Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

RANKING AND SELECTION WITH COVARIATES

Haihui Shen

Department of Management Sciences
City University of Hong Kong
Kowloon, HONG KONG

L. Jeff Hong

Department of Economics and Finance
and Department of Management Sciences
City University of Hong Kong
Kowloon, HONG KONG

Xiaowei Zhang

Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology
Clear Water Bay, HONG KONG

ABSTRACT

We consider a new ranking and selection problem in which the performance of each alternative depends on some observable random covariates. The best alternative is thus not constant but depends on the values of the covariates. Assuming a linear model that relates the mean performance of an alternative and the covariates, we design selection procedures producing policies that represent the best alternative as a function in the covariates. We prove that the selection procedures can provide certain statistical guarantee, which is defined via a nontrivial generalization of the concept of probability of correct selection that is widely used in the conventional ranking and selection setting.

1 INTRODUCTION

Ranking and selection (R&S) is a classic research problem in the simulation literature over the past decades. It represents a broad class of decision-making problems in practice that involve a finite set of competing alternatives whose performances are unknown but can be estimated via simulation. The goal in general is to design a selection procedure that selects the best alternative eventually with certain statistical guarantee; see Kim and Nelson (2006) for a review on the subject.

In this paper, however, we introduce a new R&S problem in which the performance of an alternative depends on some observable random covariates. Hence, the best alternative is not constant but varies as a function in the covariates. A motivation for us to consider this new problem stems from recent advances in customization of decision-making in various areas such as online advertising and healthcare. It is conceivable that by leveraging personal information as covariates, more informative decisions can be made, even tailored to each individual customer or patient. For example, it is found in marketing research that companies can boost both profits and customer satisfaction by sending customers advertisements or promotions that are highly customized based on their demographic and transactional characteristics; see Arora et al. (2008). As another example, it is demonstrated in Yap, Carden, and Kaye (2009) and Kim et al. (2011) that the efficacy of chemotherapy treatments depends on biometrics of the patients such as gene expression and cancer biomarker, and thus personalizing the treatment regimen can improve the treatment outcome significantly.

In the new setting, what a selection procedure produces eventually is a policy that relates the covariates with the best alternative. Like the conventional R&S, this new problem of R&S with covariates is still offline, in contrast to those online problems like multi-armed bandit. Once such decision policy is produced, it can be applied subsequently to select the best alternative after observing the values of the covariates. The new R&S problem reflects a shift in viewpoint regarding the role of simulation in decision-making. We now view simulation as a tool for system control, instead of a tool for system design as in the conventional R&S setting. Such a change in perspective is discussed extensively in the recent article of Nelson (2016) to motivate the so-called simulation analytics.

We assume that the mean performance of an alternative is linear in the covariates and that the variance of the simulation errors may depend on the covariates. A formally similar linear model is adopted in Negoescu, Frazier, and Powell (2011) to solve a R&S problem in the context of drug discovery. In particular, the linear model in their work forms a linear projection from the space of alternatives to the space of attributes. By doing so, the unknown quantities to estimate become the coefficients of the attributes instead of the mean performances of the alternatives. The number of unknown quantities and thus the computational complexity can be reduced dramatically if the number of attributes is much smaller than the number of of alternatives. However, the constituents in their linear model do not vary in the way that the covariates do in our setting. Hence, their R&S problem is still in the conventional sense. The selection procedure designed there produces the best alternative as a static decision rather than the kind of decision policy that we seek.

There are two main approaches to solve a R&S problem, i.e. the frequentist approach and the Bayesian approach; see Branke, Chick, and Schmidt (2007) for comparisons among various selection procedures following either approach. We follow the frequentist approach in this paper and define statistical guarantee via probability of correct selection (PCS). Nevertheless, the presence of covariates complicates the definition of PCS. Specifically, the concept of correct selection in the presence of covariates is a *conditional* event given the values of the covariates, thereby suggesting a conditional PCS. We then define two forms of *unconditional* PCS, one by taking expectation with respect to the distribution of the covariates, whereas the other by taking the minimum over the support of the covariates. The latter is obviously more conservative than the former. Due to the possible dependence of the simulation errors on the covariates, we design selection procedures separately depending on whether such dependence does exist. The procedures are two-stage procedures in a similar form of the Rinott's procedure (Rinott 1978).

The remaining of the paper is organized as follows. In Section 2, we formulate the new R&S problem with covariates and define two forms of PCS accordingly. In Section 3, we present a selection procedure assuming that the simulation errors are independent of the covariates and establish its statistical validity. In Section 4, we address the more general case in which the variances of the simulation errors are not constant but dependent on the covariates. In Section 5, we discuss the so-called least-favorable configuration that generalizes the same concept in the conventional R&S setting. We present the numerical experiments in Section 6 and conclude in Section 7.

2 PROBLEM FORMULATION

Consider a collection of k distinctive simulated alternatives. Suppose that the performance of each alternative depends on $\boldsymbol{X}_c = (X_1, \cdots, X_d)^\intercal$, a vector of observable random covariates with arbitrary multivariate distribution and support $\Theta_c \subseteq \mathbb{R}^d$. For notation simplicity, let $\boldsymbol{X} \coloneqq (1, X_1, \cdots, X_d)^\intercal$ be the degenerate random vector with support $\Theta \coloneqq \{1\} \times \Theta_c$. For each $i = 1, \cdots, k$, let $Y_{i\ell}$ denote the ℓth sample from alternative i, $\ell \geq 1$. Let $\boldsymbol{x}$ be the realization of $\boldsymbol{X}$, and we write $Y_{i\ell}(\boldsymbol{x})$ throughout the paper to emphasize the dependence on $\boldsymbol{x}$. We assume the following linear model.

Assumption 1 For each $i = 1, \cdots, k$ and $\ell = 1, 2, \cdots$, conditioning on $\boldsymbol{X} = \boldsymbol{x}$,

$$Y_{i\ell}(\boldsymbol{x}) = \boldsymbol{x}^\intercal \boldsymbol{\beta}_i + \varepsilon_{i\ell}(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1}, \cdots, \beta_{id})^\intercal \in \mathbb{R}^{d+1}$ is a vector of unknown parameters and $\varepsilon_{i\ell}(\boldsymbol{x})$ is normally distributed with mean 0 and variance $\sigma_i^2(\boldsymbol{x})$; moreover, $\varepsilon_{i\ell}(\boldsymbol{x})$ is independent of $\varepsilon_{i'\ell'}(\boldsymbol{x}')$ for any $(i,\ell,\boldsymbol{x}) \neq (i',\ell',\boldsymbol{x}')$.

Our goal is to select the alternative with the largest mean conditioning on $\boldsymbol{X}$, i.e., to find

$$i^*(\boldsymbol{X}) := \arg\max_{1\le i\le k} \{\boldsymbol{X}^\intercal \boldsymbol{\beta}_i | \boldsymbol{X}\}.$$

Let $\widehat{i^*}(\boldsymbol{X})$ denote the selected alternative based on a selection procedure. Clearly, there is no guarantee that $\widehat{i^*}(\boldsymbol{X}) = i^*(\boldsymbol{X})$ with certainty given a finite computational budget, due to the random simulation errors. We therefore adopt an indifference-zone (IZ) formulation as follows. Let $\delta > 0$ be a prespecified IZ parameter and define the event of correct selection (CS) as $\left\{\boldsymbol{X}^\intercal \boldsymbol{\beta}_{i^*(\boldsymbol{X})} - \boldsymbol{X}^\intercal \boldsymbol{\beta}_{\widehat{i^*}(\boldsymbol{X})} < \delta \middle| \boldsymbol{X}\right\}$. However, the presence of covariates in our formulation complicates the definition of PCS. We first define the *conditional* PCS

$$\mathrm{PCS}(\boldsymbol{X}) := \Pr\left\{\boldsymbol{X}^\intercal \boldsymbol{\beta}_{i^*(\boldsymbol{X})} - \boldsymbol{X}^\intercal \boldsymbol{\beta}_{\widehat{i^*}(\boldsymbol{X})} < \delta \middle| \boldsymbol{X}\right\}, \tag{2}$$

where the probability is taken with respect to the distribution of the samples that are used by the selection procedure to produce $\widehat{i^*}(\boldsymbol{X})$. Then, we introduce two forms of *unconditional* PCS that may be of interest under different circumstances, namely

$$\mathrm{PCS}_{\mathrm{E}} := \mathrm{E}\left[\mathrm{PCS}(\boldsymbol{X})\right], \tag{3}$$

where the expectation is taken with respect to the distribution of $\boldsymbol{X}$, and

$$\mathrm{PCS}_{\min} := \min_{\boldsymbol{x}\in\Theta} \mathrm{PCS}(\boldsymbol{x}). \tag{4}$$

With the unconditional PCS defined, we aim to design selection procedures that guarantee $\mathrm{PCS}_{\mathrm{E}} \ge 1-\alpha$ or $\mathrm{PCS}_{\min} \ge 1-\alpha$ for some user-specified $1/k < 1-\alpha < 1$.

Remark 1 $\mathrm{PCS}_{\min}$ represents a more conservative criterion than $\mathrm{PCS}_{\mathrm{E}}$ since $\mathrm{PCS}_{\mathrm{E}} \ge \mathrm{PCS}_{\min}$ by definition. Hereafter, when $\mathrm{PCS}_{\min}$ is desired we additionally assume that Θ is a compact set. In practice when the distribution of $\boldsymbol{X}$ is known or can be credibly estimated from the historical data, one is able to use $\mathrm{PCS}_{\mathrm{E}}$. When the distribution is totally unknown and there is no historical data at all except knowing that $\boldsymbol{X}$ varies in some bounded range, one may consider to use $\mathrm{PCS}_{\min}$ as risk protection. In another more realistic situation where only limited data of $\boldsymbol{X}$ are available, it seems more reasonable to use the estimated distribution while taking the input uncertainly in account when calculating $\mathrm{PCS}_{\mathrm{E}}$. This is certainly an interesting topic for future study.

Remark 2 In the conventional R&S setting, we would have considered the best alternative to be $i^\dagger$, where $i^\dagger := \arg\max_{1\le i\le k}\{\mathrm{E}[\boldsymbol{X}]^\intercal \boldsymbol{\beta}_i\}$, regardless of the realized value of $\boldsymbol{X}$. Consider the conceptual situation where the true answers of $i^\dagger$ and $i^*(\boldsymbol{X})$ are exactly known to us and used to guide the selection. Notice that $\mathrm{E}\left[\boldsymbol{X}^\intercal \boldsymbol{\beta}_{i^*(\boldsymbol{X})}\right] \ge \mathrm{E}\left[\boldsymbol{X}^\intercal \boldsymbol{\beta}_{i^\dagger}\right] = \mathrm{E}[\boldsymbol{X}]^\intercal \boldsymbol{\beta}_{i^\dagger}$. This suggests that in the presence of covariates, selecting an alternative *after* observing their values is deemed to be better than making such a decision *before* the observation (or, ignoring the covariates), demonstrating the advantage of the new R&S setting.

As a first attempt to address R&S with covariates, we consider the *fixed design* setting as follows. We choose and fix m design points $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m \in \Theta$. Let $\mathscr{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m)^\intercal$. We assume that we can simulate alternative i at design point $\boldsymbol{x}_j$ as many times as we want, for each $i = 1, \cdots, k$ and $j = 1, \cdots, m$. It is worth emphasizing that such fixed design setting is suitable in the simulation context, while in the context of statistical learning, it may not be applicable and instead some random design approach is required. We further make the following technical assumption throughout the remaining of the paper.

Assumption 2 $\mathscr{X}$ has full column rank, i.e. $\mathscr{X}^\intercal \mathscr{X}$ is nonsingular.

To reveal the practical use of the proposed setting of R&S with covariates under fixed design, we now describe an example of personalized medicine in the treatment of Barrett's esophagus, a precursor to esophageal cancer. Here the different alternatives represent the different treatment regimens. The covariates represent each individual's attributes such as age, grade of dysplasia, response to drugs, etc. Since there is a simulation model developed by Choi et al. (2014) to be used to characterize the development of esophageal cancer, we can fix some design matrix of covariates and then run simulations on each point for each alternative. When the simulation is completed offline, we hope to obtain a decision policy which can be executed online, i.e., for any real patient arriving with specific attributes, the policy suggests the best treatment regimen for him.

It turns out that in order to achieve the statistical guarantee defined via $\mathrm{PCS_E}$ or $\mathrm{PCS_{min}}$, selection procedures need to be designed separately depending on whether the variances of the simulation errors are constants relative to the covariates. We therefore differentiate the case of *homoscedastic* errors and that of *heteroscedastic* errors. Notice that this analogizes the difference between the ordinary least squares method and the generalized least squares method in linear regression models.

3 PROCEDURE WITH HOMOSCEDASTIC ERRORS

Assumption 3 $\sigma_i^2(\boldsymbol{x}) \equiv \sigma_i^2 < \infty$ for $\boldsymbol{x} \in \Theta$ and $i = 1, \cdots, k$.

3.1 The Procedure

We develop a two-stage procedure for fixed design under homoscedastic errors, denoted as Procedure FDHom. The goal of the procedure is that, once finished, it produces a policy to select the best alternative given the realized values of covariates with certain PCS guarantee. It is described as follows:

Step 0. Specify the IZ parameter $\delta > 0$ and PCS requirement $1-\alpha$. Determine the number of design points m and the appropriate design matrix $\mathscr{X}$. Determine n_0, the number of batches in the first stage. Calculate the critical constant h. $h = h_{\mathrm{E}}$ if $\mathrm{PCS_E} \geq 1-\alpha$ is required or $h = h_{\mathrm{min}}$ if $\mathrm{PCS_{min}} \geq 1-\alpha$ is required, which satisfies

$$\mathrm{E}\left\{\int_0^\infty \left[\int_0^\infty \Phi\left(\frac{1}{\sqrt{1/y+1/x}}\frac{h_{\mathrm{E}}}{\sqrt{(n_0m-1-d)\boldsymbol{X}^\intercal(\mathscr{X}^\intercal\mathscr{X})^{-1}\boldsymbol{X}}}\right) f(x)dx\right]^{k-1} f(y)dy\right\} = 1-\alpha, \tag{5}$$

or,

$$\min_{\boldsymbol{x}\in\Theta}\left\{\int_0^\infty \left[\int_0^\infty \Phi\left(\frac{1}{\sqrt{1/y+1/x}}\frac{h_{\mathrm{min}}}{\sqrt{(n_0m-1-d)\boldsymbol{x}^\intercal(\mathscr{X}^\intercal\mathscr{X})^{-1}\boldsymbol{x}}}\right) f(x)dx\right]^{k-1} f(y)dy\right\} = 1-\alpha, \tag{6}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function (cdf), $f(\cdot)$ is the probability density function (pdf) of the chi-squared random variable with n_0m-1-d degrees of freedom, and the expectation is taken with respect to the distribution of the covariates $\boldsymbol{X}$.

Step 1. For all $i = 1, \cdots, k$, take n_0 batches of observations on the fixed design matrix $\mathscr{X}$ and denote them as $\boldsymbol{Y}_{i1} = (Y_{i1}(\boldsymbol{x}_1), \cdots, Y_{i1}(\boldsymbol{x}_m))^\intercal, \cdots, \boldsymbol{Y}_{in_0} = (Y_{in_0}(\boldsymbol{x}_1), \cdots, Y_{in_0}(\boldsymbol{x}_m))^\intercal$. For all $i = 1, \cdots, k$, let $\widehat{\boldsymbol{\beta}}_i(n_0) = \frac{1}{n_0}(\mathscr{X}^\intercal\mathscr{X})^{-1}\mathscr{X}^\intercal\sum_{\ell=1}^{n_0}\boldsymbol{Y}_{i\ell}$, and $S_i^2 = \frac{1}{n_0m-1-d}\sum_{\ell=1}^{n_0}(\boldsymbol{Y}_{i\ell} - \mathscr{X}\widehat{\boldsymbol{\beta}}_i(n_0))^\intercal(\boldsymbol{Y}_{i\ell} - \mathscr{X}\widehat{\boldsymbol{\beta}}_i(n_0))$. Furthermore, let

$$N_i = \max\left\{\left\lceil\frac{h^2S_i^2}{\delta^2}\right\rceil, n_0\right\}. \tag{7}$$

Step 2. For all $i = 1, \cdots, k$, take $N_i - n_0$ batches of observations on the fixed design matrix $\mathscr{X}$ and denote them as $\boldsymbol{Y}_{i,n_0+1}, \cdots, \boldsymbol{Y}_{iN_i}$. For all $i = 1, \cdots, k$, let $\widehat{\boldsymbol{\beta}}_i = \frac{1}{N_i}(\mathscr{X}^\intercal \mathscr{X})^{-1}\mathscr{X}^\intercal \sum_{\ell=1}^{N_i} \boldsymbol{Y}_{i\ell}$. Then, the selected alternative conditioning on $\boldsymbol{X}$ is given by $\widehat{i}^*(\boldsymbol{X}) = \arg\max_{1\le i\le k}\{\boldsymbol{X}^\intercal \widehat{\boldsymbol{\beta}}_i | \boldsymbol{X}\}$.

3.2 Statistical Validity

We have the following statistical validity of the Procedure FDHom:

Theorem 1 Under Assumptions 1, 2, and 3, the Procedure FDHom ensures that the $\mathrm{PCS}_{\mathrm{E}}$ defined in (3) or the $\mathrm{PCS}_{\min}$ defined in (4) is at least $1-\alpha$, i.e., $\mathrm{PCS}_{\mathrm{E}} \ge 1-\alpha$ or $\mathrm{PCS}_{\min} \ge 1-\alpha$.

Theorem 1 can be proved based on the following Lemma 1 and Lemma 2. Lemma 1 is an extension from a result in Stein (1945), which is the foundation of the traditional R&S procedures for unknown variances (see also Kim and Nelson (2006, §3.1)). Lemma 2 is due to Slepian (1962).

Lemma 1 Suppose that $\boldsymbol{Y} = \mathscr{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathscr{X} \in \mathbb{R}^{m\times d}$ is a fixed known matrix, $\boldsymbol{\beta} \in \mathbb{R}^d$ is a fixed but unknown vector, and $\boldsymbol{\varepsilon}$ is a $m\times 1$ random vector such that $\boldsymbol{\varepsilon} \sim \mathscr{N}(\boldsymbol{0}_m, \sigma^2 \mathscr{I}_m)$, where $\boldsymbol{0}_m$ is a $m\times 1$ vector with all 0's, and $\mathscr{I}_m$ is an identity matrix of size m. $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_n, \boldsymbol{Y}_{n+1}, \boldsymbol{Y}_{n+2}, \cdots$, are independent samples of $\boldsymbol{Y}$. Suppose T is a set of random variables and T is independent of $\sum_{\ell=1}^{n}\boldsymbol{Y}_\ell$ and of $\boldsymbol{Y}_{n+1}, \boldsymbol{Y}_{n+2}, \cdots$. Suppose $N \ge n$ is an integer valued function only of T. Define $\widehat{\boldsymbol{\beta}} = \frac{1}{N}(\mathscr{X}^\intercal\mathscr{X})^{-1}\mathscr{X}^\intercal\sum_{\ell=1}^{N}\boldsymbol{Y}_\ell$.

(i) For any given vector $\boldsymbol{x} \in \mathbb{R}^d$, conditioning on T,

$$\boldsymbol{x}^\intercal\widehat{\boldsymbol{\beta}} \sim \mathscr{N}\left(\boldsymbol{x}^\intercal\boldsymbol{\beta}, \frac{\sigma^2}{N}\boldsymbol{x}^\intercal(\mathscr{X}^\intercal\mathscr{X})^{-1}\boldsymbol{x}\right).$$

(ii) For any given vector $\boldsymbol{x} \in \mathbb{R}^d$,

$$V := \sqrt{N}\frac{\boldsymbol{x}^\intercal\widehat{\boldsymbol{\beta}} - \boldsymbol{x}^\intercal\boldsymbol{\beta}}{\sigma\sqrt{\boldsymbol{x}^\intercal(\mathscr{X}^\intercal\mathscr{X})^{-1}\boldsymbol{x}}},$$

is independent of T, and furthermore $V \sim \mathscr{N}(0,1)$.

Lemma 2 (Slepian 1962) Let $(Z_1, \cdots, Z_k)$ have a k-variate normal distribution with finite mean vector and covariance matrix whose (i,j)th element is ρ_{ij} for $i, j = 1, \cdots, k$. Let $c_1, \cdots, c_k$ be some constants. If $\rho_{ij} \ge 0$ for all $i, j = 1, \cdots, k$ and $i \ne j$, then

$$\Pr\left\{\bigcap_{i=1}^{k}(Z_i \ge c_i)\right\} \ge \prod_{i=1}^{k}\Pr\{Z_i \ge c_i\}.$$

Here we provide a sketch of the proof of Theorem 1, which briefly shows how Lemma 1 and Lemma 2 facilitate the proof and where equations (5) and (6) arise from. Conditioning on $\boldsymbol{X}$, without loss of generality, we denote the best alternative as alternative 1, i.e., $i^*(\boldsymbol{X}) = 1$, and $\Omega(\boldsymbol{X}) := \{i = 2, \cdots, k : \boldsymbol{X}^\intercal\boldsymbol{\beta}_1 - \boldsymbol{X}^\intercal\boldsymbol{\beta}_i \ge \delta | \boldsymbol{X}\}$. By the result (i) of Lemma 1 and N_i defined in (7), we can have for any $i \in \Omega(\boldsymbol{X})$,

$$\Pr\left\{\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i > 0 \middle| \boldsymbol{X}, S_1^2, S_i^2\right\} \ge \Phi\left(\frac{1}{\sqrt{\sigma_1^2/S_1^2 + \sigma_i^2/S_i^2}}\frac{h}{\sqrt{\boldsymbol{X}^\intercal(\mathscr{X}^\intercal\mathscr{X})^{-1}\boldsymbol{X}}}\right).$$

Then by applying Lemma 2, we will have

$$\mathrm{PCS}(\boldsymbol{X}) \ge \mathrm{E}\left[\prod_{i\in\Omega(\boldsymbol{X})}\Phi\left(\frac{1}{\sqrt{\sigma_1^2/S_1^2 + \sigma_i^2/S_i^2}}\frac{h}{\sqrt{\boldsymbol{X}^\intercal(\mathscr{X}^\intercal\mathscr{X})^{-1}\boldsymbol{X}}}\right)\middle|\boldsymbol{X}\right].$$

Notice that $(n_0m-1-d)S_i^2/\sigma_i^2$, for $i \in \{1,\Omega(\boldsymbol{X})\}$, are independent chi-squared random variables with n_0m-1-d degrees of freedom, and the cardinality of $\Omega(\boldsymbol{X})$ is at most $k-1$. Hence, we can obtain that

$$\mathrm{PCS}(\boldsymbol{X}) \geq \int_0^\infty \left[\int_0^\infty \Phi\left(\frac{1}{\sqrt{1/y+1/x}}\frac{h}{\sqrt{(n_0m-1-d)\boldsymbol{X}^\intercal(\mathscr{X}^\intercal\mathscr{X})^{-1}\boldsymbol{X}}}\right) f(x)dx\right]^{k-1} f(y)dy.$$

Finally, $\mathrm{PCS_E} \geq 1-\alpha$ or $\mathrm{PCS_{min}} \geq 1-\alpha$ is guaranteed due to (5) or (6).

4 PROCEDURE WITH HETEROSCEDASTIC ERRORS

Assumption 4 $\sigma_i^2(\boldsymbol{x}) < \infty$, for $\boldsymbol{x} \in \Theta$ and $i = 1,\cdots,k$.

4.1 The Procedure

We develop a two-stage procedure for fixed design under heteroscedastic errors, denoted as Procedure FDHet. The goal of the procedure is that, once finished, it produces a policy to select the best alternative given the realized values of covariates with certain PCS guarantee. It is described as follows:

Step 0. Specify the IZ parameter $\delta > 0$ and PCS requirement $1-\alpha$. Determine the number of design points m and the appropriate design matrix $\mathscr{X}$. Determine n_0, the number of batches in the first stage. Calculate the critical constant h. $h = h_{\mathrm{E}}$ if $\mathrm{PCS_E} \geq 1-\alpha$ is required or $h = h_{\mathrm{min}}$ if $\mathrm{PCS_{min}} \geq 1-\alpha$ is required, which satisfies

$$\mathrm{E}\left\{\int_0^\infty \left[\int_0^\infty \Phi\left(\frac{1}{\sqrt{1/y+1/x}}\frac{h_{\mathrm{E}}}{\sqrt{(n_0-1)\boldsymbol{X}^\intercal(\mathscr{X}^\intercal\mathscr{X})^{-1}\boldsymbol{X}}}\right) f_{(1)}(x)dx\right]^{k-1} f_{(1)}(y)dy\right\} = 1-\alpha, \tag{8}$$

or,

$$\min_{\boldsymbol{x}\in\Theta}\left\{\int_0^\infty \left[\int_0^\infty \Phi\left(\frac{1}{\sqrt{1/y+1/x}}\frac{h_{\mathrm{min}}}{\sqrt{(n_0-1)\boldsymbol{x}^\intercal(\mathscr{X}^\intercal\mathscr{X})^{-1}\boldsymbol{x}}}\right) f_{(1)}(x)dx\right]^{k-1} f_{(1)}(y)dy\right\} = 1-\alpha, \tag{9}$$

where $f_{(1)}(\cdot)$ is the pdf of the smallest order statistic of m independently and identically distributed (i.i.d.) chi-squared random variables with n_0-1 degrees of freedom, i.e.,

$$f_{(1)}(t) = mf(t)(1-F(t))^{m-1},$$

with $f(\cdot)$ and $F(\cdot)$ here denoting the pdf and cdf of the chi-squared random variable with n_0-1 degrees of freedom, and the expectation is taken with respect to the distribution of the covariates $\boldsymbol{X}$.

Step 1. For all $i = 1,\cdots,k$, take n_0 batches of observations on the fixed design matrix $\mathscr{X}$ and denote them as $Y_{i1}(\boldsymbol{x}_1),\cdots,Y_{i1}(\boldsymbol{x}_m),\cdots,Y_{in_0}(\boldsymbol{x}_1),\cdots,Y_{in_0}(\boldsymbol{x}_m)$. For all $i = 1,\cdots,k$, let

$$\overline{Y}_{i1} = \frac{1}{n_0}\sum_{\ell=1}^{n_0} Y_{i\ell}(\boldsymbol{x}_1),\cdots,\overline{Y}_{im} = \frac{1}{n_0}\sum_{\ell=1}^{n_0} Y_{i\ell}(\boldsymbol{x}_m),$$

$$S_{i1}^2 = \frac{1}{n_0-1}\sum_{\ell=1}^{n_0}\left(Y_{i\ell}(\boldsymbol{x}_1)-\overline{Y}_{i1}\right)^2,\cdots,S_{im}^2 = \frac{1}{n_0-1}\sum_{\ell=1}^{n_0}\left(Y_{i\ell}(\boldsymbol{x}_m)-\overline{Y}_{im}\right)^2.$$

Furthermore, let

$$N_{i1} = \max\left\{\left\lceil\frac{h^2S_{i1}^2}{\delta^2}\right\rceil, n_0\right\},\cdots,N_{im} = \max\left\{\left\lceil\frac{h^2S_{im}^2}{\delta^2}\right\rceil, n_0\right\}. \tag{10}$$

Step 2. For all $i = 1, \cdots, k$, take $N_{i1} - n_0$ observations on the design point $\boldsymbol{x}_1$ and denote them as $Y_{i,n_0+1}(\boldsymbol{x}_1), \cdots, Y_{iN_{i1}}(\boldsymbol{x}_1)$; $\cdots$; take $N_{im} - n_0$ observations on the design point $\boldsymbol{x}_m$ and denote them as $Y_{i,n_0+1}(\boldsymbol{x}_m), \cdots, Y_{iN_{im}}(\boldsymbol{x}_m)$. For all $i = 1, \cdots, k$, let $\widehat{Y}_{i1} = \frac{1}{N_{i1}} \sum_{\ell=1}^{N_{i1}} Y_{i\ell}(\boldsymbol{x}_1), \cdots, \widehat{Y}_{im} = \frac{1}{N_{im}} \sum_{\ell=1}^{N_{im}} Y_{i\ell}(\boldsymbol{x}_m)$, and $\widehat{\boldsymbol{Y}}_i = (\widehat{Y}_{i1}, \cdots, \widehat{Y}_{im})^\intercal$. Furthermore, let $\widehat{\boldsymbol{\beta}}_i = (\mathscr{X}^\intercal \mathscr{X})^{-1} \mathscr{X}^\intercal \widehat{\boldsymbol{Y}}_i$. Then, the selected alternative conditioning on $\boldsymbol{X}$ is given by $\widehat{i}^*(\boldsymbol{X}) = \arg\max_{1 \le i \le k} \{\boldsymbol{X}^\intercal \widehat{\boldsymbol{\beta}}_i | \boldsymbol{X}\}$.

4.2 Statistical Validity

We have the following statistical validity of the Procedure FDHet:

Theorem 2 Under Assumptions 1, 2, and 4, the Procedure FDHet ensures that the $\mathrm{PCS}_{\mathrm{E}}$ defined in (3) or the $\mathrm{PCS}_{\min}$ defined in (4) is at least $1-\alpha$, i.e., $\mathrm{PCS}_{\mathrm{E}} \ge 1-\alpha$ or $\mathrm{PCS}_{\min} \ge 1-\alpha$.

Theorem 2 can be proved based on the Lemma 2 and the following Lemma 3, which is an extension (a more general version) of Lemma 1.

Lemma 3 Suppose that

$$Y(\boldsymbol{x}_1) = \boldsymbol{x}_1^\intercal \boldsymbol{\beta} + \varepsilon(\boldsymbol{x}_1), \cdots, Y(\boldsymbol{x}_m) = \boldsymbol{x}_m^\intercal \boldsymbol{\beta} + \varepsilon(\boldsymbol{x}_m),$$

where $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m \in \mathbb{R}^d$ are m fixed known vectors, $\boldsymbol{\beta} \in \mathbb{R}^d$ is a fixed but unknown vector. Suppose $\varepsilon(\boldsymbol{x}_1), \cdots, \varepsilon(\boldsymbol{x}_m)$ are independent random variables, and $\varepsilon(\boldsymbol{x}_1) \sim \mathscr{N}(0, \sigma_1^2), \cdots, \varepsilon(\boldsymbol{x}_m) \sim \mathscr{N}(0, \sigma_m^2)$. $Y_1(\boldsymbol{x}_j)$, $\cdots$, $Y_n(\boldsymbol{x}_j)$, $Y_{n+1}(\boldsymbol{x}_j), Y_{n+2}(\boldsymbol{x}_j), \cdots$, are independent samples of $Y(\boldsymbol{x}_j)$, for all $j = 1, \cdots, m$. Suppose T is a set of random variables and T is independent of $\sum_{\ell=1}^{n} Y_\ell(\boldsymbol{x}_j)$ and of $Y_{n+1}(\boldsymbol{x}_j), Y_{n+2}(\boldsymbol{x}_j), \cdots$, for all $j = 1, \cdots, m$. Suppose $N_1, \cdots, N_m \ge n$ are all integer valued functions only of T. Let $\widehat{Y}_j = \frac{1}{N_j} \sum_{\ell=1}^{N_j} Y_\ell(\boldsymbol{x}_j)$, for all $j = 1, \cdots, m$. Let $\widehat{\boldsymbol{Y}} = (\widehat{Y}_1, \cdots, \widehat{Y}_m)^\intercal$, and $\mathscr{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m)^\intercal$. Define $\widehat{\boldsymbol{\beta}} = (\mathscr{X}^\intercal \mathscr{X})^{-1} \mathscr{X}^\intercal \widehat{\boldsymbol{Y}}$, and $\Sigma = \mathrm{Diag}\left\{\sigma_1^2/N_1, \cdots, \sigma_m^2/N_m\right\}$.

(i) For any given vector $\boldsymbol{x} \in \mathbb{R}^d$, conditioning on T,

$$\boldsymbol{x}^\intercal \widehat{\boldsymbol{\beta}} \sim \mathscr{N}(\boldsymbol{x}^\intercal \boldsymbol{\beta}, \boldsymbol{x}^\intercal (\mathscr{X}^\intercal \mathscr{X})^{-1} \mathscr{X}^\intercal \Sigma \mathscr{X} (\mathscr{X}^\intercal \mathscr{X})^{-1} \boldsymbol{x}).$$

(ii) For any given vector $\boldsymbol{x} \in \mathbb{R}^d$,

$$V := \frac{\boldsymbol{x}^\intercal \widehat{\boldsymbol{\beta}} - \boldsymbol{x}^\intercal \boldsymbol{\beta}}{\sqrt{\boldsymbol{x}^\intercal (\mathscr{X}^\intercal \mathscr{X})^{-1} \mathscr{X}^\intercal \Sigma \mathscr{X} (\mathscr{X}^\intercal \mathscr{X})^{-1} \boldsymbol{x}}},$$

is independent of T, and furthermore $V \sim \mathscr{N}(0,1)$.

The proof of Theorem 2 shares the similar steps as in the proof of Theorem 1, but the covariance matrix on the design points is now a general diagonal matrix. With the help of Lemma 2, Lemma 3 and the introduced smallest order statistics, similar result of $\mathrm{PCS}(\boldsymbol{X})$ as in the proof of Theorem 1 can be obtained. Finally, $\mathrm{PCS}_{\mathrm{E}} \ge 1-\alpha$ or $\mathrm{PCS}_{\min} \ge 1-\alpha$ is guaranteed due to (8) or (9).

4.3 Implementation Guide

In Procedure FDHom (or FDHet), h_{E} and $h_{\min}$ can only be determined by solving (5) (or (8)) and (6) (or (9)) numerically. In our implementation, the integration (including the expectation) is computed by the MATLAB built-in numerical integration functions `integral` and `trapz`, and the h_{E} and $h_{\min}$ are solved by the MATLAB built-in root finding function `fzero`. It is worth mentioning that the trapezoidal method based numerical integration such as `integral` and `trapz` will suffer from the curse of dimensionality. So if the dimension of covariates, d, is high, it is certainly preferable to use Monte Carlo method to approximate the expectation.

When solving the minimization problem in (6) and (9), actually only the extreme points of the convex hull of Θ need to be checked. For example, when Θ_c is some (continuous or discrete) hyper-rectangle with dimension d, only the 2^d corner points need to be compared.

Like in the linear regression, the homoscedasticity assumption simplifies the mathematical and computational treatment, but may often be violated in reality. When the Procedure FDHom is misused in the heteroscedastic errors case, the targeted PCS may fail to be guaranteed. This is because the aggregated sample variance may underestimate the variance at some design points, leading to insufficient sample allocation on those points. On the other hand, if the Procedure FDHet is used in the homoscedastic errors case, the resulting sample size will be over-conservative. The intuitive explanation is that when the errors are indeed homoscedastic, the samples on all different design points can be aggregated to calculate the sample variance, which is a better estimator (i.e., with smaller variance) of the common variance of homoscedastic errors than the separately computed sample variance on each design point. These trade-off effects will be well reflected in the later numerical experiments in Section 6.

The above discussion provides us a simple guiding principle of choosing procedure in practice. When the realistic problem we are facing is very close to the homoscedastic errors case, or not that close but we can tolerate small amount of deviation below the designed PCS, we may consider to apply Procedure FDHom. If the errors are notably heteroscedastic and we are very strict on the designed PCS, we had better choose Procedure FDHet.

5 LEAST-FAVORABLE CONFIGURATION

For traditional R&S problem, there is a concept called least-favorable configuration (LFC), which is a set of alternative configurations where the designed PCS is the most difficult to deliver. When talking about (and searching for) the LFC, people usually focus specifically on the mean configuration, and the other configurations such as variances are kept the same. To many procedures, for example, the two-stage procedures like Rinott's procedure (Rinott 1978), and the sequential procedures like KN family procedures (Kim and Nelson 2001, Hong 2006), it is well known that given k and variance configuration, the LFC is the *slippage configuration (SC)*, i.e., $\mu_1 - \delta = \mu_i$, for $i = 2,3,\cdots,k$, where μ_i is the mean of alternative i for $i = 1,\cdots,k$.

While in the case of R&S with covariates, each alternative is no longer a single population, but the idea of SC can be easily extended to this case. We define some configuration which is analogous to SC and call it the *generalized slippage configuration (GSC)*:

$$\beta_{10} - \delta = \beta_{i0},\ \beta_{1j} = \beta_{ij},\ \text{for } j = 1,\cdots,d \text{ and } i = 2,3,\cdots,k. \tag{11}$$

We can see that when all $\beta_{ij} = 0$ for $j = 1,\cdots,d$ and $i = 1,\cdots,k$, the GSC reduces to SC. Not surprisingly, we find that the GSC is the LFC of R&S with covariates to many possible procedures (including ours), as stated in the following Theorem 3.

Theorem 3 For R&S with covariates defined under Assumption 1, GSC is the LFC for a selection procedure on fixed design matrix $\mathscr{X}$ with the following features:

(i) On the design point $\boldsymbol{x}_j$ at alternative i, only the average of N_{ij} observations, which is denoted as $\widehat{Y}_{ij}$ for $j = 1,\cdots,m$ and $i = 1,\cdots,k$, is used when finally selecting the alternative.

(ii) N_{ij} is independent of the mean configuration. Moreover, conditioning on all N_{ij}, $\widehat{Y}_{ij}$ is independent of each other and $\widehat{Y}_{ij} \sim \mathscr{N}\left(\boldsymbol{x}_j^\intercal \boldsymbol{\beta}_i, \sigma_i^2(\boldsymbol{x}_j)/N_{ij}\right)$, for $j = 1,\cdots,m$ and $i = 1,\cdots,k$.

(iii) Let $\widehat{\boldsymbol{Y}}_i = (\widehat{Y}_{i1},\cdots,\widehat{Y}_{im})^\intercal$ and $\widehat{\boldsymbol{\beta}}_i = (\mathscr{X}^\top \mathscr{X})^{-1}\mathscr{X}^\intercal \widehat{\boldsymbol{Y}}_i$ for $i = 1,\cdots,k$. The selected alternative conditioning on $\boldsymbol{X}$ is given by $\widehat{i}^*(\boldsymbol{X}) = \arg\max_{1\le i\le k}\left\{\boldsymbol{X}^\intercal \widehat{\boldsymbol{\beta}}_i \middle| \boldsymbol{X}\right\}$.

Obviously the proposed procedures FDHom and FDHet possess those features specified in Theorem 3, so we have the following Corollary 4 immediately.

Corollary 4 GSC is the LFC for the Procedure FDHom and Procedure FDHet.

6 NUMERICAL EXPERIMENTS

In this section we show the results of extensive numerical evaluations of Procedures FDHom and FDHet. To avoid the overly large number of cases arising from the full factorial combinations of various factors, such as the dimension, number of alternatives, mean configuration, variance configuration, etc., we create a benchmark case, and then investigate the effect of each factor by varying that factor while keeping others unchanged.

First we introduce the configuration of the benchmark case. Notice that some configurations are defined in a generic way, which means later they will change accordingly in the comparing cases. We let $d = 3$ and $k = 5$ in the benchmark case. They will vary later in other cases. The covariates, $X_1, \cdots, X_d$, are i.i.d. Unif$[0,1]$ random variables. For the fixed design, we apply the factorial sampling on $(X_1, \cdots, X_d)^\intercal$ by taking $m = 2^d$ design points in the set $\{0, 0.5\} \times \cdots \times \{0, 0.5\}$, which is a Cartesian product of d identical set $\{0, 0.5\}$. For the mean configuration, the GSC is used. For the variance configuration, we let $\sigma_i^2(\boldsymbol{x}) \equiv \sigma_i^2$, which is the case of homoscedastic errors inside the alternative. Besides, we let $\sigma_1 = \cdots = \sigma_k = 10$, which means there are equal variances among alternatives.

Then we design 8 comparing cases. They are all based on the benchmark case, and each time only one factor deviates from the benchmark. The details are as follows: (1) Set $k = 2$. (2) Set $k = 8$. (3) We no longer consider the GSC for the mean configuration, instead we randomly generate all components of $\boldsymbol{\beta}_i$ from Unif$[0,5]$, for $i = 1, \cdots, 5$. (4) Consider the increasing variances (IV) among alternatives, i.e., $\sigma_1 = 5$, $\sigma_2 = 7.5$, $\sigma_3 = 10$, $\sigma_4 = 12.5$, $\sigma_5 = 15$. (5) Consider the decreasing variances (DV) among alternatives, i.e., $\sigma_1 = 15$, $\sigma_2 = 12.5$, $\sigma_3 = 10$, $\sigma_4 = 7.5$, $\sigma_5 = 5$. (6) Consider the case of heteroscedastic errors inside the alternative, i.e., let $\sigma_i^2(\boldsymbol{x}) = 100(\boldsymbol{x}^\intercal \boldsymbol{\beta}_i)^2$, for $i = 1, \cdots, 5$. (7) Set $d = 1$. (8) Set $d = 5$. Notice that Cases (1) and (2) only change the number of alternatives. In Cases (3)-(6) the configuration of alternatives is changed, where the Case (3) changes the mean configuration and the Cases (4)-(6) change the variance configuration. It is worth noting that the Cases (4) and (5) only change the variances of errors among the alternatives, while the errors for individual alternatives are still homoscedastic. The Cases (7) and (8) change the dimensionality of covariates.

In all cases including the benchmark case and the 8 comparing cases, we set $\alpha = 0.05$ (i.e., PCS=95%), $\delta = 1$ and $n_0 = 50$. Here we also consider the two different forms of PCS, i.e., the PCS_E and PCS_min. No matter when which form of PCS is designed for, we evaluate both the real PCS_E and real PCS_min for reference. For each form of PCS and each case, 10^4 macro replications are carried out for Procedures FDHom and FDHet respectively. To evaluate the real PCS_E, since all random components of $\boldsymbol{X}$ are i.i.d. Unif$[0,1]$ random variables, 10^5 samples of $\boldsymbol{X}$ are uniformly drawn and the proportion of correct selections is calculated. Then the proportions over 10^4 macro replications are averaged as the real PCS_E. To evaluate the real PCS_min, only the selection result on the "worst point" matters. So the proportion of correct selections on the worst point over 10^4 macro replications is calculated as the real PCS_min. We also report the averaged total sample size in each case using each procedure. The results are shown in Tables 1 and 2.

Overall, the results show that the Procedure FDHom can deliver the designed PCS_E or PCS_min in all cases where the errors are homoscedastic, and the Procedure FDHet can always deliver the designed PCS_E or PCS_min in any case. But when the Procedure FDHom is misused in the heteroscedastic errors case (Case (6)), the designed PCS is not achieved. On the other hand, when the Procedure FDHet is misused in the homoscedastic errors case (all cases except Case (6)), it will be over-conservative. It is reflected in the larger sample size (which is determined by the larger h) and larger real PCS than those in the Procedure FDHom. Even for the Procedure FDHom in the homoscedastic errors case, the real PCS is also larger than the designed PCS, especially when the mean configuration is not GSC (Case (3)). It means the procedure requires more samples than necessary. This conservativeness is a common drawback of almost all the traditional R&S procedures developed under the IZ formulation.

Table 1: Results when PCS_E is designed to be 95%.

	Procedure FDHom				Procedure FDHet			
Case	h_E	Sample	PCS_E	PCS_min	h_E	Sample	PCS_E	PCS_min
(0) Benchmark	3.423	46865	0.9610	0.7439	4.034	65138	0.9801	0.8080
(1) $k=2$	2.363	8947	0.9501	0.8084	2.781	12380	0.9702	0.8517
(2) $k=8$	3.822	93542	0.9650	0.7246	4.510	130200	0.9842	0.8052
(3) Non-GSC	3.423	46865	0.9987	0.9410	4.034	65138	0.9994	0.9615
(4) IV	3.423	52698	0.9618	0.7549	4.034	73265	0.9807	0.8147
(5) DV	3.423	52720	0.9614	0.7501	4.034	73246	0.9806	0.8114
(6) Het	3.423	58626	0.9232	0.6336	4.034	81555	0.9846	0.8591
(7) $d=1$	4.612	21288	0.9593	0.7941	4.924	24266	0.9662	0.8223
(8) $d=5$	2.141	73428	0.9656	0.7446	2.710	117630	0.9895	0.8379

Table 2: Results when PCS_min is designed to be 95%.

	Procedure FDHom				Procedure FDHet			
Case	h_min	Sample	PCS_E	PCS_min	h_min	Sample	PCS_E	PCS_min
(0) Benchmark	5.927	140540	0.9989	0.9594	6.990	195340	0.9997	0.9825
(1) $k=2$	4.362	30447	0.9958	0.9466	5.132	42164	0.9987	0.9701
(2) $k=8$	6.481	268750	0.9993	0.9642	7.651	374720	0.9999	0.9849
(3) Non-GSC	5.927	140540	1.0000	0.9958	6.990	195340	1.0000	0.9981
(4) IV	5.927	158140	0.9989	0.9574	6.990	219870	0.9998	0.9862
(5) DV	5.927	158100	0.9990	0.9617	6.990	219740	0.9998	0.9826
(6) Het	5.927	175700	0.9952	0.8999	6.990	244490	0.9999	0.9899
(7) $d=1$	7.155	51161	0.9954	0.9600	7.648	58493	0.9971	0.9708
(8) $d=5$	3.792	230220	0.9994	0.9539	4.804	369310	1.0000	0.9907

When the number of alternative increases, we can see the sample size (even when averaged to each alternative) gets larger, which is caused by the larger value of h. For the effect of dimensionality, nothing much can be concluded because the number of design points is also increasing with the dimensionality. But at least we can find that the real PCS increases not too fast when the dimensionality increases, which tells that the two procedures may still work properly for moderately large dimensions of covariates. Another observation from comparing Tables 1 and 2 is that, when the PCS_min is designed, the sample size is about three times of what is required if the PCS_E is designed. Besides, when the PCS_min is designed, the real PCS_E is almost 1 for every case. This suggests that in practice the PCS_min criterion may be too conservative to apply unless when it is indeed necessary.

7 CONCLUSIONS

In this paper we introduce and formulate the R&S problem with covariates by assuming a linear model. After defining the unconditional PCS, we design Procedure FDHom and Procedure FDHet, which are statistical valid under the homoscedastic errors and heteroscedastic errors respectively. The LFC for the

R&S with covariates is also investigated. The numerical experiments demonstrate the validity of the two designed selection procedures and also show their differences.

ACKNOWLEDGMENTS

We gratefully acknowledge the support from the Hong Kong Research Grants Council under grants GRF 16203214 and GRF 11270116.

REFERENCES

Arora, N., X. Dreze, A. Ghose, J. D. Hess, R. Iyengar, B. Jing, Y. Joshi, V. Kumar, N. Lurie, S. Neslin et al. 2008. "Putting One-to-one Marketing to Work: Personalization, Customization, and Choice". *Marketing Letters* 19 (3-4): 305.

Branke, J., S. E. Chick, and C. Schmidt. 2007. "Selecting A Selection Procedure". *Manag. Sci.* 53 (12): 1916–1932.

Choi, S. E., K. E. Perzan, A. C. Tramontano, C. Y. Kong, and C. Hur. 2014. "Statins and Aspirin for Chemoprevention in Barrett's Esophagus: Results of a Cost-effectiveness Analysis". *Cancer Prevention Research* 7 (3): 341–350.

Hong, L. J. 2006. "Fully Sequential Indifference-zone Selection Procedures with Variance-dependent Sampling". *Naval Research Logistics (NRL)* 53 (5): 464–476.

Kim, E. S., R. S. Herbst, I. I. Wistuba, J. J. Lee, G. R. Blumenschein, A. Tsao, D. J. Stewart, M. E. Hicks, J. Erasmus, S. Gupta et al. 2011. "The BATTLE Trial: Personalizing Therapy for Lung Cancer". *Cancer discovery* 1 (1): 44–53.

Kim, S.-H., and B. L. Nelson. 2001. "A Fully Sequential Procedure for Indifference-zone Selection in Simulation". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 11 (3): 251–273.

Kim, S.-H., and B. L. Nelson. 2006. "Selecting the Best System". In *Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. L. Nelson, Volume 13, Chapter 17, 501–534. Elsevier.

Negoescu, D. M., P. I. Frazier, and W. B. Powell. 2011. "The Knowledge-gradient Algorithm for Sequencing Experiments in Drug Discovery". *INFORMS J. Comput.* 23 (3): 346–363.

Nelson, B. L. 2016. "Some Tactical Problems in Digital Simulation for the Next 10 Years". *Journal of Simulation* 10 (1): 2–11.

Rinott, Y. 1978. "On Two-stage Selection Procedures and Related Probability-inequalities". *Communications in Statistics - Theory and methods* 7 (8): 799–811.

Slepian, D. 1962. "The One-Sided Barrier Problem for Gaussian Noise". *Bell System Technical Journal* 41 (2): 463–501.

Stein, C. 1945. "A Two-sample Test for a Linear Hypothesis Whose Power is Independent of the Variance". *The Annals of Mathematical Statistics* 16 (3): 243–258.

Yap, T. A., C. P. Carden, and S. B. Kaye. 2009. "Beyond Chemotherapy: Targeted Therapies in Ovarian Cancer". *Nature Reviews Cancer* 9 (3): 167–181.

AUTHOR BIOGRAPHIES

HAIHUI SHEN is a Ph.D. student in the Department of Management Sciences in the College of Business at City University of Hong Kong. His research interests include ranking & selection, discrete optimization via simulation and kriging. His email address is haihui.shen@my.cityu.edu.hk.

L. JEFF HONG is a chair professor in the Department of Economics and Finance, and the Department of Management Sciences, in the College of Business at City University of Hong Kong. His research interests include Monte Carlo methods, financial engineering, and stochastic optimization. He is currently an associate editor for *Operations Research*, *Naval Research Logistics* and *ACM Transactions on Modeling*

and Computer Simulation. His email address is jeffhong@cityu.edu.hk.

XIAOWEI ZHANG is an assistant professor in the Department of Industrial Engineering and Logistics Management at the Hong Kong University of Science and Technology. He received his Ph.D. in Operations Research from Stanford University in 2011. He is a member of INFORMS and his research interests include simulation optimization, input uncertainty, rare-event simulation, and financial engineering. His email address is xiaoweiz@ust.hk.