# Comparison Slice Inverse Regression Method with Machine Learning Techniques in Multivariate Data

**Omar A. abd Alwahab**[*1]

[1] Statistic Department, College of Administration and Economics, University of Diyala, Iraq
Email: omaradil.d87@gmail.com

**Abstract**

In this study, the research aims to use some methods that deal with several independent variables with a dependent variable, where two methods were used to deal with, which is the method of slice inverse regression (SIR), which is considered a non-classical method, and two methods of machine learning, which is (TLBO, PSO), which is most popular of the teaching methods machine learning, the work of (SIR), (TLBO, PSO) is based on making reduced linear combinations of a partial set of the original explanatory variables, which may suffer from the problem of heterogeneity and the problem of multicollinearity between most of the explanatory variables. These new combinations of linear compounds resulting from the two methods will reduce the largest number of explanatory variables to reach one or more new dimensions called the effective dimension. The root mean square error criterion will be used to compare the two methods to indicate the preference of the methods.

**Keywords:** multivariate; Slice inverse regression; machine learning; Projection Pursuit

## 1. Introduction

When we intend to study and analyse data of economic, medical, agricultural, financial, and other phenomena, prior knowledge of these phenomena must be available, in other words, knowledge of the type of their data, which is often quantitative, and this requires building an appropriate mathematical model that represents causal relationships (causal or behavioural function)[17]. Among its factors is the best representation, which is called the description stage, in order to adopt the appropriate analysis, which then enables us to make many decisions regarding the most important indications and characteristics (characteristics) related to those phenomena, and those indications mentioned above are called parameters[18]. One of the most important models statistical analysis is what is called the analysis of regression models, and there are two different approaches to dealing with these models, and for each approach or method there are conditions or restrictions[12]. The first method is: the method of parametric regression, which assumes that the sample comes from a specific population from one of the known probability distributions[4], and the parameters are estimated using different methods because of the ease of inferential operations and the strength of the parametric tests, the parametric methods remained dominant in regression analysis for a period of time[7]. The erroneous assumption of the parametric distribution leads to erroneous conclusions and inconsistent estimates, as well as because it does not fit complex data. For these reasons, researchers resort to the second method, which is the non-parametric or semi-parametric method, to analyse data as well as complex data and to assess the legitimacy of the supposed parametric model[6]. The crowding of data in the spaces represented by it with the limited variables it represents, then the traditional methods will fail to find a good estimation of the parameters, so this problem must be dealt with directly[15]. In order to do so, this problem must be eliminated (or at least mitigated) by finding appropriate solutions to it through the use of some methods that lead to obtaining accurate results[14]. Often, methods that integrate (or compress) variables are used without loss of any information from the data, and this is called Dimensionality Reduction: RD. The

DOI: https://doi.org/10.54216/IJNS.200204
Received: May 17, 2022 Accepted: January 18, 2023

common goal of all these methods used is to reduce the dimensions of the data (or compress it), while preserving the content of the information inherent in it, regardless of the methods of analysing it and drawing conclusions from it[16].

**2. Parametric Regression Model:**

The Parametric Regression Model (PRM) model indicates the existence of a relationship between two variables or several X's independent variable or explanatory variable[3]. Which is believed to affect the dependent variable Y or the response, and this model requires the availability of Certain conditions in the data, such as: knowledge of the distribution of observations and the distribution of errors, or prior knowledge of the indicative formula that controls the causal relationship between variables. The model is represented by the following formula[9]:

$$f(X_i, \beta) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_z X_{iz} + \epsilon_i \qquad (1)$$

$f(X_i, \beta)$: represents a linear function of the independent variables.

β : represent unknown parameters.

$\epsilon_i$ : represents the random error and assuming $E(\epsilon)=0$ , var $(\epsilon)=\sigma^2 I$

**3. Nonparametric Regression Model**

The (NPRM) model enjoys high flexibility, as it does not require the availability of conditions as in the parametric regression model, which made the non-parametric regression model desirable for researchers, because real data are not always ideal, as the model was represented by the following formula[19]:

$$y_i = m(\chi_i) + \epsilon_i \qquad (2)$$

whereas :

$y_i$ : represents the response variable.

m ($\chi_i$): represents a suitable smoothing function, which does not contain parameters and is estimated by a non- parametric method.

$\epsilon_i$ : represents the random error, with mean $E(\epsilon)=0$ , and variance var $(\epsilon)=\sigma^2$ .

**4. variables reduction**

In the field of parametric or non-parametric regression, when the number of observations or data is very large[6], we need to use certain methods to deal with it, and to facilitate dealing with high-dimensional data and working on dimension reduction[8]. Where there are two methods of reduction, the first method: it increasing the independent variables that do not It affects the response variable[14], thus deleting it from the model and making its coefficient equal to zero ($\beta = 0$), and thus reducing the number of variables, and this is of course not possible when the relationships between the variables are not known in the first place[2]. As for the second method: it does not delete any of the variables categorically, but rather tries to keep as much information as possible for the purpose of benefiting from it, so it intends to use it within a linear component on the original data, in other words, the original data will be replaced by linear combinations of the original variables, and the method that[3]. These linear structures are chosen based on the method of reducing or reducing the dimensions used[12].

**5. Slice Inverse Regression Method**

The basis of the inverse slide regression (SIR) method is the inversion of the causal relationship in the traditional (classical) regression analysis to study the relationship between the response variable (Y) and the explanatory variable (X) represented by E(Y/X)[9]. Which represents the regression of the output variable (Y) against many input variables (X) to the regression of E(X/Y), i.e. making the variable (Y) represent the independent variable and the variable (X) represent the dependent variable[2]. This method works on decomposing the model into multiple segments according to the values of (Y) and then carrying out different statistical operations for each segment, as it works on integrating (composition) the information of all segments and obtaining the latent roots from which the largest is chosen to represent the effective trends (e.d.r). of (SIR), and its dispersion is proportional to the dispersion of the original (X)

DOI: https://doi.org/10.54216/IJNS.200204

Received: May 17, 2022 Accepted: January 18, 2023

variables, and the inverse regression curve cannot be straight, and this curvature plays an important role in finding trends (e.d.r)[18]. But if the inverse regression curve is straight, then we may not be able to find more than one direction[16].

The Sliced Inverse Regression (SIR) method is based on finding effective trend estimates that serve as parameters (Bk's) in which the data is transformed into the reduced form and replaces the original data for ease of dealing with it[3]. Sliced inverse (SIR) is more like a nonparametric and semiparametric regression model as it is represented by the following formula[2]:

$$y = f(\beta'_1 X, \beta'_2 X, \cdots, \beta'_k X, \epsilon_i) \quad \dots \quad (3)$$

whereas:

f : represents an unknown function.

$B_k$: represent the parameters of the reduced form, and that: K=1 , …, m .

X: information matrix of degree (n*p).

n: represents the number of witnesses.

P: the number of explanatory variables.

$\epsilon_i$: represents the random error and is independent of X.

The slice inverse regression method is based on a set of transformations and computational methods that are summarized in the following algorithm, as follows:

### 5.1 Basic Algorithm (SIR)

This algorithm is summarized in the following steps[9]:

1- Entering the explanatory variables (X) and the dependent variable (Y), as each row in the table represents an observation.

2- Arranging the explanatory variables X and the dependent variable Y in ascending order according to the dependent variable (Y).

Then, the general arithmetic mean and the general covariance and covariance matrix are calculated as follows:

$$\overline{X}_i = \frac{\sum_{j=1}^{N} X_{ij}}{N} \quad (4)$$

$$\hat{\Sigma}_x = \frac{\sum_i^n (X_i - \overline{X})(X_i - \overline{X})'}{N} \quad (5)$$

3- The matrix is divided horizontally into H, then the arithmetic mean is calculated for each segment as follows:

$$\bar{x}h = \frac{\sum_{i=1}^{nh} xhi}{nh} \dots \quad (6)$$

Whereas :

$\bar{x}h$: represents the mean of the segment h .

$n_h$: represents the slice size h.

The variance, variance and covariance matrix of the matrix are calculated as the averages of the explanatory variables $\Sigma_{\bar{x}}$and the matrix is calculated according to the following formula:

$$\hat{\Sigma}_{\bar{X}_h} = \frac{\sum_{h=1}^{H} N_h(\bar{x}_h - \bar{X})(\bar{x}_h - \bar{X})'}{N} \quad (7)$$

DOI: https://doi.org/10.54216/IJNS.200204

Received: May 17, 2022 Accepted: January 18, 2023

$$\sum_{i=1}^{N} \frac{X_i}{N} = \bar{X} \quad , \quad \sum_{i=1}^{nh} \frac{X_{hi}}{nh} = \bar{X}_h \qquad (8)$$

The values and potential vectors of the matrix $\Sigma_{\bar{x}}$ are calculated, which we will call $\Sigma_{\eta}$, where the column is selected from the matrix resulting from the formation of the potential vectors, which will be called $\eta_{k}$, and k represent the number of columns of the latent vector matrix corresponding to the potential value that is greater than or equal to (0.5).

6- The matrix $\Sigma_{x}^{-1/2}$ is found by the following formula:

$$\Sigma_{x}^{-1/2} = D * \Sigma_{\eta}^{-1/2} * D' \qquad \dots \qquad (9)$$

whereas:

D: represents the eigenvector matrix.

$\Sigma_{\eta}$ : represents the diagonal latent value matrix.

7- $\hat{B}_{k}$ is calculated by :

$$\hat{\beta}_{k} = \Sigma_{x}^{-1/2} * \eta_{k} \qquad (10)$$

whereas :

$B_{j}$: represents the new SIR vector.

8- Calculation of E(X|y) by (ols) method.

### 5.2 Modified (SIR) Algorithm

It was suggested that the average potential vectors be calculated based on the covariance matrix and the covariance of each segment, then take the highest potential vector that corresponds to the highest eigenvalue, then multiply by the original (X) matrix. This algorithm is summarized in the following steps:

1- Entering the explanatory variables (X) and the dependent variable (Y) , where each row in the data represents an observation.

2- Arranging the explanatory variables (X) and the dependent variable (Y) in ascending order according to the dependent variable (Y), then the general arithmetic mean, the variance matrix and the general covariance are calculated as in equations (4) and (5).

3- The matrix is divided horizontally into (H) slice, and then the arithmetic mean is calculated for each slice.

4- The covariance matrix and the covariance matrix are calculated for each slice (H) from formula:

$$\hat{V}_{h} = \frac{\sum_{y_i \epsilon I_h} (\tilde{x}_i - \hat{m}_h)(\tilde{x}_i - \hat{m}_h)'}{n\hat{p}_h - 1} \qquad (11)$$

5- The latent values (Eigen values) and the latent vectors (Eigen vectors) are calculated for the matrix ($\hat{V}_h$) for each slice, which we will call ($\Sigma_{\eta}$), where the column is selected from the matrix (eigen vectors), which we will call ($\eta_{k}$) and that (k) represents a number Matrix columns (Eigen vectors) corresponding to the value (Eigen values) that is greater than or equal to (0.5).

6- A matrix ($\Sigma_{x}^{-1/2}$) is found through equation (9).

7- $\hat{B}j$ is calculated through equation (10) where $B\hat{}_{j}$ represents the new (SIR) vector.

8- Calculation of E(X|y) by (ols) method.

## 6. projection pursuit regression method

DOI: https://doi.org/10.54216/IJNS.200204
Received: May 17, 2022 Accepted: January 18, 2023

It is necessary to know first the meaning of Projection Pursuit, before talking about the method of regression of successive projections (PPR)[11]. Machine learning (ML) algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so[6]. As the proposed projection (PP) was proposed for the first time by the scientist (Kruska, 1969) and this method is used in order to find the important projections For the data[20]. That the term projection indicates that we are looking for the data that has been projected, and the term Pursuit means continuing successively to find a good projection for the purpose of performing the regression until obtaining the least possible error, the method of successive projections (PP) was used It is widely used in high-dimensional data in order to separate data sources randomly[15]. Thus be important in analyzing independent compounds, and the successive projection algorithm (PP) does not neglect any of the variables that have a weak relationship to the subject, i.e. (variables with Weak information), and this data is analyzed after it is divided into groups, clusters, or surfaces[10]. Its night separately and each separately, and one of the most important advantages of (PP) is that it is suitable for scattered data sets (scattered) in high-dimensional space [5].

The scientist (Friedman, 1981) later expanded the idea and added successive projection regression (PPR)[20], and the (PPR) method is considered one of the statistical techniques that are concerned with finding the most important projections in the multidimensional data, and with finding each projection the data shrinks by means of the composite[21]. The length of the projection, and the process is repeated to find good projections until the best projections are obtained, and the most important advantage of it is that it is one of the few methods that can overcome the problem of dimensionality or the curse of dimensions (curse of dimensionality) resulting from the space of high dimensions, and that the basic idea of the successive projection regression (PPR) is to model the surface of the regression as a sum of non-linear functions of the linear combinations of the variables, which is expressed by the following function[19] :

$$Y = \sum_{j=1}^{M} \vartheta_j(\beta_j' X) \qquad (12)$$

Since:

Y: represents the response variable.

$B_j$: represents the parameters, where j=1,…,M.

$\vartheta_j$: represents a specific smoothing function.

X: represents the matrix of explanatory variables with a degree (n*p), as:

P: represents the number of explanatory variables.

n: represents the sample size.

Many of the classic (classical) multivariate analysis methods are special cases of the PPR method, and examples of them are the principal component analysis (principal analysis) and the discriminant analysis. PPR is a generalization of the Additive Model, and it deals with models that are in the following form:

$$Y = \beta_0 + \sum_{j=1}^{M} \vartheta_j\left(\beta_j' X\right) + \epsilon_i \qquad (13)$$

Since:

Y: represents the (dependent) response variable.

$\beta_0$: represents the constant limit.

$\Theta_j$: represents a specific smoothing function (function).

$\beta_j' X$: It is a projection (X) on the vector unit (P) and that ($\beta_j$) are vectors (representing the linear structures of the explanatory variable) and that each vector contains unknown parameters to obtain the best model.

DOI: https://doi.org/10.54216/IJNS.200204

Received: May 17, 2022 Accepted: January 18, 2023

M: represents the number of smoothed functions that are used to construct the explanatory variable.

$\epsilon_j$: represents the random error.

Therefore, the (PPR) method is considered a general method because it is suitable for non-linear functions of any linear combination of (X), and for this reason the model is considered more useful for prediction, and the successive projection regression method (PPR) tends to work with explanatory variables when they are proportional, that is Which can be combined in one linear structure, and that the choice of the linear structure is equivalent to choosing the projection of one dimension of (X)[15].

6.1 The Basic PPR Algorithm

Friedman and Stuetzle (1981b) proposed the following algorithm for the purpose of estimating the response function f(x) for the data $\{(x_{1,1},...,x_{1,M},y_1)\},\ldots,\{(x_{n,1},., x_{n,M},y_n)\}$ and as follows [20]:

1- Determine $r_i^{[0]} = y_i$

2- We calculate the coefficient of determination for each (j = 1, …, maximize).

$$R^2_{[j]} = 1 - \frac{\sum_{i=1}^{n}\left(r_i^{[j-1]} - \hat{\vartheta}_j(\hat{\beta}_{[j]}^T x_i)\right)^2}{\sum_{i=1}^{n}\left(r_i^{[j-1]}\right)^2} \qquad (14)$$

And that j: represents the number of projections.

By making a change to the parameters ( varying over the parameters )

$\hat{\beta}_{[j]} \in \Re^p\left(\left\|\beta_{[j]}\right\| = 1\right)$ and the univariate regression function ( $\hat{\vartheta}_{[j]}$ ), the initial parameters are calculated by an iterative method (Back fitting).

3- Make ( $r_i^{[j]} = r_i^{[j-1]} - \hat{\vartheta}_{[j]}(\hat{\beta}_{[j]}^T x_i)$ ) and then we continue to repeat the second step until the value of ( $R^2_{[j]}$ ) is reduced and that the small value of ( $R^2_{[j]}$ ) means that ( $\hat{\vartheta}_{[j]}(\hat{\beta}_{[j]}^T x_i)$ ) is almost the zero function, and thus we cannot get another useful direction, and this algorithm leads to estimating the response function by:

$$\hat{f}_M(x) = \sum_{j=1}^{M} \hat{\vartheta}_{[j]}(\hat{\beta}_{[j]}^T x) \qquad (15)$$

**6.1 TLBO Algorithms for Method (PPR)**

1- Make (M = 1) and make the remainder ($r_i = y_i$).

2- Finding the initial ( $\beta_j$) values, which are obtained by one of the numerical methods (TLBO) [22].

3- Make ( $\beta$) and predict the independent variable (X) in one dimension as ($Z_i = \beta' X_i$).

4- One of the smoothing methods ($\vartheta_j(Zi)$) is introduced for the current residuals (univariate nonparametric regression for the ($Z_i$) residuals).

5- Estimate the values of ($\beta_j$) using one of the estimation methods, let it be (ols).

6- Find the vector ($\beta_M$) that maximizes ($I_{(B)}$) and matches ($\vartheta_{\beta_M}(\beta_M' X)$).

7- Make (M=M+1) for the remainder, as in equation

$$r_{i+1} = r_i - \vartheta_\beta(\beta' X_i) \qquad (16)$$

DOI: https://doi.org/10.54216/IJNS.200204
Received: May 17, 2022 Accepted: January 18, 2023

then go back to step (3) and stop at the lowest error specified previously.

**6.3 PSO Algorithms for Method (PPR)**

1- Make (M = 1) and make the remainder ($r_i = y_i$).

2- Finding the initial ( $\beta_j$) values, which are obtained by one of the numerical methods (PSO) [1].

3- Make ( $\beta$) and predict the independent variable (X) in one dimension as ($Z_i = \beta' X_i$).

4- One of the smoothing methods ($\vartheta_j(Zi)$) is introduced for the current residuals (univariate nonparametric regression for the ($Z_i$) residuals).

5- Estimate the values of ($\beta_j$) using one of the estimation methods, let it be (ols).

6- Find the vector ($\beta_M$) that maximizes ($I_{(B)}$) and matches ($\vartheta_{\beta_M}(\beta_M' X)$).

7- Make (M=M+1) for the remainder, as in equation (16) then go back to step (3) and stop at the lowest error specified previously.

**7. comparison criteria**

There are several criteria that measure the amount of quality in estimating nonparametric regression functions, and the criteria used are[12]:

6.1 Root Of Mean Squared Error(RMSE )

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad (17)$$

**8. Simulation Concept**

Simulation can be defined as imitating the behavior of a real system by building a simple model that represents this system[11], and then testing it performance of that system[14]. To approximate the concept, we can say that simulation is a process of imitating a real phenomenon through the existence of a set of mathematical equations. In theory[12,16], any phenomenon can be shortened in the form or converted into mathematical data or equations that can be simulated[5]. Simulation is also expressed as a mathematical method that works to find similar data[14]. For real data in order to solve the problems that fall within the framework of the research and that arise when generating different sizes with different parameters in multiple models of data[22], as a sample of the theoretical community is designed to represent the phenomenon instead of the real community[1].

**8.1 Generating Exploring Variables**

A number of simulation experiments were conducted, as three numbers of explanatory variables were generated (3, 5, 7) in order to fit the reality of the problem under study, and 6 different sizes of samples were used, which are (n = 50, 100, 200, 400, 800, 1000) with (18) experiments were carried out with a frequency of (1000) for each experiment, where the numbers of explanatory variables were distributed among these experiments, and that the generation of data was a natural generation with an arithmetic mean and variance, as shown in Table No. (1):

Table 1: The order of generating the initial variables according to variables with sizes of the samples

| explanatory variables | sizes of samples | | | | | |
|---|---|---|---|---|---|---|
| 3 | 50 | 100 | 200 | 400 | 800 | 1000 |
| 5 | 50 | 100 | 200 | 400 | 800 | 1000 |
| 7 | 50 | 100 | 200 | 400 | 800 | 1000 |

DOI: https://doi.org/10.54216/IJNS.200204

Received: May 17, 2022 Accepted: January 18, 2023

And that the data simulation was by generating variables and generating an initial correlation matrix that is the cause of the problem of multicollinearity in the data, through this matrix the amount of correlation between the variables is controlled and represents the data that suffers from the problem of multicollinearity and thus suffers from the problem of dimensionality and the simulation steps are summarized As follows :

First: the initial generation of data:

The explanatory variables were initially generated naturally with an average (m = 0) and a variance (sigma = 0.5).

$$X \sim N(0 , 0.5) \quad (18)$$

Second: the generation of multiple explanatory variables:

The multiple variables were generated with the arithmetic mean vector (M) and the variance matrix (SIGMA) based on the initial generation of the variables in (first), as the average of each variable is calculated, which will represent the arithmetic mean vector (M), and the covariance matrix is calculated for the initial generation, which will represent a matrix (SIGMA).

$$X \sim \mathrm{MN}(\underline{\mathrm{M}}, \mathrm{SIGMA}) \quad (19)$$

**8.2 Simulation Algorithm**

The following algorithm shows the mechanism of generating explanatory variables that suffer from the problem of collinearity:

1. The initial generation of variables by entering the number of required observations and the number of explanatory variables as in Paragraph (8.1) first.

2. Calculation of the arithmetic mean vector and covariance matrix from step (1).

3. Multivariate generation using the information of Step No. (2).

4. Generate the correlation matrix.

5. Computing the final multivariate by multiplying the correlation matrix resulting from step number (4) by the variables matrix in step number (3).

**8.3 simulation results**

Comparison of the best results for the basic and proposed methods, and methods for bypassing the dimensionality at correlation $\rho$ (0.3, 0.5, 0.8) and for observation sizes (50, 100, 200, 400, 800, 1000) with model

$$y = \frac{x_1}{0.5+(x_2+1.5)^2} + (1 + x_2)^2 + \sigma * \delta i \quad (20)$$

where $\delta i \sim N(0, \sigma^2)$ and $\sigma$ is standard deviation.

Table 2: Shows the results of a comparison between the methods for 3 variables

| method | | RMSE | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | n | 50 | 100 | 200 | 400 | 800 | 1000 |
| 1 | SIR | 0.3 | 0.6335 | 0.6064 | 0.5925 | 0.5711 | 0.5758 | 0.5622 |
| | | 0.5 | 0.6645 | 0.5973 | 0.5915 | 0.5739 | 0.5538 | 0.5623 |
| | | 0.8 | 0.6620 | 0.6017 | 0.5896 | 0.5814 | 0.5729 | 0.5538 |

DOI: https://doi.org/10.54216/IJNS.200204
Received: May 17, 2022 Accepted: January 18, 2023

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2 | Modified SIR | 0.3 | 0.1524 | 0.3195 | 0.3388 | 0.3498 | 0.3445 | 0.3497 |
| | | 0.5 | 0.2301 | 0.3215 | 0.3433 | 0.3442 | 0.3439 | 0.3595 |
| | | 0.8 | 0.2363 | 0.3329 | 0.3451 | 0.3471 | 0.3571 | 0.3520 |
| 3 | TLPO | 0.2 | 1.1341 | 1.8590 | 1.6073 | 1.4492 | 1.4342 | 1.3088 |
| | | 0.5 | 1.0407 | 1.8727 | 1.6205 | 1.5302 | 1.4234 | 1.3714 |
| | | 0.9 | 1.1405 | 1.9547 | 1.6531 | 1.4609 | 1.4050 | 1.3508 |
| 4 | PSO | 0.3 | 0.6737 | 0.6830 | 0.6915 | 0.7141 | 0.4801 | 0.4828 |
| | | 0.5 | 0.6595 | 0.6782 | 0.7032 | 0.7161 | 0.7183 | 0.7216 |
| | | 0.8 | 0.6596 | 0.6781 | 0.7035 | 0.6881 | 0.7055 | 0.7178 |

The results of the comparison of the methods in Table No. (2) at correlation ($\rho$ = 0.3) indicate that the method (Modified SIR) has the lowest (RMSE) at sizes (50, 100, 200, 400, 800 and 1000) observations. The results of comparing the methods in Table No. () at correlation ($\rho$ = 0.5) indicate that the method (Modified SIR) has the lowest (RMSE) at sizes (50, 100, 200, 400, 800 and 1000) observations. The results of the comparison of methods in Table No. (34) at correlation ($\rho$ = 0.8) indicate that the method (Modified SIR) has the lowest (RMSE) at sizes (50, 100, 200, 400, 800 and 1000) observations.

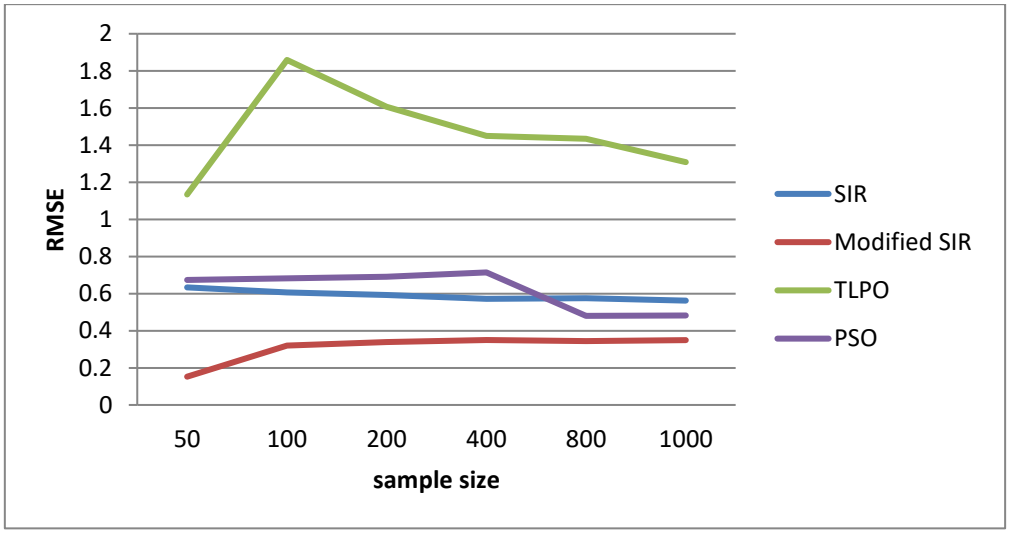


Figure 1 : RMSE for 3 variables with P=0.3

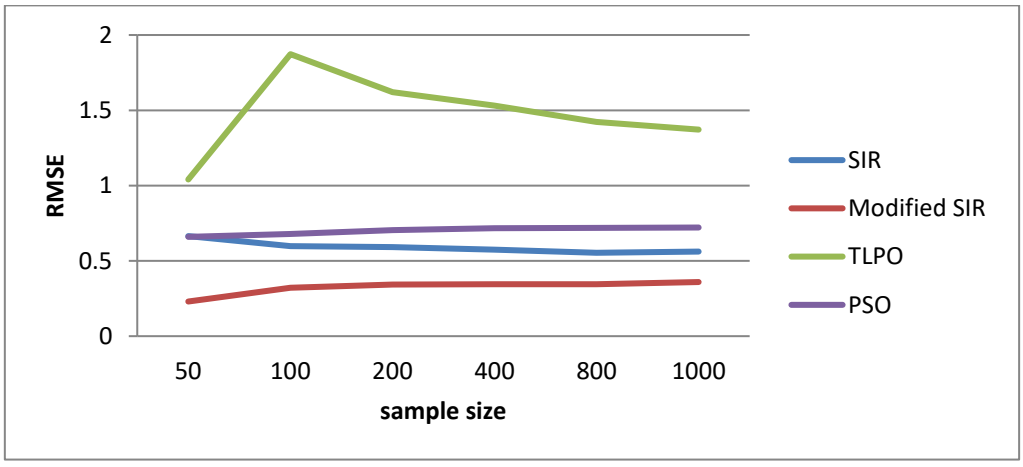


Figure 2 : RMSE for 3 variables with P=0.5

DOI: https://doi.org/10.54216/IJNS.200204
Received: May 17, 2022 Accepted: January 18, 2023

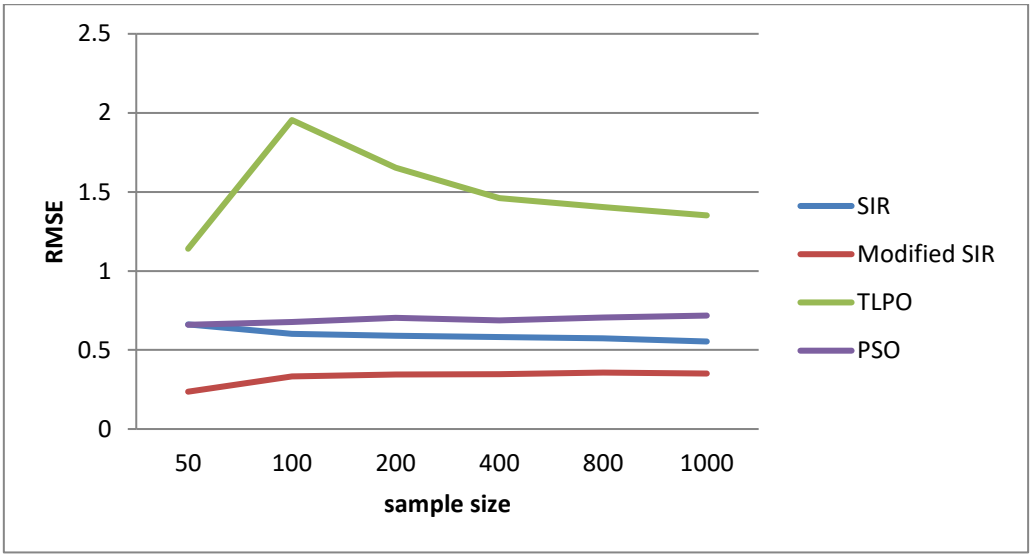


Figure 3 : RMSE for 3 variables with P=0.8

Table 3: Shows the results of a comparison between the methods for 5 variables

| method | | RMSE | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | n | 50 | 100 | 200 | 400 | 800 | 1000 |
| 1 | SIR | 0.3 | 0.6153 | 0.5723 | 0.5637 | 0.5392 | 0.5521 | 0.5495 |
| | | 0.5 | 0.5969 | 0.5758 | 0.5486 | 0.5405 | 0.5495 | 0.5287 |
| | | 0.8 | 0.6024 | 0.5588 | 0.5530 | 0.5484 | 0.5498 | 0.5370 |
| 2 | Modified SIR | 0.3 | 0.1640 | 0.2261 | 0.2405 | 0.2497 | 0.2536 | 0.2697 |
| | | 0.5 | 0.1749 | 0.2208 | 0.2352 | 0.2511 | 0.2530 | 0.2576 |
| | | 0.8 | 0.1867 | 0.2206 | 0.2409 | 0.2517 | 0.2536 | 0.2660 |
| 3 | TLPO | 0.2 | 3.0444 | 1.8584 | 1.5614 | 1.4759 | 1.3815 | 1.3616 |
| | | 0.5 | 3.0522 | 1.8816 | 1.5802 | 1.4434 | 1.4149 | 1.3937 |
| | | 0.9 | 3.0491 | 1.9043 | 1.5714 | 1.4675 | 1.3815 | 1.3619 |
| 4 | PSO | 0.3 | 0.8562 | 0.8539 | 0.8674 | 0.8665 | 0.8643 | 0.8667 |
| | | 0.5 | 0.8558 | 0.8532 | 0.8666 | 0.8662 | 0.8638 | 0.8656 |
| | | 0.8 | 0.8556 | 0.8532 | 0.8668 | 0.8706 | 0.8654 | 0.8652 |

The results of the comparison of the methods in Table No. (3) at correlation ($\rho$ = 0.3) indicate that the method (Modified SIR) has the lowest (RMSE) at sizes (50, 100, 200, 400, 800 and 1000) observations. The results of comparing the methods in Table No. () at correlation ($\rho$ = 0.5) indicate that the method (Modified SIR) has the lowest (RMSE) at sizes (50, 100, 200, 400, 800 and 1000) observations. The results of the comparison of methods in Table No. (34) at correlation ($\rho$ = 0.8) indicate that the method (Modified SIR) has the lowest (RMSE) at sizes (50, 100, 200, 400, 800 and 1000) observations.

DOI: https://doi.org/10.54216/IJNS.200204
Received: May 17, 2022 Accepted: January 18, 2023

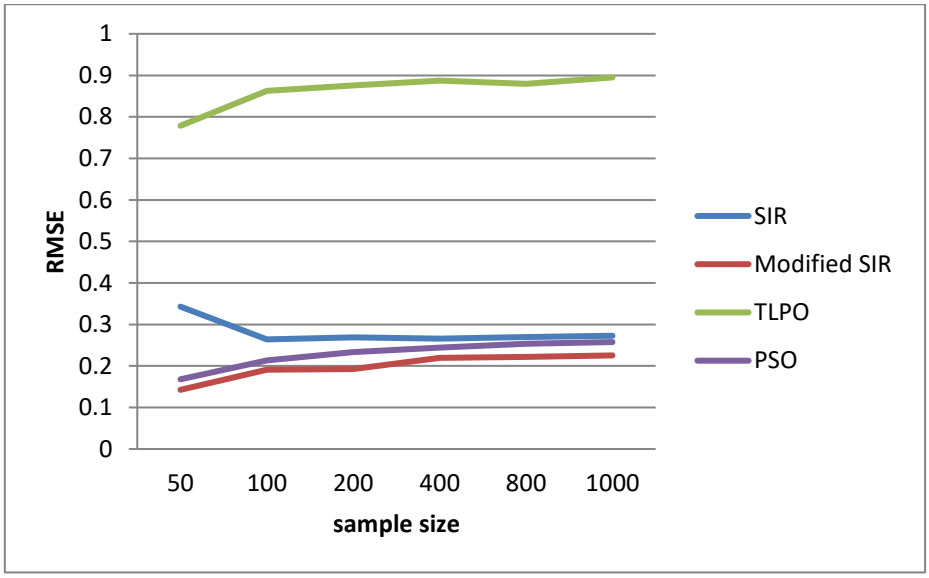


Figure 4 : RMSE for 5 variables with P=0.3

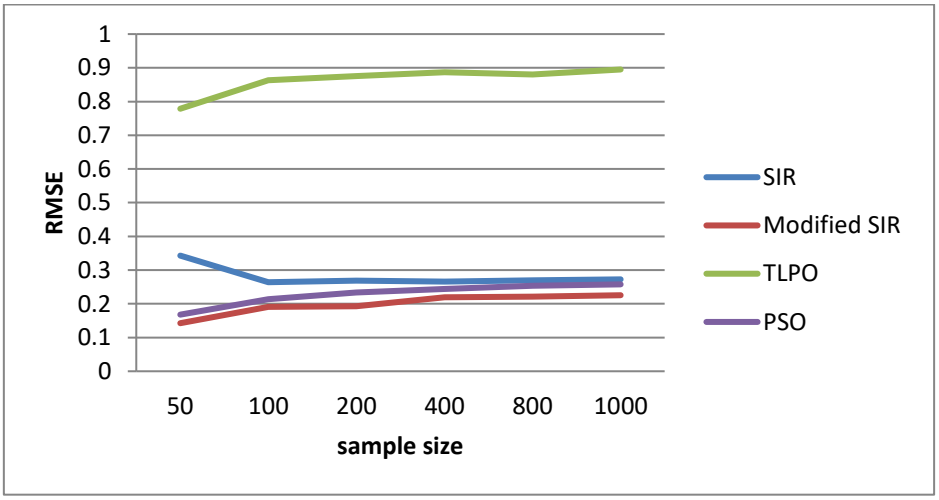


Figure 5 : RMSE for 5 variables with P=0.5

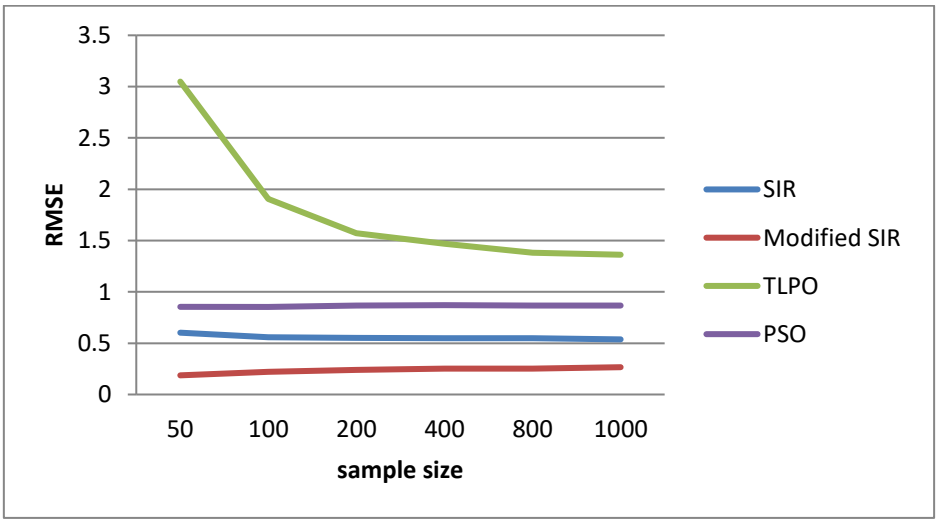


Figure 6 : RMSE for 5 variables with P=0.8

DOI: https://doi.org/10.54216/IJNS.200204
Received: May 17, 2022 Accepted: January 18, 2023

Table 4: Shows the results of a comparison between the methods for 7 variables

| method | | RMSE | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | n | 50 | 100 | 200 | 400 | 800 | 1000 |
| 1 | SIR | 0.3 | 0.3423 | 0.2619 | 0.2672 | 0.2643 | 0.2685 | 0.2715 |
| | | 0.5 | 0.3430 | 0.2624 | 0.2682 | 0.2657 | 0.2696 | 0.2725 |
| | | 0.8 | 0.3430 | 0.2641 | 0.2686 | 0.2661 | 0.2698 | 0.2726 |
| 2 | Modified SIR | 0.3 | 0.1337 | 0.1833 | 0.1956 | 0.2122 | 0.2160 | 0.2239 |
| | | 0.5 | 0.1355 | 0.1904 | 0.1989 | 0.2190 | 0.2214 | 0.2254 |
| | | 0.8 | 0.1426 | 0.1914 | 0.1929 | 0.2195 | 0.2217 | 0.2255 |
| 3 | TLPO | 0.2 | 0.7784 | 0.8624 | 0.8746 | 0.8878 | 0.8790 | 0.8941 |
| | | 0.5 | 0.7754 | 0.8620 | 0.8754 | 0.8871 | 0.8795 | 0.8947 |
| | | 0.9 | 0.7786 | 0.8627 | 0.8754 | 0.8872 | 0.8799 | 0.8950 |
| 4 | PSO | 0.3 | 0.7562 | 0.8118 | 0.8482 | 0.8632 | 0.8894 | 0.8790 |
| | | 0.5 | 0.7454 | 0.8128 | 0.8598 | 0.8574 | 0.8662 | 0.8631 |
| | | 0.8 | 0.1680 | 0.2135 | 0.2338 | 0.2445 | 0.2535 | 0.2577 |

The results of the comparison of the methods in Table No. (3) at correlation ($\rho$ = 0.3) indicate that the method (Modified SIR) has the lowest (RMSE) at sizes (50, 100, 200, 400, 800 and 1000) observations. The results of comparing the methods in Table No. () at correlation ($\rho$ = 0.5) indicate that the method (Modified SIR) has the lowest (RMSE) at sizes (50, 100, 200, 400, 800 and 1000) observations. The results of the comparison of methods in Table No. (34) at correlation ($\rho$ = 0.8) indicate that the method (Modified SIR) has the lowest (RMSE) at sizes (50, 100, 200, 400, 800 and 1000) observations.

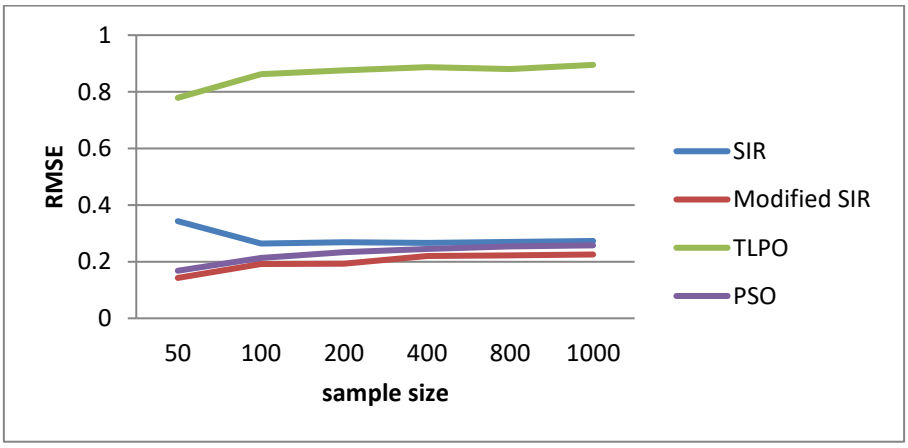


Figure 7 : RMSE for 7 variables with P=0.3

DOI: https://doi.org/10.54216/IJNS.200204
Received: May 17, 2022 Accepted: January 18, 2023

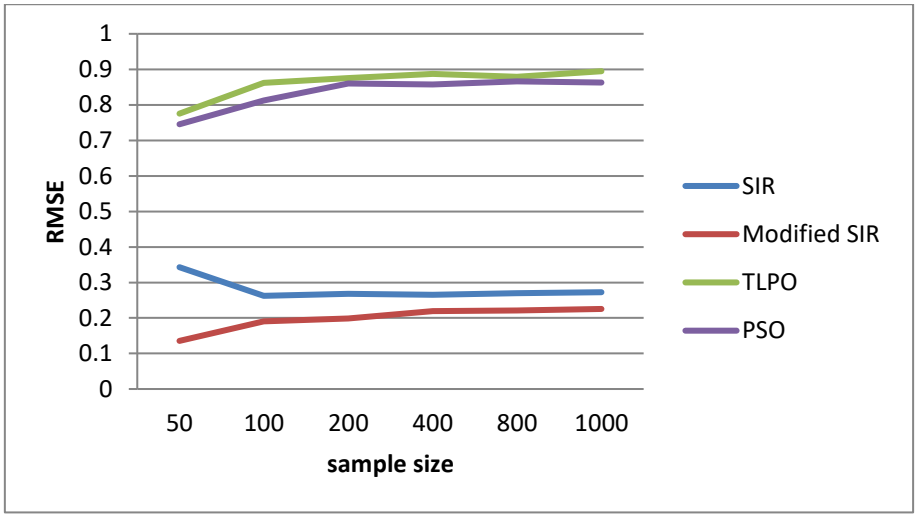


Figure 8 : RMSE for 7 variables with P=0.5

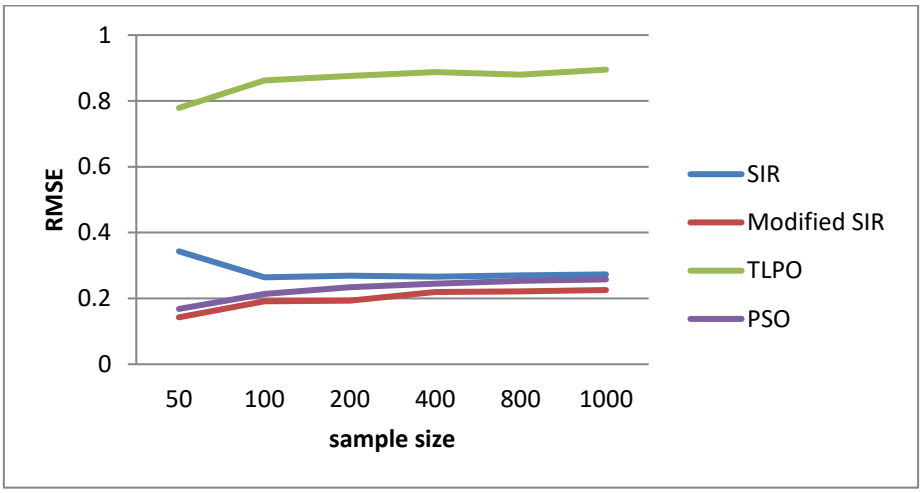


Figure 9 : RMSE for 7 variables with P=0.8

**9. Conclusion**

In the light of the results of the experimental and applied side, a number of conclusions were reached, summarized as follows:

1- When generating data with an average (0) and a standard deviation of (0.5) for three different correlation values, the problem of linear multiplicity is achieved, which leads to the occurrence of the problem of dimensionality and for a group of small, medium and large observation sizes, and for the numbers of explanatory variables are (3, 5, 7) It was noted that the (Modified SIR) method achieved accurate results, if it gave less (RMSE), followed by the (SIR) method, and it gave results close to the (Modified SIR) method, because the two methods depend on finding the best linear components of the slices and thus give accurate results .

2- We notice that when changing the value of tuning parameter (H), which represents the number of slices, it has a significant impact in making (RMSE) as low as possible, as we notice the variation in the results when changing the value of (H), and it is also noted that the value of the lower (RMSE) was It is often at H = 5) and (H = 10), that is, when the value of (H) is equal to the number of explanatory variables or double their number, and the importance of (H) has been indicated.

3- We note that the (SIR) method obtained good results and was solved after the proposed methods, because it relied on an unconventional method in finding the shorthand parameters.

DOI: https://doi.org/10.54216/IJNS.200204
Received: May 17, 2022 Accepted: January 18, 2023

4- We note that the methods of overcoming the problem of remoteness have fallen behind the methods of dealing with the problem of remoteness on the experimental side, and this difference has clearly appeared on the applied side.

5- We note from the above results that all slat regression methods can be used as detection methods for data that contain problems, and whenever they give unsatisfactory results, this indicates that the data is free from problems.

**Funding:** "This research received no external funding"

**Conflicts of Interest:** "The authors declare no conflict of interest."

**References**

[1] Ai-Qin Mu , De-Xin Cao , Xiao-Hug Wang ; (2009)" Amodified particle swarm optimization algorithm ", WCICA, Vol. (1) , No. (2) , PP. (151-155).
[2] Alexander Mitev , Michael Marefat , Dongsheng Ma and Janet Wang ; (2008) " Parameter reduction for variability analysis by Sliced Inverse Regression (SIR) Method " , IET, Vol.(2) , Issue.1 , PP.(16-22).
[3] Ali Akenani and Tahir R.Dikheel ; (2016) " Sparse Sliced Inverse Quantile Regression " ,Mathematics and Statistics ,PP. (192-200).
[4] Ali Gannoun , Christiane Guinot and Jerome Saracco ; (2004) " Reference Curve estimation Via alternating Sliced Inverse Regression " , Environmetrics , Vol.(15) , PP.(81-99).
[5] Anil Kumar Ghosh and Smarajit Bose ; (2004) " Backfitting Neural Network " , Computational Statistic , No.(19) , PP.(193-210).
[6] Arunasakthi .K. and Kamatchipriya.L ; (2014) " A Review on Linear and Non-Linear Dimensionality Reduction Techniques " Machine Learning and Applications, Vol.(1) , No.(1) , PP.(65-76).
[7] Charles B.Roosen and Trevor J.Hastie ; (1993) " Logistic Response Projection Pursuit " ,AT&T Bell Laboratories , PP.(1-25).
[8] Christopher J.C. Burges ;(2009) " Dimension Reduction : A Guided Tour " , Foundations and Trends , Vol.(2) ,No.(4) , PP.(275-365).
[9] Cook .R.D and Weisberg .S. ; (1991) " Sliced Inverse Regression for Dimension Redaction , comment " ,JASA , PP.(328-332).
[10] D. Napoleon and S. Pavalakodi ; (2011) " A New method for Dimensionality Reduction using K-means Clustering algorithm for High Dimensional Data set " , Computer Applications , Vol.(13) , No.(7) , PP.(41-46).
[11] Dan Xu and Hualing Deng ; (2012) " Application of Projection Pursuit Regression in the Forecast of the Percentage of Damaged seeds of Leguminivora Glycinivorella " , Advances in System Science and Applications ,Vol.(12) , No.(1) .
[12] El-Dereny , M. Rashwan , N. ; (2011) " Solving Multicollinearity Problem Using Ridge Regression Models ",Int. J. Contemp. Math. Sciences , Vol. (12) , PP.(585-600).
[13] Francis Bach ; (2017) " Breaking the Curse of Dimensionality with convex Neural Networks " , JMLR , Vol.(18) ,Issue.(1) ,PP.(1-53).
[14] G.E.Hinton and R.R Salakhutdinor ; (2006) " Reducing the Dimensionality of data with Neural Network" , Science , Vol.(313) ,Issue.(5786) ,PP.(504-507).
[15] Guo Qianjian and Yang Jianguo ; (2011) " Application of Projection Pursuit Regression to thermal error Modeling of CNC Machine Tool " , AMT , Nol.(55), Issue.(5) , PP,(623-629).
[16] Guy Martial Nkiet ; (2008) " Consistent Estimation of the Dimensionality in Sliced Inverse Regression" , Annals of the Institute of Statistical Mathematics , Vol.(60) ,Issue.( 2 ) , PP.( 257–271).
[17] Hamparsum Bozdogan ; (2004) "Statistical Data Mining and Knowledge Discovery " , CRC ,Ch.13 .
[18] Han .Ming Wu ; (2008) " Kernel Sliced Inverse Regression with Applications to Classification ' , Computational and Graphical Statistics , Vol.(17) , Issue.(3) , No.(3) , PP.(590-610).
[19] Hardle .W ;(1990) " Applied Nonparametric Regression " , Cambridge University Press. http://econpapers.repec.org/scripts/redir..pdf.
[20] Jerome H. Friedman , Werne Stuetzie ; (1981) " Projection Pursuit Regression ", JASA ,Vol.(76) , No.(376) , PP. (817-823) .
[21] Jerome H.Friedman and John W.Tukey ; (1974) " A Projection Pursuit Algorithm for Exploratory data analysis " , IEEE , Vol.(23) ,No.(90)

DOI: https://doi.org/10.54216/IJNS.200204
Received: May 17, 2022 Accepted: January 18, 2023

[22] R.V.Rao , V.J.Savsaini , D.P.Vakharia ; (2011) "Teaching Learning Based Optimization : Anovel method for constrained mechanical design ,PP. (303-315), No.(43).

DOI: https://doi.org/10.54216/IJNS.200204
Received: May 17, 2022 Accepted: January 18, 2023