Iraqi Journal of Statistical Science (32) 2020

pp [37-52]

Generalized ridge estimator shrinkage estimation based on particle swarm optimization algorithm

Qamar Abdul Kareem* **Dr Zakariya Yahya Algamal****

qamar.kareem95@gmail.com **zakariya.algamal@uomosul.edu.iq**

Abstract:

It is well-known that in the presence of multicollinearity, the ridge estimator is an alternative to the ordinary least square (OLS) estimator. The generalized ridge estimator (GRE) is a generalization of the ridge estimator. However, the efficiency of the GRE depends on appropriately choosing the shrinkage parameter matrix which is involved in the GRE. In this paper, a particle swarm optimization method, which is a metaheuristic continuous algorithm, is proposed to estimate the shrinkage parameter matrix. The simulation study and real application results show the superior performance of the proposed method in terms of prediction error.

Keywords: Multicollinearity; shrinkage parameter; generalized ridge estimator; particle swarm optimization.

1. Introduction

Regression modeling is a widely applied strategy for studying several real data problems. In the linear regression model, the response variable is considered as continuous and reasonably assumed to follow the normal distribution. In linear regression models, it is assumed that the correlations among the explanatory variables are not high (Alheety and Kibria, 2014,Alkhamisi and Shukur, 2007,Dorugade, 2014). However, this assumption is not always held in practice. In the linear regression model, the ordinary least squares (OLS) estimator is the best estimator among all linear and unbiased estimators. However, under multicollinearity, the OLS estimator becomes unhelpful due to its large variance.

The ridge estimator (RE) (Hoerl and Kennard, 1970) has been consistently demonstrated to be an alternative to the OLS when multicollinearity exists. The RE can shrink all the regression coefficients toward zero to reduce the large variance (Asar and Genç, 2015,Algamal, 2018). The generalized ridge estimator (GRE) has also been considered as a generalization of the RE. The performance of the GRE fully depends on the values of the shrinkage parameter matrix. Accordingly, appropriate choosing the shrinkage parameter matrix is an important part of applying the GRE. Numerous approaches are available in the literature for estimating the shrinkage parameter (Khalaf and Shukur, 2005,Allen, 1974,Muniz and Kibria, 2009,Kibria, 2003,Hamed et al., 2013,Alkhamisi et al., 2006,Hefnawy and Farag, 2014).

In recent years, numerous nature-inspired algorithms have been successfully introduced and applied as random search strategies for solving a number of optimization problems. The Particle swarm optimization algorithm is a comparatively recent population-based algorithm that is inspired by swarm intelligence.

* Researcher / College of Computer science and Mathematics / Mosul University.
**Department of Statistics and Informatics, University of Mosul, Mosul, Iraq.

Received date:16 /4 /2020 **Accepted date: 4 /5 / 2020** **Published data 1/12/ 2020**

In this paper, the particle swarm optimization algorithm is proposed to estimate the values of the shrinkage parameter matrix in the GRE. Our proposed approach will efficiently help to find the best values with high prediction accuracy. The superiority of our proposed approach in different simulated examples and a real data application is proved.

2. Generalized ridge estimator

Suppose that we have a data set $\{(y_i, \mathbf{x}_i)\}_{i=1}^{n}$ where $y_i \in \mathbb{R}$ is a response variable and $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ip}) \in \mathbb{R}^p$ represents a p-dimensional explanatory variable vector. Without loss of generality, it is assumed that the response variable is centered and the explanatory variables are standardized.

Consider the following linear regression model,

$$\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of observations of the response variable, $\mathbf{X} = (\mathbf{x_1}, ..., \mathbf{x_p})$ is an $n \times p$ known design matrix of explanatory variables, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ is a $p \times 1$ vector of unknown regression coefficients, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors with mean 0 and variance σ^2. Using the OLS method, the parameter estimation of Eq. (1) is given by

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X^T X})^{-1} \mathbf{X^T y}. \tag{2}$$

The OLS estimator is unbiased and has minimum variance among all linear unbiased estimators. However, in the presence of multicollinearity, the $\mathbf{X^T X}$ matrix is nearly singular, which makes the OLS estimator unstable due to the large variance. To reduce the effects of the multicollinearity, the RE (Hoerl and Kennard, 1970), which is the most commonly used method, adds a positive shrinkage parameter, k, to the main diagonal of the $\mathbf{X^T X}$ matrix. The RE is defined as

$$\hat{\boldsymbol{\beta}}_{RE} = (\mathbf{X^T X} + k\ \mathbf{I})^{-1} \mathbf{X^T y}, \tag{3}$$

where $\mathbf{I}$ is the identity matrix with dimension $p \times p$. The estimator $\hat{\boldsymbol{\beta}}_{RE}$ is biased but more stable and has less mean square error. The shrinkage parameter, k, controls the shrinkage of $\boldsymbol{\beta}$ toward zero. The OLS estimator can be considered as a special estimator from the RE with $k = 0$. For a large value of k, the RE yields greater shrinkage approaching zero (Yang and Emura, 2017). Rewriting Eq. (1) as the canonical form introduced by [4], we obtain

$$\mathbf{y} = \mathbf{Z\alpha} + \boldsymbol{\varepsilon} \tag{4}$$

where $\mathbf{Z} = \mathbf{XW}$, $\mathbf{W}$ is a $p \times p$ matrix such that $\mathbf{Z^T Z} = \mathbf{W^T X^T X W}$ will imply $\mathbf{Z^T Z} = \boldsymbol{\Lambda} = diag(\lambda_1, \lambda_1, ..., \lambda_p)$ where $\boldsymbol{\Lambda}$ is a diagonal matrix with the Eigen values of $\mathbf{X^T X}$ and $\boldsymbol{\alpha} = \mathbf{W^T \beta}$. Then, the OLS estimator of $\boldsymbol{\alpha}$ is given by

$$\hat{\boldsymbol{\alpha}}_{\mathbf{LS}} = \boldsymbol{\Lambda}^{-1} \mathbf{Z^T y}. \tag{5}$$

Therefore the OLS estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \mathbf{W} \hat{\boldsymbol{\alpha}}_{\mathbf{LS}}. \tag{6}$$

The RE estimator given in Eq (3) is rewritten by as (Hoerl and Kennard, 1970)

$$\hat{\boldsymbol{\beta}}_{RE} = (\boldsymbol{\Lambda} + k\,\mathbf{I})^{-1}\mathbf{Z}^{\mathbf{T}}\mathbf{y} = \mathbf{A}^{-1}\mathbf{Z}^{\mathbf{T}}\mathbf{y} \tag{7}$$

where $\mathbf{A} = \boldsymbol{\Lambda} + k\,\mathbf{I}$.

Related to Eq. (3) and Eq.(5), the mean square error (MSE) is

$$MSE(\hat{\boldsymbol{\beta}}_{RE}) = \hat{\sigma}^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + k)^2} \tag{8}$$

The GRE is suggested by [4] to generalize the ridge estimator. Several researchers deal with the GRE (Loesgen, 1990,Trenkler and Toutenburg, 1990,Ohtani, 1995). The difference between the RE and the GRE is that there are p values of k for the GRE estimator such that (Hoerl and Kennard, 1970)

$$\hat{\boldsymbol{\beta}}_{GRE} = (\boldsymbol{\Lambda} + \mathbf{K})^{-1}\mathbf{Z}^{\mathbf{T}}\mathbf{y} = \mathbf{B}^{-1}\mathbf{Z}^{\mathbf{T}}\mathbf{y} \tag{9}$$

where $\mathbf{K} = diag(k_1, k_2, ..., k_p)$ and $\mathbf{B} = \boldsymbol{\Lambda} + \mathbf{K}$. The MSE of the GRE estimator, which is less than when using the RE and OLS, is

$$MSE(\hat{\boldsymbol{\beta}}_{GRE}) = \hat{\sigma}^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k_j)^2} + k_j^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + k_j)^2} \tag{10}$$

where $\hat{\sigma}^2 = \dfrac{\mathbf{y}^{\mathbf{T}}\mathbf{y} - \hat{\boldsymbol{\alpha}}_{\mathbf{OLS}}\mathbf{Z}\mathbf{y}}{n - p - 1}$ (Hoerl and Kennard, 1970).

Since the ridge parameter is the key to reducing the multicollinearity, there are several ways to determine this value. Researchers suggest several ways to choose the optimal k, such as (Hocking et al., 1976) (HK), (Nomura, 1988) (N), (Troskie and Chalton, 1996) (TC), (Firinguetti, 1999) (F), (Alkhamisi and Shukur, 2007) (HSL), (Batah et al., 2008) (SB), (Al-Hassan, 2010) (AH), (Dorugade and Kashid, 2010) (DK), (Månsson et al., 2010) (M), (Dorugade, 2014) (D), (Asar et al., 2014) (AS) and (Bhat and Raju, 2017) (SV1,SV2). These methods can be defined as:

$$\hat{k}_{j(HK)} = \frac{\hat{\sigma}^2}{\hat{\beta}_j^2} \tag{11}$$

$$\hat{k}_{j(N)} = \frac{\hat{\sigma}^2}{\hat{\beta}_j^2}\left\{1 + \left[1 + k_j(\hat{\beta}_j^2 / \hat{\sigma}^2)^{1/2}\right]\right\} \tag{12}$$

$$\hat{k}_{j(TC)} = \frac{k_j \hat{\sigma}^2}{k_j \hat{\beta}_j^2 + \hat{\sigma}^2} \tag{13}$$

$$\hat{k}_{j(F)} = \frac{k_j \hat{\sigma}^2}{k_j \hat{\beta}_j^2 + (n-p)\hat{\sigma}^2} \tag{14}$$

$$\hat{k}_{j(HSL)} = \hat{\sigma}^2 \frac{\sum_{j=1}^{p} (\mathrm{k}_j \hat{\beta}_j^2)^2}{(\sum_{j=1}^{p} (\mathrm{k}_j \hat{\beta}_j^2))^2} \tag{15}$$

$$\hat{k}_{j(AH)} = \hat{\sigma}^2 \frac{\sum_{j=1}^{p} (\mathrm{k}_j \hat{\beta}_j^2)^2}{(\sum_{j=1}^{p} (\mathrm{k}_j \hat{\beta}_j^2))^2} + \frac{1}{\lambda_{\max}} \tag{16}$$

$$\hat{k}_{j(D)} = \frac{\hat{\sigma}^2}{k_{\max} \hat{\beta}_j^2} \tag{17}$$

$$\hat{k}_{j(\mathrm{SB})} = \frac{k_j \hat{\sigma}^2}{k_j \hat{\beta}_j^2 + \hat{\sigma}^2} + \frac{1}{k_{\max}} \tag{18}$$

$$\hat{k}_{j(DK)} = Max\left(0, \frac{p\hat{\sigma}^2}{\hat{\beta}_j^2} - \frac{1}{n(VIF_j)_{Max}}\right) \tag{19}$$

$$\hat{k}_{j(\mathrm{SV1})} = \frac{p\hat{\sigma}^2}{\hat{\beta}_j^2} + \frac{1}{\lambda_{Max} \hat{\beta}_j^2} \tag{20}$$

$$\hat{k}_{j(\mathrm{SV2})} = \frac{p\hat{\sigma}^2}{\hat{\beta}_j^2} + \frac{1}{2\left(\sqrt{\lambda_{Max} / \lambda_{Min}}\right)^2} \tag{21}$$

$$\hat{k}_{j(M)} = \frac{1}{\frac{\lambda_{Max} \hat{\beta}_j^2}{(n-p)\hat{\sigma}^2 + \lambda_{Max} \hat{\beta}_j^2}} \tag{22}$$

$$\hat{k}_{j(AS)} = \frac{\hat{\sigma}^2}{\hat{\beta}_j^2} + \frac{1}{k_j} \tag{23}$$

3. The proposed methods

The efficiency of the ridge regression model strongly depends on appropriately choosing the shrinkage parameter. A choice of shrinkage parameter that is too small leads to overfitting the GRE, while a shrinkage parameter that is too large shrinks $\boldsymbol{\beta}$ by too much, making a bias-variance tradeoff.

Particle swarm optimization (PSO) is a nature- inspired metaheuristic algorithm that was originally proposed by Kennedy and Eberhart (1995) for solving continuous optimization problems.

PSO is inspired by the social or collective behavior of animals such as bird flocking and fish schooling. PSO, when compared with other computation intelligence-based algorithms, has several advantages, such as simple implementation, computationally higher efficiency, fewer parameters to tune, scalability and flexibility, robustness. For instance, compared with the genetic algorithm, there is no crossover and mutation genetic operation (Chen et al., 2014,Kiran, 2017,Lin et al., 2008,Lu et al., 2009,Zhou and Dickerson, 2014).

PSO performs the searching using a population, which is called swarm, of particles. Each particle has three features: (1) position, (2) velocity, and (3) fitness value. In PSO, each particle can be represented as a candidate solution (position) in the search space. The particles fly through the search space by their own efforts and in cooperation with other particles, and they follow the best solutions they have achieved (local best solutions), as well as tracking the best solutions that they found (the best global solution) (Cervantes et al., 2017,Lai et al., 2016,Mirjalili and Lewis, 2013,Wen et al., 2011).

Mathematically, the search space is assumed to be D-dimensional and there are m particles in the swarm where $d = 1, 2\ldots, D$. During the movement, each particle has a position vector $k_j = \{k_1, k_2, \ldots, k_p\}$ with a velocity vector $\mathbf{v}_k = \{v_{k1}, v_{k2}, \ldots, v_{kd}\}$. In the PSO algorithm, the best position, which gives the best fitness value for the particle k, is called best previous position and is denoted as $Pbest_k = \{Pbest_{k1}, Pbest_{k2}, ..., Pbest_{kd}\}$. The best position found by all particles is called the global best, which is denoted as $Gbest_k = \{Gbest_{k1}, Gbest_{k2}, ..., Gbest_{kd}\}$. In each iteration, the PSO algorithm searches for the optimal solution by updating the position and the velocity of the k^{th} particle according to the following two equations:

$$v_{kd}^{t+1} = z \times v_{kd}^{t} + c_1 \times r_1 \times \left(Pbest_{kd}^{t} - x_{kd}^{t}\right) + c_2 \times r_2 \times \left(Gbest_{kd}^{t} - x_{kd}^{t}\right), \quad (24)$$

$$x_{kd}^{t+1} = x_{kd}^{t} + v_{kd}^{t+1}, \quad (25)$$

where t denotes the iteration in the algorithm, and z is the inertia weight which is used to balance between the global search and the local search. In addition, c_1 (the cognition learning factor) and c_2 (social learning factor) are the acceleration coefficients. r_1 and r_2 are random values selected from a uniform distribution with (0,1).

In this paper, a PSO algorithm is proposed to determine the shrinkage parameter matrix. The proposed method will efficiently help to reduce the MSE. The parameter configurations for our proposed method are presented as follows.

1- The number of particles, m, is set to 50, and the number of iterations is $t_{max}=100$. The acceleration coefficients c_1 and c_2 are set within the range [2, 4]. The c_1 and c_2 are updated during the iteration according to the following equations:

$$c_1 = c_{1,\min} + \frac{t}{t_{\max}}(c_{1,\max} - c_{1,\min}), \quad (26)$$

$$c_2 = c_{2,\min} + \frac{t}{t_{\max}}(c_{2,\max} - c_{2,\min}). \quad (27)$$

Further, the minimum and maximum values for the inertial weight are $z_{\min} = 0.2$ and $z_{\max} = 0.9$. The inertial weight is updated according to the following equation:

$$z = z_{\max} - \frac{t}{t_{\max}}(z_{\max} - z_{\min}). \quad (28)$$

2- The positions of each particle are randomly determined. The position of a particle represents the shrinkage parameters, k_j. Here, the dimension of each particle is the number of explanatory variables. The initial positions of the particles are generated from a uniform distribution within the range [0,100].
3- The initial velocities of each particle are generated from a uniform distribution within the range [0, 4].
4- The fitness function which is the MSE as indicated in Eq (10) is calculated.
5- The velocities and positions are updated using Eq. (22) and Eq. (23), respectively.
6- Steps 4 and 5 are repeated until a $t_{\max}$ is reached.

4. Monte Carlo simulation results

In this section, a comprehensive simulation study was conducted to evaluate the performance of the proposed methods (Alheety and Kibria, 2014,Alkhamisi and Shukur, 2007,Hoerl and Kennard, 1970,Yang and Emura, 2017,Hocking, Speed and Lynn, 1976,Troskie and Chalton, 1996,Firinguetti, 1999,Batah, Ramanathan and Gore, 2008,Al-Hassan, 2010,Dorugade and Kashid, 2010,Månsson, Shukur and Golam Kibria, 2010,Asar, Karaibrahimoğlu and Genç, 2014,Bhat and Raju, 2017,Bhat et al., 2016,Bhat, 2016,Hocking, 1976). Following McDonald and Galarneau (1975), the explanatory variables with different degrees of multicollinearity are generated by

$$x_{ij} = (1-\rho^2)^{1l2} w_{ij} + \rho w_{ip}, \; i = 1,2,...,n, \quad j = 1,2,...,p, \quad (29)$$

where ρ^2 represents the correlation between the explanatory variables, and w_{ij} are independent standard normal pseudo-random numbers. The response variable is generated by

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \varepsilon_i, \quad (30)$$

where ε_i are independent and identically normal distributed pseudo-random numbers with zero mean and variance σ^2. Because the sample size has direct impact on the prediction accuracy, three representative values of the sample size are considered: 30, 50 and 150. In addition, the number of explanatory variables is considered as $p \in \{4,8,12\}$. Further, because we are interested in the effect of multicollinearity, in which the degrees of correlation are considered more important, three values of the pairwise correlation are considered with $\rho = \{0.90, 0.95, 0.99\}$. Besides, the value of σ^2 is 1.

For a combination of these different values of n, p, ρ, the generated data are repeated 5000 times and the averaged mean squared errors (MSE) is calculated as

$$MSE(\hat{\boldsymbol{\beta}}) = \frac{1}{5000}\sum_{r=1}^{5000}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \quad (31)$$

where $\hat{\boldsymbol{\beta}}$ is the ridge estimator obtained by the method. In addition, the bias is calculated as

$$bias = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \quad (32)$$

The MSE and bias values from the Monte Carlo simulation study are reported in Tables 1 – 3. Several results can be obtained as follows:

1- The simulation results indicate that the PSO method of selecting K is superior to the other used selection methods for all combinations of n, p, and ρ in terms of MSE. We can see that the PSO method has smaller MSE and significantly lower MSE than others.
2- It is seen from Tables 1 – 3 that the $\hat{\boldsymbol{\beta}}_{PSO}$ estimator using the PSO method is usually more efficient than the OLS estimator for all values of n, p and when multicollinearity is high or severe.
3- In terms of ρ values, there is an increase in the MSE values when the correlation degree increases regardless of the value of n and p.
4- Regarding the number of explanatory variables, it is easily seen that there is a negative impact of MSE, where there is an increase in their values when the p increases from four explanatory variables to twelve explanatory variables.
5- With respect to the value of n, the MSE values decrease when n increases, regardless of the value of ρ and p.
6- All the selection methods of K are superior to the OLS estimator in terms of MSE.
7- According to the bias results, it was clearly seen that the proposed methods yielded smallest bias among the other estimating methods.

Table 1: Average MSE when n=30

		OLS	PSO	HK	KN	TC	f	HSL	AH	D	SB	DK	SV1	SV2	AS	M
p=4	**r=0.90**	1.829	0.812	0.914	0.953	0.889	0.85	0.875	0.875	0.844	0.889	0.914	0.938	0.938	0.916	0.891
	r=0.95	1.908	0.824	0.978	1.008	0.958	0.923	0.951	0.952	0.93	0.958	0.977	1.001	1.001	0.986	0.973
	r=0.99	1.888	0.812	0.942	0.954	0.925	0.914	0.94	0.941	0.916	0.925	0.941	0.961	0.961	0.971	0.955
p=8	**r=0.90**	2.702	0.627	0.841	0.914	0.78	0.725	0.75	0.751	0.733	0.78	0.84	0.916	0.916	0.846	0.794
	r=0.95	2.722	0.710	0.849	0.908	0.784	0.74	0.786	0.786	0.741	0.784	0.848	0.918	0.918	0.866	0.805
	r=0.99	2.751	0.734	0.871	0.899	0.824	0.794	0.864	0.864	0.79	0.824	0.871	0.931	0.932	0.929	0.885
p=12	**r=0.90**	3.593	0.612	0.794	0.9	0.702	0.622	0.644	0.644	0.633	0.702	0.792	0.911	0.911	0.801	0.715
	r=0.95	3.635	0.626	0.795	0.87	0.713	0.656	0.705	0.706	0.66	0.713	0.794	0.911	0.911	0.82	0.741
	r=0.99	3.615	0.681	0.792	0.835	0.709	0.642	0.8	0.801	0.698	0.71	0.791	0.905	0.905	0.907	0.831

Table 2: Average MSE when n=50

		OLS	PSO	HK	KN	TC	f	HSL	AH	D	SB	DK	SV1	SV2	AS	M
p=4	**r=0.90**	1.926	0.912	1.006	1.044	0.976	0.93	0.947	0.947	0.941	0.976	1.006	1.028	1.028	1.007	0.982
	r=0.95	1.934	0.924	0.996	1.022	0.978	0.95	0.964	0.964	0.952	0.978	0.996	1.015	1.015	0.998	0.982
	r=0.99	1.918	0.932	0.959	0.968	0.947	0.937	0.96	0.96	0.935	0.947	0.958	0.971	0.971	0.969	0.96
p=8	**r=0.90**	2.826	0.732	0.929	1.008	0.874	0.829	0.839	0.839	0.84	0.874	0.928	0.992	0.992	0.931	0.891
	r=0.95	2.835	0.755	0.923	0.97	0.885	0.85	0.869	0.869	0.851	0.885	0.922	0.968	0.968	0.928	0.893
	r=0.99	2.819	0.762	0.905	0.927	0.87	0.846	0.888	0.888	0.845	0.87	0.904	0.945	0.945	0.928	0.902
p=12	**r=0.90**	3.765	0.711	0.894	0.98	0.829	0.774	0.797	0.797	0.783	0.829	0.894	0.975	0.975	0.897	0.831
	r=0.95	3.767	0.720	0.889	0.959	0.826	0.785	0.815	0.815	0.78	0.826	0.889	0.969	0.969	0.899	0.838
	r=0.99	3.755	0.736	0.86	0.896	0.806	0.772	0.855	0.856	0.773	0.806	0.859	0.94	0.94	0.906	0.859

Table 3: Average MSE when n=150

		OLS	PSO	HK	KN	TC	f	HSL	AH	D	SB	DK	SV1	SV2	AS	M
p=4	r=0.90	1.942	0.913	1.008	1.034	0.984	0.943	0.946	0.946	0.952	0.984	1.008	1.022	1.022	1.008	0.991
	r=0.95	1.982	0.923	1.013	1.029	1	0.983	0.994	0.994	0.987	1.001	1.013	1.022	1.022	1.013	1.006
	r=0.99	1.967	0.959	0.983	0.988	0.978	0.971	0.983	0.983	0.971	0.978	0.983	0.987	0.987	0.983	0.981
p=8	r=0.90	2.948	0.848	0.998	1.032	0.975	0.948	0.957	0.957	0.953	0.975	0.998	1.021	1.021	0.998	0.978
	r=0.95	2.939	0.871	0.971	1.003	0.955	0.94	0.951	0.951	0.942	0.955	0.971	0.994	0.994	0.971	0.956
	r=0.99	2.946	0.875	0.974	0.989	0.961	0.951	0.975	0.975	0.95	0.961	0.974	0.991	0.991	0.978	0.967
p=12	r=0.90	3.922	0.884	0.975	1.022	0.944	0.922	0.933	0.933	0.924	0.944	0.974	1.012	1.012	0.975	0.944
	r=0.95	3.925	0.917	0.971	1.009	0.946	0.926	0.942	0.942	0.927	0.946	0.971	1.003	1.003	0.972	0.947
	r=0.99	3.913	0.927	0.951	0.974	0.929	0.916	0.947	0.947	0.916	0.929	0.951	0.981	0.988	0.956	0.937

Table 4: Average bias when n=30

		PSO	HK	KN	TC	f	HSL	AH	D	SB	DK	SV1	SV2	AS	M
p=4	r=0.90	0.0183	0.0081	0.0092	0.0096	0.0089	0.0085	0.0088	0.0088	0.0085	0.0089	0.0092	0.0094	0.0094	0.0092
	r=0.95	0.0191	0.0083	0.0098	0.0101	0.0096	0.0093	0.0095	0.0095	0.0093	0.0096	0.0098	0.01	0.01	0.0099
	r=0.99	0.0189	0.0081	0.0094	0.0096	0.0093	0.0092	0.0094	0.0094	0.0092	0.0093	0.0094	0.0096	0.0096	0.0097
p=8	r=0.90	0.027	0.0063	0.0084	0.0092	0.0078	0.0073	0.0075	0.0075	0.0074	0.0078	0.0084	0.0092	0.0092	0.0085
	r=0.95	0.0272	0.0071	0.0085	0.0091	0.0079	0.0074	0.0079	0.0079	0.0074	0.0079	0.0085	0.0092	0.0092	0.0087
	r=0.99	0.0275	0.0074	0.0087	0.009	0.0083	0.008	0.0087	0.0087	0.0079	0.0083	0.0087	0.0093	0.0093	0.0093
p=12	r=0.90	0.036	0.0061	0.008	0.009	0.007	0.0062	0.0065	0.0065	0.0064	0.007	0.0079	0.0091	0.0091	0.008
	r=0.95	0.0364	0.0063	0.008	0.0087	0.0072	0.0066	0.0071	0.0071	0.0066	0.0072	0.008	0.0091	0.0091	0.0082
	r=0.99	0.0362	0.0068	0.0079	0.0084	0.0071	0.0064	0.008	0.008	0.007	0.0071	0.0079	0.0091	0.0091	0.0091

Table 5: Average bias when n=50

		PSO	HK	KN	TC	f	HSL	AH	D	SB	DK	SV1	SV2	AS	M
p=4	**r=0.90**	0.0193	0.0091	0.0101	0.0105	0.0098	0.0093	0.0095	0.0095	0.0094	0.0098	0.0101	0.0103	0.0103	0.0101
	r=0.95	0.0194	0.0093	0.01	0.0102	0.0098	0.0095	0.0097	0.0097	0.0095	0.0098	0.01	0.0102	0.0102	0.01
	r=0.99	0.0192	0.0093	0.0096	0.0097	0.0095	0.0094	0.0096	0.0096	0.0094	0.0095	0.0096	0.0097	0.0097	0.0097
p=8	**r=0.90**	0.0283	0.0073	0.0093	0.0101	0.0088	0.0083	0.0084	0.0084	0.0084	0.0088	0.0093	0.0099	0.0099	0.0093
	r=0.95	0.0284	0.0076	0.0093	0.0097	0.0089	0.0085	0.0087	0.0087	0.0085	0.0089	0.0092	0.0097	0.0097	0.0093
	r=0.99	0.0282	0.0076	0.0091	0.0093	0.0087	0.0085	0.0089	0.0089	0.0085	0.0087	0.0091	0.0095	0.0095	0.0093
p=12	**r=0.90**	0.0377	0.0071	0.009	0.0098	0.0083	0.0078	0.008	0.008	0.0079	0.0083	0.009	0.0098	0.0098	0.009
	r=0.95	0.0377	0.0072	0.0089	0.0096	0.0083	0.0079	0.0082	0.0082	0.0078	0.0083	0.0089	0.0097	0.0097	0.009
	r=0.99	0.0376	0.0074	0.0086	0.009	0.0081	0.0077	0.0086	0.0086	0.0078	0.0081	0.0086	0.0094	0.0094	0.0091

Table 6: Average bias when n=150

		PSO	HK	KN	TC	f	HSL	AH	D	SB	DK	SV1	SV2	AS	M
p=4	r=0.90	0.0194	0.0092	0.0101	0.0104	0.0099	0.0095	0.0095	0.0095	0.0095	0.0099	0.0101	0.0102	0.0102	0.0101
	r=0.95	0.0198	0.0093	0.0102	0.0103	0.01	0.0099	0.01	0.01	0.0099	0.01	0.0102	0.0102	0.0102	0.0102
	r=0.99	0.0197	0.0096	0.0099	0.0099	0.0098	0.0097	0.0099	0.0099	0.0097	0.0098	0.0099	0.0099	0.0099	0.0099
p=8	r=0.90	0.0295	0.0085	0.01	0.0103	0.0098	0.0095	0.0096	0.0096	0.0096	0.0098	0.01	0.0102	0.0102	0.01
	r=0.95	0.0294	0.0087	0.0097	0.0101	0.0096	0.0094	0.0095	0.0095	0.0094	0.0096	0.0097	0.01	0.01	0.0097
	r=0.99	0.0295	0.0088	0.0098	0.0099	0.0096	0.0095	0.0098	0.0098	0.0095	0.0096	0.0098	0.0099	0.0099	0.0098
p=12	r=0.90	0.0392	0.0089	0.0098	0.0102	0.0095	0.0092	0.0094	0.0094	0.0093	0.0095	0.0098	0.0101	0.0101	0.0098
	r=0.95	0.0393	0.0092	0.0097	0.0101	0.0095	0.0093	0.0094	0.0094	0.0093	0.0095	0.0097	0.0101	0.0101	0.0097
	r=0.99	0.0392	0.0093	0.0095	0.0098	0.0093	0.0092	0.0095	0.0095	0.0092	0.0093	0.0095	0.0098	0.0099	0.0096

5. Real application results

To evaluate the predictive performance of the proposed method and to compare its performance with the other methods used in a real data application, the Portland cement dataset is employed. The Portland cement dataset became a standard dataset to examine and to remedy multicollinearity (Woods et al., 1932,Chen and Emura, 2017). It was widely used by numerous researchers. This dataset comes from an experimental investigation of heat evolved during the setting and hardening of Portland cements of varied composition and the dependence of this heat on the percentages of four compounds in the clinkers from which the cement was produced. There are 13 observations of heat evolved in calories per gram of cement (y), tricalcium aluminate (x_1), tetracalcium silicate (x_2), tetracalcium alumino ferrite (x_3), and dicalcium silicate (x_4).

Before fitting the linear regression model, the explanatory variables and the response variable are standardized. Then, eigenvalues of $X'X$ matrix are calculated with $k_1 = 26.8284$, $k_2 = 18.9127$, $k_3 = 2.2392$, and $k_4 = 0.0194$ resulting in a condition number $\sqrt{k_1 / k_4} = 1376.8810$. Thus, multicollinearity exists. As a result, using the RE and the GRE will be more suitable than the OLS. The predictive performance for each method used is computed using the predicted MSE, $\text{PMSE} = (1/n)\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, and the results are given in Table 7.

It is apparent from Table 4 that there is an improvement of the predictive capability of PSO compared with the other methods used, where PSO significantly reduces the the PMSE. The reduction of MSE using PSO was 11.350%, 10.892%, 10.625%, 10.552%, 10.007%, 10.888%, 9.842%, 11.101%, 9.842%, 11.101%, 9.843%, 10.969%, 9.945%, 10.056%, 10.431%, and 10.213% compared with OLS, HK, KN, TC, f, HSL, AH, D, SB, DK, SV1, SV2, AS , and M, respectively.

Table 7: Real application results for the used methods

Method	PMSE
OLS	9303.049
PSO	8247.127
HK	9255.305
KN	9227.561
TC	9220.055
f	9164.227
HSL	9254.874
AH	9147.469
D	9277.002
SB	9147.553
DK	9263.284
SV1	9157.882
SV2	9169.254
AS	9207.598
M	9185.278

6. Conclusion

In this paper, a new shrinkage parameter selection of the generalized ridge estimator, which depends on employing the particle swarm optimization algorithm, was proposed. This proposed method allows us to handle multicollinearity with decreasing the variability of shrinkage parameter selection. Simulation and results demonstrate that the proposed method outperformed several classical methods. Furthermore, the results proved that the proposed method is more efficient than the method of Hoerl and Kennard (1970).

7. Acknowledgment

The authors are very grateful to the University of Mosul/ College of Computer Sciences and Mathematics for their provided facilities, which helped to improve the quality of this work.

References

Alheety M and Kibria BG. A generalized stochastic restricted ridge regression estimator. Communications in Statistics-Theory and Methods. 2014; 43: 4415-4427.

Alkhamisi MA and Shukur G. A Monte Carlo study of recent ridge parameters. Communications in Statistics—Simulation and Computation®. 2007; 36: 535-547.

Dorugade A. New ridge parameters for ridge regression. Journal of the Association of Arab Universities for Basic and Applied Sciences. 2014; 15: 94-99.

Hoerl AE and Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics. 1970; 12: 55-67.

Asar Y and Genç A. New shrinkage parameters for the Liu-type logistic estimators. Communications in Statistics - Simulation and Computation. 2015; 45: 1094-1103.

Algamal ZY. Shrinkage parameter selection via modified cross-validation approach for ridge regression model. Communications in Statistics - Simulation and Computation. 2018: 1-9.

Khalaf G and Shukur G. Choosing ridge parameter for regression problems. Communications in Statistics - Theory and Methods. 2005; 34: 1177-1182.

Allen DM. The relationship between variable selection and data agumentation and a method for prediction. technometrics. 1974; 16: 125-127.

Muniz G and Kibria BMG. On some ridge regression estimators: An empirical comparisons. Communications in Statistics - Simulation and Computation. 2009; 38: 621-630.

Kibria BMG. Performance of some new ridge regression estimators. Communications in Statistics - Simulation and Computation. 2003; 32: 419-435.

Hamed R,Hefnawy AEL and Farag A. Selection of the ridge parameter using mathematical programming. Communications in Statistics - Simulation and Computation. 2013; 42: 1409-1432.

Alkhamisi M,Khalaf G and Shukur G. Some modifications for choosing ridge parameters. Communications in Statistics - Theory and Methods. 2006; 35: 2005-2020.

Hefnawy AE and Farag A. A combined nonlinear programming model and Kibria method for choosing ridge parameter regression. Communications in Statistics - Simulation and Computation. 2014; 43: 1442-1470.

Yang S-P and Emura T. A Bayesian approach with generalized ridge estimation for high-dimensional regression and testing. Communications in Statistics-Simulation and Computation. 2017; 46: 6083-6105.

Loesgen K. A generalization and Bayesian interpretation of ridge-type estimators with good prior means. Statistical Papers. 1990; 31: 147.

Trenkler G and Toutenburg H. Mean squared error matrix comparisons between biased estimators—An overview of recent results. Statistical Papers. 1990; 31: 165.

Ohtani K. Generalized ridge regression estimators under the LINEX loss function. Statistical Papers. 1995; 36: 99-110.

Hocking RR,Speed F and Lynn M. A class of biased estimators in linear regression. Technometrics. 1976; 18: 425-437.

Nomura M. On the almost unbiased ridge regression estimator. Communications in Statistics-Simulation and Computation. 1988; 17: 729-743.

Troskie C and Chalton D. Multidimensional statistical analysis and theory of random matrices, Proceedings of the Sixth Lukacs Symposium, eds. Gupta, AK and VL Girko1996, pp 273-292.

Firinguetti L. A generalized ridge regression estimator and its finite sample properties: A generalized ridge regression estimator. Communications in Statistics-Theory and Methods. 1999; 28: 1217-1229.

Batah FSM,Ramanathan TV and Gore SD. The efficiency of modified jackknife and ridge type regression estimators: a comparison. Surveys in Mathematics & its Applications. 2008; 3.

Al-Hassan YM. Performance of a new ridge regression estimator. Journal of the Association of Arab Universities for Basic and Applied Sciences. 2010; 9: 23-26.

Dorugade A and Kashid D. Alternative method for choosing ridge parameter for regression. Applied Mathematical Sciences. 2010; 4: 447-456.

Månsson K,Shukur G and Golam Kibria B. A simulation study of some ridge regression estimators under different distributional assumptions. Communications in Statistics-Simulation and Computation. 2010; 39: 1639-1670.

Asar Y,Karaibrahimoğlu A and Genç A. Modified ridge regression parameters: A comparative Monte Carlo study. Hacettepe Journal of Mathematics and Statistics. 2014; 43: 827-841.

Bhat S and Raju V. A class of generalized ridge estimators. Communications in Statistics-Simulation and Computation. 2017; 46: 5105-5112.

Kennedy J and Eberhart RC. Particle swarm optimization. Proceedings of IEEE Conference on Neural Network. 1995; 4: 1942–1948.

Chen K-H,Wang K-J,Wang K-M and Angelia M-A. Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data. Applied Soft Computing. 2014; 24: 773-780.

Kiran MS. Particle swarm optimization with a new update mechanism. Applied Soft Computing. 2017; 60: 670-678.

Lin S-W,Ying K-C,Chen S-C and Lee Z-J. Particle swarm optimization for parameter determination and feature selection of support vector machines. Expert Systems with Applications. 2008; 35: 1817-1824.

Lu Y,Wang S,Li S and Zhou C. Particle swarm optimizer for variable weighting in clustering high-dimensional data. Machine Learning. 2009; 82: 43-70.

Zhou W and Dickerson JA. A novel class dependent feature selection method for cancer biomarker discovery. Computers in Biology and Medicine. 2014; 47: 66-75.

Cervantes J,Garcia-Lamont F,Rodriguez L,López A,Castilla JR and Trueba A. PSO-based method for SVM classification on skewed data sets. Neurocomputing. 2017; 228: 187-197.

Lai C-M,Yeh W-C and Chang C-Y. Gene selection using information gain and improved simplified swarm optimization. Neurocomputing. 2016; 218: 331-338.

Mirjalili S and Lewis A. S-shaped versus V-shaped transfer functions for binary Particle Swarm Optimization. Swarm and Evolutionary Computation. 2013; 9: 1-14.

Wen JH,Zhong KJ,Tang LJ,Jiang JH,Wu HL,Shen GL and Yu RQ. Adaptive variable-weighted support vector machine as optimized by particle swarm optimization algorithm with application of QSAR studies. Talanta. 2011; 84: 13-18.

Bhat S,Vidya R and Parameshwar VP. Maximum Likelihood Estimation of Parameters in a Mixture Model. Communications in Statistics-Simulation and Computation. 2016; 45: 1776-1784.

Bhat SS. A comparative study on the performance of new ridge estimators. Pakistan Journal of Statistics and Operation Research. 2016; 12: 317-325.

Hocking RR. A Biometrics invited paper. The analysis and selection of variables in linear regression. Biometrics. 1976; 32: 1-49.

McDonald GC and Galarneau DI. A Monte Carlo evaluation of some ridge-type estimators. Journal of the American Statistical Association. 1975; 70: 407-416.

Woods H,Steinour HH and Starke HR. Effect of composition of Portland cement on heat evolved during hardening. Industrial & Engineering Chemistry. 1932; 24: 1207-1214.

Chen A-C and Emura T. A modified Liu-type estimator with an intercept term under mixture experiments. Communications in Statistics-Theory and Methods. 2017; 46: 6645-6667.