The Annals of Mathematical Statistics
1971, Vol. 41, No. 1, 46–58

ERRORS IN THE FACTOR LEVELS AND EXPERIMENTAL DESIGN

BY NORMAN R. DRAPER[1] AND WILLIAM J. BEGGS[2]

University of Wisconsin

1. Introduction and summary. Suppose that an experimenter is interested in determining the relationship between a response η and several "independent" variables, $x_1, x_2, \cdots, x_K$. The x-variables may be controlled by the experimenter or observed without control. Suppose, further, that these K "independent" variables represent all the factors that contribute to the response η, and that the exact relationship between η and the x's is

$$\eta = \phi(\mathbf{x}), \tag{1.1}$$

where $\mathbf{x} = (x_1, x_2, \cdots, x_K)'$. The function $\phi(\mathbf{x})$ is called the *response function* and, geometrically, it defines a surface in K-space which is called the *response surface*.

In the real world, however, we rarely know the exact relationship, or all the variables which affect that relationship. One way of proceeding then is to graduate, or approximate to, the true relationship by a polynomial function, linear in some unknown parameters to be estimated and of some selected order in the independent variables. Under the tentative assumption of the validity of this linear model (which we can justify on the basis of a Taylor expansion of ϕ), we can perform experiments, fit the model using regression techniques, and then apply standard statistical procedures to determine whether this model appears adequate. Since in practice we do not know all of the K factors which affect the response, we usually select $k(< K)$ variables which we believe might have significant effects. This selection may be made on the basis of prior knowledge, or a preliminary experiment may be performed to screen the important variables out of a larger set of possible variables. We can write our graduating model as follows:

$$y_u = f(x_{1u}, x_{2u}, \cdots, x_{ku} \mid \beta_0, \beta_1, \cdots, \beta_p) + \varepsilon_u, \tag{1.2}$$

where y_u is the observed response at the uth trial, x_{iu} is the setting of ith input variable for the uth trial, β_j is the jth parameter to be estimated, ε_u is the error involved in observing y on the uth trial, $u = 1, 2, \cdots, N$; $i = 1, 2, \cdots, k < K$;

Received June 16, 1969; revised February 16, 1970.

[1] Author was supported (1) partially by the Office of Naval Research under Contract Nonr-1202(17), Project NR 042-222, (2) for the fall semester 1967, at Imperial College, London, through funds provided to the University Research Committee by the Wisconsin Alumni Research Foundation, and (3) for the fall semester 1968, at Imperial College, London, by the National Science Foundation, Grant GP-7842.

[2] Author was supported (1) by the Office of Education, National Defense Education Act Fellowship, Title IV, and (2) by the Wisconsin Alumni Research Foundation. He is now at the Bettis Atomic Power Laboratory, Westinghouse Electric Corporation, Pittsburgh.

and $j = 0, 1, 2, \cdots, p$. The errors ε_u arise in one or more of the following ways:

(i) The true response η may be observed with error.
(ii) The function $f(x_{1u}, x_{2u}, \cdots, x_{ku})$ may not be the correct model.
(iii) The observations on the independent variables may contain errors.

Once an experimenter has chosen a polynomial model of suitable order, the problem arises as how best to choose the settings for the independent variables over which he has control. A particular selection of settings, or factor levels, at which observations are to be taken is called a *design*. Designs are usually selected to satisfy some desirable criteria chosen by the experimenter.

In this paper we consider the problem of design selection when there are errors in the factor levels as well as in the response. This problem has received little attention; in fact it appears that only Box (1963) has tackled it. Box first gives conditions under which the variance of a linear combination of the observations, say $L = \sum \alpha_u y_u$, is estimated unbiasedly. Next, he shows that these conditions are satisfied by two-level factorial and fractional factorial designs for first order models.

In Section 2, we present a criterion for selecting designs for experiments in which errors occur in the factor levels. In Section 3 we make a simplifying assumption and an approximation which leads to an approximate criterion, and this is applied to the first order model case in Section 4.

2. Development of a design criterion.

2.1. *Selected and exact design matrices.* In some region of interest R we graduate the true relationship (1.1) by the *linear* model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_X + \boldsymbol{\varepsilon}_y, \tag{2.1.1}$$

where $\mathbf{y}$ is an $N \times 1$ vector of observations, $\mathbf{X} = (\mathbf{1}, \mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_p)$ is an $N \times (p+1)$ design matrix of full rank, $\mathbf{1}$ is the $N \times 1$ unit vector, $\mathbf{f}_j = (f_{j1}, f_{j2}, \cdots, f_{jN})'$, $j = 1, 2, \cdots, k$, $f_{ju} = f_j(x_{1u}, x_{2u}, \cdots, x_{ku})$, $u = 1, 2, \cdots, N$, $\boldsymbol{\beta}_X$ is a $(p+1) \times 1$ vector of parameters, and $\boldsymbol{\varepsilon}_y$ is an $N \times 1$ vector of observational errors on $\mathbf{y}$. We shall assume

$$E(\boldsymbol{\varepsilon}_y) = \mathbf{0}, \qquad \mathrm{Var}(\boldsymbol{\varepsilon}_y) = \sigma^2_{y \mid X}\mathbf{I} \tag{2.1.2}$$

Under the standard assumption that no errors occur in setting the factors at their designated levels (i.e., the uth scaled level for x_i is *exactly* x_{iu}), Gauss' Theorem applies and the least squares method yields minimum variance, unbiased estimates of the parameters given by

$$\hat{\boldsymbol{\beta}}_X = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{2.1.3}$$

Let $\mathbf{x}_0$ be a vector which lies in the space spanned by the columns of $\mathbf{X}$. In particular, $\mathbf{x}_0$ may be one of the rows of $\mathbf{X}$, i.e.,

$$\mathbf{x}_{\cdot u} = (1, f_{1u}, f_{2u}, \cdots, f_{pu}). \tag{2.1.4}$$

The fitted or predicted value of the response at the point $\mathbf{x}_0$, obtained under the assumption of no errors in the factor levels, is given by

$$\hat{y}_X(\mathbf{x}_0) = \mathbf{x}_0'\hat{\boldsymbol{\beta}}_X. \tag{2.1.5}$$

It is the standard assumption of the absence of errors in the factor levels which is in question here. However, even if the x's are in error, the procedure given above, leading to the estimate (2.1.3), is valid as a first approximation to the correct analysis, which would employ the exact settings achieved by the independent variables. Assuming that errors do occur in the factor levels, we write

$$z_{iu} = x_{iu} + \varepsilon_{iu}, \tag{2.1.6}$$

where, for the uth run, z_{iu} is the *exact* setting of factor i actually achieved, x_{iu} is the selected setting of factor i, and ε_{iu} is the error involved.

If we knew the errors ε_{iu} and hence the factor levels z_{iu} exactly, the usual least squares analysis would be appropriate and would yield minimum variance unbiased estimates of the parameters of the model. We would have the model

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_y^*, \tag{2.1.7}$$

where $\mathbf{y}$ is the same $N \times 1$ vector of observations as before, $\mathbf{Z} = (\mathbf{1}, \mathbf{f}_1^*, \mathbf{f}_2^*, \cdots, \mathbf{f}_p^*)$ is the *exact* $N \times (p+1)$ design matrix, $\mathbf{f}_j^* = (f_{j1}^*, f_{j2}^*, \cdots, f_{jN}^*)'$, $f_{ju}^* = f_j(z_{1u}, z_{2u}, \cdots, z_{ku})$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)'$ is a $(p+1) \times 1$ vector of unknown parameters, and $\boldsymbol{\varepsilon}_y^*$ is an $N \times 1$ vector of observational errors on $\mathbf{y}$. We shall assume

$$E(\boldsymbol{\varepsilon}_y^*) = \mathbf{0} \quad \text{and} \quad \mathrm{Var}(\boldsymbol{\varepsilon}_y^*) = \sigma_y^2\mathbf{I}. \tag{2.1.8}$$

Note that (2.1.7) involves the *exact* factor levels $\mathbf{Z}$ while (2.1.1) involves the *selected* factor levels $\mathbf{X}$. Thus $\boldsymbol{\beta}_X$ and $\boldsymbol{\beta}$ are not in general the same unless all $\varepsilon_{iu} = 0$. In addition, the functions of the x's, f_{ju}, have been replaced by the same functions of the z's, f_{ju}^*. We also note that, if errors do occur in the x's in (2.1.1), the observations are those occurring when the factors are actually at the z levels, i.e., the vector of observations $\mathbf{y}$ is the same in both (2.1.1) and (2.1.7).

The estimate of $\boldsymbol{\beta}$ obtained by the least squares analysis on the z's is

$$\hat{\boldsymbol{\beta}}_Z = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}, \tag{2.1.9}$$

and the fitted or predicted value of the response at the point $\mathbf{x}_0$ is

$$\hat{y}_Z(\mathbf{x}_0) = \mathbf{x}_0'\hat{\boldsymbol{\beta}}_Z. \tag{2.1.10}$$

If, indeed, errors do occur in the x's, and $\mathbf{Z}$ is the exact but unknown design matrix, then the usual least squares estimate $\hat{\boldsymbol{\beta}}_X$ (2.1.3) based on the known matrix $\mathbf{X}$ is biased; that is, since now $E(\mathbf{Y} \mid \mathbf{Z}) = \mathbf{Z}\boldsymbol{\beta}$, then

$$E(\hat{\boldsymbol{\beta}}_X \mid \mathbf{Z}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\boldsymbol{\beta} = \boldsymbol{\beta} + \mathbf{b} \tag{2.1.11}$$

where $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\beta}$ is the bias of the estimate $\hat{\boldsymbol{\beta}}_X$ induced by the errors in setting the x's, and where

$$\boldsymbol{\varepsilon} = \mathbf{Z} - \mathbf{X}, \tag{2.1.12}$$

is the *matrix of design errors*. Similarly, $\hat{y}_X(\mathbf{x}_0)$ contains a bias, since

$$E(\hat{y}_X(\mathbf{x}_0) \mid \mathbf{Z}) = \mathbf{x}_0'\boldsymbol{\beta} + \mathbf{x}_0'\mathbf{b}. \tag{2.1.13}$$

Now $\hat{y}_X$ based on the *selected* factor levels $\mathbf{X}$, will yield a different set of fitted values from $\hat{y}_Z$, based on the unknown, but *exact*, factor levels $\mathbf{Z}$. Our goal will be to select a design $\mathbf{X}$, i.e., choose factor levels x_{iu} which, in a sense to be defined below, will minimize the distance between the "biased" fitted values and the "exact" fitted values.

2.2. *Choice of a criterion.* If $\mathbf{x}_0$ is an original selected design point $\mathbf{x}_{\cdot u}$ (see (2.1.4)), we can evaluate the fitted values $\hat{y}_X(\mathbf{x}_{\cdot u})$ and $\hat{y}_Z(\mathbf{x}_{\cdot u})$, the predicted responses at the point $\mathbf{x}_{\cdot u}$ using the selected and exact design matrices respectively. The difference,

$$w_u = \hat{y}_X(\mathbf{x}_{\cdot u}) - \hat{y}_Z(\mathbf{x}_{\cdot u}), \tag{2.2.1}$$

is a measure of how closely the biased predictor $\hat{y}_X$ (given that $\mathbf{Z}$ is the exact design matrix) approaches the unbiased and correct predictor $\hat{y}_Z$ at the point $\mathbf{x}_{\cdot u}$. By summing over all the selected design points $\mathbf{x}_{\cdot u}$ in $\mathbf{X}$, we can use, as a measure of deviation, the sum of squares

$$\sum_{u=1}^{N} w_u^2. \tag{2.2.2}$$

Note that w_u is a function of the observations $\mathbf{y}$ (which are subject to observational errors and factor errors) as well as the selected factor levels. In theory, we could repeat the experiment at the same z-values and obtain a distribution of observations at each design point. We first average (2.2.2) over these distributions for fixed ε_{iu}'s, and hence for fixed $\boldsymbol{\varepsilon}$, to obtain the measure of discrepancy

$$E_{\mathbf{y} \mid \boldsymbol{\varepsilon}}\left[\sum_u w_u^2\right]. \tag{2.2.3}$$

Now (2.2.3) is a function of the actual factor errors ε_{iu}. Averaging (2.2.3) over the distributions of the elements of $\boldsymbol{\varepsilon}$, we obtain the quantity

$$G = E_{\boldsymbol{\varepsilon}} E_{\mathbf{y} \mid \boldsymbol{\varepsilon}}\left[\sum_u w_u^2\right] = E_{\boldsymbol{\varepsilon}} E_{\mathbf{y} \mid \boldsymbol{\varepsilon}}[\mathbf{w}'\mathbf{w}], \tag{2.2.4}$$

where

$$\mathbf{w} = \hat{\mathbf{y}}_X - \hat{\mathbf{y}}_Z, \tag{2.2.5}$$

and

$$\hat{\mathbf{y}}_X = (\hat{y}_X(\mathbf{x}_{\cdot 1}), \hat{y}_X(\mathbf{x}_{\cdot 2}), \cdots, \hat{y}_X(\mathbf{x}_{\cdot N}))', \tag{2.2.6}$$

and $\hat{\mathbf{y}}_Z$ is similarly defined. Thus, G represents an average measure of discrepancy between the two fitted surfaces $\hat{\mathbf{y}}_X$ and $\hat{\mathbf{y}}_Z$, averaged over the distributions first of $\mathbf{y} \mid \boldsymbol{\varepsilon}$ and then of the elements of $\boldsymbol{\varepsilon}$, and evaluated at the points $\mathbf{x}_{\cdot u}$ in $\mathbf{X}$.

By performing the expectation operations, we can develop (2.2.4) in the following way: write $\mathbf{w} = \hat{\mathbf{y}}_X - \hat{\mathbf{y}}_Z$ as

$$\mathbf{w} = \mathbf{X}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{y}. \tag{2.2.7}$$

Then, after some reduction it can be shown that

$$E_{\mathbf{y}\mid\varepsilon}[\mathbf{w}'\mathbf{w}] = \boldsymbol{\beta}'\mathbf{C}_1\boldsymbol{\beta} + \sigma_y^2 \operatorname{tr}\mathbf{C}_2, \tag{2.2.8}$$

where

$$\mathbf{C}_1 = \boldsymbol{\varepsilon}'\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}, \tag{2.2.9}$$

and

$$\mathbf{C}_2 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{X}' + \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'. \tag{2.2.10}$$

Expecting over the factor errors gives

$$G = E_\varepsilon E_{\mathbf{y}\mid\varepsilon}[\mathbf{w}'\mathbf{w}] = \boldsymbol{\beta}' E_\varepsilon[\mathbf{C}_1]\boldsymbol{\beta} + \sigma_y^2 \operatorname{tr} E_\varepsilon[\mathbf{C}_2]. \tag{2.2.11}$$

Thus, we see that the average measure of the discrepancy between the two fitted surfaces is composed of two parts. The first of these terms is a bias term induced by the errors in the x's and involves the parameters $\boldsymbol{\beta}$ of the model. The second term is a variance term, and can be shown to be the sum of the expected variances of the deviations of w_u about their means. A desired design criterion would be to choose the design matrix $\mathbf{X}$ to minimize (2.2.11). Because of the difficulties associated with obtaining the expectation of $\mathbf{C}_2$, we use instead the procedure of the next section.

3. An assumption and an approximation. Since $\mathbf{X}'\mathbf{X}$ is a positive definite symmetric matrix, we may write $\mathbf{X}'\mathbf{X} = \mathbf{P}^{-1}(\mathbf{P}')^{-1}$, where $\mathbf{P}$ is square and non-singular. It follows that

$$\begin{aligned}\mathbf{Z}'\mathbf{Z} &= \mathbf{X}'\mathbf{X} + \mathbf{X}'\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\mathbf{X} + \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\ &= \mathbf{P}^{-1}[\mathbf{I}+\mathbf{Q}](\mathbf{P}')^{-1},\end{aligned} \tag{3.1}$$

where $\mathbf{Q} = \mathbf{P}[\mathbf{X}'\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\mathbf{X} + \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}]\mathbf{P}'$.

In performing an experiment in which errors may occur in the setting of the factor levels, it is implicitly assumed that these errors are "small" in some sense. If this were not the case, we would be in such a state of uncertainty that there would be serious doubt as to the validity of any estimation results. We make, therefore, the following assumption:

ASSUMPTION. The errors ε_{iu}, $i = 1, 2, \cdots, k$, $u = 1, 2, \cdots, N$, are sufficiently small with respect to the range of the factors x_i that

(i) the matrix $\mathbf{Q}^r$ converges to the null matrix as $r \to \infty$,
(ii) terms of third and higher order powers in the errors may be ignored, and
(iii) $\sum_{r=3}^{\infty}(-1)^r\mathbf{Q}^r$ is negligible, where $\mathbf{Q}$ is defined in (3.1).

We may now make use of the following expansion result from matrix theory:

THEOREM 1. *If* $\mathbf{M}$ *is a* $p \times p$ *diagonable matrix and* c *is a scalar constant such that* $c^r \mathbf{M}^r \to \mathbf{0}$, *a null matrix, as* $r \to \infty$, *then*

$$(\mathbf{I} - c\mathbf{M})^{-1} = \sum_{r=0}^{\infty} c^r \mathbf{M}^r, \tag{3.2}$$

where, by definition, $\mathbf{M}^0 = \mathbf{I}$. (For proof, see Jennings (1964) pages 185–186.) Applying this result to $\mathbf{Z}'\mathbf{Z}$, we first note that $\mathbf{Q}$ is a symmetric matrix and hence is diagonable. Setting $\mathbf{M} = \mathbf{Q}$ and $c = -1$, we have that $\sum_{r=0}^{\infty} c^r \mathbf{Q}^r = \mathbf{I} - \mathbf{Q} + \mathbf{Q}^2 + \sum_{r=3}^{\infty} (-1)^r \mathbf{Q}^r$ converges to $(\mathbf{I}+\mathbf{Q})^{-1} = [\mathbf{I} + \mathbf{P}(\mathbf{X}'\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\mathbf{X} + \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})\mathbf{P}']^{-1}$, since, by assumption (i), $\mathbf{Q}^r \to \mathbf{0}$ as $r \to \infty$. Furthermore, from assumption (iii), we may drop $\sum_{r=3}^{\infty} (-1)^r \mathbf{Q}^r$. Recalling that $(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{P}'\mathbf{P}$, we now have

$$\begin{aligned} (\mathbf{Z}'\mathbf{Z})^{-1} &\doteq \mathbf{P}'[\mathbf{I} - \mathbf{Q} + \mathbf{Q}^2]\mathbf{P} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\{\mathbf{I} - (\mathbf{X}'\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\mathbf{X} + \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})(\mathbf{X}'\mathbf{X})^{-1} \\ &\quad + [(\mathbf{X}'\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}]^2\} + \mathbf{R}(\boldsymbol{\varepsilon}^3), \end{aligned} \tag{3.3}$$

where $\mathbf{R}(\boldsymbol{\varepsilon}^3)$ is a matrix containing terms of third and higher order powers of the errors. Under our assumption (ii) then, we drop $\mathbf{R}(\boldsymbol{\varepsilon}^3)$ from (3.3) so that we have the following approximation:

$$\begin{aligned} (\mathbf{Z}'\mathbf{Z})^{-1} \doteq [&\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\mathbf{X}) - (\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\ &+ (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\mathbf{X})](\mathbf{X}'\mathbf{X})^{-1}. \end{aligned} \tag{3.4}$$

From (2.2.10), using (2.1.12) and (3.4), and dropping terms containing third and higher order powers of errors, we can show that

$$\operatorname{tr} \mathbf{C}_2 \doteq \operatorname{tr} \boldsymbol{\varepsilon}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\varepsilon}'. \tag{3.5}$$

Our measure of the deviation of the fitted values using the selected design matrix $\mathbf{X}$ from the fitted values using the exact design matrix $\mathbf{Z}$ can thus be approximated by

$$G \doteq \boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta} + \sigma_y^2 \operatorname{tr} \mathbf{M}^{-1}\boldsymbol{\Psi}, \tag{3.6}$$

where

$$\boldsymbol{\Lambda} = E_{\varepsilon}[\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}], \tag{3.7}$$

$$\mathbf{M} = N^{-1}\mathbf{X}'\mathbf{X}, \tag{3.8}$$

and

$$\boldsymbol{\Psi} = N^{-1}E_{\varepsilon}[\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}], \tag{3.9}$$

and where terms of third and higher order in the factor errors have been ignored. Our proposed design criterion will now be to choose the design matrix $\mathbf{X}$ to minimize (3.6).

4. The case of a first order model. We now examine (3.6) for a first order model, and see what conditions on the design moments are required in order to minimize it.

4.1. *Notation.* By definition, for a first order model,

$$\mathbf{X} = (\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k), \tag{4.1.1}$$

$$\mathbf{Z} = (\mathbf{1}, \mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_k) = (\mathbf{1}, \mathbf{x}_1 + \boldsymbol{\varepsilon}_1, \mathbf{x}_2 + \boldsymbol{\varepsilon}_2, \cdots, \mathbf{x}_k + \boldsymbol{\varepsilon}_k), \tag{4.1.2}$$

and

$$\boldsymbol{\varepsilon} = \mathbf{Z} - \mathbf{X} = (\mathbf{0}, \boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_k), \tag{4.1.3}$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iN})'$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iN})'$.

The elements of the design error matrix $\boldsymbol{\varepsilon}$ consist, then, simply of the factor errors ε_{iu}. We assume that

$$E[\varepsilon_{iu}] = 0, \tag{4.1.4}$$

$$\begin{aligned} E[\varepsilon_{iu}\varepsilon_{jv}] &= \sigma_i^2, && i = j, u = v, \\ &= \rho_{ij}\sigma_i\sigma_j, && i \neq j, u = v. \\ &= 0, && u \neq v, \end{aligned} \tag{4.1.5}$$

$i, j = 1, 2, \cdots, k$; $u, v = 1, 2, \cdots, N$. We also assume that the factor errors are uncorrelated with the observational errors ε_{yu} on the response. That is,

$$E[\varepsilon_{iu}\varepsilon_{yv}] = 0, \text{ all } i, u, v. \tag{4.1.6}$$

4.2. *The bias term and the matrix* $\boldsymbol{\Psi}$. We first note that the bias portion $G_B = \boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta}$ of G, where $\boldsymbol{\Lambda}$ is defined by (3.7) is unaffected by choice of the design. In general, $\boldsymbol{\Lambda}$ is a $(k+1)\times(k+1)$ matrix with (i,j)th element

$$\begin{aligned} \lambda_{ij} &= E[\boldsymbol{\varepsilon}_i'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}_j] \\ &= \operatorname{tr}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{ij}, \qquad i,j = 1,2,\cdots k, \end{aligned} \tag{4.2.1}$$

where $\boldsymbol{\varepsilon}_i$ is the ith column of $\boldsymbol{\varepsilon}$, and $\boldsymbol{\Sigma}_{ij}$ is the $N \times N$ covariance matrix of the random vectors $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\varepsilon}_j$. From (4.1.5)

$$\begin{aligned} \boldsymbol{\Sigma}_{ij} &= \sigma_i^2\mathbf{I}, && i = j, \\ &= \rho_{ij}\sigma_i\sigma_j\mathbf{I}, && i \neq j. \end{aligned} \tag{4.2.2}$$

Hence

$$\begin{aligned} \lambda_{ij} &= (k+1)\sigma_i^2, && i = j, \\ &= (k+1)\rho_{ij}\sigma_i\sigma_j, && i \neq j, \\ &= 0, && i \text{ or } j = 0. \end{aligned} \tag{4.2.3}$$

Thus we see that G_B does not depend upon the selected factor levels. We may thus attempt to minimize G in (3.6) with respect to the selected design by minimizing $G_V = \sigma_y^2 \operatorname{tr}\mathbf{M}^{-1}\boldsymbol{\Psi}$, the variance portion of (3.6), alone.

G_V contains the $(k+1)\times(k+1)$ matrix $\varepsilon'\varepsilon$ whose (i,j)th element is $\varepsilon_i'\varepsilon_j$. Hence the (i,j)th element of $E(\varepsilon'\varepsilon)$ is $\sigma_{ij} = \mathrm{tr}\boldsymbol{\Sigma}_{ij}$. From (3.9) and (4.2.2),

$$\boldsymbol{\Psi} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1k}\sigma_1\sigma_k \\ 0 & \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2k}\sigma_2\sigma_k \\ \vdots & & & & \\ 0 & \rho_{1k}\sigma_1\sigma_k & \rho_{2k}\sigma_2\sigma_k & \cdots & \sigma_k^2 \end{bmatrix}. \tag{4.2.4}$$

4.3. *The case of uncorrelated errors.* We now assume that the factor errors are uncorrelated; i.e., $E[\varepsilon_{iu}\varepsilon_{ju}] = 0$, $i \neq j$. Thus

$$\boldsymbol{\Psi} = \mathrm{diagonal}(0, \sigma_1^2, \sigma_2^2, \cdots, \sigma_k^2) \tag{4.3.1}$$

and

$$\begin{aligned} G_V &= \sigma_y^2 \,\mathrm{tr}\,\mathbf{M}^{-1}\boldsymbol{\Psi} \\ &= \sigma_y^2 \textstyle\sum_{i=1}^{k} b_{ii}\sigma_i^2, \end{aligned} \tag{4.3.2}$$

where b_{ii} is the (i,i)th element in $\mathbf{M}^{-1}$, and M is the moment matrix

$$\mathbf{M} = \begin{bmatrix} 1 & (1) & (2) & \cdots & (k) \\ (1) & (11) & (12) & \cdots & (1k) \\ (2) & (12) & (22) & \cdots & (2k) \\ \vdots & \vdots & \vdots & & \vdots \\ (k) & (1k) & (2k) & \cdots & (kk) \end{bmatrix}, \tag{4.3.3}$$

where $(i) = \sum_{u=1}^{N} x_{iu}/N$, $(ii) = \sum_{u=1}^{N} x_{iu}^2/N$, and $(ij) = \sum_{u=1}^{N} x_{iu}x_{ju}/N$, $i \neq j$. Our object is to determine what designs will minimize G_V. We restrict our attention to choosing the odd moments (i), (ij), $i \neq j$, while holding the even second order moments (ii) fixed at specified values. This sort of procedure is adopted because a more "spread-out" design, i.e., a design which covers a greater portion of the region of interest, will always reduce G_V. What we seek, however, are the relationships between design moments when the "spread" of the design is fixed. The whole design can always be expanded later to further reduce the value of G_V, if this were desired.

For the case $k = 1$, where we have only one independent factor x, there are no (ij) moments and G_V is simply $\sigma_1^2/\{(11)-(1)^2\}$. This is obviously minimized by setting $(1) = 0$. The best class of designs with respect to the odd moments when $k = 1$ then, consists of those for which the x's sum to zero.

For $k \geqq 2$, we can proceed in a manner equivalent to the method used by Box (1952) for minimization of the diagonal elements of $(\mathbf{X}'\mathbf{X})^{-1}$. Let the determinant of $\mathbf{M}$ be expressed by the Cauchy expansion, that is,

$$|\mathbf{M}| = |\mathbf{D}_{ii}|\,[(ii) - \mathbf{s}_i'\mathbf{D}_{ii}^{-1}\mathbf{s}_i],$$

where $\mathbf{D}_{ii}$ is the matrix obtained by deleting the ith row and column from $\mathbf{M}$, $|\mathbf{D}_{ii}|$ is the determinant of $\mathbf{D}_{ii}$, and $\mathbf{s}_i = [(i), (1i), (2i), \cdots, (i-1, i), (i+1, i), \cdots, (ki)]'$. It follows that

$$b_{ii} = 1/\{(ii) - \mathbf{s}_i' \mathbf{D}_{ii}^{-1} \mathbf{s}_i\}. \tag{4.3.5}$$

Since $\mathbf{M} = (\mathbf{X}'\mathbf{X})/N$ is positive definite, any principal submatrix $\mathbf{D}_{ii}$ is also positive definite. Hence, b_{ii} is minimized when $\mathbf{s}_i = \mathbf{0}$, for $i = 1, 2, \cdots, k$. It follows directly that (4.3.2) is minimized by setting all $\mathbf{s}_i = \mathbf{0}$, i.e., by setting equal to zero all first and odd second order moments of the design. The minimized value of G_V is then

$$\sigma_y^2 \sum_{i=1}^{k} [\sigma_i^2/(ii)]. \tag{4.3.6}$$

4.4. *Correlated factor errors: first order moments.* In Section 4.3, we assumed that the factor errors were all uncorrelated with one another. If this is not so, the form of G_V is still $G_V = \sigma_y^2 \operatorname{tr} \mathbf{M}^{-1} \boldsymbol{\Psi}$, but $\boldsymbol{\Psi}$ is now defined by (4.2.4). Let

$$\mathbf{M} = \begin{bmatrix} 1 & \mathbf{s}' \\ \mathbf{s} & \mathbf{M}_{22} \end{bmatrix}, \tag{4.4.1}$$

where $\mathbf{s}$ is a $k \times 1$ vector of first order moments, and $\mathbf{M}_{22}$ is a $k \times k$ positive definite matrix of second order moments. Similarly, let

$$\boldsymbol{\Psi} = \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Psi}_{22} \end{bmatrix}, \tag{4.4.2}$$

where $\boldsymbol{\Psi}_{22} = \sum_{u=1}^{N} E[\mathbf{e}_u \mathbf{e}_u'/N]$, and $\mathbf{e}_u = (e_{1u}, e_{2u}, \cdots, e_{ku})'$. Then

$$\operatorname{tr} \mathbf{M}^{-1} \boldsymbol{\Psi} = E[\textstyle\sum_{u=1}^{N} \mathbf{e}_u' (\mathbf{M}_{22} - \mathbf{s}\mathbf{s}')^{-1} \mathbf{e}_u / N]. \tag{4.4.3}$$

Now since $\mathbf{M}$ is positive definite, $(\mathbf{M}_{22} - \mathbf{s}\mathbf{s}')$ is positive definite, and hence $(\mathbf{M}_{22} - \mathbf{s}\mathbf{s}')^{-1} = \mathbf{M}_{22}^{-1} + \sum_{r=1}^{\infty} (\mathbf{M}_{22}^{-1} \mathbf{s}\mathbf{s}')^r \mathbf{M}_{22}^{-1}$. The positive definite quadratic form in (4.4.3) may now be written

$$\mathbf{e}_u' \mathbf{M}_{22}^{-1} \mathbf{e}_u + \mathbf{e}_u' [\textstyle\sum_{r=1}^{\infty} (\mathbf{M}_{22}^{-1} \mathbf{s}\mathbf{s}')^r \mathbf{M}_{22}^{-1}] \mathbf{e}_u > 0. \tag{4.4.4}$$

The matrix in the second term of (4.4.4) however, is symmetric and at least positive semi-definite. Therefore, (4.4.4) is minimized when the second term is zero. This is accomplished only when $\mathbf{s} = \mathbf{0}$, i.e., only when all the first order moments are equal to zero.

4.5. *Correlated factor errors: odd second order moments.* For $k = 2$, when $(1) = (2) = 0$, G_V is minimized when

$$(12) = \{A - [A^2 - B^2(11)(22)]^{\frac{1}{2}}\}/B \tag{4.5.1}$$

where $A = (22)\sigma_1^2 + (11)\sigma_2^2$, $B = 2\rho\sigma_1\sigma_2$. For $k > 2$,

$$G_V = \sigma_y^2 [\textstyle\sum_{i=1}^{k} |\mathbf{D}_{ii}| \sigma_1^2 + 2 \sum\sum_{i<j} (-1)^{i+j} |\mathbf{D}_{ij}| \rho_{ij} \sigma_i \sigma_j] |\mathbf{M}^{-1}|, \tag{4.5.2}$$

where $|\mathbf{D}_{ij}|$ is the minor of the (i, j)th element of $\mathbf{M}$. We wish to find the values of the $q = k(k-1)/2$ odd second order moments which minimize G_V. Unfortunately,

even for $k = 3$, the problem is cumbersome and numerical solution appears necessary. For this reason, we now turn our attention to some special cases for which ana ytic solutions can be obtained.

Special case 1. *All factor error correlations equal.* For the case $\rho_{ij} = \rho \neq 0$, $(ii) = c$, $\sigma_i = \sigma$, and $(ij) = x$, $i \neq j = 1, 2, \cdots, k$, we find that G_V is minimized for

$$x = \frac{c[-1+\{(1-\rho)[1+(k-1)\rho]\}^{\frac{1}{2}}]}{(k-2)-(k-1)\rho}, \qquad \rho \neq (k-2)/(k-1),$$

$$= \frac{c(k-2)}{2(k-1)}, \qquad \rho = (k-2)/(k-1),$$

where $-1/(k-1) < \rho < 1$.

Special case 2. *Groups of factor errors: correlations zero between groups, equal within groups.* Let the factor errors be grouped such that

$$\rho_{i_r, j_s} = 0, \qquad \text{if } i_r \text{ and } j_s \text{ belong to different groups, } r \text{ and } s,$$

$$= \rho_r, \qquad \text{if } r = s \text{ and } i_r, j_r \text{ belong to group } r,$$

$i_r = 1, 2, \cdots, k_r$, $\sum_{r=1}^{m} k_r = k$, and $r \neq s = 1, 2, \cdots, m$. That is, we now have k factor errors divided into m groups of errors which are uncorrelated between groups, but are identically correlated within each group.

We further require that $\sigma_{i_r} = \sigma_r$, $(i_r i_r) = c_r$, $(i_r j_r) = x_r$, $i_r \neq j_r$, and $(i_r j_s) = 0$, $r \neq s$. Under these conditions, G_V is minimized for

$$x_r = \frac{c_r[-1+\{(1-\rho_r)[1+(k_r-1)\rho_r]\}^{\frac{1}{2}}]}{(k_r-2)-(k_r-1)\rho_r}, \qquad \rho_r \neq (k_r-2)/(k_r-1),$$

$$= \frac{c_r(k_r-2)}{2(k_r-1)}, \qquad \rho_r = (k_r-2)/(k_r-1),$$

$$-1/(k_r-1) < \rho_r < 1, k_r > 1, r = 1, 2, \cdots, m.$$

Special case 3. *Two groups of errors: correlations zero within groups, equal between groups.* Let the factor errors belong to two groups, denoted by I and II, of size k_1 and k_2 respectively, such that

$$E[\varepsilon_i \varepsilon_j] = \sigma_1^2, \qquad i = j \text{ belongs to I},$$

$$= \sigma_2^2, \qquad i = j \text{ belongs to II},$$

$$= \rho\sigma_1\sigma_2, \qquad i \text{ belongs to I}, j \text{ belongs to II},$$

$$= 0, \qquad \text{otherwise.}$$

Furthermore, let

$$(ii) = c_1, \qquad i \text{ belongs to I}, i = 1, 2, \cdots, k_1,$$

$$= c_2, \qquad i \text{ belongs to II}, i = 1, 2, \cdots, k_2,$$

and

$$(ij) = x, \qquad i \text{ belongs to I, } j \text{ belongs to II,}$$
$$= 0, \qquad i,j \text{ belong to same group.}$$

Under these assumptions, G_V is minimized for

$$x = [c_2\sigma_1^2 + c_1\sigma_2^2 - \{c_2\sigma_1^4 + 2\sigma_1^2\sigma_2^2 c_1 c_2(1-2k_1k_2\rho^2) + c_1\sigma_2^4\}^{\frac{1}{2}}]/[2k_1k_2\rho\sigma_1\sigma_2],$$

where $0 < \rho^2 < (k_1k_2)^{-1}$.

4.6. *Examples.* Table 1 and Table 2 show best moment values derived from the foregoing work for the situations specified. The moments in Table 2b were obtained by numerical minimization of $\operatorname{tr}\mathbf{M}^{-1}\boldsymbol{\Psi}$, and the others were evaluated from

TABLE 1

Case $k = 2$, $\rho \neq 0$, $(ii) = 1$: odd second order moments and percentage reduction for a first order model

(a) $\sigma_1^2:\sigma_2^2 = 1:1$			(b) $\sigma_1^2:\sigma_2^2 = 1:4$		
ρ	(12)	R	ρ	(12)	R
$\pm.2$	$\pm.101$	1.0	$\pm.2$	$\pm.081$	0.6
$\pm.3$	$\pm.153$	2.3	$\pm.3$	$\pm.122$	1.5
$\pm.5$	$\pm.268$	6.7	$\pm.5$	$\pm.209$	4.2
$\pm.7$	$\pm.408$	14.3	$\pm.7$	$\pm.306$	8.6
$\pm.9$	$\pm.627$	28.2	$\pm.9$	$\pm.425$	15.3

formulas in Section 4.5. The R-values show the percentage reduction $100(F_0 - F_m)/F_0$, where F_m is the appropriate minimizing value of G_V, and F_0 is the value of G_V for an orthogonal design, i.e., (4.3.6). We see that R increases with $|\rho|$ and, for given ρ, is higher when all variances are equal. Thus, when $|\rho|$ is high, a worthwhile improvement in G_V can be obtained by choosing a non-orthogonal design rather than an orthogonal one, but for small $|\rho|$, it is clear that the orthogonal design is robust.

Table 3 shows, for the two factor situations specified, best values of a, b, c, and d for the four point design $(-a, -b)$, $(c, -d)$, $(-c, d)$, (a, b). These points, whose coordinates are chosen to satisfy the required moment conditions, are simple amendments of the 2^2 factorial points. Table 4 shows, for the three-factor situations specified, best designs which are similar amendments of the 2^3 factorial. For additional examples, see Beggs (1969).

Acknowledgment. The authors greatly appreciate the constructive comments of a referee.

TABLE 2

Case $k = 3$, $\rho_{ij} = \rho \neq 0$, *(ii)* $= 1$: *odd second order moments and percentage reduction for a first order model*

(a) $\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 1:1:1$

ρ	(12) = (13) = (23)	R
−.4	−.262	12.0
−.2	−.108	2.3
.2	.097	1.8
.4	.196	7.1
.6	.310	16.1
.8	.465	30.2

(b) $\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 1:1:4$

ρ	(12)	(13) = (23)	R
−.4	−.336	−.179	8.0
−.2	−.119	−.083	1.6
.2	.091	.079	1.3
.4	.174	.160	5.0
.6	.250	.250	11.1
.8	.337	.356	20.1

TABLE 3

Case $k = 2$, (11) = (22) = 1: *amended* 2^2 *factorial point coordinates*

(a) $\sigma_1^2 = \sigma_2^2$

ρ*	(12)	$a = b$	$c = d$
.2	.101	1.05	.95
.5	.268	1.13	.86
.9	.627	1.28	.61

* When ρ is the negative of the values shown, change the sign of (12) and switch the headings "$a = b$" and "$c = d$".

(b) $4\sigma_1^2 = \sigma_2^2$

ρ	(12)	a	b	c	d
.2	.081	1.03	1.05	0.97	0.95
−.2	−.081	0.97	0.95	1.03	1.05
.5	.209	1.08	1.11	0.91	0.87
−.5	−.209	0.92	0.86	1.07	1.12
.9	.425	1.17	1.22	0.80	0.71
−.9	−.425	0.84	0.67	1.14	1.24

TABLE 4

Case $k = 3$, (11) = (22) = (33) = 1: *amended first order designs*

(a) $\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 1:1:1$, $\rho_{12} = \rho_{13} = \rho_{23} = 0.4$, (12) = (13) = (23) = 0.196

x_1	x_2	x_3
−1.180	−1.180	−1.180
0.802	−0.991	−0.991
−0.991	0.802	−0.991
0.991	0.991	−0.802
−0.991	−0.991	0.802
0.991	−0.802	0.991
−0.802	0.991	0.991
1.180	1.180	1.180

(b) $\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 1:1:4$, $\rho_{12} = \rho_{13} = \rho_{23} = 0.4$, (12) = 0.174, (13) = (23) = 0.160

x_1	x_2	x_3
−1.215	−1.215	−1.013
0.756	−1.062	−1.000
−1.062	0.756	−1.000
0.909	0.909	−0.987
−0.909	−0.909	0.987
1.062	−0.756	1.000
−0.756	1.062	1.000
1.215	1.215	1.013

REFERENCES

[1] BEGGS, W. J. (1969). Errors in the factor levels and experimental design. Ph.D. Dissertation, University of Wisconsin.

[2] BOX, G. E. P. (1962). Multifactor designs of first order. *Biometrika* **39** 49–57.

[3] BOX, G. E. P. (1963). The effect of errors in the factor levels and experimental design. *Technometrics* **5** 247–262.

[4] JENNINGS, W. (1964). *First Course in Numerical Methods*. Macmillan, New York.