**Journal of Sciences, Islamic Republic of Iran 26(4): 355 - 366 (2015)** http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

# Detection of Outliers and Influential Observations in Linear Ridge Measurement Error Models with Stochastic Linear Restrictions

F. Ghapani, A. R. Rasekh*, M. R. Akhoond and B. Babadi

*Department of Statistics, Faculty of Mathematical Sciences and Computer, Shahid Chamran University, Ahvaz, Islamic Republic of Iran*

Received: 22 April 2015 / Revised: 31 October 2015 / Accepted: 18 November 2015

## Abstract

The aim of this paper is to propose some diagnostic methods in linear ridge measurement error models with stochastic linear restrictions using the corrected likelihood. Based on the bias-corrected estimation of model parameters, diagnostic measures are developed to identify outlying and influential observations. In addition, we derive the corrected score test statistic for outliers detection based on mean shift outlier models. The analogues of Cook's distance and likelihood distance are proposed to determine influential observations based on case deletion model. A parametric bootstrap procedure is used to obtain empirical distributions of the test statistics and a simulation study has been given to show the performance of the score test statistic. Finally, the proposed diagnostic procedures are illustrated on a numerical example to show the theoretical results.

**Keywords**: Case deletion, Corrected likelihood, Influential observations, Mean shift outlier model, outliers.

## Introduction

One of the basic assumptions in regression analysis is that all the observations are correctly observed. However, in many applications the observations are recorded with measurement errors. The presence of measurement errors in the observations violates the essential properties of estimators. An important issue in the area of measurement errors is to find the consistent estimators of the parameters. Several approaches have been developed for the measurement error problems (see, e.g., Cheng and Van Ness [1], Fuller [2] for more details). In order to correct the effects of measurement error on parameters estimation, Nakamura [3] considered an approach based on the correction of score function. This approach makes it possible to do inference as well as estimation of parameters without additional assumption.

Another standard assumption in the linear regression analysis is that all the explanatory variables are linearly independent. When this assumption is violated and the columns of the regression matrix are nearly dependent, the problem of collinearity enters into the data and the existence of collinearity in the linear regression model can lead to the method of least squares generally produces poor estimates of parameters. In order to

* Corresponding author: Tel: +986133331043; Fax: +986133331043; Email: Rasekh_a@scu.ac.ir

resolve this problem, several approaches have been suggested, among them, the ridge regression estimator was proposed by Hoerl and Kennard [4] to overcome the problem of collinearity for the estimation of regression parameters (see Belsley et al. [5], Mason and Gunst [6] and Belsley [7] for more details).

Another popular technique to overcome the collinearity problem is to consider parameter estimation in addition to the sample information such as some exact or stochastic linear restrictions on the unknown parameters (Rao et al. [8]). When such prior information can be expressed in the form of exact linear restrictions binding the regression coefficients, the restricted least squares estimator is used. The restricted least squares estimator is unbiased, consistent, satisfies the given linear restrictions on regression coefficients and has smaller variability around mean than the ordinary least square estimator when there is no measurement error in the data (see, e.g. Toutenburg [9], Rao et al. [8]). When prior information comes to stochastic linear restrictions, Durbin [10], Theil and Goldberger [11] and Theil [12] proposed the ordinary mixed estimator by combing the sample model with stochastic linear restrictions (see, e.g., Toutenburg [9]; Rao et al. [8], for more details). Sarkar [13] introduced a new restricted estimator by combining the restricted least square estimator with ordinary ridge estimator. Kaciranlar et al. [14] compared the estimator introduced by Sarkar [13] and the modified ridge regression estimator based on prior information proposed by Swindel [15]. Yalian and Yang [16] derived the stochastic restricted ridge estimator. He and Wu [17] proposed a new estimator to combat the collinearity in the linear model when there were stochastic linear restrictions on the regression coefficients. Wu and Liu [18] considered several estimators for estimating the stochastic restricted ridge regression estimators. Alkhamisi and MacNeill [19] derived the necessary and sufficient conditions for superiority of the restricted ridge estimator over the restricted least squares estimator by trace of the mean square error criterion. In this paper an attempt is made to find ridge estimators in measurement error models with stochastic linear restrictions using the appropriate corrected log-likelihood of Nakamura [3].

Outliers and influential data are observations that appear inconsistent with the other observations of a data set and can have more influence on the different aspects of the statistical analysis. Therefore, it is important to consider influential points in data analysis. In order to detect these kinds of observations, various methods, including case deletion model (CDM) and mean shift outlier model (MSOM), have been proposed in the literature (Cook and Weisberg, [20]). Wang [21] discussed the linear regression model with the random constraints and showed that the CDM is equivalent to the MSOM based on general least square estimate. In measurement error models, Kelly [22] and Wellman and Gunst [23] studied diagnostics methods. Zhong et al. [24] considered CDM and MSOM, using the corrected log-likelihood of Nakamura [3]. Zare and Rasekh [25] obtained diagnostic methods, including MSOM, for linear mixed measurement error models based on the corrected score function of Nakamura [3]. Babadi et al. [26] studied a variance shift model for a linear measurement error model using the corrected likelihood of Nakamura [3].

Walker and Birch [27] studied the influence of observations in ridge regression by case deletion, and the dependence of several influence measures derived from case deletion on the ridge parameter was studied. Rasekh and Mohtashami [28] extended results from Rasekh and Fieller [29] and derived influence function of ridge estimate in measurement error models using case deletion. Rasekh [30] assessed the local influence of observations on the ridge estimate in the measurement error models.

The plan of this paper is as follows. In preliminary section the measurement error model, the needed notations and some preliminaries are presented. The main results of the paper, including the stochastic restricted ridge estimation based on the corrected log-likelihood of Nakamura [3] and it's properties, the diagnostic models and the corrected score test for detecting outliers are proposed in the main results section. Furthermore, case deletion diagnostics for detecting influential points and a parametric bootstrap procedure for generating the empirical distribution of the given statistics are developed. Finally, a simulation study and an illustrative example of the realdata are performed.

***Prelimaniers***

Consider the linear measurement error model:

$$(1) \quad \begin{aligned} \boldsymbol{y} &= \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \\ \boldsymbol{X} &= \boldsymbol{Z} + \Delta, \end{aligned}$$

where $\boldsymbol{y} = (y_1, y_2, ..., y_n)'$ is an $n \times 1$ vector of response variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, $\boldsymbol{Z}$ is an $n \times p$ matrix of unobservable values of explanatory variables which can be observed through the matrix $\boldsymbol{X}$ with the measurement error $\Delta' = [\delta_1, \delta_2, ..., \delta_n]$, where $\delta_i$, $i = 1, ..., n$ are $p \times 1$ uncorrelated random vectors with $E(\delta_i) = \mathbf{0}$ and

$Var(\delta_i) = \Lambda$. We assume that the common variance $\Lambda$ of measurement errors associated with the explanatory variables is known. Also $\varepsilon$ is an $n \times 1$ vector of unobservable random errors with $E(\varepsilon) = \mathbf{0}$ and $Var(\varepsilon) = \sigma^2 \boldsymbol{I}_n$. We assume that $\varepsilon$ and $\Delta$ are mutually independent and we denote the *i*th rows of matrices $\boldsymbol{Z}$ and $\boldsymbol{X}$ with $z_i'$ and $x_i'$, respectively. For model (1), the log-likelihood and the appropriate corrected log-likelihood are given by

$$l(\beta,\sigma^2,\boldsymbol{Z},\boldsymbol{y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}[(\boldsymbol{y}-\boldsymbol{Z}\beta)'(\boldsymbol{y}-\boldsymbol{Z}\beta)],$$
$$l^*(\beta,\sigma^2,\boldsymbol{X},\boldsymbol{y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}[(\boldsymbol{y}-\boldsymbol{X}\beta)'(\boldsymbol{y}-\boldsymbol{X}\beta) - n\beta'\Lambda\beta].$$

Let $E^*$ denotes the conditional mean with respect to $\boldsymbol{X}$ given $\boldsymbol{y}$. The corrected log-likelihood $l^*(\beta,\sigma^2,\boldsymbol{X},\boldsymbol{y})$ should satisfy

$$E^*\left[\frac{\partial}{\partial\beta} l^*(\beta,\sigma^2,\boldsymbol{X},\boldsymbol{y})\right] = \frac{\partial}{\partial\beta} l(\beta,\sigma^2,\boldsymbol{Z},\boldsymbol{y})$$
and
$$E^*\left[\frac{\partial}{\partial\sigma^2} l^*(\beta,\sigma^2,\boldsymbol{X},\boldsymbol{y})\right] = \frac{\partial}{\partial\sigma^2} l(\beta,\sigma^2,\boldsymbol{Z},\boldsymbol{y})\cdot$$

The corrected score estimates of model (1) will be obtained with differentiating from $l^*(\beta,\sigma^2,\boldsymbol{X},\boldsymbol{y})$ with respect to the $\beta$ and $\sigma^2$ and are given by $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X} - n\Lambda)^{-1}\boldsymbol{X}'\boldsymbol{y}$ and $\hat{\sigma}^2 = \frac{1}{n}\left[(\boldsymbol{y}-\boldsymbol{X}\hat{\beta})'(\boldsymbol{y}-\boldsymbol{X}\hat{\beta})' - n\hat{\beta}'\Lambda\hat{\beta}\right]$ (see, Nakamura [3]).

## Results

***Measurement error models with stochastic linear restrictions***

We assume that the vector of parameters $\beta$ is subject to the following stochastic linear restrictions:

(2) $\boldsymbol{r} = \boldsymbol{R}\beta + \boldsymbol{e}$,

where $\boldsymbol{r}$ is a $q \times 1$ observable random vector, $\boldsymbol{R}$ is a $q \times p$ matrix of known constants of rank $q$ for $q < p$, $\boldsymbol{e}$ is a $q \times 1$ error vector with $E(\boldsymbol{e}) = \mathbf{0}$ and $Var(\boldsymbol{e}) = \sigma^2\boldsymbol{W}$, $\boldsymbol{W}$ is a positive definite matrix of known elements. Furthermore, we assume that random vector $\boldsymbol{e}$ is independent of $\varepsilon$ and $\Delta$. The log-likelihood and the appropriate corrected log-likelihood for model (1) with stochastic linear restrictions (2) are given by

$$l(\beta,\sigma^2,\boldsymbol{Z},\boldsymbol{y},\boldsymbol{r}) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log|\boldsymbol{W}|$$
$$-\frac{1}{2\sigma^2}\left[(\boldsymbol{y}-\boldsymbol{Z}\beta)'(\boldsymbol{y}-\boldsymbol{Z}\beta) + (\boldsymbol{r}-\boldsymbol{R}\beta)'\boldsymbol{W}^{-1}(\boldsymbol{r}-\boldsymbol{R}\beta)\right],$$
$$l^*(\beta,\sigma^2,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r}) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log|\boldsymbol{W}|$$
$$-\frac{1}{2\sigma^2}\left[(\boldsymbol{y}-\boldsymbol{X}\beta)'(\boldsymbol{y}-\boldsymbol{X}\beta) - n\beta'\Lambda\beta + (\boldsymbol{r}-\boldsymbol{R}\beta)'\boldsymbol{W}^{-1}(\boldsymbol{r}-\boldsymbol{R}\beta)\right]$$

respectiveely, where, $N = n + q$, and these have the following properties:

$$E^*\left[\frac{\partial l^*(\beta,\sigma^2,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r})}{\partial\beta}\right] = \frac{\partial l(\beta,\sigma^2,\boldsymbol{Z},\boldsymbol{y},\boldsymbol{r})}{\partial\beta}$$

$$E^*\left[\frac{\partial l^*(\beta,\sigma^2,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r})}{\partial\sigma^2}\right] = \frac{\partial l(\beta,\sigma^2,\boldsymbol{Z},\boldsymbol{y},\boldsymbol{r})}{\partial\sigma^2}.$$

The corrected score estimates of $\beta$ and $\sigma^2$ for model (1) with stochastic linear restrictions (2) will be obtained with differentiating from $l^*(\beta,\sigma^2,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r})$ with respect to the $\beta$ and $\sigma^2$. We call these estimators, denoted by $\hat{\beta}_r$ and $\hat{\sigma}_r^2$, respectively, as the mixed estimators. Then we have

$$\hat{\beta}_r = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R} - n\Lambda)^{-1}(\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{r})$$
$$\hat{\sigma}_r^2 = \frac{1}{N}\left[(\boldsymbol{y}-\boldsymbol{X}\hat{\beta}_r)'(\boldsymbol{y}-\boldsymbol{X}\hat{\beta}_r) - n\hat{\beta}_r'\Lambda\hat{\beta}_r + (\boldsymbol{r}-\boldsymbol{R}\hat{\beta}_r)'\boldsymbol{W}^{-1}(\boldsymbol{r}-\boldsymbol{R}\hat{\beta}_r)\right]$$
$$= \frac{1}{N}\left[\boldsymbol{y}'\boldsymbol{y} - \hat{\beta}_r'\boldsymbol{X}'\boldsymbol{y} - \hat{\beta}_r'\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{r} + \boldsymbol{r}'\boldsymbol{W}^{-1}\boldsymbol{r}\right].$$

***Ridge estimation under the stochastic linear restrictions***

To reduce the effect of collinearity, we propose the ridge estimator of $\beta$ under the stochastic linear restrictions. We consider the augmented model

(3) $\boldsymbol{u} = \boldsymbol{U}\beta + \eta$

in which

$$\boldsymbol{u}_{m\times 1} = \begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{r}\\ \mathbf{0}\end{bmatrix},\quad \boldsymbol{U}_{m\times p} = \begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{R}\\ \sqrt{k}\boldsymbol{I}_p\end{bmatrix},\quad \eta_{m\times 1} = \begin{bmatrix}\varepsilon - \Delta\beta\\ \boldsymbol{e}\\ \varphi\end{bmatrix}$$

for $m = n + q + p$, and $\eta$ is a random vector with $E(\eta) = \mathbf{0}$ and

$$Var(\eta) = BlockDiag\left[(\sigma^2 + \beta'\Lambda\beta)\boldsymbol{I}_n, \sigma^2\boldsymbol{W}, \sigma^2\boldsymbol{I}_p\right].$$

Here, the parameter $k > 0$ denotes the ridge parameter,

$\boldsymbol{\varphi}$ is an error vector with $E(\boldsymbol{\varphi})=\mathbf{0}$ and $Var(\boldsymbol{\varphi})=\sigma^2\boldsymbol{I}_p$. For model (3), the appropriate corrected log- likelihood is given by

$$l^*(\boldsymbol{\beta},\sigma^2,k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r})=-\frac{m}{2}\log(2\pi\sigma^2)-\frac{1}{2}\log|\boldsymbol{W}| -\frac{1}{2\sigma^2}\left[(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})-n\boldsymbol{\beta}'\Lambda\boldsymbol{\beta}+(\boldsymbol{r}-\boldsymbol{R\beta})'\boldsymbol{W}^{-1}(\boldsymbol{r}-\boldsymbol{R\beta})+k\boldsymbol{\beta}'\boldsymbol{\beta}\right].$$

The corrected score estimates of $\boldsymbol{\beta}$ and $\sigma^2$ will be obtained with differentiating from the corrected log-likelihood of model (3) with respect to the $\boldsymbol{\beta}$ and $\sigma^2$. We call these estimators as mixed ridge estimators and are denoted by $\hat{\boldsymbol{\beta}}_r(k)$ and $\hat{\sigma}_r^2(k)$, respectively. Then we obtain

(4)

$$\hat{\boldsymbol{\beta}}_r(k)=(\boldsymbol{X'X}+\boldsymbol{R'W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)^{-1}(\boldsymbol{X'y}+\boldsymbol{R'W}^{-1}\boldsymbol{r}),\ \ k>0,$$

$$\hat{\sigma}_r^2(k)=\frac{1}{m}\left[\boldsymbol{y'y}-\hat{\boldsymbol{\beta}}_r'(k)\boldsymbol{X'y}-\hat{\boldsymbol{\beta}}_r'(k)\boldsymbol{R'W}^{-1}\boldsymbol{r}+\boldsymbol{r'W}^{-1}\boldsymbol{r}\right].$$

Using matrix results (see, Rao et al. [8], Theorem A.18), we can write

$$\hat{\boldsymbol{\beta}}_r(k)=\hat{\boldsymbol{\beta}}(k)+(\boldsymbol{X'X}-n\Lambda+k\boldsymbol{I}_p)^{-1}\boldsymbol{R}'\left[I+\boldsymbol{R}(\boldsymbol{X'X}-n\Lambda+k\boldsymbol{I}_p)^{-1}\boldsymbol{R}'\right]^{-1}\left[r-\boldsymbol{R}\hat{\boldsymbol{\beta}}(k)\right],$$

where $\hat{\boldsymbol{\beta}}(k)=(\boldsymbol{X'X}-n\Lambda+k\boldsymbol{I}_p)^{-1}\boldsymbol{X'y}$.

Furthermore, as $k$ tends to zero, $\hat{\boldsymbol{\beta}}_r(k)$ approaches to the estimator $\hat{\boldsymbol{\beta}}_r$.

***Asymptotic properties of the ridge estimator***

The exact distribution and finite sample properties of the corrected score function estimates shown in (4) are difficult to derive. So we propose to employ the large sample asymptotic approximation theory to study the asymptotic distribution of the estimators. We assume that all the derivatives related to the log-likelihood exist and the parameter $\boldsymbol{\beta}$ is identifiable. We also assume that as $n$ tends to infinity, the limits of $n^{-1}(\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R})$ and $n^{-1}(\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R}+k\boldsymbol{I}_p)$ exist and $E_0$ denotes the global expectation taken at the true value $\boldsymbol{\beta}_0$.

**Theorem 1:** $\hat{\boldsymbol{\beta}}_r(k)$ has asymptotic normal distribution with mean $\boldsymbol{M}_k\boldsymbol{M}_0^{-1}\boldsymbol{\beta}_0$ and covariance matrix $AVar\left[\hat{\boldsymbol{\beta}}_r(k)\right]=\boldsymbol{S}_k=\boldsymbol{M}_k(\boldsymbol{B}+\sigma^2\boldsymbol{M}_0^{-1})\boldsymbol{M}_k$, where $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$,

$\boldsymbol{M}_k=(\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R}+k\boldsymbol{I}_p)^{-1}$,

$\boldsymbol{M}_0^{-1}=\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R}$ and

$\boldsymbol{B}=(n\sigma^2+\boldsymbol{\beta}_0'\boldsymbol{Z'Z}\boldsymbol{\beta}_0)\Lambda$.

**Proof:** Since $E(\boldsymbol{X'X})=\boldsymbol{Z'Z}+n\Lambda$, by Fung et al. [31] we have

(5) $\boldsymbol{X'X}=\boldsymbol{Z'Z}+n\Lambda+\mathbf{O}_p(n^{\frac{1}{2}})$.

Then we can write

(6)

$$n^{-1}(\boldsymbol{X'X}+\boldsymbol{R'W}^{-1}\boldsymbol{R}+k\boldsymbol{I}_p)=n^{-1}(\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R}+k\boldsymbol{I}_p)+\Lambda+\mathbf{O}_p(n^{-\frac{1}{2}})$$

It follows from (4) and (6) that

(7)

$$\begin{aligned}\sqrt{n}\hat{\boldsymbol{\beta}}_r(k)&=\left[n^{-1}(\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R}+k\boldsymbol{I}_p)+\mathbf{O}_p(n^{-\frac{1}{2}})\right]^{-1}n^{-\frac{1}{2}}(\boldsymbol{X'y}+\boldsymbol{R'W}^{-1}\boldsymbol{r})\\&=\left[\boldsymbol{I}_p+\mathbf{O}_p(n^{-\frac{1}{2}})\right]^{-1}\left[n^{-1}(\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R}+k\boldsymbol{I}_p)\right]^{-1}n^{-\frac{1}{2}}(\boldsymbol{X'y}+\boldsymbol{R'W}^{-1}\boldsymbol{r})\\&=\left[\boldsymbol{I}_p+\mathbf{O}_p(n^{-\frac{1}{2}})\right]\left[n^{-1}(\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R}+k\boldsymbol{I}_p)\right]^{-1}n^{-\frac{1}{2}}(\boldsymbol{X'y}+\boldsymbol{R'W}^{-1}\boldsymbol{r}),\end{aligned}$$

where $\left[\boldsymbol{I}_p+\mathbf{O}_p(n^{-\frac{1}{2}})\right]^{-1}=\boldsymbol{I}_p+\mathbf{O}_p(n^{-\frac{1}{2}})$ is obtained from Taylor series expansion. Moreover, since the limit of $\boldsymbol{C}=n^{-1}(\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R}+k\boldsymbol{I}_p)$ exists, then (7) can be written as

(8) $\sqrt{n}\hat{\boldsymbol{\beta}}_r(k)=\boldsymbol{C}^{-1}\xi+\boldsymbol{C}^{-1}\xi\mathbf{O}_p(n^{-\frac{1}{2}})$,

where $\xi=n^{-\frac{1}{2}}(\boldsymbol{X'y}+\boldsymbol{R'W}^{-1}\boldsymbol{r})$ is asymptotically normal (Fung et al. [31]). It follows from $E_0(\boldsymbol{X'y}+\boldsymbol{R'W}^{-1}r)=(\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R})\boldsymbol{\beta}_0$ that $E_0(\xi)=n^{-\frac{1}{2}}(\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R})\boldsymbol{\beta}_0$ and by Fung et al. [31], we can write

$$n^{-1}(\boldsymbol{X'y}+\boldsymbol{R'W}^{-1}\boldsymbol{r})=n^{-1}(\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R})\boldsymbol{\beta}_0+\boldsymbol{O}_p(n^{-\frac{1}{2}}).$$

Therefore, we have $\xi=\sqrt{n}\left[n^{-1}(\boldsymbol{Z'Z}+\boldsymbol{R'W}^{-1}\boldsymbol{R})\right]\boldsymbol{\beta}_0+\boldsymbol{O}_p(1)$, then (8) can be written as

(9) $\sqrt{n}\hat{\boldsymbol{\beta}}_r(k)=\boldsymbol{C}^{-1}\xi+\boldsymbol{O}_p(1)$.

Consequently, $\sqrt{n}\left[\hat{\boldsymbol{\beta}}_r(k)-\boldsymbol{M}_k\boldsymbol{M}_0^{-1}\boldsymbol{\beta}_0\right]$ has asymptotically normal distribution with zero mean. Furthermore, from (9) we have

$AVar\left[\sqrt{n}\hat{\beta}_r(k)\right] = \boldsymbol{C}^{-1}Var(\xi)\boldsymbol{C}^{-1}$. The variance of $\xi$ can be obtained from

$$Var(\xi) = E_{\tilde{y}}\left[Var(\xi|\tilde{\boldsymbol{y}})\right] + Var_{\tilde{y}}\left[E(\xi|\tilde{\boldsymbol{y}})\right]$$
$$= n^{-1}E_{\tilde{y}}(\boldsymbol{y}'\boldsymbol{y}\Lambda) + n^{-1}Var_{\tilde{y}}(\boldsymbol{Z}'\boldsymbol{y} + \boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{r}),$$

where $E_{\tilde{y}}$ and $\mathrm{var}_{\tilde{y}}$ denote the expectation and variance with respect to the random vector $\tilde{\boldsymbol{y}}' = (\boldsymbol{y}', \boldsymbol{r}')$. Since $E_0(y'y) = n\sigma^2 + \beta_0'\boldsymbol{Z}'\boldsymbol{Z}\beta_0$ and $Var_{\tilde{y}}(\boldsymbol{Z}'\boldsymbol{y} + \boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{r}) = \sigma^2\boldsymbol{M}_0^{-1}$, therefore, $Var(\xi) = n^{-1}(\boldsymbol{B} + \sigma^2\boldsymbol{M}_0^{-1})$ whose limit exists as $n$ tends to infinity by the assumptions. Thus, $AVar\left[\hat{\beta}_r(k)\right] = \boldsymbol{S}_k = \boldsymbol{M}_k(\boldsymbol{B} + \sigma^2\boldsymbol{M}_0^{-1})\boldsymbol{M}_k$, this completes the proof.

**Corollary:** Since the limit of $n^{-1}(\boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R})$ exists as $n$ tends to infinity and $k = 0$, we have $\boldsymbol{S}_0^{-\frac{1}{2}}(\hat{\beta}_r - \beta_0) \xrightarrow{D} N(\boldsymbol{0}, \boldsymbol{I}_P)$, where $\boldsymbol{S}_0 = \boldsymbol{M}_0(\boldsymbol{B} + \sigma^2\boldsymbol{M}_0^{-1})\boldsymbol{M}_0$ is the asymptotic variance of $\hat{\beta}_r$.

***Choice of the ridge parameter***

We use a slight extension of the mean squared error matrix (**MSEM**) criterion, considered by Ozkale [32] to study the superiority of $\hat{\beta}_r(k)$ over $\hat{\beta}_r$. The **MSEM** of an estimator $\hat{\beta}$ of $\beta$ is defined as

$$\mathbf{MSEM}(\hat{\beta}) = E(\hat{\beta} - \beta)'(\hat{\beta} - \beta) = Var(\hat{\beta}) + Bias(\hat{\beta})Bias(\hat{\beta})',$$

where $Bias(\hat{\beta}) = E(\hat{\beta}) - \beta$ denotes the bias vector. For any two given estimators $\hat{\beta}_1$ and $\hat{\beta}_2$, the estimator $\hat{\beta}_2$ is said to be superior to $\hat{\beta}_1$ under the **MSEM** criterion if $\boldsymbol{D}(\hat{\beta}_1, \hat{\beta}_2) = \mathbf{MSEM}(\hat{\beta}_1) - \mathbf{MSEM}(\hat{\beta}_2)$, is a nonnegative definite (*n.n.d.*) matrix, i.e. $\boldsymbol{D}(\hat{\beta}_1, \hat{\beta}_2) \geq \boldsymbol{0}$. If $\boldsymbol{D}(\hat{\beta}_1, \hat{\beta}_2)$ is positive definite (p.d.), $\hat{\beta}_2$ is said to be strongly superior to $\hat{\beta}_1$, i.e. $\boldsymbol{D}(\hat{\beta}_1, \hat{\beta}_2) > \boldsymbol{0}$. We can obtain the asymptotic **MSEM** of the estimators $\hat{\beta}_r$ and $\hat{\beta}_r(k)$ as follows

(10) $\mathbf{AMSEM}(\hat{\beta}_r) = \boldsymbol{M}_0(\boldsymbol{B} + \sigma^2\boldsymbol{M}_0^{-1})\boldsymbol{M}_0$

(11)
$\mathbf{AMSEM}\left[\hat{\beta}_r(k)\right] = \boldsymbol{M}_k(\boldsymbol{B} + \sigma^2\boldsymbol{M}_0^{-1})\boldsymbol{M}_k + k^2\boldsymbol{M}_k\beta_0\beta_0'\boldsymbol{M}_k$.

We are now interested in knowing under which conditions $\hat{\beta}_r(k)$ is better than $\hat{\beta}_r$. For this, we investigate the difference

$$\boldsymbol{D}\left[\hat{\beta}_r, \hat{\beta}_r(k)\right] = \mathbf{AMSEM}(\hat{\beta}_r) - \mathbf{AMSEM}\left[\hat{\beta}_r(k)\right].$$

When $\boldsymbol{D}\left[\hat{\beta}_r, \hat{\beta}_r(k)\right]$ is a p.d. matrix, $\hat{\beta}_r(k)$ is preferred to $\hat{\beta}_r$. From (10) and (11), we find the matrix difference as

(12)
$\boldsymbol{D}\left[\hat{\beta}_r, \hat{\beta}_r(k)\right] = \boldsymbol{M}_k\left[2\sigma^2k\boldsymbol{I}_p + k^2\sigma^2\boldsymbol{M}_0 + 2kB\boldsymbol{M}_0 + k^2\boldsymbol{M}_0B\boldsymbol{M}_0 - k^2\beta_0\beta_0'\right]\boldsymbol{M}_k$.

Note that $k^2\sigma^2\boldsymbol{M}_0 + 2kB\boldsymbol{M}_0 + k^2\boldsymbol{M}_0B\boldsymbol{M}_0$ is p.d. Therefore, using Farebrother [33] we have that (12) is p.d. if $2k\sigma^2 - k^2\beta_0\beta_0'I_p$ is positive semi-definite (p.s.d.). Thus a sufficient condition is $k \leq \frac{2\sigma^2}{\beta_0'\beta_0}$. We may replace the unknown parameters in $k$ by appropriate estimators to obtain $\hat{k} = \frac{2\hat{\sigma}_r^2}{\hat{\beta}_r'\hat{\beta}_r}$.

***Mean shift outlier model and score test statistic***

The MSOM is a common approach for detecting outlier observations (Cook and Weisberg, [20]). Suppose that the *i*th case is a candidate for an outlier, then MSOM in model (3) can be represented as

(13)
$$y_j = \boldsymbol{z}_j'\beta + \varepsilon_j, \; j = 1, \ldots, n, \; j \neq i,$$
$$y_i = \boldsymbol{z}_i'\beta + \gamma + \varepsilon_i,$$
$$\boldsymbol{x}_l' = \boldsymbol{z}_l' + \delta_l', \; l = 1, \ldots, n, \text{ subject to}$$
$$\boldsymbol{r} = \boldsymbol{R}\beta + \boldsymbol{e}, \text{with}$$
$$\boldsymbol{0} = \sqrt{k}\beta + \varphi, \qquad k > 0,$$

where $\gamma$ is an extra parameter, which describes the outlier in *i*th case (Cook and Weisberg, [20]). It is easily seen that, the nonzero value of $\gamma$ implies that the *i*th case may be an outlier. To detect outliers, we estimate the parameter $\beta$ and an outlier test can be formulated as a test of the null hypothesis that $\gamma = 0$. The corrected score estimates of $\beta$, $\sigma^2$ and $\gamma$ in MSOM (13) are denoted by $\hat{\beta}_m(k)$, $\hat{\sigma}_m^2(k)$ and $\hat{\gamma}_m(k)$, respectively, where $m$ indicate the estimate of parameters in MSOM. We rearrange the elements of the $\boldsymbol{y}$ and $\boldsymbol{X}$ so that the $i$th deleted case to be in the first row. Then we have $\boldsymbol{y}=\begin{bmatrix} y_i \\ \boldsymbol{y}_{(i)} \end{bmatrix}$ and $\boldsymbol{X}=\begin{bmatrix} \boldsymbol{x}_i' \\ \boldsymbol{X}_{(i)} \end{bmatrix}$. Therefore, the corrected log-likelihood of (13) is given by

$$\begin{aligned} & l_m^*(\boldsymbol{\beta},\sigma^2,\gamma,k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r}) \\ &= -\frac{m}{2}\log(2\pi\sigma^2)-\frac{1}{2}\log|\boldsymbol{W}|-\frac{1}{2\sigma^2}\Big[(\boldsymbol{y}_{(i)}-\boldsymbol{X}_{(i)}\boldsymbol{\beta})'(\boldsymbol{y}_{(i)}-\boldsymbol{X}_{(i)}\boldsymbol{\beta}) \\ &\quad +(y_i-\boldsymbol{x}_i'\boldsymbol{\beta}-\gamma)^2-n\boldsymbol{\beta}'\Lambda\boldsymbol{\beta}+(\boldsymbol{r}-\boldsymbol{R}\boldsymbol{\beta})'\boldsymbol{W}^{-1}(\boldsymbol{r}-\boldsymbol{R}\boldsymbol{\beta})+k\,\boldsymbol{\beta}'\boldsymbol{\beta}\Big]. \end{aligned}$$

The corrected score estimates $\hat{\beta}_m(k)$, $\hat{\gamma}_m(k)$ and $\hat{\sigma}_m^2(k)$ are derived with differentiating of $l_m^*$ with respect to $\beta$, $\gamma$ and $\sigma^2$, respectively. Therefore, we have

$$\begin{aligned} \hat{\boldsymbol{\beta}}_m(k) &= (\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)^{-1}\left[\boldsymbol{X}'\boldsymbol{y}-\boldsymbol{x}_i\hat{\gamma}_m(k)+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{r}\right] \\ &= \hat{\boldsymbol{\beta}}_r(k)-(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)^{-1}\boldsymbol{x}_i\hat{\gamma}_m(k) \end{aligned}$$

$$\hat{\gamma}_m(k)=y_i-\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_m(k).$$

Replacing $\hat{\beta}_m(k)$ into $\hat{\gamma}_m(k)$ we obtain

$$\hat{\gamma}_m(k)=y_i-\boldsymbol{x}_i'\left\{\hat{\boldsymbol{\beta}}_r(k)-\left[(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)^{-1}\boldsymbol{x}_i\hat{\gamma}_m(k)\right]\right\}=\frac{\hat{v}_i}{c_i}.$$

in which

$c_i=1-\boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)^{-1}\boldsymbol{x}_i$ and $\hat{v}_i=y_i-\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r(k)$. Furthermore, for the estimate of $\sigma^2$ in model (13) we have

$$\begin{aligned} m\hat{\sigma}_m^2(k) &= \left[\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_m(k)\right]'\left[\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_m(k)\right]-\left[y_i-\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_m(k)\right]^2+\left[y_i-\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_m(k)-\hat{\gamma}_m(k)\right]^2 \\ &\quad -n\hat{\boldsymbol{\beta}}_m'(k)\Lambda\hat{\boldsymbol{\beta}}_m(k)+\left[\boldsymbol{r}-\boldsymbol{R}\hat{\boldsymbol{\beta}}_m(k)\right]'\boldsymbol{W}^{-1}\left[\boldsymbol{r}-\boldsymbol{R}\hat{\boldsymbol{\beta}}_m(k)\right]+k\,\hat{\boldsymbol{\beta}}_m'(k)\hat{\boldsymbol{\beta}}_m(k) \\ &= m\hat{\sigma}_r^2(k)+\hat{v}_i^2\left(\frac{1-c_i}{c_i^2}\right)-\hat{v}_i^2-2\hat{v}_i^2\left(\frac{1-c_i}{c_i}\right)-\hat{v}_i^2\left(\frac{1-c_i}{c_i}\right)^2. \end{aligned}$$

Therefore, we have

$$\hat{\sigma}_m^2=\hat{\sigma}_r^2(k)\left[1-m^{-1}t_i^2\left(1+\frac{\hat{\boldsymbol{\beta}}_r'(k)\Lambda\hat{\boldsymbol{\beta}}_r(k)}{\hat{\sigma}_r^2(k)}\right)\right],$$ where

$t_i=\dfrac{\hat{v}_i}{\hat{\sigma}_v\sqrt{c_i}}$ is the $i$th studentized residual of the model and $\hat{\sigma}_v^2=\hat{\sigma}_r^2(k)+\hat{\boldsymbol{\beta}}_r'(k)\Lambda\hat{\boldsymbol{\beta}}_r(k)$ is an estimate of the variance of $v_i=y_i-\boldsymbol{x}_i'\boldsymbol{\beta}$.

The score test statistic for the $i$th observation ($SC_i$) based on the corrected observed information matrix $J(\beta,\gamma)$ of the MSOM, for testing $H_0:\ \gamma=0$ versus $H_1:\ \gamma\neq 0$, is given by

$$(14)\quad SC_i=\left[\frac{\partial l_m^*(\boldsymbol{\beta},\sigma^2,\gamma,k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r})}{\partial\gamma}\right]^2 J^{\gamma\gamma},$$

where $J^{\gamma\gamma}$ is the lower right corner of $J^{-1}(\beta,\gamma)$. Substituting $\gamma=0$, $\beta=\hat{\beta}_r(k)$ and $\sigma^2=\hat{\sigma}_r^2(k)$ into the (14), we have

$$\frac{\partial l_m^*(\boldsymbol{\beta},\sigma^2,\gamma,k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r})}{\partial\gamma}=\frac{1}{\hat{\sigma}_r^2(k)}\left[\boldsymbol{y}_i-\boldsymbol{x}'\hat{\boldsymbol{\beta}}_r(k)\right]'$$

$$J(\beta,\gamma)=\frac{1}{\hat{\sigma}_r^2(k)}\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_P & \boldsymbol{x}_i \\ \boldsymbol{x}_i' & 1 \end{bmatrix}$$

and $J^{\gamma\gamma}=\dfrac{\hat{\sigma}_r^2(k)}{\mathrm{c}_i}$. Then under $H_0:\gamma=0$

$$SC_i=\frac{1}{\hat{\sigma}_r^2(k)\mathrm{c}_i}\left[\boldsymbol{y}_i-\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r(k)\right]^2=t_i^2\frac{\hat{\sigma}_v^2}{\hat{\sigma}_r^2(k)}.$$

We can also write $SC_i=t_i^2\left[1+\dfrac{\hat{\boldsymbol{\beta}}_r'(k)\Lambda\hat{\boldsymbol{\beta}}_r(k)}{\hat{\sigma}_r^2(k)}\right]$.

Therefore, the score test statistic is a multiple of the square of studentized residual of the model that is an adequate diagnostic statistic as often used in linear regression diagnostics. If $H_0$ is rejected, then the $i$th case may not come from the original model and so is an outlier.

***Case deletion model***

To quantify the effects of deleting the $i$th observation on $\hat{\beta}_r(k)$ and $\hat{\sigma}_r^2(k)$, a fundamental approach is called CDM with the $i$th observation deleted and can be represented as

$$\begin{aligned} & \boldsymbol{y}_{(i)}=\boldsymbol{Z}_{(i)}\boldsymbol{\beta}+\boldsymbol{\varepsilon}_{(i)}, \text{ where} \\ (15)\quad & \boldsymbol{X}_{(i)}=\boldsymbol{Z}_{(i)}+\Delta_{(i)},\quad i=1,\ldots,n,\quad \text{subject to} \\ & \boldsymbol{r}=\boldsymbol{R}\boldsymbol{\beta}+\boldsymbol{e}, \text{ with} \\ & \boldsymbol{0}=\sqrt{k}\,\boldsymbol{\beta}+\boldsymbol{\varphi}. \end{aligned}$$

The corrected log-likelihood for model (15) is given by

$$\begin{aligned} l_c^*(\boldsymbol{\beta},\sigma^2,k,\boldsymbol{X}_{(i)},\boldsymbol{y}_{(i)},\boldsymbol{r}) &= -\frac{m-1}{2}\log(2\pi\sigma^2)-\frac{1}{2}\log|\boldsymbol{W}|-\frac{1}{2\sigma^2}\big[(\boldsymbol{y}_{(i)}-\boldsymbol{X}_{(i)}\boldsymbol{\beta})'(\boldsymbol{y}_{(i)}-\boldsymbol{X}_{(i)}\boldsymbol{\beta}) \\ &\quad -(n-1)\boldsymbol{\beta}'\Lambda\boldsymbol{\beta}+(\boldsymbol{r}-\boldsymbol{R\beta})'\boldsymbol{W}^{-1}(\boldsymbol{r}-\boldsymbol{R\beta})+k\boldsymbol{\beta}'\boldsymbol{\beta}\big].\end{aligned}$$

The estimate of $\boldsymbol{\beta}$ will be obtained with differentiating of $l_c^*$ with respect to $\boldsymbol{\beta}$ and is denoted by $\hat{\boldsymbol{\beta}}_{(i)}(k)$. Then we have

(16)

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{(i)}(k) &= \left[\boldsymbol{X}'\boldsymbol{X}-\boldsymbol{x}_i\boldsymbol{x}_i'+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-(n-1)\Lambda+k\boldsymbol{I}_p\right]^{-1}(\boldsymbol{X}'\boldsymbol{y}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{r}-\boldsymbol{x}_i\boldsymbol{y}_i) \\ &= (\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p-\boldsymbol{x}_i\boldsymbol{x}_i')^{-1}(\boldsymbol{X}'\boldsymbol{y}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{r}-\boldsymbol{x}_i\boldsymbol{y}_i)+\boldsymbol{O}_p(n^{-1}).\end{aligned}$$

However, using matrix results (see, Rao et al. [8], Theorem A.18) we have

$$\begin{aligned}(\boldsymbol{X}'\boldsymbol{X}&+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p-\boldsymbol{x}_i\boldsymbol{x}_i')^{-1} \\ &= (\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)^{-1}+(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+\boldsymbol{I}_p)^{-1}\boldsymbol{x}_i \\ &\quad\cdot\left[1-\boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)^{-1}\boldsymbol{x}_i\right]^{-1}\boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)^{-1}.\end{aligned}$$

With substituting the above expression in $\hat{\boldsymbol{\beta}}_{(i)}(k)$, we obtain

$$\hat{\boldsymbol{\beta}}_{(i)}(k)=\hat{\boldsymbol{\beta}}_r(k)-(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)^{-1}\boldsymbol{x}_i\frac{\hat{v}_i}{c_i}+\boldsymbol{O}_p(n^{-1}).$$

Taking the differential of $l_c^*$ with respect to $\sigma^2$, we have

$$\begin{aligned}(m-1)\hat{\sigma}^2_{(i)}(k) &= \boldsymbol{y}'_{(i)}\boldsymbol{y}_{(i)}-\hat{\boldsymbol{\beta}}'_{(i)}(k)\boldsymbol{X}'_{(i)}\boldsymbol{y}_{(i)}+\boldsymbol{r}'\boldsymbol{W}^{-1}\boldsymbol{r}-\hat{\boldsymbol{\beta}}'_{(i)}(k)\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{r} \\ &= m\hat{\sigma}_r^2(k)-\frac{\hat{v}_i^2}{c_i}+O_p(1)\end{aligned}$$

or

$$\hat{\sigma}^2_{(i)}(k)=(m-1)^{-1}\hat{\sigma}_r^2(k)\left\{m-t_i^2\left[1+\frac{\hat{\boldsymbol{\beta}}_r'(k)\Lambda\hat{\boldsymbol{\beta}}_r(k)}{\hat{\sigma}_r^2(k)}\right]\right\}+O_p(m^{-1}).$$

***Analogous of generalized Cook's distance***

As a measure of influence, an appropriate measure would be Cook's distance [34]. Cook's distance is used by statisticians to detect influential observations in the data set. It is based on the difference between two estimators, one includes the *i*th observation in the data set; the other excludes the *i*th observation. The Cook's distance statistic for model (15) can be defined analogously by $CD_i(\boldsymbol{\beta})=\left[\hat{\boldsymbol{\beta}}_r(k)-\hat{\boldsymbol{\beta}}_{(i)}(k)\right]'\boldsymbol{M}\left[\hat{\boldsymbol{\beta}}_r(k)-\hat{\boldsymbol{\beta}}_{(i)}(k)\right]$.

Where, $\boldsymbol{M}=\hat{I}^*(\boldsymbol{\beta})=\frac{1}{\hat{\sigma}_r^2(k)}(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)$ and $\hat{I}^*(\boldsymbol{\beta})$ is an estimate of the corrected observed information matrix $I^*(\boldsymbol{\beta})=-\frac{\partial l^*(\boldsymbol{\beta},\sigma^2,k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}$ for $\boldsymbol{\beta}$. Then we have

$$CD_i(\boldsymbol{\beta})=\frac{\left[\hat{\boldsymbol{\beta}}_r(k)-\hat{\boldsymbol{\beta}}_{(i)}(k)\right]'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)\left[\hat{\boldsymbol{\beta}}_r(k)-\hat{\boldsymbol{\beta}}_{(i)}(k)\right]}{\hat{\sigma}_r^2(k)}.$$

We can get

$$CD_i(\boldsymbol{\beta})=\left[\frac{\boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)\boldsymbol{x}_i}{\hat{\sigma}_r^2(k)}\right]\frac{\hat{v}_i^2}{c_i^2}+O_p(n^{-1})=\frac{(1-c_i)\hat{v}_i^2}{\hat{\sigma}_r^2(k)c_i^2}+O_p(n^{-1})$$

Substituting $SC_i=\frac{\hat{v}_i^2}{\hat{\sigma}_r^2(k)c_i}$ into the above expression, we have

$$CD_i(\boldsymbol{\beta})=\frac{(1-c_i)SC_i}{c_i}+O_p(n^{-1})=\frac{1-c_i}{c_i}t_i^2\left[1+\frac{\hat{\boldsymbol{\beta}}_r'(k)\Lambda\hat{\boldsymbol{\beta}}_r(k)}{\hat{\sigma}_r^2(k)}\right]+O_p(n^{-1})$$

Cases for which $CD_i(\boldsymbol{\beta})$'s are large have substantial influence on both the estimates of $\boldsymbol{\beta}$, and deletion of them may result in important changes in conclusions.

***Likelihood distance***

The likelihood distance is a popular measure to assess the influence of the *i*th observation on corrected score estimate. We consider the corrected log-likelihood evaluated at $\left[\hat{\boldsymbol{\beta}}_r(k),\hat{\sigma}_r^2(k)\right]$ and $\left[\hat{\boldsymbol{\beta}}_{(i)}(k),\hat{\sigma}_r^2(k)\right]$, then a measure of the influence of the *i*th observation on $\hat{\boldsymbol{\beta}}$ can be defined as

$$LD_i(\boldsymbol{\beta})=2\left\{l^*\left[\hat{\boldsymbol{\beta}}_r(k),\hat{\sigma}_r^2(k),k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r}\right]-l^*\left[\hat{\boldsymbol{\beta}}_{(i)}(k),\hat{\sigma}_r^2(k),k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r}\right]\right\}$$

Taylor expansion of $l^*\left[\hat{\boldsymbol{\beta}}_{(i)}(k),\hat{\sigma}_r^2(k),k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r}\right]$ at $\hat{\boldsymbol{\beta}}_r(k)$ gives

$$\begin{aligned} l^*&\left[\hat{\boldsymbol{\beta}}_{(i)}(k),\hat{\sigma}_r^2(k),k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r}\right]= \\ l^*&\left[\hat{\boldsymbol{\beta}}_r(k),\hat{\sigma}_r^2(k),k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r}\right]+\left[\frac{\partial l^*(\boldsymbol{\beta},\sigma^2,k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r})}{\partial\boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_r(k),\sigma^2=\hat{\sigma}_r^2(k)}\right]'\left[\hat{\boldsymbol{\beta}}_{(i)}(k)-\hat{\boldsymbol{\beta}}_r(k)\right] \\ &+\frac{1}{2}\left[\hat{\boldsymbol{\beta}}_{(i)}(k)-\hat{\boldsymbol{\beta}}_r(k)\right]'\left[\frac{\partial^2 l^*(\boldsymbol{\beta},\sigma^2,k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_r(k),\sigma^2=\hat{\sigma}_r^2(k)}\right]\left[\hat{\boldsymbol{\beta}}_{(i)}(k)-\hat{\boldsymbol{\beta}}_r(k)\right].\end{aligned}$$

We have

$$\left[\frac{\partial l^*(\boldsymbol{\beta},\sigma^2,k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r})}{\partial\boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_r(k),\sigma^2=\hat{\sigma}_r^2(k)}\right]=\boldsymbol{0}$$

$$\left[-\frac{\partial^2 l^*(\boldsymbol{\beta},\sigma^2,k,\boldsymbol{X},\boldsymbol{y},\boldsymbol{r})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_r(k),\sigma^2=\hat{\sigma}_r^2(k)}\right]=\frac{\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p}{\hat{\sigma}_r^2(k)}$$

and so

$$LD_i(\beta)=\hat{\sigma}_r^{-2}(k)\left[\hat{\boldsymbol{\beta}}_r(k)-\hat{\boldsymbol{\beta}}_{(i)}(k)\right]'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R}-n\Lambda+k\boldsymbol{I}_p)\left[\hat{\boldsymbol{\beta}}_r(k)-\hat{\boldsymbol{\beta}}_{(i)}(k)\right].$$

As seen, we have $LD_i(\boldsymbol{\beta})=CD_i(\boldsymbol{\beta})$.

### *Empirical distribution*

The following simulation procedure (see, Lin et al. [35] and Rebai et al. [36]) will be used to obtain the empirical distributions of the $SC_i$ and $CD_i$ test statistics under the hypothesis of no outliers and influential observations exist:

Step1: We fit the model (1) with stochastic linear restrictions (2) to the data and estimate ridge parameters. A mixed ridge estimate of $\boldsymbol{Z}$ can be derived as, $\hat{\boldsymbol{Z}}_r(k)=\boldsymbol{X}+\hat{\sigma}_v^{-2}\hat{v}\hat{\boldsymbol{\beta}}_r'(k)\Lambda$ (see, Rasekh [30]).

Step 2a: Generate a new data vector

$$\boldsymbol{y}^*=\hat{\boldsymbol{Z}}_r(k)\hat{\boldsymbol{\beta}}_r(k)+\boldsymbol{\varepsilon}^*,$$
$$\hat{\boldsymbol{X}}=\hat{\boldsymbol{Z}}_r(k)+\Delta,$$

where $\boldsymbol{\varepsilon}^*$ is randomly generated as $N\left[\boldsymbol{0},\hat{\sigma}_r^2(k)\boldsymbol{I}_n\right]$ and $\Delta$ is randomly generated as $N(\boldsymbol{0},\boldsymbol{I}_n\otimes\Lambda)$.

Step 2b: Generate a new data vector $\boldsymbol{r}^*=\boldsymbol{R}\hat{\boldsymbol{\beta}}_r(k)+\boldsymbol{e}^*$,

where $\boldsymbol{e}^*$ is randomly generated as $N\left[\boldsymbol{0},\hat{\sigma}_r^2(k)\boldsymbol{I}_q\right]$ and $\boldsymbol{R}$ is a known matrix.

Step 3: Compute the test statistics $SC_i$ and $CD_i$ for $i=1,2,\ldots,n$ and save the order statistics of the set $(SC_i$ and $CD_i:1,2,\ldots,n)$.

Step 4: Repeat steps 2 and 3, $M$ times, for $M$ reasonably large. This generates an empirical distribution of size $M$ for each order statistic.

Step 5: Calculate the $100(1-\alpha)$ percentile for each order statistic for the level $\alpha$.

The percentile of the *i*th order statistic can be considered as a threshold for the *i*th largest value of the test statistics from the original data. If the $i$ largest values of the test statistic $SC_i$ (or $CD_i$) from the original data all exceed their respective thresholds, then it is concluded that these data are all outliers (or influential) observations.

### *Simulation Study*

In this section, we conducted a parametric bootstrap simulation to present the performance of the score test statistic in terms of type I error and power of test. We generated the *j*th set of simulated data as

$$(17)\quad \begin{aligned}\boldsymbol{y}_j&=\boldsymbol{Z}\boldsymbol{\beta}+\boldsymbol{\varepsilon}_j, \qquad j=1,\ldots,1000,\\ \boldsymbol{r}_j&=\boldsymbol{R}\boldsymbol{\beta}+\boldsymbol{e}_j,\end{aligned}$$

where, $\boldsymbol{y}_j=(y_{1j},\ldots,y_{nj})'$, $\boldsymbol{Z}=(\boldsymbol{z}^{(1)},\boldsymbol{z}^{(2)},\boldsymbol{z}^{(3)})$ and $\boldsymbol{z}^{(s)}=(z_{1s},\ldots,z_{ns})'$, $s=1,2,3$, $\boldsymbol{\varepsilon}_j$ is rewritten in accordance with $\boldsymbol{y}_j$. Furthermore, $\boldsymbol{R}$ is a known matrix, $\boldsymbol{r}_j=(r_{1j},r_{2j})'$ and $\boldsymbol{e}_j$ is rewritten in accordance with $\boldsymbol{r}_j$. We consider the following combinations for simulation: $n$=50 or 100,

$\boldsymbol{\beta}=(\beta_1,\beta_2,\beta_3)'=(2.5,2,0.8)'$ or

$(\beta_1,\beta_2,\beta_3)'=(3.5,2.75,0.5)'$, $z^{(1)}\sim U(10,100)$,

$z^{(2)}\sim U(10,100)$, $z^{(3)}=z^{(2)}+\upsilon$,

$\varepsilon_{ij}\sim N(0,\sigma^2)$, $e_{ij}\sim N(0,\sigma^2)$, $\upsilon\sim N(0,\sigma_1^2)$

for $i=1,\ldots,n$, $\sigma^2=0.09$ or $0.25$, $\sigma_1^2=1$,

$\Lambda=\text{diag}(0.1,0.1,0.1)$ or

$\Lambda=\text{diag}(0.15,0.15,0.15)$ and $R=\begin{bmatrix}1&1&3\\2&1&1\end{bmatrix}$.

The simulation study was carried out using the R software (The R codes are available from the second author upon request). For each simulated data set, we derived $\hat{k}$, $\hat{\boldsymbol{\beta}}_r(k)$, $\hat{\sigma}_r^2(k)$ and the score test statistic for the first observation. The choice of the first observation was arbitrary. To generate an empirical distribution of the test statistic under the null hypothesis, the data sets for $h=1,\ldots,2500$ were simulated as

$$\boldsymbol{y}_{jh}^*=\hat{\boldsymbol{Z}}_{rj}(k)\hat{\boldsymbol{\beta}}_{rj}(k)+\boldsymbol{\varepsilon}_{jh}^*,$$
$$\boldsymbol{r}_{jh}^*=\boldsymbol{R}\hat{\boldsymbol{\beta}}_{rj}(k)+\boldsymbol{e}_{jh}^*,$$

where $\boldsymbol{\varepsilon}_{jh}^*$ and $\boldsymbol{e}_{jh}^*$ have normal distribution with zero mean and variances $\hat{\sigma}_{rj}^2(k)\boldsymbol{I}_n$ and $\hat{\sigma}_{rj}^2(k)\boldsymbol{I}_q$, respectively. Also $\hat{\boldsymbol{\beta}}_{rj}(k)$, $\hat{\boldsymbol{Z}}_{rj}(k)$ and $\hat{\sigma}_{rj}^2(k)$ are the corrected mixed ridge estimates of $\boldsymbol{\beta}$, $\boldsymbol{Z}$ and $\sigma^2$ from model (17). The score test statistic was performed for the first observation of each simulated data and $100(1-\alpha)$ percentile from the empirical distribution of test statistic was used as threshold value of the test

**Table 1**. Type I error ( $\alpha = 0.1$ ) of score test statistic for a mean shift model with different combination of parameters $\beta$ , $\sigma^2$ and $\Lambda$ .

| $n$ | $\beta$ | $\sigma^2$ | Score test | |
|---|---|---|---|---|
| | | | $\Lambda = diag\,(0.1, 0.1, 0.1)$ | $\Lambda = diag\,(0.15, 0.15, 0.15)$ |
| 50 | (2.5,2,0.8) | 0.09 | 0.120 | 0.093 |
| | | 0.25 | 0.089 | 0.083 |
| | (3.5,2.75,0.5) | 0.09 | 0.113 | 0.130 |
| | | 0.25 | 0.106 | 0.096 |
| 100 | (2.5,2,0.8) | 0.09 | 0.101 | 0.121 |
| | | 0.25 | 0.076 | 0.081 |
| | (3.5,2.75,0.5) | 0.09 | 0.106 | 0.131 |
| | | 0.25 | 0.091 | 0.100 |

statistic of the model (17). The probability of a type I error estimate for the score test statistic and $\alpha = 0.1$ was calculated as the number of data sets for which the score test statistic exceeded the $100(1-\alpha)$ percentile of the empirical distribution, divided by the number of replicates (Table 1). It appears that in general the type I error of score test statistic for different combinations of parameters are close to the nominal value of $0.1$ . We found no substantial difference between result of both values of $\Lambda$ .

In order to evaluate the relative sensitivity of the score test statistic, we introduce the shift values 1, 3, and 5 for the first observation and again for each combination of parameters, 1000 data sets are generated from the following model

$$\begin{aligned} \boldsymbol{y}_j &= \boldsymbol{Z}\boldsymbol{\beta} + \gamma \boldsymbol{d} + \boldsymbol{\varepsilon}_j, \\ \boldsymbol{r}_j &= \boldsymbol{R}\boldsymbol{\beta} + \boldsymbol{e}_j, \end{aligned} \qquad j = 1, \ldots, 1000,$$

for $\gamma = 1,\ 3$ or 5 where $\boldsymbol{d}$ is an $n \times 1$ vector with value 1 in the first element and zero elsewhere.

Again, for each simulated data set, we derive the ridge estimate of parameters and the score test statistic for the first observation. The power of the score test statistic was calculated as the number of data sets for which the score test statistic exceeded the $100(1-\alpha)$ percentile of the empirical distribution, divided by the number of replicates. The results are presented in Table 2.

A review on the results of this table shows that with increase of the displacement, $\gamma$ , the power of the score test statistic, increases in general. Moreover, we can see that power of the test also increase as sample size increases, while with the increase of $\Lambda$ , the power of the test will decreases slitghtly.

***Example: Egyptian pottery data***

Diagnostic measures developed in the previous sections are applied to a real data set, which is known as the Egyptian pottery data. Briefly, this data set arises from an extensive archaeological survey of pottery production and distribution in the ancient Egyptian city of Al-Amarna. The data consist of measurements of chemical contents (mineral elements) made on many samples of pottery using two different techniques, NAA and ICP (see Smith et al, [37] for description of techniques). The set of pottery was collected from different locations around the city. In general, two type of clay were used to make the ancient Egyptian pottery- Silt and Marl. In addition, archaeologists have classified some sherds as imports from North African countries.

The group structure among the objects arises from two main sources, fabric code and location of pottery. Both of these subdivisions are important to the archaeologists. Consequently, according to this group structure, the selected vessels have been divided into 27 groups and one group of imported vessels is selected as stochastic linear restrictions. In each group, there are different numbers of vessels from the same fabric code and provenance, which can essentially be regarded as replicated observations. Among all mineral elements, our interest is in the relation between Na measured with NAA versus mineral elements Na, Al, K, V, Cr and Mn measured with ICP. The data set is available from the second author upon request.

Rasekh [38] analyzed this data set and fitted a functional measurement error model. In addition, Rasekh [22] considered the same data set and realized that there is collinearity among the explanatory

**Table 2.** The power of score test statistic for a mean shift model with different combination of parameters $\beta$, $\sigma^2$, $\gamma$ and $\Lambda$.

| $n$ | $\beta$ | $\sigma^2$ | $\gamma$ | Power of Score test | |
|---|---|---|---|---|---|
| | | | | $\Lambda = diag(0.1, 0.1, 0.1)$ | $\Lambda = diag(0.15, 0.15, 0.15)$ |
| 50 | (2.5,2,0.8) | 0.09 | 1 | 0.108 | 0.089 |
| | | | 3 | 0.266 | 0.184 |
| | | | 5 | 0.761 | 0.479 |
| | | 0.25 | 1 | 0.084 | 0.087 |
| | | | 3 | 0.284 | 0.174 |
| | | | 5 | 0.808 | 0.489 |
| | (3.5,2.75,0.5) | 0.09 | 1 | 0.103 | 0.121 |
| | | | 3 | 0.164 | 0.155 |
| | | | 5 | 0.376 | 0.276 |
| | | 0.25 | 1 | 0.081 | 0.100 |
| | | | 3 | 0.173 | 0.114 |
| | | | 5 | 0.445 | 0.254 |
| 100 | (2.5,2,0.8) | 0.09 | 1 | 0.145 | 0.141 |
| | | | 3 | 0.625 | 0.430 |
| | | | 5 | 0.993 | 0.915 |
| | | 0.25 | 1 | 0.130 | 0.106 |
| | | | 3 | 0.590 | 0.407 |
| | | | 5 | 0.977 | 0.898 |
| | (3.5,2.75,0.5) | 0.09 | 1 | 0.128 | 0.131 |
| | | | 3 | 0.404 | 0.277 |
| | | | 5 | 0.855 | 0.650 |
| | | 0.25 | 1 | 0.112 | 0.110 |
| | | | 3 | 0.366 | 0.250 |
| | | | 5 | 0.800 | 0.614 |

variables. He fitted a ridge measurement error model and studied the local influence of minor perturbation on the ridge estimate in the measurement error model.

In this section we analysed this data set using the ridge measurement error model with stochastic linear restrictions given in (3). Also, the score test statistic and Cook's distance were calculated for each group under model (3) and 10000 simulated data sets were generated from the fitted model under the null hypothesis of no outliers and influential groups exist. In each simulation, model (3) was fitted for each group and the test statistics were sorted and used to generate the empirical distribution of the order statistics for each test.

Figures 1 give plots of the test statistics from the original data and 95th percentile from the empirical distribution of the first, second and third largest values for each test statistic. Figure 1(a) shows that the score test statistic for group 14 is larger than the 95th percentile of the distribution of the corresponding order statistic. So we conclude that this data set contains only one outlier. On the other hand, Figure 1(b) shows that the Cook's distance for group 14 is larger than the 95th percentile of the distribution of the first order statistic. Therefore, group 14 has also more influence on the estimate of $\beta$.

The corrected score estimates of the full data set and with only group 14 deleted, are given in Table 3. As seen, after deleting group 14, the effect of Al and Cr

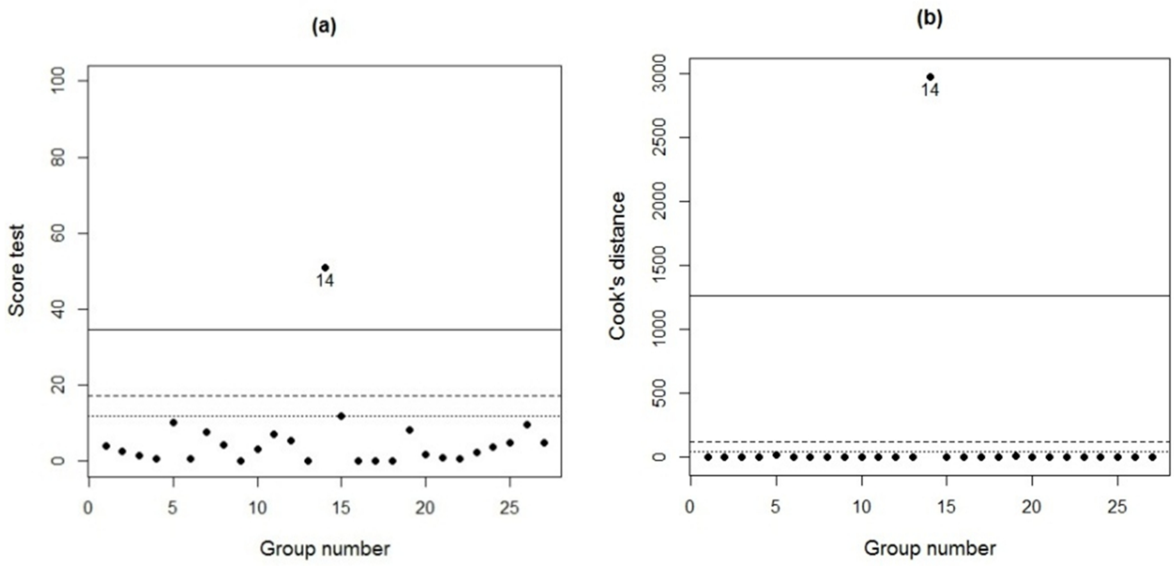


**Figure 1**. Score test statistic for each group (a) and Cook's distance for each group (b), with 95th percentile of the empirical distribution under $H_0$ shown for the 1st (solid line), 2nd (dashed line) and 3rd (dotted line) order statistics for each test.

**Table 3.** Corrected score estimates for the Egyptian pottery data.

| Variable | Full data | Group 14 deleted |
|---|---|---|
| Na | 0.07167761 | 0.53244249 |
| Al | -1.34278991 | -2.30027994 |
| K | 4.08898127 | 6.97760931 |
| V | 0.08152407 | 0.08922238 |
| Cr | -0.09668435 | -0.11011112 |
| Mn | 44.19227355 | 71.19221040 |
| $\sigma^2$ | 0.1867507 | 0.1828421 |

have been less than before. Also after deleting group 14, the effect of NA, K and Mn have been more than before.

## Discussion

We derived the estimate of parameters in the linear ridge measurement error model with stochastic linear restrictions based on the corrected likelihood of Nakamura [3], and we investigated the performance of the mixed ridge estimators over the mixed estimators of the parameters, by the variance and the MSE matrix criteria. We extended the case deletion and mean shift outlier models of the proposed model. We derived the score test statistic for testing that an observation stands out as a possible outlier. In addition, we derived analogous of Cook's distance and likelihood distance for detecting influential observations of the proposed model. The performance of the score test statistic is studied using a parametric bootstrap simulation. It was found out that with the increase of shift value the power of the score test statistic increase.

## Acknowledgments

The authors would like to thank the editor and anonymous referees for several helpful comments and suggestions, which resulted in a significant improvement in the presentation of this paper.

## References

1. Cheng C.L. and Van Ness J.W. *Statistical regression with reasurement error*. Arnold, London, (1999).
2. Fuller W.A. *Measurement error models*. Wiley, New York, (1987).
3. Nakamura T. Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, **77**: 127–137 (1990).
4. Hoerl A.E. and Kennard R.W. Ridge regression: biased estimation for non-orthogonal problems. *Technometrics*, **12**: 69–82 (1970).
5. Belsley D.A., Kuh E. and Welsch R.E. *Regression diagnostics: identifying data and sourcesof collinearity*. Wiley, New York, (1980).
6. Mason R.L. and Gunst R.F. Outlier-induced collinearities. *Technometrics,* **27**: 401–407 (1985).
7. Belsley D.A. Conditioning diagnostics: Collinearity and Weak Data in Regression. John Wiley & Sons, Inc. New York, (1991).
8. Rao C.R., Toutenburg H., Shalabh and Heumann C. *Linear models and enerslizations*, *Least squares and alternatives*, Springer, Berlin, (2008).

9. Toutenburg H. *Prior information in linear models*, John Wiley and Sons, New York, (1982).
10. Durbin J. A note on regression when there I extraneous information about one of the coefficients. *J. Amer. Stat. Assoc*., **48**: 799–808 (1953).
11. Theil H. and Goldberger A.S. On pure and mixed estimation in econometrics. *International Eeconomic Rreview*, **2**: 65–78 (1961).
12. Theil H. The use of incomplete prior information in regression analysis. *J. Amer. Stat. Assoc*., **58**: 401–414 (1963).
13. Sarkar N. A new estimator combining the ridge regression and the restricted least squares methods of estimation. *Comm. Stat. Theory Methods,* **21**: 1987–2000 (1992).
14. Kaciranlar S., Sakallioglus S. and Akdeniz F. Mean squared error comparisons of the modified ridge regression estimator and the restricted ridge regression estimator. *Comm. Stat. Theory Methods*, **27**: 131–138 (1998).
15. Swindel B.F. Good estimators based on prior information. *Comm. Stat. Theory Methods*, **5**: 1065–1075 (1976).
16. Yalian Li. and·Yang Hu. A new stochastic mixed ridge estimator in linear regression model. *Stat Papers*, **51**: 315–323 (2010).
17. He D.J. and Wu Y. A stochastic restricted principal components regression estimator in the linear model. *The Scientific World Journal*, Article ID: 231506 (2014).
18. Wu J.B. and Liu C.L. Performance of some stochastic restricted ridge estimator in linear regression model. *Journal of Applied Mathematics*, Article ID: 508793 (2014).
19. Alkhamisi M.A. and MacNeill I. B. Recent results in ridge regression methods, METRON, (2015).
20. Cook R.D. and Weisberg S. *Residuals and influence in regression.* Chapman & Hall, London, (1982).
21. Wang J. Statistical diagnosis of linear regression model with the random constraints and bayes method. Nanjing University of Science and Technology, Nanjing (2007).
22. Kelly G. The influence function in the errors in variables problem. *Ann. Stat*., **12**: 87–100 (1984).
23. Wellman J.M. and Gunst R.F. Influence diagnostics for linear measurement error models. *Biometrika*, **78**: 373–380 (1991).
24. Zhong X.P., Wei B.C. and Fung W.K. Influence analysis for linear measurement error models. *Ann. Inst. Math. Stat*., **52**: 367–379 (2000).
25. Zare K. and Rasekh A. Diagnostic measures for linear mixed measurement error models, *Sort.*, **35**: 125–144 (2011).
26. Babadi B., Rasekh A., Rasekhi A.A., Zare K. and Zadkarami M.R. Variance shift model for detection of outliers in the linear measurement error model. Hindawi Publishing Corporation, Article ID 396875, 9 pageshttp://dx.doi.org/10.1155/2014/396875 (2014).
27. Walker E. and Birch J.B. Influence measures in ridge regression. *Technometrics*, **30**: 221–227 (1988).
28. Rasekh A.R. and Mohtashami G. Assessing influence on ridge estimates in functional measurement error models. Pub. Inst. Stat. Univ. Paris. LI. Fasc. **1**–**2**: 97–109 (2004).
29. Rasekh A.R. and Fieller N.R.J. Influence functions in functional measurement error models with replicated data. *Statistics*, **37**: 169–178 (2003).
30. Rasekh A.R. Local influence in measurement error models with ridge estimate, *Comput. Stat. Data Anal.* **50**: 2822–2834 (2006).
31. Fung W.K., Zhong X.P. and Wei B.C. On estimation and influence diagnostics in linear mixed measurement error models. *American Journal of Mathematical and Management Science*, **23**: 37–59 (2003).
32. Ozkale M.R. A stochastic restricted ridge regression estimator. *J. Multivar Anal*., **100**: 1706–1716 (2009).
33. Farebrother R.W. Further results on the mean square error of ridge regression. *J. Roy. Stat. Soc.* Ser. B **38**: 248–250 (1976).
34. Cook R.D. Detection of influential observationin linear regression. *Technometrics*, **19**: 15–18 (1977).
35. Lin D.Y., Wei L.J. and Ying Z. Checking the Cox model using cumulative sums of martingale-based residuals. *Biometrika*, **80**: 557–572 (1993).
36. Rebai A., Goffinet B. and Mangin B. Approximate thresholds of interval mapping tests for QTL detection. *Genetics* **138**: 235–240 (1994).
37. Smith N.D.M., Hart F.A., Symond R.D. and Walsh J.N. Analysis of Roman pottery from Colchester by Industively Coupled Plasma Spectrometry. In: Slatet, E. A., Tate, J.O.(Eds.), Science and Archaeology Glasgow ,vol. **196(I)**. B.A.R.: 41–55 (1987).
38. Rasekh A.R. Ridge estimation in functional measurement error models, *Annales DE L'I.S.U.P., Pub. De L' Institut de Statistique de L' Universite de Paris,* XXXXV, Fascicule **2**–**3**: 47–59 (2001).