Journal of Engineering and Applied Sciences 14 (10): 3345-3353, 2019
ISSN: 1816-949X

Using Particle Swarm Optimization Algorithm to Address the Multicollinearity Problem

Sabah Manfi Redha, Adila Abdullatif and Inaam Aboud Hussain
Department of Statistics, College of Administration and Economics,
University of Baghdad, Baghdad, Iraq

Abstract: The problem of the multicollinearity problem is one of the important problems that occur in the data which address the existence of the linear relationship between the independent variables. The aim of this study is to address the problem of the multicollinearity problem using particle swarm optimization algorithm or so-called intelligence of the squadron. The variables were generated with different sample sizes for small and large samples (10, 30, 100 and 200) as well as the correlation coefficients between the independent variables (0.85, 0.90 and 0.99) a program was written in MATLAB R 2013 a Version, 8.1, Na had reached the superiority of particle swarm optimization algorithm in all sizes and samples of all correlation coefficients, the comparison has been using the average error boxes (Mean Square Error (MSE)).

Key words: The multicollinearity problem, particle swarm optimization algorithm, multicollinearity, MATLAB, important problems, relationship

INTRODUCTION

The regression is one of the most important methods of statistical and most used in many fields of scientific and analytical and when using this method in the analysis of variables it is facing many of the problems because of the correlation variables have relations between them and affect the accuracy of the desired results including the problem of multi-linear multicollinarity which affect the estimation of the parameters of the model when independent variables are associated with a complete or almost complete linear relationship. This is because the inverse of the matrix can not be found because the matrix variable will have a value equal to zero. This problem is solved by using the particle swarm algorithm to eliminate the linear correlation between the independent variables and obtain accurate results. The particle swarm algorithm is one of the most modern and advanced algorithms inspired by nature, characterized by high efficiency and speed of performance.

The aim of research: The study aims to process the problem of linear multiplicity by comparing the methods of multicollinearity using the particle swarm algorithm to eliminate the linear correlation between the independent variables and also to obtain accurate results.

Theoretical side

Multicollinearity problem: This problem occurs when 2 or more independent variables are associated with a linear relationship. When the value of one variable is equal to all data and whether the data is taken by independent variables and the variable is dependent on the time series or intermittent the hypothesis is that there is no linear relationship or a nearly linear relationship between the independent variables. The number of estimated parameters should be less than the size of the problem observations (Kazem and Dulaimi, 1988):

$$\text{rank}\left(x\right) = k+1<n \tag{1}$$

That first scientist who discover the problem of linear multiplicity and the extent of its influence on the assessment of the parameters of the model is Fisher in 1934.

MATERIALS AND METHODS

Reasons of multicollinearity problem: If the linear model parameters is estimated, the result will be:

$$\hat{b} = \left(x'\,x\right)^{-1} x'\,y \tag{2}$$

Corresponding Author: Sabah Manfi Redha, Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq

where, the matrix (($X^{'X}$ P (k) requires finding the inverse of the matrix and the inverse matrix is equal to zero as a result of mathematical operations such as the division by zero, so, the data will reject the estimated model beause it contains a complete linear relationship between variables (Taha, 2014).

Testing of the existence of multicollinearity problem: There are several tests to find out the existence of multicollinearity problem such as (Taha, 2014):

- Frisch test
- Farrar-Glauber test

The processing of multicollinearity problem: There are several ways to process multicollinearity that process the existence of a linear relation between independent variables such as, Ridge Regression (RR) this method is one of the methods which is used to process a model has multicollinearity problem by adding a positive value which it's value is between zero and one to the diagonal elements of matrix ($X^{'X}$) and obtaining more accurate model capabilities (Hayawi, 2010). The method of combining time series with CT data the method of Primary Components (PC). The main components method for processing a model that has a multicollinearity problem is to convert the linked original variables to new unlinked variables andthese new variables called the main compounds.

Particle swarm optimization algorithm: It is a scientific technique developed to improve problem solutions and obtain approximated solutions for the optimal solution of these problems. It is a one of the most advanced developmental fields in the field of artificial intelligence. This algorithm is developed by the scientists Kennedy and Eberhart in 1995. It is based on the PSO idea of the behavioral and social behavior of swarm or groups of fish through the idea of searching for food. The swarm looking for food from one place to another and that some birds in the swarm have the ability to distinguish the smell of food is strong and effective with information about the best place where a supplier of food because some birds in the swarm send information between them during the research phase and search the best place for food. When the swarm of birds explore a good place which has a good quality of food they will take the advantage of this place to get the best food. So, the research of the algorithm will be with 2 operations, the search process and the process of redundancy for the best existing solutions within the specified search area. Thus, it can be used PSO algorithm to solve problems related to optimization problems and the problems that change over time, etc. because of the properties the algorithm that help to solve these problems (Rini *et al*., 2011).

Basic components of Particle Swarm Optimization (PSO): The bird squadron algorithm consists of the population of the squadron called particles and denotes n which consists of n = (n_1, n_2 and n_i) that move within the squad for the search area determined by the type of problem and the multiple dimensions and the search for good initial solutions particles are based on their own expertise and also rely on the experiments and experiments of their adjacent particles inside the squadron. The PSO algorithm is created from the number of randomly assigned particles in the search area and the squadron particles are based on velocity particle formation consisting of V_i^t = (V_1^t , V_ (2)^t, ..., V_i^t) and the position particle position which consists of X_i^t = (X_1^t, X_2^t, ..., X_i^t) as it is updated based on previous cases The best position of the particle itself is P_ (best, I)^t and the best position of the particles in the whole squadron and symbolized by G_ (best, I)^t and by the dimensions of problem d which consists of d = (d_1, d_2, ..., d_j) and the speed and position of each particle are adjusted according to the update equations shown as follows (Lee *et al*., 2008):

$$V_i^{t+1} = V_i^t+c_1\ r_1^t\ \left(P_{best,\ i}^t-X_i^t\right)+c_2\ r_2^t\left(G_{best,\ i}^t-X_i^t\right) \qquad (3)$$

$$X_i^{t+1} = X_i^t+V_i^{t+1} \qquad (4)$$

Where:

V_i^t = The particle i velocity of the particle represents the dimension j and the frequency t

X_i^t = The particle i position in the squared is the dimension j and the frequency t (c_1, c_2) the transaction constant represents the acceleration

c_1 = The cognitive component and

c_2 = The social component

r_1^t, r_2^t = The random numbers distributed by regular distribution are within the period (1, 0)

t = The frequency assigned by the type of problem

$P_{best,\ i}^t$ = The best position of a particle itself and is called the best local position

$G_{best,\ i}^t$ = The best position of particles i in the whole squadron and is called the best global position

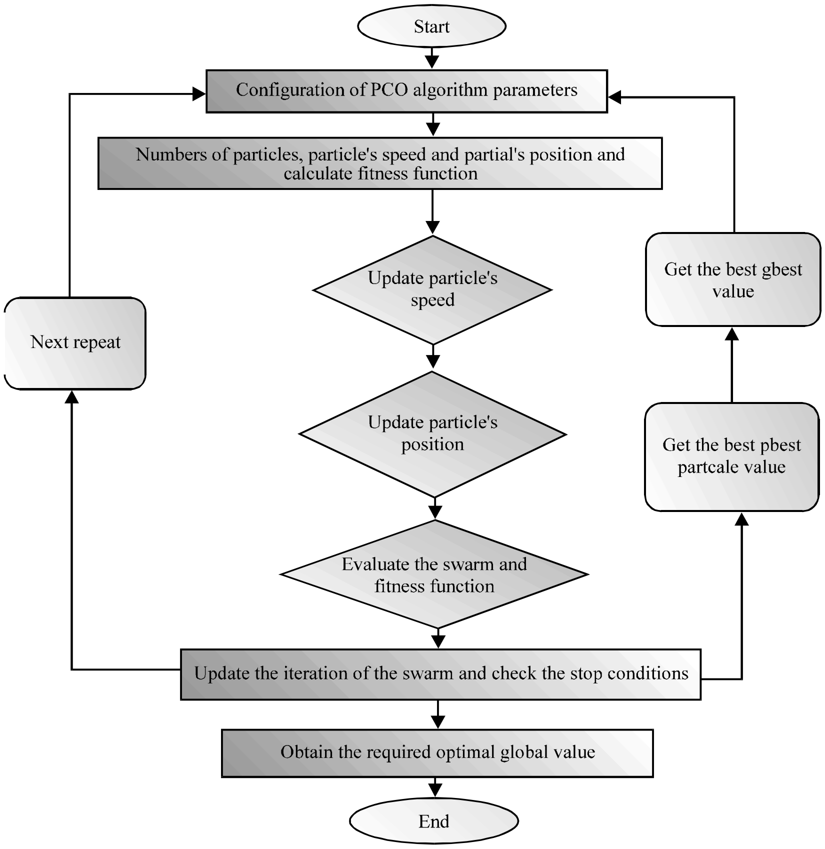
Start
Configuration of PCO algorithm parameters
Numbers of particles, particle's speed and partial's position and calculate fitness function
Update particle's speed
Update particle's position
Evaluate the swarm and fitness function
Update the iteration of the swarm and check the stop conditions
Obtain the required optimal global value
End
Next repeat
Get the best gbest value
Get the best pbest partcale value

Fig. 1: The basic procedures of the swarm

Kennedy and Ehrhardt have introduced the BPSO dual bird squadron algorithm the problem is that the velocity particle is converted to a probability and the position of the particle particle is converted to a probability that takes a value of zero or one (Lee *et al*., 2008) where it can be calculated from the equation shown as follows:

$$x_{ij}^{t} = \begin{cases} 1 \text{ if } u_{ij}^{t} < S_{ij}^{t} \\ 1 \text{ if } u_{ij}^{t} < S_{ij}^{t} \end{cases} \qquad (5)$$

where, u_{ij}^{t} is a random variable distributed according to the regular distribution within the period (0, 1) and S_{ij}^{t} is The x-function is used to convert the particle's velocity to a probability where it can be calculated from the equation as follows (Fig. 1):

$$S_{ij}^{t} = \frac{1}{1+e^{-v_{ij}^{t+1}}} \qquad (6)$$

RESULTS AND DISCUSSION

Parameters of Particle Swarm Optimization algorithm (PSO): There are basic parameters of the bird swarm algorithm affect their research and have a strong impact on the performance efficiency of the PSO algorithm (Satyobroto, 2011) and these parameters help to improve the search in the problem area the parameters are shown as follows, the size of the squadron, the number of particles bird flock and the number of particles should be appropriate to solve the problem and covers the search area and corresponds to the number of duplicates (Satyobroto, 2011).

Number of duplicates, represents the number of repetitions algorithm which is determined by the type of problem and after the number of repetitions specified this leads to obtain the desired results to solve the problem (Satyobroto, 2011).

Acceleration coefficients represent positive acceleration constants that help to maintain the cognitive and social components of particle velocity where C1 represents the cognitive component of the particle itself and C2 represents the social component of the particle with its vicinity (Fig. 2-4).

The weight of inertia represents the new parameter added to the PSO algorithm to improve its performance and efficiency. It was added by Shay and Eberhard for the particle speed update (Eq. 3) and symbolized by W (8) shown as follows:

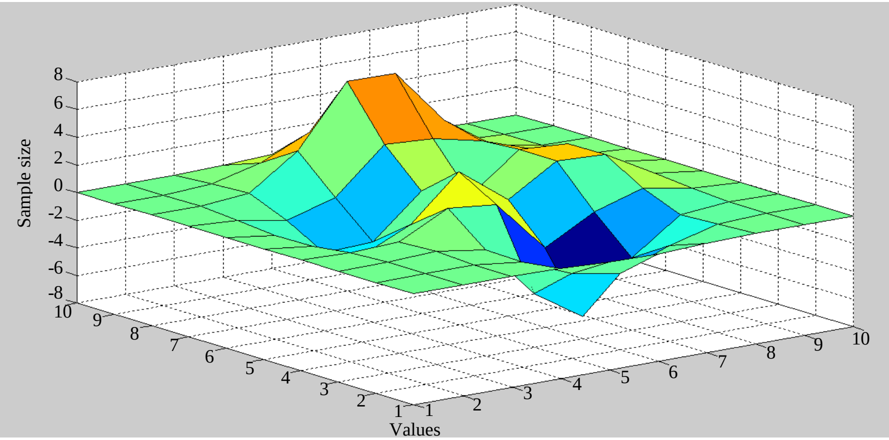

Fig. 2: Sample size n = 10 and row = 0.85; Best fitness ever found

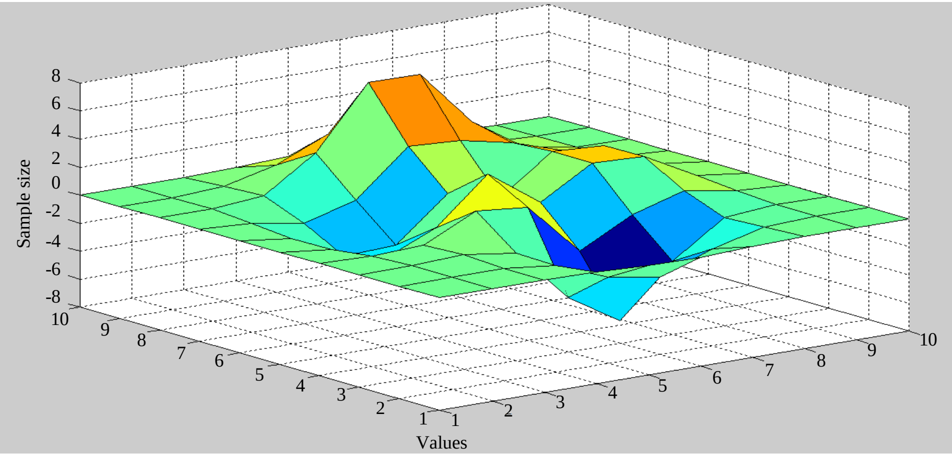

Fig. 3: Sample size n = 10 and row = 0.90; Best fitness ever found

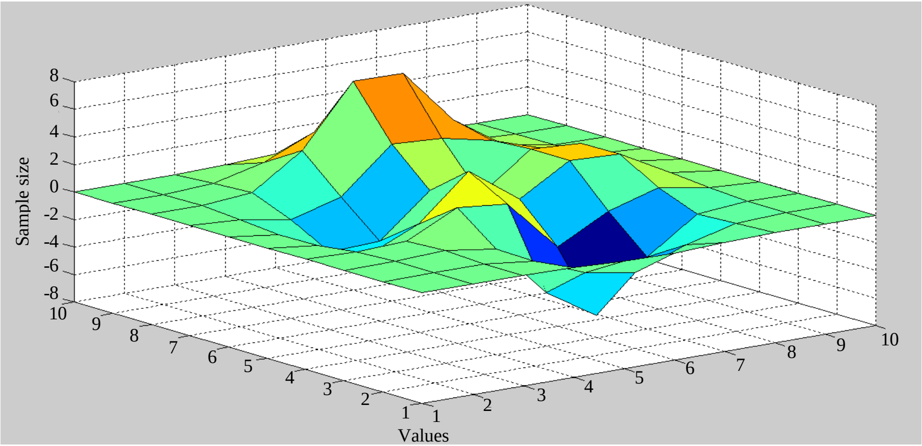

Fig. 4: Sample size n = 10 and row = 0.99; Best fitness ever found

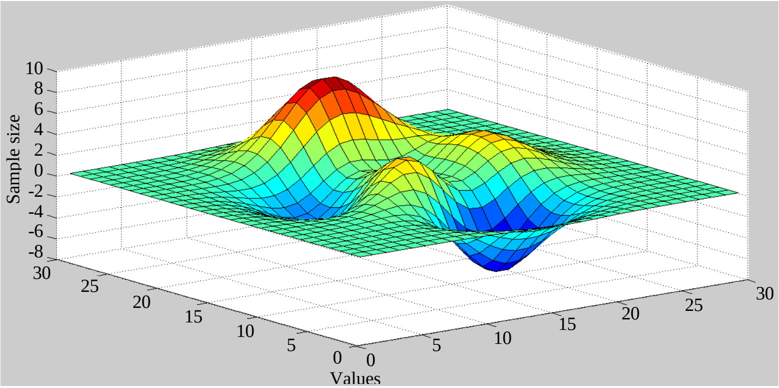

Fig. 5: Sample size n = 30 and row = 0.85; Best fitness ever found

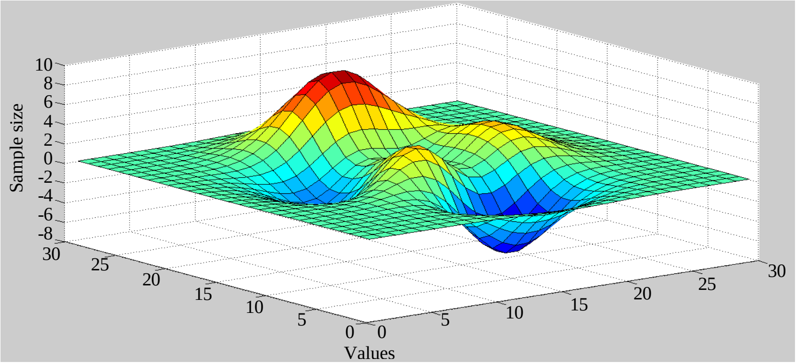

Fig. 6: Sample size n = 30 and row = 0.90; Best fitness ever found

$$V_i^{t+1} = W\ V_i^t + c_1\ r_1^t \left(P_{best,\ i}^t - X_i^t\right) + c_2\ r_2^t \left(P_{best,\ i}^t - X_i^t\right) \tag{7}$$

where, the value of the weight of inertia is calculated according to the equation shown as follows (Fig. 5 and 6):

$$W^{\,t+1} = W_{max} - \left(\frac{W_{max} - W_{min}}{T_{max}}\right)t, \quad W_{max} > W_{min} \tag{8}$$

Where:

W_{max} = The upper limit of the weight of the inertia
W_{min} = The minimum value of the inertia
t = The number of specific duplicates of the problem
T_{max} = The upper limit of the number of specified iterations

Coefficient: The new p arameter that was added to the particle velocity modernization Eq. 3 by the world by Clerk and symbolized by K has a very important utility for controlling the process of exploration and the process of exploiting the particles inside the squadron to ensure particle convergence (Satyobroto, 2011) as follows (Fig. 7-13):

$$V_i^{t+1} = K\left[V_i^t + c_1\ r_1^t \left(P_{best,\ i}^t - X_i^t\right) + c_2\ r_2^t \left(P_{best,\ i}^t - X_i^t\right)\right] \tag{9}$$

The value of K is calculated by Eq. 5:

$$K = \frac{2}{\left|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}\right|}, \ \varphi > 4 \tag{10}$$

$$\varphi = \varphi_1 + \varphi_2 \tag{11}$$

$$\varphi_1 = c_1 r_1 \tag{12}$$

$$\varphi_2 = c_2 r_2 \tag{13}$$

Tobologia (neighboring particles) represents a very important component developed by Kennedy

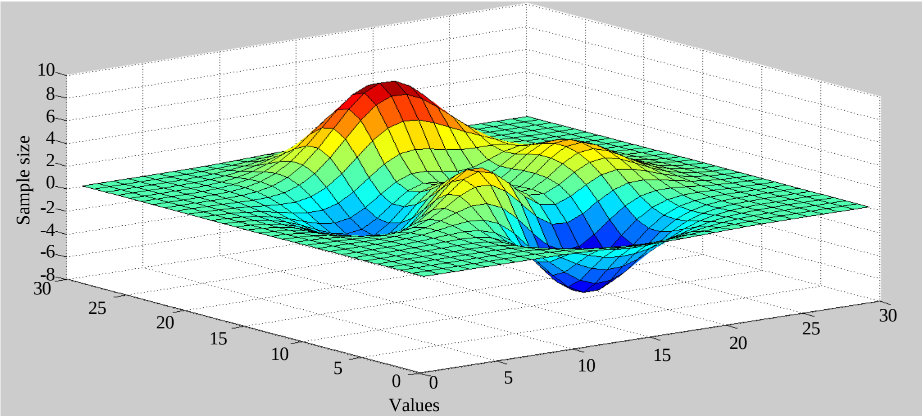

Fig. 7: Sample size n = 30 and row = 0.99; Best fitness ever found

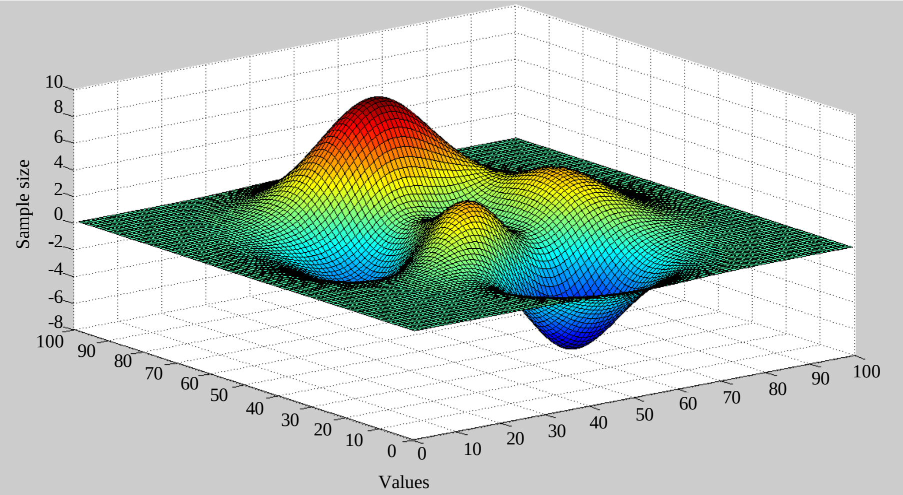

Fig. 8: Sample size n = 100 and row = 0.85; Best fitness ever found

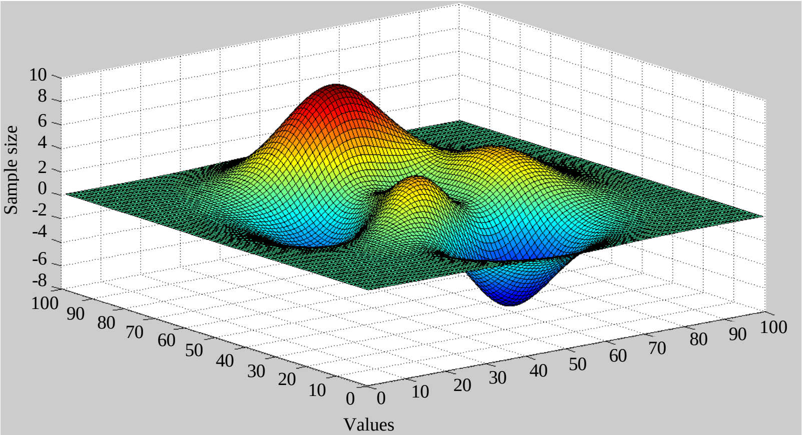

Fig. 9: Sample size n = 100 and row = 0.90; Best fitness ever found

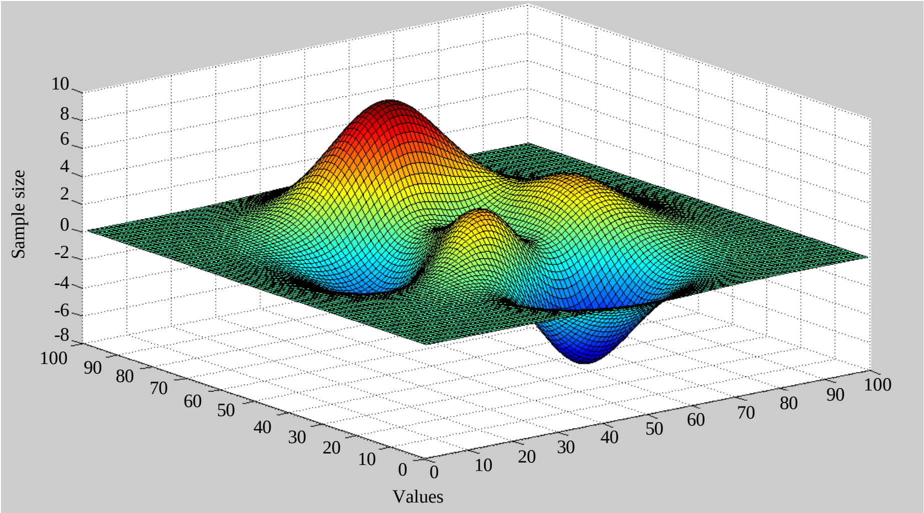

Fig. 10: Sample size n = 100 and row = 0.99; Best fitness ever found

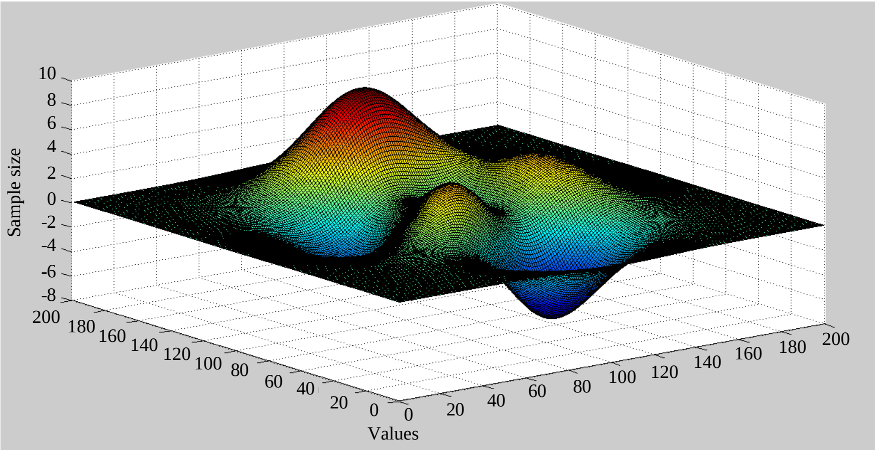

Fig. 11: Sample size n = 200 and row = 0.85; Best fitness ever found

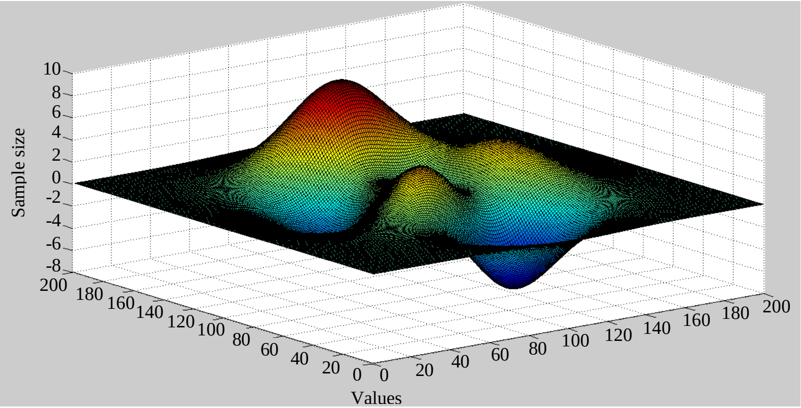

Fig. 12: Sample size n = 200 and row = 0.90; Best fitness ever found

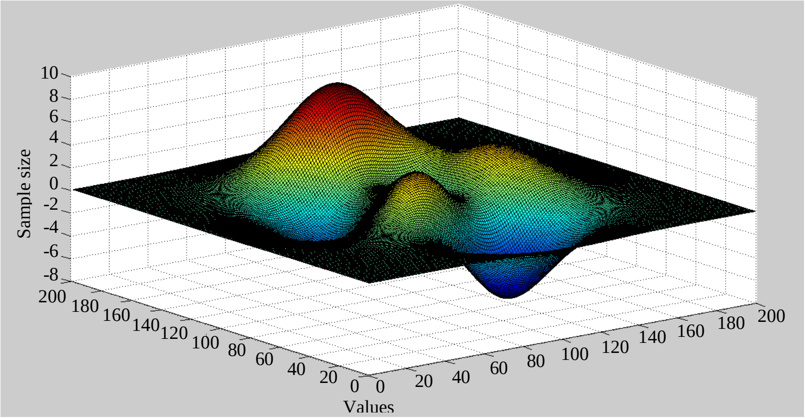

Fig. 13: Represents (sample size n = 200 and row = 0.99); Best fitness ever found

Table 1: Represents (mse values when n = 10 and sigma = 0.5)

PSObest	PSO (RR)	RR	PSO (PC)	PC	PSO (OLS)	OLS	Row
PSO (RR)	0.0121	0.0330	1.4960	1.7554	3.2583	3.5108	0.85
PSO (RR)	0.0123	0.0379	1.4960	1.7554	3.1045	3.8619	0.90
PSO (RR)	0.0089	0.0577	2.2329	3.5108	7.9985	9.4793	0.99

Table 2: Represents (mse values when n = 30 and sigma = 0.5)

Psobest	PSO (RR)	RR	PSO (PC)	PC	PSO (OLS)	OLS	Row
PSO (RR)	0.0284	0.0387	1.2654	3.4459	0.0503	1.9256	0.85
PSO (RR)	0.0285	0.0390	0.8856	1.0135	3.5798	8.1079	0.90
PSO (RR)	0.0282	0.0438	6.1274	7.0944	2.9690	4.7127	0.99

Table 3: Represents (mse values when n = 100 and sigma = 0.5)

Psobest	PSO (RR)	RR	PSO (PC)	PC	PSO (OLS)	OLS	Row
PSO (RR)	0.0021	0.0021	1.1957	2.8866	4.2050	8.8818	0.85
PSO (RR)	0.0021	0.0021	0.0375	1.3323	2.1092	2.4425	0.90
PSO (RR)	0.0018	0.0018	2.9852	5.1070	4.9914	6.6613	0.99

Table 4: Represents (mse values when n = 200 and sigma = 0.5)

PSObest	PSO (RR)	RR	PSO (Pc)	PC	PSO (OLS)	OLS	Row
PSO (RR)	0.0019	0.0019	5.9968	7.2617	4.9465	6.2804	0.85
PSO (RR)	0.0018	0.0018	2.4807	3.7682	0.5362	1.0088	0.90
PSO (RR)	0.0016	0.0016	0.7954	1.0520	0.5362	1.0088	0.99

because it determines how the spread of particles in the area of research and there are three topology next to the main used in PSO are the forms (the wheel, the circle or ring, stars) (Sun *et al*., 2011).

The basic procedures of Particle Swarm Optimization (PSO): The basic procedures of particle swarm optimization algorithm can be summarized in several steps as shown in Fig. 1 as follows (Satyobroto, 2011; Sun *et al*., 2011).

Practical side: This aspect of the study deals with the comparison between the multicollinearity methods and particle swarm optimization algorithm. The data were obtained by using simulations at different sample sizes (10, 30, 100 and 200). A program was programmed with MATLAB R2013aVersion 8.1 for this purpose, described as follows. When the sample size is 10, the results are shown in Table 1 when the sample size is 30, the results are shown in Table 2 when the sample size is 100, the results are shown in Table 3 when the sample size is 200, the results are shown in Table 4.

CONCLUSION

The results of the algorithm are better than the results of the methods used to solve the problem of multicollinearity. At the correlation coefficient (0.85, 0.90 and 0.99) and the sample sizes (10, 30, 100 and 200) the best results of particle swarm optimization were found in the ridge regression method at the correlation coefficient (0.99).

RECOMMENDATIONS

We recommend using advanced algorithms to solve and process the problem multicollinearity because it gives the best results in a fast high resolution and speed.

REFERENCES

Hayawi, H.A.M., 2010. Estimation of case-space models using the method of regression of character with application. Iraqi J. Stat. Sci., 18: 155-176.

Kazem, A.H. and M.M.A. Dulaimi, 1988. Introduction to Linear Regression Analysis. University of Baghdad, Baghdad, Iraq,.

Lee, S., S. Soak, S. Oh, W. Pedrycz and M. Jeon, 2008. Modified binary particle swarm optimization. Prog. Nat. Sci., 18: 1161-1166.

Rini, D.P., S.M. Shamsuddin and S.S. Yuhaniz, 2011. Particle swarm optimization: Technique, system and challenges. Intl. J. Comput. Appl., 14: 19-26.

Satyobroto, T., 2011. Mathematicle modelling and applications of particle swarm optimization. Master Thesis, Blekinge Institute of Technology, University in Karlskrona, Sweden.

Sun, J., C.H. Lai and X.J. Wu, 2011. Particle Swarm Optimisation: Classical and Quantum Perspectives. CRC Press, Boca Raton, Florida, USA., ISBN:9781439835760, Pages: 419.

Taha, M.M.A., 2014. The Treatment of the Problem of Multiple Linear Interference of the Nile Blue Company for Packaging and Printing using the Regression of the Letter (1986-2010). University of Sudan, Khartoum, Sudan,.