The Annals of Statistics
1983, Vol. 11, No. 1, 225–239

MINIMAXITY FOR RANDOMIZED DESIGNS: SOME GENERAL RESULTS*

BY KER-CHAU LI

Purdue University

In many design settings where model violations are present, a "stochastic" minimaxity for many standard randomization procedures is demonstrated. This result requires no special analytic properties of the loss function and estimators. Next, under the squared loss and with the restriction to the use of linear estimators, a recipe for finding a randomized strategy is given. As a special case, randomizing an A-optimal design in the standard manner and using the least squares estimates yields a minimax strategy in most cases. These results generalize some aspects of Wu (1981).

1. Introduction. The role of randomization in the design of experiments has been discussed in numerous papers (see the references given in Wu, 1981). As it was summarized by Wu, the most popular of the arguments favoring the use of randomization are the following: it provides a solid basis for statistical inference; it ensures impartiality; it is a source of robustness against model inadequacies. Most of the literature has been addressed to the first and the second arguments.

While the third argument on the model robustness aspect of randomization has already been well accepted, Wu (1981) seems to be the first work devoted to giving it a formal definition and rigorous justification. For some basic design setups in comparative experiments where T treatments are to be assigned to N experiment units, Wu argued that since the experimenter's information about the model is never perfect, there is always the possibility that the "true" model deviates from the assumed model. Thus if G is the collection of all possible "true" models, he defined the concept of model-robustness with respect to G in terms of minimizing the maximum possible mean squared error of the corresponding best linear unbiased estimator (for the assumed model) over G. For the use of the model-robustness notion in other contexts, see, for example, Box and Draper (1959) and Huber (1975). Some randomized designs, including the balanced completely randomized design (coined by Wu), the randomized complete block design and the randomized Latin square design, were shown to be model-robust with respect to any G which possesses an appropriate invariance property in each setting. Furthermore, Wu compared some randomized designs in terms of maximum squared bias. In this paper, we shall discuss only the minimax results. Basically we adopt Wu's general framework on the model-violation consideration; i.e., G will be invariant in an appropriate sense. But we shall extend the results to quite general design settings after a careful study of Wu's ideas.

This paper is composed of two parts. The first part (Section 2) discusses the minimaxity of some commonly-used randomization procedures. The results obtained here are the generalization of Wu's Proposition 1, Theorem 4, and the minimaxity for the randomized complete block design and BIB designs. In these results, the competing class of designs was restricted to those having the same treatment replication numbers of each block (the block design case) and for each row and each column (Latin squares case). We shall make the same restriction in this section. For example, in the block design case, the block-

---

Received June 1981; revised June 1982.

[1] Research sponsored in part by the National Science Foundation under Grant No. MCS-78-25301. Revision supported by National Science Foundation under grant No. MCS-8200631.

AMS 1980 *subject classification.* Primary 62K99; secondary, 62A05, 62C20.

Key words and phrases. A-optimality, doubly transitive groups, H-uniform randomization, randomization procedures, randomized strategies, robustness.

treatment incidence matrix is fixed (but arbitrary). (The optimum choice of a block-treatment incidence matrix will be considered in Section 3). However, it is easier to convey these results in terms of a rigorously-defined notion of randomization procedures. Our definition is based on the observation that when applying any randomization procedure (in the usual sense, for example, the complete randomization) to a non-randomized design, the possible realized designs are those with the same treatment replication numbers as those for the original design. Thus by comparing different randomization procedures only, we avoid the more complicated problem of choosing an optimal non-randomized design to be randomized. Roughly speaking, if the class of all possible true models is invariant under a group H, then the randomization procedure (H-uniform randomization, to be defined later) generated by the uniform probability measure on H is minimax in the sense that for any fixed non-randomized design d and estimator δ, the maximum risk of applying a randomization procedure to d and using the corresponding permuted version of the estimator δ is minimized when H-uniform randomization is used. We do not require any special analytic property of δ; thus non-linear estimators (which are sometimes proposed for guarding against distribution violations) are allowed. The loss function could be arbitrary, athough some invariant properties should be satisfied to ensure that we are not estimating any feature of the nuisance parameters such as unit effects, etc.

An important observation leading to our broadened results is that to prove Wu's Proposition 1, no explicit expression of the risk functions is needed. The basic idea is the standard concept that "suitable invariance" implies "minimaxity", due to the Hunt-Stein Theorem (see Lehmann, 1971). However, the result of Blackwell and Girshick (1954), which shows the minimaxity of simple random sampling, is more relevant. This is because unlike the cases where the Hunt-Stein Theorem usually applies, what our group actually transforms are the nuisance parameters (unit effects, etc.), *not* the parameters of interest! Furthermore, there is one special feature about the manner of evaluating the randomized decision rules which makes our results different from any earlier results. Recall that in the standard decision theory, after defining the risk for a non-randomized rule, the risk for a randomized rule is defined to be the mean of the risks of its possible realized rules. However, it is quite obvious that instead of the means, several other location measures such as medians, quantiles, etc., may also be used to assess a randomized rule provided that the possible mathematical difficulties can be removed. In other words, ideally we should compare the randomized rules according to the stochastic orderings of their random (due to randomization) risks. Our minimaxity results in Section 2 are established under such considerations. Therefore, they provide a very sound basis for using randomization procedures in guarding against model-violations.

The second part of this paper concerns the choice of a randomized design under the same model-violation considerations as in Section 2 but *without any restriction to the competing class.* However, we do require that the estimators be linear (but not necessarily the least squares) and the loss function be the squared one. Furthermore, we evaluate the performance of a randomized rule by defining its risk in the standard way; i.e., by considering the mean risk only. These restrictions seem to be unavoidable for obtaining useful results since the usual work on optimal experimental designs (which assumed no model-violations) is based on these assumptions. Among the three commonly-used design criteria, (A, D, and E criteria), our results are most closely related to the A-criterion. In the block design (the two-way heterogeneity design, respectively) settings, we show that the randomized strategy (i.e., design and estimator) which first chooses an A-optimal design and then randomizes it in the standard way, i.e., randomizes completely the blocks and the units within each block (rows and columns, respectively), and uses the usual least squares estimators, is minimax.

These results extend Wu's Theorem 1 and Theorem 3 which justified *randomization* as well as *balance* from the model-robustness viewpoint for the no-blocking setup. We also justify the use of least squares estimators in the appropriate randomization procedures. In using A-optimal designs, we assume (by the loss function) that all treatments are of equal interest. Recently, there have been considerable research interests on designs for comparing test treatments to a control treatment, Bechhofer and Tamhane (1981). In such cases, the loss function should reflect the relative importance of the control and the test treatments. In general, if the loss function is of the form $(L\alpha - \mathbf{a})' (L\alpha - \mathbf{a})$ where L is a $p \times T$ matrix and $\mathbf{a}$ is $p \times 1$ vector estimating $L\alpha$, which was referred to as a linear criterion or L-criterion (see Kiefer, 1974, Karlin and Studden, 1966, or Fedorov, 1972), then our results show that to obtain the minimax randomized strategy one only has to first search for the corresponding optimal designs for the "assumed" (or ideal) model. After finding an optimal one, then we should randomize it in the standard way and use the standard least squares estimates. Having seen such results for the one-way and two-way settings, one is easily led to the conclusion that similar results should hold for the k-way settings. Unfortunately, this is true only when the complicated model which assumes the existence of all higher order (up to k–1) interactions among the units is considered. For the usual additivity model, the minimax randomized design may depend on the actual form of the class G of possible models if $k \geq 3$.

Furthermore, our counter-example shows that for certain invariant class G, randomizing the most symmetric design may sometimes be inferior to randomizing a less symmetric design. This example illustrates the need for rigorous justification in applying randomization in various settings. Like elsewhere, a careless application of Hunt-Stein's idea or any related concepts may incur misleading conclusions. But the crucial issue involved here is not the compactness of the transformation group on the class G (permutation groups are always finite and hence compact). The issue is how the group works. Usually, the transformation group used here can be naturally decomposed into some basic subgroups. (For example, in two-way heterogeneity settings, the transformation group involved is the product of a row permutation group and a column permutation group). The relation between the orbits of these subgroups and the block effects, row, column effects, or higher order interaction effects turns out to be a very important consideration in obtaining the results. (Such a consideration was implicit in Cheng and Li, 1980). If the orbits of the subgroups correspond to block or interaction effects then our minimax results hold. (For example, the orbits of the row permutation group correspond to the column effects and the orbits of the column permutation group correspond to the row effects). However, for the $k \geq 3$ way settings, the orbits of each subgroup which permute the levels of one factor will correspond to (k–1)-interaction effects among factors which are not assumed in the usual additivity models. This explains why we need a complicated model to ensure minimaxity. Also, as a simple consequence, we obtain other randomization procedures which are generated by groups of very small orders and are of the same efficiency as the commonly-used ones when the squared loss function is assumed. Section 4 is devoted to the proofs.

2. Minimax randomization procedures under general loss functions. Suppose T treatments are to be assigned to N experimental units. A (non-randomized) design is a function d from $\{1, \cdots, N\}$ to $\{1, \cdots, T\}$ with the uth unit receiving treatment $d(u)$. Let D be the class of all designs. In this paper, instead of defining a randomized design as a probability measure on D, we shall conveniently treat it as a *random element* with the nonrandomized designs as possible realizations. Denote the ith treatment effect by α_i and let $\alpha = (\alpha_1, \cdots, \alpha_T)'$. We now first present a simple example to illustrate the general results we shall obtain. This example was already considered by Wu.

EXAMPLE 1. *No blocking.* Suppose the yield (or response) y_u of the uth unit satisfies the following additivity assumption:

$$y_u = \alpha_{d(u)} + g_u + \varepsilon_u, \quad u = 1, \cdots, N, \tag{2.1}$$

where g_u is the uth unit effect and ε_u is the random error with mean 0. In the ideal case we would assume $g_u = 0$ (or a constant) and the random errors are homogeneous and uncorrelated. But this certainly is not a good situation for justifying the use of randomization. In fact, there is always the possibility that the "true" model deviates from the ideal one. Let G be the set of all possible $\mathbf{g} = (g_1, \cdots, g_n)'$. Let $\mathscr{E}$ be the set of all possible probability measures of $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)'$. To reflect the vagueness of the experimenter's knowledge, we require that G and $\mathscr{E}$ are invariant under the group H of all permutations on $\{1, \cdots, N\}$; i.e., $\mathbf{g} \in G \Rightarrow \pi\mathbf{g} \in G$ and $\xi \in \mathscr{E} \Rightarrow \pi\xi \in \mathscr{E}$ for all $\pi \in H$, where $\pi\mathbf{g} = (g_{\pi^{-1}(1)}, \cdots, g_{\pi^{-1}(N)})'$ and $\pi(\xi)$ is the probability measure of $\pi\boldsymbol{\varepsilon}$ when ξ is the probability measure of $\boldsymbol{\varepsilon}$. We denote the triple $(\boldsymbol{\alpha}, \mathbf{g}, \xi)$ by s, let S be the set of all possible s, and write $\pi s = (\boldsymbol{\alpha}, \pi\mathbf{g}, \pi\xi)$. Thus we have

$$s \in S \Rightarrow \pi s \in S \text{ for all } \pi \in H. \tag{2.2}$$

Now we shall define the concept of a randomization procedure rigorously. Recall that when applying any randomization procedure (e.g., the complete randomization) to a given non-randomized design d, the possible realized designs will have the same replication numbers as those for the original design d. Since it is clear to see that $H(d) = \{\pi d \mid \pi \in H\}$ is the class of designs with the same replication numbers as d, we define a randomization procedure to be a *a function Φ on D such that $\Phi(d)$ is a random element with possible realizations in $H(d)$ for any $d \in D$.* In particular, the complete randomization is a function which maps $d \in D$ to $\mathbf{h}d$, where $\mathbf{h}$ *is the random permutation generated by the uniform distribution on H.* Thus we denote the complete randomization by $\mathbf{h}$. We shall demonstrate a minimax property for $\mathbf{h}$, after discussing the problem of choice of estimators and loss functions and the problem of evaluating a randomized strategy.

The loss function $\mathscr{L}$ considered in this section does not need any special analytic property. We only require $\mathscr{L}$ to be invariant in the following sense:

$$\mathscr{L}(\pi(s), a) = \mathscr{L}(s, a) \text{ for any } \pi \in H, s \in S \text{ and any } a \text{ in the action space } \mathscr{A}. \tag{2.3}$$

This invariance requirement amounts to claiming that what we estimate depends only on $\boldsymbol{\alpha}$ and in no way on $\mathbf{g}$ or ξ. For instance, we may take $\mathscr{L}(s, a) = (L\boldsymbol{\alpha} - a)'(L\boldsymbol{\alpha} - a)$ where L is a $p \times T$ matrix and $a \in \mathscr{A} = R^p$.

The choice of estimators should also be invariant under H in the following sense. Suppose for design d, an estimator δ, which is a function mapping $\mathbf{y} = (y_1, \cdots, y_N)'$ to an element in $\mathscr{A}$, is used. Then we require that for design πd, the estimator $\pi\delta(\mathbf{y}) = \delta(\pi^{-1}(\mathbf{y}))$ should also be used. This is a reasonable restriction, similar to that imposed by Blackwell and Girshick (1954) in justifying simple random sampling, because when there is no model-violation, the distribution of $\pi^{-1}(\mathbf{y})$ under design πd is the same as the distribution of $\mathbf{y}$ under design d. Thus when a randomization procedure ϕ is applied to design d for which estimator δ is used, the realized strategy (i.e., design and estimator) is determined and will be denoted by $\phi(d, \delta)$.

Now we discuss the problem of evaluating a randomized strategy. As usual, the risk of a non-randomized design and estimator (d, δ) under $s \in S$ is defined by $r(d, \delta; s) = E\,\mathscr{L}(s, \delta(\mathbf{y}))$. But we do not assess a randomized strategy merely by its expected risks. Instead, we consider the class $\mathscr{F}$ of real functions f on the class of all probability measures of R such that $f(p\mu_1 + (1-p)\mu_2) \le \max\{f(\mu_1), f(\mu_2)\}$ for any $0 \le p \le 1$ and any probability measures μ_1 and μ_2 on R. We also write $f(X) = f(\mu)$ if X is a random variable with the probability measure μ. This broad class $\mathscr{F}$ includes the mean, median or any quantile (all given a convenient definition if not unique) functionals of random variables. The following result is what we shall prove:

Under (2.1)–(2.3), the complete randomization $\mathbf{h}$ is minimax in the sense that it minimizes $\max f_{s \in S}(r(\phi(d, \delta); s))$ over all randomization procedures ϕ, for any $f \in \mathscr{F}$ and any design d and estimator δ.

The above statement follows from Theorem 2.1 below. We call this a "stochastic" minimax property for $\mathbf{h}$ for the following reasons.

For any $t \in R$ and any random variable X, define $f_t(X) = P(X > t)$. Observe that for two random variable X and Y, X is usually said to be stochastically at least as large as Y if $f_t(X) \ge f_t(Y)$ for any $t \in R$. Also, it is clear that $\mathscr{F}$ contains any f_t. Therefore the minimax result preserves the genuine spirit of stochastic ordering. When taking f to be the mean functional, our "stochastic" minimaxity result is then reduced to a form with the standard sense of risks for randomized strategies. This standard sense of minimaxity was already explored by Wu with the use of the squared loss that reflects the experimenter's equal interests among all treatment effects (i.e., A-criterion) and the use of the least squares estimates. But Wu further justified balance for treatment replication numbers. We shall take up the same task in Section 3 for more complicated designs.

Now, we shall generalize the above notions and results to other design settings such as block design or higher-way heterogeneity settings. This extension requires only an abstraction of the framework of Example 1.

Let S be a set of possible "true states" of nature, and H be a permutation group on $\{1, \cdots, N\}$. Assume that for any $\pi \in H$ and $s \in S$, πs is well-defined and (2.2) holds. (Implicitly, π will transform only the nuisance parameter part of s such as unit effects, etc.) Replace the model assumption (2.1) by the following:

(2.4) for $\pi \in H$, $d \in D$ and $s \in S$, the yield $\mathbf{y}$ under πs and design πd has a probability measure equal to that of $\pi\mathbf{z}$ where $\mathbf{z}$ is the yield under s and design d.

The loss function should satisfy (2.3). The definition of randomization procedures is the same as before and we take Φ_H to be the class of all randomization procedures. We also require the choice of estimators to be invariant under H. The randomization procedure $\mathbf{h}$ will be referred to as the H-uniform randomization. We have the following.

THEOREM 2.1. *Under* (2.2)–(2.4), *the H-uniform randomization* $\mathbf{h}$ *is minimax in the sense that it minimizes* $\max_{s\in S} f(r(\phi(d, \delta); s))$ *over all randomization procedures* $\phi \in \Phi_H$ *for any* $f \in \mathscr{F}$, $d \in D$ *and estimator* δ.

The proof of this theorem will be given in Section 4. We now present two examples to illustrate the application. These examples were already treated by Wu, but our theorem strengthens his results.

EXAMPLE 2. *Block design setup.* Suppose the $N = \sum_{b=1}^{B} N_b$ units are arranged into b blocks with sizes $N_1, \cdots, N_B$ respectively. Consider the model:

$$y_u = \alpha_{d(u)} + \beta_b + g_u + \varepsilon_u, \quad u = 1, \cdots, N, \tag{2.5}$$

where d is the design, g_u is the uth unit effect, ε_u is the random error, β_b is the bth block effect with unit u in block b. Take H to be the group of all permutations within blocks. The class G of all possible $\mathbf{g}$ and the class $\mathscr{E}$ of all possible error distributions ξ are assumed to be invariant under H. Now take $s = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{g}, \xi)$ where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_B)'$. Observe that $\pi s = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \pi\mathbf{g}, \pi\xi)$; (2.2) and (2.4) are satisfied; (2.3) amounts to claiming that what we estimate depends only on $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ but not on $\mathbf{g}$ or ξ; for any design d, $H(d)$ is the class of designs possessing the same treatment replication numbers for each block as those of d; $\mathbf{h}$ is the procedure of complete randomization within blocks. Applying Theorem 2.1, we obtain a stochastic minimax property for $\mathbf{h}$.

When the block sizes are equal, i.e., $N_1 = \cdots = N_B$, we may consider a larger group H generated by all permutations within blocks and all block permutations. Observe that if $\pi = \pi_1 \cdot \pi_2$ where π_1 is a permutation within blocks and π_2 is a block permutation, then $\pi s = (\boldsymbol{\alpha}, \pi_2\boldsymbol{\beta}, \pi g, \pi\xi)$; (2.3) claims that what we are interested in depends only on $\boldsymbol{\alpha}$ and not on $\boldsymbol{\beta}$, $\mathbf{g}$, or ξ; the H-uniform randomization $\mathbf{h}$ is the procedure of completely randomizing the blocks and the units within each block. Thus if the model is invariant under H, then $\mathbf{h}$ will be a minimax randomization procedure.

EXAMPLE 3. *Two-way heterogeneity setup.* The $N = \ell_1\ell_2$ units are now arranged in an $\ell_1 \times \ell_2$ array. Suppose the model is

$$y_{ij} = \alpha_{d(i,j)} + \beta_i + r_j + g_{ij} + \varepsilon_{ij},\ i = 1, \cdots, \ell_1, j = 1, \cdots, \ell_2, \tag{2.6}$$

where β_i is the ith row effect and r_j is the jth column effect, g_{ij} is the (i, j)th unit effect and ε_{ij} is the random error. Take H to be the group generated by all row permutations and column permutations. Let $s = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{r}, \mathbf{g}, \xi)$ and $\pi s = (\boldsymbol{\alpha}, \pi_1\boldsymbol{\beta}, \pi_2\mathbf{r}, \pi\mathbf{g}, \pi\xi)$ where $\pi = \pi_1 \cdot \pi_2$ and π_1, π_2 are row and column permutations respectively. The H-uniform randomization $\mathbf{h}$ is the procedure of completely randomizing the rows and the columns. Thus our theorem gives a minimaxity for this procedure. It is interesting to note that under the squared loss $\sum_{t,s} \{\alpha_t - \alpha_s - (\hat{\alpha}_t - \hat{\alpha}_s)\}^2$ (A-criterion) and by using the least squares estimates $\{\hat{\alpha}_t\}$, Wu showed that randomly permuting the rows (or the columns) of a Latin square is of the same efficiency as that of permuting both the rows and the columns and treatment numbers. This greatly simplifies the Fisher-Yates "recipe" of randomization procedure for Latin squares. However, if not for the Latin square with the A-criterion, the mean functional, and the least squares estimator, permuting the rows (or the columns) only will typically be inferior to permuting both rows and columns. Thus, the latter is necessary in general cases.

3. Minimax randomized designs under the squared loss. In Section 2, we justify several commonly-used randomization procedures from the viewpoint of robustness against model-violations. However, to which designs should these procedures be actually applied and what estimators should be used are still not solved. By the knowledge of the classical optimal designs, we would expect the solution to be dependent on the criterion used. Thus to successfully attack these problems, in this section we shall merely focus on the situation where the loss function is a squared one and the mean functional is used to assess randomized strategies. Moreover, we shall only consider the linear estimators (but not necessarily the least squares ones). These restrictions are necessary because we need explicit expressions for the risks of randomized strategies. Confining to these and under the H-invariant considerations of Section 2, we shall obtain some general results which reduce the problem of finding a minimax strategy to the classical problem of finding an optimal design under an ideal model. For the cases where the classical optimal design theory has provided a solution, say d^0, our results then provide a minimax strategy $\mathbf{h}(d^0, \delta^0)$, where $\mathbf{h}$ is the H-uniform randomization procedure and δ^0 is the (weighted) least squares estimators; in other words, H-uniformly randomizing a classical optimal design (for suitable criterion) and using the (weighted) least squares estimators is a minimax randomized strategy when the class of possible true models is H-invariant. Specifically, we consider the following setting.

Suppose T treatments are to be assigned to the $N = \sum_{b=1}^{B} N_b$ units which are classified into B blocks, where N_b is the bth block size. Within block b, the $N_b = \prod_{i=1}^{n(b)} \ell_i^{(b)}$ units are arranged according to $n(b)$ factors so that when $n(b) \geq 2$ they form an $n(b)$-dimensional hyper-rectangle of size $\ell_1^{(b)} \times \cdots \times \ell_{n(b)}^{(b)}$, where $\ell_i^{(b)}$ is the number of levels of the ith factor in block b, and when $n(b) = 1$, the $N_b = \ell_1^{(b)}$ units are assumed to be of the same level. To avoid trivialities, we assume $\ell_i^{(b)} \geq 2$. The uth unit, when it falls in block b, is now labeled by $(i_1^{(b)}, \cdots, i_j^{(b)}, \cdots, i_{n(b)}^{(b)})$, where $1 \leq i_j^{(b)} \leq \ell_j^{(b)}$. Assume that

$$y_u = \alpha_{d(u)} + \beta_b + \textstyle\sum_{j=1}^{n(b)} \beta^{(b)}_{(i_1^{(b)}, \ldots, \hat{i}_j^{(b)}, \ldots, i_{n(b)}^{(b)})} + g_u + \varepsilon_u, \tag{3.1}$$

where β_b is the bth block effect and

$$\beta^{(b)}_{(i_1^{(b)}, \ldots, \hat{i}_j^{(b)}, \ldots i_{n(b)}^{(b)})}$$

is the interaction effect of all but the jth factor in block b at levels $i_1^{(b)}, \cdots, i_{j-1}^{(b)}, i_{j+1}^{(b)}, \cdots, i_{n(b)}^{(b)}$ respectively. Note that all the lower level interactions of factors in the same block are implicit in this model. The set G of all possible $\mathbf{g}$ and the set $\mathscr{E}$ of all possible error distributions are assumed to be invariant under a group H to be specified below.

For any b and j such that $1 \leq b \leq B$ and $1 \leq j \leq n(b)$, take $H_j'^{(b)}$ to be a doubly transitive group on $\{1, \cdots, \ell_j^{(b)}\}$; i.e., for any $k_1, k_2, k_3, k_4 \in \{1, \cdots, \ell_j^{(b)}\}$ such that $k_1 \neq k_2$ and $k_3 \neq k_4$, there exists some $\pi' \in G_j'^{(b)}$ such that $\pi'(k_1) = k_3$ and $\pi'(k_2) = k_4$. For any $\pi' \in H_j'^{(b)}$, define a permutation π on $\{1, \cdots, N\}$ by letting $\pi(u) = u$ for $u \notin$ block b, and $\pi(u) =$

$(i_1^{(b)}, \cdots, i_{j-1}^{(b)}, \pi'(i_j^{(b)}), i_{j+1}^{(b)}, \cdots, i_{n(b)}^{(b)})$ when $u = (i_1^{(b)}, \cdots, i_{n(b)}^{(b)})$. Let H_j^b be the group of all such π derived from $\pi' \in H_j'^{(b)}$. It is clear that an element in H_j^b permutes the levels of factor j in block b and any two H_j^b commute. Let H be the group generated by all H_j^b, i.e.,

$$H = \prod_{b=1}^{B} \prod_{j=1}^{n(b)} H_j^b. \tag{3.2}$$

Note that (3.1) is reduced to (2.1) ((2.5); (2.6), respectively) when $B = 1$ and $n(1) = 1$ ($n(b) = 1$ for $1 \le b \le B$; $B = 1$ and $n(1) = 2$, respectively). It is also clear that the H's considered in these examples are of the form (3.2).

We are interested in estimating p contrasts among the treatment effects and want to use a squared loss function and linear estimators only. More explicitly, take $s = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{g}, \xi)$ where $\boldsymbol{\beta}$ is the vector of block and interaction effects, and define

$$\mathscr{L}(s, \mathbf{a}) = (L\boldsymbol{\alpha} - \mathbf{a})'(L\boldsymbol{\alpha} - \mathbf{a}), \tag{3.3}$$

where L is a $p \times T$ matrix with zero row sums, and $\mathbf{a} \in \mathscr{A} = R^p$. Also denote a linear estimator by a $p \times N$ matrix δ and a randomized strategy by $(\mathbf{d}, \boldsymbol{\delta})$; and recall the definition of the risk function r. We shall find a minimax randomized strategy that achieves

$$\min_{(\mathbf{d},\delta)} \max_{s \in S} Er(\mathbf{d}, \boldsymbol{\delta}; s). \tag{3.4}$$

To make risks finite, we assume that $\mathscr{E}$ contains only the probability measure ξ with finite second moments. Since only linear estimators are considered, the risks depends on the covariance matrix $\mathbf{V}$ of ξ. Thus hereafter we replace ξ by $\mathbf{V}$ and let $\mathscr{V}$ be the set of all possible $\mathbf{V}$'s.

The minimax solution of (3.4) is related to the following classical optimal design problem. Set $g_u = 0$ in (3.1). Assume that ε_u's are uncorrelated, with means 0 and known (up to a constant) variances σ_b^2 where b is the label of the block containing unit u. Under such an ideal model and the squared loss (3.3), it is clear that no randomization is necessary. Also, the best linear unbiased estimator (b.l.u.e.) is the weighted least squares one. Using the b.l.u.e., we reduce the problem to finding a design which minimizes

$$\text{trace}[(L, \mathbf{0})\{X' \operatorname{diag}(\sigma_b^{-2})X\}^{-}(L, \mathbf{0})'],$$

where X is the usual design matrix, $\operatorname{diag}(\sigma_b^{-2})$ is the inverse of the covariance matrix of ε_u's, $\mathbf{0}$ is the zero matrix, and A^- denotes any generalized inverse of A. These were called the L- or linear criteria in the optimal design literature. Denote any optimum design under this criterion by d^0. For the case where the σ_b^2's are equal and L is chosen so that (3.3) becomes $\mathscr{L}(s, a) = \sum_{i,j=1} (\alpha_i - \alpha_j - a_i + a_j)^2$, the linear criterion is often called the A-criterion. A-optimal designs have been found in many settings; for example, the balanced block designs (Kiefer, 1958), some group divisible designs in the block design settings (Cheng, 1978a), Generalized Youden designs (G. Y. D.) (Kiefer, 1975). When there exists a control treatment (say, the first treatment is a control), one may want to use an L for which the loss function (3.3) becomes $\mathscr{L}(\mathbf{s}, \mathbf{a}) = \sum_{i=2}^{T} (\alpha_i - \alpha_1 - a_i + a_1)^2$. Several balanced treatment incomplete block designs are found to be optimal under this criterion (Notz, 1981).

The following is a recipe for finding a minimax solution of (3.4): (i) suitably define the σ_b's by some feature of G and $\mathscr{V}$; (ii) find an L-optimal design d^0; (iii) H-uniformly randomize d^0 and use the b.l.u.e. δ^0.

In short, $\mathbf{h}(d^0, \delta^0)$ is minimax. To define σ_b we need some notation. The cardinal number of a set (or a group) A is denoted by $\#A$. For each $b \in \{1, \cdots, B\}$, let $\Lambda^b = \{\gamma \mid \gamma \subset \{1, \cdots, n(b)\}\}$ and define $H_\gamma^b = \prod_{j \in \gamma} H_j^b$ for $\gamma \in \Lambda^b$. For $\mathbf{g} \in R^N$ define $\mathbf{g}^\gamma = (-1)^k \sum_{\pi \in H_\gamma^b} \pi\mathbf{g}/\#H_\gamma^b$, where $k = \#\gamma$. Now let $\mathbf{g}_{[b]}$ be the projection of $\mathbf{g} \in R^N$ on block b, .e., the uth coordinate of $\mathbf{g}_{[b]}$ equals that of $\mathbf{g}$ or 0, depending on whether u falls in block b or not. Then define $\tilde{\mathbf{g}}_{[b]} = \sum_{\gamma \in \Lambda^b} \mathbf{g}_{[b]}^\gamma$. For a $N \times N$ matrix $\mathbf{V}$ with the uth column vector $\mathbf{v}_u$, let $\mathbf{V}_{[b]}^*$ be the $N \times N$ matrix with the uth column vector $\tilde{\mathbf{v}}_{u[b]}$ and define $\tilde{\mathbf{V}}_{[b]} = ((\mathbf{V}_{[b]}^*)')_{[b]}^*$. Finally, define

$$\sigma_b^2 = c_b \cdot \max_{(\mathbf{g},\mathbf{V}) \in G \times \mathscr{V}} \{\| \tilde{\mathbf{g}}_{[b]} \|^2 + \text{trace } \tilde{\mathbf{V}}_{[b]}\}, \qquad b = 1, \cdots, B, \tag{3.5}$$

where $\|\cdot\|$ is the Euclidean norm and

$$c_b = \frac{1}{N_b} \cdot \textstyle\sum_{\gamma\in\Lambda^b} \prod_{j\in\gamma} (\ell_j^{(b)} - 1)^{-1}$$

(the product over an empty set is 1).

THEOREM 3.1. *Suppose* (3.1)–(3.5) *hold. If there exists a* $(\mathbf{g}^0, \mathbf{V}^0) \in G \times \mathscr{V}$ *which simultaneously achieves all the maxima of* (3.5), *then* $\mathbf{h}(d^0, \delta^0)$ *is a randomized strategy which achieves* (3.4).

Now we apply this theorem to the settings of the three examples of Section 3. For the settings of Examples 1 and 3, we only have one block; i.e., $B = 1$. This greatly simplifies the matter. We do not have to verify the existence of $(\mathbf{g}^0, \mathbf{V}^0)$ and the b.l.u.e. is the least squares estimate. Thus in Example 1, a minimax randomized strategy is to completely randomize an L-optimal design and to use the least squares estimator. In particular, when the A-criterion is assumed, the balanced completely randomized design together with the least squares estimator is a minimax strategy. This slightly strengthens Wu's Theorem 1 and Theorem 3 which justified balance as well as randomization but the use of the least squares estimator was assumed. For Example 3, completely randomizing the rows and the columns of a G.Y.D. (whenever existent) together with the use of the least squares estimator is a minimax strategy. In general, when k-way setting is assumed and all higher order interactions are present so that (3.1) holds with $B = 1$, the minimax randomized strategy can be found in a similar way. However, for $k \geq 3$ if the setting does not include the interaction effects, then Theorem 3.1 does not apply. This is demonstrated in the following example.

EXAMPLE 4. Suppose 8 experimental units are classified by 3 factors. Each factor has 2 levels (high and low). Thus each unit can be labeled by (i, j, k) where $i, j, k = 1$ or 2. Suppose there are only two treatments. Instead of (3.1), we consider the following simpler additivity assumption:

$$\mathbf{y}_u = \alpha_{d(u)} + \beta_{1i} + \beta_{2j} + \beta_{3k} + g_u + \varepsilon_u, \tag{3.6}$$

where β_{1i} is the first factor's ith level effect, and β_{2j}, β_{3k} are defined similarly. This model is valid when the interaction effects are known to be negligible. The class G of all possible unit effects and the class $\mathscr{V}$ of all possible covariance matrices for the random errors are again assumed to be invariant under the group H of all the permutations of the factor levels. Since (3.6) is not of the form (3.1), Theorem 3.1 does not apply. In fact, for the A-criterion, the associated classical design problem (i.e., finding an A-optimal design for model (3.6) when $g_u = 0$ and the ε_u's are uncorrelated with a common variance) has the following two solutions d_1 and d_2:

$$d_1: \quad \begin{array}{|c|c|}\hline A & B \\ \hline B & A \\ \hline\end{array} \quad \begin{array}{|c|c|}\hline B & A \\ \hline A & B \\ \hline\end{array} \qquad\qquad d_2: \quad \begin{array}{|c|c|}\hline A & B \\ \hline B & A \\ \hline\end{array} \quad \begin{array}{|c|c|}\hline A & B \\ \hline B & A \\ \hline\end{array}$$

where A and B denote the treatment labels; for each design, units in the first and the second squares are those with the third factor at low level and high level, respectively. If the conclusion of Theorem 3.1 were true, then we would expect that randomly permuting the factor levels for d_1 and d_2 (and using least squares estimators) would yield the same maximum risks for any H-invariant G and $\mathscr{V}$ because they should be both minimax. However, the following two special G's disprove this assertion:

$$\text{(i)} \quad G_1 = \left\{ \left(\begin{array}{|c|c|}\hline 1 & 0 \\ \hline 0 & 1 \\ \hline\end{array}, \begin{array}{|c|c|}\hline 1 & 0 \\ \hline 0 & 1 \\ \hline\end{array} \right), \left(\begin{array}{|c|c|}\hline 0 & 1 \\ \hline 1 & 0 \\ \hline\end{array}, \begin{array}{|c|c|}\hline 0 & 1 \\ \hline 1 & 0 \\ \hline\end{array} \right) \right\}$$

$$\text{(ii)} \quad G_2 = \left\{ \left(\begin{array}{|c|c|}\hline 1 & 0 \\ \hline 0 & 1 \\ \hline\end{array}, \begin{array}{|c|c|}\hline 0 & 1 \\ \hline 1 & 0 \\ \hline\end{array} \right), \left(\begin{array}{|c|c|}\hline 0 & 1 \\ \hline 1 & 0 \\ \hline\end{array}, \begin{array}{|c|c|}\hline 1 & 0 \\ \hline 0 & 1 \\ \hline\end{array} \right) \right\}$$

Both G_1 and G_2 are H-invariant. To simplify the computation, assume that the random errors vanish. Consider the case $G = G_1$. It is easy to find that the maximum mean squared error for the strategy of H-uniformly randomizing d_1(respectively d_2) and using the least squares estimator is equal to 0 (respectively 1). Thus d_1 performs better than d_2 when $G = G_1$. However, similar observation leads to the opposite conclusion (i.e., d_2 is better than d_1) when $G = G_2$. Intuitively, d_1 should always be more desirable than d_2 since it possesses better symmetry properties. However, the above discussion shows that this intuition may sometimes be misleading. This also demonstrates the importance of Wu's rigorous treatment on the justification of the role of randomization from the model-robustness viewpoint although it seems to have been well accepted.

For the k-way ($k \geq 3$) settings without interaction effects, Cheng (1978b, 1980) showed that the Youden hyperrectangles (Y.H.R.) are A-optimal. (The two designs d_1, d_2 in Example 4 are both Y.H.R.'s). However, due to the above consideration, H-uniformly randomizing a Y.H.R. does not necessarily provide a minimax strategy. Moreover, the actual minimax strategy may depend on the actual form of G. This then creates many difficulties in finding a solution and we have no satisfactory answer yet.

We turn to the block design settings of Example 2. The H here contains all permutations within each block. It is clear that

$$\tilde{\mathbf{g}}_{[b]} = \mathbf{g}_{[b]} - \frac{1}{N_b} \cdot \left(\textstyle\sum_{u \in \text{block}\, b} g_u \right) \mathbf{1}_{[b]}$$

where $\mathbf{1}$ is the vector of ones, $c_b = 1/(N_b - 1)$, and with $\mathbf{V} = (v_{uu'}) \in \mathscr{V}$,

$$\text{trace}\, \tilde{\mathbf{V}}_{[b]} = \textstyle\sum_{u \in \text{block}\, b} v_{uu} - \frac{1}{N_b} \sum_{u,u' \in \text{block}\, b} v_{uu'}.$$

Thus for any specified G, it is not hard to actually compute σ_b^2. But the conclusion of Theorem 3.1 may or may not be true, depending on whether the σ_b^2's are achieved by a common $(\mathbf{g}^0, \mathbf{V}^0)$ or not. If we take, for instance,

$$G = G_1 = \{\mathbf{g} : \|\mathbf{g}\| \leq k\},\ G = G_2 = \{\mathbf{g} : |g_u| \leq k \text{ for each } u\},$$

or

$$G = G_3 = \{\mathbf{g} \,|\, \text{for each } b, \textstyle\sum_{u \in \text{block}\, b} g_u = 0 \quad \text{and} \quad |g_u| \leq k_b \text{ for } u \in \text{block } b\},$$

where k and the k_b's are constants, then the σ_b^2's can be achieved by a $\mathbf{g}^0$. Thus a minimax solution can be found by the general recipe given before. In particular, consider the case where block sizes are equal, $\mathscr{V}$ contains only the identity matrix, and A-criterion is desired. If the σ_b^2's are equal (for instance $G = G_1$, $G = G_2$, or $G = G_3$ with equal k_b's), then randomizing the units within each block of an A-optimal block design (e.g. a B.B.D.) is minimax. Note that in such cases we do not need to randomize the blocks. However, if we take, for instance,

$$G = G_4 = \cup_{b=1}^{B} \{\mathbf{g} \,|\, |g_u| \leq k \text{ for } u \in \text{block } b \text{ and } g_u = 0 \text{ elsewhere}\},$$

then for different blocks their σ_b^2's are achieved by different $\mathbf{g}$'s. For such a case, only randomizing the units within each block is not minimax. The common sense suggests that one should randomize both the blocks and the units within each block. But to justify this, we need to consider a larger transformation group H^0 which contains both the original group H and a group H_c that permutes the blocks. This enlarged group H^0 cannot be represented as the form of (3.2) because H_c and H do not commute. Therefore we need a different theorem to handle this case. The following development is mostly motivated by this consideration.

Suppose that $N_1 = N_2 = \cdots = N_B$, $n(1) = n(2) = \cdots = n(B) = n$, and $\ell_j^{(1)} = \ell_j^{(2)} = \cdots = \ell_j^{(B)} = \ell_j$, $j = 1, 2, \cdots, n$. Take a transitive group H_c' on $\{1, \cdots, B\}$; i.e., for any $k_1, k_2 \in \{1, \cdots, B\}$, there exists some $\pi' \in H_c'$ such that $\pi'(k_1) = k_2$. For any $\pi' \in H_c'$, define a permutation π on $\{1, \cdots, N\}$ by $\pi(i_1^{(b)}, \cdots, i_n^{(b)}) = (i_1^{(\pi'(b))}, \cdots, i_n^{(\pi'(b))})$; clearly, π is a block permutation. Let H_c be the group of all such π's. Now define H^0 to be a group generated by the H of (3.2) and H_c. Denote the H^0-uniform randomization by $\mathbf{h}^0$.

THEOREM 3.2. *Suppose* (3.1) *and* (3.3) *hold. If G and* $\mathscr{V}$ *are invariant under* H^0, *then* $\mathbf{h}^0(d^0, \delta^0)$ *achieves* (3.4) *where* d^0 *is an L-optimal design* (*defined before*) *with* $\sigma_1^2 = \cdots = \sigma_B^2$, *and* δ^0 *is the least squares estimator.*

Applying this theorem to the block design setups of Example 2 with equal block sizes, we see that applying the standard randomization procedure to an A-optimal design and using the least squares estimator is minimax. Moreover, it suggests a simpler randomization procedure; i.e., instead of completely randomly permuting the blocks we may just randomly rotate the blocks. This is because we may take H_c' to be a cyclic group. This simplified procedure is not only easier to implement but also enjoys at least as many robustness properties as the standard one when the squared loss is concerned. To see this, we simply observe that if a $(G, \mathscr{V})$ is invariant under the group of all block permutations and the permutations within each block, then it is also invariant under a smaller group that contains only a cyclic group of block permutations and any permutations within each block. Similar argument also applies to the cases covered by Theorem 3.1 and we conclude that randomization procedure generated by a doubly transitive group is as good as the complete randomization. But in general a doubly transitive group with a simple form and a small order is not easy to obtain. For some particularly simple cases (e.g., a product of two cyclic groups may sometimes be doubly transitive), see Burnside (1911).

4. Proofs. To save space, some of the proofs are only outlined. For details, see Li (1981).

PROOF OF THEOREM 2.1. It is clear that we need only to show that for any $f \in \mathscr{F}$ and $s_0 \in S$,

$$\min_\phi \max_{\pi\in H} f(r(\phi(d, \delta); \pi s_0)) = \max_{\pi\in H} f(r(\mathbf{h}(d, \delta); \pi s_0)). \tag{4.1}$$

Observe that for any $\pi \in H$, $r(\pi d, \pi\delta; \pi s_0) = r(d, \delta; s_0)$; this is due to the invariance properties of the model (2.4), the estimator and the loss function (2.3). Define a real function l on H by $l(\pi) = r(\pi d, \pi\delta; s_0)$. Then,

$$r(\phi(d, \delta); \pi s_0) = r(\pi^{-1}\phi(d, \delta); s_0) = l(\pi^{-1}\phi).$$

We may write (4.1) as $\min_\phi \max_{\pi\in H} f(l(\pi\phi)) = \max_{\pi\in H} f(l(\pi\mathbf{h}))$; equivalently,

$$\min_{\mu\in\mathscr{M}} \max_{\pi\in H} f(l(\pi\mu)) = \max_{\pi\in H} f(l(\pi\mu_0)), \tag{4.2}$$

where μ_0 is the uniform distribution on H, $\mathscr{M}$ is the class of all probability measures on H, and $l(\pi\mu)$ is the distribution of $l(\pi\phi))$ when ϕ has probability measure μ.

Obviously, for any μ,

$$\frac{1}{\#H}\sum_{\pi\in H} \pi\mu = \mu_0 \quad \text{and} \quad \frac{1}{\#H}\sum_{\pi\in H} l(\pi\mu) = l(\mu_0).$$

Hence for any $\mu \in \mathscr{M}$ and $f \in \mathscr{F}$, $f(l(\mu_0)) \le \max_{\pi\in H} f(l(\pi\mu))$ by the definition of $\mathscr{F}$.

Therefore (4.2) holds since $\pi\mu_0 = \mu_0$ for any $\pi \in H$. □

PROOF OF THEOREM 3.1. To proceed, a sequence of lemmas will be presented first.

LEMMA 4.1. *Under* (3.1)–(3.3), *we have*

$$\min_{(\mathbf{d},\delta)} \max_{s\in S} Er(\mathbf{d}, \boldsymbol{\delta}; s) \ge \max_{s\in S} \min_{(d,\delta)} Er(\mathbf{h}(d, \delta); s). \tag{4.3}$$

Roughly speaking, this lemma suggests that in order to find a randomized strategy achieving (3.4) we may first choose a suitable non-randomized strategy (d, δ); then we apply the H-uniform randomization on the design d, and use the estimator accordingly. We now proceed to evaluate $Er(\mathbf{h}(d, \delta); s)$. Write $\mathbf{Y} = (y_1, \cdots, y_N)'$ and $\boldsymbol{\alpha}_d = (\alpha_{d(1)}, \cdots, \alpha_{d(N)})'$; let $\mathscr{B}$ be $E\mathbf{Y}$ under $\boldsymbol{\alpha} = 0$ and $\mathbf{g} = 0$ ($\mathscr{B}$ depends only on $\boldsymbol{\beta}$); for any $\mathbf{g} \in R^N$, define $\bar{\mathbf{g}} = \sum_{\pi \in H} \pi\mathbf{g}/\#H$; for any symmetric matrix $\mathbf{V}$, define

$$\bar{\mathbf{V}} = \frac{1}{\#H} \sum_{\pi \in H} \pi(\mathbf{V}) \text{ where } \pi(\mathbf{V}) = (v_{\pi^{-1}(i)\pi^{-1}(j)}) \text{ for } \mathbf{V} = (v_{ij}).$$

LEMMA 4.2. *For any* (d, δ),

$$Er(\mathbf{h}(d, \delta); s) = \| L\boldsymbol{\alpha} - \delta(\boldsymbol{\alpha}_d + \mathscr{B} + \bar{\mathbf{g}}) \|^2 + \operatorname{trace}\delta(\overline{(\mathbf{g} - \bar{\mathbf{g}})(\mathbf{g} - \bar{\mathbf{g}})'} + \bar{\mathbf{V}})\delta',$$

where $s = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{g}, \mathbf{V})$.

To avoid having an infinite maximum risk, it is necessary for the estimator to satisfy an unbiasedness condition when $\mathbf{g} = 0$. This will be made explicit by the notion of orbits. The orbit of an element u under a group K is the set $\{\pi(u) \mid \pi \in K\}$ and is denoted by $K(u)$. For any subset $A \subset \{1, \cdots, N\}$, define $\mathbf{1}_A$ to be the vector in R^N with the ith coordinate equal to 1 or 0 depending on whether $i \in A$ or not. Thus for $n(b) \geq 2$, $1 \leq j \leq n(b)$ and $u = (i_1, \cdots, i_{n(b)})$, $\mathbf{1}_{H_j^b(u)}$ is the expectation of $\mathbf{Y}$ when $\beta^{(b)}_{(i_1, \cdots, \hat{i}_j, \cdots, i_{n(b)})} = 1$ and all other parameter values are 0. Similarly, for $n(b) = 1$, $\mathbf{1}_{H_1^b(u)}$ is the expectation of $\mathbf{Y}$ when $\beta_b = 1$ and all other parameter values are 0. Let

$$W = \{\mathbf{a} \mid \mathbf{a} \in R^N, \mathbf{a}'\mathbf{1}_{H_j^b(u)} = 0, u \in \text{block } b, \quad b = 1, \cdots, B \text{ and } j = 1, \cdots, n(b)\}.$$

LEMMA 4.3. *A necessary condition for* $\max_{s \in S} Er(\mathbf{h}(d, \delta); s)$ *to be finite is that each column vector of* δ' *belongs to* W *and* $\delta\boldsymbol{\alpha}_d = L\boldsymbol{\alpha}$ *for all* $\boldsymbol{\alpha} \in R^T$.

Let U be the class of any non-randomized strategies (d, δ) such that δ satisfies the necessary condition in Lemma 4.3. Without loss of generality, we may restrict the randomized strategies to have supports on U. The following lemma will be used to simplify $Er(\mathbf{h}(d, \delta), s)$. Recall some notations from Section 3.

LEMMA 4.4. *For any* $\mathbf{a} \in W$, $\mathbf{g} \in R^N$, *and any symmetric* $N \times N$ *matrix* $\mathbf{V}$, *we have*

$$\mathbf{a}'\bar{\mathbf{g}} = 0, \tag{4.4}$$

$$\mathbf{a}'\overline{(\mathbf{g} - \bar{\mathbf{g}})(\mathbf{g} - \bar{\mathbf{g}})'}\mathbf{a} = \mathbf{a}'\overline{\mathbf{g}\mathbf{g}'}\mathbf{a}, \tag{4.5}$$

$$\mathbf{a}'\overline{\mathbf{g}\mathbf{g}'}\, \mathbf{a} = \textstyle\sum_{b=1}^{B} c_b \| \tilde{\mathbf{g}}_{[b]} \|^2 \cdot \| \mathbf{a}_{[b]} \|^2, \tag{4.6}$$

and

$$\mathbf{a}'\bar{\mathbf{V}}\mathbf{a} = \textstyle\sum_{b=1}^{B} c_b \operatorname{trace} \tilde{\mathbf{V}}_{[b]} \cdot \| \mathbf{a}_{[b]} \|^2. \tag{4.7}$$

We now combine these lemmas to complete the proof of Theorem 3.1.

First, by Lemmas 4.2, 4.3 and 4.4, we obtain

$$Er(\mathbf{h}(d, \delta); s) = \textstyle\sum_{b=1}^{B} c_b(\| \tilde{\mathbf{g}}_{[b]} \|^2 + \operatorname{trace} \tilde{\mathbf{V}}_{[b]}) \cdot \operatorname{trace} \delta_{[b]}(\delta_{[b]})' \tag{4.8}$$

where $\delta_{[b]}$ is a $p \times N_b$ matrix with each row vector equal to the component of the corresponding row vector of δ on the block b.

Next, by Lemma 4.1, (4.8) and the assumptions of Theorem 3.1, we have

$$\begin{aligned}
\min_{(\mathbf{d}, \delta)} \max_{s \in S} Er(\mathbf{d}, \boldsymbol{\delta}; s) &\geq \max_{s \in S} \min_{(d, \delta)} Er(\mathbf{h}(d, \delta); s) \\
&\geq \max_{s \in S} \min_{(d, \delta) \in U} \textstyle\sum_{b=1}^{B} c_b(\| \tilde{\mathbf{g}}_{[b]} \|^2 + \operatorname{trace} \tilde{\mathbf{V}}_{[b]}) \cdot \operatorname{trace}(\delta_{[b]}\delta'_{[b]}) \\
&\geq \min_{(d, \delta) \in U} \textstyle\sum_{b=1}^{B} c_b(\| \tilde{\mathbf{g}}^0_{[b]} \|^2 + \operatorname{trace} \tilde{\mathbf{V}}^0_{[b]}) \cdot \operatorname{trace}(\delta_{[b]}\delta'_{[b]}) \\
&= \min_{(d, \delta) \in U} \textstyle\sum_{b=1}^{B} \sigma_b^2 \operatorname{trace}(\delta_{[b]}\delta'_{[b]}).
\end{aligned}$$

By the definition of (d^0, δ^0), we see that the above minimum is achieved by (d^0, δ^0). On the other hand, for any $s \in S$,

$$Er(\mathbf{h}(d^0, \delta^0); s) \le \textstyle\sum_{b=1}^{B} \sigma_b^2 \operatorname{trace}(\delta_{[b]}^0 \delta_{[b]}^{0\prime}).$$

Therefore, $\max_{s\in S} Er(\mathbf{h}(d^0, \delta^0); s) = \min_{(\mathbf{d},\delta)} \max_{s\in S} Er(\mathbf{d}, \boldsymbol{\delta}; s)$ and Theorem 3.1 is proved.
It remains to establish Lemmas 4.1–4.4.

Proof of Lemma 4.1. For any $s^* \in S$ and any $(\mathbf{d}, \boldsymbol{\delta})$, if $\mathbf{d}$ has probability measure μ,

$$\max_{s\in S} Er(\mathbf{d}, \boldsymbol{\delta}; s) \ge \max_{\pi\in H} Er(\mathbf{d}, \boldsymbol{\delta}; \pi^{-1}s^*) \ge \frac{1}{\#H} \textstyle\sum_{\pi\in H} Er(\mathbf{d}, \boldsymbol{\delta}; \pi^{-1}s^*)$$

$$= \textstyle\sum_{(d,\delta)} \mu(d) \cdot \sum_{\pi\in H} r(\pi d, \pi\delta; s^*)/\#H$$

$$= \textstyle\sum_{(d,\delta)} \mu(d) \cdot Er(\mathbf{h}(d, \delta); s^*) \ge \min_{(d,\delta)} Er(\mathbf{h}(d, \delta); s^*).$$

It follows that $\min_{(d,\delta)} \max_{s\in S} Er(\mathbf{d}, \boldsymbol{\delta}; s) \ge \min_{(d,\delta)} Er(\mathbf{h}(d, \delta); s^*)$ for any $s^* \in S$. Hence the lemma holds. □

Proof of Lemma 4.2. By (3.1), (3.3), we have

$$\begin{aligned} Er(\mathbf{h}(d, \delta); s) &= Er(d, \delta; \mathbf{h}^{-1}s) \\ &= E[\| L\alpha - \delta \cdot (\alpha_d + \mathscr{B} + \mathbf{h}^{-1}\mathbf{g}) \|^2 + \operatorname{trace}\{\delta \mathbf{h}^{-1}(\mathbf{V})\delta'\}] \\ &= \| L\alpha - \delta \cdot (\alpha_d + \mathscr{B} + E\mathbf{h}^{-1}\mathbf{g}) \|^2 + \operatorname{trace}\{\delta \operatorname{Cov}(\mathbf{h}^{-1}\mathbf{g})\delta'\} + \operatorname{trace}(\delta\bar{\mathbf{V}}\delta') \\ &= \| L\alpha - \delta \cdot (\alpha_d + \mathscr{B} + \bar{\mathbf{g}}) \|^2 + \operatorname{trace}\{\delta(\mathbf{g} - \bar{\mathbf{g}})(\mathbf{g} - \bar{\mathbf{g}})'\delta'\} + \operatorname{trace}(\delta\bar{\mathbf{V}}\delta'). \square \end{aligned}$$

Proof of Lemma 4.3. In view of Lemma 4.2, the proof becomes straightforward. □

Proof of Lemma 4.4. Observe that

$$\pi\bar{\mathbf{g}} = \bar{\mathbf{g}} \quad \text{for any} \quad \pi \in H. \tag{4.9}$$

Take π to be any permutation in H_j^b and we see that $\bar{\mathbf{g}}$ is a constant for any coordinate in $H_j^b(u)$ where $u \in$ block b. By the definition of W, we see that (4.4) holds.
Next, to verify (4.5), it suffices to show $\mathbf{a}'\bar{\mathbf{g}}\,\overline{\mathbf{g}'}\mathbf{a} = 0 = \mathbf{a}'\overline{\mathbf{g}}\,\bar{\mathbf{g}}'\mathbf{a}$. Now,

$$\mathbf{a}'\overline{\bar{\mathbf{g}}\mathbf{g}'}\mathbf{a} = \frac{1}{\#H} \textstyle\sum_{\pi\in H} \mathbf{a}'\pi\bar{\mathbf{g}}(\pi\mathbf{g})'\mathbf{a} = \frac{1}{\#H} \sum_{\pi\in H} \mathbf{a}'\bar{\mathbf{g}}(\pi\mathbf{g})'\mathbf{a} = 0,$$

where the second equality holds by (4.9) and the last equality is due to (4.4). A similar argument completes the proof of (4.5).
Next, we need some lemmas to prove (4.6) and (4.7).

Lemma 4.5. *For any* $\mathbf{a} \in W$, $\mathbf{a}'\overline{\mathbf{g}\mathbf{g}'}\mathbf{a} = \sum_{b=1}^{B} \mathbf{a}_{[b]}'\overline{\mathbf{g}_{[b]}\mathbf{g}_{[b]}'}\mathbf{a}_{[b]}$.

This lemma suggests that we may assume $B = 1$ in proving (4.6) without any loss of generality. Thus we delete the block label b from all the notations hereafter (e.g., $H_i = H_j^b$, $\tilde{\mathbf{g}} = \tilde{\mathbf{g}}_{[b]}$, $n(b) = n$, $H_\gamma^b = H_\gamma$, etc.)

Lemma 4.6. *For* $\gamma \ne \varnothing$, $(\tilde{\mathbf{g}})^\gamma = 0$.

Lemma 4.7. *For any* $\mathbf{a} \in W$, $\mathbf{a}'\overline{\mathbf{g}\mathbf{g}'}\mathbf{a} = \mathbf{a}'\overline{\tilde{\mathbf{g}}\tilde{\mathbf{g}}'}\mathbf{a}$.

In view of Lemma 4.6 and Lemma 4.7, to establish (4.6), we may assume $\mathbf{g}^\gamma = 0$ for $\gamma \ne \varnothing$ and show that $\mathbf{a}'\overline{\mathbf{g}\mathbf{g}'}\mathbf{a} = c \,\| \mathbf{g} \|^2 \,\| \mathbf{a} \|^2$ where $c = \sum_\gamma \prod_{j\in\gamma}(\ell_j - 1)^{-1}/N$. Thus, fixing u

$\in [1, \cdots, N\}$, denoting the uth column of $\mathbf{gg}'$ by $\mathbf{w} = (w_1, \cdots, w_N)'$ and writing $\mathbf{a} = (a_1, \cdots, a_N)'$, it suffices to show that

$$\mathbf{w}'\mathbf{a} = c \, \| \mathbf{g} \|^2 a_u. \tag{4.10}$$

To proceed, define $U_\gamma = H_\gamma(u) - \cup_{j\in\gamma} H_{\gamma-\{j\}}(u)$ for $\gamma \neq \varnothing$. Evidently,

$$\mathbf{1}_{U_\gamma \cup \{j\}} = \textstyle\sum_{u' \in U_\gamma} \mathbf{1}_{H_j(u')} - \mathbf{1}_{U_\gamma}, \tag{4.11}$$

and

$$U_\gamma \cap U_{\gamma'} = \varnothing \quad \text{for any} \quad \gamma \neq \gamma'. \tag{4.12}$$

The following lemma is crucial to get (4.10).

LEMMA 4.8. *For any $u', u'' \in U_\gamma$, we have $w_{u'} = w_{u''}$. This constant equals $(-1)^k N^{-1} \| \mathbf{g} \|^2 \prod_{j\in\gamma} (\ell_j - 1)^{-1}$, where $k = \#\gamma$. In particular, we have $w_u = N^{-1} \| \mathbf{g} \|^2$.*

Write $\lambda_\gamma = (-1)^k N^{-1} \| \mathbf{g} \|^2 \prod_{j\in\gamma} (\ell_j - 1)^{-1}$. By (4.12) and Lemma 4.8, we get $\mathbf{w}'\mathbf{a} = \sum_{\gamma\neq\varnothing} \lambda_\gamma \mathbf{1}'_{U_\gamma}\mathbf{a} + N^{-1} \| \mathbf{g} \|^2 a_u$. Hence it suffices to show that $\mathbf{1}'_{U_\gamma}\mathbf{a} = (-1)^k a_u$, where $k = \#\gamma$. This will be proved by mathematical induction. When $\gamma = \{j\}$, by the definition of W, $0 = \mathbf{1}'_{H_j(u)}\mathbf{a} = \mathbf{1}'_{U_\gamma}\mathbf{a} + \alpha a_u$. Thus our assertion is true for $k = 1$. Suppose it is true for $\#\gamma = k$. For $j \notin \gamma$, by (4.11) and the induction hypothesis we have

$$\mathbf{1}'_{U_\gamma \cup \{j\}}\mathbf{a} = \textstyle\sum_{u' \in U_\gamma} \mathbf{1}'_{H_j(u')}\mathbf{a} - \mathbf{1}'_{U_\gamma}\mathbf{a} = 0 - \mathbf{1}'_{U_\gamma}\mathbf{a} = (-1)^{k+1} a_u.$$

Therefore (4.10) is established and so is (4.6).

Turning now to (4.7), let $\mathbf{V}^{1/2}$ be any square root of $\mathbf{V}$. Denote the uth column vector of $\mathbf{V}^{1/2}$ by $\mathbf{e}_u$. By (4.6), we have

$$\mathbf{a}'\bar{\mathbf{V}}\mathbf{a} = \mathbf{a}' \overline{\textstyle\sum_{u=1}^N \mathbf{e}_u\mathbf{e}'_u}\mathbf{a} = \sum_{u=1}^N \mathbf{a}'\overline{\mathbf{e}_u\mathbf{e}'_u}\mathbf{a} = \sum_{b=1}^B \sum_{u=1}^N c_b \, \| \tilde{\mathbf{e}}_{u[b]} \|^2 \cdot \mathbf{a}'_{[b]}\mathbf{a}_{[b]}.$$

Thus, it suffices to show that trace $\tilde{\mathbf{V}}_{[b]} = \sum_{u=1}^N \| \tilde{\mathbf{e}}_{u[b]} \|^2$. Let A_b be the $N \times N$ matrix such that $A_b\mathbf{g} = \tilde{\mathbf{g}}_{[b]}$ for any $\mathbf{g}$. Now we have

$$\begin{aligned}\sum_{u=1}^N \| \tilde{\mathbf{e}}_{u[b]} \|^2 &= \text{trace} \sum_{u=1}^N \tilde{\mathbf{e}}_{u[b]}\tilde{\mathbf{e}}'_{u[b]} = \text{trace} \sum_{u=1}^N A_b\mathbf{e}_u\mathbf{e}'_u A'_b = \text{trace}\, A_b\mathbf{V}A'_b \\ &= \text{trace}\, \mathbf{V}^*_{[b]}A'_b = \text{trace}\, A_b(\mathbf{V}^*_{[b]})' = \text{trace}\, \tilde{\mathbf{V}}_{[b]}.\end{aligned}$$

Hence (4.7) is established. The proof of Lemma 4.4 is complete. □

PROOF OF LEMMA 4.5. An argument similar to that used in proving (4.4) leads to the conclusion that for any u, u' in different blocks b, b', the (u, u')th cell of $\mathbf{gg}'$ equals the $(u, \pi(u'))$th cell, for any $\pi \in H_j^{b'}$, $j = 1, \cdots, n(b')$. In view of this and by the definition of W, Lemma 4.5 follows. □

PROOF OF LEMMA 4.6. This can be verified by using mathematical induction and observing the following two facts:

$$(\mathbf{g}_1 + \mathbf{g}_2)^\gamma = \mathbf{g}_1^\gamma + \mathbf{g}_2^\gamma \quad \text{for any} \quad \mathbf{g}_1, \mathbf{g}_2 \in R^N; \tag{4.13}$$

$$\begin{aligned}(\mathbf{g}^\gamma)^{\{j\}} &= \mathbf{g}^{\gamma\cup\{j\}} \qquad \text{if} \quad j \notin \gamma, \\ &= -\mathbf{g}^\gamma \qquad \text{if} \quad j \in \gamma.\end{aligned} \tag{4.14}$$

□

PROOF OF LEMMA 4.7. Write $\tilde{\mathbf{g}} = \mathbf{g} + \sum_{\gamma\neq\varnothing} \mathbf{g}^\gamma$ and compare both sides of the equation to be established in Lemma 4.7. It suffices to show that $\mathbf{a}'\pi\mathbf{g}^\gamma = 0$ for any $\pi \in H$, any $\gamma \neq \varnothing$ and any $\mathbf{a} \in W$. Since W is invariant under H, we may only show $\mathbf{a}'\mathbf{g}^\gamma = 0$ for any $\mathbf{a} \in W$. Now, this assertion can be verified by using mathematical induction, (4.14), and the definition of W. □

PROOF OF LEMMA 4.8. Since each H_j is doubly transitive, it is clear that for any u' and $u'' \in U_{\{j\}}$ (which implies $u \neq u'$ and $u \neq u''$) there exists a π such that $\pi u = u$, $\pi u' = u''$. Now, since $\mathbf{gg}' = \pi\mathbf{gg}'\pi$, we get $w_{u'} = w_{u''}$ by comparing the (u, u')th cells of these two matrices. Thus the first statement is proved for $\#\gamma = 1$. For general γ, the proof is similar. We now compute the constant for each U_γ by mathematical induction.

First, the diagonal elements of $\mathbf{gg}'$ are the same due to the transitivity of the group $\prod_{j=1}^n H_j$. This constant is easily verified to be $N^{-1}\|\mathbf{g}\|^2$, by considering the trace of $\mathbf{gg}'$. Hence we have shown $w_u = N^{-1}\|\mathbf{g}\|^2$.

Next, let $\mathbf{z} = (z_1, \cdots, z_n)' = \mathbf{hg}$. Clearly, $E\mathbf{zz}' = \mathbf{gg}'$ and $w_{u'} = Ez_u z_{u'}$. By the assumption that $\mathbf{g}^{\{j\}} = 0$, it follows that $\sum_{\pi\in H_j} z_{\pi(u)} = 0$. Thus $-z_u^2 = \sum_{u'\in U_{\{j\}}} z_u z_{u'}$. Taking the expectations on both sides, we obtain $-w_u = \sum_{u'\in U_{\{j\}}} w_{u'} = (\ell_j - 1)w_{u'}$. Hence $w_{u'} = -(\ell_j - 1)^{-1}\cdot N^{-1}\cdot\|\mathbf{g}\|^2$, as desired.

Next, suppose that our lemma is true for some γ. We shall find the desired constant for $\gamma\cup\{j\}$ where $j \not\in \gamma$. For any $u' \in U_{\gamma\cup\{j\}}$, there exist some $\pi_1 \in H_\gamma$ and some $\pi_2 \in H_j$ such that $\pi_1(u) \in U_\gamma$ and $\pi_2\pi_1(u) = u'$. Now, $\mathbf{g}^{\{j\}} = 0$ implies $\sum_{\pi\in H_j} z_{\pi\pi_1(u)} = 0$ and thus we get $z_u z_{\pi_1(u)} = -\sum_{\pi\in H_j - \{\mathscr{I}\}} z_u z_{\pi\pi_1(u)}$, where $\mathscr{I}$ is the identity permutation. Observe that $\pi\pi_1(u) \in U_{\gamma\cup\{j\}}$ for any $\pi \in H_j - \{\mathscr{I}\}$ and take the expectations on both sides of the last equality. It follows that $w_{\pi_1(u)} = -(\ell_j - 1)w_{u'}$. Hence

$$w_{u'} = -(\ell_j - 1)^{-1}w_{\pi_1(u)} = (-1)^{k+1}N^{-1}\|\mathbf{g}\|^2\prod_{i\in\gamma\cup\{j\}}(\ell_i - 1)^{-1},$$

where the last equality is due to the induction hypothesis. The proof for Lemma 4.8 is now complete. □

PROOF OF THEOREM 3.2. Define

$$\mathbf{g}^0 = \frac{1}{\#H^0}\sum_{\pi\in H^0}\pi\mathbf{g} \quad\text{and}\quad \mathbf{V}^0 = \frac{1}{\#H^0}\sum_{\pi\in H^0}\pi(\mathbf{V}).$$

By arguments similar to those in the proof of Lemmas 4.1–4.3, we see that

$$Er(\mathbf{h}^0(d, \delta); s) = \|\delta\mathbf{g}^0\|^2 + \operatorname{trace}[\delta\{(\mathbf{g} - \mathbf{g}^0)(\mathbf{g} - \mathbf{g}^0)' + \mathbf{V}\}^0\delta'], \tag{4.15}$$

for (d, δ) such that the maximum risk is finite. Write $\mathbf{M} = (\mathbf{g} - \mathbf{g}^0)(\mathbf{g} - \mathbf{g}^0)' + \mathbf{V}$. We claim that for any $\mathbf{a} \in W$,

$$\mathbf{a}'\mathbf{g}^0 = 0 \tag{4.16}$$

and

$$\mathbf{a}'\mathbf{M}^0\mathbf{a} = \lambda\|\mathbf{a}\|^2 \tag{4.17}$$

for a constant λ depending on $\mathbf{M}$ but not on $\mathbf{a}$.

First, observe that

$$\mathbf{g}^0 = \frac{1}{\#H_c}\sum_{\pi\in H_c}\pi\bar{\mathbf{g}}.$$

Then (4.16) follows easily from (4.4). Next,

$$\begin{aligned}
\mathbf{a}'\mathbf{M}^0\mathbf{a} &= \frac{1}{\#H_c}\sum_{\pi\in H_c}\mathbf{a}'\pi\bar{\mathbf{M}}\pi^{-1}\mathbf{a} \qquad \text{(where } \pi \text{ is treated as a matrix)}\\
&= \frac{1}{\#H_c}\sum_{\pi\in H_c}\sum_{b=1}^B c_b \operatorname{trace}\tilde{\mathbf{M}}_{[b]}\cdot\|(\pi^{-1}\mathbf{a})_{[b]}\|^2 \qquad \text{(by (4.7))}\\
&= \sum_{b=1}^B c_b \operatorname{trace}\tilde{\mathbf{M}}_{[b]}\cdot\left(\frac{1}{\#H_c}\sum_{\pi\in H_c}\|\mathbf{a}_{[\pi^{-1}(b)]}\|^2\right)\\
&= \sum_{b=1}^B c_b \operatorname{trace}\tilde{\mathbf{M}}_{[b]}\cdot(B^{-1}\|\mathbf{a}\|^2)\\
&= \left(B^{-1}\sum_{b=1}^B c_b \operatorname{trace}\tilde{\mathbf{M}}_{[b]}\right)\cdot\|\mathbf{a}\|^2 \equiv \lambda\cdot\|\mathbf{a}\|^2,
\end{aligned}$$

where the fourth equality is due to the transitivity of H'_c. Thus (4.17) holds. From (4.15)-(4.17), we get

$$Er(\mathbf{h}^0(d, \delta); s) = \lambda(s)\mathrm{trace}(\delta\delta'),$$

where $\lambda(s)$ is some positive constant depending on s. The rest of the proof is similar to that of Theorem 3.1.

Acknowledgments. This paper is based on parts of my Ph.D. thesis. I am indebted to my advisor, Professor Jack Kiefer, for guidance, encouragement, and many valuable suggestions. I sincerely thank an associate editor, a referee, Professor C. S. Cheng and Professor C. F. Wu for providing me with numerous useful comments which greatly improved the presentation and quality of this paper. Finally, this paper is dedicated to the memory of Professor Jack Kiefer, a great teacher, advisor, and friend.

REFERENCES

BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.

BECHHOFER, R. E. and TAMHANE, A. C. (1981). Incomplete block designs for comparing treatments with a control: general theory. *Technometrics* **23** 45–58.

BOX, G. E. P. and DRAPER, N. R. (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **54** 622–654.

BURNSIDE, W. (1911). *Theory of Group with Finite Order*, 2nd ed., Cambridge University Press, London.

CHENG, C. S. (1978a). Optimality of certain asymmetrical experimental designs. *Ann. Statist.* **6** 1239–1261.

CHENG, C. S. (1978b). Optimal designs for the elimination of multi-way heterogeneity. *Ann. Statist.* **6** 1262–1272.

CHENG, C. S. (1980). Optimal designs for the elimination of multi-way heterogeneity. II. *IMS Bull.* 80t–57.

CHENG, C. S. and LI, K. C. (1981). Identifying the covariance structures equivalent to those with homoscedastic uncorrelated errors for the estimation and design problems. Unpublished manuscript.

FEDOROV, V. V. (1972). *Theory of Optimal Experiments*. Academic, New York.

HUBER, P. J. (1975). Robustness and designs, in *A Survey of Statistical Design and Linear Models*, (J. N. Srivastava, ed.), 287–301. North Holland, Amsterdam.

KARLIN, S. and STUDDEN, W. J. (1966). Optimal experimental designs. *Ann. Math. Statist.* **37** 783–815.

KIEFER, J. (1958). On the nonrandomized optimality and randomized non-optimality of symmetrical designs. *Ann. Math. Statist.* **29** 675–699.

KIEFER, J. (1974). General equivalence theory for optimum designs (approximation theory). *Ann. Statist.* **2** 849–879.

KIEFER, J. (1975). Constructions and optimality of generalized Youden designs, in *A Survey of Statistical Designs and Linear Models* (J. N. Srivastava, ed). 333–353. North Holland, Amsterdam.

LEHMANN, E. (1971). *Testing Statistical Hypotheses*. Wiley, New York.

LI, K. C. (1981). Minimax property for randomized designs—a unified approach. Mimeograph Series #81-44, Department of Statistics, Purdue University.

NOTZ, W. (1981). Optimal incomplete block designs for comparing treatments to a control. Mimeograph Series #81-34, Department of Statistics, Purdue University.

WU, C. F. (1981). On the robustness and efficiency of some randomized designs. *Ann. Statist.* **9** 1168–1177.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907