International Journal of Innovative Computing, Information and Control
Volume **12**, Number **3**, June **2016** ISSN 1349-4198
pp. 689–706

ANALYSIS ON THE CONSISTENT RELATIONSHIPS BETWEEN RIDGE ESTIMATOR AND THE OTHER TYPICAL BIASED ESTIMATORS

Xionglin Luo, Yuanlong Yue and Xin Zuo

Department of Automation
China University of Petroleum
No. 18, Fuxue Road, Changping District, Beijing 102249, P. R. China
{ luoxl; yueyuanlong; zuox }@cup.edu.cn

Received October 2015; revised February 2016

Abstract. *Ridge estimator widely applied in chemometrics has been proposed to improve the least squares estimator variance by trading bias for variance. Our contribution is a comparative analysis on the relationships between ridge estimator and the other typical biased estimators regarding the effect of bias on the decrease in variance and the extent of its effect. We show that, although the biasing parameters take on various expressions, when the variances of the different biased estimators applied to enhance LS precision are the same then the biases must also be the same. For any two alternatives, the change rates of the variances with respect to the biases always take on identical regulation as long as the values of the variances or the biases are the same. This conclusion implies that the curve of the changing rate of the variance with respect to the bias for any given estimator always has the same shape. It has nothing to do with the estimator selected in the analytical chemistry field.*
Keywords: Chemometrics, Biased estimator, Ridge estimator, Consistency analysis, Relationship between variance and bias

1. Introduction.

1.1. Literature review.

Parameter estimation based on the linear model $y = Hx + w$ is commonly used in academic research and engineering communities in process control, chemometrics, system identification and so on [1,2]. The classic least squares estimator (LS) was proposed by Gauss and Legendre in the 1800s [3]. In the 1900s, Gauss-Markov theorem [4] and Fisher information matrix [5] confirmed that LS is the optimal estimator in the linear unbiased estimation class. However, we note that the Gauss-Markov theorem and the Fisher information matrix merely guarantee that the LS possesses the optimization property only in the linear unbiased framework and not for the entire estimation process [6]. Actually, if the variance of LS is larger or beyond the acceptable maximum, then LS would not work.

Statisticians proposed numerous novel estimators to improve its variance [7]. Among these alternatives, the linear biased estimation class, proposed in the 1950s by Stein, is the most effective and practicable. The essence of the linear biased estimator is that it introduces relatively small bias to pursue biased alternatives superior to the LS in terms of the mean squared error (MSE) criterion. The reason for adopting the MSE criterion is that the MSE is always equivalent to variance plus the squared bias given any estimation method. Because the bias of the LS is zero, its MSE is the same as the variance. This phenomenon also confirms that the variance applied to evaluate the validity of an unbiased estimation class is reasonable. Under the condition of the MSE criterion, if we obtain one biased estimator superior to the LS, then the biased estimator's variance must be less than that of the LS.

For more than half a century, great progress has been achieved in terms of linear biased estimation theory. In 1956, Stein proposed an admissible estimator shortened for a spherically symmetrical one [8]. The essence of this alternative is a biased estimator heralding the beginning of the biased estimation theory. Soon after, in 1961, the James-Stein estimator was developed by James and Stein [9]. Sclove further developed the James-Stein estimator, extensively researching the shrunken LS estimator [10]. Horel and Kennard discussed the ridge analysis and its application in the regression field [11] and obtained a biased estimation method named the ridge estimator [12]. At present, the ridge estimator is accepted as an approach for overcoming the LS drawback and is commonly applied in process measurement [13], chemometrics [14] and economics modelling [15]. Based on the ridge estimator, a large number of improved estimated methods have been successively proposed. In 1976, Swindel proposed the modified ridge estimator by synthesizing priori information of the unknown parameter into the ridge estimator [16]. Liu, resorting to the merits of the Stein shrunken estimator and the ridge estimator, proposed the Liu estimator and the Liu-type estimator [17,18], respectively. Among the above alternatives, the ridge estimator and the Liu estimator are used most widely. While the illustrated estimation methods are mainly suitable for specific estimation problems, statisticians have worked to develop hybrid estimators having the merits of the proposed estimators.

Skallıoglu and Kaciranlar proposed the k-d estimator [19], substituting the ridge estimator for the priori information of the Liu estimator. Duran and Akdeniz in 2012 proposed a modifed jackknifng Liu-type estimator [20], which is a combinational estimator consisting of a general Liu estimator and a modified-general Liu estimator [21]. Li and Yang pointed out that the modified ridge estimator is a convex combination of a priori information and the LS, and they proposed the modified Liu estimator [22] in 2012. Ozkale and Kaciranlar in 2007 researched the two-parameter estimator [23], which is a combinational estimator based on the ridge and Liu ones. In 2008, Batah et al. proposed a general jackknifing ridge estimator and a modified jackknifing general ridge estimator [24]. At present, novel methods and theories for linear biased estimation are reported continuously by research institutes year after year [25,26]. Nevertheless, compared with the LS, the theoretical basis for the biased estimation has not yet been refined. Researchers must continue their efforts to make progress in linear biased estimation in terms of application and theory.

1.2. **Our contributions.** The above biased alternatives are all superior to the LS with respect to the MSE by the introduction of different biasing parameters. However, whether or not all the above improved estimation methods have the same regulation is a question worthy of consideration. Also it is worth noting that almost all the published superior biased estimators focus on the MSE condition under which the improved biased estimators are superior to the LS. Now since the MSE is equivalent to the variance plus the squared bias, it cannot reveal the influence arising from the change process of the bias on the degree of improvement in the variance. Since the purpose of introducing the bias into the estimation process is to decrease the variance, what we more concern about is the variance change rate with respect to the bias. With regard to the above problems, this paper carries out a comparative analysis on the relationships between ridge estimator and the other typical biased estimators regarding the effect of bias on the decrease in variance and the extent of its effect. The idea of this paper is stimulated by the process of searching a better biased estimator than ridge estimator. The article is mainly limited to several classical improved biased estimators, i.e., the ridge estimator, Liu estimator, combinational estimators with the priori information and jackknifing estimators. The most significant contribution of this manuscript is the exploration of the common regulation among these outstanding biased alternatives. Specifically, for any existing biased estimators dominating the LS, the value scopes of the variances, biases, and the estimated unknown parameters are always the same and are independent of the selected estimating methods as long as the values of the model parameters are the same.

The paper is organized as follows. In Section 2, we provide an overview of our problem. The proof that the variance change rate of the general ridge estimator is a lower convex function with respect to its bias is presented in Section 3. Sections 4 to 7 analyze the relationships between general ridge estimator and the other typical biased estimators, and conclude that the homologous characteristic is the common law restricting all the proposed biased estimators. We demonstrate through an example, in Section 8, that the change rate of variance with respect to the bias is always the same for every proposed alternative and has no relation to the estimator selected. The paper is concluded in Section 9.

2. **Preliminary.** Assuming that the relationship between the estimated parameter vector and the data is described by the following linear model

$$y = Hx + w \tag{1}$$

where $y \in R^n$ is the observed data, $x \in R^m$ is the unknown parameter vector, and H is an $n \times m$ matrix of observation on the regression of rank m, we always assume that $n \geq m$, w is an $n \times 1$ vector of independent and identically distributed random errors with mean zero and variance matrix $\mathrm{Var}(w) = \sigma^2 \mathrm{I}_n$, and I_n is an identity matrix of order $n \times n$. According to the Gauss-Markov theorem, the least squares estimator of x generated from model (1) is shown as the following:

$$\hat{x}_{\mathrm{LS}} = \left(H^{\mathrm{T}} H\right)^{-1} H^{\mathrm{T}} y \tag{2}$$

The variance matrix of $\hat{x}_{\mathrm{LS}}$ is given as

$$\mathrm{Var}(\hat{x}_{\mathrm{LS}}) = \sigma^2 \left(H^{\mathrm{T}} H\right)^{-1} \tag{3}$$

Let $Q = H^{\mathrm{T}} H$, so Q is a symmetric matrix having an eigenvalue-eigenvector decomposition in the form $Q = G \Lambda G^{\mathrm{T}}$, where G is an orthogonal matrix and Λ is a diagonal matrix. In this article, we assume that $Z = HG$, $\alpha = G^{\mathrm{T}} x$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ are the ordered eigenvalues of Q, $G = (\rho_1, \rho_2, \cdots, \rho_n)$ is the eigenvector matrix of Q whose columns constitute the eigenvectors corresponding to the eigenvalues, so we get the following representation

$$Z^{\mathrm{T}} Z = G^{\mathrm{T}} H^{\mathrm{T}} H G = G^{\mathrm{T}} G \Lambda G^{\mathrm{T}} G = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \tag{4}$$

For the sake of convenience, we rewrite model (1) in canonical form:

$$y = Z\alpha + w \tag{5}$$

The least squares estimator of α and its variance matrix are given by the following, respectively:

$$\hat{\alpha}_{\mathrm{LS}} = \left(Z^{\mathrm{T}} Z\right)^{-1} Z^{\mathrm{T}} y = \Lambda^{-1} Z^{\mathrm{T}} y \tag{6}$$

$$\mathrm{Var}(\hat{\alpha}_{\mathrm{LS}}) = \sigma^2 \Lambda^{-1} \tag{7}$$

In the following sections, a detailed comparison of biased estimators is made, based on the canonical model. Because statisticians not only encounter the problem of choosing between LS and a biased estimator, but must also choose between two biased estimators. We now make a bold assumption, if the proposed linear biased estimators are characterized by mutual transformation, then we can immediately conclude that these alternatives to LS are homologous, and thus only one biased estimator is required, among those proposed, to completely describe the superiority of all the above improved estimation methods.

3. **General Ridge Estimator.** Referring to ridge analysis and its application to regression problems in chemical processes, Hoerl and Kennard in 1970 proposed a general ridge estimator (GRE) written as

$$\hat{x}_{\mathrm{GRE}} = \left(H^{\mathrm{T}}H + G^{\mathrm{T}}K_{\mathrm{GRE}}G\right)^{-1} H^{\mathrm{T}}y \tag{8}$$

where $K_{\mathrm{GRE}} = \mathrm{diag}(k^i_{\mathrm{GRE}})$, $i = 1, 2, \cdots, n$, $k^i_{\mathrm{GRE}} > 0$ is a biasing parameter matrix. Corresponding to $\hat{\alpha}_{\mathrm{LS}}$, this GRE can be rewritten in canonical form

$$\hat{\alpha}_{\mathrm{GRE}} = (\Lambda + K_{\mathrm{GRE}})^{-1}Z^{\mathrm{T}}y = (\Lambda + K_{\mathrm{GRE}})^{-1}\Lambda\hat{\alpha}_{\mathrm{LS}} \tag{9}$$

The bias vector and the variance matrix of the GRE are computed as follows, respectively

$$\mathrm{Bias}(\hat{\alpha}_{\mathrm{GRE}}) = \mathrm{E}(\hat{\alpha}_{\mathrm{GRE}}) - \alpha = (\Lambda + K_{\mathrm{GRE}})^{-1}\Lambda\alpha - \alpha = -(\Lambda + K_{\mathrm{GRE}})^{-1}K_{\mathrm{GRE}}\alpha \tag{10}$$

$$\mathrm{Var}(\hat{\alpha}_{\mathrm{GRE}}) = [(\Lambda + K_{\mathrm{GRE}})^{-1}\Lambda]\mathrm{Var}(\hat{\alpha}_{\mathrm{LS}})[(\Lambda + K_{\mathrm{GRE}})^{-1}\Lambda]^{\mathrm{T}} = \sigma^2(\Lambda + K_{\mathrm{GRE}})^{-1}\Lambda(\Lambda + K_{\mathrm{GRE}})^{-\mathrm{T}} \tag{11}$$

where $\mathrm{E}(\hat{\alpha}_{\mathrm{GRE}})$ is the expectation of $\hat{\alpha}_{\mathrm{GRE}}$. Let α_i, $i = 1, 2, \cdots, n$, be the ith element of vector α. According to Equation (11), the general ridge estimator of α_i is given by

$$\hat{\alpha}^i_{\mathrm{GRE}} = \frac{k^i_{\mathrm{GRE}}}{\lambda_i + k^i_{\mathrm{GRE}}}\hat{\alpha}^i_{\mathrm{LS}} \tag{12}$$

Combining Equations (10) and (11), we can easily compute the MSE that measures the squared bias and variance of the biased estimator $\hat{\alpha}^i_{\mathrm{GRE}}$ simultaneously.

$$\mathrm{MSE}(\hat{\alpha}^i_{\mathrm{GRE}}) = \mathrm{Var}(\hat{\alpha}^i_{\mathrm{GRE}}) + \mathrm{Bias}(\hat{\alpha}^i_{\mathrm{GRE}})^2 = \frac{\lambda^i}{\lambda_i + k^i_{\mathrm{GRE}}}\mathrm{Var}(\hat{\alpha}^i_{\mathrm{LS}}) + \left(\frac{k^i_{\mathrm{GRE}}}{\lambda_i + k^i_{\mathrm{GRE}}}\alpha_i\right)^2 \tag{13}$$

If α^i_{GRE} is superior to α^i_{LS}, then the following inequality must be established

$$\left(\frac{\lambda^i}{\lambda_i + k^i_{\mathrm{GRE}}}\right)^2 \mathrm{Var}\left(\hat{\alpha}^i_{\mathrm{LS}}\right) + \left(\frac{k^i_{\mathrm{GRE}}}{\lambda_i + k^i_{\mathrm{GRE}}}\alpha_i\right)^2 \le \mathrm{Var}\left(\hat{\alpha}^i_{\mathrm{LS}}\right) \tag{14}$$

Considering Equation (7), the solution of Equation (14) is given by

$$0 < k^i_{\mathrm{GRE}} < \frac{2\lambda_i\sigma^2}{\lambda_i\alpha_i^2 - \sigma^2} \tag{15}$$

Theorem 3.1. *Let $\alpha_i \neq 0$, $i = 1, 2, \cdots, n$. When $k^i_{\mathrm{GRE}} \to 0^+$, the change rate of the variance of α^i_{GRE} with respect to its bias approaches $-\infty$; $0 < k^i_{\mathrm{GRE}} < \frac{2\lambda_i\sigma^2}{\lambda_i\alpha_i^2 - \sigma^2}$, and the variance of α^i_{GRE} is a lower convex function with respect to its bias.*

Proof: Combining the variance and the squared bias, formulate a parameter equation with regard to k^i_{GRE}

$$\begin{cases} F_1 = \dfrac{\lambda^i}{(\lambda_i + k^i_{\mathrm{GRE}})^2}\sigma^2 \\ F_2 = \left(\dfrac{k^i_{\mathrm{GRE}}}{\lambda_i + k^i_{\mathrm{GRE}}}\alpha_i\right)^2 \end{cases} \tag{16}$$

The F_1 first-order derivative of the F_2 is given by

$$\frac{\mathrm{d}F_1}{\mathrm{d}F_2} = -\frac{\sigma^2}{k^i_{\mathrm{GRE}}\alpha_i^2} \tag{17}$$

For $k^i_{\mathrm{GRE}} > 0$ and $\alpha_i \neq 0$, we know that $\frac{\mathrm{d}F_1}{\mathrm{d}F_2}$ is always negative. In particular, when $k^i_{\mathrm{GRE}} \to 0^+$, $\frac{\mathrm{d}F_1}{\mathrm{d}F_2}$ approaches $-\infty$, i.e., the change rate of the variance to the squared bias is at maximum. The F_1 second-order derivative of F_2 is given by

$$\frac{\mathrm{d}^2 F_1}{\mathrm{d}F_2^2} = \frac{F_1'' F_2' - F_1' F_2''}{(F_2')^3} = \frac{(\lambda_i + k^i_{\mathrm{GRE}})^3 \sigma^2}{2\lambda_i (k^i_{\mathrm{GRE}})^3 \alpha_i^4} \tag{18}$$

for $k^i_{\mathrm{GRE}} > 0$ and $\alpha_i \neq 0$, so $\frac{\mathrm{d}^2 F_1}{\mathrm{d}F_2^2}$ is always positive. According to the properties of the convex function, the proof of Theorem 3.1 is completed.

This theorem is derived simply; however, it further illustrates that the values of variance and the change rate of the variance to the squared bias decreases within the scope of $0 < k^i_{\mathrm{GRE}} < \frac{2\lambda_i\sigma^2}{\lambda_i\alpha_i^2 - \sigma^2}$. The assumption $\alpha_i \neq 0$ in Theorem 3.1 is explained as follows: because an unknown parameter always exists in the estimation process, the assumption $\alpha_i \neq 0$ is always satisfied. A particularly noteworthy phenomenon is that $\mathrm{Bias}(\alpha^i_{\mathrm{GRE}})$ is almost zero given $k^i_{\mathrm{GRE}} \to 0^+$. The comprehensive consideration of this phenomenon and Equation (16) means that both the bias and its change rate referring to k^i_{GRE} approaches 0, and there is the maximum change rate of α^i_{GRE} variance. This conclusion indicates that a biased estimator remains unbiased when $k^i_{\mathrm{GRE}} \to 0^+$, but the variance is markedly decreased. Meanwhile, Theorem 3.1 illustrates that the improving effect of bias on variance weakens gradually.

Based on Theorem 3.1, we can obtain accurate information about the change rate of variance to bias relationship. For any two different biased estimators superior to LS, if we assume that the variances have the same values and it must also be true that the biases have the same values, then Theorem 3.1 always exists independent of the biased estimator chosen.

Now we can calculate any point within the scope of $0 < k^i_{\mathrm{GRE}} < \frac{2\lambda_i\sigma^2}{\lambda_i\alpha_i^2 - \sigma^2}$ denoted by $k^i_{\mathrm{GRE}}(w)$

$$k^i_{\mathrm{GRE}}(w) = \frac{2\lambda_i\sigma^2}{\lambda_i\alpha_i^2 - \sigma^2 w} \tag{19}$$

where $w \in [1{+}\infty)$ is a tuning parameter. The variance and the bias of $\hat{\alpha}^i_{\mathrm{GRE}}$ corresponding to $k^i_{\mathrm{GRE}}(w)$, respectively, are given by

$$\mathrm{Var}(\hat{\alpha}^i_{\mathrm{GRE}})|_{k^i_{\mathrm{GRE}} = k^i_{\mathrm{GRE}}(w)} = \left[\frac{\lambda^i}{\lambda_i + k^i_{\mathrm{GRE}}(w)}\right]^2 \mathrm{Var}(\hat{\alpha}^i_{\mathrm{LS}}) = \left[\frac{(\lambda_i\alpha_i^2 - \sigma^2)w}{(\lambda_i\alpha_i^2 - \sigma^2)w + 2\sigma^2}\right]^2 \frac{\sigma^2}{\lambda_i} \tag{20}$$

$$\mathrm{Bias}(\hat{\alpha}^i_{\mathrm{GRE}})|_{k^i_{\mathrm{GRE}} = k^i_{\mathrm{GRE}}(w)} = -\frac{k^i_{\mathrm{GRE}}(w)}{\lambda_i + k^i_{\mathrm{GRE}}(w)}\alpha_i = -\frac{2\sigma^2}{(\lambda_i\alpha_i^2 - \sigma^2)w + 2\sigma^2}\alpha_i \tag{21}$$

Since $k^i_{\mathrm{GRE}}(w)$ is any point within $0 < k^i_{\mathrm{GRE}} < \frac{2\lambda_i\sigma^2}{\lambda_i\alpha_i^2 - \sigma^2}$, Equations (20) and (21) also comply with the conclusion made in Theorem 3.1. With respect to the other biased estimators in Section 1.1, if we can invariably find the variance or the bias whose value is equivalent to their counterparts in Equations (20) or (21), then it is reasonable to conclude that these conventional alternatives to LS are also fit for the conclusion in Theorem 3.1.

4. **General Liu Estimator.** Combining the Stein estimator with LS, Liu in 1993 proposed the general Liu estimator (GLE) defined by

$$\hat{x}_{\mathrm{GLE}} = \left(H^{\mathrm{T}} H + \mathrm{I}\right)^{-1} \left(H^{\mathrm{T}} y + D\hat{x}_{\mathrm{LS}}\right) \tag{22}$$

where $D_{\text{GLE}} = \text{diag}(d^i_{\text{GLE}})$, $i = 1, 2, \cdots, n$, $0 < d^i_{\text{GLE}} < 1$ is a GLE biasing parameter matrix. Referring to Equation (9), we get the following representation

$$\hat{\alpha}_{\text{GLE}} = (\Lambda + \text{I})^{-1}(\Lambda + D)\hat{\alpha}_{\text{LS}} \tag{23}$$

For the α_i, $i = 1, 2, \cdots, n$, the variance and the bias of $\hat{\alpha}^i_{\text{GLE}}$ are computed respectively as follows:

$$\text{Var}\left(\hat{\alpha}^i_{\text{GLE}}\right) = \left(\frac{\lambda_i + d^i_{\text{GLE}}}{\lambda_i + 1}\right)^2 \text{Var}\left(\hat{\alpha}^i_{\text{LS}}\right) \tag{24}$$

$$\text{Bias}(\hat{\alpha}^i_{\text{GLE}}) = \frac{\lambda_i + d^i_{\text{GLE}}}{\lambda_i + 1}\alpha_i - \alpha_i = \frac{d^i_{\text{GLE}} - 1}{\lambda_i + 1}\alpha_i \tag{25}$$

Similar to the GRE, that $\hat{\alpha}^i_{\text{GLE}}$ is superior to $\hat{\alpha}^i_{\text{LS}}$ in the sense of MSE implies that the following inequality must be true:

$$\left(\frac{\lambda_i + d^i_{\text{GLE}}}{\lambda_i + 1}\right)^2 \text{Var}\left(\hat{\alpha}^i_{\text{LS}}\right) + \left(\frac{d^i_{\text{GLE}} - 1}{\lambda_i + 1}\alpha_i\right)^2 < \text{Var}(\hat{\alpha}^i_{\text{LS}}) \tag{26}$$

Considering Equation (7), we can get the scope of d^i_{GLE} by solving the inequality (26)

$$\frac{\lambda_i\alpha_i^2 - (2\lambda_i + 1)\sigma^2}{\lambda_i\alpha_i^2 + \sigma^2} < d^i_{\text{GLE}} < 1 \tag{27}$$

In the following, we show that the value range of the biasing parameter d^i_{GLE} can be determined by k^i_{GRE}, $i = 1, 2, \cdots, n$. According to Equation (23), the GLE of α_i is written as

$$\hat{\alpha}^i_{\text{GLE}} = \frac{\lambda_i + d^i_{\text{GLE}}}{\lambda_i + 1}\hat{\alpha}^i_{\text{LS}} \tag{28}$$

Based on Equations (12) and (28), we assume that

$$\frac{\lambda_i + d^i_{\text{GLE}}}{\lambda_i + 1} = \frac{\lambda_i}{\lambda_i + k^i_{\text{GRE}}} \tag{29}$$

Rewriting Equation (29), then d^i_{GLE} can be expressed as the function of k^i_{GRE}

$$d^i_{\text{GLE}} = \frac{\lambda_i(\lambda_i + 1)}{\lambda_i + k^i_{\text{GRE}}} - \lambda_i = \frac{\lambda_i(1 - k^i_{\text{GRE}})}{\lambda_i + k^i_{\text{GRE}}} \tag{30}$$

Analyzing Equation (30), we know that the values of the function of d^i_{GLE} with respect to k^i_{GRE} are monotonically decreasing. According to Equation (15), the value scope of d^i_{GLE} is obtained as

$$\frac{\lambda_i\alpha_i^2 - (2\lambda_i + 1)\sigma^2}{\lambda_i\alpha_i^2 + \sigma^2} < d^i_{\text{GLE}} < 1 \tag{31}$$

Comparing Equation (31) with Equation (27), it is obvious that the biasing parameter d^i_{GLE} can be calculated by solving Equation (29). Indeed, this conversion is inevitable. The reason for the phenomenon is that $\hat{\alpha}^i_{\text{GLE}}$ and $\hat{\alpha}^i_{\text{GRE}}$ have the same value scopes constrained by Equations (15) and (27), respectively. Substituting the left end point of $\frac{\lambda_i\alpha_i^2 - (2\lambda_i+1)\sigma^2}{\lambda_i\alpha_i^2 + \sigma^2} < d^i_{\text{GLE}} < 1$ into Equation (28), we obtain $\hat{\alpha}^i_{\text{GLE}} = \frac{\lambda_i\alpha_i^2 - \sigma^2}{\lambda_i\alpha_i^2 + \sigma^2}\hat{\alpha}^i_{\text{LS}}$. Similarly, the $\hat{\alpha}^i_{\text{GRE}}$ determined by the right end point of $0 < k^i_{\text{GRE}} < \frac{2\lambda_i\sigma^2}{\lambda_i\alpha_i^2 - \sigma^2}$ is $\hat{\alpha}^i_{\text{GRE}} = \frac{\lambda_i\alpha_i^2 - \sigma^2}{\lambda_i\alpha_i^2 + \sigma^2}\hat{\alpha}^i_{\text{LS}}$. Substituting $d^i_{\text{GLE}} = 1$ and $k^i_{\text{GRE}} = 0$ into Equations (12) and (28), respectively, then $\hat{\alpha}^i_{\text{GLE}}$ and $\hat{\alpha}^i_{\text{GRE}}$ are equivalent to $\hat{\alpha}^i_{\text{LS}}$.

Let us further calculate the variance and the bias of $\hat{\alpha}^i_{\text{GLE}}$ under the condition of the biasing parameter d^i_{GLE} determined by Equation (30). Because the above analysis points out that the relationship between d^i_{GLE} and k^i_{GRE} is one of mutual transformations, we can calculate any point value of d^i_{GLE} denoted by $d^i_{\mathrm{GLE}}(w)$ within $\frac{\lambda_i\alpha_i^2-(2\lambda_i+1)\sigma^2}{\lambda_i\alpha_i^2+\sigma^2} < d^i_{\mathrm{GLE}} < 1$ by substituting the parameter $k^i_{\mathrm{GRE}}(w)$ into Equation (30), i.e.,

$$d^i_{\mathrm{GLE}}(w) = \frac{\lambda_i[1-k^i_{\mathrm{GRE}}(w)]}{\lambda_i + k^i_{\mathrm{GRE}}(w)} = \frac{(\lambda_i\alpha_i^2-\sigma^2)w - 2\lambda_i\sigma^2}{(\lambda_i\alpha_i^2-\sigma^2)w + 2\lambda\sigma^2} \tag{32}$$

Up to now, the variance and the bias corresponding to $d^i_{\mathrm{GLE}}(w)$ are respectively computed as

$$\mathrm{Var}(\hat{\alpha}^i_{\mathrm{GLE}})|_{d^i_{\mathrm{GLE}}=d^i_{\mathrm{GLE}}(w)} = \left[\frac{\lambda_i + d^i_{\mathrm{GLE}}(w)}{\lambda_i+1}\right]^2 \mathrm{Var}(\hat{\alpha}^i_{\mathrm{LS}}) = \left[\frac{(\lambda_i\alpha_i^2-\sigma^2)w}{(\lambda_i\alpha_i^2-\sigma^2)w+2\sigma^2}\right]^2 \frac{\sigma^2}{\lambda_i} \tag{33}$$

$$\mathrm{Bias}(\hat{\alpha}^i_{\mathrm{GRE}})|_{d^i_{\mathrm{GLE}}=d^i_{\mathrm{GLE}}(w)} = -\frac{1-d^i_{\mathrm{GLE}}(w)}{\lambda_i+1}\alpha_i = -\frac{2\sigma^2}{(\lambda_i\alpha_i^2-\sigma^2)w+2\sigma^2}\alpha_i \tag{34}$$

The comparison of Equations (33) and (34) with Equations (20) and (21), correspondingly, illustrates that the variance and the biases of GLE and GRE are the same, given Equation (29). This conclusion ensures that the relationship of the variance and the bias of GLE also satisfy Theorem 3.1. Although GRE and GLE are two very different biased estimators, the change rate of the variance of GLE with respect to its bias is completely consistent with that of GRE.

5. **Priori Information Estimators.**

In the following discussion, several typical priori information estimators are described and compared with GRE in terms of their scope of variance, bias and estimated values.

5.1. **Modified ridge estimator.**

Swindel in 1976 proposed the modified ridge estimator (MRE). Let $T_k = \left(H^{\mathrm{T}}H + k\mathrm{I}\right)^{-1} H^{\mathrm{T}}H = \mathrm{I} - k\left(H^{-\mathrm{T}}H + k\mathrm{I}\right)^{-1}$, then MRE is obtained as

$$\hat{x}_{MRE} = \left(H^{\mathrm{T}}H + k\mathrm{I}\right)^{-1}\left(H^{\mathrm{T}}y + kb_0\right) \tag{35}$$

In canonical form, the MRE is rewritten as

$$\hat{\alpha}_{MRE} = (\Lambda + k\mathrm{I})^{-1}\left(Z^{\mathrm{T}}y + kb\right) \tag{36}$$

where b_0 and b are the priori information about the unknown estimated parameter, and $b = Gb_0$. We substitute $K_{\mathrm{MGRE}} = \mathrm{diag}(k^i_{\mathrm{MGRE}})$, $i = 1, 2, \cdots, n$ for $k\mathrm{I}$ in Equations (35) and (36). Taking $b = T\hat{\alpha}_{\mathrm{LS}}$ as the priori information for the MGRE, where $T = \mathrm{diag}(t_i)$, $0 < t_i < 1$, $i = 1, 2, \cdots, n$ is a diagonal shrunken matrix, then Equation (36) is transformed as

$$\hat{\alpha}_{\mathrm{MGRE}} = (\Lambda + K_{\mathrm{MGRE}})^{-1}(\Lambda + K_{\mathrm{MGRE}}T)\hat{\alpha}_{\mathrm{LS}} \tag{37}$$

The variance and the bias of the MGRE for α_i, $i = 1, 2, \cdots, n$ are respectively obtained as

$$\mathrm{Var}(\hat{\alpha}^i_{\mathrm{MGRE}}) = \left(\frac{\lambda_i + k^i_{\mathrm{MGRE}}t_i}{\lambda_i + k^i_{\mathrm{MGRE}}}\right)^2 \mathrm{Var}(\hat{\alpha}^i_{\mathrm{LS}}) \tag{38}$$

$$\mathrm{Bias}(\hat{\alpha}^i_{\mathrm{MGRE}}) = \frac{k^i_{\mathrm{MGRE}}(t_i-1)}{\lambda_i + k^i_{\mathrm{MGRE}}}\alpha_i \tag{39}$$

Similar to the solving process of k^i_{GRE}, we can get the scope of k^i_{MGRE} as the following

$$0 < k^i_{\mathrm{MGRE}} < \frac{2\lambda_i\sigma^2}{(1-t_i)\lambda_i\alpha_i^2 - (1+t_i)\sigma^2} \tag{40}$$

The scope of $\hat{\alpha}^i_{\mathrm{MGRE}}$ determined by k^i_{MGRE} is

$$\frac{\lambda_i\alpha_i^2-\sigma^2}{\lambda_i\alpha_i^2+\sigma^2}\hat{\alpha}^i_{\mathrm{LS}} < \hat{\alpha}^i_{\mathrm{MGRE}} < \hat{\alpha}^i_{\mathrm{LS}} \tag{41}$$

According to Equation (37), the MGRE of α_i is written as

$$\hat{\alpha}^i_{\mathrm{MGRE}} = \frac{\lambda_i + k^i_{\mathrm{MGRE}} t_i}{\lambda_i + k^i_{\mathrm{MGRE}}}\hat{\alpha}^i_{\mathrm{LS}} \tag{42}$$

Referring to Equations (29) and (42), k^i_{MGRE} can be expressed as the function of k^i_{GRE}

$$k^i_{\mathrm{MGRE}} = \frac{\lambda_i k^i_{\mathrm{GRE}}}{\lambda_i - t_i(\lambda_i + k^i_{\mathrm{GRE}})} \tag{43}$$

and $k^i_{\mathrm{MGRE}}(w)$ can be obtained as

$$k^i_{\mathrm{MGRE}}(w) = \frac{2\lambda_i\sigma^2}{(\lambda_i\alpha_i^2-\sigma^2)w - t_i[(\lambda_i\alpha_i^2-\sigma^2)w+2\sigma^2]} \tag{44}$$

So the variance and the bias corresponding to $k^i_{\mathrm{MGRE}}(w)$ are respectively given by

$$\begin{aligned}\mathrm{Var}(\hat{\alpha}^i_{\mathrm{MGRE}})|_{k^i_{\mathrm{MGRE}}=k^i_{\mathrm{MGRE}}(w)} &= \left[\frac{\lambda_i + k^i_{\mathrm{MGRE}}(w)t_i}{\lambda_i + k^i_{\mathrm{MGRE}}(w)}\right]^2 \mathrm{Var}(\hat{\alpha}^i_{\mathrm{LS}}) \\ &= \left[\frac{(\lambda_i\alpha_i^2-\sigma^2)w}{(\lambda_i\alpha_i^2-\sigma^2)w+2\sigma^2}\right]^2\frac{\sigma^2}{\lambda_i}\end{aligned} \tag{45}$$

$$\mathrm{Bias}(\hat{\alpha}^i_{\mathrm{MGRE}})|_{k^i_{\mathrm{MGRE}}=k^i_{\mathrm{MGRE}}(w)} = \frac{k^i_{\mathrm{MGRE}}(w)(t_i-1)}{\lambda_i + k^i_{\mathrm{MGRE}}(w)}\alpha_i = -\frac{2\sigma^2}{(\lambda_i\alpha_i^2-\sigma^2)w+2\sigma^2}\alpha_i \tag{46}$$

Comparing Equations (45) and (46) with Equations (20) and (21), correspondingly, illustrates that the relationship of the variance and the bias of the MGRE also satisfy Theorem 3.1. It is interesting that T has no effect on the variance, the bias and the scope of $\hat{\alpha}^i_{\mathrm{MGRE}}$.

5.2. **Modified Liu estimator.** Li and Yang introduced a new estimator named the modified Liu estimator (MLE). Let $F_d = \left(H^{\mathrm{T}}H+\mathrm{I}\right)^{-1}\left(H^{\mathrm{T}}H+d\mathrm{I}\right) = \mathrm{I}-(1-d)\left(H^{\mathrm{T}}H+\mathrm{I}\right)^{-1}$, so the MLE is obtained simply by the convex combination of $\hat{x}_{\mathrm{LS}}$ and the priori information denoted by b_0:

$$\hat{x}_{\mathrm{MLE}} = \left(H^{\mathrm{T}}H+\mathrm{I}\right)^{-1}\left[\left(H^{\mathrm{T}}H+d\mathrm{I}\right)\right]\hat{x}_{\mathrm{LS}} + (\mathrm{I}-d\mathrm{I})b_0 \tag{47}$$

Similar to the MGRE, the modified general Liu estimator (MGLE) will be used for analysis in this manuscript. Let $D_{\mathrm{MGLE}} = \mathrm{diag}(d^i_{\mathrm{MGLE}})$, $i=1,2,\cdots,n$, $0<d^i_{\mathrm{MGLE}}<1$, the MGLE in canonical form, to be

$$\hat{\alpha}_{\mathrm{MGLE}} = (\Lambda+\mathrm{I})^{-1}[(\Lambda+D_{\mathrm{MGLE}})\hat{x}_{\mathrm{LS}} + (\mathrm{I}-D_{\mathrm{MGLE}})b] \tag{48}$$

where $b = Gb_0$ is the priori information. Let $b = T\hat{\alpha}_{\mathrm{LS}}$, $T=\mathrm{diag}(t_i)$; $0<t_i<1$, $i=1,2,\cdots,n$ be the shrunken matrix. According to Equation (48), the MGLE of α_i, $i=1,2,\cdots,n$ is given by

$$\hat{\alpha}^i_{\mathrm{MGLE}} = \frac{\lambda_i + t_i + (1-t_i)d^i_{\mathrm{MGLE}}}{\lambda_i+1}\hat{\alpha}^i_{\mathrm{LS}} \tag{49}$$

The comparison of Equations (49) and (28) implies that the structure of the MGLE is similar to that of the GLE, so the procedure for comparing the MGLE and GRE is omitted for the sake of convenience. d^i_{MGLE} can be expressed as a function with respect to k^i_{GRE} and t_i.

$$d^i_{\text{MGLE}} = \frac{\lambda_i(1-t_i) - k^i_{\text{GRE}}(\lambda_i + t_i)}{(1-t_i)(\lambda_i + k^i_{\text{GRE}})} \tag{50}$$

5.3. **Two-parameter estimator.** Ozkale and Kaciranlar in 2007 proposed the two-parameter estimator denoted by $x_{(k,d)}$.

$$\hat{x}_{(k,d)} = \left(H^{\mathrm{T}}H + k_{(k,d)}\mathrm{I}\right)^{-1}\left(H^{\mathrm{T}}H + k_{(k,d)}d_{(k,d)}\mathrm{I}\right)\hat{x}_{\text{LS}} \tag{51}$$

In canonical form, the MRE is rewritten as

$$\hat{\alpha}_{(k,d)} = \left(\Lambda + k_{(k,d)}\mathrm{I}\right)^{-1}\left(\Lambda + k_{(k,d)}d_{(k,d)}\mathrm{I}\right)\hat{\alpha}_{\text{LS}} \tag{52}$$

where $k_{(k,d)} > 0$ and $0 < d_{(k,d)} < 1$. If we substitute $K_{(k,d)} = \text{diag}\left(k^i_{(k,d)}\right)$ and $D_{(k,d)} = \text{diag}\left(d^i_{(k,d)}\right)$, $i = 1, 2, \cdots, n$, for $k_{(k,d)}$ and $d_{(k,d)}$, where $k^i_{(k,d)} > 0$ and $0 < d^i_{(k,d)} < 1$, then Equation (52) is rewritten as

$$\hat{\alpha}_{(k,d)} = \left(\Lambda + K_{(k,d)}\right)^{-1}\left(\Lambda + K_{(k,d)}D_{(k,d)}\right)\hat{\alpha}_{\text{LS}} \tag{53}$$

Comparing Equation (53) with Equation (44), we know that $\hat{\alpha}_{(k,d)}$ is a special case of $\hat{\alpha}_{\text{MGRE}}$, where the former simply takes $D_{(k,d)}\hat{\alpha}_{\text{LS}}$ as the priori information instead of $T\hat{\alpha}_{\text{LS}}$, so $k^i_{(k,d)}$ can be expressed as a function of k^i_{GRE} directly.

$$k^i_{(k,d)} = \frac{\lambda_i k^i_{\text{GRE}}}{\lambda_i - d^i_{(k,d)}(\lambda_i + k^i_{\text{GRE}})} \tag{54}$$

6. **Jackknifing Estimators.** Jackknifing estimators are all the non-liner transformation on the LS, and the base distinctions among these alternatives are the various matrices.

6.1. **Jackknifing general Liu estimation.** Nyquist and Batah proposed the jackknifing general Liu estimator (JGLE) denoted by $\hat{\alpha}_{\text{JGLE}}$ in canonical form. Let $F_D = \text{diag}(f_i) = (\Lambda + \mathrm{I})^{-1}(\Lambda + D)$, $i = 1, 2, \cdots, n$, so $\hat{\alpha}_{\text{JGLE}}$ is given by

$$\hat{\alpha}_{\text{JGLE}} = (\mathrm{I} - F_D)F_D\hat{\alpha}_{\text{LS}} \tag{55}$$

where D is a positive definite diagonal matrix with nonzero elements d_i, $0 < d_i < 1$, $i = 1, 2, \cdots, n$. Componentwise estimators in Equation (55) are in the following form:

$$\hat{\alpha}^i_{\text{JGLE}} = \frac{f_i}{2 - f_i}\hat{\alpha}^i_{\text{LS}} \tag{56}$$

The variance and the bias of $\hat{\alpha}^i_{\text{JGLE}}$ are, respectively

$$\text{Var}(\hat{\alpha}^i_{\text{JGLE}}) = \left(\frac{f_i}{2 - f_i}\right)^2 \text{Var}(\hat{\alpha}^i_{\text{LS}}) \tag{57}$$

$$\text{Bias}(\hat{\alpha}^i_{\text{JGLE}}) = \frac{2f_i - 2}{2 - f_i}\alpha_i \tag{58}$$

Similar to the solving process of k^i_{GRE}, we can get the scope of f_i as the following

$$\frac{\alpha_i^2\lambda_i - \sigma^2}{\alpha_i^2\lambda_i} < f_i < 1 \tag{59}$$

Referring to Eqsuation (12) and (59), f_i can be expressed as the function of k^i_{GRE}

$$f_i = \frac{2\lambda_i}{2\lambda_i + k^i_{\text{GRE}}} \tag{60}$$

and $f_i(w)$ can be obtained as

$$f_i(w) = \frac{2\lambda_i}{2\lambda_i + k^i_{\mathrm{GRE}}(w)} = \frac{(\lambda_i\alpha_i^2 - \sigma^2)w}{(\lambda_i\alpha_i^2 - \sigma^2)w + \sigma_i^2} \tag{61}$$

The variance and the bias with respect to $f_i(w)$ are respectively calculated as

$$\mathrm{Var}(\hat{\alpha}^i_{\mathrm{JGLE}})|_{f_i=f_i(w)} = \left[\frac{f_i(w)}{2-f_i(w)}\right]^2 \mathrm{Var}(\hat{\alpha}^i_{\mathrm{LS}}) = \left[\frac{(\lambda_i\alpha_i^2 - \sigma^2)w}{(\lambda_i\alpha_i^2 - \sigma^2)w + 2\sigma^2}\right]^2 \frac{\sigma^2}{\lambda_i} \tag{62}$$

$$\mathrm{Bias}(\hat{\alpha}^i_{\mathrm{JGLE}})|_{f_i=f_i(w)} = \frac{2f_i(w)-2}{2-f_i(w)}\alpha_i = -\frac{2\sigma^2}{(\lambda_i\alpha_i^2 - \sigma^2)w + 2\sigma^2}\alpha_i \tag{63}$$

The JGLE variance in Equation (62) and the bias in Equation (63) are the same as that in Equations (20) and (21), respectively. This result guarantees that the change rate of the variance with respect to the bias of the JGLE also satisfies Theorem 3.1.

6.2. **Modified jackknifing general Liu estimator.** Esra proposed the modified jackknifing general Liu estimator (MJL) denoted by $\hat{\alpha}_{\mathrm{MJL}}$.

$$\hat{\alpha}_{\mathrm{MJL}} = (2\mathrm{I} - F_D)F_D^2\hat{\alpha}_{\mathrm{LS}} \tag{64}$$

Comparing $\hat{\alpha}_{\mathrm{MJL}}$ with $\hat{\alpha}_{\mathrm{JGLE}}$, it is obvious that the MJL is similar to the JGLE, but by inserting the GLE instead of the LS, f^i_{MJL} can be also expressed as a function of k^i_{GRE}.

$$f^i_{\mathrm{MJL}} = \frac{-\lambda_i + \sqrt{9\lambda_i^2 + 8\lambda_i k^i_{\mathrm{GRE}}}}{2(\lambda_i + k^i_{\mathrm{GRE}})} \tag{65}$$

6.3. **Jackknifing general ridge estimator and its modification form.** Khurana introduced the jackknifing generalized ridge estimator (JGRE) and the modified JGRE (MJR). Let $A_{\mathrm{MJR}} = \Lambda + K_{\mathrm{MJR}}$ and $K_{\mathrm{JGRE}} = \mathrm{diag}(k^i_{\mathrm{JGRE}})$, $K_{\mathrm{MJR}} = \mathrm{diag}(k^i_{\mathrm{MJR}})$, $i = 1, 2, \cdots, n$, where k^i_{JGRE} and k^i_{MJR} are all nonnegative, then the JGRE and the MJR are respectively written as

$$\hat{\alpha}_{\mathrm{JGRE}} = \left[\mathrm{I} - (A^{-1}_{\mathrm{JGRE}}K_{\mathrm{JGRE}})^2\right]\hat{\alpha}_{\mathrm{LS}} \tag{66}$$

$$\hat{\alpha}_{\mathrm{MJR}} = [\mathrm{I} - (A^{-1}_{\mathrm{MJR}}K_{\mathrm{MJR}})^2][\mathrm{I} - A^{-1}_{\mathrm{MJR}}K_{\mathrm{MJR}}]\hat{\alpha}_{\mathrm{LS}} \tag{67}$$

The componentwise estimators in Equations (66) and (67) have the following forms:

$$\hat{\alpha}^i_{\mathrm{JGRE}} = \frac{2k^i_{\mathrm{JGRE}} + \lambda_i}{k^i_{\mathrm{JGRE}} + \lambda_i}\hat{\alpha}^i_{\mathrm{GRE}} \tag{68}$$

$$\hat{\alpha}^i_{\mathrm{MJR}} = \frac{\lambda_i}{k^i_{\mathrm{MJR}} + \lambda_i}\hat{\alpha}^i_{\mathrm{JGRE}} \tag{69}$$

$\hat{\alpha}^i_{\mathrm{JGRE}}$ and $\hat{\alpha}^i_{\mathrm{MJR}}$ can be seen as the nonlinear transformations of $\hat{\alpha}^i_{\mathrm{GRE}}$ and $\hat{\alpha}^i_{\mathrm{JGRE}}$. Similar to the MJL, k^i_{MJR} can be expressed as a function of k^i_{GRE}.

$$k^i_{\mathrm{MJR}} = k^i_{\mathrm{GRE}} + \sqrt{k^i_{\mathrm{GRE}}(k^i_{\mathrm{GRE}} + \lambda_i)} \tag{70}$$

7. **Consistent Analysis of the Construction Features of Biased Estimators.** In the preceding sections, we compared eight types of biasing estimators. The comparative results consistently illustrate that if the variances of these improved estimators are the same, then the biases are the same as well. This conclusion implies that the rates of the variance's change with respect to the bias are all the same, although they present in various forms. Meanwhile, the value ranges of these alternatives are also equal. So the homologous characteristic of the biasing estimator class is a common law restricting all the proposed alternatives to the LS. This consistent character is due to the fact that these dominant methods originate from the shrunken transformation to the LS, and the shrunken operators are non-negative and less than 1 as shown in Table 1.

TABLE 1. The value scope of the shrunken coefficient

Biased Estimators	Biasing Parameter	Shrunken Coefficient
α^i_{GRE}	$k^i_{\mathrm{GRE}} > 0$	$0 < \frac{\lambda_i}{\lambda_i + k^i_{\mathrm{GRE}}} < 1$
α^i_{GLE}	$0 < d^i_{\mathrm{GLE}} < 1$	$0 < \frac{\lambda_i + d^i_{\mathrm{GLE}}}{\lambda_i + 1} < 1$
α^i_{MGRE}	$0 < t_i < 1 \quad 0 < k^i_{\mathrm{MGRE}} < 1$	$0 < \frac{\lambda_i + k^i_{\mathrm{MGRE}} t_i}{\lambda_i + k^i_{\mathrm{MGRE}}} < 1$
α^i_{MGLE}	$0 < d^i_{\mathrm{MGLE}} < 1$ and $0 < t_i < 1$	$0 < \frac{\lambda_i + t_i + (1-t_i) d^i_{\mathrm{MGLE}}}{\lambda_i + 1} < 1$
$\alpha^i_{(k,d)}$	$k^i_{(k,d)} > 0$ and $0 < d_{(k,d)} < 1$	$0 < \frac{\lambda_i + k^i_{(k,d)} d^i_{(k,d)}}{\lambda_i + k^i_{(k,d)}} < 1$
$\hat{\alpha}^i_{\mathrm{JGLE}}$	$0 < f_i < 1$	$0 < \frac{f_i}{2 - f_i} < 1$
$\hat{\alpha}^i_{\mathrm{MJL}}$	$0 < f_i < 1$	$0 < \frac{f_i^2}{2 - f_i} < 1$
$\hat{\alpha}^i_{\mathrm{JGRE}}$	$k^i_{\mathrm{JGRE}} > 0$	$0 < \frac{(2k^i_{\mathrm{JGRE}} + \lambda_i)\lambda_i}{(k^i_{\mathrm{JGRE}} + \lambda_i)^2} < 1$

The common characteristic constraining the value scope of the shrunken coefficient is that the shrunken coefficient essentially shrinks the components of the LS towards zero, but does not change the sign convention. Further summarizing Sections 3 to 7, as long as we know one kind of biased estimator, the value scopes of the biasing parameter, shrunken matrix, and estimator value for any other biased estimator can be determined. In other words, it is only necessary to research the GRE properties in terms of variance, bias, biasing parameter and its superiority to the LS, because the others can be easily obtained.

8. **Numerical Analysis.** Following our completion of the comparative analysis by componentwise estimators, for the sake of convenience and to illustrate the core question, we take single unknown parameter estimation as an example, focusing on the biasing parameters' effect on the variance, bias and MSE of the different estimators researched in this paper.

Figure 1 and Figure 2 illustrate that the variance and the bias are monotonically decreasing and monotonically increasing functions with respect to the biasing parameters, respectively. However, the change rate of the variance or bias curve to the biasing parameter for distinctive alternatives to LS displays diverse features, even when the variance or bias of the different estimators is equal. Comparing the variance of any biased estimator to that of LS, the former is always superior to the latter, but sacrifices its unbiasedness.

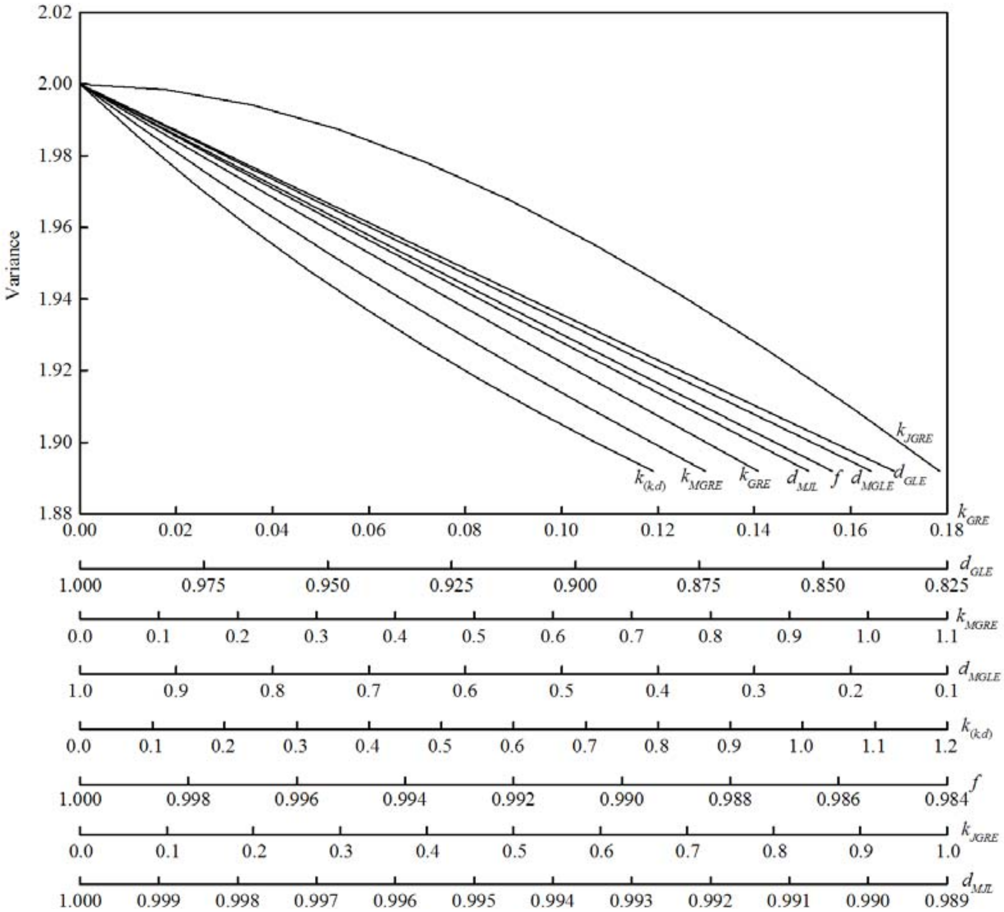

FIGURE 1. The curve of variance to biasing parameter

Figure 3 shows the curves of the MSE to the biasing parameter and confirms that any biased estimator is superior to the LS in the sense of MSE. When the maximum of the bias is acceptable, Figure 1 and Figure 3 imply that the biased estimator is always satisfactory. However, we must pay more attention to the change rate of the variance with respect to the bias as depicted in Figure 4. Analyzing Figure 4 for every dominant estimator, we immediately obtain that the curves of the variances' change rate with respect to the biases are always the same and have no relation to the estimator selected. This conclusion is coincident with Theorem 3.1. The reason for the consistency regulation is caused by the mutual conversation of the shrunken matrix acting on the LS among these biased estimators. Furthermore, the minimum MSEs with respect to different biased estimators are always the same, because the variance and the bias occur in pairs in any alternative methods. However, the point of the biasing parameter with respect to the optimal MSE is not to correspond to the extreme point of either variance or bias, i.e., the biasing parameter value with respect to the MSE is inconsistent with that of variance and bias. The cause of their being non-simultaneous is that the bias's influence on the degree of improvement in variance as shown by Figure 4 is degraded with the change in the scope of the biasing parameter. So the significant role of Figure 4 reflects the common law in which the biased estimators abide. The consistency regulation refined from Figure 4 guarantees that there is no doubt regarding the superiority of all the biased estimators being equivalent to each other; any one of them can be used to enhance precision directly.

The following numerical analysis concentrates on the mutual transformation of the biasing parameters between two different estimators under the same variance or bias

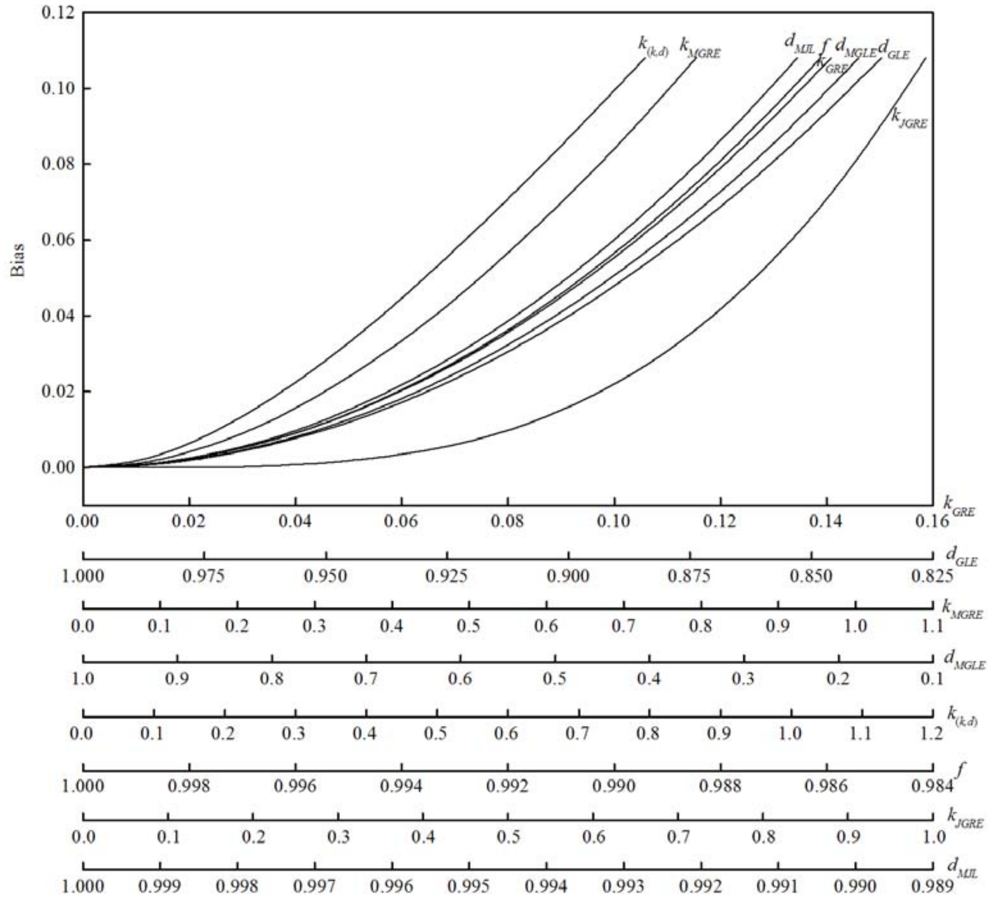

FIGURE 2. The curve of bias to biasing parameter

condition. We take ten numerical points within the scope of the GRE's biasing parameter at the same step and their corresponding variances and biases with regard to these samples can be easily calculated. Based on the conversation functions of the biasing parameters, we can consistently obtain ten points of the other estimators' biasing parameters and their corresponding variances and biases. Figures 5 to 7 are the curves of the same values of variance, bias and MSE with respect to the biasing parameters of different biased estimators. Analyzing the abscissas in any of the figures in Figures 5 to 7, we know that the change styles are distinguishable, i.e., either linear or nonlinear. The reason for this distinction is that different superior estimators have different shrunken transformation matrices. The far left abscissa in Figure 5 corresponds to the LS variance; however, the values of these end points are not always different. The distinctive form of shrunken matrices is the direct reason for this difference. The bias values generated from the biasing parameter samples, which correspond to the variance samples, are shown in Figure 6. It is obvious that the biasing parameter samples determined by the biases are identical to those determined by the variances. So we can choose either the variance or bias indexes to discuss the properties of biased estimators. Similarly, a comparison of the value of the biasing parameter samples in terms of MSE and of variance or bias can be obtained directly. Figure 8 illustrates the biasing parameters' relationships between the GRE and the other biased estimators. We can determine the values of the other estimators' biasing parameters by referring to that of the GRE. The GRE biasing parameter is regarded as the bridge applied for the biasing parameters conversation between any two other estimators. At the same time, it leads to the consistency regulation.

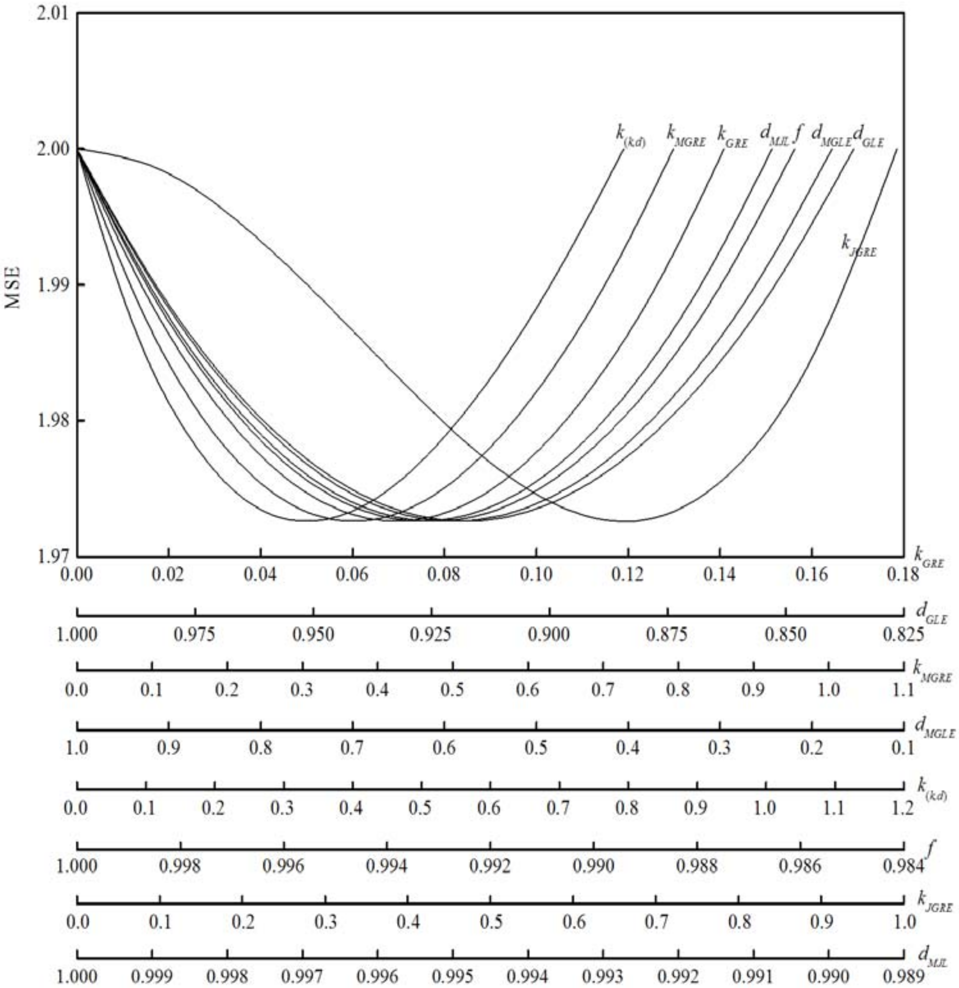

FIGURE 3. The curve of MSE to biasing parameter

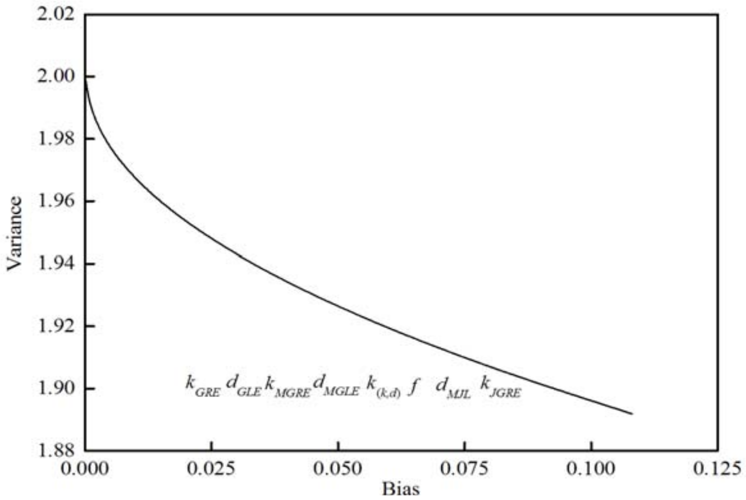

FIGURE 4. The curve of variance to bias

9. **Conclusions.** The paper investigates the consistency regulation to which the existing biased estimators abide. The variance change rate with respect to bias is taken as a criterion for evaluating the bias function regarding the improvement on the variance. From our thorough analysis, we conclude that for the proposed biased estimators superior to

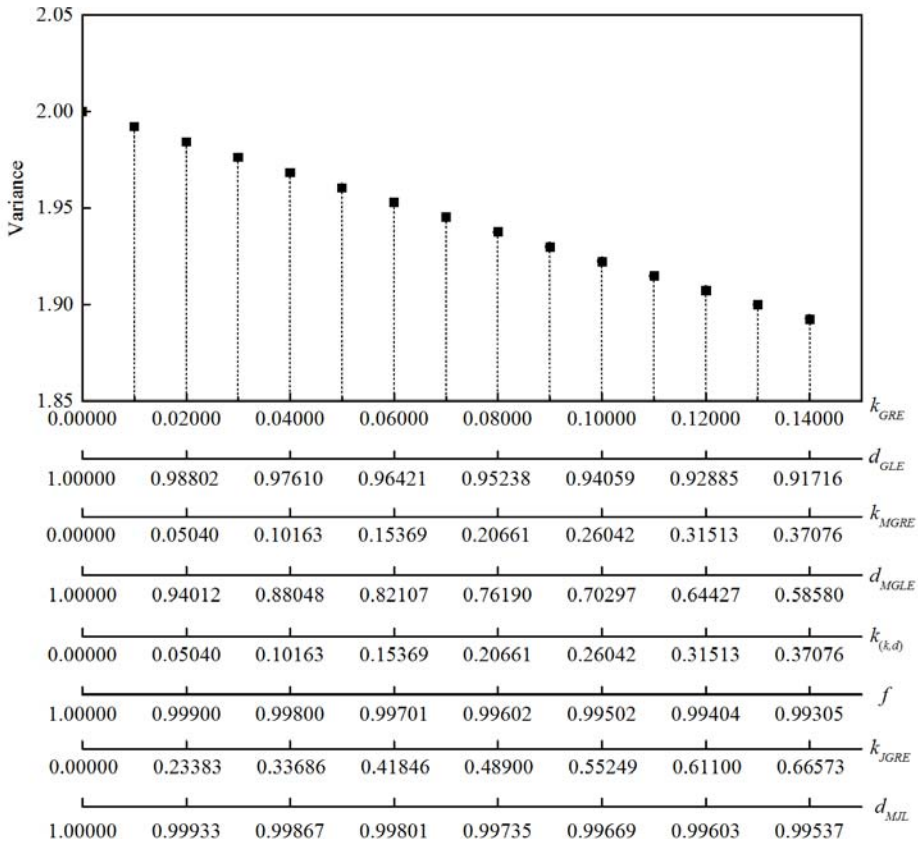

FIGURE 5. The inter-conversion of biasing parameters

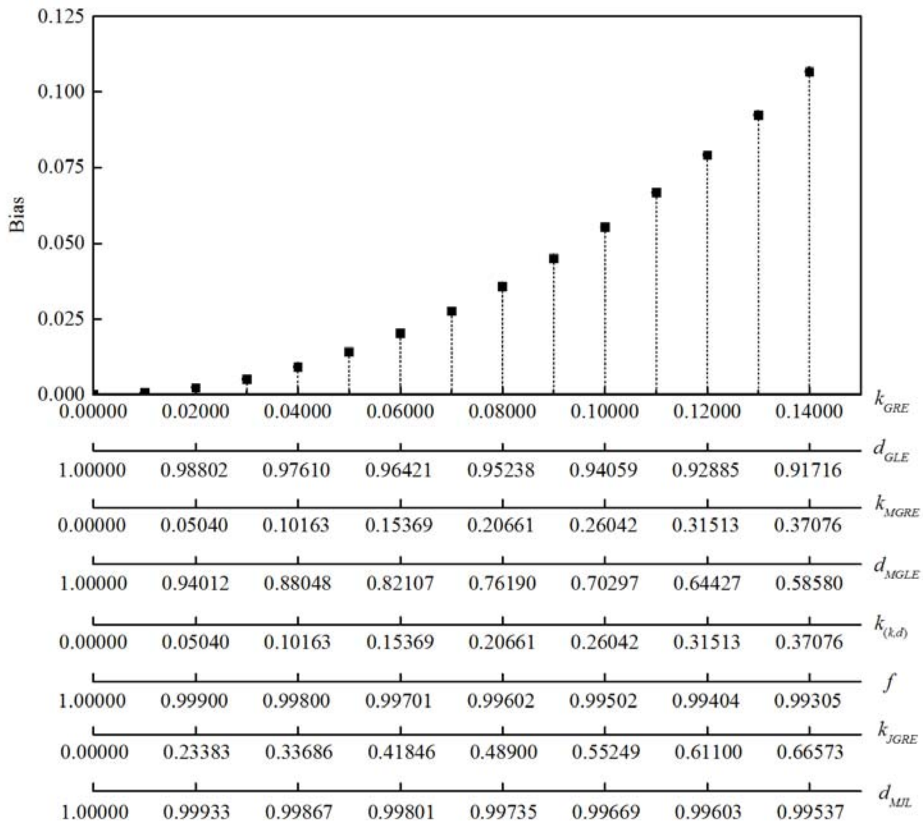

FIGURE 6. The inter-conversion of biasing parameters

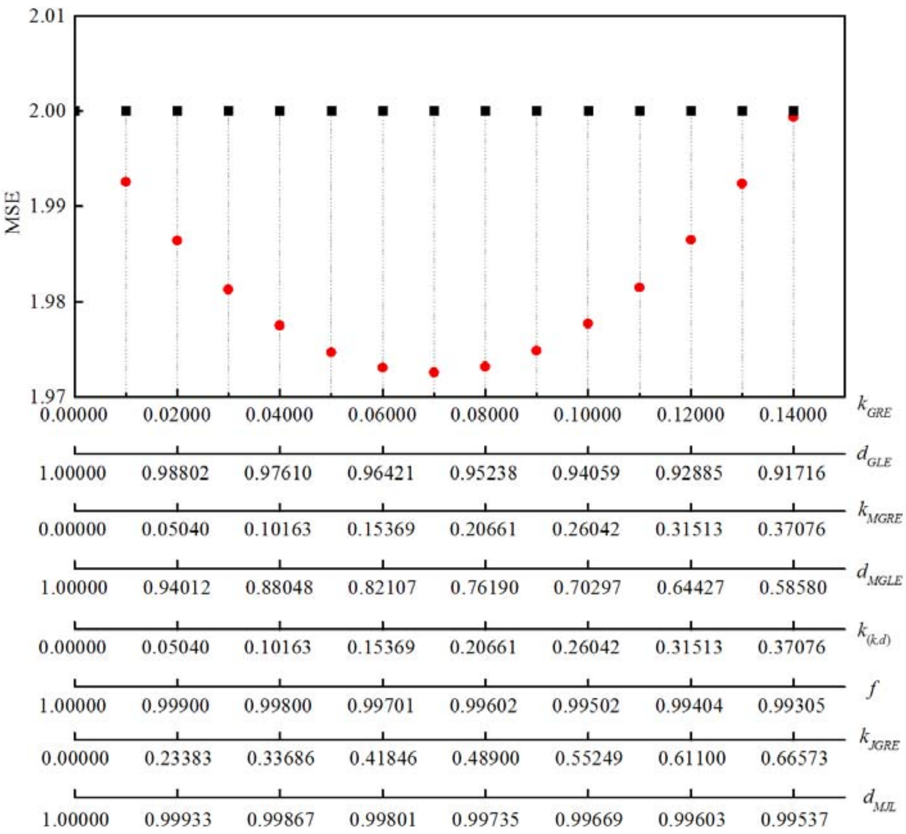

FIGURE 7. The inter-conversion of biasing parameters

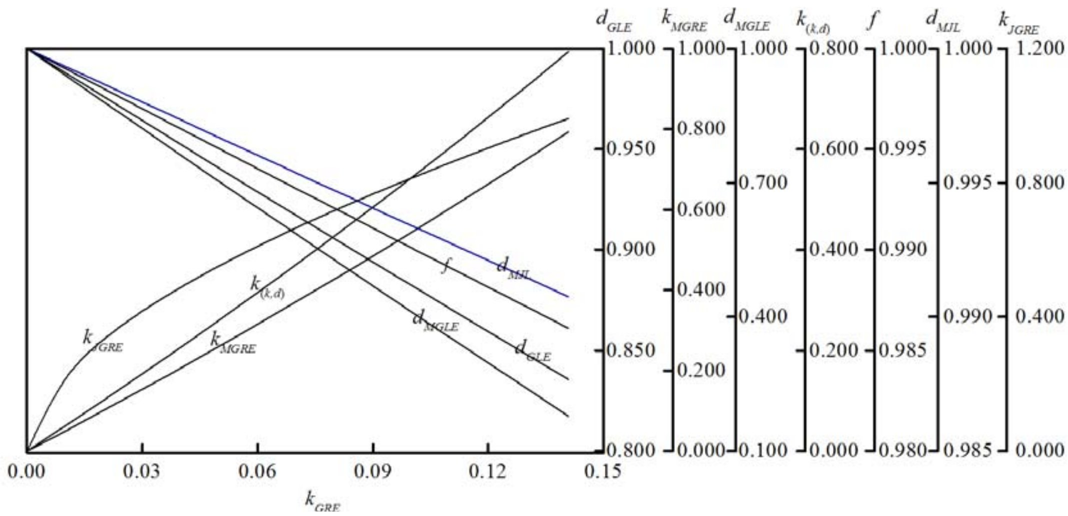

FIGURE 8. The inter-conversion of biasing parameters

the LS, the value scope of all the dominant estimators are coincident with each other. The biases are the same as long as the variances are the same. These two merit properties confirm that research on the characteristics of only one biased estimator is required, then those of the other biased estimators can be obtained without a complex solution procedure. The values' scopes of all the biased estimators are all the same. For any existing biased estimators that dominate the LS, the value scopes of the variances, biases, and the estimated unknown parameters are always the same and are independent of the selected estimating methods as long as the values of model parameters are the same.

So researchers may employ any of the biased estimators to enhance the precision of the estimation process. Today, the biased estimator theory is relatively mature, but if we do not balance biased estimator methods and application both in research and academia, we run the risk of ending up with an uninteresting appendix of statistics. In future work, we are determined to further popularize the engineering application of biased estimators in the communities of process control, chemometrics, system identification and so on, and especially in the chemometrics field where there are always correlative data in ill-conditioned situations underlying the spectroscopic process.

Acknowledgment. This work was supported by the National Basic Research Program of China (2012CB720500). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd Edition, Springer, New York, 1998.
[2] S. J. Haswell and A. D. Walmsley, Chemometrics: The issues of measurement and modeling, *Anal. Chim. Acta.*, vol.400, pp.399-412, 1999.
[3] A. M. Legendre, *Nouvelles Méthodes Pour La détermination Des Orbites Des Comètes: Avec Un Supplément Contenant Divers Perfectionnemens De Ces Méthodes Et Leur Application Aux Deux Comètesde*, 1805.
[4] Y. C. Eldar, A. Ben-Tal and A. Nemirovski, Robust mean-squared error estimation in the presence of model uncertainties, *IEEE Trans. Signal Process.*, vol.53, no.1, pp.168-181, 2005.
[5] Y. C. Eldar, Uniformly improving the Cramér-Rao bound and maximum-likelihood estimation, *IEEE Trans. Signal Process.*, vol.54, no.8, pp.2943-2956, 2006.
[6] Z. Ben-Haim and Y. C. Eldar, Blind minimax estimation, *IEEE Trans. Inf. Theory*, vol.53, no.9, pp.3145-3157, 2007.
[7] D. W. Marquaridt, Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation, *Technometrics*, vol.12, no.3, pp.591-612, 1970.
[8] C. Stein, Inadmissibility of the usual estimator for mean of multivariate normal distribution, *Proc. of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, vol.1, no.399, pp.197-206, 1956.
[9] W. James and C. Stein, Estimation with quadratic loss, *Proc. of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol.1, pp.311-319, 1961.
[10] S. L. Sclove, Improved estimators for coefficients in linear regression, *J. Amer. Statist. Assoc.*, vol.63, pp.596-606, 1968.
[11] A. E. Hoerl, Application of ridge analysis to regression problems, *Chem. Eng. Prog.*, vol.58, no.3, pp.54-59, 1962.
[12] A. E. Hoerl and R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, vol.12, no.1, pp.55-67, 1970.
[13] Y. L. Yue, X. Zuo and X. L. Luo, Confidence level based on ridge estimator in process measurement and it application, *Chin. J. Chem. Eng.*, vol.21, no.10, pp.1144-1154, 2013.
[14] A. Höskuldsson, Common framework for linear regression, *Chemom. Intell. Lab. Syst.*, vol.146, pp.250-262, 2015.
[15] G. Tutz and H. Binder, Boosting ridge regression, *Comput. Statist. Data Anal.*, vol.51, pp.6044-6059, 2007.
[16] B. F. Swindel, Good ridge estimators based on prior information, *Comm. Stat. Theory Methods*, vol.5, no.11, pp.1065-1075, 1976.
[17] K. J. Liu, A new class of some biased regression estimators, *Comm. Stat. Theory Methods*, vol.22, no.2, pp.393-402, 1993.
[18] K. J. Liu, Using Liu-type estimator to combat collinearity, *Comm. Stat. Theory Methods*, vol.32, no.5, pp.1009-1020, 2003.
[19] S. Skallıoglu and S. Kaciranlar, A new biased estimator based on the ridge estimation, *Stat Papers*, vol.49, no.4, pp.669-689, 2008.
[20] E. A. Duran and F. Akdeniz, Efficiency of the modified jackknifed Liu-type estimator, *Stat Papers*, vol.53, no.2, pp.265-280, 2012.

[21] D. V. Hinkley, On jackknifing in unbalanced situations, *Technometrics*, vol.19, no.3, pp.285-292, 1977.

[22] Y. L. Li and H. Yang, A new Liu-type estimator in linear regression model, *Stat Papers*, vol.53, no.2, pp.427-437, 2012.

[23] M. R. Ozkale and S. Kaciranlar, Superiority of the r-d class estimator over some estimators by the mean square error matrix criterion, *Stat. Probab. Lett.*, vol.77, no.4, pp.438-446, 2007.

[24] F. S. M. Batah, T. V. Ramanathan and S. D. Gore, The efficiency of modified jackknife and ridge type regression estimators: A comparison, *Surv. Math. Appl.*, vol.3, pp.111-122, 2008.

[25] M. Khurana, Y. P. Chaubey and S. Chandra, Jackknifing the ridge regression estimator: A revisit, *Technical Report*, Con-cordia University, 2012.

[26] K. Mansson, On ridge estimators for the negative binomial regression model, *Ecol. Model.*, vol.29, no.2, pp.178-184, 2012.