that sub-group 2's first system of each configuration is equal to sub-group 1's first system of each configuration multiply a constant C such that $(\sigma_1 C/\mu_1 C)/(\sigma_2/\mu_2) = (\sigma_1/\mu_1)/(\sigma_2/\mu_2)$. When we compare the PCS and SATO lines from the left side to right side, the PCS can always keep above the guaranteed con-



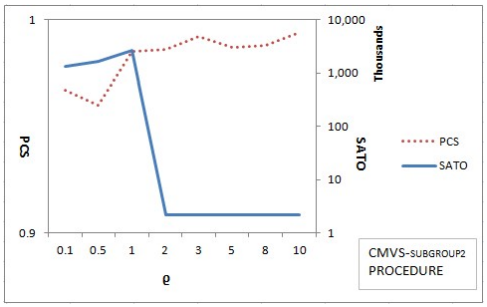
(a) Results from sub-group 1 for procedure MVS.



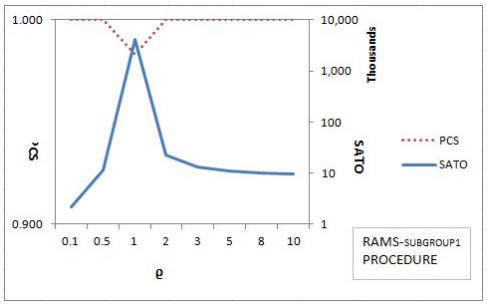
(b) Results from sub-group 2 for procedure MVS.



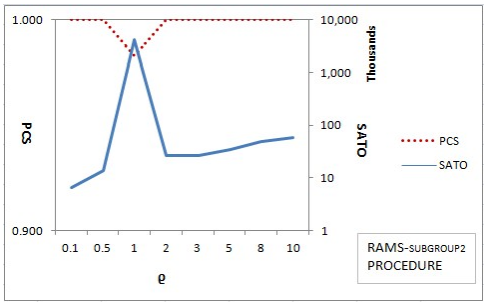
(c) Results from sub-group 1 for procedure CMVS.



(d) Results from sub-group 2 for procedure CMVS.



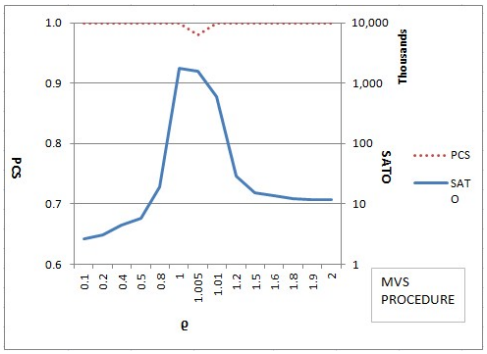
(e) Results from sub-group 1 for procedure RAMS.



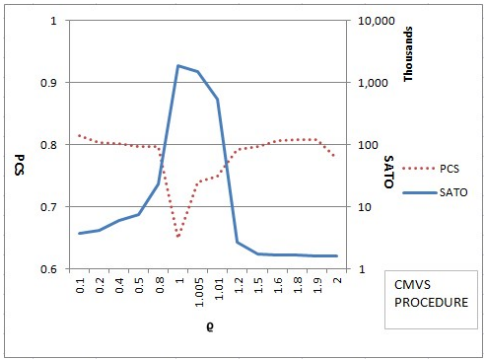
(f) Results from sub-group 2 for procedure RAMS.

Figure 5.4: Results from group 1

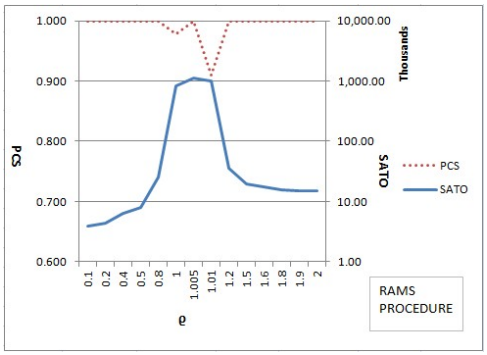
fidence level but the SATO are different even for the same ρ value. For procedure MVS (from figures (5.4a) to (5.4b)), when the first system is multiplied by a positive integer, if $\rho < 1$, less SATO is needed and if $\rho > 1$, more SATO is needed. The larger the positive integer multiplied, the larger SATO needed. For procedure CMVS, the



(a) Results from group 2 for procedure MVS



(b) Results from group 2 for procedure CMVS



(c) Results from group 2 for procedure RAMS

Figure 5.5: Results from group 2

changes depend on the value of standard system's variance. For procedure RAMS, the results are similar to procedure MVS, but the change of SATO is smaller than procedure MVS. All of configurations in group 1 are listed in Appendix.

In group 2's systems comparison, we keep mean value of the second system for each configuration unchanged, and when the mean value of the first system increases, the ρ increases. In Figure (5.5), the results are from group 2. It is seen that the pattern of the SATO is similar to the sub-group 1's, and the peak point is around $\rho = 1$. However, not all of the PCS can reach the guaranteed confidence level 0.95, e.g. all configurations in procedure CMVS (figure (5.5b)), and the configuration in procedure

RAMS which has the attribution of $\delta_x = L_1 - L_2$. In order to understand the inefficiency of the CMVS procedure, separating comparisons between the means and variances are conducted. The experiments indicate that for exponential distribution system comparisons, the variance comparisons cannot always satisfy the guaranteed PCS. In addition, for RAMS procedure, when the difference of L between two systems is equal to δ_x value, the PCS are difficult to satisfy the pre-specified confidence level for exponential distributed system comparison. In figure (5.6), the x axis denotes the δ_x values. It is seen that when the δ_x is equal to 0.01 and the difference of L's between two systems is equal to $\delta = 0.01$, the PCS is only 0.85.

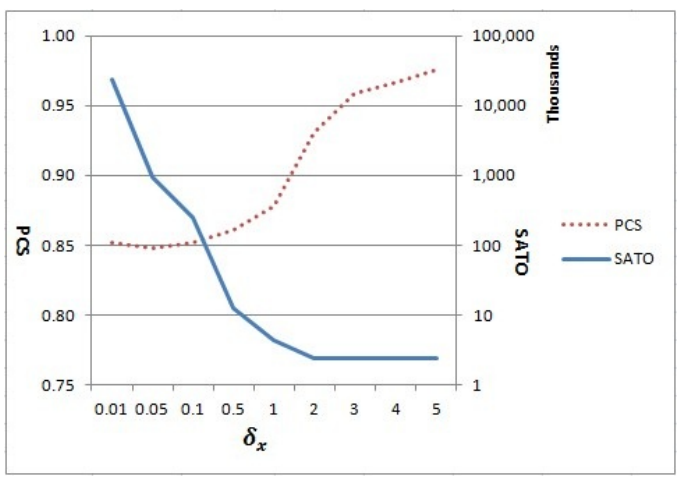


Figure 5.6: RAMS procedure when systems are exponential and the differences between L's are all equal to δ_x.

Chapter 6

Conclusion

In the traditional mean selection criterion, each alternative system is measured and compared through only the mean value which is a measure of the average behavior. However, variance, a measure of risk, should be considered based on the preference of the decision makers. Thus, in this thesis three different selection criteria are presented based on both the mean and variance measures. For each of the three criteria, a selection procedure is illustrated with statistical correct selection validity.

In the MVS procedure, the mean and variance of each system are compared with other systems simultaneously. The system with a larger mean and smaller variance is defined as the best system. Two systems, if one of them has a better mean value but a worse variance, are nondominant. Two systems are equal, if they have the same mean and variance values. The application of the MVS procedure may result in a nondominated set with more than one system. In the CMVS procedure, a variance constraint is set to control the risk for the decision makers. A system whose variance is smaller than the standard system is a feasible system. We compare the mean values of the feasible systems and select the system with the largest mean value as the best system. In this procedure, the efficient set generally includes only one system. In the

RAMS procedure, according to the Bamoul's $L = \mu - G\sigma$ criterion, the simulated systems are compared. A system with the largest L value is the best system. Two systems which have the same L values are considered equal. The computational efficiency of each procedure is shown by experiments in the end of this thesis.

However, there are some some potential improvements in these procedures. First, large number of observations may be needed in the RAMS procedure because of the difficulty in the accurate estimation of the quantiles. Second, in the case of finite number of decimal digits of q_T, the quantile window method seems too complex on solving these problems. Third, in our procedures, the simulated systems are assumed to follow a normal distribution, which may be not be satisfied in general. These are the problems that we will tackle in our future studies.

Appendix A

Results for Experiments of Section 5.5

The configurations results are listed as follow: the configurations of Group 1: sub-group 1 are in Table (A.1), the configurations of Group 1: sub-group 2 are in Table (A.2), the configurations of Group 2 are in Table (A.3), and the configurations, where the differences of L's between two systems are equal to δ_x, are in Table (A.4).

Table A.1: Configurations of sub-group 1, ($\delta = 0.05, \delta_x = 0.05, R = 1.1, \sigma_0 = 10, G = 1$).

ρ	Sys. 1	Sys.2	MVS		CMVS		RAMS	
			PCS	SATO	PCS	SATO	PCS	SATO
0.1	N[100,10]	N[10,10]	0.993	2557	0.993	3099	1.00	2202
0.5	N[20,10]	N[10,10]	0.998	4718	0.994	7029	1 .000	11583
1	N[10,10]	N[10,10]	0.995	1781579	0.985	2651258	0.983	4220809
2	N[5,10]	N[10,10]	1.000	9398	0.998	13901	1.000	22310
3	N[10/3,10]	N[10,10]	0.999	7040	0.990	10475	1.000	12916
5	N[2,10]	N[10,10]	0.997	5868	0.984	8706	1.000	10910
8	N[10/8,10]	N[10,10]	0.997	5371	0.990	8026	1.000	10074
10	N[1,10]	N[10,10]	0.998	5232	0.987	6429	1.000	9777

Table A.2: Configurations of sub-group 2, ($\delta = 0.05, \delta_x = 0.05, R = 1.1, \sigma_0 = 10, G = 1$).

ρ	Sys. 1	Sys.2	MVS		CMVS		RAMS	
			PCS	SATO	PCS	SATO	PCS	SATO
0.1	N[10,1]	N[10,10]	1.000	941846	0.967	13504400	1.000	6660
0.5	N[10,5]	N[10,10]	0.990	1103786	0.960	1635913	1 .000	13994
1	N[10,10]	N[10,10]	0.995	1781579	0.985	2651258	0.983	4220809
2	N[10,20]	N[10,10]	0.990	4417608	0.986	2216	1.000	27070
3	N[10,30]	N[10,10]	0.980	9061709	0.992	2194	1.000	26987
5	N[10,50]	N[10,10]	0.990	23871494	0.987	2183	1.000	34867
8	N[10,80]	N[10,10]	0.990	57626673	0.988	2168	1.000	49381
10	N[10,100]	N[10,10]	0.990	91029874	0.994	2140	1.000	59484

Table A.3: Configurations of Group 2, ($\delta = 0.05, \delta_x = 0.05, R = 1.1, \sigma_0 = 10, G = 0.5$).

ρ	Sys. 1	Sys.2	MVS		CMVS		RAMS	
			PCS	SATO	PCS	SATO	PCS	SATO
0.1	EXP[1]	EXP[10]	1.000	2623	0.814	3685	1.000	3932
0.2	EXP[2]	EXP[10]	1.000	3052	0.803	4134	1 .000	4426
0.4	EXP[4]	EXP[10]	1.000	4521	0.802	5896	1.000	6255
0.5	EXP[5]	EXP[10]	1.000	5868	0.797	7452	1.000	7941
0.8	EXP[8]	EXP[10]	1.000	19271	0.797	23626	1.000	25453
1	EXP[10]	EXP[10]	1.000	1734007	0.650	1890000	0.979	823120
1.005	EXP[10.05]	EXP[10]	0.980	1601185	0.740	1515611	1.000	1132496
1.01	EXP[10.1]	EXP[10]	1.000	587863	0.750	539981	0.910	1020530
1.2	EXP[12]	EXP[10]	1.000	28784	0.793	2679	1.000	36047
1.4	EXP[14]	EXP[10]	1.000	15228	0.797	1715	1.000	19238
1.6	EXP[16]	EXP[10]	1.000	13922	0.807	1685	1.000	17618
1.8	EXP[18]	EXP[10]	1.000	12426	0.808	1659	1.000	15764
1.9	EXP[19]	EXP[10]	1.000	12023	0.808	1608	1.000	15237
2	EXP[20]	EXP[10]	1.000	11702	0.778	1636	1.000	14897

Table A.4: Configurations of the differences of Ls between two systems are equal to δ_x, $(G = 0.5)$.

δ_x	Sys. 1	Sys.2	PCS	SATO
0.01	EXP[10.02]	EXP[10]	0.852	23783320
0.05	EXP[10.1]	EXP[10]	0.847	958927
0.1	EXP[10.2]	EXP[10]	0.852	245081
0.5	EXP[11]	EXP[10]	0.869	12279
1	EXP[12]	EXP[10]	0.877	4348
2	EXP[14]	EXP[10]	0.931	2436
3	EXP[16]	EXP[10]	0.959	2404
4	EXP[18]	EXP[10]	0.967	2440
5	EXP[20]	EXP[10]	0.976	2438

Appendix B

Matlab Code for Procedure MVS

```
%script file: k systems based on mean and variance test.
%user-defined parameters:
K = 2; % the number of systems
delta = 0.5; % mean indifference zone
R = 1.1; % variance indifference zone
mean = [10/8, 10];
sigma = [10, 10];
corr_S_I = [0,1];
corr_S_E = [0,0;0,0]; % upper triangular matrix
corr_S_N = [0,0;0,0];% upper triangular matrix
alpha = 0.5;
beta = 0.05;
alpha_k = alpha/(k*(k-1));
beta_k = beta/(K*(K-1));
replication = 10000;
%procedure parameters:
```

```
d = 3*delta/8;
R_square = R^2;
lambda = 1+0.7*(R-1);
n_0 = floor(1-log(alpha_k)/log(lambda))+1;  %variables:
correct_selection_count = 0;
total_sample_size = 0;
%main code:
for l=1:replication
    %initialization:
    X= zeros(K,n_0);
    for i = 1: K
    X(i,:) = normrnd(mean(i),sigma(i),1, n_0); % original sample, (K,n_0)
    end
    average_X = sum(X,2)/n_0; % average of the observations from each system, (K,1)
    sum_X_sqre = sum(X^2,2); % sum of the square of the observations from each
system, (K,1)
    total_sample_size = total_sample_size + K*n_0; % add n_0 observation for each
system
    var_X = (sum_X_sqre-n_0*average_X^2)/(n_0-1); % (K,1)
    %decision variables:
    var_H0 = zeros(K); % H_0: sigma1 = sigma2, upper triangular matrix
    var_H1 = zeros(K); % H_1: sigma1 > sigma2, upper triangular matrix
    var_H2 = zeros(K); % H_2: sigma1 < sigma2, upper triangular matrix
    mean_H0 = zeros(K); % H_0: mean1 = mean2, upper triangular matrix
    mean_H1 = zeros(K); % H_1: mean1 > mean2, upper triangular matrix
    mean_H2 = zeros(K); % H_2: mean1 < mean2, upper triangular matrix
```

```
S_I = ones(1,K); % set of systems under consideration
S_D = zeros(1,K); % set of decided systems
S_N = zeros(K); % set of nondominant system pairs, upper triangular matrix
S_E = zeros(K);
r = n_0; % stage number
sum_diff_X = zeros(K);
for i=1:K
for j=1:K
if i < j
%mean comparison:
sum_diff_X(i,j) = sum(X(i,:)-X(j,:));
IB_mean_ij = r*(delta - d) - (var_X(i) + var_X(j))*((1/beta_k)^(2/(r-1)) -1)*(r-1) /
(4*d);
OB_mean_ij = r*d + (var_X(i) + var_X(j))*((2/alpha_k)^(2/(r-1)) -1)*(r-1) / (4*d);
if sum_diff_X(i,j) > -IB_mean_ij && sum_diff_X(i,j) < IB_mean_ij %H0: mean_i =
mean_j
mean_H0(i,j)= 1;
elseif sum_diff_X(i,j) > OB_mean_ij %H1: mean_i > mean_j
mean_H1(i,j)= 1;
elseif sum_diff_X(i,j) < -OB_mean_ij %H2: mean_i < mean_j
mean_H2(i,j)= 1;
end
%variance comparison
IB_sigma_ij = lambda * (var_X(i) / var_X(j)) * (lambda - beta_k^(1/(r-1)))
/ (lambda*beta_k^(1/(r-1))-1);
IB_sigma_ji = lambda * (var_X(j) / var_X(i)) * (lambda - beta_k^(1/(r-1)))
```

/ (lambda*beta_k$^{(1/(r-1))}$-1);

OB_sigma_ij = (var_X(i) / var_X(j)) * (lambda - (2/alpha_k)$^{(1/(r-1))}$) / (lambda* (lambda*(2/alpha_k)$^{(1/(r-1))}$-1));

OB_sigma_ji = (var_X(j) / var_X(i)) * (lambda - (2/alpha_k)$^{(1/(r-1))}$) / (lambda * (lambda*(2/alpha_k)$^{(1/(r-1))}$-1));

if IB_sigma_ij < R_square && IB_sigma_ji < R_square % H0:

var_i = var_j

var_H0(i,j)= 1;

elseif OB_sigma_ij > 1 && r > floor(1-log(alpha_k/2)/log(lambda)) % H1:

var_i > var_j

var_H1(i,j)= 1;

elseif OB_sigma_ji > 1 && r > floor(1-log(alpha_k/2)/log(lambda)) % H2:

var_i < var_j

var_H2(i,j)= 1;

end

if (var_H1(i,j)==1 && mean_H0(i,j)==1) ‖ (var_H0(i,j)==1

&& mean_H2(i,j)==1) ‖ (var_H1(i,j)==1 && mean_H2(i,j)==1)

S_I(i) = 0;

S_N(i,:) = 0;

S_N(:,i) = 0;

S_E(i,:) = 0;

S_E(:,i) = 0;

elseif (var_H2(i,j)==1 && mean_H0(i,j)==1) ‖ (var_H0(i,j)==1 &&

mean_H1(i,j)==1) ‖ (var_H2(i,j)==1 && mean_H1(i,j)==1)

S_I(j) = 0;

S_N(j,:) = 0;

```
S_N(:,j) = 0;
S_E(j,:) = 0;
S_E(:,j) = 0;
elseif (var_H1(i,j)==1 && mean_H1(i,j)==1)
|| (var_H2(i,j)==1 && mean_H2(i,j)==1)
S_N(i,j) = 1;
elseif var_H0(i,j)==1 && mean_H0(i,j)==1
S_E(i,j) = 1;
end
end
end
end
if sum(S_I)==1 || ((sum(S_I) > 1) && ( sum(S_I)
*(sum(S_I)-1)/2 == sum(sum(S_N)) + sum(sum(S_E)) ))
sampling_needed = 0;
else
sampling_needed = 1;
end
while sampling_needed
r=r+1;
for i=1:K % update S_D
for j=1:K
if i =j && S_I(i)==1 && S_I(j)==1
ii = min(i,j); % S_N and S_E are upper triangular matrices
jj = max(i,j);
if S_N(ii,jj)==1 or S_E(ii,jj)==1
```

```
S_D(i) = 1;
else
S_D(i) = 0;
break;
end
end
end
end
X_new = zeros(K,1);
for i=1:K
if S_I(i)==1 && S_D(i)==0
X_new(i) = normrnd(mean(i),sigma(i),1);
total_sample_size = total_sample_size+1;
end
end
average_X = (average_X*(r-1)+X_new)/r; % (K,1)
sum_X_sqre = sum_X_sqre + X_new^2; % (K,1)
var_X = (sum_X_sqre-r*average_X^2)/(r-1); % (K,1)
for i=1:K
for j=1:K
if i<j && S_I(i)==1 && S_I(j)==1 && S_N(i,j)==0 && S_E(i,j)==0
% If the final decision has not been made yet, the mean and variance
% tests can be renewed with the new observations.
mean_H0(i,j)=0; mean_H1(i,j)=0; mean_H2(i,j)=0;
var_H0(i,j)=0; var_H1(i,j)=0; var_H2(i,j)=0;
%mean comparison
```

sum_diff_X(i,j) = sum_diff_X(i,j) + X_new(i)-X_new(j);

IB_mean_ij = r*(delta - d) - (var_X(i) + var_X(j))*((1/beta_k)$^{(2/(r-1))}$ -1)*(r-1) / (4*d);

OB_mean_ij = r*d + (var_X(i) + var_X(j))*((2/alpha_k)$^{(2/(r-1))}$ -1)*(r-1) / (4*d);

if sum_diff_X(i,j) > -IB_mean_ij && sum_diff_X(i,j) < IB_mean_ij %H0: mean_i = mean_j

mean_H0(i,j)= 1;

elseif sum_diff_X(i,j) > OB_mean_ij %H1: mean_i > mean_j

mean_H1(i,j)= 1;

elseif sum_diff_X(i,j) < -OB_mean_ij %H2: mean_i < mean_j

mean_H2(i,j)= 1;

end

%variance comparison

IB_sigma_ij = lambda * (var_X(i) / var_X(j)) * (lambda - beta_k$^{(1/(r-1))}$)

/ (lambda*beta_k$^{(1/(r-1))}$-1);

IB_sigma_ji = lambda * (var_X(j) / var_X(i)) * (lambda - beta_$k^{(1/(r-1))}$)

/ (lambda*beta_k$^{(1/(r-1))}$-1);

OB_sigma_ij = (var_X(i) / var_X(j)) * (lambda - (2/alpha_k)$^{(1/(r-1))}$) / (lambda * (lambda*(2/alpha_k)$^{(1/(r-1))}$-1));

OB_sigma_ji = (var_X(j) / var_X(i)) * (lambda - (2/alpha_k)$^{(1/(r-1))}$) / (lambda * (lambda*(2/alpha_k)$^{(1/(r-1))}$-1));

if IB_sigma_ij < R_square && IB_sigma_ji < R_square % H0: var_i = var_j

var_H0(i,j)= 1;

elseif OB_sigma_ij > 1 && r > floor(1-log(alpha_k/2)/log(lambda)) % H1: var_i > var_j

var_H1(i,j)= 1;

```
elseif OB_sigma_ji > 1 && r > floor(1-log(alpha_k/2)/log(lambda)) % H2: var_i
< var_j
var_H2(i,j)= 1;
end
%elimination
if (var_H1(i,j)==1 && mean_H0(i,j)==1) || (var_H0(i,j)==1
&& mean_H2(i,j)==1) || (var_H1(i,j)==1 && mean_H2(i,j)==1)
S_I(i) = 0;
S_N(i,:) = 0;
S_N(:,i) = 0;
S_E(i,:) = 0;
S_E(:,i) = 0;
elseif (var_H2(i,j)==1 && mean_H0(i,j)==1) || (var_H0(i,j)==1
&& mean_H1(i,j)==1) || (var_H2(i,j)==1 && mean_H1(i,j)==1)
S_I(j) = 0;
S_N(j,:) = 0;
S_N(:,j) = 0;
S_E(:,j) = 0;
S_E(j,:) = 0;
elseif (var_H1(i,j)==1 && mean_H1(i,j)==1)
|| (var_H2(i,j)==1 && mean_H2(i,j)==1)
S_N(i,j) = 1;
elseif var_H0(i,j)==1 && mean_H0(i,j)==1
S_E(i,j) = 1;
end
end
```

```
end
end
if sum(S_I)==1 || ((sum(S_I) > 1) && ( sum(S_I)*(sum(S_I)-1)/2
== sum(sum(S_N)) + sum(sum(S_E)) ))
sampling_needed = 0;
else
sampling_needed = 1;
end
end
correct_selection = 0;
if sum(corr_S_I==S_I)==K
correct_selection = 1;
% If more than one system survives, we need to be able to determine the equality
and nondominance relationships correctly
if sum(S_I)>1 && sum(sum(S_E==corr_S_E)) =K^2
|| sum(sum(S_N==corr_S_N)) =K^2
correct_selection = 0;
end
end
correct_selection_count = correct_selection_count + correct_selection;
end
%output
fprint('the percentage of correct selection is: %6.4f \n', correct_selection_count/replication);
fprintf('the average sample size is: %6.4f \n', total_sample_size/replication);
```

Appendix C

Matlab Code for Procedure CMVS

```
%script file:K systems based on constrained mean and variance.
%user-defined parameters
K = 2; % the number of systems
delta = 0.05; % mean indifference zone
R = 1.1; % variance indifference zone
stand_mean = 10; % standard system
stand_sigma = 10; % standard system
mean = [10, 10];
sigma = [5, 10];
corr_S_F = [1, 1]; % upper triangular matrix
corr_S_N = [0, 1 ; 0, 0];% upper triangular matrix
alpha = 0.05;
beta = 0.05;
alpha_k1 = alpha/(2*K);
```

```
beta_k1 = beta/(2*K);
alpha_k2 = alpha/(K*(K-1));
beta_k2 = beta/(K*(K-1));
replication = 100;
% procedure parameters d = 3*delta/8;
R_square = R^2;
lambda = 1 + 0.7*(R-1);
n_0 = floor(1-log(alpha_k1)/log(lambda))+1;
% variables
correct_selection_count = 0;
total_sample_size = 0;
for l = 1:replication
    % initialization X = zeros(K,n_0);
    for i = 1: K
    X(i,:) = normrnd(mean(i),sigma(i),1, n_0); % original sample, (K,n_0)
    end
    X_average = sum(X,2)/n_0; % average of the observations from each system, (K,1)
    X_sum_square = sum(X.^2,2); % sum of the square of the observations from each
system, (K,1)
    X_var = (X_sum_square - n_0*X_average.^2)/(n_0-1); % (K,1)
    total_sample_size = total_sample_size + K*n_0; % add n_0 observation for each
system
    X_0 = normrnd(stand_mean,stand_sigma,1, n_0); % standard system
    X_0_average = sum(X_0)/n_0;
    X_0_sum_square = sum(X_0.^2);
    X_0_var = (X_0_sum_square - n_0*X_0_average^2)/(n_0-1);
```

```
total_sample_size = total_sample_size + n_0;
% decision variables
mean_H0 = zeros(K); % H_0: mean1 = mean2, upper triangular matrix
mean_H1 = zeros(K); % H_1: mean1 > mean2, upper triangular matrix
mean_H2 = zeros(K); % H_2: mean1 < mean2, upper triangular matrix
S_I = ones(1,K); % set of systems whose feasibility test has not been concluded yet
S_F = zeros(1,K); % set of feasible systems
S_N = zeros(K); % set of nondominant system pairs, upper triangular matrix
S_D = zeros(1,K); % set of decided systems
r = n_0; % stage number
% feasibility check
for i = 1:K
LB_sigma_i = lambda*(X_var(i)/X_0_var)*(lambda-beta_k1^(1/(r-1)))
/(lambda*beta_k1^(1/(r-1))-1);
UB_sigma_i = (X_var(i)/X_0_var)*(lambda-(1/alpha_k1)^(1/(r-1)))
/ (lambda*(lambda*(1/alpha_k1)^(1/(r-1))-1));
if LB_sigma_i<R_square
S_I(i) = 0;
S_F(i) = 1;
end
if UB_sigma_i>1
S_I(i) = 0;
end
end
% mean comparison
```

```
sum_diff_X = zeros(K);
for i=1:K
for j=1:K
if i<j
sum_diff_X(i,j) = sum(X(i,:)-X(j,:));
IB_mean_ij = r*(delta - d) - (X_var(i) + X_var(j))*((1/beta_k2)^(2/(r-1)) -1)*(r-1) / (4*d);
OB_mean_ij = r*d + (X_var(i) + X_var(j))*((2/alpha_k2)^(2/(r - 1)) -1)*(r-1) / (4*d);
if S_F(i)==1 && S_F(j)==1 && S_N(i,j)==0
if sum_diff_X(i,j) ¿ -IB_mean_ij && sum_diff_X(i,j) < IB_mean_ij % H0: mean_i = mean_j
mean_H0(i,j) = 1;
S_N(i,j) = 1;
elseif sum_diff_X(i,j) > OB_mean_ij % H1: mean_i > mean_j
mean_H1(i,j) = 1;
S_F(j) = 0;
S_N(j,:) = 0;
S_N(:,j) = 0;
elseif sum_diff_X(i,j) < -OB_mean_ij % H2: mean_i < mean_j
mean_H2(i,j) = 1;
S_F(i) = 0;
S_N(i,:) = 0;
S_N(:,i) = 0;
end
elseif S_F(i)==1 && S_I(j)==1
```

```
if sum_diff_X(i,j) > OB_mean_ij % H1: mean_i >mean_j
mean_H1(i,j) = 1;
S_I(j) = 0;
end
elseif S_I(i)==1 && S_F(j)==1
if sum_diff_X(i,j) < -OB_mean_ij % H2: mean_i < mean_j
mean_H2(i,j) = 1;
S_I(i) = 0;
end
end
end
end
end
if sum(S_I) > 0
add_more_observation_X0 = 1;
add_more_observation = 1;
elseif (sum(S_I)==0) && ( sum(S_F)*(sum(S_F)-1)/2 == sum(sum(S_N)) )
add_more_observation_X0 = 0;
add_more_observation = 0;
else
add_more_observation_X0 = 0;
add_more_observation = 1;
end
while add_more_observation_X0 || add_more_observation
r = r+1;
% no more sampling is needed from decided systems
```

```
if sum(S_I) == 0
for i = 1:K
for j = 1:K
if i =j && S_F(i)==1 && S_F(j)==1
ii = min(i,j);
jj = max(i,j);
if S_N(ii,jj) == 1
S_D(i) = 1;
else
S_D(i) = 0;
break;
end
end
end
end
end
add_more_observation_X0 = 1;
if add_more_observation_X0 == 1
X_0_new = normrnd(stand_mean,stand_sigma,1 ,1); %standard system
X_0_average = (X_0_average*(r-1)+X_0_new)/r;
X_0_sum_square = X_0_sum_square + X_0_new^2;
X_0_var = (X_0_sum_square - r*X_0_average^2)/(r-1);
total_sample_size = total_sample_size +1;
end
X_new = zeros(1,K);
for i = 1:K
```

```
if (S_I(i)==1 || S_F(i)==1) && S_D(i)==0
X_new(i) = normrnd(mean(i),sigma(i),1,1);
X_average(i) = (X_average(i)*(r-1) + X_new(i))/r ;
X_sum_square(i) = X_sum_square(i) + X_new(i)^2;
X_var(i) = (X_sum_square(i) - r*X_average(i)^2)/(r-1);
total_sample_size = total_sample_size +1;
end
end
% feasibility check
for i = 1:K
if S_I(i)==1
LB_sigma_i = lambda*(X_var(i)/X_0_var)*(lambda-beta_k1^(1/(r-1)))
/(lambda*beta_k1^(1/(r-1))-1);
UB_sigma_i = (X_var(i)/X_0_var)*(lambda-(1/alpha_k1)^(1/(r-1)))
/ (lambda*(lambda*(1/alpha_k1)^(1/(r-1))-1));
if LB_sigma_i<R_square
S_I(i) = 0;
S_F(i) = 1;
end
if UB_sigma_i>1
S_I(i) = 0;
end
end
end
% mean comparison
for i=1:K
```

```
for j=1:K
if i<j
sum_diff_X(i,j) = sum_diff_X(i,j)+ (X_new(i)-X_new(j));
IB_mean_ij = r*(delta - d) - (X_var(i) + X_var(j))*((1/beta_k2)^(2/(r-1)) -1)*(r-1)
/ (4*d);
OB_mean_ij = r*d + (X_var(i) + X_var(j))*((2/alpha_k2)^(2/(r-1)) -1)*(r-1) / (4*d);
if S_F(i)==1 && S_F(j)==1 && S_N(i,j)==0
if sum_diff_X(i,j) > -IB_mean_ij && sum_diff_X(i,j) < IB_mean_ij % H0: mean_i
= mean_j
mean_H0(i,j) = 1;
S_N(i,j) = 1;
elseif sum_diff_X(i,j) > OB_mean_ij % H1: mean_i > mean_j
mean_H1(i,j) = 1;
S_F(j) = 0;
S_N(j,:) = 0;
S_N(:,j) = 0;
elseif sum_diff_X(i,j) < -OB_mean_ij % H2: mean_i < mean_j
mean_H2(i,j) = 1;
S_F(i) = 0;
S_N(i,:) = 0;
S_N(:,i) = 0;
end
elseif S_F(i)==1 && S_I(j)==1
if sum_diff_X(i,j) > OB_mean_ij % H1: mean_i > mean_j
mean_H1(i,j) = 1;
S_I(j) = 0;
```

```
end
elseif S_I(i)==1 && S_F(j)==1
if sum_diff_X(i,j) < -OB_mean_ij % H2: mean_i < mean_j
mean_H2(i,j) = 1;
S_I(i) = 0;
end
end
end
end
end
if sum(S_I) > 0
add_more_observation_X0 = 1;
add_more_observation = 1;
elseif (sum(S_I)==0) && ( sum(S_F)*(sum(S_F)-1)/2 == sum(sum(S_N)) )
add_more_observation_X0 = 0;
add_more_observation = 0;
else
add_more_observation_X0 = 0;
add_more_observation = 1;
end
end
correct_selection = 0;
if sum(corr_S_F==S_F)==K
correct_selection = 1;
% If more than one system survives, we need to be able to determine the non-dominance relationships correctly
```

```
    if sum(S_F)>1 && sum(sum(S_N==corr_S_N)) =K^2
    correct_selection = 0;
    end
    end
    correct_selection_count = correct_selection_count + correct_selection;
    end
    %output
fprintf('the percentage of correct selection is: %6.4f \n', correct_selection_count/replication);
fprintf('the average sample size is: %6.4f \n', total_sample_size/replication);
```

Appendix D

Matlab Code for Procedure RAMS

```
%script file:K systems based on mean minus G sigma.
% user-defined parameters
K = 2;
delta_x = 0.05;
G = 0.5;
R=1.1;
mean = [1,10];
sigma = [10,10];
corr_S_I = [0 1];
corr_S_E = [0, 0; 0, 0]; % upper triangular matrix
alpha = 0.05;
beta = 0.05;
alpha_k = alpha/(K*(K-1)); % Quantile test is performed at points q_l and q_u
beta_k = alpha/(K*(K-1));
```

```
replication = 10000;
% procedure parameters
d = 3*delta_x/8;
n_0 = 10;
q_t = normcdf(-G,0,1); % target quantile
% variables
correct_selection_count = 0;
total_sample_size = 0;
for l= 1:replication
    N = 10; % This is multiplied by 10 in line 46.
    % variables
    S_I = ones(1,K);
    S_E = zeros(K); % upper triangular matrix
    S_D = zeros(1,K);
    change_N = 1;
    while change_N
    % initialization
    N = N*10;
    fprintf('%d \n', N);
    q_l = floor(q_t*N)/N;
    q_u = ceil(q_t*N)/N;
    Q = [q_l q_u];
    x_q = zeros(K,length(Q),n_0);
    for i=1:K
    if S_I(i)==1 && S_D(i)==0
    x = normrnd(mean(i),sigma(i),[N,n_0]); % original sample, (N,n_0)
```

```
x_sorted = sort(x,1); % [N,n_0]
weight = floor(N*Q+0.5); % (length(Q),1)
for q = 1:length(Q)
x_q(i,q,:) = x_sorted(weight(q),:)*(0.5+weight(q)-N*Q(q)) +
x_sorted(weight(q)+1,:)*(0.5-weight(q)+N*Q(q));
end
x_q_avg = sum(x_q,3)/n_0; % K by length(Q)
x_q_sum_sqr = sum(x_q.^2,3); % K by length(Q)
x_q_var = var(x_q,0,3); % K by length(Q)
end
end
total_sample_size = total_sample_size + K*N*n_0;
sum_diff_x_q = zeros(K,K,length(Q));
% decision variables
L_H0 = zeros(K,K,length(Q)); % upper triangular matrix
L_H1 = zeros(K,K,length(Q)); % upper triangular matrix
L_H2 = zeros(K,K,length(Q)); % upper triangular matrix
r = n_0;
for i = 1:K
for j = 1:K
if i<j && S_I(i)==1 && S_I(j)==1 && S_E(i,j)==0
% L_q comparison
for q = 1:length(Q)
sum_diff_x_q(i,j,q) = r*(x_q_avg(i,q)-x_q_avg(j,q));
IB_q_ij = r*(delta_x - d) - (x_q_var(i,q)+x_q_var(j,q))*((1/beta_k)^(2/(r-1))-1)*(r-
1)/(4*d);
```

```
OB_q_ij = r*d + (x_q_var(i,q)+x_q_var(j,q))*((2/alpha_k)^(2/(r-1))-1)*(r-1)/(4*d);
if sum_diff_x_q(i,j,q) > -IB_q_ij && sum_diff_x_q(i,j,q) < IB_q_ij %H0: x_q_i =
x_q_j
L_H0(i,j,q) = 1;
elseif sum_diff_x_q(i,j,q) > OB_q_ij %H1: x_q_i > x_q_j
L_H1(i,j,q) = 1;
elseif sum_diff_x_q(i,j,q) < -OB_q_ij %H2: x_q_i <x_q_j
L_H2(i,j,q) = 1;
end
end
% decision
if sum(L_H0(i,j,:)) == length(Q)
S_E(i,j) = 1;
elseif sum(L_H1(i,j,:)) == length(Q)
S_I(j) = 0;
S_E(:,j) = 0; % These four lines are needed for the Termination if loop
S_E(j,:) = 0;
L_H0(:,j,:)=0; L_H1(:,j,:)=0; L_H2(:,j,:)=0;
L_H0(j,:,:)=0; L_H1(j,:,:)=0; L_H2(j,:,:)=0;
elseif sum(L_H2(i,j,:)) == length(Q)
S_I(i) = 0;
S_E(i,:) = 0;
S_E(:,i) = 0;
L_H0(:,i,:)=0; L_H1(:,i,:)=0; L_H2(:,i,:)=0;
L_H0(i,:,:)=0; L_H1(i,:,:)=0; L_H2(i,:,:)=0;
end
```

```
end
end
end
count_S_I = sum(S_I);
count_L_H0 = sum(sum(sum(L_H0)));
count_L_H1 = sum(sum(sum(L_H1)));
count_L_H2 = sum(sum(sum(L_H2)));
count_S_E = sum(sum(S_E));
if count_S_I==1 || count_S_E==count_S_I*(count_S_I-1)/2 % termination
add_more_observation = 0;
change_N = 0;
elseif (count_L_H0+count_L_H1+count_L_H2)/2+count_S_E == count_S_I
*(count_S_I-1)/2
add_more_observation = 0;
change_N = 1;
else
add_more_observation = 1;
change_N = 0;
end
while add_more_observation
% decision variables
L_H0 = zeros(K,K,length(Q)); % upper triangular matrix
L_H1 = zeros(K,K,length(Q)); % upper triangular matrix
L_H2 = zeros(K,K,length(Q)); % upper triangular matrix
% update S_D
for i = 1:K
```

```
for j = 1:K
if i =j && S_I(i) == 1 && S_I(j) == 1
ii = min(i,j);
jj = max(i,j);
if S_E(ii,jj) == 1
S_D(i) = 1;
else
S_D(i) = 0;
break;
end
end
end
end
r = r+1;
x_q = zeros(K,length(Q));
for i=1:K
if S_I(i)== 1 && S_D(i) == 0
x = normrnd(mean(i),sigma(i),[N,1]); % original sample, (N,1);
x_sorted = sort(x,1);
for q = 1:length(Q)
x_q(i,q) = x_sorted(weight(q))*(0.5+weight(q)-N*Q(q)) + x_sorted(weight(q)+1)
*(0.5-weight(q)+N*Q(q));
end
end
total_sample_size = total_sample_size + N;
end
```

```
x_q_avg = ((r-1)*x_q_avg + x_q) / r; % K by length(Q)
x_q_sum_sqr = x_q_sum_sqr + x_q.^2; % K by length(Q)
x_q_var = (x_q_sum_sqr - r * x_q_avg.^2) / (r-1); % K by length(Q)
for i = 1:K
for j = 1:K
if i<j && S_I(i) == 1 && S_I(j) == 1 && S_E(i,j)==0
% L_q comparison
for q = 1:length(Q)
sum_diff_x_q(i,j,q) = r*(x_q_avg(i,q)-x_q_avg(j,q));
IB_q_ij = r*(delta_x-d) - (x_q_var(i,q)+x_q_var(j,q))*
((1/beta_k)^(2/(r-1))-1)*(r-1)/(4*d);
OB_q_ij = r*d + (x_q_var(i,q) + x_q_var(j,q))*
((2/alpha_k)^(2/(r-1))-1)*(r-1)/(4*d);
if sum_diff_x_q(i,j,q) > -IB_q_ij && sum_diff_x_q(i,j,q) < IB_q_ij
L_H0(i,j,q) = 1;
elseif sum_diff_x_q(i,j,q) > OB_q_ij %H1: x_q_i > x_q_j
L_H1(i,j,q) = 1;
elseif sum_diff_x_q(i,j,q) < -OB_q_ij %H2: x_q_i < x_q_j
L_H2(i,j,q) = 1;
end
end
% decision
if sum(L_H0(i,j,:)) == length(Q)
S_E(i,j) = 1;
elseif sum(L_H1(i,j,:)) == length(Q)
S_I(j) = 0;
```

```
S_E(:,j) = 0;
S_E(j,:) = 0;
L_H0(:,j,:)=0; L_H1(:,j,:)=0; L_H2(:,j,:)=0;
L_H0(j,:,:)=0; L_H1(j,:,:)=0; L_H2(j,:,:)=0;
elseif sum(L_H2(i,j,:)) == length(Q)
S_I(i) = 0;
S_E(i,:) = 0;
S_E(:,i) = 0;
L_H0(:,i,:)=0; L_H1(:,i,:)=0; L_H2(:,i,:)=0;
L_H0(i,:,:)=0; L_H1(i,:,:)=0; L_H2(i,:,:)=0;
end
end
end
count_S_I = sum(S_I);
count_L_H0 = sum(sum(sum(L_H0)));
count_L_H1 = sum(sum(sum(L_H1)));
count_L_H2 = sum(sum(sum(L_H2)));
count_S_E = sum(sum(S_E));
if count_S_I== 1 || count_S_E==count_S_I*(count_S_I-1)/2 % termination
add_more_observation = 0;
change_N = 0;
elseif (count_L_H0+count_L_H1+count_L_H2)/2+count_S_E == count_S_I*
(count_S_I-1)/2
add_more_observation = 0;
change_N = 1;
else
```

```
add_more_observation = 1;
change_N = 0;
end
end % while add_more_observation
end % while change_N
correct_selection = 0;
if sum(corr_S_I==S_I)==K
correct_selection = 1;
% If more than one system survives, we need to be able to determine the non-dominance relationships correctly
if sum(S_I)>1 && sum(sum(S_E==corr_S_E)) =K^2
correct_selection = 0;
end
end
correct_selection_count = correct_selection_count + correct_selection;
end
%output
fprintf('the percentage of correct selection is: %6.4f \n', correct_selection_count/replication);
fprintf('the average sample size is: %6.4f \n', total_sample_size/replication);
```

Bibliography

[1] S. Andradóttir and S.H. Kim. Fully sequential procedures for comparing constrained systems via simulation. *Naval Research Logistics (NRL)*, 57(5):403–421, 2010.

[2] D. Batur and F.F. Choobineh. Do mean-based ranking and selection procedures consider systems' risk? *Winter Simulation Conference (WSC), Proceedings of the 2009*, pages 423–433, 2009.

[3] D. Batur and F.F. Choobineh. Mean-Variance based ranking and selection. *Winter Simulation Conference (WSC), Proceedings of the 2010*, pages 1160–1166, 2010.

[4] W.J. Baumol. An expected gain-confidence limit criterion for portfolio selection. *Management science*, (1):174–182, 1963.

[5] R.E. Bechhofer. A single-sample multiple decision procedure for ranking means of normal populations with known variances. *The Annals of Mathematical Statistics*, 25(1):16–39, 1954.

[6] R.E. Bechhofer, T.J. Santner, and D.M. Goldsman. Design and analysis of experiments for statistical selection, screening, and multiple comparisons, 1995.

[7] C.H. Chen. An effective approach to smartly allocate computing budget for discrete event simulation. *Decision and Control, 1995., Proceedings of the 34th IEEE Conference on*, 3:2598–2603, 1995.

[8] C.H. Chen, H.C. Chen, and L. Dai. A gradient approach for smartly allocating computing budget for discrete event simulation. *Proceedings of the 28th conference on Winter simulation*, pages 398–405, 1996.

[9] Chun-hung Chen, Donghai He, and M Fu. Efficient simulation budget allocation for selecting an optimal subset. *INFORMS Journal on Computing*, 2010.

[10] H.C. Chen, L. Dai, C.H. Chen, and E. Yücesan. New development of optimal computing budget allocation for discrete event simulation. *Proceedings of the 29th conference on Winter simulation*, pages 334–341, 1997.

[11] S.E. Chick. Selecting the best system: A decision-theoretic approach. *Proceedings of the 29th conference on Winter simulation*, pages 326–333, 1997.

[12] E.J. Dudewicz. The heteroscedastic method. *American Journal of Mathematical and Management Sciences*, (5):179–197, 1996.

[13] D. Goldsman. A multinomial ranking and selection procedure: Simulation and applications. *Winter Simulation Conference (WSC), Proceedings of the 1984*, pages 259–264, 1984.

[14] D. Goldsman, S.H. Kim, W.S. Marshall, and B.L. Nelson. Ranking and selection for steady-state simulation: Procedures and perspectives. *INFORMS Journal on Computing*, 14(1):2–19, 2002.

[15] D. Goldsman and B.L. Nelson. *Comparing systems via simulation.* Wiley Online Library, 1998.

[16] S.S. Gupta. On a decision rule for a problem in ranking means. *Institute of Statistic*, 1956.

[17] Christopher Healey and Seong-hee Kim. Proceedings of the 2007 Winter Simulation Conference S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds. *Area*, pages 522–529, 2007.

[18] S.G. Henderson. Stochastic computer simulation. *management science: simulation*, pages 1–18, 2006.

[19] K. Inoue and S.E. Chick. Comparison of Bayesian and frequentist assessments of uncertainty for selecting the best system. *Proceedings of the 30th conference on Winter simulation*, pages 727–734, 1998.

[20] S.-H. Kim. Comparison with a standard via fully sequential procedures. *ACM Transactions on Modeling and Computer*, (15):155–174, 2005.

[21] S.-H. Kim and B.L. Nelson. A fully sequential procedure for indifference-zone selection in simulation, 2001.

[22] S.H. Kim and B.L. Nelson. On the asymptotic validity of fully sequential selection procedures for steady-state simulation. *Operations Research-Baltimore then Linthicum-*, 54(3):475, 2006.

[23] L.W. Koenig. A procedure for selecting a subset of size m containing the l best of k independent normal populations, with applications to simulation. *Communications in Statistics-Simulation*, pages 719–734, 1985.

[24] H. Levy. *Stochastic dominance: investment decision making under uncertainty*, volume 12. Springer Verlag, 2006.

[25] H.M. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.

[26] H.M. Markowitz. Portfolio Selection: Efficient Diversification of Investments. *wiley*, 1959.

[27] B.L. Nelson. Comparisons with a standard in simulation experiments. *Management Science*, (47):449–463, 2001.

[28] BL Nelson. Control variates for screening, selection, and estimation of the best. *ACM Transactions on Modeling and Computer*, 2006.

[29] B.L. Nelson, J. Swann, and D. Goldsman. Simple procedures for selecting the best simulated system when the number of alternatives is large. *Operations Research*, (49):950–963, 2001.

[30] R.S. Parish. Comparison of quantile estimators in normal sampling. *Biometrics*, 46(1):247–254, 1990.

[31] E. Parzen. Nonparametric statistical data modeling. *Journal of the American Statistical Association*, 74:105–121, 1979.

[32] E. Paulson. A Sequential Procedure for Selecting the Population with the Largest Mean from k Normal Populations. *The Annals of Mathematical Statistics*, 35(1):174–180, 1964.

[33] E. Paulson. Sequential estimation and closed sequential decision procedures. *The Annals of Mathematical Statistics*, 35(3):1048–1058, 1964.

[34] Y. Rinott. On two-stage selection procedures and related probability-inequalities. *Communications in Statistics-Theory and Methods*, 7(8):799–811, 1978.

[35] L.J. Robison and P˜.J. Barry. Risk Efficiency Using Stochastic Dominance and Expected Gain-Confidence Limits. *The Journal of Finance*, 33(4):1244–1249, 1978.

[36] T. Santer and A. Tamhame. Designing experiments for selecting a normal population with a large mean and a small variance. *Design of Experiments: Ranking and Selection*, pages 179–198, 1983.

[37] J.R. Swisher and S.H. Jacobson. A survey of ranking, selection, and multiple comparison procedures for discrete-event simulation. *Proceedings of the 31st conference on Winter simulation: Simulationa bridge to the future-Volume 1*, pages 492–501, 1999.

[38] S Teng and LH Lee. Integration of indifference-zone with multi-objective computing budget allocation. *European Journal of Operational Research*, 2010.

[39] SC Tsai. Fully sequential selection procedures with control variates. *IIE Transactions*, 2009.

[40] SC Tsai and BL Nelson. Combined screening and selection of the best with control variates. *Advancing the Frontiers of Simulation*, 133(Cv), 2009.