Journal of Mathematical Modeling
Vol. 8, No. 2, 2020, pp. 157-176

JMM

Generalized two-parameter estimator in linear regression model

Amir Zeinal[*]
Department of Statistics, Faculty of Mathematical Sciences, University of Guilan, Rasht, Iran

Email(s): amirzeinal@guilan.ac.ir

Abstract. In this paper, a new two-parameter estimator is proposed. This estimator is a generalization of two-parameter (TP) estimator introduced by Özakle and Kaçiranlar (The restricted and unrestricted two-parameter estimator, Commun. Statist. Theor. Meth. **36** (2007) 2707–2725) and includes the ordinary least squares (OLS), the ridge and the generalized Liu estimators, as special cases. Here, the performance of this new estimator over the TP estimator is theoretically investigated in terms of quadratic bias (QB) criterion and its performance over the OLS and TP estimators is also studied in terms of mean squared error matrix (MSEM) criterion. Furthermore, the estimation of the biasing parameters is obtained, a numerical example is given and a simulation study is done as well.

Keywords: Generalized Liu estimator, Lagrange method, mean squared error, ridge estimator, two-parameter estimator.
AMS Subject Classification 2010: 34A34, 65L05.

1 Introduction

Consider the linear regression model with

$$Y = X\beta + \varepsilon, \tag{1}$$

where Y is an $n \times 1$ vector of responses, X is an $n \times p$ matrix of the explanatory variables and of full rank p ($p \leq n$), β is a $p \times 1$ vector of

[*]Corresponding author.
Received: 3 November 2020 / Revised: 7 March 2020 / Accepted: 14 March 2020.
DOI: 10.22124/jmm.2020.14903.1353

unknown parameters and ε is an $n\times 1$ vector of error terms with expectation $E(\varepsilon)=0$ and covariance matrix $\mathrm{Cov}(\varepsilon)=\sigma^2 I$.

The OLS estimator, that is,

$$\hat{\beta}_{OLS}=\left(X'X\right)^{-1}X'Y,$$

is often used to estimate β in model (1). In regression analysis, researchers often face the problem of multicollinearity. In the presence of multicollinearity, the OLS estimator performs weakly. Multicollinearity is defined as the existence of nearly linear dependency between explanatory variables. When there is multicollinearity, we have $|X'X|\to 0$. This causes the vector $\hat{\beta}_{OLS}$ to have entries with big absolute value. As well, with regard to $\mathrm{Cov}\left(\hat{\beta}\right)=\sigma^2(X'X)^{-1}$, the multicollinearity causes the variance of the estimators for regression parameters to be very big, which in turn will results in wide confidence intervals for parameters. To solve this problem, various biased estimators have been presented.

Stein [11] and James and Stein [5] suggested a biased estimator, named Stein-James estimator or the shrunk least squares estimator. Massy [9] introduced the principal component regression (PCR) estimator. For minimizing $(Y-X\beta)'(Y-X\beta)$, Hoerl and Kennard [4] considered the restriction of $\beta'\beta=c$, in which c is a constant value, to overcome the problem of multicollinearity. That is, using the Lagrange method, they minimized the following expression

$$(Y-X\beta)'(Y-X\beta)+k\left(\beta'\beta-c\right),$$

where $k\geq 0$ is the Lagrangian multiplier. As a result, they achieved the ridge regression (RR) estimator as follows

$$\hat{\beta}(k)=\left(X'X+kI\right)^{-1}X'Y.$$

Parameter $k\geq 0$ is also called biasing parameter.

Liu [7] obtained the estimator of β by shrinking $\hat{\beta}_{OLS}$, namely, by considering the following equation

$$d\hat{\beta}_{OLS}=\beta+\varepsilon_1,$$

where $0<d<1$. Instead of minimizing just the expression $(Y-X\beta)'$ $(Y-X\beta)$, he recommended to minimize the following expression,

$$\begin{aligned}(Y-X\beta)'(Y-X\beta)+\varepsilon_1'\varepsilon_1 &= (Y-X\beta)'(Y-X\beta)\\ &\quad+\left(d\hat{\beta}_{OLS}-\beta\right)'\left(d\hat{\beta}_{OLS}-\beta\right).\end{aligned}$$

Consequently, he achieved the following estimator

$$\hat{\beta}_d = \left(X'X + I\right)^{-1} \left(X'Y + d\hat{\beta}_{OLS}\right).$$

He also introduced the generalized formula for this estimator by considering the following equation,

$$D\hat{\beta}_{OLS} = \beta + \varepsilon_2,$$

where $D = \operatorname{diag}(d_1, d_2, \ldots, d_p)$, $d_i \geq 0$. That is, by minimizing the following expression,

$$(Y - X\beta)'(Y - X\beta) + \left(D\hat{\beta}_{OLS} - \beta\right)' \left(D\hat{\beta}_{OLS} - \beta\right),$$

he achieved the generalized estimator as follows

$$\hat{\beta}_{GD} = \left(X'X + I\right)^{-1} \left(X'Y + D\hat{\beta}_{OLS}\right).$$

Özakle and Kaçiranlar [10] combined ridge and Liu estimators and achieved the two-parameter (TP) estimator. They recommended to minimize the following objective function,

$$(Y - X\beta)'(Y - X\beta) + k\left[\left(d\hat{\beta}_{OLS} - \beta\right)' \left(d\hat{\beta}_{OLS} - \beta\right) - c\right].$$

The estimator they proposed is as follows

$$\hat{\beta}(k,d) = \left(X'X + kI\right)^{-1} \left(X'Y + kd\hat{\beta}_{OLS}\right).$$

where $k > 0$ and $0 < d < 1$. In this paper, the above-mentioned estimator will be generalized.

The rest of the paper is as follows. In Section 2 the generalized two-parameter (GTP) estimator is introduced. The performance of the proposed estimator with respect to quadratic bias (QB) and mean squared error matrix (MSEM) criteria is discussed in Section 3 and a method was presented to choose the biasing parameters in Section 4. To compare this estimator with TP and OLS estimators, a numerical example is given and a simulation study is done in Sections 5 and 6, respectively. The conclusion is given in Section 7.

2 The proposed estimator

To simplify the consideration about the linear model, the canonical form is often used. A symmetric matrix $S = X'X$ has a spectral decomposition of the form $S = P\Lambda P'$, where P is an orthogonal matrix and Λ is a real diagonal matrix. The diagonal elements of Λ are the eigenvalues of S and the column vectors of P are eigenvectors of S. The orthogonal version of the standard multiple linear regression model is

$$Y = XPP'\beta + \varepsilon = Z\alpha + \varepsilon,$$

where $Z = XP$, $\alpha = P'\beta$ and $Z'Z = \Lambda$. The ordinary LS estimator of α is given by

$$\hat{\alpha}_{OLS} = \left(Z'Z\right)^{-1} Z'Y = \Lambda^{-1} Z'Y. \tag{2}$$

The two-parameter estimator introduced by Özakle and Kaçiranlar [10] is defined as

$$\begin{aligned} \hat{\alpha}(k,d) &= (\Lambda + kI)^{-1}\left(Z'Y + kd\hat{\alpha}_{OLS}\right) \\ &= (\Lambda + kI)^{-1}(\Lambda + kdI)\hat{\alpha}_{OLS}. \end{aligned} \tag{3}$$

This estimator is derived by minimizing $(Y - Z\alpha)'(Y - Z\alpha)$ subject to $(\alpha - d\hat{\alpha}_{OLS})'(\alpha - d\hat{\alpha}_{OLS}) = c$, that is by minimizing

$$(Y - Z\alpha)'(Y - Z\alpha) + k\left[(\alpha - d\hat{\alpha}_{OLS})'(\alpha - d\hat{\alpha}_{OLS}) - c\right],$$

where c is a constant and k is a the Lagrangian multiplier.

Here, by replacing d with $D = \mathrm{diag}\,(d_1, d_2, \ldots, d_p)$, the GTP estimator will be obtained. That is, the following expression will be minimized

$$(Y - Z\alpha)'(Y - Z\alpha) + k\left[(\alpha - D\hat{\alpha}_{OLS})'(\alpha - D\hat{\alpha}_{OLS}) - c\right]. \tag{4}$$

Differentiating function (4) with respect to α will lead to

$$\left(Z'Z + kI\right)\alpha = Z'Y + kD\hat{\alpha}_{OLS}.$$

Consequently,

$$\hat{\alpha}(k,D) = (\Lambda + kI)^{-1}\left(Z'Y + kD\hat{\alpha}_{OLS}\right) = (\Lambda + kI)^{-1}(\Lambda + kD)\,\hat{\alpha}_{OLS}, \tag{5}$$

where $k > 0$ and $0 < d_i < 1$, $i = 1, 2, \ldots, p$.

Different estimators are derived from $\hat{\alpha}(k, D)$ as follows

(I) $\lim\limits_{D \to I} \hat{\alpha}(k, D) = \hat{\alpha}_{OLS}$.

(II) $\lim\limits_{k \to 0} \hat{\alpha}(k, D) = \hat{\alpha}_{OLS}$.

(III) $\lim\limits_{D \to 0} \hat{\alpha}(k, D) = (\Lambda + kI)^{-1} \Lambda \hat{\alpha}_{OLS} = (\Lambda + kI)^{-1} Z'Y$, which is RR estimator.

(IV) $\hat{\alpha}\,(k, dI) = \hat{\alpha}\,(k, d)$, which is TP estimator.

(V) $\hat{\alpha}(1, D) = (\Lambda + I)^{-1} (\Lambda + D)\, \hat{\alpha}_{OLS} = \hat{\alpha}_{GD}$, which is generalized Liu estimator.

(VI) $\hat{\alpha}(1, dI) = (\Lambda + I)^{-1} (\Lambda + dI)\, \hat{\alpha}_{OLS} = \hat{\alpha}_d$, which is Liu estimator.

3 The performance of the new estimator by QB and MSEM criteria

3.1 QB criterion

The QB of an estimator such as $\hat{\alpha}$ is defined as

$$\mathrm{QB}(\hat{\alpha}) = \mathrm{Bias}(\hat{\alpha})' \mathrm{Bias}(\hat{\alpha}), \tag{6}$$

where $\mathrm{Bias}(\hat{\alpha}) = E(\hat{\alpha}) - \alpha$.

Theorem 1. *If $d < \min\{d_i,\ i = 1, 2, \ldots, p\}$, then*

$$\mathrm{QB}\left(\hat{\alpha}\,(k, D)\right) < \mathrm{QB}\left(\hat{\alpha}\,(k, d)\right).$$

Proof. From Eqs. (3) and (5), it is concluded that

$$\mathrm{Bias}\left(\hat{\alpha}(k, d)\right) = k(d - 1)(\Lambda + kI)^{-1} \alpha, \tag{7}$$
$$\mathrm{Bias}\left(\hat{\alpha}(k, D)\right) = k(D - I)(\Lambda + kI)^{-1} \alpha. \tag{8}$$

Then, from Eq. (6), it is deduced that

$$\mathrm{QB}\left(\hat{\alpha}\,(k, d)\right) = k^2 (d - 1)^2 \sum_{i=1}^{p} \frac{\alpha_i^2}{(\lambda_i + k)^2}$$
$$\mathrm{QB}\left(\hat{\alpha}\,(k, D)\right) = k^2 \sum_{i=1}^{p} \frac{(d_i - 1)^2 \alpha_i^2}{(\lambda_i + k)^2}. \tag{9}$$

Consequently,

$$QB\left(\hat{\alpha}\left(k,d\right)\right)-QB\left(\hat{\alpha}\left(k,D\right)\right)=k^2\sum_{i=1}^{p}\frac{\left[(d-1)^2-(d_i-1)^2\right]\alpha_i^2}{(\lambda_i+k)^2}.$$

Noticing that $0<d<1$, $0<d_i<1$, $i=1,\ldots,p$, the proof is completed. □

3.2 MSEM criterion

The MSEM of an estimator such as $\hat{\alpha}$ is defined as

$$\mathrm{MSEM}(\hat{\alpha})=\mathrm{Cov}(\hat{\alpha})+\mathrm{Bias}(\hat{\alpha})\mathrm{Bias}(\hat{\alpha})' \tag{10}$$

Lemma 1. *(Farebrother [2]) Let M be a positive definite matrix, namely $M>0$, and let l be a some vector, then $M-ll'>0$ if and only if $l'M^{-1}l<1$.*

Lemma 2. *(Trenkler and Toutenburg [12]) Let $\hat{\alpha}_j=A_jY$, $j=1,2$ be two competing estimators of α. Suppose that $E=\mathrm{Cov}(\hat{\alpha}_1)-\mathrm{Cov}(\hat{\alpha}_2)>0$, where $\mathrm{Cov}(\hat{\alpha}_j)$, $j=1,2$ denotes the covariance matrix of $\hat{\alpha}_j$, $j=1,2$, then $\Delta(\hat{\alpha}_1,\hat{\alpha}_2)=\mathrm{MSEM}(\hat{\alpha}_1)-\mathrm{MSEM}(\hat{\alpha}_2)>0$ if and only if $b_2'(E+b_1b_1')b_2<1$, where $\mathrm{MSEM}(\hat{\alpha}_j)$ and b_j denote the mean squared error matrix and bias vector of $\hat{\alpha}_j$, respectively.*

Theorem 2. *If $k>0$ and $0<d_i<1$, $i=1,\ldots,p$, then*

$$\mathrm{MSEM}(\hat{\alpha}_{OLS})-\mathrm{MSEM}(\hat{\alpha}(k,D))>0,$$

if and only if

$$k\alpha'\left[2I+k(I+D)\Lambda^{-1}\right]^{-1}(I-D)\alpha<\sigma^2.$$

Proof. It is well-known that

$$\mathrm{Bias}(\hat{\alpha}_{OLS}) = 0, \tag{11}$$

$$\mathrm{Cov}(\hat{\alpha}_{OLS}) = \sigma^2\Lambda^{-1}. \tag{12}$$

From Eqs. (5) and (12), it is concluded that

$$\mathrm{Cov}(\hat{\alpha}(k,D))=\sigma^2(\Lambda+kI)^{-1}(\Lambda+kD)\Lambda^{-1}(\Lambda+kD)(\Lambda+kI)^{-1}. \tag{13}$$

Now, from Eqs. (12) and (13), the following equation is obtained:

$$\begin{aligned} E &= \mathrm{Cov}(\hat{\alpha}_{OLS}) - \mathrm{Cov}(\hat{\alpha}(k, D)) \\ &= \sigma^2(\Lambda + kI)^{-1}\big[(\Lambda + kI)\Lambda^{-1}(\Lambda + kI) \\ &\quad - (\Lambda + kD)\Lambda^{-1}(\Lambda + kD)\big](\Lambda + kI)^{-1} \\ &= k\sigma^2(\Lambda + kI)^{-1}(I - D)\left[2I + k(I + D)\Lambda^{-1}\right](\Lambda + kI)^{-1}. \end{aligned} \tag{14}$$

Consequently, using the equations (8), (11), (14) and Lemma 2, the proof is completed. □

Theorem 3. *Let $k > 0$, $0 < d < 1$, $0 < d_i < 1$, $i = 1, \dots, p$ and $d > \max\{d_i,\ i = 1, \dots, p\}$. Then*

$$\mathrm{MSEM}(\hat{\alpha}(k, d)) - \mathrm{MSEM}(\hat{\alpha}(k, D)) > 0,$$

if

$$k\alpha'(I - D)(dI - D)\left[2I + k(dI + D)\Lambda^{-1}\right](I - D)\alpha < \sigma^2.$$

Proof. From Eqs. (3), (12), it is concluded that

$$\mathrm{Cov}(\hat{\alpha}(k, d)) = \sigma^2(\Lambda + kI)^{-1}(\Lambda + kdI)\Lambda^{-1}(\Lambda + kdI)(\Lambda + kI)^{-1}. \tag{15}$$

Now, from Eqs. (7), (10) and (15), it is derived that

$$\begin{aligned} \mathrm{MSEM}(\hat{\alpha}(k, d)) = {} & \sigma^2(\Lambda + kI)^{-1}(\Lambda + kdI)\Lambda^{-1}(\Lambda + kdI)(\Lambda + kI)^{-1} \\ & - k^2(d - 1)^2(\Lambda + kI)^{-1}\alpha\alpha'(\Lambda + kI)^{-1}. \end{aligned} \tag{16}$$

On the other hand, from Eqs. (8), (10) and (13), it is resulted that

$$\begin{aligned} \mathrm{MSEM}(\hat{\alpha}(k, D)) = {} & \sigma^2(\Lambda + kI)^{-1}(\Lambda + kD)\Lambda^{-1}(\Lambda + kD)(\Lambda + kI)^{-1} \\ & - k^2(D - I)(\Lambda + kI)^{-1}\alpha\alpha'(\Lambda + kI)^{-1}(D - I). \end{aligned} \tag{17}$$

Consequently, from Eqs. (16) and (17), it is concluded that

$$\begin{aligned} & \mathrm{MSEM}(\hat{\alpha}(k, d)) - \mathrm{MSEM}(\hat{\alpha}(k, D)) \\ & \quad = (\Lambda + kI)^{-1}\Big\{\sigma^2\big[(\Lambda + kdI)\Lambda^{-1}(\Lambda + kdI) - (\Lambda + kD)\Lambda^{-1}(\Lambda + kD)\big] \\ & \qquad + k^2(d - 1)^2\alpha\alpha' - k^2(D - I)\alpha\alpha'(D - I)\Big\}(\Lambda + kI)^{-1}. \end{aligned}$$

It is obvious that $k^2(d - 1)^2\alpha\alpha' > 0$. Therefore,

$$\mathrm{MSEM}(\hat{\alpha}(k, d)) - \mathrm{MSEM}(\hat{\alpha}(k, D)) > 0,$$

if

$$\sigma^2\left[(\Lambda+kdI)\Lambda^{-1}(\Lambda+kdI)-(\Lambda+kD)\Lambda^{-1}(\Lambda+kD)\right] \\ -k^2(D-I)\alpha\alpha'(D-I)>0. \tag{18}$$

Denoting

$$M=\sigma^2\left[(\Lambda+kdI)\Lambda^{-1}(\Lambda+kdI)-(\Lambda+kD)\Lambda^{-1}(\Lambda+kD)\right] \\ =\sigma^2\mathrm{diag}\left\{\frac{(\lambda_i+kd)^2-(\lambda_i+kd_i)^2}{\lambda_i}\right\}_{i=1}^{p},$$

if $d>\max\{d_i,\ i=1,\dots,p\}$, then $M>0$. Consequently, using Lemma 1, condition (18) is valid if and only if

$$k^2\alpha'(D-I)M^{-1}(D-I)\alpha<1.$$

Therefore, the proof is completed. □

4 Selection of the parameters k and d_i, $i=1,2,\dots,p$

The optimal values for the parameters of an estimator such as $\hat{\alpha}$ can be derived by minimizing the scalar mean squared error (MSE) of $\hat{\alpha}$, which is defined as

$$\mathrm{MSE}(\hat{\alpha})=E\left[(\hat{\alpha}-\alpha)'(\hat{\alpha}-\alpha)\right]=\mathrm{tr}\left[\mathrm{MSEM}(\hat{\alpha})\right]. \tag{19}$$

Then Eqs. (6), (10) and (19) will result in

$$\mathrm{MSE}(\hat{\alpha})=\mathrm{tr}\left[\mathrm{Cov}(\hat{\alpha})\right]+\mathrm{QB}(\hat{\alpha}). \tag{20}$$

Consequently, from equations (9), (13) and (20), it is concluded that

$$\mathrm{MSE}\left(\hat{\alpha}\left(k,D\right)\right)=\sum_{i=1}^{p}\frac{\sigma^2(\lambda_i+kd_i)^2+k^2(d_i-1)^2\alpha_i^2\lambda_i}{\lambda_i(\lambda_i+k)^2}.$$

Now, the optimal values for k and d_i, $i=1,\dots,p$, from the following function will be obtained

$$f(k,d_1,d_2,\dots,d_p)=\mathrm{MSE}(\hat{\alpha}(k,D)).$$

The values of d_i, $i=1,2,\dots,p$, which minimize $f\left(k,d_1,d_2,\dots,d_p\right)$ for fixed k value can be obtained by differentiating $f\left(k,d_1,d_2,\dots,d_p\right)$ with respect to d_i, $i=1,2,\dots,p$.

$$\frac{\partial f\left(k,d_1,d_2,\dots,d_p\right)}{\partial d_i}=\frac{2\sigma^2k\left(\lambda_i+kd_i\right)+2k^2(d_i-1)\,\alpha_i^2\lambda_i}{\lambda_i(\lambda_i+k)^2},\quad i=1,2,\dots.p,$$

and equating them to zero. After the unknown parameters σ^2 and α_i's are replaced with their unbiased estimators, the optimal estimator of d_i, $i = 1, 2, \ldots, p$ for fixed k value, will be obtained as follows

$$\hat{d}_{iopt} = \frac{\left(k\hat{\alpha}_i^2 - \hat{\sigma}^2\right)\lambda_i}{k\left(\hat{\sigma}^2 + \hat{\alpha}_i^2\lambda_i\right)}, \quad i = 1, 2, \ldots, p. \tag{21}$$

The k value, which minimizes the function $f(k, d_1, d_2, \ldots, d_p)$, can be found by differentiating $f(k, d_1, d_2, \ldots, d_p)$ with respect to k when d_i's, $i = 1, 2, \ldots, p$, are fixed

$$\frac{\partial f(k, d_1, d_2, \ldots, d_p)}{\partial k} = \sum_{i=1}^{p} \frac{2\sigma^2(\lambda_i + kd_i)(d_i - 1) + 2k(d_i - 1)\alpha_i^2\lambda_i}{(\lambda_i + k)^3},$$

and equating it to zero. Using the idea suggested by Hoerl and Kennard [4], by equating the numerator of $\frac{\partial f(k,d_1,d_2,\ldots,d_p)}{\partial k}$ to zero, the value of k can be derived as follows

$$k = \frac{\sigma^2}{\alpha_i^2 - d_i\left(\frac{\sigma^2}{\lambda_i} + \alpha_i^2\right)}, \quad i = 1, 2, \ldots, p.$$

By replacing α_i, $i = 1, 2, \ldots, p$, and σ^2 values with their unbiased estimators, the optimal values of k for fixed d_i, $i = 1, 2, \ldots, p$, values will be obtained as follows

$$\hat{k} = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2 - d_i\left(\frac{\hat{\sigma}^2}{\lambda_i} + \hat{\alpha}_i^2\right)}, \quad i = 1, 2, \ldots, p.$$

Using the idea suggested by Kiabria [6], the arithmetic mean of abovementioned $\hat{k}$ values, the optimal estimator of k for fixed d_i, $i = 1, 2, \ldots, p$ values, will be obtained as follows

$$\hat{k}_{opt} = \frac{1}{p}\sum_{i=1}^{p} \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2 - d_i\left(\frac{\hat{\sigma}^2}{\lambda_i} + \hat{\alpha}_i^2\right)}. \tag{22}$$

Theorem 4. *If*

$$\hat{d}_i < \frac{\hat{\alpha}_i^2}{\frac{\hat{\sigma}^2}{\lambda_i} + \hat{\alpha}_i^2}, \quad i = 1, 2, \ldots, p, \tag{23}$$

then $\hat{k}_{opt}$ is always positive.

Proof. From (22), it is concluded. □

The selection of the estimators of the parameters k and d_i, $i = 1, 2, \ldots, p$, in $\hat{\beta}(k, D)$ can be obtained by applying the following iterative method.
Step 1. Calculate $\hat{d}_i$, $i = 1, 2, \ldots, p$ from (23).
Step 2. Estimate $\hat{k}_{opt}$ from (22) by using $\hat{d}_i$, $i = 1, 2, \ldots, p$, in Step 1.
Step 3. Obtain $\hat{d}_{iopt}$, $i = 1, 2, \ldots, p$ from (21) by using $\hat{k}_{opt}$ in Step 2.
Step 4. If $\hat{d}_{iopt}$, $i = 1, 2, \ldots, p$ is negative, use $\hat{d}_{iopt} = \hat{d}_i$, $i = 1, 2, \ldots, p$.

5 Numerical example

In order to illustrate the performance of the new estimator, the dataset originally due to Gruber [3], and later discussed by Akdeniz and Erol [1], is considered. Data found in economics are often multicollinear. Table 1 gives Total National Research and Development Expenditures-as a percent of Gross National Product by country: 1972-1986. It represents the relationship between the dependent variable Y, the percentage spent by the United States, and the four other independent variables X_1, X_2, X_3 and X_4. The variables X_1, X_2, X_3 and X_4, respectively, represent the percentage spent by France, the percentage spent by West Germany, the percentage spent by Japan, and the percentage spent by the former Soviet Union.

Table 1: The percentage of Gross National Product.

Year	Y	X_1	X_2	X_3	X_4
1972	2.3	1.9	2.2	1.9	3.7
1975	2.2	1.8	2.2	2.0	3.8
1979	2.2	1.8	2.4	2.1	3.6
1980	2.3	1.8	2.4	2.2	3.8
1981	2.4	2.0	2.5	2.3	3.8
1982	2.5	2.1	2.6	2.4	3.7
1983	2.6	2.1	2.6	2.6	3.8
1984	2.6	2.2	2.6	2.6	4.0
1985	2.7	2.3	2.8	2.8	3.7
1986	2.7	2.3	2.7	2.8	3.8

By considering $X = [\mathbf{1}, X_1, X_2, X_3, X_4]$, where $\mathbf{1}$ is a 10×1 vector in which all elements are 1, the eigenvalues of $X'X$ are obtained as follows

$$\lambda_1 = 312.9320,\ \lambda_2 = 0.7536,\ \lambda_3 = 0.0453,\ \lambda_4 = 0.0372,\ \lambda_5 = 0.0019$$

with $\hat{\sigma}^2 = 0.0016$. Consequently, the condition number is obtained 1.647×10^5, which suggests the presence of very severe collinearity.

In Table 2, the estimated QB and MSE of OLS, TP and GTP estimators are presented. To obtain these values, first the theoretical values of the QB and MSE of the estimators were used and then σ^2 and α_i, $i = 1, \dots, p$ were replaced with their unbiased estimators and at last the estimated optimal of their other parameters were used.

Table 2: Comparing the estimators.

	EMSE	EQB
OLS	0.9566	0
TP	0.4278	0.2342
GTP	0.3472	0.2182

6 The Monte Carlo simulation

The explanatory variables are generated following McDonald [8]

$$x_{ij} = (1 - \rho^2)^{\frac{1}{2}} Z_{ij} + \rho Z_{i\,p+1}, \quad i = 1, 2, \dots, n, \ j = 1, 2, \dots, p,$$

where Z_{ij}'s are independent standard normal pseudo-random numbers and ρ is specified so that the theoretical correlation between any two explanatory variables is given by ρ^2. Six different sets of correlations are considered corresponding to $\rho = 0.5$, 0.6, 0.7, 0.8, 0.9, 0.95, and twenty different values of $\sigma^2 = 0.01, 0.05, \dots, 4, 5$, will be studied, too.

Dependent variables y_i, $i = 1, 2, \dots, n$, are generated by the following equation:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

Here, $n = 30$, $p = 4$, $\beta_1 = 0.2$, $\beta_2 = 0.3$, $\beta_3 = 0.4$ and $\beta_4 = 0.5$ are considered. Also ε_i's are independent normal Pseudo-random numbers with mean 0 and variance σ^2. For a choice of ρ and σ^2, the simulation is repeated 10000 times. The estimated mean squared error (EMSE) is calculated for $\hat{\alpha}_{OLS}$, $\hat{\alpha}(k, d)$ and $\hat{\alpha}(k, D)$ as follows

$$\text{EMSE}(\hat{\alpha}) = \frac{1}{10000} \sum_{r=1}^{10000} (\hat{\alpha}_{(r)} - \alpha)'(\hat{\alpha}_{(r)} - \alpha).$$

The estimated bias (EB) is calculated for $\hat{\alpha}(k, d)$ and $\hat{\alpha}(k, D)$ as follows

$$\text{EB}(\hat{\alpha}) = \frac{1}{10000} \sum_{r=1}^{10000} (\hat{\alpha}_{(r)} - \alpha).$$

Table 3: EMSE, $\rho = 0.5$.

	σ^2				
	0.01	0.05	0.1	0.2	0.3
OLS	0.0018	0.0091	0.0183	0.0367	0.0555
TP	0.0018	0.0090	0.0181	0.0357	0.0532
GTP	0.0019	0.0092	0.0176	0.0330	0.0468
	σ^2				
	0.4	0.5	0.6	0.7	0.8
OLS	0.0733	0.0914	0.1100	0.1264	0.1458
TP	0.0689	0.0843	0.0996	0.1127	0.1267
GTP	0.0595	0.0711	0.0830	0.0942	0.1057
	σ^2				
	0.9	1	1.25	1.5	1.75
OLS	0.1652	0.1858	0.2287	0.2738	0.3262
TP	0.1420	0.1567	0.1854	0.2127	0.2447
GTP	0.1182	0.1302	0.1555	0.1826	0.2138
	σ^2				
	2	2.5	3	4	5
OLS	0.3622	0.4604	0.5556	0.7180	0.9053
TP	0.2634	0.3129	0.3626	0.4298	0.5071
GTP	0.2326	0.2874	0.3434	0.4284	0.5276

For each replication, the values k and d_i, $i = 1, 2, \ldots, p$, and the corresponding $\hat{\alpha}(k, D)$ are estimated by using the method in Section 4. Also the values k, d and the corresponding $\hat{\alpha}(k, d)$ are estimated by using the method presented by Özakle and Kaçiranlar [10].

The estimated MSE (EMSE) of the OLS, TP and GTP estimators for different values of ρ and σ^2 are presented in Tables 3–8. Also the estimated QB (EQB), obtained from EB, corresponding to $\hat{\alpha}(k, d)$ and $\hat{\alpha}(k, D)$, for different values of ρ and σ^2, are presented in Tables 9–14.

In Tables 3 and 4, when $\rho = 0.5$ and $\rho = 0.6$, respectively, the GTP estimator has a better performance than the TP estimator for $0.1 \leq \sigma^2 \leq 4$ and has a better performance than the OLS estimator for $0.1 \leq \sigma^2 \leq 5$.

In Tables 5 and 6, when $\rho = 0.7$ and $\rho = 0.8$, respectively, the GTP estimator has a better performance than the TP estimator for $0.05 \leq \sigma^2 \leq 4$ and has a better performance than the OLS estimator for $0.05 \leq \sigma^2 \leq 5$.

In Table 7, when $\rho = 0.9$, the GTP estimator has a better performance than the TP estimator for $0.05 \leq \sigma^2 \leq 2.5$ and has a better performance than the OLS estimator for $0.05 \leq \sigma^2 \leq 5$.

In Table 8, when $\rho = 0.95$, the GTP estimator has a better performance than the TP estimator for $0.01 \leq \sigma^2 \leq 1.5$ and has a better performance than the OLS estimator for $0.01 \leq \sigma^2 \leq 5$.

Table 4: EMSE, $\rho = 0.6$.

	σ^2				
	0.01	0.05	0.1	0.2	0.3
OLS	0.0021	0.0103	0.0208	0.0411	0.0619
TP	0.0021	0.0102	0.0204	0.0396	0.0585
GTP	0.0022	0.0103	0.0193	0.0347	0.0487
	σ^2				
	0.4	0.5	0.6	0.7	0.8
OLS	0.0824	0.1045	0.1254	0.1437	0.1636
TP	0.0761	0.0941	0.1108	0.1245	0.1391
GTP	0.0613	0.0749	0.0870	0.0970	0.1078
	σ^2				
	0.9	1	1.25	1.5	1.75
OLS	0.1845	0.2073	0.2560	0.3094	0.3610
TP	0.1537	0.1700	0.2009	0.2337	0.2614
GTP	0.1199	0.1332	0.1600	0.1893	0.2157
	σ^2				
	2	2.5	3	4	5
OLS	0.4133	0.5141	0.6268	0.8169	1.0440
TP	0.2911	0.3362	0.3926	0.4719	0.5679
GTP	0.2456	0.2973	0.3595	0.4582	0.5807

Table 5: EMSE, $\rho = 0.7$.

	σ^2				
	0.01	0.05	0.1	0.2	0.3
OLS	0.0025	0.0125	0.0250	0.0494	0.0760
TP	0.0025	0.0124	0.0243	0.0469	0.0699
GTP	0.0026	0.0122	0.0221	0.0387	0.0545
	σ^2				
	0.4	0.5	0.6	0.7	0.8
OLS	0.1005	0.1246	0.1505	0.1766	0.1989
TP	0.0899	0.1087	0.1285	0.1475	0.1627
GTP	0.0684	0.0814	0.0952	0.1098	0.1214
	σ^2				
	0.9	1	1.25	1.5	1.75
OLS	0.2265	0.2501	0.3175	0.3815	0.4366
TP	0.1812	0.1959	0.2366	0.2727	0.2994
GTP	0.1356	0.1480	0.1845	0.2179	0.2437
	σ^2				
	2	2.5	3	4	5
OLS	0.5012	0.6235	0.7578	0.9966	1.2540
TP	0.3329	0.3896	0.4515	0.5509	0.6530
GTP	0.2771	0.3409	0.4113	0.5336	0.6657

Table 6: EMSE, $\rho = 0.8$.

	σ^2				
	0.01	0.05	0.1	0.2	0.3
OLS	0.0035	0.0174	0.0345	0.0696	0.1049
TP	0.0034	0.0170	0.0330	0.0640	0.0929
GTP	0.0036	0.0162	0.0281	0.0493	0.0682
	σ^2				
	0.4	0.5	0.6	0.7	0.8
OLS	0.1399	0.1719	0.2054	0.2418	0.2779
TP	0.1194	0.1421	0.1652	0.1893	0.2109
GTP	0.0864	0.1024	0.1191	0.1378	0.1557
	σ^2				
	0.9	1	1.25	1.5	1.75
OLS	0.3130	0.3529	0.4329	0.5196	0.6963
TP	0.2327	0.2554	0.2985	0.3413	0.3826
GTP	0.1744	0.1946	0.2328	0.2767	0.3173
	σ^2				
	2	2.5	3	4	5
OLS	0.6905	0.8631	1.0451	1.3872	1.7167
TP	0.4201	0.4941	0.5738	0.7076	0.8322
GTP	0.3584	0.4447	0.5348	0.7056	0.8693

Table 7: EMSE, $\rho = 0.9$.

	σ^2				
	0.01	0.05	0.1	0.2	0.3
OLS	0.0063	0.0323	0.0634	0.1277	0.1926
TP	0.0063	0.0310	0.0585	0.1097	0.1556
GTP	0.0065	0.0266	0.0446	0.0778	0.1093
	σ^2				
	0.4	0.5	0.6	0.7	0.8
OLS	0.2605	0.3171	0.3907	0.4493	0.5133
TP	0.1996	0.2326	0.2741	0.3048	0.3374
GTP	0.1429	0.1682	0.2062	0.2328	0.2642
	σ^2				
	0.9	1	1.25	1.5	1.75
OLS	0.5746	0.6318	0.8073	0.9640	1.1204
TP	0.3669	0.3928	0.4726	0.5399	0.6031
GTP	0.2927	0.3185	0.4053	0.4799	0.5540
	σ^2				
	2	2.5	3	4	5
OLS	1.2873	1.6007	1.9231	2.5386	3.2162
TP	0.6738	0.7895	0.9195	1.1399	1.3935
GTP	0.6358	0.7789	0.9314	1.2235	1.5529

Table 8: EMSE, $\rho = 0.95$.

	σ^2				
	0.01	0.05	0.1	0.2	0.3
OLS	0.0125	0.0628	0.1242	0.2466	0.3708
TP	0.0123	0.0579	0.1066	0.1899	0.2620
GTP	0.0120	0.0439	0.0751	0.1338	0.1920
	σ^2				
	0.4	0.5	0.6	0.7	0.8
OLS	0.4951	0.6261	0.7319	0.8592	1.0028
TP	0.3284	0.3938	0.4420	0.4966	0.5580
GTP	0.2527	0.3166	0.3632	0.4241	0.4915
	σ^2				
	0.9	1	1.25	1.5	1.75
OLS	1.1372	1.2399	1.5273	1.8669	2.1765
TP	0.6164	0.6570	0.7662	0.8982	1.0131
GTP	0.5541	0.6013	0.7372	0.8965	1.0440
	σ^2				
	2	2.5	3	4	5
OLS	2.4564	3.1136	3.7396	4.9859	6.2105
TP	1.1211	1.3606	1.5950	2.0464	2.4761
GTP	1.1773	1.4913	1.7899	2.3847	2.9546

Table 9: EQB, $\rho = 0.5$.

	σ^2			
	0.01	0.05	0.1	0.2
TP	3.3577×10^{-6}	6.7450×10^{-5}	2.8859×10^{-4}	9.5487×10^{-5}
GTP	7.3560×10^{-7}	7.9619×10^{-6}	1.9536×10^{-5}	4.1078×10^{-5}
	σ^2			
	0.3	0.4	0.5	0.6
TP	0.0021	0.0031	0.0043	0.0058
GTP	7.1923×10^{-5}	9.2023×10^{-5}	1.3334×10^{-4}	1.6937×10^{-4}
	σ^2			
	0.7	0.8	0.9	1
TP	0.0072	0.0084	0.0098	0.0115
GTP	2.4676×10^{-4}	2.8040×10^{-4}	3.7432×10^{-4}	4.4332×10^{-4}
	σ^2			
	1.25	1.5	1.75	2
TP	0.0141	0.0164	0.0198	0.0223
GTP	6.7685×10^{-4}	8.9358×10^{-4}	0.0013	0.0015
	σ^2			
	2.5	3	4	5
TP	0.0250	0.0312	0.0358	0.0393
GTP	0.0024	0.0032	0.0048	0.0061

In Table 9, when $\rho = 0.5$, for $0.01 \leq \sigma^2 \leq 5$,

$$2.32 \leq \frac{\mathrm{EQB}(\hat{\alpha}_{TP})}{\mathrm{EQB}(\hat{\alpha}_{GTP})} \leq 34.24,$$

and

$$mean\left(\frac{\mathrm{EQB}(\hat{\alpha}_{TP})}{\mathrm{EQB}(\hat{\alpha}_{GTP})}\right) = 18.71.$$

Table 10: EQB, $\rho = 0.6$.

	σ^2			
	0.01	0.05	0.1	0.2
TP	5.8584×10^{-6}	1.3871×10^{-4}	5.0453×10^{-4}	0.0018
GTP	3.0249×10^{-6}	3.0536×10^{-5}	6.3256×10^{-5}	1.1362×10^{-4}
	σ^2			
	0.3	0.4	0.5	0.6
TP	0.0035	0.0057	0.0078	0.0099
GTP	1.7377×10^{-4}	2.0933×10^{-4}	2.6146×10^{-4}	2.9586×10^{-4}
	σ^2			
	0.7	0.8	0.9	1
TP	0.0124	0.0146	0.0172	0.0195
GTP	3.7222×10^{-4}	4.2906×10^{-4}	5.5307×10^{-4}	6.8264×10^{-4}
	σ^2			
	1.25	1.5	1.75	2
TP	0.0249	0.0294	0.0340	0.0377
GTP	9.5795×10^{-4}	0.0012	0.0014	0.0020
	σ^2			
	2.5	3	4	5
TP	0.0439	0.0497	0.0586	0.0635
GTP	0.0028	0.0038	0.0058	0.0080

In Table 10, when $\rho = 0.6$, for $0.01 \leq \sigma^2 \leq 5$,

$$1.94 \leq \frac{\mathrm{EQB}(\hat{\alpha}_{TP})}{\mathrm{EQB}(\hat{\alpha}_{GTP})} \leq 34.03,$$

and

$$mean\left(\frac{\mathrm{EQB}(\hat{\alpha}_{TP})}{\mathrm{EQB}(\hat{\alpha}_{GTP})}\right) = 20.42.$$

In Table 11, when $\rho = 0.7$, for $0.01 \leq \sigma^2 \leq 5$,

$$1.36 \leq \frac{\mathrm{EQB}(\hat{\alpha}_{TP})}{\mathrm{EQB}(\hat{\alpha}_{GTP})} \leq 39.03,$$

Table 11: EQB, $\rho = 0.7$.

	σ^2			
	0.01	0.05	0.1	0.2
TP	1.0332×10^{-5}	2.4312×10^{-4}	8.6973×10^{-4}	0.0031
GTP	7.5740×10^{-6}	7.4227×10^{-5}	1.3660×10^{-4}	2.4667×10^{-4}
	σ^2			
	0.3	0.4	0.5	0.6
TP	0.0060	0.0091	0.0125	0.0159
GTP	3.3316×10^{-4}	3.5397×10^{-4}	3.7043×10^{-4}	4.5644×10^{-4}
	σ^2			
	0.7	0.8	0.9	1
TP	0.0194	0.0229	0.0263	0.0293
GTP	5.6417×10^{-4}	6.2990×10^{-4}	7.3330×10^{-4}	7.5062×10^{-4}
	σ^2			
	1.25	1.5	1.75	2
TP	0.0361	0.0425	0.0482	0.0534
GTP	9.8979×10^{-4}	0.0013	0.0015	0.0019
	σ^2			
	2.5	3	4	5
TP	0.0621	0.0668	0.0796	0.0861
GTP	0.0026	0.0041	0.0059	0.0080

Table 12: EQB, $\rho = 0.8$.

	σ^2			
	0.01	0.05	0.1	0.2
TP	2.1964×10^{-5}	4.8116×10^{-4}	0.0017	0.0055
GTP	1.8833×10^{-5}	1.0045×10^{-4}	2.5618×10^{-4}	3.0485×10^{-4}
	σ^2			
	0.3	0.4	0.5	0.6
TP	0.0104	0.0153	0.0202	0.0254
GTP	3.9374×10^{-4}	4.9255×10^{-4}	4.9342×10^{-4}	5.8792×10^{-4}
	σ^2			
	0.7	0.8	0.9	1
TP	0.0305	0.0345	0.0386	0.0425
GTP	6.8475×10^{-4}	6.8511×10^{-4}	7.3412×10^{-4}	7.7025×10^{-4}
	σ^2			
	1.25	1.5	1.75	2
TP	0.0515	0.0592	0.0658	0.0715
GTP	9.9231×10^{-4}	0.0012	0.0014	0.0020
	σ^2			
	2.5	3	4	5
TP	0.0807	0.0872	0.0977	0.1060
GTP	0.0027	0.0031	0.0050	0.0068

and

$$mean\left(\frac{\mathrm{EQB}(\hat{\alpha}_{TP})}{\mathrm{EQB}(\hat{\alpha}_{GTP})}\right) = 23.76.$$

In Table 12, when $\rho = 0.8$, for $0.01 \leq \sigma^2 \leq 5$,

$$1.16 \leq \frac{\mathrm{EQB}(\hat{\alpha}_{TP})}{\mathrm{EQB}(\hat{\alpha}_{GTP})} \leq 55.18,$$

and

$$mean\left(\frac{\mathrm{EQB}(\hat{\alpha}_{TP})}{\mathrm{EQB}(\hat{\alpha}_{GTP})}\right) = 32.6.$$

Table 13: EQB, $\rho = 0.9$.

	σ^2			
	0.01	0.05	0.1	0.2
TP	7.1488×10^{-5}	0.0015	0.0049	0.0140
GTP	5.8316×10^{-5}	2.4421×10^{-4}	3.3347×10^{-4}	4.7426×10^{-4}
	σ^2			
	0.3	0.4	0.5	0.6
TP	0.0235	0.0324	0.0409	0.0478
GTP	4.9281×10^{-4}	5.0516×10^{-4}	5.7657×10^{-4}	5.8582×10^{-4}
	σ^2			
	0.7	0.8	0.9	1
TP	0.0542	0.0605	0.0649	0.0706
GTP	5.9802×10^{-4}	6.3600×10^{-4}	6.5398×10^{-4}	7.3646×10^{-4}
	σ^2			
	1.25	1.5	1.75	2
TP	0.0801	0.0883	0.0943	0.0993
GTP	9.5805×10^{-4}	0.0011	0.0012	0.0015
	σ^2			
	2.5	3	4	5
TP	0.1075	0.1144	0.1216	0.1277
GTP	0.0017	0.0025	0.0040	0.0054

In Table 13, when $\rho = 0.9$, for $0.01 \leq \sigma^2 \leq 5$,

$$1.23 \leq \frac{\mathrm{EQB}(\hat{\alpha}_{TP})}{\mathrm{EQB}(\hat{\alpha}_{GTP})} \leq 99.24,$$

and

$$mean\left(\frac{\mathrm{EQB}(\hat{\alpha}_{TP})}{\mathrm{EQB}(\hat{\alpha}_{GTP})}\right) = 58.43.$$

In Table 14, when $\rho = 0.95$, for $0.01 \leq \sigma^2 \leq 5$,

$$1.73 \leq \frac{\mathrm{EQB}(\hat{\alpha}_{TP})}{\mathrm{EQB}(\hat{\alpha}_{GTP})} \leq 128.94,$$

Table 14: EQB, $\rho = 0.95$.

	σ^2			
	0.01	0.05	0.1	0.2
TP	2.6593×10^{-4}	0.0047	0.0133	0.0313
GTP	1.5399×10^{-4}	3.7387×10^{-4}	4.3908×10^{-4}	4.6131×10^{-4}
	σ^2			
	0.3	0.4	0.5	0.6
TP	0.0468	0.0594	0.0691	0.0783
GTP	4.7877×10^{-4}	6.5256×10^{-4}	6.5815×10^{-4}	7.0489×10^{-4}
	σ^2			
	0.7	0.8	0.9	1
TP	0.0848	0.0902	0.0950	0.0993
GTP	7.2367×10^{-4}	7.4055×10^{-4}	7.7354×10^{-4}	9.1609×10^{-4}
	σ^2			
	1.25	1.5	1.75	2
TP	0.1082	0.1145	0.1197	0.1239
GTP	8.3913×10^{-4}	9.2828×10^{-4}	0.0013	0.0014
	σ^2			
	2.5	3	4	5
TP	0.1302	0.1337	0.1391	0.1429
GTP	0.0021	0.0024	0.0038	0.0050

and

$$mean\left(\frac{\text{EQB}(\hat{\alpha}_{TP})}{\text{EQB}(\hat{\alpha}_{GTP})}\right) = 80.16.$$

From Tables 3–14, it is concluded that, for each value of ρ, as σ^2 increases, the EMSE and EQB of the estimators will increase. From Tables 9–14, it is concluded that the relative superiority of the $\hat{\alpha}_{GTP}$ over the $\hat{\alpha}_{TP}$, in the sense of EQB, mostly increases as ρ increases.

7 Conclusion

In this paper, a two type parameter estimator was introduced and then its performance over the two-parameter (TP) estimator in terms of QB criterion was theoretically investigated and it was theoretically compared with the TP and OLS estimators in terms of MSEM criterion. Moreover, the estimation of the biasing parameters was presented, a numerical example was given, and a simulation study was done to compare the performance of the GTP estimator with the TP estimator in terms of EQB criterion, and to compare the performance with the TP and OLS estimators in terms of EMSE criterion.

References

[1] F. Akdeniz and H. Erol, *Mean squared error matrix comparisons of some biased estimators in linear regression*, Commun. Statist. Theor. Meth. **32** (2003) 2389–2413.

[2] R.W. Farebrother, *Further results on the mean square error of ridge regression*, J. Roy. Stat. Soc. B **38** (1976) 248–250.

[3] M.H.J. Gruber, *Improving Efficiency by Shrinkage: The James–Stein and Ridge Regression Estimators*, New york: Marcell Dekker (1998).

[4] A.E. Hoerl and R.W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics **12** (1970) 55–67.

[5] W. James and C. Stein, *Estimation with quadratic loss*, Fourth Berkely Symp. **1** (1961) 361–379.

[6] B.M.G. Kibria, *Performance of some new ridge regression estimators*, Commun. Statist. Simul. Computat. **32** (2003) 419–435.

[7] K. Liu, *A new class of biased estimate in linear regression*, Commun. Statist. Theor. Meth **22** (1993) 393–402.

[8] G.C. McDonald and D.I. Galarneau, *A Monte Carlo evaluation of some ridge-type estimators*, J. Amer. Statist. Assoc **20** (1975) 407–416.

[9] W.F. Massy, *Principal component regression in explanatory statistical research*, J. Amer. Statist. Assoc. **60** (1965) 234–256.

[10] M.R. Özakle and S. Kaçiranlar, *The restricted and unrestricted two-parameter estimator*, Commun. Statist. Theor. Meth. **36** (2007) 2707–2725.

[11] C. Stein, *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*, Proc. Third Berkely Symp. **1** (1956) 197–206.

[12] G. Trenkler and H. Toutenburg, *Mean squared error matrix comparisons between biased estimators: an overview of recent results*, Statist. Pap. **31** (1990) 165–179.