Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 806471, 9 pages
http://dx.doi.org/10.1155/2014/806471

Research Article

Graphical Evaluation of the Ridge-Type Robust Regression Estimators in Mixture Experiments

Ali Erkoc,[1] Esra Emiroglu,[2] and Kadri Ulas Akay[2]

[1] *Department of Statistics, Faculty of Science and Letters, Mimar Sinan Fine Arts University, Sisli, 34380 Istanbul, Turkey*
[2] *Department of Mathematics, Faculty of Science, Istanbul University, Beyazit, 34134 Istanbul, Turkey*

Correspondence should be addressed to Kadri Ulas Akay; kulas@istanbul.edu.tr

Received 27 March 2014; Revised 17 July 2014; Accepted 31 July 2014; Published 18 August 2014

Academic Editor: Rui C. Marques

Copyright © 2014 Ali Erkoc et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In mixture experiments, estimation of the parameters is generally based on ordinary least squares (OLS). However, in the presence of multicollinearity and outliers, OLS can result in very poor estimates. In this case, effects due to the combined outlier-multicollinearity problem can be reduced to certain extent by using alternative approaches. One of these approaches is to use biased-robust regression techniques for the estimation of parameters. In this paper, we evaluate various ridge-type robust estimators in the cases where there are multicollinearity and outliers during the analysis of mixture experiments. Also, for selection of biasing parameter, we use fraction of design space plots for evaluating the effect of the ridge-type robust estimators with respect to the scaled mean squared error of prediction. The suggested graphical approach is illustrated on Hald cement data set.

1. Introduction

Mixture experiments, a special study area of response surface methodology (RSM), are performed in many areas of product development and improvement. The purpose of mixture experiments is to model the blending surface with some forms of the mathematical equation. In general, for modeling well-behaved mixture system, the specific models such as Scheffé canonical polynomials are used in developing the appropriate response model. The parameter estimates of the Scheffé canonical polynomials are obtained by the ordinary least squares (OLS) method. However, in mixture experiments, multicollinearity is more frequent than the other experimental studies because of the constraints of components composing the mixture. Therefore, OLS estimates of the Scheffé canonical polynomials may be statistically insignificant with wrong sign and their variances are large so they may be far from the true values due to constraints on the mixture components [1–3]. To reduce the effect of multicollinearity in the analysis of mixture experiments, variable transformation or ordinary ridge regression (ORR), suggested by Hoerl and Kennard [4], is widely used as a classical method.

In mixture experiments, the detailed application of ORR estimator is given by St. John [1]. To evaluate the performance of ORR estimator, the graphical approaches apart from the analytical methods have been proposed in mixture experiments. These graphical approaches are based on plots of the scaled prediction variance (SPV) trace, which is suggested by Vining et al. [5]. The plot of SPV traces is used to compare the quality of prediction for designs in mixture experiments [5, 6]. Using the plot of SPV traces, Jang and Yoon [7] suggested a graphical method for evaluating and selecting an appropriate biasing parameter k that balances the bias-variance tradeoff. With this approach, the SPV values are plotted along prediction rays that correspond to Cox directions. On the other hand, Jang and Anderson-Cook [8] used fraction of design space (FDS) plots, which is developed by Zahran et al. [9], and violin plots to illustrate and evaluate the effect of ORR estimator with respect to the SPV values and to guide the decision about the value of biasing parameter k.

In mixture experiments, apart from the problems due to multicollinearity, another important problem which needs attention in parameter estimation is the presence of outliers in the data set [3, 10, 11]. In this case, outliers affect the efficiency of OLS estimates [12]. To overcome this problem, robust

estimation methods like M-estimator, least median of squares (LMS) estimator, least trimmed sum of squares (LTS) estimator, S-estimator, and MM-estimator are proposed (see [12–16]). However, although robust estimators are the resistant techniques used to obtain the parameter estimates which are not affected by outliers, they are frequently unstable when the design matrix is ill-conditioned [12]. Therefore, effects due to both outliers and multicollinearity in the analysis of mixture experiments can be reduced to a certain extent by using alternative approaches. One of these approaches is to use biased-robust estimators for the estimation of the coefficients. In literature, it is recommended to use alternative ridge-type robust estimators for both outliers and multicollinearity problem in data set (see [17–22]).

The purpose of this study is to apply and evaluate various ridge-type robust estimators in the analysis of mixture experiments. Also, the performance of ridge-type robust estimators is dependent on the biasing parameter. In this case, to determine the biasing parameter apart from the analytical methods, alternative plots which are used in comparison and evaluation of the designs and which are based on the SPV values or the scaled mean squared error of prediction (SMSEP) values can be used. Considering the SPV values in evaluating and comparing designs is important, but bias properties of designs should also be considered [23]. Therefore, our study will focus on the use of FDS plots as a graphical tool for evaluating the effect of various ridge-type robust estimators with respect to the SMSEP values and for guiding the decision about the value of the biasing parameter.

This paper is organized as follows. In the next section, the concise information about the analysis of mixture experiments with robust methods is given. In Section 3, the ridge-type robust estimators based on the M-estimator, GM-estimator, and MM-estimator are introduced. In Section 4, the FDS plots are introduced for the selection of biasing parameter k. The suggested approach is illustrated on Hald cement data in Section 5. Finally, the obtained results are given in Section 6.

2. Mixture Experiments

In mixture experiments, the measured response depends only on the proportions of the components present in the mixture and not on the amount of the mixture. For example, the measured response might be the strength of concrete which is a mixture of cement, sand, and water, or it might be octane rating of a blend of gasoline. The aim of these experiments is to find optimal component proportions that provide desired values of some product performance characteristics [24].

Let x_i represent the proportion of the ith component present in a q component mixture; then

$$0 \le x_i \le 1, \ i = 1, 2, \ldots, q, \quad \sum_{i=1}^{q} x_i = 1. \tag{1}$$

The mixture constraints (1) produce a regular $(q-1)$-dimensional simplex experimental region. However, the experimental region (1) is sometimes subjected to additional constraints on individual mixture components of the form

$$0 \le L_i \le x_i \le U_i \le 1 \quad i = 1, 2, \ldots, q, \tag{2}$$

where L_i and U_i denote lower and upper bounds, respectively. In general, restriction (2) reduces the experimental region given by (1) to an irregular $(q-1)$-dimensional hyperpolyhedron. Mixture experimental designs are discussed by Cornell [24] for regular and irregular simplex regions.

In general, the presence of natural constraints or/and some additional linear constraints between the components requires the use of specific polynomial models such as Scheffé canonical polynomials for the modeling of the mixture system [25]. For modeling well-behaved mixture system, generally the Scheffé canonical polynomials are sufficient [24]. The first- and second-degree Scheffé canonical polynomial are given as

$$\begin{aligned} y &= \sum_{i=1}^{q} \beta_i x_i \\ y &= \sum_{i=1}^{q} \beta_i x_i + \sum_{i=1}^{q} \sum_{i<j}^{q} \beta_{ij} x_i x_j. \end{aligned} \tag{3}$$

As usual, we can represent the Scheffé canonical polynomial models in matrix form by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{4}$$

where $\mathbf{y}$ is $n \times 1$ vector of observations on the response variable, $\mathbf{X}$ is $n \times p(\ge q)$ matrix, where p is number of the terms in the model, $\boldsymbol{\beta}$ is $p \times 1$ vector of parameters to be estimated, and $\boldsymbol{\varepsilon}$ is $n \times 1$ vector of random errors satisfying $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$. The vector of unknown parameters $\boldsymbol{\beta}$ can be generally estimated using OLS method. The OLS estimator for $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}. \tag{5}$$

When the observations $\mathbf{y}$ in model (4) are normally distributed, the OLS is a good parameter estimation procedure in the sense that it produces an estimator of the parameter vector $\boldsymbol{\beta}$ that has good statistical properties [12]. For example, $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$ and also the maximum-likelihood estimator of $\boldsymbol{\beta}$. On the other hand, there are many situations where the distribution of the response variable is nonnormal. Under conditions of nonnormal distributions, particularly heavy-tailed distributions, OLS estimator no longer has these desirable properties. These heavy-tailed distributions tend to generate outliers, and these outliers may have improper effect on the OLS estimates. To deal with this, several robust-to-outliers methods have been proposed in the literature [12, 16].

Robust regression methods are well-designed statistical techniques to produce regression coefficients which are not sensitive to outliers and nonnormal error distributions [13]. The most popular of all robust estimation techniques is M

estimation, proposed by Huber [26]. The M-estimator is defined as

$$\widehat{\boldsymbol{\beta}}_M = \arg\min_{\boldsymbol{\beta}} \sum_{u=1}^{n} \rho\left(\frac{e_u(\boldsymbol{\beta})}{s}\right), \tag{6}$$

where e_u denote the residuals as $e_u = e_u(\widehat{\boldsymbol{\beta}}) = y_u - \mathbf{x}_u'\widehat{\boldsymbol{\beta}}$, where $\mathbf{x}_u'$ denote the uth row of $\mathbf{X}$, and s is a robust estimate of scale. The scale s is required to gain scale equivariance and either can be an external scale estimate or can be estimated simultaneously. The function ρ is related to the likelihood function for an appropriate choice of the error distribution. Different choices for ρ lead to different variants of M-estimators. As with most M-estimation procedures, both the Huber weight function and the Tukey biweight function are two common choices [16].

To minimize (6), equate the first partial derivatives of ρ with respect to each parameter which leads to p equations of the form

$$\sum_{u=1}^{n} x_{uj}\psi\left(\frac{e_u(\boldsymbol{\beta})}{s}\right) = 0, \quad j = 1, 2, \ldots, p, \tag{7}$$

where $\psi(\cdot)$ is the partial derivative of ρ with respect to parameters. In general, the ψ function in (7), a nonlinear function, must be solved by iterative methods. The most common solution is iteratively reweighted least squares (IRLS), resulting in an estimator

$$\widehat{\boldsymbol{\beta}}_M = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}, \tag{8}$$

where $\mathbf{W}$ is an $n \times n$ diagonal matrix of weights $w_u = \psi(e_u/s)/(e_u/s)$, $u = 1, 2, \ldots, n$. IRLS estimation procedure updates $\mathbf{W}$ matrix until a convergence criterion has been met [12].

On the other hand, M-estimators are not robust to high leverage points. As a remedy to this problem, bounded-influence generalized M-estimators (GM-estimators) have been proposed. The most commonly used GM-estimator is the Schweppe-type estimator. Handschin et al. [27] proposed an objective function with associated normal equations of the form

$$\sum_{u=1}^{n} \psi\left(\frac{e_u(\boldsymbol{\beta})}{\pi_u s}\right)\pi_u \mathbf{x}_u = 0, \tag{9}$$

where, for appropriate values of $\pi_u = \pi_u(\mathbf{x}_u)$, the GM-estimator can downweight outliers with high leverage points. Several suggestions for π_u have been made that involve typical OLS outlier diagnostics, including studentized residuals or DFFITS [28]. The resulting estimator is (8), except the weights which become $w_u = \psi(e_u/\pi_u s)/(e_u/\pi_u s)$. In the literature, several GM estimation approaches are suggested using various combinations of GM components (objective function, initial estimate, scale estimate, π function, etc.) [29, 30].

The breakdown point and efficiency are two practical concerns that should be taken into account when selecting a robust estimation procedure [12, 16]. According to these properties, GM-estimators possess the same efficiency and asymptotic distributional properties as M-estimators. But their breakdown points may not be much better than the worst case $1/n$. In these cases, both M-estimation and GM-estimation can be improved by starting with a good initial estimate. In general, an alternative desirable initial estimator with a high breakdown point can be S-estimator [29, 30].

S-estimator proposed by Rousseeuw and Yohai [14] is based on minimizing the dispersion of the residuals. The objective function for S-estimation is

$$\widehat{\boldsymbol{\beta}}_S = \arg\min_{\boldsymbol{\beta}} \widehat{\sigma}_S\left(e_1(\boldsymbol{\beta}), \ldots, e_n(\boldsymbol{\beta})\right), \tag{10}$$

where the dispersion function $\widehat{\sigma}_S$ is found implicitly as the solution to

$$\frac{1}{n}\sum_{u=1}^{n} \rho\left(\frac{e_u(\boldsymbol{\beta})}{\widehat{\sigma}_S}\right) = \kappa, \tag{11}$$

where κ is a tuning constant defined as $\kappa = E_\Phi[\rho(e)]$ and Φ represent the standard normal distribution. The solution $\widehat{\sigma}_S$ is called an M-estimator of scale. But it is impossible for S-estimator to achieve a high breakdown point as well as a high efficiency [16].

Other most commonly used robust estimators, proposed by Yohai [15], are MM-estimators, which combine high breakdown with high asymptotic efficiency. Estimates are a modification of the typical M-estimates and are obtained through a three-step process [16]. Let $\widehat{\boldsymbol{\beta}}_S$ be an S-estimator, and let $\widehat{\sigma}_S$ be the corresponding M-estimator of scale, solving (11) for $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}_S$. The MM-estimator is then defined as local solution of

$$\widehat{\boldsymbol{\beta}}_{\mathrm{MM}} = \arg\min_{\boldsymbol{\beta}} \sum_{u=1}^{n} \rho\left(\frac{e_u(\boldsymbol{\beta})}{\widehat{\sigma}_S}\right) \tag{12}$$

obtained with IRLS and initial value $\widehat{\boldsymbol{\beta}}_S$. The initial S-estimator guarantees a high breakdown point and the final MM-estimate a high Gaussian efficiency. In both the initial S-estimator and the final MM-estimator, a standard choice of ρ is Tukey biweight function. The tuning constant κ can be set to 1.547 for the S-estimator to guarantee a 50% breakdown point, and it can be set to 4.685 for the second step MM-estimator in (12) to guarantee a 95% efficiency of the final estimator [16].

3. Ridge-Type Robust Estimators

Multicollinearity and outliers are two main problems for researchers using regression estimation methods. In the presence of multicollinearity, several alternative estimation techniques are proposed. One of them is the ORR estimator which was proposed by Hoerl and Kennard [4]. The ORR estimator which is the most popular method for dealing with multicollinearity is defined by

$$\widehat{\boldsymbol{\beta}}_{\mathrm{ORR}} = \left(\mathbf{X}'\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}, \quad k > 0, \tag{13}$$

where k is called biasing parameter. Since $\widehat{\boldsymbol{\beta}}_{\mathrm{ORR}}$ depends on $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$, $\widehat{\boldsymbol{\beta}}_{\mathrm{ORR}}$ is sensitive to outliers. Therefore, outliers may have also improperly effect on the ridge estimates. The problem is further complicated when both outliers and multicollinearity are present in the data [31].

In recent years, major efforts have been made to obtain reliable estimates especially in the presence of outliers and also multicollinearity. When both outliers and multicollinearity occur in a data set, it would seem beneficial to combine methods designed to deal with these problems individually. It is hoped that the resulting ridge-type robust estimators will be resistant to outliers and less affected by multicollinearity [19, 21, 31].

A ridge-type M-estimator, suggested by Silvapulle [19], is defined by

$$\widehat{\boldsymbol{\beta}}_{\mathrm{RM}} = \left(\mathbf{X}'\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}}_{M} = \mathbf{Z}\widehat{\boldsymbol{\beta}}_{M}, \quad k > 0, \tag{14}$$

where $\widehat{\boldsymbol{\beta}}_{M}$ is the M-estimators of $\boldsymbol{\beta}$. Silvapulle [19] shows that $\mathrm{MSE}[\widehat{\boldsymbol{\beta}}_{\mathrm{RM}}] < \mathrm{MSE}[\widehat{\boldsymbol{\beta}}_{\mathrm{ORR}}]$ is obtained under some conditions and a positive k_1 exists which depends on the unknown parameters such that $\mathrm{MSE}[\widehat{\boldsymbol{\beta}}_{\mathrm{RM}}] < \mathrm{MSE}[\widehat{\boldsymbol{\beta}}_{M}]$, for $0 < k < k_1$.

On the other hand, Maronna [22] stated that $\mathbf{Z}$ matrix and $\widehat{\boldsymbol{\beta}}_{M}$ estimator in (14) are also sensitive to leverage points. Therefore, a ridge-type robust estimator which is not sensitive to both x- and y-outliers is given by

$$\widehat{\boldsymbol{\beta}}_{\mathrm{RR}} = \left(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\widehat{\boldsymbol{\beta}}_{R}, \quad k > 0, \tag{15}$$

where $\widehat{\boldsymbol{\beta}}_{R}$ is one of the robust estimators (such as GM-estimator or MM-estimator) of $\boldsymbol{\beta}$. In this case, the use of ridge-type robust estimators is a good method for obtaining more stable parameter estimates for the model.

The ridge-type robust estimator given in (15) depends on the biasing parameter k. For selection of biasing parameter, Silvapulle [19] proposed the two different types of adaptive choices based on Hoerl et al. [32] (abbreviated to HKB) and Lawless and Wang [33] (abbreviated to LW). The same results can be extended to $\widehat{\boldsymbol{\beta}}_{\mathrm{RR}}$ estimators. The resulting two choices of k are

$$\widehat{k}_{\mathrm{HKB}} = \frac{p\widehat{A}^2}{\left\|\widehat{\boldsymbol{\beta}}_{R}\right\|^2}, \qquad \widehat{k}_{\mathrm{LW}} = \frac{p\widehat{A}^2}{\sum_{i=1}^{p}\lambda_i\beta_{Ri}^2}, \tag{16}$$

where $\widehat{A}^2 = s^2\{(n-p)^{-1}\sum\psi^2(e_u/s)\}\{n^{-1}\sum\psi'(e_u/s)\}^{-2}$ and λ_j is the jth eigenvalue of $\mathbf{X}'\mathbf{W}\mathbf{X}$. Also for ORR estimator, there are several methods for selecting the biasing parameter k in literature. Most of these methods have been reviewed by Clark and Troskie [34]. Recommended methods can be similarly extended to the ridge-type robust estimators.

4. The Graphical Method for Evaluating Biasing Parameter

Due to the fact that there is not a definite rule for selection of biasing parameter, the graphical approaches developed for the ORR estimator can be used to select the biasing parameter of the ridge-type robust estimators. Different than analytical methods, determination of the biasing parameter with graphical approach is useful since it helps the researcher to examine the prediction properties of design based on the different biasing parameters visually. The prediction properties of a design, specifically the SMSEP values, also depend on the fitted model. Therefore multicollinearity due to the fitted model also influences the performance of the prediction properties of design completely. In this case, to examine the graphical approaches developed for comparison of the performance of designs based on the different biasing parameter for a single design is a useful method [7, 8].

For using the suggested methods on the analysis of mixture experiments, firstly we define the predicted response at $\mathbf{x}_u$. When we use $\widehat{\boldsymbol{\beta}}_{\mathrm{RR}}$ estimator, the predicted response value at $\mathbf{x}_u$ is given as

$$\widehat{y}_k(\mathbf{x}_u) = \mathbf{x}_u'\widehat{\boldsymbol{\beta}}_{\mathrm{RR}} = \mathbf{x}_u'\left(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\widehat{\boldsymbol{\beta}}_{R}. \tag{17}$$

The mean squared error of $\widehat{y}_k(\mathbf{x}_u)$, also known as the mean squared error of prediction (MSEP), is given by

$$\mathrm{MSE}\left[\widehat{y}_k(\mathbf{x}_u)\right] = \mathrm{var}\left[\widehat{y}_k(\mathbf{x}_u)\right] + \left\{\mathrm{Bias}\left[\widehat{y}_k(\mathbf{x}_u)\right]\right\}^2, \tag{18}$$

where $\mathrm{Bias}[\widehat{y}_k(\mathbf{x}_u)] = E[\widehat{y}_k(\mathbf{x}_u)] - y(\mathbf{x}_u)$ is the bias associated with estimating $\widehat{y}(\mathbf{x}_u)$.

The variance of $\widehat{y}_k(\mathbf{x}_u)$, also known as the prediction variance, is given by

$$\begin{aligned}\mathrm{var}\left[\widehat{y}_k(\mathbf{x}_u)\right] = {} & \mathbf{x}_u'\left(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\mathrm{var}\left[\widehat{\boldsymbol{\beta}}_{R}\right] \\ & \times \mathbf{X}'\mathbf{W}\mathbf{X}\left(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{x}_u,\end{aligned} \tag{19}$$

where $\mathrm{var}[\widehat{\boldsymbol{\beta}}_{R}]$ is the variance-covariance matrix of $\widehat{\boldsymbol{\beta}}_{R}$. There are various approximations for determination of covariance matrix of the robust estimators in the literature [13]. Huber [13] has shown that a reasonable approximation for the covariance matrix of $\widehat{\boldsymbol{\beta}}_{R}$ is $\mathrm{var}[\widehat{\boldsymbol{\beta}}_{R}] \approx K^2\widehat{A}^2(\mathbf{X}'\mathbf{X})^{-1}$, where $K = 1 + (p/n)(\mathrm{var}(\psi'(e_u))/[E(\psi'(e_u))]^2)$ is correction factor.

Using (17), the bias of $\widehat{y}_k(\mathbf{x}_u)$ at $\mathbf{x}_u$ is approximately given by

$$\mathrm{Bias}\left[\widehat{y}_u(\mathbf{x}_u)\right] \approx \mathbf{x}_u'\left[\left(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{X} - \mathbf{I}\right]\boldsymbol{\beta}. \tag{20}$$

By combining the result in (19) and (20), an approximate expression for the MSE of $\widehat{y}_k(\mathbf{x}_u)$ is given by

$$\begin{aligned}\mathrm{MSE}\left[\widehat{y}_k(\mathbf{x}_u)\right] \approx {} & K^2A^2\mathbf{x}_u'\left(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ & \times\mathbf{X}'\mathbf{W}\mathbf{X}\left(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{x}_u + k^2\mathbf{x}_u' \\ & \times\left(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I}\right)^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{x}_u.\end{aligned} \tag{21}$$

Note that $\mathrm{MSE}[\widehat{y}_k(\mathbf{x}_u)]$ is a function of the biasing parameter k, the location of mixture $\mathbf{x}_u$ in the design space, the model matrix, and the weights matrix. It should also be noted that

MSE$[\widehat{y}_k(\mathbf{x}_u)]$ depends on $\boldsymbol{\beta}$, which is unknown. To calculate the MSE$[\widehat{y}_k(\mathbf{x}_u)]$ in (21), the unknown parameters can be replaced by suitable estimates of the parameters.

When design comparison focuses on prediction, the MSEP is generally scaled by n/σ^2 to create a scale-free quantity that reflects the design size. The scaled mean squared error of prediction (SMSEP) at the uth observation for a ridge-type robust estimator is written as

$$\text{SMSE}\left[\widehat{y}_k\left(\mathbf{x}_u\right)\right] = \frac{n}{s^2}\text{MSE}\left[\widehat{y}_k\left(\mathbf{x}_u\right)\right], \tag{22}$$

where n denotes the design size. The multiplication by n serves to adjust the MSE for the total number of observations. This practice, which is quite common in RSM, makes it possible to use the SMSEP as a design criterion for comparing several designs [6].

In an experimental region, determination of the mixture points $\mathbf{x}_u$, SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ values along with a constant n value can be compared for different k values. Instead of comparing SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ values arbitrarily, a graphical approach which examines the distribution of the values of SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ on the experimental region can be used. So, an appropriate $k > 0$ value which achieves SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ < SMSE$[\widehat{y}(\mathbf{x}_u)]$ condition on the experimental region can be determined.

To evaluate the distribution of SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ values on the constrained experimental design, the FDS plots which are used in comparison and evaluation of the designs can be used. The main advantage of the use of FDS plots is to display characteristics of SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ values throughout a multidimensional design region on a two-dimensional graph. Therefore, we can use the FDS plots as a graphical tool for evaluating the effect of ridge-type robust estimators with respect to the SMSEP values and for selecting an appropriate value of the biasing parameter k. In this case, instead of just considering a single FDS curve for a design, we examine the impact of the biasing parameter on the MSEP performance throughout the experimental region.

To calculate the FDS, two sampling methods with nearly same results are generally explored to sample mixture points $\mathbf{x}_u$ from the design space [35]. The first uses shrunken regions by Piepel et al. [36] and randomly selects a number of points on each shrinkage level proportional to that level. The second method selects points completely at random that fit within the constraints of the region [35]. In this study, we will use second method for generation of the sample points on the experimental region. Once the points are generated, the SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ values are calculated for each point. The empirical cumulative distribution function is then calculated for the SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ values and plotted on the FDS plot. The minimum SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ value is shown at FDS of 0 and the maximum value is shown at the fraction 1. Along the FDS curve, the fraction of the design space at or below a particular SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ value can be determined. A good design starts with small values and remains flat throughout. Therefore, with FDS plot, we see the relationship between the different biasing parameters k and we see the pattern of SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ changing as k increases. By examining several different values of k, we can determine the smallest possible value of k, which stabilizes the SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ distribution.

5. Example: The Hald Cement Data

To evaluate the performances of ridge-type robust estimators, we will draw on a mixture experiment given by Piepel and Redgate [37]. Piepel and Redgate [37] pointed out that the data from a cement example published by Hald [38] can be analyzed in the framework of the mixture experiments. The purpose of the experiment is to investigate the influence of the components on the heat of hardening (calories/gram), measured after the cements cured for 180 days. The data given in Table 1 contains the x_i values (weight fractions, wf) computed from the 13 oxide weight percent clinker compositions. Additional information on the analyzing of the Hald cement experiment can be found in Piepel and Redgate [37] and Smith [3].

When we use first-degree Scheffé canonical polynomial, the obtained model is as follows:

$$\widehat{y} = \underset{(33.64)}{-433.7x_1} + \underset{(64.35)}{54.9\,x_2} - \underset{(32.21)}{252.7x_3} + \underset{(88.58)}{56.7\,x_4} + \underset{(16.76)}{309.2x_5}, \tag{23}$$

where the quantities in parentheses are estimated as standard errors of the coefficient estimates. Statistics for detecting influential observations for the Hald cement data set are given in Table 2.

Based on the result of Table 2, e_8 ordinary residual seems suspiciously large. Also, observation 3 has a high leverage value. The parameter estimates of the robust estimators are given in Table 3. In Table 3, the quantities in parentheses are estimated as standard errors of the coefficient estimates and AICR is the robust version of Akaike Information Criterion which is used for model selection [16].

Table 4 displays the final weights for robust estimators. In M-estimation, there is considerably downweight only at point 3. In case of GM-estimation, both point 3 and point 8 are significantly downweight. On the other hand, all observations are moderately downweight in MM-estimation. But, we note that neither point 3 nor point 8 is significantly downweight in MM-estimation, despite the fact that these points have the largest Cook's distance and ordinary residual, respectively.

In addition, the condition number of $\mathbf{X}'\mathbf{X}$ matrix is 101.034. So, the design matrix $\mathbf{X}$ is moderately ill-conditioned. Therefore, due to the multicollinearity and outliers, the ridge-type GM-estimator or the ridge-type MM-estimator can be used. For the ridge-type GM-estimator, we can calculate the biasing parameter as $\widehat{k}_{\text{HKB}} \approx \widehat{k}_{\text{LW}} = 0.00001$. For the ridge-type MM-estimator, the biasing parameter can be estimated as $\widehat{k}_{\text{HKB}} = 0.00007$ and $\widehat{k}_{\text{LW}} = 0.00002$ based on the methods proposed by Silvapulle [19].

To evaluate the biasing parameter k for the ridge-type GM-estimator and the ridge-type MM-estimator, randomly generated mixture points on the experimental region are considered. For different k values, the FDS plots of SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ values based on $\widehat{\boldsymbol{\beta}}_{\text{GM}}$ and $\widehat{\boldsymbol{\beta}}_{\text{MM}}$ estimators are given in Figures 1 and 2, respectively.

In Figures 1 and 2, as the biasing parameter k increases, the distribution of SMSE$[\widehat{y}_k(\mathbf{x}_u)]$ values initially decreases gradually and then increases dramatically. In this case, we

Table 1: Hald cement data.

ID	x_1 SiO_2 (wf)	x_2 Al_2O_3 (wf)	x_3 Fe_2O_3 (wf)	x_4 MgO (wf)	x_5 CaO (wf)	y H_{180} (cal/g)
1	0.2742	0.0376	0.0198	0.0248	0.6436	78.5
2	0.2600	0.0350	0.0510	0.0230	0.6310	74.3
3	0.2181	0.0568	0.0279	0.0498	0.6474	104.3
4	0.2465	0.0581	0.0281	0.0240	0.6433	87.6
5	0.2500	0.0390	0.0210	0.0240	0.6660	95.9
6	0.2226	0.0619	0.0279	0.0239	0.6637	109.2
7	0.2098	0.0462	0.0572	0.0211	0.6657	102.7
8	0.2357	0.0481	0.0722	0.0221	0.6219	72.5
9	0.2220	0.0464	0.0616	0.0232	0.6468	93.1
10	0.2129	0.0876	0.0119	0.0249	0.6627	115.9
11	0.2252	0.0501	0.0751	0.0220	0.6276	83.8
12	0.2132	0.0611	0.0290	0.0260	0.6707	113.3
13	0.2183	0.0558	0.0269	0.0239	0.6751	109.4

Table 2: Statistics for detecting influential observations for the Hald cement data.

ID	y	$\widehat{y}$	e_i	r_i	$e_{(i)}$	h_{ii}	Cook's Distance	DFITS
1	78.5	78.532	−0.032	−0.020	−0.019	0.489	0.000	−0.018
2	74.3	72.703	1.596	0.856	0.840	0.286	0.059	0.533
3	104.3	104.495	−0.195	−0.863	−0.848	0.989	14.074	−8.240
4	87.6	89.487	−1.887	−0.979	−0.976	0.236	0.059	−0.543
5	95.9	95.721	0.178	0.101	0.095	0.362	0.001	0.071
6	109.2	106.403	2.796	1.370	1.465	0.144	0.063	0.603
7	102.7	104.108	−1.408	−0.835	−0.817	0.415	0.099	−0.689
8	72.5	75.722	−3.222	−1.845	−2.277	0.373	0.406	−1.758
9	93.1	92.049	1.050	0.530	0.505	0.195	0.013	0.249
10	115.9	115.777	0.122	0.101	0.094	0.699	0.004	0.144
11	83.8	81.430	2.369	1.342	1.426	0.359	0.202	1.069
12	113.3	112.417	0.882	0.446	0.422	0.197	0.009	0.209
13	109.4	111.649	−2.249	−1.176	−1.210	0.249	0.092	−0.697

can select a reasonably positive small value of k at which the distribution of $\mathrm{SMSE}[\widehat{y}_k(\mathbf{x}_u)]$ is stable and $\mathrm{SMSE}[\widehat{y}_k(\mathbf{x}_u)] < \mathrm{SMSE}[\widehat{y}(\mathbf{x}_u)]$ condition is achieved. For the ridge-type GM-estimator, the biasing parameter k is approximately chosen as $k = 0.00001$. Also, for the ridge-type MM-estimator, we can select the biasing parameter k as $k = 0.0001$. The obtained models along with $\mathrm{MSE}[\widehat{y}_k(\mathbf{x}_u)]$ values are given in Table 5.

In Tables 3 and 5, we observe that $\mathrm{MSE}[\widehat{y}_{k=0.00001}(\mathbf{x}_u)] < \mathrm{MSE}[\widehat{y}(\mathbf{x}_u)]$ is achieved for the ridge-type GM-estimator. Also, when we select the biasing parameter k as $k = 0.0001$, the same result is obtained for the ridge-type MM-estimator. A more comprehensive comparison of the ridge-type GM-estimator and the ridge-type MM-estimator under MSEP sense is given in Figure 3.

From Figure 3, it can be seen that $\mathrm{MSE}[\widehat{y}_k(\mathbf{x}_u)]_{(\widehat{\boldsymbol{\beta}}_{\mathrm{GM}})}$ is smaller than $\mathrm{MSE}[\widehat{y}(\mathbf{x}_u)]_{(\widehat{\boldsymbol{\beta}}_{\mathrm{GM}})}$ for $k < 0.00012$. Also, $\mathrm{MSE}[\widehat{y}_k(\mathbf{x}_u)]_{(\widehat{\boldsymbol{\beta}}_{\mathrm{MM}})}$ is smaller than $\mathrm{MSE}[\widehat{y}(\mathbf{x}_u)]_{(\widehat{\boldsymbol{\beta}}_{\mathrm{MM}})}$ for $k < 0.00029$. In addition, when we select the biasing parameter k as $k < 0.0006$, $\mathrm{MSE}[\widehat{y}_k(\mathbf{x}_u)]_{(\widehat{\boldsymbol{\beta}}_{\mathrm{GM}})}$ is smaller than $\mathrm{MSE}[\widehat{y}_k(\mathbf{x}_u)]_{(\widehat{\boldsymbol{\beta}}_{\mathrm{MM}})}$. For the values of biasing parameter k larger than $k > 0.0006$, $\mathrm{MSE}[\widehat{y}_k(\mathbf{x}_u)]_{(\widehat{\boldsymbol{\beta}}_{\mathrm{MM}})}$ is smaller than $\mathrm{MSE}[\widehat{y}_k(\mathbf{x}_u)]_{(\widehat{\boldsymbol{\beta}}_{\mathrm{GM}})}$. As a result, which ridge-type robust-estimator is the best seems to be dependent on the robust estimator of $\boldsymbol{\beta}$ and the choice of biasing parameter k. We have to replace these unknown parameters by some suitable estimates in practice. Therefore, the results of the graphical findings may change.

6. Conclusions

In this study, we used various ridge-type robust estimators to reduce the effect of the multicollinearity and outliers on the model parameters to certain extent in the analysis of

Table 3: The parameter estimates of the robust estimators.

$\widehat{\boldsymbol{\beta}}_R$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{s}$	AICR	MSE$[\widehat{y}(\mathbf{x}_u)]$
$\widehat{\boldsymbol{\beta}}_M$	−431.244 (35.823)	59.539 (68.522)	−248.274 (34.294)	57.195 (94.323)	307.774 (17.845)	2.118	15.196	27.599
$\widehat{\boldsymbol{\beta}}_{\mathrm{GM}}$	−444.933 (27.238)	44.524 (52.10)	−222.962 (26.075)	677.678 (71.717)	290.069 (13.568)	1.335	72.534	15.956
$\widehat{\boldsymbol{\beta}}_{\mathrm{MM}}$	−432.122 (35.199)	56.987 (67.328)	−250.218 (33.697)	57.286 (92.680)	308.365 (17.534)	2.519	10.304	26.690

Table 4: Final weights from robust regressions.

ID	M-estimate	GM-estimate	MM-estimate
1	1	1	0.9999
2	1	1	0.9669
3	1	0.0015	0.9994
4	1	1	0.9476
5	1	1	0.9993
6	1	1	0.8896
7	1	1	0.9715
8	0.8268	0.4595	0.8456
9	1	1	0.9856
10	1	1	0.9997
11	1	1	0.9281
12	1	1	0.9879
13	1	1	0.9314

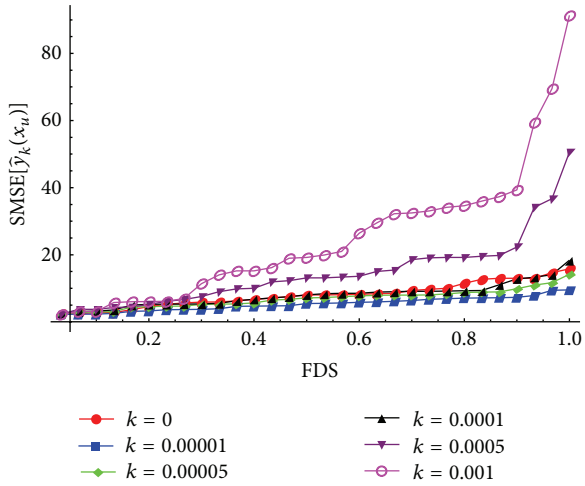

Figure 1: The FDS plots for different k values according to the randomly generated mixture, where $\widehat{\boldsymbol{\beta}}_R = \widehat{\boldsymbol{\beta}}_{\mathrm{GM}}$.

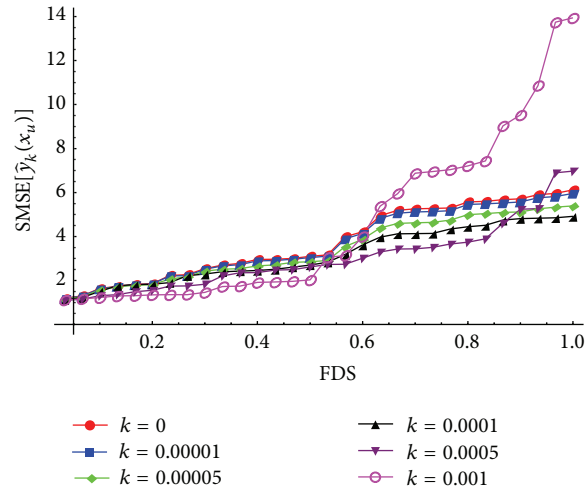

Figure 2: The FDS plots for different k values according to the randomly generated mixture, where $\widehat{\boldsymbol{\beta}}_R = \widehat{\boldsymbol{\beta}}_{\mathrm{MM}}$.

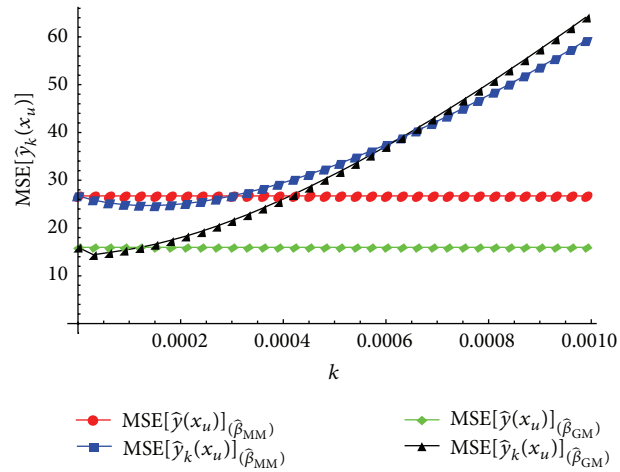

Figure 3: Comparison of the ridge-type GM-estimator and the ridge-type MM-estimator under MSEP sense.

mixture experiments. We observed that the performance of the ridge-type robust estimator depends on the robust estimator of $\boldsymbol{\beta}$ and the biasing parameter k. In this study, as robust estimator of $\boldsymbol{\beta}$, the M-estimator, GM-estimator, and MM-estimator were discussed. By selecting the robust estimators which are less sensitive to points outlying in the x-and/or y-direction will improve the performance of the ridge-type robust-estimator. On the other hand, depending on the biasing parameter, to evaluate the performance of the ridge-type robust estimators, we adapted extensive enlargement of the graphical approach suggested for ORR estimator. For evaluating the biasing parameter, the FDS plots which are based on the SMSEP values are considered. As known in literature, the MSEP is a more powerful criterion for comparison of the designs than the prediction variance. Because MSEP incorporates the prediction variance and the prediction bias associated with the fitted model, the results obtained by FDS plots which are based on the SMSEP values were observed to be consistent with the other analytic

Table 5: The parameter estimates of the ridge-type robust estimators.

$\widehat{\boldsymbol{\beta}}_{\mathrm{RR}}$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\mathrm{MSE}[\widehat{y}_k(\mathbf{x}_u)]_{(\widehat{\boldsymbol{\beta}}_{\mathrm{RR}})}$
$\widehat{\boldsymbol{\beta}}_{\mathrm{RR}}(\widehat{\boldsymbol{\beta}}_{\mathrm{GM}})_{(k=0.00001)}$	−431.634 (27.121)	62.523 (51.491)	−230.555 (25.808)	304.331 (32.155)	297.865 (13.274)	14.124
$\widehat{\boldsymbol{\beta}}_{\mathrm{RR}}(\widehat{\boldsymbol{\beta}}_{\mathrm{MM}})_{(k=0.0001)}$	−419.691 (35.011)	72.054 (66.697)	−243.282 (33.578)	55.259 (91.264)	302.402 (17.385)	26.382

approaches used in literature. Therefore, for evaluation of biasing parameter, the FDS plots which are based on the SMSEP values can also be used for both the mixture models affected by multicollinearity and outliers and other biased-robust estimators.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to acknowledge the editor and referees for their constructive and valuable suggestions and comments that have improved the presentation of the paper greatly. This work was supported by Scientific Research Projects Coordination Unit of Istanbul University, Project no. UDP-24923.

References

[1] R. C. St. John, "Experiments with mixtures, ill-conditioning, and ridge regression," *Journal of Quality Technology*, vol. 16, no. 2, pp. 81–96, 1984.

[2] J. A. Cornell and J. W. Gorman, "Two new mixture models: living with collinearity but removing its influence," *Journal of Quality Technology*, vol. 35, no. 1, pp. 78–88, 2003.

[3] W. F. Smith, *Experimental Design for Formulation*, vol. 15 of *ASA-SIAM Series on Statistics and Applied Probability*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, USA, 2005.

[4] A. E. Hoerl and R. V. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[5] G. G. Vining, J. A. Cornell, and R. H. Myers, "A graphical approach for evaluating mixture designs," *Applied Statistics*, vol. 42, no. 1, pp. 127–138, 1993.

[6] A. I. Khuri, J. M. Harrison, and J. A. Cornell, "Using quantile plots of the prediction variance for comparing designs for a constrained mixture region: an application involving a fertilizer experiment," *Journal of the Royal Statistical Society C: Applied Statistics*, vol. 48, no. 4, pp. 521–532, 1999.

[7] D. Jang and M. Yoon, "Graphical methods for evaluating ridge regression estimator in mixture experiments," *Communications in Statistics B: Simulation and Computation*, vol. 26, no. 3, pp. 1049–1061, 1997.

[8] D. Jang and C. M. Anderson-Cook, "Fraction of design space plots for evaluating ridge estimators in mixture experiments," *Quality and Reliability Engineering International*, vol. 27, no. 1, pp. 27–34, 2011.

[9] A. Zahran, C. M. Anderson-Cook, and R. H. Myers, "Fraction of design space to assess prediction capability of response surface designs," *Journal of Quality Technology*, vol. 35, no. 4, pp. 377–386, 2003.

[10] D. C. Montgomery and S. R. Voth, "Multicollinearity and leverage in mixture experiments," *Journal of Quality Technology*, vol. 26, no. 2, pp. 96–108, 1994.

[11] O. Coskuntuncel, *Karma Denemelerde ve Modellerde Robust İstatistiksel Analizler [Ph.D. thesis]*, Cukurova Universitesi FBE, Adana, Turkey, 2005.

[12] D. C. Montgomery, E. A. Peck, and G. Vining, *Introduction to Linear Regression Analysis*, John Wiley & Sons, New York, NY, USA, 3rd edition, 2001.

[13] P. J. Huber, *Robust statistics*, John Wiley & Sons, New York, NY. USA, 1981.

[14] P. J. Rousseeuw and V. J. Yohai, "Robust regression by means of S-estimators," in *Robust and Nonlinear Time Series Analysis*, J. Franke, W. Härdle, and R. D. Martin, Eds., Springer, New York, NY, USA, 1984.

[15] V. J. Yohai, "High breakdown-point and high efficiency robust estimates for regression," *The Annals of Statistics*, vol. 15, no. 2, pp. 642–656, 1987.

[16] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*, Wiley Series in Probability and Statistics, John Wiley & Sons, London, UK, 2006.

[17] P. W. Holland, "Weighted ridge regression: combining ridge and robust regression methods," *NBER Working Paper Series*, vol. 11, pp. 1–19, 1973.

[18] R. G. Askin and D. C. Montgomery, "Augmented robust estimators," *Technometrics*, vol. 22, no. 3, pp. 333–341, 1980.

[19] M. J. Silvapulle, "Robust ridge regression based on an *M*-estimator," *The Australian Journal of Statistics*, vol. 33, no. 3, pp. 319–333, 1991.

[20] O. Arslan and N. Billor, "Robust ridge regression estimation based on the GM-estimators," *Journal of Mathematical and Computational Science*, vol. 9, no. 1, pp. 1–9, 1996.

[21] J. R. Simpson and D. C. Montgomery, "A biased-robust regression technique for the combined outlier-multicollinearity problem," *Journal of Statistical Computation and Simulation*, vol. 56, no. 1, pp. 1–22, 1996.

[22] R. A. Maronna, "Robust ridge regression for high-dimensional data," *Technometrics*, vol. 53, no. 1, pp. 44–53, 2011.

[23] G. Piepel, "Discussion of "Response surface design evaluation and comparison" by C.M. Anderson-Cook, C.M. Borror , and D.C. Montgomery," *Journal of Statistical Planning and Inference*, vol. 139, pp. 653–656, 2009.

[24] J. A. Cornell, *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*, John Wiley & Sons, New York, NY, USA, 3rd edition, 2002.

[25] H. Scheffé, "Experiments with mixtures," *Journal of the Royal Statistical Society B*, vol. 20, pp. 344–360, 1958.

[26] P. J. Huber, "Robust estimation of a location parameter," *Annals of Mathematical Statistics*, vol. 35, pp. 73–101, 1964.

[27] E. Handschin, F. C. Schweppe, J. Kohlas, and A. R. Flechter, "Bad data analysis for power system state estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. 94, no. 2, pp. 329–337, 1975.

[28] W. S. Krasker and R. E. Welsch, "Efficient bounded-influence regression estimation," *Journal of the American Statistical Association*, vol. 77, no. 379, pp. 595–604, 1982.

[29] J. R. Simpson and D. C. Montgomery, "The development and evaluation of alternative generalized *M*-estimation techniques," *Communications in Statistics—Simulation and Computation*, vol. 27, no. 4, pp. 999–1018, 1998.

[30] J. R. Simpson and D. C. Montgomery, "A performance-based assessment of robust regression methods," *Communications in Statistics*, vol. 27, no. 4, pp. 1031–1049, 1998.

[31] R. C. Pfaffenberger and T. E. Dielman, "A comparison of regression estimators when both multicollinearity and outliers are present," in *Robust Regression: Analysis and Applications*, K. Lawrence and J. Arthur, Eds., pp. 243–270, Marcel Dekker, New York, NY, USA, 1990.

[32] A. E. Hoerl, R. W. Kennard, and K. F. Baldwin, "Ridge regression: some simulations," *Communications in Statistics—Theory and Methods*, vol. 4, no. 2, pp. 105–123, 1975.

[33] J. F. Lawless and P. Wang, "A simulation study of ridge and other regression estimators," *Communications in Statistics: Theory and Methods*, vol. 5, no. 4, pp. 307–323, 1976.

[34] A. E. Clark and C. G. Troskie, "Ridge regression—a simulation study," *Communications in Statistics. Simulation and Computation*, vol. 35, no. 3, pp. 605–619, 2006.

[35] H. B. Goldfarb, C. M. Anderson-Cook, C. M. Borror, and D. C. Montgomery, "Fraction of design space plots for assessing mixture and mixture-process designs," *Journal of Quality Technology*, vol. 36, no. 2, pp. 169–179, 2004.

[36] G. Piepel, C. M. Anderson, and P. E. Redgate, "Variance dispersion graphs for evaluating and comparing experimental designs on polyhedral regions," Tech. Rep. PNWD-SA-4218, Batelle, Pacific Northwest Laboratories, Richland, Wash, USA, 1995.

[37] G. Piepel and T. Redgate, "A mixture experiment analysis of the hald cement data," *American Statistician*, vol. 52, no. 1, pp. 23–30, 1998.

[38] A. Hald, *Statistical Theory with Engineering Applications*, John Wiley & Sons, New York, NY, USA, 1952.