The Korean Communications
in Statistics Vol. 6, No. 1, 1999
pp. 1- 10

A Graphical Method for Evaluating the Mixture Component Effects of Ridge Regression Estimator in Mixture Experiments

Dae-Heung Jang[1)]

Abstract

When the component proportions in mixture experiments are restricted by lower and upper bounds, multicollinearity appears all too frequently. The ridge regression can be used to stabilize the coefficient estimates in the fitted model. I propose a graphical method for evaluating the mixture component effects of ridge regression estimator with respect to the prediction variance and the prediction bias.

1. Introduction

In mixture experiments, the measured response is assumed to depend only on the relative proportions of the components present in the mixture. For mixture experiments, if we let x_i represent the proportion of the ith component in the mixture where the number of components is q, then

$$\sum_{i=1}^{q} x_i = 1,$$

where $0 \le x_i \le 1,\ i = 1, 2, \cdots, q$. The experimental region is a regular $(q-1)$-dimensional simplex. When additional constraints are imposed on the proportions in the form of lower and upper bounds

$$0 < L_i \le x_i \le U_i < 1\ , \quad \text{where } i = 1, 2, \cdots, q, \tag{1}$$

the experimental region becomes a subregion of the simplex.

Typically, mixture models are of the Scheffé type where the first-order model is

$$y = \sum_{i=1}^{q} \beta_i x_i + \varepsilon$$

and the second-order model is

$$y = \sum_{i=1}^{q} \beta_i x_i + \sum \sum_{i<j}^{q} \beta_{ij} x_i x_j + \varepsilon$$

where y is observed response and ε is random error. Scheffé models can be expressed in matrix notation as

1) Professor, Division of Mathematical Science, Pukyong National University, Pusan, 608-737, Korea.

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

where $\underset{\sim}{y} = (y_1, y_2, \cdots, y_n)'$ is the vector of observed responses, X is the $n \times p (\leq n)$ matrix of the component proportions and cross-products between the proportions depending on the model, $\underset{\sim}{\beta}$ is the $p \times 1$ vector of parameters which appear in the chosen model, and $\underset{\sim}{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)'$ is the vector of random errors associated with $\underset{\sim}{y}$. Here p is the number of parameters in the model, and the response at any location $\underset{\sim}{x}$ in the region of interest is

$$y(\underset{\sim}{x}) = \underset{\sim}{x}_0{}' \underset{\sim}{\beta} + \varepsilon$$

where $\underset{\sim}{x}_0{}' = (x_1, x_2, \cdots, x_q)$ for a first-order model and $\underset{\sim}{x}_0{}' = (x_1, x_2, \cdots, x_q, x_1 x_2, \cdots, x_{q-1} x_q)$ for a second-order model.

Assuming the assumptions for ordinary least squares method are met, the vector of unknown parameters is estimated by

$$\underset{\sim}{b} = (X'X)^{-1} X' \underset{\sim}{y}$$

where $\underset{\sim}{b}' = (b_1, b_2, \cdots, b_q)$ for a first-order model and $\underset{\sim}{b}' = (b_1, b_2, \cdots, b_q, b_{12}, \cdots, b_{1q}, \cdots, b_{q-1,q})$ for a second-order model. Under the assumption that $\underset{\sim}{\varepsilon} \sim (\underset{\sim}{0}, \sigma^2 I)$, the estimated value of the response at any location $\underset{\sim}{x}$ in the region of interest is

$$\hat{y}(\underset{\sim}{x}) = \underset{\sim}{x}_0{}' \underset{\sim}{b}. \tag{2}$$

Unstable coefficient estimates arise from what is known as multicollinearity. Multicollinearity is a condition among the set of p regressor variables $x_1, x_2, \cdots, x_p$ in the model. If there exists an approximate linear dependence among the columns of X, then we have the condition usually identified as ill-conditioning or multicollinearity.

Various techniques for rectifying or reducing the effects of multicollinearity have been proposed. A prevailing technique is the ridge regression although there are some criticisms(for example, See Draper and Smith(1981)). Hoerl and Kennard(1970 a, b), and Marquardt(1970) have suggested problems associated with a ridge regression estimator. The ridge regression estimator for the parameters in the first order and second order polynomial models are calculated using the formula

$$\underset{\sim}{b}(k) = (X'X + kI)^{-1} X' \underset{\sim}{y}, \tag{3}$$

where k is a constant and usually $0 < k < 1$.

The purpose of this paper is to suggest a graphical method for evaluating the mixture component effects of the ridge regression estimator with respect to the prediction variance and the prediction bias. In Section 2, I propose a graphical method for evaluating the mixture component effects of the ridge regression estimator with respect to the prediction variance and the prediction bias. In Section 3, I give numerical examples. In Section 4, I draw conclusion.

2. A Graphical Method for Evaluating the Ridge Regression Estimator

Multicollinearity is introduced into the model-fitting exercise when trying to do mixture experiments with the built-in mixture contraint, and additional constraints on components or linear combinations of components. Therefore, we can use ridge regression estimator for overcoming multicollinearity.

From (3), the variance-covariance matrix of ridge regression estimator $\underline{b}(k)$, is

$$Var[\underline{b}(k)] = \sigma^2 (X'X+kI)^{-1}X'X(X'X+kI)^{-1}.$$

And, the predicted response value at $\underline{x}$ is

$$\begin{aligned} \hat{y}_k(\underline{x}) &= \underline{x}_0{}'\underline{b}(k) \\ &= \underline{x}_0{}'(X'X+kI)^{-1}X'\underline{y}. \end{aligned} \tag{4}$$

Therefore, the prediction variance at $\underline{x}$ is

$$Var[\hat{y}_k(\underline{x})] = \sigma^2 \underline{x}_0{}'(X'X+kI)^{-1}X'X(X'X+kI)^{-1}\underline{x}_0. \tag{5}$$

And, the prediction bias at $\underline{x}$ is

$$bias[\hat{y}_k(\underline{x})] = -k\,\underline{x}_0{}'(X'X+kI)^{-1}\underline{\beta}.$$

Thus, the squared prediction bias at $\underline{x}$ is

$$bias^2[\hat{y}_k(\underline{x})] = k^2\underline{\beta}'(X'X+kI)^{-1}\underline{x}_0\,\underline{x}_0{}'(X'X+kI)^{-1}\underline{\beta}. \tag{6}$$

This bias is not model bias, but estimator bias under the assumption that the model is correct. Let $MSE[\hat{y}_k(\underline{x})]$ be the mean square error of the prediction at $\underline{x}$. We then note that

$$MSE[\hat{y}_k(\underline{x})] = Var[\hat{y}_k(\underline{x})] + bias^2[\hat{y}_k(\underline{x})].$$

By Hoerl and Kennard(1970a), we can obtain the following facts.

Fact 1. For all $k>0$, $Var[\hat{y}(\underline{x})] > Var[\hat{y}_k(\underline{x})]$.

Fact 2. There exists $k>0$ such that $MSE[\hat{y}(\underline{x})] > MSE[\hat{y}_k(\underline{x})]$.

Through fact 1 and fact 2, we know that ridge regression estimator is superior to least squares estimator from standpoint of the prediction variance and the MSE when multicollinearity exists.

$bias^2[\hat{y}_k(\underline{x})]$ is a function not only of the particular location in the design region but also the unknown vector $\underline{\beta}$. A method for overcoming this difficulty is the use of the Euclidean norm of the vector $\underline{\beta}$. Thus, we can consider the following lemma 1.

Lemma 1. For a particular location in the design region, the maximum squared prediction bias, given the constraint that Euclidean norm of the vector $\underline{\beta}$, $|\underline{\beta}| = 1$, is

$$\max_{|\beta|=1} bias^2[\hat{y}_k(x)] = k^2 tr[A(k)] \tag{7}$$

where $A(k) = (X'X + kI)^{-1} x_0 \; x_0'(X'X + kI)^{-1}$.

Proof. $bias^2[\hat{y}_k(x)]$ is a quadratic form in β. Thus, we know that

$$\max_{|\beta|=1} \frac{1}{k^2} bias^2[\hat{y}_k(x)] = \lambda_1$$

where λ_1 is the largest eigenvalue of $A(k)$. Because $A(k)$ has rank 1 and is positive semidefinite,

$$tr[A(k)] = \sum_{i=1}^{p} \lambda_i = \lambda_1,$$

where λ_i are the eigenvalues of $A(k)$. Thus,

$$\max_{|\beta|=1} bias^2[\hat{y}_k(x)] = k^2 tr[A(k)].$$

$\max_{|\beta|=1} bias^2[\hat{y}_k(x)]$ is the maximum over the possible values for β. Therefore, we can $\max_{|\beta|=1} bias^2[\hat{y}_k(x)]$ as the numerical measure of the squared prediction bias at x.

When the mixture component proportions are restricted by lower and upper bounds of the form (1), these restrictions make the reference mixture (or overall centroid), the centroid of constrained simplex. When measuring the effect of component with respect to the predicted response value, the prediction variance, and the squared prediction bias, and a reference mixture other than the centroid of the simplex is to be used, Cox direction is generally appropriate. Cox direction of component i is an imaginary line projected from the reference mixture to the vertex $x_i = 1$. Cox direction was introduced by Cox (1971). Let us denote the proportions of the q components at the reference mixture by $c' = (c_1, c_2, \cdots, c_q)$ where $\sum_{i=1}^{q} c_i = 1$. When the proportion c_i of component i is changed by an amount Δ_i in Cox direction, so that the the new proportion becomes

$$x_i = c_i + \Delta_i, \tag{8}$$

the proportions of the remaining $q-1$ components resulting from the c_i in the ith component, are

$$x_j = c_j \frac{1 - c_i - \Delta_i}{1 - c_i}, \qquad j = 1, 2, \cdots, q, \quad j \neq i. \tag{9}$$

Note that the ratio of the proportions for components j and k, where x_j and x_k are defined by (9), is the same value as the ratio of components j and k at the reference mixture.

Let

$$V_k(x) = \frac{Var[\ \hat{y}_k(x)]}{\sigma^2} = x_0'(X'X+kI)^{-1}X'X(X'X+kI)^{-1}x_0 \tag{10}$$

and

$$B_k(x) = \max_{|\beta|=1} bias^2[\ \hat{y}_k(x)] = k^2 tr[A(k)]. \tag{11}$$

Using the idea of Cornell(1990) and Vining, Cornell, and Myers(1993), Jang and Yoon(1997) proposed the response trace and the prediction variance trace as a tool for evaluating the mixture component effects of the ridge regression estimator.

With the response trace and the prediction variance trace, I propose the prediction bias trace as a tool for evaluating the mixture component effects of the ridge regression estimator. The plot of $B_k(x)$ along Cox direction of each component for some k, the prediction bias trace, can be used to give comprehensive picture of the behavior of the maximum squared prediction bias charge due to ridge regression over Cox direction of each component under constrained region. These graphs - the response trace, the prediction variance trace, and the prediction bias trace - can be used to examine the mixture component effects of the ridge regression estimator on mixture designs with respect to the predicted value, the prediction variance, and the squared prediction bias, respectively.

3. Numerical Examples

My first example is taken from McLean and Anderson (1966). The purpose of the experiment was to find the combination of the proportions of magnesium (x_1), sodium nitrate (x_2), strontium nitrate (x_3), and binder (x_4) for producing flare with maximum illumination. McLean and Anderson (1966) suggested the 15-point extreme vertices design consisting of the eight extreme vertices, the centroids of the six faces and the overall centroid of the region along with the flare illumination data. A second-order polynomial was fit to the data. The component ranges are $0.40 \le x_1 \le 0.60$, $0.10 \le x_2 \le 0.50$, $0.10 \le x_3 \le 0.50$ and $0.03 \le x_4 \le 0.08$. We can use ridge regression because of multicollinearity in this example.

Using (10) and (11), we can draw Figure 1. Figure 1 compares the prediction variance traces and the prediction bias traces for the extreme vertices design for several k. As k increases, the prediction variance decrease gradually and the maximum squared prediction bias values increase gradually, but that when k is greater than 0.004, this decreasing trend in the prediction variance becomes very weak. When looking at these traces, it is important to keep in mind that as the prediction variance values are decreasing, regression estimators become more stable. Especially, the lowest prediction variance value of each component is located at the nearby center of each range when k is increasing. But, when k is increasing, the lowest maximum squared prediction bias values of component x_1, x_2 and x_3 are gradually located toward the boundaries of each component range, respectively, and the lowest maximum squared prediction bias value of component x_4 is located at the nearby center of the range of component x_4.

Cornell (1990) presented 15-point D-optimal design as a computer-generated design which is generated by ACED program and compete with extreme vertices design, consisting the eight extreme vertices, the centroids of the two faces and the midpoints of five edges of the region.

Using (10) and (11), we can draw Figure 2. Figure 2 compares the prediction variance traces and the prediction bias traces for the D-optimal design for several k.

From Figure 1 and Figure 2, we see that the D-optimal design is similar to the extreme vertices design with respect to the prediction variance when $k=0.004 \sim 0.005$. But, the D-optimal design is different from extreme vertices design with respect to the maximum squared prediction bias, especially in component x_2 and x_3 according to the increase of k.

My second example is taken from Snee (1975). The objective of the study was to determine the amount of additive (x_1) needed in three lubricant blends (x_2, x_3, x_4) so that a certain critical physical property would attain a desired level. Snee suggested 18-point design consisting of the ten extreme vertices, one mid-edge point, six face centroids, and the overall centroid. A second-order polynomial was fit to the data. The component ranges are $0.07 \le x_1 \le 0.18$, $0.0 \le x_2 \le 0.30$, $0.37 \le x_3 \le 0.70$ and $0.0 \le x_4 \le 0.15$. We can use the ridge regression because of multicollinearity in this example.

Using (10) and (11), we can draw Figure 3. Figure 3 compares the prediction variance traces and the prediction bias traces for the Snee design for several k. We can ascertain that as k increases, the prediction variance values decrease gradually and the maximum squared prediction bias values increase gradually, but that when k is greater are than 0.003, this decreasing trend becomes very weak. From Figure 3, we see that the lowest prediction variance value of each component is located at the nearby center of each range when k is increasing. But, when k increases, the lowest maximum squared prediction bias values of component x_2 and x_3 are gradually located toward the boundaries of each component range, respectively, and the lowest maximum squared prediction bias values of component x_1 and x_4 are located at the nearby center of each component range, respectively.

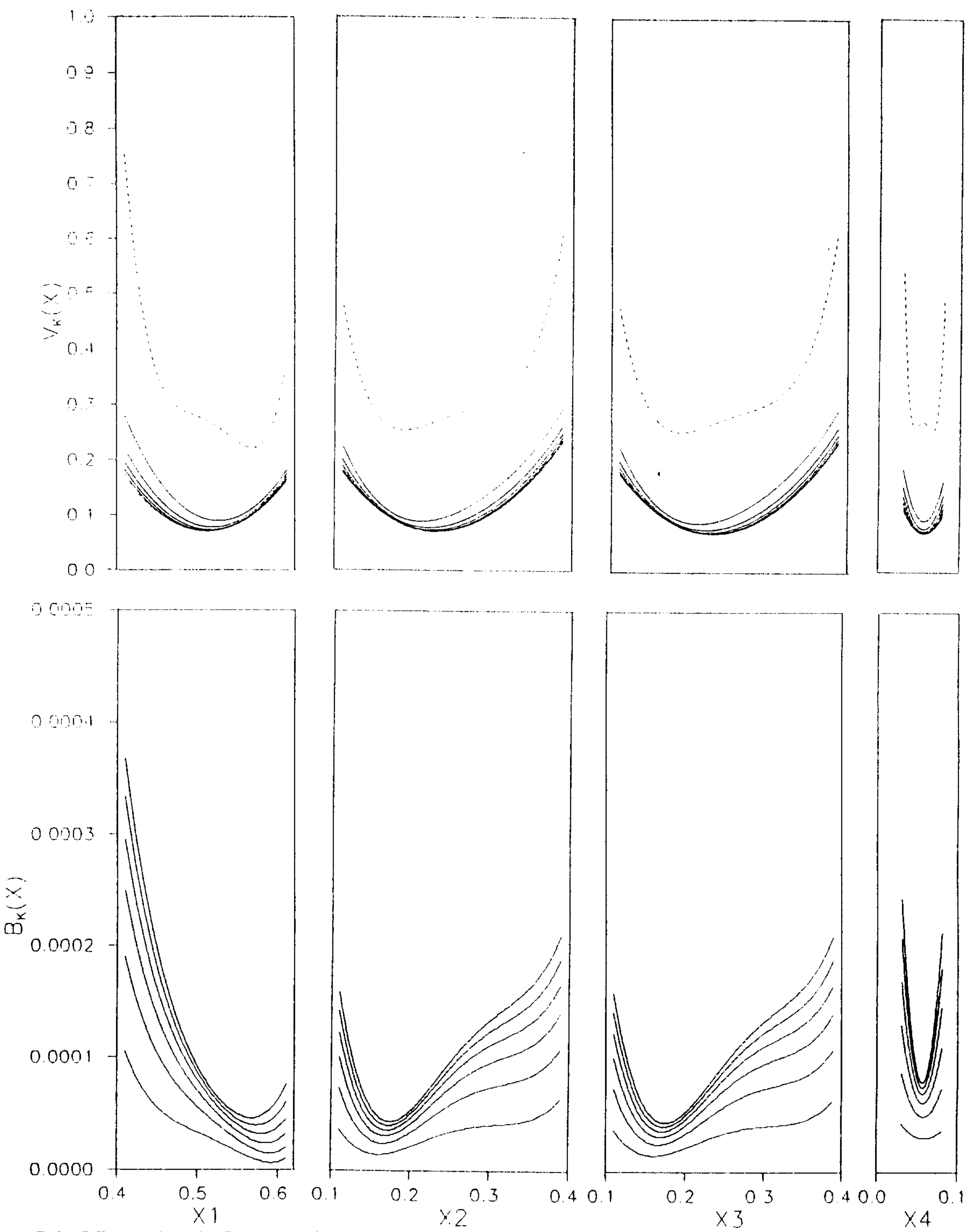

FIGURE 1. COMPARISON OF THE PREDICTION VARIANCE AND BIAS TRACES FOR THE EXTREME VERTICES DESIGN USING RIDGE REGRESSION (-----:K=0.000, ——:K=0.001,0.002,0.003,0.004,0.005, 0.006(VARIANCE:FROM TOP TO BOTTOM, BIAS:FROM BOTTOM TO TOP))

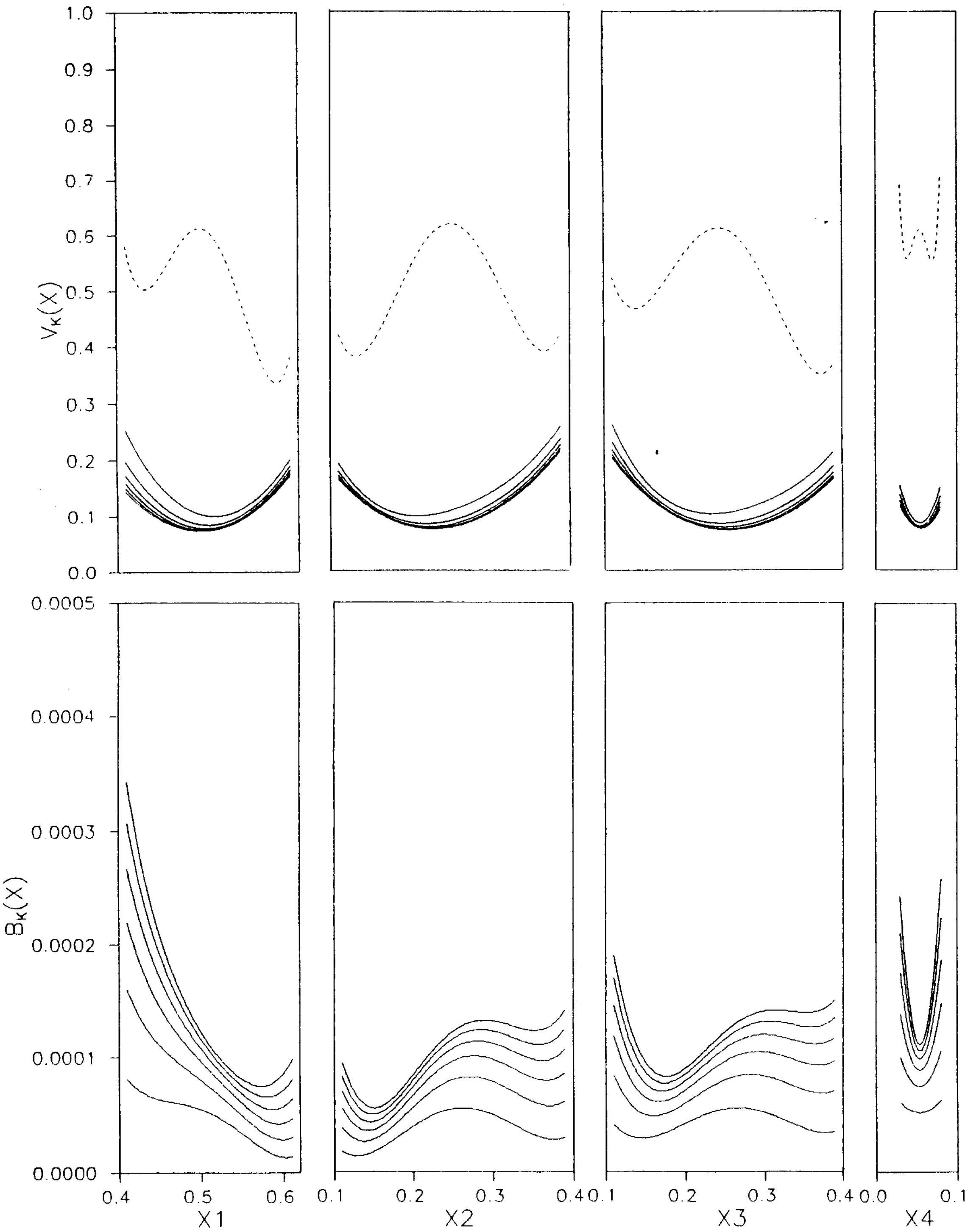

FIGURE 2. COMPARISON OF THE PREDICTION VARIANCE AND BIAS TRACES FOR THE D-optimal DESIGN USING RIDGE REGRESSION (-----: K=0.000, ———:K=0.001, 0.002, 0.003, 0.004, 0.005, 0.006 (VARIANCE:FROM TOP TO BOTTOM, BIAS:FROM BOTTOM TO TOP))

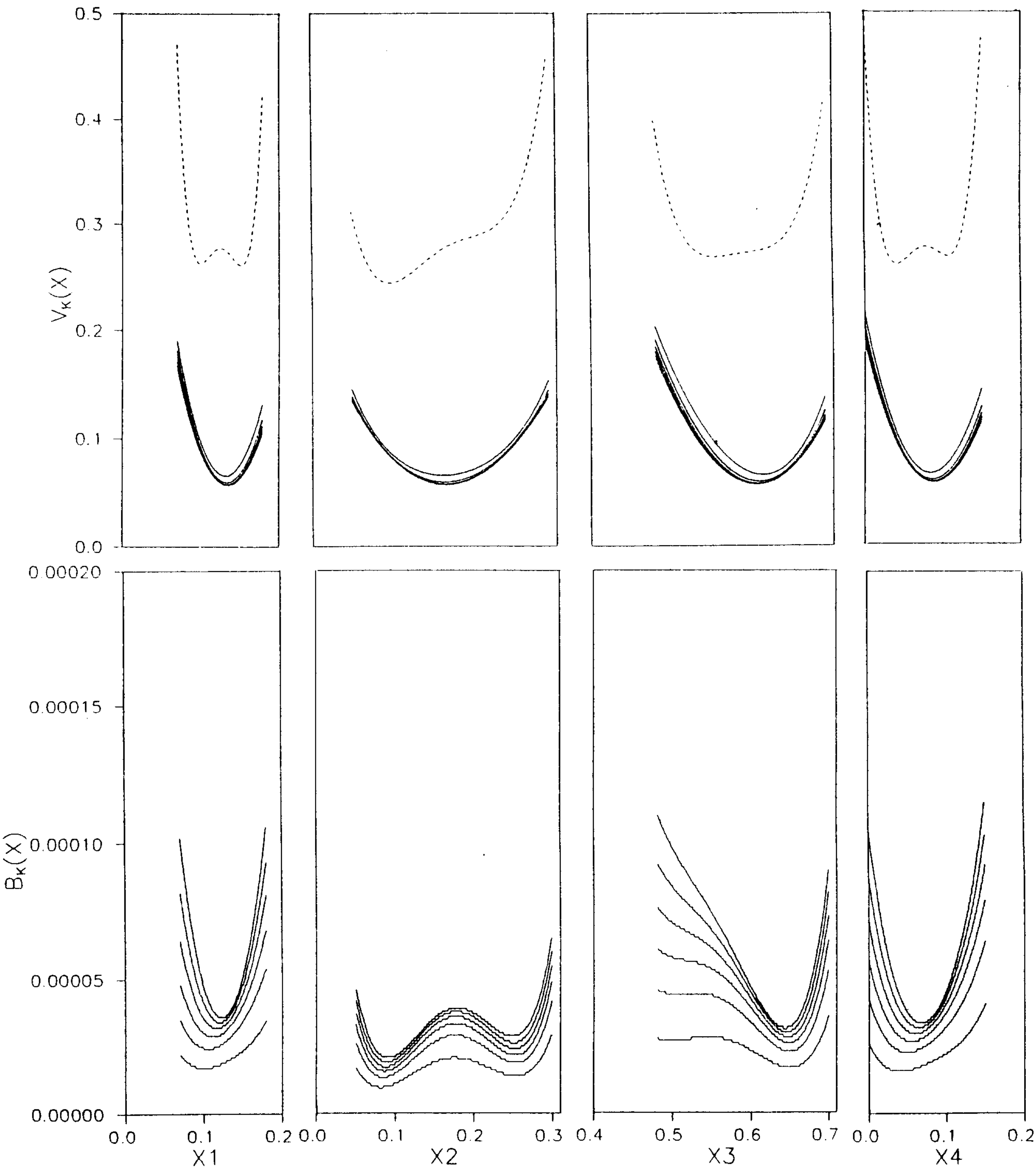

FIGURE 3. COMPARISON OF THE PREDICTION VARIANCE AND BIAS TRACES FOR THE Snee DESIGN USING RIDGE REGRESSION (------:K=0.000, ———:K=0.001, 0.002, 0.003, 0.004, 0.005, 0.006 (VARIANCE:FROM TOP TO BOTTOM, BIAS:FROM BOTTOM TO TOP))

4. Conclusion

In this paper, a graphical method for evaluating ridge regression estimator have been proposed. For mixture experiments, this graphical method - the prediction variance trace and the prediction bias trace - can be used as a tool for evaluating the mixture component effects of the ridge regression estimator.

References

[1] Cornell, J. A. (1990). *Experiments with Mixture : Designs, Models, and the Analysis of Mixture Data,* 2nd ed., John Wiley & Sons, Inc., New York.

[2] Cox, D. R. (1971). A Note on Polynomial Response Functions for Mixtures, *Biometrika,* **58,** 155 - 159.

[3] Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis,* 2nd ed., John Wiley & Sons, Inc., New York, 319 - 325.

[4] Hoerl, A. E. and Kennard, R. V. (1970a). Ridge Regression : Biased Estimation for Nonorthogonal Problems, *Technometrics,* **12,** 55 - 67.

[5] ——————(1970b). Ridge Regression : Applications to Nonorthogonal Problems, *Technometrics,* **12,** 69 - 82.

[6] Jang, D. H., and Yoon, M. (1997). Graphical Methods for Evaluating Ridge Regression Estimator in Mixture Experiments, *Communications in Statistics - Simulation and Computation,* **26,** 1049 - 1061.

[7] Marquardt, D. W. (1970). Generalized Inverse, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation, *Technometrics,* **12,** 591 - 612.

[8] McLean, R. A. and Anderson, V. L. (1966). Extreme Vertices Design of Mixture Experiments, *Technometrics,* **8,** 447 - 454.

[9] Snee, R. D. (1975). Experimental Designs for Quadratic Models in Constrained Mixture Spaces, *Technometrics,* **17,** 149 - 159.

[10] Vining, G. G., Cornell, I. A. and Myers, R. H. (1993). A Graphical Approach for Evaluating Mixture Designs, *Applied Statistics,* **42,** 127 - 138.