Pan, Santner:

Screening Procedures to Identify Robust Product or Process Designs Using Fractional Factorial Experiments

Sonderforschungsbereich 386, Paper 92 (1997)

Online unter: http://epub.ub.uni-muenchen.de/

Projektpartner

Screening Procedures to Identify Robust Product or Process Designs Using Fractional Factorial Experiments

Guohua Pan
Department of Mathematical Sciences
Oakland University
Rochester, MI 48309

Thomas J. Santner[1]
Department of Statistics
Ohio State University
Columbus, OH, 43210

Abstract

In many quality improvement experiments, there are one or more "control" factors that can be modified to determine a final product design or manufacturing process, and one or more "environmental" (or " noise") factors that vary under field or manufacturing conditions. In many applications, the product design or process design is considered seriously flawed if its performance is poor for any level of the environmental factor. For example, if a particular prosthetic heart valve design has poor fluid flow characteristics for certain flow rates, then a manufacturer will not want to put this design into production. Thus this paper considers cases when it is appropriate to measure a product's quality to be its *worst* performance over the levels of the environmental factor. We consider the frequently occurring case of combined-array experiments and extend the subset selection methodology of Gupta (1956, 1965) to provide statistical screening procedures to identify product designs that maximize the worst case performance of the design over the environmental conditions for such experiments. A case study is provided to illustrate the proposed procedures.

Keywords and phrases: Combined-array, Inner array, Minimax approach, Outer array, Product-array, Quality improvement, Response model, Screening, Simulation, Subset selection, Variance reduction.

[1]Written while visiting Sonderforschungsbereich 386, Ludwig-Maximilians-Universität München, 80539 München

1 Introduction

Taguchi (1986) emphasizes two types of factors that effect product quality in his pioneering work on product and process improvement. The first are "control factors" which are those factors that can be (easily) manipulated by the manufacturer; sometimes these are called "manufacturing" or "engineering" factors. The second are "noise factors" which are those variables that represent either different environmental conditions that affect the performance of a product in the field *or* (uncontrollable) variability in component parts or raw materials that affect the performance of an end-product. By identifying conditions of the control factors under which the mean product quality is (relatively) independent of the noise factors, the product or process can be made "robust."

For experiments to determine such conditions, Taguchi advocates using statistical designs that are products of highly fractionated orthogonal arrays in the control and noise factors. A number of other authors (Shoemaker, Tsui, and Wu 1991; Nair et al. 1992; Myers, Khuri, and Vining 1992, for example) have proposed alternatives to the Taguchi methodology, particularly the use of combined-arrays in the control and noise factors. At the expense of confounding higher-order interactions, carefully chosen combined-arrays allow the experimenter to determine interactions *among* the control factors and interactions *among* the noise factors, as well as the critical control factor $\times$ noise factor interactions that allow one to minimize the effect of noise factors in product quality. Thus the basic viewpoint that Taguchi advocates has been applied widely and with many successes (Taguchi and Phadke 1984).

This paper considers applications where it is appropriate to use the worst possible performance of a product under the different environments as a performance or quality index. This criterion is natural in situations where a low response at *any* level of the noise factor can have potentially serious consequence. Seat belts or heart valves that fail catastrophically under rare, though non-negligible, sets of operating conditions must be identified early in the product design cycle. We extend the subset selection methodology introduced in Gupta (1956, 1965) for balanced one-way layouts, to selection of a subset containing the *control* factor combination that maximizes the worst performance over the levels of the *noise* variables. Such procedures are proposed for data collected using (fractional) combined-array experiments. Pan and Santner (1996) consider this criterion for the case of complete experiments conducted under a variety of randomization restrictions; Santner and Pan (1997) present a case study involving a 2^{5-1} combined-array experiment with three control factors and two noise factors. This paper develops procedures for *arbitrary* combined-array experiments, including Taguchi's cross-arrays as special cases. Bechhofer, Santner, and Goldsman (1995) give an overview of selection and screening methodology and present procedures to accomplish other important experimental goals.

Section 2 presents the basic model, goal, and assumptions regarding the associated combined-array experiment. Section 3 introduces the proposed subset selection procedure for a class of general models; it presents a theorem that gives the least favorable configuration and associated value of the probability of correct selection. In particular, the critical value required to implement the procedure is identified. Section 4 analyzes an integrated circuit example using the proposed method. Some generalizations and caveats are presented in the final section.

2 The Model and Confidence Requirement

We suppose that an experiment has been conducted in which there are $p+q$ control factors and $r+s$ noise factors. We assume a known model holds for the matrix of true mean responses in which p of the control factors interact with r of the noise factors, q of the control factors have *no interactions* with noise factors, and s noise factors have *no interactions* with control factors. Of special importance is the case when all the factors are at two levels, but nothing in the development below requires this assumption. We introduce the following notation to distinguish these types of control and noise factors.

Notation	Interpretation
$\mathrm{C}_1^1,\ldots,\mathrm{C}_p^1$	Control Factors that *interact* with Noise Factors
$\mathrm{C}_1^2,\ldots,\mathrm{C}_q^2$	Control Factors that *do not interact* with Noise Factors
$\mathrm{N}_1^1,\ldots,\mathrm{N}_r^1$	Noise Factors that *interact* with Control Factors
$\mathrm{N}_1^2,\ldots,\mathrm{N}_s^2$	Noise Factors that *do not interact* with Control Factors

Let $\boldsymbol{i}^1=(i_1^1,\ldots,i_p^1)$ denote the $1\times p$ vector of indices for the levels of the C^1–type control factors, $\boldsymbol{i}^2=(i_2^1,\ldots,i_q^2)$ denote the $1\times q$ vector of indices for the levels of the C^2–type control factors, $\boldsymbol{j}^1=(j_1^1,\ldots,j_r^1)$ be the $1\times r$ vector of indices for the levels of the N^1–type noise factors, and $\boldsymbol{j}^2=(j_1^2,\ldots,j_s^2)$ denote the $1\times s$ vector of indices for the levels of the N^2–type noise factors. Suppose that $\boldsymbol{i}^1\in\mathcal{I}^1$, $\boldsymbol{i}^2\in\mathcal{I}^2$, $\boldsymbol{j}^1\in\mathcal{J}^1$, and $\boldsymbol{j}^2\in\mathcal{J}^2$ for a (hypothetical) complete factorial experiment in these factors; thus $\mathcal{I}^1$ is a cross product of the p index sets corresponding to the p C^1–type control factors and similarly for $\mathcal{I}^2$, $\mathcal{J}^1$, and $\mathcal{J}^2$. For example, in the case of an experiment with each factor at two levels, we have $\mathcal{I}^1=\{0,1\}^p$, $\mathcal{I}^2=\{0,1\}^q$, $\mathcal{J}^1=\{0,1\}^r$, and $\mathcal{J}^2=\{0,1\}^s$.

Let $\boldsymbol{i}=(\boldsymbol{i}^1,\boldsymbol{i}^2)$ and $\boldsymbol{j}=(\boldsymbol{j}^1,\boldsymbol{j}^2)$ denote the index vectors for the entire set of control and noise factors, respectively. Then $\mu_{\boldsymbol{i},\boldsymbol{j}}$ denotes the mean response when the control factors are at level $\boldsymbol{i}$ and the noise factors are at level $\boldsymbol{j}$; here $(\boldsymbol{i},\boldsymbol{j})\in\mathcal{I}\times\mathcal{J}$ where $\mathcal{I}$ is the cross product of $\mathcal{I}^1$ and $\mathcal{I}^2$ and $\mathcal{J}$ is the cross product of $\mathcal{J}^1$ and $\mathcal{J}^2$. We adopt the convention that $\boldsymbol{\mu}=(\mu_{\boldsymbol{i},\boldsymbol{j}})$ is arranged as an $|\mathcal{I}|\times|\mathcal{J}|$ matrix where $|\cdot|$ denotes the number of elements in a set and $\boldsymbol{i}$ and $\boldsymbol{j}$ are arranged *lexicographically*. Thus each row of $\boldsymbol{\mu}$ corresponds to a single setting of the $p+q$ control factors and each column corresponds to a single setting of the noise factors. The lexicographic ordering will be used in the Kronecker product formulas for the mean response in terms the model components (2.4). This notation is illustrated in the following example.

Example 1: Box and Jones (1992) discuss a taste-testing experiment that is typical of those used in the food industry to evaluate cake recipes. The experiment involves five factors, each at two levels. Three of the factors (S=shortening, E=egg powder and F=flour) are control factors because they can be varied by the manufacturer. The remaining two factors (T=baking temperature and Z=baking time) are noise factors because they are determined by the consumer; temperature controls in ovens can be considerably biased and consumers both under and overbake pre-packaged mixes. Throughout this and other examples involving 2^{n-p} experiments, the subscript denoting the level of a factor is taken to be zero (unity) when the factor is at its low (high) level.

Santner and Pan (1996) study subset selection procedure for this set-up under the empirically derived assumption that the model

$$\mu_{ijk\ell m} = m_0 + S_i + F_j + E_k + T_\ell + Z_m + (ST)_{i\ell}. \tag{2.1}$$

holds for all $(ijk\ell m)$. Here the terms S_i, F_j, E_k, T_ℓ, and Z_m are the shortening, flour, egg, temperature, and time main effects, respectively. For model (2.1) we have $(p =)$ 1 control factor, S, that interacts with $(r =)$ 1 noise factor, T; there are $(q =)$ 2 control factors, F and E, that do not interact with any noise factors, and $(s =)$ 1 noise variable, Z, that does not interact with any control variable.

The matrix $\boldsymbol{\mu}$ of means for this case is $8 \times 4 = 2^{1+2} \times 2^{1+1}$ with entries ordered as follows

$$\boldsymbol{\mu} = \begin{array}{c} \\ \left[\begin{array}{cccc} \mu_{00000} & \mu_{00001} & \mu_{00010} & \mu_{00011} \\ \mu_{00100} & \mu_{00101} & \mu_{00110} & \mu_{00111} \\ \mu_{01000} & \mu_{01001} & \mu_{01010} & \mu_{01011} \\ \mu_{01100} & \mu_{01101} & \mu_{01110} & \mu_{01111} \\ \mu_{10000} & \mu_{10001} & \mu_{10010} & \mu_{10011} \\ \mu_{10100} & \mu_{10101} & \mu_{10110} & \mu_{10111} \\ \mu_{11000} & \mu_{11001} & \mu_{11010} & \mu_{11011} \\ \mu_{11100} & \mu_{11101} & \mu_{11110} & \mu_{11111} \end{array}\right] \\ \\ (T,Z) \end{array} \begin{array}{c} (S,F,E) \\ (0,0,0) \\ (0,0,1) \\ (0,1,0) \\ (0,1,1) \\ (1,0,0) \\ (1,0,1) \\ (1,1,0) \\ (1,1,1) \\ \\ (0,0)\ \ (0,1)\ \ (1,0)\ \ (1,1) \end{array}$$

Notice that lexicographic order produces a row order in which the first four rows are at the low level of (the interacting factor) S, and whose second four rows are at the high level of S. Similarly, the columns are ordered so that the first two columns are at the low level of (the interacting factor) T, and the last two columns at the high level of T. □

For each combination of control factors we are interested in the worst mean performance of the response over the levels of the noise factors. Formally,

$$\xi_{\boldsymbol{i}} = \min_{\boldsymbol{j}} \boldsymbol{\mu}_{\boldsymbol{i},\boldsymbol{j}}$$

gives the worst performance for the product/process design defined by control factor combination $\boldsymbol{i}$. We denote the ordered $\xi_{\boldsymbol{i}}$ corresponding to the $|\mathcal{I}|$ product designs by

$$\xi_{[1]} \leq \cdots \leq \xi_{[|\mathcal{I}|]}. \tag{2.2}$$

Our goal is to find a screening procedure that selects a subset of the control factor combinations so as to contain the product or process design associated with $\xi_{[|\mathcal{I}|]}$. We desire our procedure to achieve this goal subject to the following performance requirement.

Confidence Requirement: Given α with $0 < \alpha < 1$, we desire that

$$P_{\boldsymbol{\mu}} \{CS\} \geq 1 - \alpha \tag{2.3}$$

for all $\boldsymbol{\mu}$ satisfying Model (2.4) where CS denotes the event that the selected subset contains the control factor combination associated with $\xi_{[|\mathcal{I}|]}$.

Throughout, we assume the model

$$\mu_{\boldsymbol{i},\boldsymbol{j}} = m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) + m_C(\boldsymbol{i}^2) + m_N(\boldsymbol{j}^2) \tag{2.4}$$

holds for each $(\boldsymbol{i}, \boldsymbol{j}) \in \mathcal{I} \times \mathcal{J}$ where

$$\begin{aligned} m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) &= m_0 + \sum_{Q_1^{CN}} (C^1_{i_1^\star} C^1_{i_2^\star} \cdots C^1_{i_{p^*}^\star} N^1_{j_1^\star} N^1_{j_2^\star} \cdots N^1_{j_{r^*}^\star})_{\boldsymbol{i}^1,\boldsymbol{j}^1}, \\ m_C(\boldsymbol{i}^2) &= \sum_{Q_2^C} (C^2_{i_1^\star} C^2_{i_2^\star} \cdots C^2_{i_{q^*}^\star})_{\boldsymbol{i}^2}, \\ m_N(\boldsymbol{j}^2) &= \sum_{Q_2^N} (N^2_{j_1^\star} N^2_{j_2^\star} \cdots N^2_{j_{s^*}^\star})_{\boldsymbol{j}^2}. \end{aligned} \tag{2.5}$$

Notice, that we group the overall mean, m_0, with the control $\times$ noise interaction terms. Here Q_2^C, Q_2^N, and Q_1^{CN} are nonempty sets that identify the main effects and interactions among the C^2–type control factors, the main effects and interactions among the N^2–type noise factors, and the main effects and interactions among the C^1–type control factors and N^1–type noise factors, respectively. For notational simplicity we use, for example, the subscript $\boldsymbol{i}^2$ in the term $(C^2_{i_1^\star} C^2_{i_2^\star} \cdots C^2_{i_{q^*}^\star})_{\boldsymbol{i}^2}$ with the understanding the this function depends only on the q^* components $1 \le i_1^\star < i_2^\star < i_{q^*}^\star \le q$ of $\boldsymbol{i}^2$. Each of $m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1)$, $m_C(\boldsymbol{i}^2)$, and $m_N(\boldsymbol{j}^2)$ is a linear combination of the elements of $\boldsymbol{\mu}$; we do not assume that these linear combinations are orthonormal.

We assume that the observations come from a combined-array experiment with observations

$$Y_{\boldsymbol{i},\boldsymbol{j}} = \mu_{\boldsymbol{i},\boldsymbol{j}} + \epsilon_{\boldsymbol{i},\boldsymbol{j}} \quad (\boldsymbol{i}, \boldsymbol{j}) \in \mathcal{D} \tag{2.6}$$

for each $(\boldsymbol{i}, \boldsymbol{j}) \in \mathcal{D} \subseteq \mathcal{I} \times \mathcal{J}$ where the $|\mathcal{D}|$ measurement errors $\epsilon_{\boldsymbol{i},\boldsymbol{j}}$ are independent $N(0, \sigma_\epsilon^2)$ variables and *all main effects and interactions* in the mean model (2.4) are confounded only with model terms that are zero, i.e., terms not in the model (2.4).

Example 1 (Continued): The quality of the recipe $(S, F, E) = (i, j, k)$ is specified by the row minimum

$$\xi_{ijk} = \min\{\mu_{ijk00}, \mu_{ijk01}, \mu_{ijk10}, \mu_{ijk11}\}$$

where $\mu_{ijk\ell m}$ satisfies (2.1). For this model we have $m_C(j, k) = F_j + E_k$, $m_N(m) = Z_m$, and $m_{CN}(i, \ell) = m_0 + S_i + T_\ell + (ST)_{i\ell}$. Our goal is to identify a subset of recipes that contains the best (S, F, E) combination, i.e., the recipe associated with $\xi_{[8]} = \max_{i,j,k} \xi_{ijk}$.

The data used by Santner and Pan (1997) in conjunction with model (2.1) came from a 2^{5-1} experiment with defining contrast $I = SFETZ$; thus the main effects are confounded with 4-way interactions and the 2-way interactions are confounded with 3-way interactions

all of which are assumed to be zero. The observed data have the following structure:

$$\begin{bmatrix} & Y_{00001} & Y_{00010} & \\ Y_{00100} & & & Y_{00111} \\ Y_{01000} & & & Y_{01011} \\ & Y_{01101} & Y_{01110} & \\ Y_{10000} & & & Y_{10011} \\ & Y_{10101} & Y_{10110} & \\ & Y_{11001} & Y_{11010} & \\ Y_{11100} & & & Y_{11111} \end{bmatrix}$$

$$= \begin{bmatrix} & \mu_{00001} & \mu_{00010} & \\ \mu_{00100} & & & \mu_{00111} \\ \mu_{01000} & & & \mu_{01011} \\ & \mu_{01101} & \mu_{01110} & \\ \mu_{10000} & & & \mu_{10011} \\ & \mu_{10101} & \mu_{10110} & \\ & \mu_{11001} & \mu_{11010} & \\ \mu_{11100} & & & \mu_{11111} \end{bmatrix} + \begin{bmatrix} & \epsilon_{00001} & \epsilon_{00010} & \\ \epsilon_{00100} & & & \epsilon_{00111} \\ \epsilon_{01000} & & & \epsilon_{01011} \\ & \epsilon_{011101} & \epsilon_{01110} & \\ \epsilon_{10000} & & & \epsilon_{10011} \\ & \epsilon_{101101} & \epsilon_{101110} & \\ & \epsilon_{11001} & \epsilon_{11010} & \\ \epsilon_{11100} & & & \epsilon_{11111} \end{bmatrix}.$$

We will denote the three matrices above by $\boldsymbol{Y}_{\mathcal{D}}$, $\boldsymbol{\mu}_{\mathcal{D}}$ and $\boldsymbol{\epsilon}_{\mathcal{D}}$, respectively. The matrix equation above is then $\boldsymbol{Y}_{\mathcal{D}} = \boldsymbol{\mu}_{\mathcal{D}} + \boldsymbol{\epsilon}_{\mathcal{D}}$. This notation emphasizes that observations are only made at the design points in $\mathcal{D}$. The 16 observations collected in the experiment provide a 9 $(= 16 - 7)$ degree of freedom chi-square estimator of σ_ϵ^2. $\square$

In general, we will use $\boldsymbol{\mu}_{\mathcal{D}}$ to denote the $|\mathcal{I}^1| \times |\mathcal{J}^1|$ matrix obtained by deleting all the entries in $\boldsymbol{\mu}$ with indices not belonging to $\mathcal{D}$. We use $\boldsymbol{Y}_{\mathcal{D}}$ and $\boldsymbol{\epsilon}_{\mathcal{D}}$ to denote the conformably ordered matrix of observations, $Y_{\boldsymbol{i},\boldsymbol{j}}$, and errors, $\epsilon_{\boldsymbol{i},\boldsymbol{j}}$, respectively. Lastly, we assume that there is an estimator S^2 of σ_ϵ^2, for which $\nu S^2 \sim \chi_\nu^2$ and that S^2 is independent of $\boldsymbol{Y}_{\mathcal{D}}$. Ordinarily, such a chi-square estimator would be available when the number of observations, $|\mathcal{D}|$, is larger than the number of parameters estimated in the model (2.4).

For simplicity, we assume above that there is one replicate of the design $\mathcal{D}$. This is the most common case in quality improvement experimentation. However, the theory and methods developed in this paper extend straightforwardly to situations where replicates of the design $\mathcal{D}$ are observed. The case study in Section 4 illustrates such a situation.

3 A Screening Procedure

The procedure we use is based on the ordinary least square (OLS) estimator of $\boldsymbol{\mu}$ from Model (2.4)–(2.6). For each $(\boldsymbol{i},\boldsymbol{j}) \in \mathcal{I} \times \mathcal{J}$, let $\widehat{\mu}_{\boldsymbol{i},\boldsymbol{j}}$ denote the OLS estimator of $\mu_{\boldsymbol{i},\boldsymbol{j}}$ based on the data from the fractional factorial design in (2.6). We estimate $\xi_{\boldsymbol{i}}$ by

$$\widehat{\xi}_{\boldsymbol{i}} = \min_{\boldsymbol{j}} \{\widehat{\mu}_{\boldsymbol{i},\boldsymbol{j}}\} \quad (\boldsymbol{i} \in \mathcal{I}).$$

Let

$$\widehat{\xi}_{[1]} \leq \cdots \leq \widehat{\xi}_{[|\mathcal{I}|]}$$

denote the ordered $\widehat{\xi}_{\boldsymbol{i}}$. We propose the following procedure to select a subset of the levels of the control factor.

Procedure $\mathcal{G}$: Select control factor combination $\boldsymbol{i}$ if and only if

$$\widehat{\xi}_{\boldsymbol{i}} \geq \widehat{\xi}_{[|\mathcal{I}|]} - h\,S$$

where h is chosen as in Theorems 3.1, 3.2 and 3.3. Let $\boldsymbol{m}_{CN} = [m_{CN}(\boldsymbol{i},\boldsymbol{j})]_{\boldsymbol{i}\in\mathcal{I}^1,\boldsymbol{j}\in\mathcal{J}^1}$ be the $|\mathcal{I}^1| \times |\mathcal{J}^1|$ matrix of $m_{CN}(\boldsymbol{i},\boldsymbol{j})$'s, where the indices $\boldsymbol{i}^1$ and $\boldsymbol{j}^1$ are ordered lexicographically. Let $\boldsymbol{V}_{CN} = \{\boldsymbol{m}_{CN}\}$ be the linear space consisting of all $\boldsymbol{m}_{CN}$ satisfying the model (2.5). The next theorem describes the Least Favorable Configuration for $\mathcal{G}$ for a large class of models.

Theorem 3.1 *Assume that Model (2.4)–(2.6) is symmetric with respect to the C^1, N^1, C^2 and N^2 type factors, respectively, in that*
(1) If the term $(C^1_{i^\star_1}C^1_{i^\star_2}\cdots C^1_{i^\star_{p^*}}N^1_{j^\star_1}N^1_{j^\star_2}\cdots N^1_{j^\star_{r^*}})_{\boldsymbol{i}^1,\boldsymbol{j}^1}$ is in $m_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1)$, then all other terms involving p^* of the C^1 control factors and r^* of the N^1 noise factors are also in $m_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1)$;
(2) If the term $(C^2_{i^\star_1}C^2_{i^\star_2}\cdots C^2_{i^\star_{q^*}})_{\boldsymbol{i}^2}$ is in $m_C(\boldsymbol{i}^2)$, then all other terms involving q^* of the C^2 control factors are also in $m_C(\boldsymbol{i}^2)$;
(3) if the term $(N^2_{j^\star_1}N^2_{j^\star_2}\cdots N^2_{j^\star_{s^*}})_{\boldsymbol{j}^2}$ is in $m_N(\boldsymbol{j}^2)$, then all other terms involving s^* of the N^2 noise factors are also in the model.
In addition, suppose that there exists a sequence of points $\{\boldsymbol{v}_k\}_k$ in $\boldsymbol{V}_{CN}$ for which

$$\lim_{k\to\infty}\boldsymbol{v}_k = \boldsymbol{v}_\infty \equiv \begin{bmatrix} 0 & +\infty & \cdots & +\infty \\ \vdots & \vdots & \vdots & \vdots \\ 0 & +\infty & \cdots & +\infty \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{3.7}$$

where, of course, $\boldsymbol{v}_\infty$ is $|\mathcal{I}^1| \times |\mathcal{J}^1|$. Define $\boldsymbol{\mu}_k = \boldsymbol{v}_k \otimes J_{|\mathcal{I}^2|\times|\mathcal{J}^2|}$. Then,

$$\inf_{\boldsymbol{\mu}} P_{\boldsymbol{\mu}}\{CS\} = \lim_{k\to\infty} P_{\boldsymbol{\mu}_k}\{CS\} = \lim_{k\to\infty} P_{\boldsymbol{\mu}_k}\left\{\widehat{\xi}_{\boldsymbol{I}^*} \geq \max_{\boldsymbol{i}}\widehat{\xi}_{\boldsymbol{i}} - hS\right\} \tag{3.8}$$

where $\boldsymbol{I}^*$ is the control factor combination corresponding to the last row of $\boldsymbol{\mu}_k$, i.e., each control factor is at its highest level.

Theorem 3.1 is proved in the Appendix. Heuristically, it states that the least favorable configuration of means is

$$\boldsymbol{\mu}_{LFC} = \begin{bmatrix} 0 & +\infty & \cdots & +\infty \\ \vdots & \vdots & \vdots & \vdots \\ 0 & +\infty & \cdots & +\infty \\ 0 & 0 & \cdots & 0 \end{bmatrix} \otimes J_{|\mathcal{I}^2|\times|\mathcal{J}^2|}$$

where the control combination associated with the last row, say $\boldsymbol{I}^*$, is designated as the best. To compute h, it suffices to solve the implicit equation

$$P_{\boldsymbol{\mu}_{LFC}}\left\{\widehat{\xi}_{\boldsymbol{I}^*} \geq \max_{\boldsymbol{i}}\widehat{\xi}_{\boldsymbol{i}} - hS\right\} = 1-\alpha.$$

In practice, simulation is the simplest method to obtain h. In Model (2.4)–(2.6) take model parameters $\sigma_\epsilon^2 = 1$ and $\boldsymbol{\mu} = \boldsymbol{v}_k^\star \otimes J_{|\mathcal{I}^2|\times|\mathcal{J}^2|}$ where $\boldsymbol{v}_k^\star$ can be any $\boldsymbol{v}_k^\star$ whose nonzero entries are large relative to zero, for example, nonzero entries are no smaller than 100. Then h is obtained by simulating the $100 \times (1-\alpha)$ percentile of the random variable $T = (\max_{\boldsymbol{i}} \widehat{\xi}_{\boldsymbol{i}} - \widehat{\xi}_{\boldsymbol{I}^*})/S$. A SAS macro to calculate h using simulation is under development by one of the authors (G. Pan).

Example 1 (Continued): Recall that

$$\mu_{ijk\ell m} = m_0 + S_i + F_j + E_k + T_\ell + Z_m + (ST)_{i\ell},$$

for the taste test experiment with

$$m_{CN}(i\ell) = m_0 + S_i + T_\ell + (ST)_{i\ell}.$$

The symmetry conditions (1)–(3) are automatic; also $V_{CN} = \mathbb{R}^{2\times 2}$ so that condition (3.7) also holds $\left(\text{take } v_k = \begin{bmatrix} 0 & k \\ 0 & 0 \end{bmatrix}\right)$. To compute h, one can choose $k = 100$, to effectively be ∞, and take

$$\boldsymbol{\mu}_D = \begin{bmatrix} & 0 & 100 & \\ 0 & & & 100 \\ 0 & & & 100 \\ & 0 & 100 & \\ 0 & & & 0 \\ & 0 & 0 & \\ & 0 & 0 & \\ 0 & & & 0 \end{bmatrix}.$$

Now, repeatedly generate data of the form:

$$\boldsymbol{Y}_D^* = \begin{bmatrix} & Z_1 & 100+Z_2 & \\ Z_3 & & & 100+Z_4 \\ Z_5 & & & 100+Z_6 \\ & Z_7 & 100+Z_8 & \\ Z_9 & & & Z_{10} \\ & Z_{11} & Z_{12} & \\ & Z_{13} & Z_{14} & \\ Z_{15} & & & Z_{16} \end{bmatrix},$$

where $Z_1, \cdots, Z_{16}$ are *iid* standard normal random variables and also generate a chi-square random variable V with degrees of freedom $v = 9$ and set $S = \sqrt{V/v}$. Compute $T = (\max_{\boldsymbol{i}} \widehat{\xi}_{\boldsymbol{i}} - \widehat{\xi}_{111})/S$ based on the generated data. The sample quantile based on the draws from this distribution is an estimate of h. □

Condition (3.7) is always satisfied whenever $V_{CN} = \mathbb{R}^{\mathcal{I}^1\times\mathcal{J}^1}$, but it can also be satisfied in many other cases. Example 2 illustrates such a situation and describes a general method to check condition (3.7).

Example 2: Consider a six factor example with each factor at two levels. Suppose that there are 3 control factors and 3 noise factors classified so that there are $(p, r) = (1, 2)$ interacting control and noise factors, $q = 2$ non-interacting control factors, and $s = 1$ noise non-interacting factor. The notation identifying these factors is listed in the following table:

Index	Name	Type
i	C_1^1	Interacting
j, k	C_1^2, C_2^2	Non-interacting
ℓ, m	N_1^1, N_2^1	Interacting
n	N_1^2	Non-interacting

Suppose

$$\mu_{ijk\ell mn} = m_{CN}(i\ell m) + m_C(jk) + m_N(n) \tag{3.9}$$

where

$$m_{CN}(i\ell m) = m_0 + (C_1^1)_i + (N_1^1)_\ell + (N_2^1)_m + (C_1^1 N_1^1)_{i\ell} + (C_1^1 N_2^1)_{im},$$

$m_C(jk) = (C_1^2)_j + (C_2^2)_k + (C_1^2 C_2^2)_{jk}$, and $m_N(n) = (N_1^2)_n$. Notice that the symmetry conditions (1)–(2) are immediate for this model. However, V_{CN} is six dimensional and does not span the eight dimensional space $\mathbb{R}^{2\times 4}$. However, condition (3.7) still holds for this model.

Condition (3.7) can be checked using the ANOVA model (3.9) but it is more convenient to use an equivalent regression model defined by indicator variables in which

$$x_s^t = \begin{cases} 1, & \text{if } C_s^t \text{ is at its low level} \\ 0, & \text{if } C_s^t \text{ is at its high level} \end{cases} \text{ and } z_s^t = \begin{cases} 1, & \text{if } N_s^t \text{ is at its high level} \\ 0, & \text{if } N_s^t \text{ is at its low level} \end{cases}$$

The indicator variables x_s^t and z_s^t are defined slightly differently than usual to simplify some of the terms in $\boldsymbol{v}$. The model (3.9) is equivalent to the following

$$\mu_{ijk\ell mn} = \beta_0 + \beta_1^1 x_1^1 + \gamma_1^1 z_1^1 + \gamma_2^1 z_2^1 + \delta_{11} x_1^1 z_1^1 + \delta_{12} x_1^1 z_2^1 + \beta_1^2 x_1^2 + \beta_2^2 x_2^2 + \beta_{12}^{22} x_1^2 x_2^2 + \gamma_1^2 z_1^2$$

and $\boldsymbol{V}_{CN}$ is generated by

$$v_{i\ell m} = \beta_0 + \beta_1^1 x_1^1(i) + \gamma_1^1 z_1^1(\ell) + \gamma_2^1 z_2^1(m) + \delta_{11} x_1^1(i) z_1^1(i) + \delta_{12} x_1^1(i) z_2^1(\ell)$$

where each regression coefficient ranges over $\mathbb{R}$. Thus each element in $\boldsymbol{V}_{CN}$ has form

$$\boldsymbol{v} = \begin{bmatrix} \beta_0 + \beta_1^1 & \beta_0 + \beta_1^1 + \gamma_2^1 + \delta_{12} & \beta_0 + \beta_1^1 + \gamma_1^1 + \delta_{11} & \beta_0 + \beta_1^1 + \gamma_1^1 + \gamma_2^1 + \delta_{11} + \delta_{12} \\ \beta_0 & \beta_0 + \gamma_2^1 & \beta_0 + \gamma_1^1 & \beta_0 + \gamma_1^1 + \gamma_2^1 \end{bmatrix}.$$

With this representation it is easy to find a sequence satisfying (3.7). Make the last row of $\boldsymbol{v}$ zero by taking $\beta_0 = 0$, $\gamma_1^1 = 0$, and $\gamma_2^1 = 0$. The remaining element in the first column of $\boldsymbol{v}$, v_{000}, becomes zero if $\beta_1^1 = 0$. Then the remaining two parameters, δ_{11} and δ_{12}, can be used to increase the other entries in the first row of the $\boldsymbol{v}$ matrix to infinity, for example, by letting $\delta_{11} = \delta_{12} = K$. Then the resulting $\boldsymbol{v}$ is

$$\boldsymbol{v}_K = \begin{bmatrix} 0 & K & K & 2K \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 0 & +\infty & +\infty & +\infty \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Hence, the result in Theorem 3.1 can be used to compute h. Simulation can be used to estimate h by repeatedly computing $T = (\max_{\boldsymbol{i}} \widehat{\xi}_{\boldsymbol{i}} - \widehat{\xi}_{111})/S$ from data drawn from the particular design, $\mathcal{D}$, used in the experiment and having means taken from the appropriate positions in

$$\boldsymbol{u}_{100} = \boldsymbol{v}_{100} \otimes J_{2^2 \times 2} = \begin{bmatrix} 0 & 0 & 100 & 100 & 100 & 100 & 100 & 100 \\ 0 & 0 & 100 & 100 & 100 & 100 & 100 & 100 \\ 0 & 0 & 100 & 100 & 100 & 100 & 100 & 100 \\ 0 & 0 & 100 & 100 & 100 & 100 & 100 & 100 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Notice that the random numbers are generated only at factor combinations in the set of design points $\mathcal{D}$ involved in an experiment. □

In principle, the following more general result, Theorem 3.2, can be applied to determine h for models that either do not satisfy the symmetry conditions or (3.7) in Theorem 3.1. To state this result, some additional notations must be introduced. Let $\widehat{m}_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1)$, $\widehat{m}_C(\boldsymbol{i}^2)$, and $\widehat{m}_N(\boldsymbol{j}^2)$ be the OLS estimators of $m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1)$, $m_C(\boldsymbol{i}^2)$, and $m_N(\boldsymbol{j}^2)$, respectively. Since all the estimators are unbiased, it is easy to see that

$$\begin{aligned} \widehat{\mu}_{\boldsymbol{i},\boldsymbol{j}} &= \widehat{m}_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) + \widehat{m}_C(\boldsymbol{i}^2) + \widehat{m}_N(\boldsymbol{j}^2) \\ &= \mu_{\boldsymbol{i},\boldsymbol{j}} + \overline{\epsilon}_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) + \overline{\epsilon}_C(\boldsymbol{i}^2) + \overline{\epsilon}_N(\boldsymbol{j}^2) \end{aligned}$$

where $\overline{\epsilon}_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1)$ is the linear combination of the $\epsilon_{\boldsymbol{i},\boldsymbol{j}}$'s determined by the terms in (2.5) and whose values depend only on $(\boldsymbol{i}^1, \boldsymbol{j}^1)$. Similarly $\overline{\epsilon}_C(\boldsymbol{i}^2)$ and $\overline{\epsilon}_N(\boldsymbol{j}^2)$ are linear combinations of the $\epsilon_{\boldsymbol{i},\boldsymbol{j}}$'s whose values depend on $\boldsymbol{i}^2$ and $\boldsymbol{j}^2$, respectively.

For fixed $\boldsymbol{I}^1 \in \mathcal{I}^1$, $\boldsymbol{I}^2 \in \mathcal{I}^2$ and for each $\boldsymbol{i}^1 \in \mathcal{I}^1$ fix $\boldsymbol{j}^1(\boldsymbol{i}^1) \in \mathcal{J}^1$. Define

$$P(\boldsymbol{I}^1, \boldsymbol{I}^2, \{\boldsymbol{j}^1(\cdot)\}) = P\left\{ \begin{array}{l} \overline{\epsilon}_C(\boldsymbol{I}^2) \geq \overline{\epsilon}_C(\boldsymbol{i}^2) - hS \quad \forall\, \boldsymbol{i}^2 \in \mathcal{I}^2; \\ \overline{\epsilon}_C(\boldsymbol{I}^2) + \min_{\boldsymbol{j}^1} \left\{ \overline{\epsilon}_{CN}(\boldsymbol{I}^1, \boldsymbol{j}^1) \right\} \geq \\ \overline{\epsilon}_C(\boldsymbol{i}^2) + \overline{\epsilon}_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1(\boldsymbol{i}^1)) - hS \quad \forall\, \boldsymbol{i}^1 \neq \boldsymbol{I}^1, \boldsymbol{i}^2 \in \mathcal{I}^2 \end{array} \right\}, \tag{3.10}$$

where the $|\mathcal{D}|$ error terms $\epsilon_{\boldsymbol{i},\boldsymbol{j}}$'s are independent standard normal random variables and $\nu S^2 \sim \chi^2_\nu$ is independent of the $\epsilon_{\boldsymbol{i},\boldsymbol{j}}$'s. Then, the following Theorem can be used to determine h in Procedure $\mathcal{G}$ for *any model (2.4)–(2.6).*

Theorem 3.2 *For any $\boldsymbol{\mu}$ satisfying (2.4)–(2.6),*

$$P_{\boldsymbol{\mu}}\{CS\} \geq \min_{(\boldsymbol{I}^1, \boldsymbol{I}^2) \in \mathcal{I}^1 \times \mathcal{I}^2} \min_{\boldsymbol{j}^1(\cdot)} P(\boldsymbol{I}^1, \boldsymbol{I}^2, \{\boldsymbol{j}^1(\cdot)\}) \tag{3.11}$$

Then h can be determined by solving equation that results from setting the right hand side of Equation (3.11) equal to $1 - \alpha$. However, this method can be very tedious due to the large number of choices of $\boldsymbol{I}^1 \in \mathcal{I}^1, \boldsymbol{I}^2 \in \mathcal{I}^2$ and functions $\boldsymbol{j}^1(\cdot)$. When the model (2.4)–(2.5) satisfies the symmetry conditions (1)–(3) of Theorem 3.1, all the terms on the right hand side of equation (3.11) are equal. That is,

Theorem 3.3 *When the symmetry conditions (1)–(3) in Theorem 3.1 are satisfied,*

$$P(\boldsymbol{I}^1, \boldsymbol{I}^2, \{\boldsymbol{j}^1(\cdot)\})$$

is constant for any $\boldsymbol{I}^1 \in \mathcal{I}^1, \boldsymbol{I}^2 \in \mathcal{I}^2$ *and function* $\boldsymbol{j}^1(\cdot)$.

4 A Case Study

In this section we provide a detailed analysis of an example using the proposed methodology. The example and data are described in Myers and Montgomery (1995, pages 144, 529). Five factors in a manufacturing process for integrated circuits were investigated using a 2^{5-1} design. The notation we use to describe the five factors is

Notation	Factor
I	implant dose
Z	time
T	temperature
O	oxide thickness
F	furnace position

Each factor was used at two levels. The design and the measured response: resistivity of the wafer, are shown in the following table. In the process of manufacturing integrated circuits the temperature, T, is difficult to control and is considered a noise factor; the other four factors are regarded as control factors. The primary concern is the variability in wafer resistivity due to transmitted variability in the temperature. It is of interest to choose a combination of the control factors to maximize wafer resistivity over the different temperatures. Thus one appropriate measure of performance is the smallest resistivity of a given process design over temperature.

Run	I	Z	T	O	F = IZTO	Resistivity
1	0	0	0	0	1	15.1
2	1	0	0	0	0	20.6
3	0	0	1	0	0	68.7
4	1	0	1	0	1	101.0
5	0	1	0	0	0	32.9
6	1	1	0	0	1	46.1
7	0	1	1	0	1	87.5
8	1	1	1	0	0	119.0
9	0	0	0	1	0	11.3
10	1	0	0	1	1	19.6
11	0	0	1	1	1	62.1
12	1	0	1	1	0	103.2
13	0	1	0	1	1	27.1
14	1	1	0	1	0	40.3
15	0	1	1	1	0	87.7
16	1	1	1	1	1	128.3

Past experience with this process indicated that O and F have little effect on resistivity, but these factors were included in the experiment for confirmation. It was further known that Z has only an additive effect on the resistivity. Thus the following model is postulated:

$$Y_{ijk\ell m} = \mu_{ijk} + \epsilon_{ijklm} = m_0 + I_i + Z_j + T_k + (IT)_{ik} + \epsilon_{ijk\ell m}.$$

The current data confirmed this model. The adjusted R^2 is above 99%. Residual plots, a normal probability plot, and other diagnostic measures suggest that it is reasonable to assume normally distributed measurement errors with a constant variance σ_ϵ^2.

Since factors O and F do not have significant effects, the design can be collapsed into factors I, Z, and T which results in a replicated 2^3 design in these three factors. While the theory is developed assuming a single replicate of a design $\mathcal{D}$, situations with replicates can be transformed easily to use the theory. Notice that the OLS estimator of $\boldsymbol{\mu}$ depends only on the sample means at the design points in $\mathcal{D}$. Therefore, for the collapsed design in this example, we adopt the following more succinct expression of the same model:

$$Y^\star_{ijk} = \mu_{ijk} + \epsilon_{ijk} = m_0 + I_i + Z_j + T_k + (IT)_{ik} + \epsilon^\star_{ijk}, \tag{4.12}$$

where $Y^\star_{ijk} = \overline{Y}_{ijk\cdot\cdot}$ and $\epsilon^\star_{ijk} = \overline{\epsilon}_{ijk\cdot\cdot}$. Here, the standard bar-dot notation means that an average is taken over the subscripts replaced by dots (but, of course, only at the design points in $\mathcal{D}$).

In the language of this paper, I is a C^1–type factor, Z is a C^2–type factor, and T is a N^1–type factor. In this example, there are no C^2–type or N^2–type factors. The model (4.12) satisfies the symmetry condition in Theorem (3.1). Let

$$\boldsymbol{v}_k = \begin{array}{c} \\ \left[\begin{array}{cc} 0 & k \\ 0 & 0 \end{array}\right] \\ \begin{array}{cc} 0 & 1 \end{array} \\ T \end{array} \quad \begin{array}{c} I \\ 0 \\ 1 \end{array} \qquad \text{and} \qquad \boldsymbol{\mu}_k = \boldsymbol{v}_k \otimes J_{2\times 1} = \begin{array}{c} \\ \left[\begin{array}{cc} 0 & k \\ 0 & k \\ 0 & 0 \\ 0 & 0 \end{array}\right] \\ \begin{array}{cc} 0 & 1 \end{array} \\ T \end{array} \quad \begin{array}{c} (I,Z) \\ (0,0) \\ (0,1) \\ (1,0) \\ (1,1) \end{array}$$

To compute h, we simulated data sets having mean $\boldsymbol{\mu}_{100}$, i.e., each generated data set had the form

		(I,Z)
Z_1	$100 + Z_2$	$(0,0)$
Z_3	$100 + Z_4$	$(0,1)$
Z_5	Z_6	$(1,0)$
Z_7	Z_8	$(1,1)$
0	1	
T		

where $Z_1, \ldots Z_8$ were mutually independent standard normal random variables. Model (4.12) was fit to the simulated data, then the estimated cell means $\widehat{\mu}_{ijk}$ and $\widehat{\xi}_{ij} = \min\{\widehat{\mu}_{ij0}, \widehat{\mu}_{ij1}\}$ were computed.

Next, a chi-square random variate with $\nu = 11$ degrees of freedom, V, was drawn and $S^2 = V/2v$ was computed. It is important to note that the factor 2 is used in the definition of S^2 because $\mathrm{Var}(\epsilon^\star_{ijk}) = \sigma^2_\epsilon/2$ where σ^2_ϵ is the standard deviation of the original $Y_{ijk\ell m}$ data. Then $T = (\max_{ij} \widehat{\xi}_{ij} - \widehat{\xi}_{11})/S$ is calculated. The estimated 90% confidence point of $h = 1.50$ was based on 10,000 replications of the above T calculation.

Based on the original data, $\widehat{m}_0 = 60.67$, $\widehat{I}_1 = 23.22$, $\widehat{Z}_1 = 68.06$, $(\widehat{IT})_{11} = 13.16$, and $\widehat{Z}_1 = 20.92$ so that

i	j	μ_{ij0}	μ_{ij1}	$\widehat{\xi}_{ij}$
0	0	60.66	94.69	60.66
0	1	71.12	105.15	71.12
1	0	72.27	112.88	72.27
1	1	82.73	123.34	82.73

Also $S = 3.47$ and so the yardstick is $h \times S = 1.50 \times 3.47$=5.21. Thus at the 90% confidence level, Procedure $\mathcal{G}$ selects the single design combination with factors I and T both at high levels.

5 Discussion

5.1 A Systematic Method of Verifying Theorem 3.1

In addition to the method illustrated in Example 2, there is a systematic technique of determining whether the hypothesis (3.7) of Theorem 3.1 holds in any particular application. We provide a brief sketch of the technique which involves solving a related linear programming problem. Recall that the canonical form of an LP in unknown $\boldsymbol{w} = (w_1, \ldots, w_n)$ is

$$\begin{gathered} \max \boldsymbol{c}^\top \boldsymbol{w} \\ \text{s.t.} \\ \boldsymbol{A}\boldsymbol{w} \le \boldsymbol{b} \\ \boldsymbol{w} \ge 0 \end{gathered}$$

where $\boldsymbol{c}$, $\boldsymbol{A}$, and $\boldsymbol{b}$ are given $n \times 1$, $m \times n$, and $m \times 1$ arrays.

Let $\mathcal{Z}$ (for "Zero") denote those $(\boldsymbol{i}^1, \boldsymbol{j}^1)$ combinations corresponding to elements in the first column *or* last row of $\boldsymbol{m}_{CN}$. We wish to determine whether a sequence of $\boldsymbol{m}_{CN}$ exists that satisfies Model (2.4)–(2.5), has $m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) \to +\infty$ for all $(\boldsymbol{i}^1, \boldsymbol{j}^1) \notin \mathcal{Z}$, and for which $m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) = 0$ for all $(\boldsymbol{i}^1, \boldsymbol{j}^1) \in \mathcal{Z}$. To solve this problem, we introduce the auxiliary scalar variable w whose role is to be a lower bound on the elements we wish to simultaneously drive to $+\infty$. Then we solve

$$\begin{gathered} \max\ w \\ \text{s.t.} \\ m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) = 0 \ \text{ for all } \ (\boldsymbol{i}^1, \boldsymbol{j}^1) \in \mathcal{Z} \\ w \le m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) \ \text{ for all } \ (\boldsymbol{i}^1, \boldsymbol{j}^1) \notin \mathcal{Z} \\ \textstyle\sum_{\mathcal{I}^1 \times \mathcal{J}^1} (C^1_{i^\star_1} C^1_{i^\star_2} \cdots C^1_{i^\star_{p^*}} N^1_{j^\star_1} N^1_{j^\star_2} \cdots N^1_{j^\star_{r^*}})_{\boldsymbol{i}^1, \boldsymbol{j}^1} = 0 \ \text{ for all } \ (\boldsymbol{i}^\star, \boldsymbol{j}^\star) \in Q^C_1 \times Q^N_1 \end{gathered} \tag{5.13}$$

The variables for the LP are w, m_0, and $(C^1_{i^\star_1} C^1_{i^\star_2} \cdots C^1_{i^\star_{p^*}} N^1_{j^\star_1} N^1_{j^\star_2} \cdots N^1_{j^\star_{r^*}})_{\boldsymbol{i}^1, \boldsymbol{j}^1}$ for $(\boldsymbol{i}^\star, \boldsymbol{j}^\star) \in Q_1^C \times Q_1^N$; we regard $m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1)$ as merely a notation for the linear combination

$$m_0 + \sum_{Q_1^C \times Q_1^N} (C^1_{i^\star_1} C^1_{i^\star_2} \cdots C^1_{i^\star_{p^*}} N^1_{j^\star_1} N^1_{j^\star_2} \cdots N^1_{j^\star_{r^*}})_{\boldsymbol{i}^1, \boldsymbol{j}^1}$$

The initial constraint forces the first column and last row to be zero while the second causes w to be the minimum of the $m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1)$ for $(\boldsymbol{i}^1, \boldsymbol{j}^1) \notin \mathcal{Z}$. The third set of equalities are identifiability constraints for the main effects and interactions.

In the usual way, the LP (5.13) can be reformulated in canonical form by replacing unconstrained variables z by $z^+ - z^-$ where $z^+ \geq 0$ and $z^- \geq 0$ and equality constraints $\boldsymbol{a}^\top \boldsymbol{z} = 0$ by $\boldsymbol{a}^\top \boldsymbol{z} \leq 0$ and $-\boldsymbol{a}^\top \boldsymbol{z} \geq 0$. Condition (3.7) of Theorem 3.1 holds if and only if the LP has an unconstrained optimum.

5.2 Bounded Means

In some applications, the response is bounded above by a known value. For example, in the cake mix study in Example 1, the highest taste score is 8. In the integrated circuit example of Section 4, it may be known that the highest possible resistivity can not exceed some threshold, say, 130. When there is a known upper bound on the measurements, the maxmin criterion has an alternative interpretation and the proposed procedure can be modified slightly to improve efficiency.

The maxmin criterion judges the quality of a product or process design by its lowest mean across the environmental conditions. If there is an upper bound on the means, say, U, then the design can be equivalently judged by the largest difference between the means across the environmental conditions and the upper bound U. The best product based on maxmin criterion is *also* the product that minimizes the difference from its mean and U, across the environmental conditions.

If, in addition, L is a lower bound on the of the means then Theorems 3.1, 3.2, and 3.3 can be improved by replacing each $+\infty$ by $U - L$, i.e, taking the least favorable configuration to be

$$\begin{pmatrix} 0 & U-L & \cdots & U-L \\ \vdots & \vdots & \vdots & \vdots \\ 0 & U-L & \cdots & U-L \\ 0 & 0 & \cdots & 0 \end{pmatrix} \otimes J_{|\mathcal{I}^2| \times |\mathcal{J}^2|}.$$

However, unless $U - L$ is less than $3\sigma_\epsilon$ the resulting h will be close to the one determined directly from Theorems 3.1, 3.2 or 3.3.

References

Bechhofer, R. E., Santner, T. J. and Goldsman, D. M. (1995). *Designing Experiments for Statistical Selection, Screening, and Multiple Comparisons.* New York:J. Wiley & Sons.

Box, G. E. P. (1988). Signal-to-Noise ratios, performance criteria, and transformations. *Technometrics*, **30**, 1-17.

Box, G. E. P. and Jones, S. (1992). Split-plot designs for robust product experimentation. *Jour. Applied Statistics*, **19**, 3-26.

Gupta, S. S. (1956). *On A Decision Rule for a Problem in Ranking Mean.* Ph.D. Dissertation, Institute of Statistics, University of North Carolina, Chapel Hill, NC.

Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics*, **7**, 225-245.

Myers, R. H., Khuri, A. I., and Vining, G. (1992). Response Surface Alternatives to the Taguchi Robust Parameter Design Approach, *American Statistician*, **46**, 131-139.

Nair, V. J. et al. (1992). Taguchi's parameter design: a panel discussion, *Technometrics*, **34**, 127-161.

Pan, G. and Santner, T. J. (1996). Selection and screening procedures to determine optimal product designs, submitted.

Santner, T. J. and Pan, G. (1997). The Use of Subset Selection in Combined-Array Experiments to Determine Optimal Product or Process Designs. *Advances in Statistical Decision Theory and Methodology* , Editors: Panchapakesan, S. and Balakrishnan, N., to appear.

Shoemaker, A. C., Tsui, K., and Wu, C.-F. J. (1991). Economical experimental methods for robust design, *Technometrics*, **33**, 415-427.

Taguchi, G. and Phadke, M. S. (1984). Quality engineering through design optimization, Atlanta: Conference Record, *GLOBECOM84 Meeting, IEEE Communications Society*, 1106-1113.

A Appendix: Proofs

We first establish the following lemma.

Lemma A.1

$$\inf_{\Omega} P_{\boldsymbol{\mu}}\{CS\} = \inf_{\Omega_0} P_{\boldsymbol{\mu}}\{CS\} = \inf_{\Omega_{00}} P_{\boldsymbol{\mu}}\{CS\} \tag{A.14}$$

where

$$\Omega = \{\boldsymbol{\mu} : (2.4) - (2.5) \text{ hold}\}, \quad \Omega_0 = \{\boldsymbol{\mu} \in \Omega : m_N(\boldsymbol{j}^2) \equiv 0\}, \quad \Omega_{00} = \{\boldsymbol{\mu} \in \Omega_0 : m_C(\boldsymbol{i}^2) \equiv 0\}.$$

Proof: First, recall that any possible true $\boldsymbol{\mu}$ is an $|\mathcal{I}| \times |\mathcal{J}|$ matrix that can be written uniquely in the form

$$\boldsymbol{\mu} = \boldsymbol{m}_{CN} \otimes J_{|\mathcal{I}^2|\times|\mathcal{J}^2|} + J_{|\mathcal{I}^1|\times 1} \otimes \boldsymbol{m}_C \otimes J_{1\times|\mathcal{J}|} + J_{|\mathcal{I}|\times|\mathcal{J}^1|} \otimes \boldsymbol{m}_N^{\top} \tag{A.15}$$

where $\otimes$ denotes Kronecker product, $\boldsymbol{m}_{CN} = (m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1))$ is $|\mathcal{I}^1| \times |\mathcal{J}^1|$, $\boldsymbol{m}_C = (m_C(\boldsymbol{i}^2))$ is $|\mathcal{I}^2| \times 1$, $\boldsymbol{m}_N = (m_N(\boldsymbol{j}^2))$ is $|\mathcal{J}^2| \times 1$. All components are ordered lexicographically. Notice that model component (2.5) specifies that $\boldsymbol{m}_{CN} \in \boldsymbol{V}_{CN}$, where

$$\boldsymbol{V}_{CN} = \{\boldsymbol{m}_{CN} : (2.5) \text{ holds}\}.$$

To prove the left-hand equality of (A.14), pick $\boldsymbol{\mu} \in \Omega$ with corresponding elements $(\boldsymbol{m}_{CN}, \boldsymbol{m}_C, \boldsymbol{m}_N)$ in (A.15). Define

$$\boldsymbol{\mu}^\star = \boldsymbol{\mu} - J_{|\mathcal{I}|\times|\mathcal{J}^1|} \otimes \boldsymbol{m}_N^\top.$$

If $(\boldsymbol{m}_{CN}^\star, \boldsymbol{m}_C^\star, \boldsymbol{m}_N^\star)$ correspond to $\boldsymbol{\mu}^\star$, then it is straightforward to show that $\boldsymbol{m}_{CN} = \boldsymbol{m}_{CN}^\star$, $\boldsymbol{m}_C = \boldsymbol{m}_C^\star$, and $\boldsymbol{m}_N^\star \equiv 0$ which shows that $\boldsymbol{\mu}^\star \in \Omega_0$. We claim that $P_{\boldsymbol{\mu}}\{CS\} = P_{\boldsymbol{\mu}^\star}\{CS\}$ which will obviously establish the left-hand side of (A.14).

Consider the relationship between the *population* row minimums for $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^\star$; let $\xi_{\boldsymbol{i}}$ correspond to $\boldsymbol{\mu}$ and $\xi_{\boldsymbol{i}}^\star$ correspond to $\boldsymbol{\mu}^\star$. Fix $\boldsymbol{i} = (\boldsymbol{i}^1, \boldsymbol{i}^2)$ then

$$\begin{aligned}
\xi_{\boldsymbol{i}} &= \min_{\boldsymbol{j}} \left\{ m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) + m_C(\boldsymbol{i}^2) + m_N(\boldsymbol{j}^2) \right\} \\
&= \min_{\boldsymbol{j}^1} \left\{ m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) + m_C(\boldsymbol{i}^2) \right\} + \min_{\boldsymbol{j}^2} \left\{ m_N(\boldsymbol{j}^2) \right\} \\
&= \xi_{\boldsymbol{i}}^\star + T_{pop},
\end{aligned}$$

say, where T_{pop} is independent of $\boldsymbol{i}$. Using this fact, it is easy to see that the index $[|\mathcal{I}|]$ that maximizes $\xi_{\boldsymbol{i}}$ is identical for both $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^\star$ and the relationship between the row minimums is

$$\xi_{[|\mathcal{I}|]} = \max\left\{\xi_{\boldsymbol{i}}\right\} = \max\left\{\xi_{\boldsymbol{i}}^\star + T_{pop}\right\} = \xi_{[|\mathcal{I}|]}^\star + T_{pop}.$$

Let $\boldsymbol{Y}_{\mathcal{D}} = \boldsymbol{\mu}_{\mathcal{D}} + \boldsymbol{\epsilon}_{\mathcal{D}}$ and $\boldsymbol{Y}_{\mathcal{D}}^\star = \boldsymbol{\mu}_{\mathcal{D}}^\star + \boldsymbol{\epsilon}_{\mathcal{D}}$ where the components of $\boldsymbol{\epsilon}_{\mathcal{D}}$ are independent $N(0, \sigma_\epsilon^2)$ random variables. We determine the relationship between the *estimated* row minimums for $\boldsymbol{Y}_{\mathcal{D}}$ and $\boldsymbol{Y}_{\mathcal{D}}^\star$.

This association is established by first computing the relation between the individual estimated means $\mu_{\boldsymbol{i},\boldsymbol{j}}$ under $\boldsymbol{Y}_{\mathcal{D}}$ and $\boldsymbol{Y}_{\mathcal{D}}^\star$. Let $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\mu}}^\star$ denote the estimated means using the data $\boldsymbol{Y}_{\mathcal{D}}$ and $\boldsymbol{Y}_{\mathcal{D}}^\star$, respectively. We have $\widehat{\mu}_{\boldsymbol{i},\boldsymbol{j}} = \widehat{m}_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) + \widehat{m}_C(\boldsymbol{i}^2) + \widehat{m}_N(\boldsymbol{j}^2)$ where each of the latter three terms are the sum of estimated main effects and interactions. In particular, $(\widehat{\boldsymbol{m}}_{CN}, \widehat{\boldsymbol{m}}_C, \widehat{\boldsymbol{m}}_N)$ minimizes

$$\sum_{\boldsymbol{i},\boldsymbol{j}} \left\{ Y_{\boldsymbol{i},\boldsymbol{j}} - m_{CN}^a(\boldsymbol{i}^1, \boldsymbol{j}^1) - m_C^a(\boldsymbol{i}^2) - m_N^a(\boldsymbol{j}^2) \right\}^2$$

where the superscript a denotes that these are the variables over which the minimization is to take place. Throughout the appendix, we assume that the subscripts in summations and minimizations are over the design points in $\mathcal{D}$. To derive $\widehat{\boldsymbol{\mu}}^\star$, we observe that

$$\begin{aligned}
&\sum_{\boldsymbol{i},\boldsymbol{j}} \left\{ Y_{\boldsymbol{i},\boldsymbol{j}}^\star - m_{CN}^a(\boldsymbol{i}^1, \boldsymbol{j}^1) - m_C^a(\boldsymbol{i}^2) - m_N^a(\boldsymbol{j}^2) \right\}^2 \\
= &\sum_{\boldsymbol{i},\boldsymbol{j}} \left\{ Y_{\boldsymbol{i},\boldsymbol{j}} - m_N(\boldsymbol{j}^2) - m_{CN}^a(\boldsymbol{i}^1, \boldsymbol{j}^1) - m_C^a(\boldsymbol{i}^2) - m_N^a(\boldsymbol{j}^2) \right\}^2 \\
= &\sum_{\boldsymbol{i},\boldsymbol{j}} \left\{ Y_{\boldsymbol{i},\boldsymbol{j}} - m_{CN}^a(\boldsymbol{i}^1, \boldsymbol{j}^1) - m_C^a(\boldsymbol{i}^2) - m_N^{aa}(\boldsymbol{j}^2) \right\}^2,
\end{aligned}$$

say, where $m_N^{aa}(\boldsymbol{j}^2) = m_N(\boldsymbol{j}^2) + m_N^a(\boldsymbol{j}^2)$. Thus we see that the OLSEs of the terms that comprise $\boldsymbol{\mu}^\star$ satisfy $\widehat{\boldsymbol{m}}^\star_{CN} = \widehat{\boldsymbol{m}}_{CN}$, $\widehat{\boldsymbol{m}}^\star_C = \widehat{\boldsymbol{m}}_C$, and $\widehat{\boldsymbol{m}}^\star_N = \widehat{\boldsymbol{m}}_N - \boldsymbol{m}_N$; these equations imply that $\widehat{\mu}^\star_{\boldsymbol{i},\boldsymbol{j}} = \widehat{\mu}_{\boldsymbol{i},\boldsymbol{j}} - m_N(\boldsymbol{j}^2)$.

The last equation shows two things. First, let $\widehat{\xi}_{\boldsymbol{i}}$ and $\widehat{\xi}^\star_{\boldsymbol{i}}$ denote the row minimums for $\boldsymbol{Y}_{\mathcal{D}}$ and $\boldsymbol{Y}^\star_{\mathcal{D}}$, respectively. Then, for any $\boldsymbol{i} \in \mathcal{I}$ we have

$$\begin{aligned}
\widehat{\xi}^\star_{\boldsymbol{i}} &= \min_{\boldsymbol{j}} \left\{ \widehat{m}^\star_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) + \widehat{m}^\star_C(\boldsymbol{i}^2) + \widehat{m}^\star_N(\boldsymbol{j}^2) \right\} \\
&= \min_{\boldsymbol{j}^1} \left\{ \widehat{m}_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) + \widehat{m}_C(\boldsymbol{i}^2) \right\} + \min_{\boldsymbol{j}^2} \left\{ \widehat{m}_N(\boldsymbol{j}^2) - m_N(\boldsymbol{j}^2) \right\} \\
&\quad + \min_{\boldsymbol{j}^2} \left\{ \widehat{m}_N(\boldsymbol{j}^2) \right\} - \min_{\boldsymbol{j}^2} \left\{ \widehat{m}_N(\boldsymbol{j}^2) \right\} \\
&= \min_{\boldsymbol{j}} \left\{ \widehat{m}_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) + \widehat{m}_C(\boldsymbol{i}^2) + \widehat{m}_N(\boldsymbol{j}^2) \right\} \\
&\quad + \min_{\boldsymbol{j}^2} \left\{ \widehat{m}_N(\boldsymbol{j}^2) - m_N(\boldsymbol{j}^2) \right\} - \min_{\boldsymbol{j}^2} \left\{ \widehat{m}_N(\boldsymbol{j}^2) \right\} \\
&= \widehat{\xi}_{\boldsymbol{i}} + \min_{\boldsymbol{j}^2} \left\{ \widehat{m}_N(\boldsymbol{j}^2) - m_N(\boldsymbol{j}^2) \right\} - \min_{\boldsymbol{j}^2} \left\{ \widehat{m}_N(\boldsymbol{j}^2) \right\} \\
&= \widehat{\xi}_{\boldsymbol{i}} + T_{est},
\end{aligned}$$

say.

In addition, if S^2 is the residual sum of squares based on $\boldsymbol{Y}$ used to estimate σ^2_ϵ and $(S^\star)^2$ is the corresponding sum of squares based on $\boldsymbol{Y}^\star$, then

$$\begin{aligned}
\nu(S^\star)^2 &= \sum_{\boldsymbol{i},\boldsymbol{j}} (Y^\star_{\boldsymbol{i},\boldsymbol{j}} - \widehat{\mu}^\star_{\boldsymbol{i},\boldsymbol{j}})^2 \\
&= \sum_{\boldsymbol{i},\boldsymbol{j}} (Y_{\boldsymbol{i},\boldsymbol{j}} - m_N(\boldsymbol{j}^2) - \widehat{\mu}_{\boldsymbol{i},\boldsymbol{j}} + m_N(\boldsymbol{j}^2))^2 \\
&= \sum_{\boldsymbol{i},\boldsymbol{j}} (Y_{\boldsymbol{i},\boldsymbol{j}} - \widehat{\mu}_{\boldsymbol{i},\boldsymbol{j}})^2 = \nu S^2,
\end{aligned}$$

where ν is the (common) degrees of freedom for S^2 and $(S^*)^2$.

With these relationships we can compute the PCS under $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^\star$. Let $\widehat{\xi}_{(|\mathcal{I}|)}$ be associated with $\xi_{[|\mathcal{I}|]}$ then $\widehat{\xi}^\star_{(|\mathcal{I}|)} = \widehat{\xi}_{[|\mathcal{I}|]} + T_{pop}$ is associated with $\xi^\star_{[|\mathcal{I}|]} = \xi_{[|\mathcal{I}|]} + T_{pop}$ since the same product design $[|\mathcal{I}|]$ simultaneously maximizes both $\xi_{\boldsymbol{i}}$ and $\xi^\star_{\boldsymbol{i}} = \xi_{\boldsymbol{i}} + T_{pop}$. Thus we obtain

$$\begin{aligned}
P_{\boldsymbol{\mu}}\{CS\} &= P\left\{ \widehat{\xi}_{(|\mathcal{I}|)} \geq \widehat{\xi}_{\boldsymbol{i}} - hS \ \text{ for all } \ \boldsymbol{i} \right\} \\
&= P\left\{ \widehat{\xi}_{(|\mathcal{I}|)} + T_{pop} \geq \widehat{\xi}_{\boldsymbol{i}} + T_{pop} - hS \ \text{ for all } \boldsymbol{i} \right\} \\
&= P\left\{ \widehat{\xi}^\star_{(|\mathcal{I}|)} \geq \widehat{\xi}^\star_{\boldsymbol{i}} - hS \ \text{ for all } \boldsymbol{i} \right\} \\
&= P_{\boldsymbol{\mu}^\star}\{CS\}
\end{aligned}$$

To prove the right-hand equality in (A.14), pick $\boldsymbol{\mu} \in \Omega_0$ with uniquely defined expansion terms $(\boldsymbol{m}_{CN}, \boldsymbol{m}_C)$; set

$$\boldsymbol{\mu}^\star = \boldsymbol{\mu} - J_{|\mathcal{I}^1| \times 1} \otimes \boldsymbol{m}_C \otimes J_{1 \times |\mathcal{J}|}.$$

As above, it is straightforward to calculate that the corresponding terms for $\boldsymbol{\mu}^\star$ satisfy $(\boldsymbol{m}^\star_{CN}, \boldsymbol{m}^\star_C, \boldsymbol{m}^\star_N) = (\boldsymbol{m}_{CN}, 0, 0)$ so that $\boldsymbol{\mu}^\star \in \Omega_{00}$. Thus it suffices to show $P_{\boldsymbol{\mu}}\{CS\} = P_{\boldsymbol{\mu}^\star}\{CS\}$ to prove the right-hand equality.

The row minimums for $\boldsymbol{\mu}$, $\xi_{\boldsymbol{i}}$, and for $\boldsymbol{\mu}^\star$, $\xi^\star_{\boldsymbol{i}}$, satisfy the equation

$$\begin{aligned}\xi_{\boldsymbol{i}} &= \min_{\boldsymbol{j}}\left\{m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1) + m_C(\boldsymbol{i}^2)\right\} \\ &= \min_{\boldsymbol{j}^1}\left\{m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1)\right\} + m_C(\boldsymbol{i}^2) \\ &= \xi^\star_{\boldsymbol{i}} + m_C(\boldsymbol{i}^2);\end{aligned}$$

notice that $\xi^\star_{\boldsymbol{i}}$ is independent of $\boldsymbol{i}^2$. From this calculation, the optimal row levels are determined as follows. Suppose that $(\boldsymbol{i}^1_N, \boldsymbol{i}^2_N)$ is defined by

$$m_C(\boldsymbol{i}^2_N) = \max_{\boldsymbol{i}^2} m_C(\boldsymbol{i}^2) \quad \text{and} \quad \min_{\boldsymbol{j}^1} m_{CN}(\boldsymbol{i}^1_N, \boldsymbol{j}^1) = \max_{\boldsymbol{i}^1}\left\{m_0 + \min_{\boldsymbol{j}^1} m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1)\right\}.$$

Then $(\boldsymbol{i}^1_N, \boldsymbol{i}^2_N)$ is the optimum product design under $\boldsymbol{\mu}$ and $(\boldsymbol{i}^1_N, \boldsymbol{i}^2)$ is the optimum product design under $\boldsymbol{\mu}^\star$ for any $\boldsymbol{i}^2$ and, in particular, for $\boldsymbol{i}^2 = \boldsymbol{i}^2_N$. To see this we compute

$$\begin{aligned}\max_{\boldsymbol{i}^1, \boldsymbol{i}^2} \xi_{\boldsymbol{i}^1, \boldsymbol{i}^2} &= \max_{\boldsymbol{i}^1, \boldsymbol{i}^2}\left\{\min_{\boldsymbol{j}^1}\left\{m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1)\right\} + m_C(\boldsymbol{i}^2)\right\} \\ &= \max_{\boldsymbol{i}^1}\left\{\min_{\boldsymbol{j}^1}\left\{m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1)\right\}\right\} + \max_{\boldsymbol{i}^2}\left\{m_C(\boldsymbol{i}^2)\right\} = \xi_{\boldsymbol{i}^1_N, \boldsymbol{i}^2_N}\end{aligned}$$

and

$$\begin{aligned}\max_{\boldsymbol{i}^1, \boldsymbol{i}^2} \xi^\star_{\boldsymbol{i}^1, \boldsymbol{i}^2} &= \max_{\boldsymbol{i}^1, \boldsymbol{i}^2}\left\{\min_{\boldsymbol{j}^1}\left\{m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1)\right\}\right\} \\ &= \max_{\boldsymbol{i}^1}\left\{\min_{\boldsymbol{j}^1}\left\{m_{CN}(\boldsymbol{i}^1, \boldsymbol{j}^1)\right\}\right\} = \xi^\star_{\boldsymbol{i}^1_N, \boldsymbol{i}^2}\end{aligned}$$

for any $\boldsymbol{i}^2$ and in particular for $\boldsymbol{i}^2_N$. Also note that $\xi_{\boldsymbol{i}^1_N, \boldsymbol{i}^2_N} = \xi^\star_{\boldsymbol{i}^1_N, \boldsymbol{i}^2_N} + Q_{pop}$, say, where Q_{pop} is constant.

As above, let $\boldsymbol{Y}_{\mathcal{D}} = \boldsymbol{\mu}_{\mathcal{D}} + \boldsymbol{\epsilon}_{\mathcal{D}}$ and $\boldsymbol{Y}^\star_{\mathcal{D}} = \boldsymbol{\mu}^\star_{\mathcal{D}} + \boldsymbol{\epsilon}_{\mathcal{D}}$ where the components of $\boldsymbol{\epsilon}_{\mathcal{D}}$ are independent $N(0, \sigma^2_\epsilon)$ random variables. Arguing similarly as the first part of the proof and letting quantities with(out) the superscript correspond to $\boldsymbol{Y}^\star_{\mathcal{D}}$ ($\boldsymbol{Y}_{\mathcal{D}}$), the estimated row minimums for $\boldsymbol{Y}_{\mathcal{D}}$ and $\boldsymbol{Y}^\star_{\mathcal{D}}$ can be shown to satisfy $\widehat{\xi}^\star_{\boldsymbol{i}} = \widehat{\xi}_{\boldsymbol{i}} - m_C(\boldsymbol{i}^2)$ for all $\boldsymbol{i}$ and $(S^\star)^2 = S^2$. Let $\widehat{\xi}_{(\boldsymbol{i}^1_N, \boldsymbol{i}^2_N)}$ be the estimator associated with $\xi_{\boldsymbol{i}^1_N, \boldsymbol{i}^2_N}$ and similarly for $\widehat{\xi}^\star_{(\boldsymbol{i}^1_N, \boldsymbol{i}^2_N)}$ Thus we obtain

$$\begin{aligned}P_{\boldsymbol{\mu}}\{CS\} &= P\left\{\widehat{\xi}_{(\boldsymbol{i}^1_N, \boldsymbol{i}^2_N)} \geq \widehat{\xi}_{\boldsymbol{i}} - hS \ \text{ for all } \ \boldsymbol{i}\right\} \\ &= P\left\{\widehat{\xi}^\star_{(\boldsymbol{i}^1_N, \boldsymbol{i}^2_N)} + m_C(\boldsymbol{i}^2_N) \geq \widehat{\xi}^\star_{\boldsymbol{i}} + m_C(\boldsymbol{i}^2) - hS \ \text{ for all } \ \boldsymbol{i}\right\}\end{aligned}$$

$$\begin{aligned} &\geq P\left\{\widehat{\xi}^{\star}_{(\boldsymbol{i}^1_N,\boldsymbol{i}^2_N)} \geq \widehat{\xi}^{\star}_{\boldsymbol{i}} - hS \ \text{ for all } \ \boldsymbol{i}\right\} \qquad (A.16)\\ &= P_{\boldsymbol{\mu}^\star}\{CS\} \end{aligned}$$

where (A.16) holds because $m_C(\boldsymbol{i}^2_N) \geq m_C(\boldsymbol{i}^2)$ for all $\boldsymbol{i}^2$. $\square$

Proof of Theorem 3.2 Choose $\boldsymbol{\mu} = \boldsymbol{v} \otimes J_{|\mathcal{I}^2|\times|\mathcal{J}^2|}$ with $\boldsymbol{v} = (v_{\boldsymbol{i}^1,\boldsymbol{j}^1}) \in \boldsymbol{V}_{CN}$. Then

$$\mu_{\boldsymbol{i},\boldsymbol{j}} = v_{\boldsymbol{i}^1,\boldsymbol{j}^1} \qquad (A.17)$$

for all $(\boldsymbol{i}^2,\boldsymbol{j}^2) \in \mathcal{I}^2 \otimes \mathcal{J}^2$ and

$$\xi_{\boldsymbol{i}^1,\boldsymbol{i}^2} = \min_{\boldsymbol{j}^1} v_{\boldsymbol{i}^1,\boldsymbol{j}^1} \qquad (A.18)$$

is independent of $\boldsymbol{i}^2$. Suppose that $\boldsymbol{j}^1(\boldsymbol{i}^1)$ is a function that denotes a set of column indices for which the minimum in (A.18) is attained, i.e., $\xi_{\boldsymbol{i}^1,\boldsymbol{i}^2} = v_{\boldsymbol{i}^1,\boldsymbol{j}^1(\boldsymbol{i}^1)}$. Finally, let $\boldsymbol{I} = (\boldsymbol{I}^1,\boldsymbol{I}^2) \in \mathcal{I}$ denote an optimum control setting for $\boldsymbol{\mu}$, i.e.,

$$\xi_{\boldsymbol{I}} = \max_{\boldsymbol{i}} \xi_{\boldsymbol{i}}.$$

From the development in the proof of Lemma A.1,

$$\begin{aligned} \widehat{\mu}_{\boldsymbol{i},\boldsymbol{j}} &= \widehat{m}_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1) + \widehat{m}_C(\boldsymbol{i}^2) + \widehat{m}_N(\boldsymbol{j}^2)\\ &= \mu_{\boldsymbol{i},\boldsymbol{j}} + \overline{\epsilon}_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1) + \overline{\epsilon}_C(\boldsymbol{i}^2) + \overline{\epsilon}_N(\boldsymbol{j}^2) \qquad (A.19) \end{aligned}$$

where $\mu_{\boldsymbol{i},\boldsymbol{j}} = v_{\boldsymbol{i},\boldsymbol{j}}$ by (A.17), $\overline{\epsilon}_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1)$ is the linear combination of the $\epsilon_{\boldsymbol{i},\boldsymbol{j}}$ determined by the terms in (2.5) and whose values depends only on $(\boldsymbol{i}^1,\boldsymbol{j}^1)$, and similarly each of $\overline{\epsilon}_C(\boldsymbol{i}^2)$ and $\overline{\epsilon}_N(\boldsymbol{j}^2)$ are linear combinations of the $\epsilon_{\boldsymbol{i},\boldsymbol{j}}$ whose values depend on $\boldsymbol{i}^2$ and $\boldsymbol{j}^2$, respectively. Thus we have

$$\widehat{\xi}_{\boldsymbol{i}} = \min_{\boldsymbol{j}} \widehat{\mu}_{\boldsymbol{i},\boldsymbol{j}} = \overline{\epsilon}_C(\boldsymbol{i}^2) + \min_{\boldsymbol{j}^2} \overline{\epsilon}_N(\boldsymbol{j}^2) + \min_{\boldsymbol{j}^1}\left\{v_{\boldsymbol{i}^1,\boldsymbol{j}^1} + \overline{\epsilon}_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1)\right\}. \qquad (A.20)$$

Then the probability of correct selection is

$$\begin{aligned} P_{\boldsymbol{\mu}}\{CS\} &= P_{\boldsymbol{\mu}}\left\{\widehat{\xi}_{\boldsymbol{I}} \geq \widehat{\xi}_{\boldsymbol{i}} - h\,S \ \ \forall\ \boldsymbol{i} \in \mathcal{I}\right\}\\ &= P_{\boldsymbol{\mu}}\left\{\begin{array}{l} \widehat{\xi}_{\boldsymbol{I}^1,\boldsymbol{I}^2} \geq \widehat{\xi}_{\boldsymbol{I}^1,\boldsymbol{i}^2} - h\,S \ \ \forall\ \boldsymbol{i}^2 \in \mathcal{I}^2;\\ \widehat{\xi}_{\boldsymbol{I}^1,\boldsymbol{I}^2} \geq \widehat{\xi}_{\boldsymbol{i}^1,\boldsymbol{i}^2} - h\,S \ \ \forall\ \boldsymbol{i}^1 \neq \boldsymbol{I}^1, \boldsymbol{i}^2 \in \mathcal{I}^2 \end{array}\right\} \qquad (A.21) \end{aligned}$$

$$= P\left\{\begin{array}{l} \overline{\epsilon}_C(\boldsymbol{I}^2) \geq \overline{\epsilon}_C(\boldsymbol{i}^2) - h\,S \ \ \forall\ \boldsymbol{i}^2 \in \mathcal{I}^2;\\ \overline{\epsilon}_C(\boldsymbol{I}^2) + \min_{\boldsymbol{j}^1}\left\{v_{\boldsymbol{I}^1,\boldsymbol{j}^1} + \overline{\epsilon}_{CN}(\boldsymbol{I}^1,\boldsymbol{j}^1)\right\} \geq\\ \overline{\epsilon}_C(\boldsymbol{i}^2) + \min_{\boldsymbol{j}^1}\left\{v_{\boldsymbol{i}^1,\boldsymbol{j}^1} + \overline{\epsilon}_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1)\right\} - h\,S \ \ \forall\ \boldsymbol{i}^1 \neq \boldsymbol{I}^1, \boldsymbol{i}^2 \in \mathcal{I}^2 \end{array}\right\} \qquad (A.22)$$

$$
\begin{aligned}
&\geq P\left\{\begin{array}{l}\overline{\epsilon}_C(\boldsymbol{I}^2) \geq \overline{\epsilon}_C(\boldsymbol{i}^2) - h\,S \;\;\forall\, \boldsymbol{i}^2 \in \mathcal{I}^2; \\ \overline{\epsilon}_C(\boldsymbol{I}^2) + \min_{\boldsymbol{j}^1}\left\{v_{\boldsymbol{I}^1,\boldsymbol{j}^1(\boldsymbol{I}^1)} + \overline{\epsilon}_{CN}(\boldsymbol{I}^1,\boldsymbol{j}^1)\right\} \geq \\ \overline{\epsilon}_C(\boldsymbol{i}^2) + \min_{\boldsymbol{j}^1}\left\{v_{\boldsymbol{i}^1,\boldsymbol{j}^1} + \overline{\epsilon}_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1)\right\} - h\,S \;\;\forall\, \boldsymbol{i}^1 \neq \boldsymbol{I}^1, \boldsymbol{i}^2 \in \mathcal{I}^2\end{array}\right\} \qquad \text{(A.23)} \\
&\geq P\left\{\begin{array}{l}\overline{\epsilon}_C(\boldsymbol{I}^2) \geq \overline{\epsilon}_C(\boldsymbol{i}^2) - h\,S \;\;\forall\, \boldsymbol{i}^2 \in \mathcal{I}^2; \\ \overline{\epsilon}_C(\boldsymbol{I}^2) + v_{\boldsymbol{I}^1,\boldsymbol{j}^1(\boldsymbol{I}^1)} + \min_{\boldsymbol{j}^1}\left\{\overline{\epsilon}_{CN}(\boldsymbol{I}^1,\boldsymbol{j}^1)\right\} \geq \\ \overline{\epsilon}_C(\boldsymbol{i}^2) + v_{\boldsymbol{i}^1,\boldsymbol{j}^1(\boldsymbol{i}^1)} + \overline{\epsilon}_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1(\boldsymbol{i}^1)) - h\,S \;\;\forall\, \boldsymbol{i}^1 \neq \boldsymbol{I}^1, \boldsymbol{i}^2 \in \mathcal{I}^2\end{array}\right\} \qquad \text{(A.24)} \\
&\geq P\left\{\begin{array}{l}\overline{\epsilon}_C(\boldsymbol{I}^2) \geq \overline{\epsilon}_C(\boldsymbol{i}^2) - h\,S \;\;\forall\, \boldsymbol{i}^2 \in \mathcal{I}^2; \\ \overline{\epsilon}_C(\boldsymbol{I}^2) + \min_{\boldsymbol{j}^1}\left\{\overline{\epsilon}_{CN}(\boldsymbol{I}^1,\boldsymbol{j}^1)\right\} \geq \\ \overline{\epsilon}_C(\boldsymbol{i}^2) + \overline{\epsilon}_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1(\boldsymbol{i}^1)) - h\,S \;\;\forall\, \boldsymbol{i}^1 \neq \boldsymbol{I}^1, \boldsymbol{i}^2 \in \mathcal{I}^2\end{array}\right\} \qquad \text{(A.25)} \\
&= P(\boldsymbol{I}^1, \boldsymbol{I}^2, \{\boldsymbol{j}^1(\cdot)\})
\end{aligned}
$$

where (A.22) holds by substituting (A.20) in the event (A.21), (A.23) holds because $v_{\boldsymbol{I}^1,\boldsymbol{j}^1(\boldsymbol{I}^1)} \leq v_{\boldsymbol{i}^1,\boldsymbol{j}^1}$ for all $\boldsymbol{j}^1$, (A.24) holds since $v_{\boldsymbol{i}^1,\boldsymbol{j}^1(\boldsymbol{i}^1)} + \overline{\epsilon}_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1(\boldsymbol{i}^1)) \geq \min_{\boldsymbol{j}^1}\{v_{\boldsymbol{i}^1,\boldsymbol{j}^1} + \overline{\epsilon}_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1)\}$, and (A.25) holds because $v_{\boldsymbol{I}^1,\boldsymbol{j}^1(\boldsymbol{I}^1)} \geq v_{\boldsymbol{i}^1,\boldsymbol{j}^1(\boldsymbol{i}^1)}$. □

Proof of Theorems 3.3 and 3.1 Theorem 3.3 follows from Theorem 3.2 by symmetry; when the symmetry conditions (1)–(3) of Theorem 3.1 are satisfied, the terms $\overline{\epsilon}_{CN}(\boldsymbol{i}^1,\boldsymbol{j}^1) + \overline{\epsilon}_C(\boldsymbol{i}^2) + \overline{\epsilon}_N(\boldsymbol{j}^2)$ in $\widehat{\mu}_{\boldsymbol{i},\boldsymbol{j}}$ (A.19) are exchangeable. Thus Theorem 3.1 is a direct conclusion from Theorem 3.3.