Journal of Mathematical Extension
Vol. 15, No. 4, (2021) (3)1-23
URL: https://doi.org/10.30495/JME.2021.1581
ISSN: 1735-8299
Original Research Paper

Performance of Ridge Regression Approach in Linear Measurement Error Models with Replicated Data

A. R. Ziaei
Marvdasht Branch, Islamic Azad University

K. Zare*
Marvdasht Branch, Islamic Azad University

A. Sheikhi
Shahid Bahonar University

Abstract. It is well known that bias in parameter estimates arises when there are measurement errors in the covariates of regression models. One solution for decreasing such biases is the use of prior information concerning the measurement error, which is often called replication data. In this paper, we present a ridge estimator in replicated measurement error (RMER) to overcome the multicollinearity problem in such models. The performance of RMER against some other estimators is investigated. Large sample properties of our estimator are derived and compared with other estimators using a simulation study as well as a real data set.

AMS Subject Classification: 62J05; 62J07
Keywords and Phrases: Measurement error model, Ridge regression, Multicollinearity, Corrected log-likelihood

Received: March 2020; Accepted: December 2020
*Corresponding Author

1 Introduction

One of the fundamental assumptions in regression analysis is that all the observations are correctly observed. However, in many applications this assumption is violated and the data are contaminated by measurement errors. In these models, estimation based on the above mentioned assumption leads to inconsistent estimates, meaning that the parameter estimates do not tend to the true values even in very large samples. In fact, measurement error is known to cause biased parameter estimates (Carroll et al. [6]) and lack of efficiency is a direct consequence of this misestimates. Ignoring non-negligible measurement error often leads to incorrect inferences about parameters (see for example, Cook and Campbel [8]). When the explanatory variables cannot be measured truly, some additional information is required to obtain consistent estimators of the regression coefficients. The literature introduces several approaches for finding consistent estimators. One such approach is to consider the replicated measurement error model and there are some works on this approach. Devanarayan and stefanski [10] and Nawarathna and Choudhary [21] studied a heteroscedastic measurement error models with replicate measurements. Singh et al. [28] investigated a replicated measurement error model under exact linear restrictions. Wimmer and Witkovsky [33] explored a measurement error model with replicated data in comparative calibration problem. See also Gimenez and Patat [14] and Blas et al. [2]. They did some study on Berkson measurement errors for replicated data. Dalen et al. [9] used replication to correct misclassification of a categorized exposure in binary regression. Also, Chan and Mak [7] and Isogawa [17] investigated the structural form of the replicated measurement error model under the condition of normally distributed measurement errors. Cao et al. [5] studied the multivariate measurement error models for replicated data under heavy-tailed distributions, as well. Recently, Shalabh et al.[26] did a research on the inconsistent estimator of parameter in ultrastructural measurement error model with replicated data, see also Ullah et al. [32] and Shalabh et al. [27] for non-normal measurement errors in such a case.
Replicated data contaminated with measurement errors are frequently presented in medical, economic, environmental, chemical and other fields.

Estimation in measurement error models with replicated observations, by the corrected score log-likelihood approach has been introduced by Nakamura [20]. This approach is based on the corrected log-likelihood function, which, when feasible, yields consistent estimators for the model parameters. In this paper we use the Nakamura's corrected score log-likelihood approach to estimate parameters. One of the most important advantages of this approach is that it directly allows one to find consistent and asymptotically normal estimators for the parameters of interest, voiding the problem of estimating the unobserved quantities or incidental parameters. see for example, Huang [16] and Yang et al. [34] for the recent uses of this method.

Another important assumption in a classical regression analysis is that explanatory variables are uncorrelated. When this assumption is violated, the explanatory variables are nearly dependent which refers as multicollinearity problem and yields poor estimators of interest parameters. In order to resolve this problem several approaches have been considered, among them, the Ridge regression was introduced by Horel and Kennard [15] and considers a shrinkage method to overcome the problem of multicollinearity for the estimation of regression parameters. This approach has been considered in measurement error models. See for example, Saleh and Shalabh [25] and Rasekh [23]. In this article we employ the ridge regression method to combat multicollinearity in the estimation of parameters in measurement error models with replicated data.

The organization of this paper is as follows. In Section 2, we obtain a ridge estimator in replicated measurement error models as well as some other estimators. Some large sample properties and theoretical comparison presented in Section 3 and finally, in Section 4 we present a simulation study and a real numerical application of our results.

2 Model specification and estimation

Consider the following exact relationship between the $n \times 1$ vector of study variable $\boldsymbol{\eta}$ and the $n \times p$ matrix $\boldsymbol{Z}$ of n values on each of the p explanatory variables:

$$\boldsymbol{\eta} = \alpha \boldsymbol{e_n} + \boldsymbol{Z}\boldsymbol{\beta}$$

where α is the intercept term, $\boldsymbol{e}_n$ is the vector of elements unity, and $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression coefficients.
Suppose that the observations on the study and explanatory variables, $\boldsymbol{\eta}$ and $\boldsymbol{Z}$ are contaminated with measurement errors and hence cannot be observed directly, but instead we assume that they are observed with additive measurement errors as

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{\eta} + \boldsymbol{\varepsilon}\\ \boldsymbol{X} &= \boldsymbol{Z} + \boldsymbol{\Delta}\end{aligned}$$

where $\boldsymbol{X}$ is $n \times p$ matrix whose ith row is $\boldsymbol{x}_i^T$; The $n \times p$ matrix of measurement errors $\boldsymbol{\Delta} = (\boldsymbol{\delta}_1,\ \boldsymbol{\delta}_2, \ldots,\ \boldsymbol{\delta}_n)^T$ is associated with $\boldsymbol{Z}$ and assumed to have normally distributed rows, with mean zero and covariance $\boldsymbol{\Lambda}$ and hence $\boldsymbol{\Delta} \sim N(0,\ \boldsymbol{I}_n \otimes \boldsymbol{\Lambda})$, where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix and $\otimes$ denotes the Kronecker product, which, for two arbitrary matrices $\boldsymbol{H} = (h_{ij})$ and $\boldsymbol{S} = (s_{ij})$, of dimensions $a \times b$ and $c \times d$, respectively, defined as

$$\boldsymbol{H} \otimes \boldsymbol{S} = \begin{bmatrix} h_{11}\boldsymbol{S} & h_{12}\boldsymbol{S} & \ldots & h_{1b}\boldsymbol{S} \\ h_{21}\boldsymbol{S} & h_{22}\boldsymbol{S} & \ldots & h_{2b}\boldsymbol{S} \\ \ldots & \ldots & \ldots & \ldots \\ h_{a1\boldsymbol{S}} & h_{a2}\boldsymbol{S} & \ldots & h_{ab}\boldsymbol{S} \end{bmatrix}.$$

Associated with $\boldsymbol{\eta}$, The model errors $\boldsymbol{\varepsilon} = (\varepsilon_1,\ \varepsilon_2, \ldots, \varepsilon_n)^T$ are i.i.d. normally with mean zero and variance σ^2. Furthermore, we assume that $\boldsymbol{\varepsilon}$ and $\boldsymbol{\Delta}$ are independent and $\boldsymbol{\Lambda}$ is a $p \times p$ matrix of known values with non-negative diagonal elements. For the sake of notational simplicity we assume that $\alpha = 0$, then we have the linear measurement error model as

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\\ \boldsymbol{X} &= \boldsymbol{Z} + \boldsymbol{\Delta}\end{aligned} \tag{1}$$

By considering the model (1) without measurement error term (i.e. $\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$), the log-likelihood function of $\boldsymbol{\beta}$ is as

$$\ell(\boldsymbol{\beta}, \sigma^2; \boldsymbol{Z}, \boldsymbol{y}) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\beta})$$

and based on Nakamura [20], the appropriate corrected log-likelihood function is defined as

$$\ell^*(\boldsymbol{\beta},\sigma^2;\boldsymbol{X},\boldsymbol{y}) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\{(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta}) - n\boldsymbol{\beta}^T\boldsymbol{\Lambda\beta}\}$$

such that,

$$E^*(\ell^*(\boldsymbol{\beta},\sigma^2;\boldsymbol{X},\boldsymbol{y})) = \ell(\boldsymbol{\beta},\sigma^2;\boldsymbol{Z},\boldsymbol{y})$$

where E^* is the conditional expectation with respect to $\boldsymbol{X}$ given $\boldsymbol{y}$ and $\boldsymbol{Z}$. The corrected log-likelihood function $\ell^*(\boldsymbol{\beta},\sigma^2;\boldsymbol{X},\boldsymbol{y})$ satisfies:

$$E^*(\frac{\partial}{\partial\boldsymbol{\beta}}\ell^*(\boldsymbol{\beta},\sigma^2;\boldsymbol{X},\boldsymbol{y})) = \frac{\partial}{\partial\boldsymbol{\beta}}\ell(\boldsymbol{\beta},\sigma^2;\boldsymbol{Z},\boldsymbol{y}),$$
$$E^*(\frac{\partial}{\partial\sigma^2}\ell^*(\boldsymbol{\beta},\sigma^2;\boldsymbol{X},\boldsymbol{y})) = \frac{\partial}{\partial\sigma^2}\ell(\boldsymbol{\beta},\sigma^2;\boldsymbol{Z},\boldsymbol{y})$$

and therefore by solving the equations $\frac{\partial}{\partial\boldsymbol{\beta}}\ell^*(\boldsymbol{\beta},\sigma^2;\boldsymbol{X},\boldsymbol{y}) = 0$ and $\frac{\partial}{\partial\boldsymbol{\sigma^2}}\ell^*(\boldsymbol{\beta},\sigma^2;\boldsymbol{X},\boldsymbol{y}) = 0$ the corrected score estimated of β and σ^2 respectively are given by

$$\hat{\boldsymbol{\beta}}_{ME} = \left(\boldsymbol{X}^T\boldsymbol{X} - n\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}$$
$$\hat{\sigma}^2_{ME} = \frac{1}{n}\left((\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta}) - n\boldsymbol{\beta}^T\boldsymbol{\Lambda\beta}\right)$$

where subscript ME stands for Measurement Error. In a multicollinearity problem, the suggested estimator of $\boldsymbol{\beta}$ based on a shrinkage strategy, the corrected log-likelihood function for the model (1) for $0 < k < 1$ can be defined as

$$\ell^*(\boldsymbol{\beta},\sigma^2;\boldsymbol{X},\boldsymbol{y}) =$$
$$-\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\{(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta}) - n\boldsymbol{\beta}^T\boldsymbol{\Lambda\beta} + k\boldsymbol{\beta}^T\boldsymbol{\beta}\}.$$

Differentiating from $\ell^*(\boldsymbol{\beta},\sigma^2;\boldsymbol{X},\boldsymbol{y})$ with respect to $\boldsymbol{\beta}$ and σ^2 yields the Measurement Error Ridge-type (MER) as

$$\hat{\boldsymbol{\beta}}_{MER} = \left(\boldsymbol{X}^T\boldsymbol{X} - n\,\boldsymbol{\Lambda} + k\boldsymbol{I}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y},$$
$$\hat{\sigma}^2_{MER} = \frac{1}{n}\left((\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta}) - n\boldsymbol{\beta}^T\boldsymbol{\Lambda\beta} + k\boldsymbol{\beta}^T\boldsymbol{\beta}\right)$$

which is a ridge estimator of $\boldsymbol{\beta}$ in measurement error model (1).

Now, suppose on unit i there are m replicate values $\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{im}$ of the error-prone measure of $\boldsymbol{x_i}$. Furthermore, we assume the replicate values are independent with $\boldsymbol{x}_{ij} = \boldsymbol{z}_i + \boldsymbol{\delta}_{ij}$, $j = 1, 2, \ldots, m$ and in the matrix form, $\boldsymbol{X}_j = \boldsymbol{Z} + \boldsymbol{\Delta}_j$, $j = 1, 2, \ldots, m$. So, under the normality assumptions we have $\boldsymbol{\Delta}_j \sim N(0,\ \boldsymbol{I}_n \otimes \boldsymbol{\Lambda})$. Using a direct expectation we have

$$E^*(\boldsymbol{X}_j^T \boldsymbol{X}_j) = \boldsymbol{Z}^T \boldsymbol{Z} + n\boldsymbol{\Lambda},\ j = 1, 2, \ldots, m,$$

where E^* is the conditional expectation with respect to $\boldsymbol{X}$ given $\boldsymbol{y}$ and $\boldsymbol{Z}$.

The most important benefit of replicated data in measurement error model is that it helps the researcher to find an unbiased estimate of $\boldsymbol{\Lambda}$ from replicated observations on the independent variables (see Nagelkerke [19], for more details).

Let we define $\overline{\boldsymbol{X}^T \boldsymbol{X}} = \frac{1}{m} \sum_{j=1}^{m} \boldsymbol{X_j}^T \boldsymbol{X_j}$ and $\overline{\boldsymbol{X}^T} \boldsymbol{y} = \frac{1}{m} \sum_{j=1}^{m} \boldsymbol{X_j}^T \boldsymbol{y}$. In this case, the appropriate corrected log-likelihood function is defined as

$$\ell^*(\boldsymbol{\beta}, \sigma^2; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{y}) =$$
$$-\frac{nm}{2} \ln(2\pi\sigma^2) - \frac{1}{2m\sigma^2} \sum_{j=1}^{m} \left(\sum_{i=1}^{n} \{(y_i - \boldsymbol{x}_{ij}^T \boldsymbol{\beta})^2 - \boldsymbol{\beta}^T \boldsymbol{\Lambda} \boldsymbol{\beta}\boldsymbol{\}} \right)$$

such that,

$$E^*(\ell^*(\boldsymbol{\beta}, \sigma^2; \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_m, \boldsymbol{y}) = \ell(\boldsymbol{\beta}, \sigma^2; \boldsymbol{Z}, \boldsymbol{y})$$

and by differentiating from $\ell^*(\boldsymbol{\beta}, \sigma^2; \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_m, \boldsymbol{y})$ with respect to $\boldsymbol{\beta}$ we obtain the following normal equation

$$\frac{1}{m} \sum_{j=1}^{m} \boldsymbol{X}_j^T \boldsymbol{y} - \frac{1}{m} \sum_{j=1}^{m} (\boldsymbol{X}_j^T \boldsymbol{X}_j - n\boldsymbol{\Lambda})\boldsymbol{\beta} = \mathbf{0},$$

which yields the corrected log-likelihood for the replicated measurement error estimator (RME) of $\boldsymbol{\beta}$ as

$$\hat{\boldsymbol{\beta}}_{RME} = \left(\overline{\boldsymbol{X}^T \boldsymbol{X}} - n\boldsymbol{\Lambda} \right)^{-1} \overline{\boldsymbol{X}^T} \boldsymbol{y},$$
$$\hat{\sigma}^2_{RME} = \frac{1}{nm} \sum_{j=1}^{m} \{(\boldsymbol{y} - \boldsymbol{X}_j \boldsymbol{\beta})^T (\boldsymbol{y} - \boldsymbol{X}_j \boldsymbol{\beta}) - n\boldsymbol{\beta}^T \boldsymbol{\Lambda} \boldsymbol{\beta}\}.$$

Now we are ready to introduce an estimator which employs replication data to achieve more accuracy and use the ridge penalty to overcome the multicollinearity, for $j = 1, 2, \ldots, m$ we define

$$\boldsymbol{y}^* = \begin{pmatrix} \boldsymbol{y} \\ 0 \end{pmatrix} \quad , \boldsymbol{x}_j^* = \begin{pmatrix} \boldsymbol{x}_j \\ \sqrt{k}\boldsymbol{I}_p \end{pmatrix} \quad , \boldsymbol{\varepsilon}_{ij}^* = \begin{pmatrix} \boldsymbol{\varepsilon}_{ij} \;\; - \Delta_j \boldsymbol{\beta} \\ \boldsymbol{\varphi} \end{pmatrix}$$

where the $0 < k < 1$ denotes the ridge parameter and $\boldsymbol{\varphi}$ is an error vector with $\mathrm{E}(\boldsymbol{\varphi}) = 0$ and $\mathrm{Var}(\boldsymbol{\varphi}) = \sigma^2 \boldsymbol{I}_p$. Then the appropriate corrected log-likelihood for model (1) is given by

$$\ell^*(\boldsymbol{\beta}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{y}) =$$
$$-\frac{nm}{2}\ln(2\pi\sigma^2) - \frac{1}{2m\sigma^2}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}\{(y_i - \boldsymbol{x}_{ij}^T\boldsymbol{\beta})^2 - \boldsymbol{\beta}^T\boldsymbol{\Lambda}\boldsymbol{\beta} + \boldsymbol{k}\boldsymbol{\beta}^T\boldsymbol{\beta}\}\right)$$

and by differentiating with respect to $\boldsymbol{\beta}$ we obtain the following normal equation

$$\frac{1}{m}\sum_{j=1}^{m}\boldsymbol{X}_j^T\boldsymbol{y} - \frac{1}{m}\sum_{j=1}^{m}(\boldsymbol{X}_j^T\boldsymbol{X}_j - n\boldsymbol{\Lambda} + k\boldsymbol{I})\boldsymbol{\beta} = \boldsymbol{0},$$

and hence the ridge estimator of $\boldsymbol{\beta}$ using the corrected log-likelihood method in a replicated measurement error model (RMER) is as

$$\hat{\boldsymbol{\beta}}_{RMER} = \left(\overline{\boldsymbol{X}^T\boldsymbol{X}} - n\boldsymbol{\Lambda} + k\boldsymbol{I}\right)^{-1}\overline{\boldsymbol{X}^T}\boldsymbol{y}. \tag{2}$$

Note that we can use these replications to consistently estimate the covariance matrix of measurement errors as $\hat{\sigma}^2_{RMER} = \frac{1}{nm}\sum_{j=1}^{m}\{(\boldsymbol{y} - \boldsymbol{X}_j\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}_j\boldsymbol{\beta}) - n\boldsymbol{\beta}^T\boldsymbol{\Lambda}\boldsymbol{\beta} + \boldsymbol{k}\boldsymbol{\beta}^T\boldsymbol{\beta}\}$. Also, in a similar manner we can estimate the variance of measurement errors. If we denote $\bar{\boldsymbol{x}}_{i.} = \frac{1}{m}\sum_{j=1}^{m}\boldsymbol{x}_{ij}$ as the mean of replicate measurements $\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im}$ of $\boldsymbol{z}_i$, the ith row of matrix $\boldsymbol{Z}$, then

$$\widehat{\boldsymbol{\Lambda}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i.})(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i.})^T}{n(m-1)}.$$

See e.g., Carroll et al.[6] and Buonaccorsi [3] for more details. Furthermore as shown in Rasekh and Fieller [24] an estimate of $\boldsymbol{Z}$ can be derived as $\hat{\boldsymbol{Z}} = \bar{\boldsymbol{X}} + m^{-1}\hat{\sigma}_{\nu}^{-2}\hat{\nu}\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\Lambda}}$ where $\hat{\nu} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_{\nu}^2 = n^{-1}\hat{\sigma}^2 + m^{-1}\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{X}}$ components is $\bar{\boldsymbol{x}}_{i.}$.

3 Comparison of estimators

Since the exact distribution and finite sample properties of the RMER estimator is difficult to drive, we propose to use the large sample asymptotic approximation theory to study the asymptotic distribution of the estimators.

Theorem 3.1. $\hat{\boldsymbol{\beta}}_{RMER}$ *has asymptotically normal distribution with mean and variance-covariance matrix respectively* $(\boldsymbol{Z}^T\boldsymbol{Z}+k\boldsymbol{I})^{-1}\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta}$ *and*

$$(\boldsymbol{Z}^T\boldsymbol{Z}+k\boldsymbol{I})^{-1}\left[m^{-1}\boldsymbol{\Lambda}(n\sigma^2+\boldsymbol{\beta}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta})+\sigma^2\boldsymbol{Z}^T\boldsymbol{Z}\right](\boldsymbol{Z}^T\boldsymbol{Z}+k\boldsymbol{I})^{-1}.$$

Proof. Since $E^*(\overline{\boldsymbol{X}^T\boldsymbol{X}})=\boldsymbol{Z}^T\boldsymbol{Z}+n\boldsymbol{\Lambda}$, and by Fung et al.[12] we can write

$$\overline{\boldsymbol{X}^T\boldsymbol{X}}=\boldsymbol{Z}^T\boldsymbol{Z}+n\boldsymbol{\Lambda}+O_p(n^{\frac{1}{2}}).$$

we also obtain

$$n^{-1}(\overline{\boldsymbol{X}^T\boldsymbol{X}}+k\boldsymbol{I})=n^{-1}(\boldsymbol{Z}^T\boldsymbol{Z}+k\boldsymbol{I})+\boldsymbol{\Lambda}+O_p(n^{-\frac{1}{2}})$$

from (2) we have

$$\begin{aligned}\sqrt{n}\hat{\boldsymbol{\beta}}_{RMER}&=\left(n^{-1}(\overline{\boldsymbol{X}^T\boldsymbol{X}}-n\boldsymbol{\Lambda}+k\boldsymbol{I})\right)^{-1}n^{-\frac{1}{2}}\overline{\boldsymbol{X}^T}\boldsymbol{Y}\\&=\left(n^{-1}(\boldsymbol{Z}^T\boldsymbol{Z}+k\boldsymbol{I})+O_p(n^{-\frac{1}{2}})\right)^{-1}n^{-\frac{1}{2}}\overline{\boldsymbol{X}^T}\boldsymbol{Y}\\&=\left(\boldsymbol{I}+O_p(n^{-\frac{1}{2}})\right)^{-1}\left(n^{-1}(\boldsymbol{Z}^T\boldsymbol{Z}+k\boldsymbol{I})\right)^{-1}n^{-\frac{1}{2}}\overline{\boldsymbol{X}^T}\boldsymbol{Y}\\&=\left(\boldsymbol{I}+O_p(n^{-\frac{1}{2}})\right)\left(n^{-1}(\boldsymbol{Z}^T\boldsymbol{Z}+k\boldsymbol{I})\right)^{-1}n^{-\frac{1}{2}}\overline{\boldsymbol{X}^T}\boldsymbol{Y}\end{aligned}$$

where $\{\boldsymbol{I}+O_p(n^{-\frac{1}{2}})\}^{-1}=\boldsymbol{I}+O_p(n^{-\frac{1}{2}})$ is obtained from taylor series expansion and since the limit of $\boldsymbol{C}=n^{-1}(\boldsymbol{Z}^T\boldsymbol{Z}+k\boldsymbol{I})$ exist, hence, $\sqrt{n}\hat{\boldsymbol{\beta}}_{RMER}=\boldsymbol{C}^{-1}\boldsymbol{\xi}+O_p(n^{-\frac{1}{2}})$ where $\boldsymbol{\xi}=n^{-\frac{1}{2}}\overline{\boldsymbol{X}^T}\boldsymbol{Y}$ is asymptotically normal with mean $n^{-\frac{1}{2}}\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta}$ (See for example, Fung et al.[12] and Zare and Rasekh [35]). So, we readily conclude that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{RMER}-(\boldsymbol{Z}^T\boldsymbol{Z}+k\boldsymbol{I})^{-1}\boldsymbol{Z}^T\boldsymbol{Z}\,\boldsymbol{\beta})=\boldsymbol{C}^{-1}(\boldsymbol{\xi}-E(\boldsymbol{\xi}))+O_p(n^{-\frac{1}{2}})\boldsymbol{C}^{-1}\boldsymbol{\xi}.$$

Consequently, $\sqrt{n}(\hat{\boldsymbol{\beta}}_{RMER} - (\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1}\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta})$ has asymptotically normal distribution with zero mean.
Also, from (2) the asymptotic variance (AVar) of RMER will obtains as $\text{AVar}(\sqrt{n}\hat{\boldsymbol{\beta}}_{RMER}) = \boldsymbol{C}^{-1}\text{Var}(\boldsymbol{\xi})\boldsymbol{C}^{-1}$. The variance-covariance matrix of $\boldsymbol{\xi}$ is

$$\begin{aligned}\text{Var}(\boldsymbol{\xi}) &= \text{E}_{\mathbf{y}}(\text{Var}(\boldsymbol{\xi}|\mathbf{y})) + \text{Var}_{\mathbf{y}}(\text{E}(\boldsymbol{\xi}|\mathbf{y}))\\ &= n^{-1}m^{-1}E_{\mathbf{y}}(\mathbf{y}^T\mathbf{y}\boldsymbol{\Lambda}) + n^{-1}Var_{\mathbf{y}}(\boldsymbol{Z}^T\mathbf{y})\\ &= n^{-1}m^{-1}\boldsymbol{\Lambda}(n\sigma^2 + \boldsymbol{\beta}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta}) + n^{-1}\sigma^2\boldsymbol{Z}^T\boldsymbol{Z}\\ &= n^{-1}[m^{-1}\boldsymbol{\Lambda}(n\sigma^2 + \boldsymbol{\beta}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta}) + \sigma^2\boldsymbol{Z}^T\boldsymbol{Z}].\end{aligned}$$

Thus,

$$\begin{aligned}&\text{AVar}(\hat{\boldsymbol{\beta}}_{RMER}) =\\ &(\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1}\left[m^{-1}\boldsymbol{\Lambda}(n\sigma^2 + \boldsymbol{\beta}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta}) + \sigma^2\boldsymbol{Z}^T\boldsymbol{Z}\right](\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1},\end{aligned}$$

and concequently the desired result is achieved. □

Obviously for suitable choises of k and m in $\text{AVar}(\hat{\boldsymbol{\beta}}_{RMER})$, we have the AVar of other mentioned estimators. when $k = 0$, we obtain $\text{AVar}(\hat{\boldsymbol{\beta}}_{RME})$, when $m = 1$, we obtain $\text{AVar}(\hat{\boldsymbol{\beta}}_{MER})$ and $\text{AVar}(\hat{\boldsymbol{\beta}}_{ME})$ is obtained by choosing $k = 0$ and $m = 1$.
The following corollaries are now the consequent results of the Theorem 3.1.

Corollary 3.2. *$\hat{\boldsymbol{\beta}}_{RME}$ has asymptotically normal distribution with mean $\boldsymbol{\beta}$ and variance-covariance matrix*

$$(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\left[m^{-1}\boldsymbol{\Lambda}(n\sigma^2 + \boldsymbol{\beta}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta}) + \sigma^2\boldsymbol{Z}^T\boldsymbol{Z}\right](\boldsymbol{Z}^T\boldsymbol{Z})^{-1}.$$

Corollary 3.3. *$\hat{\boldsymbol{\beta}}_{MER}$ has asymptotically normal distribution with mean and variance-covariance matrix respectively $(\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1}\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta}$ and $(\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1}\left[\boldsymbol{\Lambda}(n\sigma^2 + \boldsymbol{\beta}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta}) + \sigma^2\boldsymbol{Z}^T\boldsymbol{Z}\right](\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1}$.*

Corollary 3.4. *$\hat{\boldsymbol{\beta}}_{ME}$ has asymptotically normal distribution with mean $\boldsymbol{\beta}$ and variance-covariance matrix*

$$(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\left[\boldsymbol{\Lambda}(n\sigma^2 + \boldsymbol{\beta}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta}) + \sigma^2\boldsymbol{Z}^T\boldsymbol{Z}\right](\boldsymbol{Z}^T\boldsymbol{Z})^{-1}.$$

An extension of univariate mean square error (MSE) in multivariate manner is the mean square error matrix (MSEM) and for an arbitrary estimator, say $\hat{\boldsymbol{\theta}}$, defined as

$$\mathrm{MSEM}(\hat{\boldsymbol{\theta}}) = \mathrm{E}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T = \mathrm{Var}(\hat{\boldsymbol{\theta}}) + \mathrm{B}(\hat{\boldsymbol{\theta}})\mathrm{B}(\hat{\boldsymbol{\theta}})^T,$$

where $\mathrm{B}(\hat{\boldsymbol{\theta}}) = \mathrm{E}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}$, stands for bias vector of $\hat{\boldsymbol{\theta}}$ in estimating $\boldsymbol{\theta}$. Also, in the asymptotic case, we may have the asymptotic MSEM (AMSEM) can be obtained as

$$\mathrm{AMSEM}(\hat{\boldsymbol{\theta}}) = \mathrm{AVar}(\hat{\boldsymbol{\theta}}) + \mathrm{B_A}(\hat{\boldsymbol{\theta}})\mathrm{B_A}(\hat{\boldsymbol{\theta}})^{\mathrm{T}}$$

where $B_A(\hat{\boldsymbol{\theta}})$ is the asymptotic bias vector of $\hat{\boldsymbol{\theta}}$. In order to compare two given estimators we may have the following definition (see for example Ozkale [22]):

Definition 3.5. For two estimators $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$, it is said that $\hat{\boldsymbol{\theta}}_2$ is superior to $\hat{\boldsymbol{\theta}}_1$, with respect to AMSEM sense, if and only if $\Delta_A(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) = \mathrm{AMSEM}(\hat{\boldsymbol{\theta}}_1) - \mathrm{AMSEM}(\hat{\boldsymbol{\theta}}_2)$ is a non-negative definite matrix.

We can readily obtain the AMSEM of estimators $\hat{\boldsymbol{\beta}}_{RMER}$, $\hat{\boldsymbol{\beta}}_{MER}$, $\hat{\boldsymbol{\beta}}_{RME}$ and $\hat{\boldsymbol{\beta}}_{ME}$ as follows:

$$\begin{aligned}
\mathrm{AMSEM}(\hat{\boldsymbol{\beta}}_{\mathrm{RMER}}) &= (\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1}\mathbf{N}_m(\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1} + \mathbf{a}\mathbf{a}^T \\
\mathrm{AMSEM}(\hat{\boldsymbol{\beta}}_{\mathrm{RME}}) &= (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\mathbf{N}_m(\boldsymbol{Z}^T\boldsymbol{Z})^{-1} \\
\mathrm{AMSEM}(\hat{\boldsymbol{\beta}}_{\mathrm{MER}}) &= (\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1}\mathbf{N}_1(\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1} + \mathbf{a}\mathbf{a}^T \\
\mathrm{AMSEM}(\hat{\boldsymbol{\beta}}_{\mathrm{ME}}) &= (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\mathbf{N}_1(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}
\end{aligned}$$

where $\mathbf{N}_m = m^{-1}\boldsymbol{\Lambda}(n\sigma^2 + \boldsymbol{\beta}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta}) + \sigma^2\boldsymbol{Z}^T\boldsymbol{Z}$ and $\mathbf{a} = ((\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1}\boldsymbol{Z}^T\boldsymbol{Z} - \boldsymbol{I})\boldsymbol{\beta}$.
Note that, by using Theorem 3.1 we have

$$\begin{aligned}
&\widehat{\mathrm{AVar}}(\hat{\boldsymbol{\beta}}_{RMER}) = \\
&(\hat{\boldsymbol{Z}}^T\hat{\boldsymbol{Z}} + k\boldsymbol{I})^{-1}\left[m^{-1}\hat{\boldsymbol{\Lambda}}(n\hat{\sigma}^2 + \hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{Z}}^T\hat{\boldsymbol{Z}}\hat{\beta}) + \hat{\sigma}^2\hat{\boldsymbol{Z}}^T\hat{\boldsymbol{Z}}\right](\hat{\boldsymbol{Z}}^T\hat{\boldsymbol{Z}} + k\boldsymbol{I})^{-1},
\end{aligned}$$

and the simillar result is confirmed for $\widehat{\mathrm{AMSEM}}(\hat{\boldsymbol{\beta}}_{RMER})$.
Theorem 3.7 indicates the conditions that $\hat{\boldsymbol{\beta}}_{RMER}$ is superior to $\hat{\boldsymbol{\beta}}_{RME}$ with respect to AMSEM sense, which easily can be proved using the following Lemma (see Farebrother [11]). Other comparisons are similar.

Lemma 3.6. *Let* $\mathbf{M}$ *be a positive definite matrix, namely* $\mathbf{M} > 0$ *and* $\mathbf{c}$ *be some vector, then* $\mathbf{M} - \mathbf{c}\mathbf{c}^T \geq 0$ *if and only if* $\mathbf{c}^T\mathbf{M}^{-1}\mathbf{c} \leq 1$.

Theorem 3.7. $\hat{\boldsymbol{\beta}}_{RMER}$ *is superior to* $\hat{\boldsymbol{\beta}}_{RME}$ *with respect to* AMSEM sense if and only if $\mathbf{a}^T\mathbf{M}^{-1}\mathbf{a} < \mathbf{1}$ where $\mathbf{M} = [(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\mathbf{N}_m(\boldsymbol{Z}^T\boldsymbol{Z})^{-1} - (\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1}\mathbf{N}_m(\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1}]$.

Proof. To use lemma 3.6 we need to prove that $\mathbf{M}$ is positive definite matrix. Note that

$$\begin{aligned}
\mathbf{M} &= (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\mathbf{N}_m(\boldsymbol{Z}^T\boldsymbol{Z})^{-1} - (\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1}\mathbf{N}_m\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-1} \\
&= (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\mathbf{N}_m(\boldsymbol{Z}^T\boldsymbol{Z})^{-1} - [\mathbf{G}_k{}^{-1}(\boldsymbol{Z}^T\boldsymbol{Z})]^{-1}\mathbf{N}_m[(\boldsymbol{Z}^T\boldsymbol{Z})\mathbf{G}_k{}^{-1})]^{-1} \\
&= (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}(\mathbf{N}_m - \mathbf{G}_k\mathbf{N}_m\mathbf{G}_k)(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}
\end{aligned}$$

where $\mathbf{G}_k = [\boldsymbol{I} + k(\boldsymbol{Z}^T\boldsymbol{Z})^{-1})]^{-1}$. To prove that $\mathbf{M} > 0$ it is suffices to show that $\mathbf{N}_m - \mathbf{G}_k\mathbf{N}_m\mathbf{G}_k > 0$. Now, if we replace the value of $\mathbf{N}_m$, we will have

$$\begin{aligned}
\mathbf{N}_m - \mathbf{G}_k\mathbf{N}_m\mathbf{G}_k &= \mathbf{B} + \sigma^2(\boldsymbol{Z}^T\boldsymbol{Z}) - \mathbf{G}_k(\mathbf{B} + \sigma^2(\boldsymbol{Z}^T\boldsymbol{Z}))\mathbf{G}_k \\
&= \mathbf{B} - \mathbf{G}_k\mathbf{B}\mathbf{G}_k + \sigma^2(\boldsymbol{Z}^T\boldsymbol{Z} - \mathbf{G}_k(\boldsymbol{Z}^T\boldsymbol{Z})\mathbf{G}_k)
\end{aligned}$$

where $\mathbf{B} = m^{-1}\boldsymbol{\Lambda}(n\sigma^2 + \boldsymbol{\beta}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\beta})$. But $\mathbf{B} - \mathbf{G}_k\mathbf{B}\mathbf{G}_k > 0$ and $\sigma^2(\boldsymbol{Z}^T\boldsymbol{Z} - \mathbf{G}_k\boldsymbol{Z}^T\boldsymbol{Z}G_k) > 0$. See Ghapani and Babadi [13] for more details. □

4 Numerical study

4.1 The Monte Carlo simulation

In this section we employ the Monte Carlo simulation to compare the performance of the four estimators in the previous section with respect to their estimated mean square errors under several degrees of multicollinearity. Also, the two different values of error variances are considered, too. Following McDonald and Galarneau [18], the explanatory variables are generated by

$$z_{ij} = \sqrt{(1 - \rho^2)}w_{ij} + \rho w_{i,p+1},\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, p,$$

where w_{ij} are independent standard normal variables and ρ is the correlation between any two explanatory variables. We consider three different sets of correlation corresponding to $\rho = 0.20, 0.50, 0.80$.
In this experiment, $p = 4$ is the number of explanatory variables and $n = 50$, $n = 100$ and $n = 500$. We generated the l-th set of simulated data as

$$\begin{aligned} \boldsymbol{y_l} &= \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_l \\ \boldsymbol{X_l} &= \boldsymbol{Z} + \boldsymbol{\Delta}_l, l = 1, 2, \cdots, 2000, \end{aligned}$$

where $y_l = (y_{1l}, y_{2l}, \cdots, y_{nl})^T$ and $Z = (Z^{(1)}, Z^{(2)}, Z^{(3)}, Z^{(4)})$, $Z^{(j)} = (Z_{1j}, Z_{2j} \cdots, Z_{nj})^T$, $j = 1, 2, 3, 4$ and $\boldsymbol{\varepsilon_l} \sim N(0, \sigma^2 I_n)$ is rewritten in accordance with y_l. Furthermore, we assume that $\sigma^2 = 0.25$, $\sigma^2 = 2$, $\Lambda = \text{diag}(0.05, 0.05, 0.05, 0.05)$ and $\Lambda = \text{diag}(0.15, 0.15, 0.15, 0.15)$. For each set of explanatory variables, we consider the coefficient vector equal to $(4, 3, 2, 1)$. Then, the experiment is replicated 2000 times by generating new error terms. Once a set of explanatory and dependent variables is constructed, all variables are standardized and the estimates are determined using the standardized variables. After generating the sample, the estimated MSE (EMSE) for any estimator is calculated as follows:

$$\text{EMSE}(\hat{\boldsymbol{\beta}}) = \frac{1}{2000} \sum_{j=1}^{2000} (\hat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{\beta})^T (\hat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{\beta})$$

where $\hat{\boldsymbol{\beta}}_{(j)}$ is the estimation of $\boldsymbol{\beta}$ in the jth replication of the simulation. We use the R software version 3.4.0, R.app 1.70 and all source codes are available from the first author upon request. Note that the best value of ridge parameter is calculated from R package glmnet. The results are summarized in Table 1. This table displays the EMSE's of ME, MER, RME and RMER estimators for the various values of ρ, Λ, σ^2 and n. Also, in this table we display the value of $\mathbf{a}^T \mathbf{M}^{-1} \mathbf{a}$ to compare RME and RMER estimators with respect to Theorem 3.7.
ME and MER estimated values are calculated using one replicate of measurement error, i.e. $m = 1$, while RME and RMER are estimated through using $m = 10$ replications. By comparing the simulation results for ME, MER, RME and RMER, we observe that by increasing the level of collinearity component, ρ, Λ and σ^2 , the EMSE values of the

different estimators will increase in general. Moreover, we can see that, for all cases, the RMER estimator contains smaller EMSE values than the others. The same results is deduced by comparing RME estimator with MER and ME estimators. Also, the MER estimator has small EMSE values than the ME estimator. It is noteworthy that by increasing ρ, decreasing EMSE's for RMER estimator is more remarkable. For example, the EMSE for ME estimator is 46.485 where it is 6.7514 for RMER estimator when $\sigma^2 = 2$, $\Lambda = \text{diag}(0.15, 0.15, 0.15, 0.15)$, $\rho = 0.80$ and $n = 50$. Furthermore, for all cases by increasing the value of n, the EMSE for all estimators will decrease for all levels of Λ, σ^2 and ρ. In addition, for all cases, the value of $\mathbf{a}^T\mathbf{M}^{-1}\mathbf{a} \leqslant 1$, which represents that RMER estimator can perform better than RME estimator as we theoretically pointed out in Section 3. It is worth mentioning, in all cases the RMER estimator has smaller EMSE than the other mentioned methods. However, when the collinearity does not exist or it's level is low, this difference is not significant. Therefore, the use of new estimator is not justifiable in such cases. Also, by considering $\beta = (4, 3, 2, 1)$ and observing Tables 2, 3 and 4, we figure out that the absolute of difference between true values of β_i's and their estimated values ($|\text{bias}(\hat{\beta}_{\text{i}})|$) in RMER estimator is less than that in all other methods. For example, in Table 4, for $\sigma^2 = 2$, $\rho = 0.8$ and $\Lambda = \text{diag}(0.05, 0.05, 0.05, 0.05)$ we estimate $|\text{bias}(\hat{\beta})| = (0.2728, 0.1119, 0.0734, 0.0612)$ which means $\hat{\beta}$ is closer to β in contrast to ME, MER and RME estimates.

Table 1: EMSE values of the ME, MER, RME and RMER estimators

	σ^2	$\Lambda = \text{diag}(0.05, 0.05, 0.05, 0.05)$			$\Lambda = \text{diag}(0.15, 0.15, 0.15, 0.015)$		
		$n = 50$	$n = 100$	$n = 500$	$n = 50$	$n = 100$	$n = 500$
				$\rho = 0.2$			
ME		0.12653	0.09052	0.06757	1.03957	0.74285	0.56965
MER		0.12023	0.08812	0.06745	1.01602	0.73467	0.56828
RME	0.25	0.11665	0.08635	0.06652	0.90397	0.66869	0.55248
RMER		0.11492	0.08563	0.06639	0.89717	0.66611	0.55204
$\mathbf{a}^T\mathbf{M}^{-1}\mathbf{a}$		0.6422	0.3754	0.3085	0.5933	0.3143	0.4752
ME		0.30620	0.17689	0.08352	1.51912	0.86315	0.59543
MER		0.30045	0.17610	0.08312	1.49926	0.86027	0.59498
RME	2	0.29301	0.17115	0.08225	1.13054	0.78431	0.57478
RMER		0.29077	0.17035	0.08212	1.12225	0.78159	0.57434
$\mathbf{a}^T\mathbf{M}^{-1}\mathbf{a}$		0.5354	0.7214	0.7156	0.7303	0.5208	0.6136
				$\rho = 0.5$			
ME		0.10356	0.06880	0.04692	0.84462	0.55506	0.39812
MER		0.09902	0.06712	0.04679	0.82571	0.54904	0.39716
RME	0.25	0.09322	0.06679	0.04637	0.67581	0.48614	0.38415
RMER		0.09132	0.06601	0.04624	0.66783	0.48326	0.38369
$\mathbf{a}^T\mathbf{M}^{-1}\mathbf{a}$		0.4235	0.4574	0.2269	0.7358	0.3454	0.6702
ME		0.31714	0.16556	0.06592	1.52747	0.70492	0.42363
MER		031248	0.16469	0.06556	1.50930	0.70161	0.42316
RME	2	0.30045	0.15750	0.06415	0.93965	0.60006	0.40721
RMER		0.29787	0.15661	0.06402	0.93023	0.59704	0.40674
$\mathbf{a}^T\mathbf{M}^{-1}\mathbf{a}$		0.5489	0.7312	0.5232	0.6429	0.6274	0.7201
				$\rho = 0.8$			
ME		1.39063	1.09724	0.98734	8.67687	5.79753	3.71574
MER		1.17154	0.93690	0.79834	8.19834	4.74788	3.01004
RME	0.25	0.43715	0.38697	0.23295	6.99749	2.92181	2.29550
RMER		0.25253	0.17880	0.13226	5.57399	2.08505	1.47081
$\mathbf{a}^T\mathbf{M}^{-1}\mathbf{a}$		0.7907	0.5464	0.6269	0.5214	0.7313	0.7122
ME		1.94745	1.42028	1.17373	**46.4850**	9.42483	5.81949
MER		1.63035	1.31499	0.97284	17.6864	6.87706	4.61469
RME	2	0.92864	0.69677	0.47655	6.89490	3.93014	2.73132
RMER		0.72020	0.39167	0.17572	**6.75143**	2.88960	1.64052
$\mathbf{a}^T\mathbf{M}^{-1}\mathbf{a}$		0.7440	0.5488	0.7366	0.7636	0.7406	0.7313

Table 2: Absolute values of bias($\hat{\beta}_i$) for the ME, MER, RME and RMER when n=50

	σ^2	$\Lambda = \text{diag}(0.05, 0.05, 0.05, 0.05)$				$\Lambda = \text{diag}(0.15, 0.15, 0.15, 0.015)$			
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
					$\rho = 0.2$				
ME	0.25	0.2118	0.1486	0.0902	0.0382	0.6455	0.4717	0.2614	0.0936
MER		0.2077	0.1452	0.0898	0.0380	0.6390	0.4670	0.2598	0.0884
RME		0.1996	0.1435	0.0886	0.0352	0.5793	0.4296	0.1620	0.0882
RMER		0.1991	0.1428	0.0882	0.0350	0.5359	0.4288	0.1615	0.0741
ME	2	0.2142	0.1572	0.0944	0.0575	1.3082	0.5294	0.2936	0.1074
MER		0.2089	0.1551	0.0931	0.0570	1.3059	0.5284	0.2907	0.1064
RME		0.2033	0.1543	0.0925	0.0561	0.6439	0.4540	0.2871	0.1006
RMER		0.2023	0.1537	0.0918	0.0552	0.6409	0.4519	0.2861	0.1001
					$\rho = 0.5$				
ME	0.25	0.1982	0.1145	0.0406	0.0364	0.6255	0.3462	0.1072	0.0769
MER		0.1963	0.1135	0.0403	0.0360	0.6224	0.3445	0.1070	0.0759
RME		0.1832	0.1125	0.0398	0.0327	0.5872	0.3436	0.1008	0.0654
RMER		0.1803	0.1122	0.0390	0.0324	0.5835	0.3203	0.1005	0.0643
ME	2	0.2024	0.1269	0.0782	0.0480	0.6926	0.4549	0.1123	0.0891
MER		0.2009	0.1250	0.0752	0.0477	0.6902	0.4035	0.1120	0.0884
RME		0.1913	0.1241	0.0664	0.0464	0.6780	0.3834	0.1060	0.0873
RMER		0.1882	0.1232	0.0599	0.0458	0.6740	0.3511	0.1054	0.0860
					$\rho = 0.8$				
ME	0.25	0.9071	0.5271	0.2921	0.1767	1.8405	0.9375	0.4354	0.2712
MER		0.8506	0.5015	0.2712	0.1583	1.6612	0.8703	0.3855	0.2383
RME		0.7238	0.4147	0.1985	0.1293	1.2381	0.7242	0.3543	0.1992
RMER		0.2717	0.1202	0.0848	0.0729	1.0198	0.4981	0.2966	0.1372
ME	2	0.9734	0.5840	0.3323	0.2001	2.6653	2.2506	1.8733	1.6206
MER		0.8962	0.5341	0.2963	0.1710	1.6421	1.0429	0.7321	0.8046
RME		0.7760	0.4419	0.2114	0.1501	1.4375	0.5939	0.5840	0.3990
RMER		0.3075	0.1238	0.1009	0.0901	1.2682	0.4649	0.3178	0.1872

Table 3: Absolute values of bias($\hat{\beta}_i$) for the ME, MER, RME and RMER when n=100

	σ^2	$\Lambda = \text{diag}(0.05, 0.05, 0.05, 0.05)$				$\Lambda = \text{diag}(0.15, 0.15, 0.15, 0.015)$			
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
					$\rho = 0.2$				
ME	0.25	0.2012	0.1459	0.0876	0.0304	0.5915	0.4330	0.1048	0.0722
MER		0.2002	0.1439	0.0847	0.0293	0.5878	0.4303	0.1047	0.0715
RME		0.1909	0.1361	0.0567	0.0281	0.5567	0.3638	0.1015	0.0638
RMER		0.1891	0.1359	0.0483	0.0231	0.4882	0.3620	0.1011	0.0629
ME	2	0.2086	0.1512	0.0935	0.0367	0.6071	0.4353	0.2735	0.1024
MER		0.2076	0.1505	0.0931	0.0366	0.6032	0.4325	0.2730	0.1022
RME		0.1911	0.1465	0.0892	0.0334	0.5968	0.4270	0.2693	0.0965
RMER		0.1906	0.1457	0.0888	0.0333	0.5956	0.4261	0.2676	0.0959
					$\rho = 0.5$				
ME	0.25	0.1806	0.1141	0.0355	0.0353	0.5425	0.3291	0.0959	0.0640
MER		0.1796	0.1134	0.0353	0.0351	0.5415	0.3285	0.0955	0.0637
RME		0.1725	0.1089	0.0331	0.0326	0.5401	0.3212	0.0936	0.0584
RMER		0.1712	0.1081	0.0323	0.0317	0.5384	0.2457	0.0887	0.0577
ME	2	0.1845	0.1218	0.0679	0.0452	0.5548	0.3713	0.1034	0.0763
MER		0.1837	0.1214	0.0614	0.0441	0.5536	0.3505	0.1033	0.0759
RME		0.1789	0.1204	0.0591	0.0421	0.5476	0.3431	0.0942	0.0751
RMER		0.1710	0.1196	0.0567	0.0416	0.5459	0.2822	0.0933	0.0746
					$\rho = 0.8$				
ME	0.25	0.8493	0.5002	0.2689	0.1530	1.6735	0.8286	0.3847	0.2312
MER		0.8217	0.4891	0.2510	0.1471	1.5684	0.7384	0.3077	0.2177
RME		0.6957	0.4042	0.1886	0.1190	1.4555	0.6477	0.2527	0.2032
RMER		0.2333	0.1118	0.0764	0.0599	0.8597	0.3614	0.1611	0.1168
ME	2	0.9482	0.5727	0.3710	0.2413	1.9298	0.9463	0.4963	0.3592
MER		0.9101	0.5624	0.3266	0.2011	1.6856	0.8247	0.4054	0.3057
RME		0.8053	0.4301	0.2307	0.1322	1.5373	0.6956	0.3249	0.2551
RMER		0.2901	0.1217	0.0886	0.0788	1.0482	0.4428	0.2133	0.1498

Table 4: Absolute values of bias($\hat{\beta}_i$) for the ME, MER, RME and RMER when n=500

	σ^2	$\Lambda = \text{diag}(0.05, 0.05, 0.05, 0.05)$				$\Lambda = \text{diag}(0.15, 0.15, 0.15, 0.015)$			
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
					$\rho = 0.2$				
ME	0.25	0.1912	0.1366	0.0859	0.0296	0.5582	0.4023	0.0985	0.0707
MER		0.1910	0.1364	0.0817	0.0283	0.5576	0.4019	0.0982	0.0706
RME		0.1906	0.1171	0.0513	0.0276	0.5448	0.3033	0.0985	0.0609
RMER		0.1799	0.0854	0.0462	0.0218	0.4560	0.2032	0.0884	0.0579
ME	2	0.1918	0.1379	0.0873	0.0355	0.5607	0.4108	0.2503	0.0907
MER		0.1913	0.1376	0.0872	0.0354	0.5605	0.4106	0.2502	0.0907
RME		0.1901	0.1364	0.0868	0.0322	0.5535	0.4056	0.2491	0.0852
RMER		0.1675	0.1363	0.0867	0.0322	0.5523	0.3055	0.2393	0.0852
					$\rho = 0.5$				
ME	0.25	0.1718	0.1054	0.0348	0.0330	0.5124	0.3024	0.0951	0.0519
MER		0.1716	0.1053	0.0337	0.0329	0.5121	0.3022	0.0950	0.0512
RME		0.1713	0.1053	0.0323	0.0316	0.5108	0.3009	0.0857	0.0498
RMER		0.1631	0.1051	0.0278	0.0315	0.5031	0.2045	0.0756	0.0472
ME	2	0.1724	0.1128	0.0568	0.0412	0.5136	0.3045	0.0982	0.0684
MER		0.1721	0.1127	0.0567	0.0408	0.5134	0.3044	0.0971	0.0672
RME		0.1714	0.1118	0.0528	0.0398	0.5129	0.3040	0.0932	0.0659
RMER		0.1675	0.1100	0.0478	0.0381	0.5105	0.2067	0.0831	0.0650
					$\rho = 0.8$				
ME	0.25	0.7935	0.4599	0.2348	0.1465	1.4815	0.6738	0.3167	0.2482
MER		0.7530	0.4196	0.1947	0.1259	1.4707	0.6435	0.2858	0.2176
RME		0.6018	0.3703	0.1545	0.0971	0.9880	0.4041	0.2255	0.1748
RMER		0.2711	0.1062	0.0714	0.0493	0.9167	0.2936	0.1553	0.1037
ME	2	0.8431	0.5167	0.2941	0.1766	1.6801	0.8310	0.3911	0.2615
MER		0.7924	0.4659	0.2523	0.1440	1.5697	0.7589	0.3357	0.2366
RME		0.6765	0.4088	0.1917	0.1293	1.2229	0.4922	0.2803	0.2072
RMER		**0.2728**	**0.1119**	**0.0734**	**0.0612**	0.9316	0.3487	0.1701	0.1361

4.2 The Real data

In order to apply our theoretical results in a real data set, we apply the data set were provided by Dr. Paul Nicholson of the Department of Archaeology and Prehistory, University of Sheffield, UK. The data arises from an extensive archaeological survey of pottery production and distribution in the ancient Egyptian city of Amarna. The data consist of measurements of chemical contents (mineral elements) made on many samples of pottery. In this regard using two different techniques which are known as neutron activation analysis (NAA) and inductively coupled plasma (ICP) spectrometry (Smith et al., [30]). The pots in this example have been collected from different locations around the city and each pottery has it's own fabric code which can be recognised. Archaeologists believe that observations from pottery with the same fabric code and from the same provenance can essentially be regarded as replicates. Consequently, the data set has been divided into twenty-eight groups and in most groups there is some replication of observations. Weeding out a few fabric code with maximum two replications, we consider three initial replications of eighteen remained groups.

Among all mineral elements, we have concentrated on the relation between the Na values measured by NAA as the dependent variable versus three mineral elements Al, K, Ti measured by ICP technique as the explanatory variables. The data are not given here but they are available from the first author on request. Clearly, because of the nature of the data and the accuracy of the measuring instruments or human mistake, all explanatory variables are prone to measurment error as well as predictors. The correlations between all explanatory variables are significant, especially between two variables Al and Ti; this correlation value is equal to 0.85. Therefore we fitted a replicated measurement error model to this data set. The Ridge parameter (k) is calculated with R package glmnet and it's equal to 0.00055. First we obtain the estimates $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$ and $\hat{\boldsymbol{Z}}$ and then we calculate $\widehat{\mathrm{AVar}}$ and $\widehat{\mathrm{AMSEM}}$ for all estimators. Estimates of coefficients and their asymptotic variances (list in parentheses) using these four estimators are summarized in Table 5. From this table, we find that the $\hat{\boldsymbol{\beta}}_i^{\prime}s$ asymptotic variances of the RMER estimator are all smaller than those of other estimators. The same results are held when we compare RME estimator with ME and MER estimators. Also, with respect to Table 5, the superiority of the RMER method is also confirmed by the values of Akaike Information Criterion (AIC) where it is equal to $-2(\mathrm{log}-\mathrm{likelihood})+2\mathrm{k}+(\frac{2\mathrm{k}(\mathrm{k}+1)}{\mathrm{n}-\mathrm{k}-1})$ (for more details, see Akaike [1] and Burnham and Anderson [4]), indeed, this values are smaller than other estimators. Finaly, according to the results of Theorem 3.7, to compare RMER estimator by RME, we find the estimation of $\mathbf{a}^T\mathbf{M}^{-1}\mathbf{a} = 2.03\mathrm{E}-6$. In another words, $\hat{\boldsymbol{\beta}}_{RMER}$ is superior to $\hat{\boldsymbol{\beta}}_{RME}$ with respect to AMSEM sense. Also, this value is equal to $1.88\mathrm{E}-5$ and $2.85\mathrm{E}-6$ to compare RMER estimator by MER and ME methods, respectively.

Table 5: Estimation of regression coefficient and AIC values for Egyptian Pottery data

Estimation	ME	MER	RME	RMER
$\widehat{\beta}_1$	-0.7711(0.190)	-0.7676(0.198)	0.1115(0.037)	0.1113(0.023)
$\widehat{\beta}_2$	0.9735(0.531)	0.9724(0.566)	0.6997(0.562)	0.6995(0.289)
$\widehat{\beta}_3$	6.3535(9.979)	6.3262(10.37)	-0.6074(2.809)	-0.6058(1.554)
AIC	39.159	38.733	37.960	37.900

5 Conclusion

In this work we use the ridge regression method to combat multicollinearity in the estimation of model parameters in replicated measurement error models. Some large sample properties of our estimator are derived and compared with some other estimators using a simulation study and a real data analysis. Some developments of this article may be done elsewhere. We hope to do more works with correlated error assumption in the future. Also we are trying to use the Lasso and Elastic net methods (description of these regression models can been found in, Tibshirani [31] and Zou and Hastie [36]) to overcome the sparsity in replicated measurement error models.

References

[1] H. Akaike , A new look at the statistical model identification, *IEEE Transaction Automatic Control*, 19(6) (1974), 716-723.

[2] B. Blas, H. Bolfarine and V. H. Lachos, Heavy tailed calibration model with Berkson measurement errors for replicated data, *Chemometrics and Intelligent Laboratory systems*, 156 (2016), 21-35.

[3] J. P. Buonaccorsi, *Measurement Error Models, Methods and Applications*, Chapman and Hall/CRC, (2010).

[4] K. P. Burnham and D. R. Anderson, *Model Selection and Multi-model Inference: A practical Information-theoretic Approach*, 2nd edition, Springer, New York (2002).

[5] C. Z. Cao, J. G. Lin, J. Q. Shi, W. Wang and X. Y. Zhang, Multivariate measurement error models for replicated data under heavy-tailed distributions, *Chemometrics*, 8 (2015), 457-466.

[6] R. J. Carroll, D. Ruppert, A. Stefanski and C. M. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective*, CRC Press, Boca Raton, Fla, 2nd edition, pp. 485 (2006).

[7] L. K. Chan and T. K. Mak, On the maximum likelihood estimation of a linear structural relationship when the intercept is known, *Journal of Multivariate Analysis*, 9 (1979), 304-313.

[8] T. B. Cook and D. T. Campbell, *Quasi-experimentation: Design and Analysis Issues for Field Settings*, Chicago, 450 pp (1979).

[9] I. Dalen, J. Buonaccorsi, J. Sexton, P. Laake and M. Thoresen, Correcting for misclassification of a categorized exposure in binary regression using replication data, *Statistics in Medicine*, 28 (2009), 3386-3410.

[10] V. Devanarayan and L. A. Stefanski, Empirical simulation extrapolation for measurement error models with replicate measurements, *Statistics and Probability Letters*, 59 (2002), 219-225.

[11] R. W. Farebrother, Further results on the mean square error of ridge regression, *Journal of the royal statistical society*, 38 (1976), 248-250.

[12] W. K. Fung, X. P. Zhong and B. C. Wei, On estimation and influence diagnostics in linear mixed measurement errors models, *American Journal of Mathematical and Management Sciences*, 23 (2003), 37-59.

[13] F. Ghapani and B. Babadi, A new ridge estimator in linear measurement error model with stochastic linear restrictions, *Journal of the Iranian Statistical Society*, 15(2) (2016), 87-103.

[14] P. Gimenez and M. L. Patat, Local influence for functional comparative calibration models with replicated data, *Statistical Papers*, 55(2) (2014), 431-454.

[15] A. E. Horel and R. W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, 12 (1970), 55- 83.

[16] Y. J. Huang, Corrected score with sizable covariate measurement error: patheology and remedy, *Statistica Sinica*, 24 (2014), 357-374.

[17] Y. Isogawa, Estimating a multivariate linear structural relationship with replication, *Journal of the Royal Statistical Society, Series B*, 47 (1985), 211-215.

[18] C. Mc Donald and D. A. Galarneau, A monte carlo evaluation of some ridge-type estimators, *Journal of the American Statistical Association*, 70 (1975), 407-416.

[19] N. D. Nagelkerke, *Maximum Likelihood Estimation of Functional Relationships*, Springer, Berlin (1992).

[20] T. Nakamura, Corrected score function for errors-in-variables models: methodology and application to generalized linear models, *Biometrika*, 77 (1990), 127-137.

[21] L. S. Nawarathna and P. K. Choudhary, A heteroscedastic measurement error model for method comparison data with replicate measurements, *Statistics in Medicine*, 34(7) (2015), 1242-1258.

[22] M. R. Ozkale, A stochastic restricted ridge regression estimator, *Journal of multivariate analysis*, 100 (2009), 1706-1716.

[23] A. R. Rasekh, Local influence in measurement error models with ridge estimate, *computational Statistics and Data Analysis*, 50 (2006), 2822-2834.

[24] A. R. Rasekh and N. R. J. Fieller, Influence functions in functional measurement error models with replicated data, *Statistics*, 37(2) (2003), 169-178.

[25] E. Saleh and T. J. Shalabh, Ridge regression estimation approach to measurement error model, *Journal of Multivariate Analysis*, 123 (2014), 68-84.

[26] T. J. Shalabh and P. Liau, Immaculating the inconsistent estimator of slope parameter in measurement error model with replicated data, *Journal of Statistical Computation and Simulation*, 86 (2016), 3371-3387.

[27] T. J. Shalabh, C. M. Paudel and N. Kumar, Consistent estimation of regression parameters under replicated ultrastructural model with non normal errors, *Journal of Statistical Computation and Simulation*, 79(3) (2009), 251-274.

[28] S. Singh, K. Jain and S. Sharma, Replicated measurement error model under exact linear restrictions, *StatisticaL Paper*, 55 (2014), 253-274.

[29] S. Singh, K. Jain and S. Sharma, Using stochastic prior information in consistent estimation of regression coefficients in replicated measurement error model, *Journal of Multivariate Analysis*, 111 (2012), 198-212.

[30] D. M. Smith, F. A. Hart, R. D. Symond and J. N. Walsh, Analysis of Roman pottery from Colchester by inductively coupled plasma spectrometry, *In: Slater, E. A. and Tate, J. O. (Eds.), Science and Archaeology Glasgow 1987, Vol. 196, Issue No. i. B. A. R., Oxford*, (1988), pp. 41–55.

[31] R. Tibshirani, Regression shrinkage and selection via the LASSO, *Journal of the Royal Statistical Society, Series B, Wiley*, 58(1) (1996), 267-288.

[32] A. Ullah, T. J. Shalabh and D. Mukherjee, Consistent estimation of regression coefficients in replicated data with non-normal measurement errors, *Annals of Economics and Finance*, 2 (2001), 249-264.

[33] G. Wimmer and V. Witkovsky, Univariate linear calibration via replicated errors-in-variables model, *Journal of Statistical Computation and Simulation*, 77 (2007), 213-227.

[34] Y. P. Yang, G. R. Li and T. J. Tong, Corrected empirical likelihood for a class of generalized linear measurement error models, *Science China Mathematics*, 58 (2015), 1523-1536.

[35] K. Zare and A.R. Rasekh, Diagnostic measures for linear mixed measurement error models, *Sort*, 35 (2011), 125-144.

[36] H. Zou and T. Hastie, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society, Series B, Wiley*, 67 (2005), 301-320.

Abdul Rasoul Ziaei
Ph.D. Student of Statistics
Department of Statistics
Marvdasht Branch, Islamic Azad University
Marvdasht, Iran
E-mail: ziaei85@yahoo.com

Karim Zare
Assistant Professor of Statistics
Department of Statistics
Marvdasht Branch, Islamic Azad University
Marvdasht, Iran
E-mail: karim.zare@yahoo.com

Ayyub Sheikhi
Associate Professor of Statistics
Department of Statistics, Faculty of Mathematics and Computer
Shahid Bahonar University of Kerman
Kerman, Iran
E-mail: sheikhy.a@uk.ac.ir