Statistica Sinica 4(1994), 275-280

CHOOSING A DESIGN FOR STRAIGHT LINE FITS TO TWO CORRELATED RESPONSES

Whasoo B. Kim and Norman R. Draper

Rusan, Korea and University of Wisconsin

Abstract: Variance and bias methods, used previously to choose response surface designs in the single response situation are explored for the two response case with no common parameters. The effect of correlation between the responses on the choice of a design for fitting two straight lines is evaluated.

Key words and phrases: Multiple responses, variance and bias.

1. Introduction

Experiments are to be performed at levels of a single predictor variable x, with a *region of interest* R, coded to $(-1, 1)$. The experimental runs need not be within R in order best to explore R but could lie in a larger region of operability, O. Two responses will be fitted to the data by least squares using straight lines with no common parameters. A modest amount of quadratic bias is anticipated in each response, and the design must be chosen to balance off variance and bias errors appropriately. The responses are correlated to an extent that can be roughly specified. How does this additional knowledge affect the choice of design, compared to the zero correlation case? Weight functions could be used to give varying importance to various parts of O. Here, only a weight function which is uniform within R and zero outside it is considered.

Two types of discrepancies, variance error and bias error, can occur. The variance error occurs because of sampling error and the bias error because the graduating function is inadequate. A detailed discussion of these two types of errors is given by Box and Draper (1959, pp. 624-625) in their examination of the single response case. The choice of a suitable design is made by taking both these errors into account. Suppose that the model to be fitted is represented by the first order polynomials

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1.1}$$

and the true or feared relationships are second order polynomials

$$E(\boldsymbol{Y}) = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\Gamma}. \tag{1.2}$$

Thus, in these equations, $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$, with $\boldsymbol{y}_i = (y_{1i}, y_{2i}, \ldots, y_{Ni})'$, $i = 1, 2$, is the $N \times 2$ matrix of the observed values of the response variables; $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)'$, with $\boldsymbol{x}_u = (1, x_u)'$, for $u = 1, 2, \ldots, N$, is the $N \times 2$ matrix of the input levels; $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, with $\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i})'$, $i = 1, 2$, is the 2×2 matrix of the parameters; $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2)$, with $\boldsymbol{\epsilon}_i = (\epsilon_{1i}, \epsilon_{2i}, \ldots, \epsilon_{Ni})'$, for $i = 1, 2$, is the $N \times 2$ matrix of errors. Also $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$, with $\boldsymbol{\eta}_i = (\eta_{1i}, \ldots, \eta_{Ni})'$, for $i = 1, 2$, is the $N \times 2$ matrix of true values; $\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_N)'$, with $\boldsymbol{z}_u = (x_{u1}^2, x_{u2}^2; x_{u1}x_{u2})'$, for $u = 1, 2, \ldots, N$, is the $N \times 3$ matrix of input variables shown only in the feared model; and $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, with $\boldsymbol{\gamma}_i = (\beta_{11,i}, \beta_{22,i}; \beta_{12,i})'$, $i = 1, 2$, is the 3×2 matrix of parameters shown only in the feared model. Let the 2×2 variance-covariance matrix of the ith row vector of the error matrix, $\boldsymbol{\epsilon}_i$, be

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}. \tag{1.3}$$

Let $\hat{\boldsymbol{Y}}(\boldsymbol{x})$ and $\boldsymbol{\eta}(\boldsymbol{x})$ be defined, respectively, as fitted values of $\boldsymbol{Y}$ under the least squares estimation method and the corresponding true mean values at the point $\boldsymbol{x} = (1, x)'$. Then the design will be chosen to minimize

$$\boldsymbol{J} = N\boldsymbol{\Sigma}^{-1} \int_O w(\boldsymbol{x}) E\left\{\hat{\boldsymbol{Y}}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})\right\}' \left\{\hat{\boldsymbol{Y}}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})\right\} d\boldsymbol{x}, \tag{1.4}$$

where $w(\boldsymbol{x})$ is a weight function and $d\boldsymbol{x} = dx_1 dx_2$. In (1.4), the premultiplication factor $N\boldsymbol{\Sigma}^{-1}$ is the natural extension of the N/σ^2 factor for the univariate case. This movement of the correlation structure to the bias term is sensible because it converts the unknown parameters in $\boldsymbol{\Gamma}$ to sizes made relative to the variance structure. It is possible to introduce an additional weight function if it were thought desirable to weight the responses unequally. Using the uniform weight function within R, $w(\boldsymbol{x}) = \Omega = 1/\int_R d\boldsymbol{x}$ in R, and 0 elsewhere, (1.4) becomes

$$\begin{aligned} \boldsymbol{J} &= N\Omega \int_R \boldsymbol{x}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}\, d\boldsymbol{x}\, \boldsymbol{I}_2 \\ &+ N\Omega\boldsymbol{\Sigma}^{-1} \int_R \boldsymbol{\Gamma}'\left\{\boldsymbol{x}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} - \boldsymbol{z}'\right\}'\left\{\boldsymbol{x}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} - \boldsymbol{z}'\right\}\boldsymbol{\Gamma}\, d\boldsymbol{x} \\ &= \boldsymbol{V} + \boldsymbol{B}, \end{aligned} \tag{1.5}$$

say. The criterion $\boldsymbol{J}$, which is a 2×2 matrix, can be represented by the sum of the contributions from the two types of errors, where $\boldsymbol{V}$ explains the contribution from the variance error and $\boldsymbol{B}$ from the bias error. How can one "minimize $\boldsymbol{J}$", which is a matrix ? Here, the trace of $\boldsymbol{J}$, i.e., tr($\boldsymbol{J}$) is minimized. (Computations on the determinant of $\boldsymbol{J}$, and on the maximum eigen value of $\boldsymbol{J}$, showed very similar results.) Let us assume that the design points are centrally located in the region R, i.e.,

$$\sum_{u=1}^{N} x_u = 0. \tag{1.6}$$

2. Design Behavior

In our specific case,

$$\begin{aligned} \operatorname{tr}(\boldsymbol{J}) &= \operatorname{tr}(\boldsymbol{V}) + \operatorname{tr}(\boldsymbol{B}) = V + B \\ &= 2v(c) + b(c)m + \{d^2/(3c)\}m, \end{aligned} \tag{2.1}$$

where $c = \sum_{u=1}^{N} x_u^2/N$, $d = \sum_{u=1}^{N} x_u^3/N$, $v(c) = 1 + 1/(3c)$, $b(c) = (c - 1/3)^2 + 4/45$, where

$$m = \operatorname{trace}(N\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}) = \frac{1}{1-\rho^2}\left\{\alpha_{11,1}^2 + \alpha_{11,2}^2 - 2\rho\alpha_{11,1}\alpha_{11,2}\right\}, \tag{2.2}$$

with ρ the correlation coefficient between the two responses, and where $\alpha_{11,i} = \frac{\beta_{11,i}}{\sigma_i/\sqrt{N}}$, $i = 1, 2$, are the normalized quadratic coefficients for the two models. Because $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}$ is nonnegative definite, minimization of $\operatorname{tr}(\boldsymbol{J})$ implies that $d = 0$. Thus the function to be minimized becomes

$$2\left(1 + \frac{1}{3c}\right) + \left\{\left(c - \frac{1}{3}\right)^2 + \frac{4}{45}\right\}m. \tag{2.3}$$

The optimum c is a function of m. Table 1 shows the best $c^{1/2}$ values for some selected m, together with corresponding values of V, B, and the ratio V/B.

Table 1. Optimum values of $c^{1/2}$ given m

m	$c^{1/2}$	V	B	V/B
0.00	∞	2.00	0.00	∞
0.25	1.10	2.55	0.21	11.9
1.00	0.91	2.81	0.33	8.4
4.00	0.76	3.15	0.59	5.3
16.00	0.66	3.53	1.59	2.2
∞	0.58	4.00	∞	0.0

Note that when V and B are roughly equal, i.e. $V/B = 1$, a situation one might think "typical" if one has chosen the correct models, a design close in size to the all-bias value of $c^{1/2} = 0.58$ is called for. Even if V is several times B, not much increase in design size is indicated. To see how variations in the components of m in (2.2) affect the design, consider the cases $\alpha_{11,1} = \alpha$, $\alpha_{11,2} = \alpha/g$, $0 \leq g \leq 1$, and $-0.9 \leq \rho \leq 0.9$. It can be shown that this includes all cases, due to the essential interchangeability of $\alpha_{11,1}$ and $\alpha_{11,2}$ and the fact that changes of sign in the α's can be replaced by changes of sign in ρ. Table 2 shows optimum $c^{1/2}$ values for $\alpha = 1$. (As α rises above 1, m increases, leading to decreases in $c^{1/2}$ values (see Table 1) and a slightly changed pattern. As α falls below 1, m decreases and the $c^{1/2}$ values increase, with a slightly changed pattern. Larger α's are unlikely to be considered, in practice. If it were known that there was much bias, fitting the straight lines would no longer make sense.)

Note from Table 2, that, as g goes to 0, for $-.9 \leq \rho \leq .9$ approximately, the design gets smaller and the changes in $c^{1/2}$ with respect to changing ρ diminish, i.e. the pattern flattens out in the east-west direction. Also, the changes in the $c^{1/2}$ values in the $\rho < 0$ part are larger than those in the $\rho > 0$ portion, being larger for larger g.

The effect of ρ can be observed by considering g fixed (and recall $\alpha = 1$). Then

$$m = (1 + g^2 - 2g\rho)/\{(1 - \rho^2)g^2\}.$$

When $g = 1$, $m \propto (1 + \rho)^{-1}$, which changes more for negative ρ than for positive ρ. As $g \to 0$, m becomes extremely large for any ρ, indicating a design close to the all bias design.

The practical consequences of this are as follows. If one response is biased and the other is (relatively) not, i.e., g is small, then the correlation ρ between the responses does not matter. (It has only a slight effect on the design when ρ is close to ± 1.) As the biases become "more equal" (g rises), a known negative correlation would cause us to shrink the design compared with the $\rho = 0$ case. A positive correlation would need no shrinkage for most of the ρ range and a little shrinkage for large values of ρ. A useful general assessment, therefore, is not to worry much about changing the design (from the $\rho = 0$ design) if ρ is positive, but to make it smaller if ρ is negative. Thus picking an incorrect positive ρ does not matter. Picking ρ positive when it is really negative will lead to a design that is too widely spread.

Example. Consider the five point design $(-2a, -a, 0, a, 2a)$. This has first and third moments zero and $c = 2a^2$. Suppose it was thought that $g = \alpha = 1$. Entering Table 2 with $\rho = 0$ (no correlation) gives $c^{1/2} = 0.83$, implying $a = 0.59$. Thus the design is $(-1.18, -0.59, 0, 0.59, 1.18)$. (We recall that, in our formulation, design points can lie both inside and outside the region of interest.) Suppose, however, that the two responses were actually negatively correlated

Table 2. Optimum $c^{1/2}$ values for $\alpha = 1$ and selected values of g and ρ.

	ρ																		
g	−0.9	−0.8	−0.7	−0.6	−0.5	−0.4	−0.3	−0.2	−0.1	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.0	0.65	0.69	0.72	0.74	0.76	0.78	0.79	0.81	0.82	0.83	0.84	0.85	0.86	0.87	0.87	0.88	0.89	0.89	0.90
0.9	0.65	0.68	0.71	0.73	0.75	0.77	0.78	0.80	0.81	0.82	0.83	0.84	0.84	0.85	0.86	0.87	0.87	0.88	0.88
0.8	0.64	0.68	0.70	0.72	0.74	0.76	0.77	0.78	0.79	0.80	0.81	0.82	0.83	0.84	0.84	0.85	0.85	0.86	0.85
0.7	0.63	0.67	0.69	0.71	0.73	0.74	0.76	0.77	0.78	0.79	0.80	0.80	0.81	0.82	0.82	0.82	0.83	0.82	0.81
0.6	0.63	0.66	0.68	0.70	0.71	0.73	0.74	0.75	0.76	0.77	0.77	0.78	0.79	0.79	0.79	0.80	0.79	0.79	0.76
0.5	0.62	0.65	0.67	0.68	0.70	0.71	0.72	0.73	0.74	0.74	0.75	0.75	0.76	0.76	0.76	0.76	0.76	0.74	0.71
0.4	0.61	0.63	0.65	0.66	0.68	0.69	0.69	0.70	0.71	0.71	0.72	0.72	0.72	0.73	0.72	0.72	0.71	0.70	0.67
0.3	0.60	0.62	0.63	0.64	0.65	0.66	0.67	0.67	0.68	0.68	0.68	0.68	0.69	0.68	0.68	0.68	0.67	0.66	0.63
0.2	0.59	0.60	0.61	0.62	0.62	0.63	0.63	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.63	0.63	0.62	0.60
0.1	0.58	0.58	0.59	0.59	0.59	0.59	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.59	0.59	0.59	0.58

with $\rho = -0.8$. Table 2 indicates that, $c^{1/2} = 0.69$, whence $a = 0.49$. Thus a reduction in the design to $(-0.98, -0.49, 0, 0.49, 0.98)$ is called for here.

A similar exploration of the two responses, two predictors case provides a similar overall conclusion, that negative correlation would call for a slight reduction in the size of the design compared with the no-correlation design. Cases with three responses have three correlations and thus are more difficult to study. Numerical calculations carried out for three responses and one or two predictors show that negative correlations increase bias and thus call for smaller designs. (Technical Report 883 gives details. It is available from the second author.)

References

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.

Kim, W. B. and Draper, N. R. (1991). Response surface designs for multiple responses. University of Wisconsin, Statistics Department, Technical Report No. 883, September.

306-504 Seonkyung Apt., Gumjeong Gu, Pusan, Korea.
Department of Statistics, University of Wisconsin, 1210 West Dayton Street, Madison, WI 53706, U.S.A.

(Received June 1991; accepted July 1993)