High dimensional data challenges in estimating multiple linear regression

Jassim N Hussain
Department of Statistics, College of Administration & Economics, University of Kerbala, Iraq.
Email: jasim.nasir@uokerbala.edu.iq

Abstract. Nowadays, High dimensional data are quickly increasing in many areas because of the development of new technology which helping to collect data with a large number of variables in order to better understanding for a given phenomenon of interest. Multiple Linear Regression is a famous technique used to investigate the relationship between one dependent variable and one or more of independent variables and analyzing the effects of them. Fitting this model requests assumptions, one of them is large sample size. High dimensional data does not satisfy this assumption because the sample size is small compared to the number of explanatory variables (k). Consequently, the results of traditional methods to estimate the model can be misleading. Regularization or shrinkage techniques (e.g., LASSO) have been proposed to estimate this model in this case. Nonparametric method was proposed to estimate this model. Average mean square error and root mean square error criteria are used to assess the performance of nonparametric; LASSO and OLS methods in the case of simulation study and analyzing the real dataset. The results of simulation study and the analysis of real data set show that nonparametric regression method is outperformance of LASSO and OLS methods to fit this model with high dimensional data.

Key Words: Average Mean square error; Data reduction techniques; Nonparametric regression, Regularization methods, Variable selection method.

1. Introduction
In the last decades, high dimensional data becoming increasingly common in different areas. The cases when the number of covariates is larger than the number of observations (p > n) is named high dimensional. Consequently, traditional methods of analyzing statistical models (e.g., regression models) produce poor results and are unreliable for inference because of overfitting in such cases. On the other hand, Multiple Linear Regression (MLR) is one of the important procedure for modeling and analyzing data in many disciplines. Numerous assumptions are requested in fitting this model, the important one is large size of sample. Accordingly, the method of estimation are differing depending on the validity of the assumptions in the data set. If the requested assumptions are hold, then the traditional fitting methods (e.g., ordinary least square (OLS) or maximum likelihood (ML)) are used to fit the MLR. Meanwhile, when some of these assumptions do not satisfied, specifically if the size of the sample is small and the number of the covariates (p) may equal or greater than the size of the sample (n). This case is named as high dimensional data, with this case the results of traditional method when fitting MLR model produce poor and unreliable results for inference because of overfitting property. Also, the multicollinearity problem tends to occur when n becomes small or p becomes large. Therefore, the estimators of β’s become unstable [1].

Consequently, when analyzing the MLR with high dimensional data several alternatives methods are used including the variable selection methods, data Reduction Techniques [2,3] and The shrinkage techniques These methods give biased estimators with a smaller variance than OLS estimators [4, 5].

Several studies to explore and to explain the effect of covariates on dependent variable and other topics in regression analysis are conducted. [6] investigated the estimation of MLR parameters models...
when the original assumptions of OLS estimation are weak. Also, they introduced some MLR models with outliers and get conclusions. [7] stated that the common question is how to relate the response variable (Y) and the explanatory variables (X_i) by employing the analysis of regression. [8] was developed statistical methods that are capable to detect outliers. He suggested a robust regression to analysis data with outliers. He also, defined performance of outliers in LR and compared some of robust methods using simulation studies. [9] used a genetic algorithm to determine a set of parameters that minimizes the prediction error for MLR model. [10] extended the econometric methods to GLMs to analysis the binary, count and duration response involved in social sciences and business. He found that these methods perform well in the applications for prediction and inference with high dimensional data. [11] used simulation studies to compared the performance of Lasso, Elastic net, Ridge Regression, and Bayesian models with high dimensional data in multivariate regression. They found that the Ridge Regression model was effective in estimating parameter accurately and control over the Type I error rate. [12] investigated the performance of regularization methods to analysis the high dimensional data with different sparse and non-sparse conditions. They studied the prediction, parameter estimation and variable selection properties.

Finally, [13] stated that there are two purposes of analyzing high-dimensional data are firstly to define the relationship between the covariates and response variable for scientific objectives. Secondly to develop effective techniques that can predict the future observations accurately.

The goal of this paper is to propose nonparametric method (kernel regression (KR)) to estimate the MLR model with high dimensional data as an alternative method and compare its performance with OLS and LASSO methods.

The rest of the paper consists of Section 2 devoted to specify the model and to describe two methods of estimation. Section 3 devoted to present the results of simulation study and analyzing real dataset. Section 4 devoted to discuss the results of the study. Finally some conclusion are placed in section 5.

2. Methodology
MLR model is defined as “a form of predictive modeling technique which investigates the relationship between a dependent (response) variable (Y) and independent variables (covariates) (X_i). This technique is used for forecasting, time series modeling and finding the fundamental effect relationship between the variables” [14]. Three fundamental components determined MLR model; the first component is the number of explanatory variables, the second component is the type of response variable and the third component is the shape of regression line. Meanwhile, the main goal from any data analysis is to get the correct estimation from raw data. Consequently, the important question is “if there is a statistical relationship between a response variable (Y) and explanatory variables (X_i)”.

Therefore, answering this question means that MLR modeling is conducted in order to present this relationship. MLR is also used to make inferences about the response variable (Y) based on values of a set of explanatory variables; (X_i) [7].

Here, we use the MLR model where the response variable (Y) is a linear function of p explanatory variables (X_j) and a random error (U), this model with an intercept. MLR model of ith row of these variables is given by the following equation [15] [16]:

\[ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + u_i \]

(1)

\[ Y = X\beta + U \]

(2)

where \( Y \) is a vector of \((n \times 1)\) observed response values, \( X \) is the matrix of the explanatory variables with \((n \times p)\) rank, \( \beta \) is the vector of unknown parameters with \((p \times 1)\), and \( U \) is the vector of random error terms with \((n \times 1)\) rank as follow:

\[ Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{nk} & \cdots & x_{nk} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \]
The goal of regression analysis is to estimate the unknown parameters $\beta$’s and make inference about the future values of response variable.

If the assumptions of MLR model are holding the ordinary least squares (OLS) method is used to estimate the regression parameters ($\beta$) using the formula [16]:

$$\hat{\beta}_{ols} = (X'X)^{-1}X'Y$$  \hspace{1cm} (3)

Consequently, the estimated MLR model is:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}, \hspace{0.5cm} i = 1, 2, \ldots, n; \hspace{0.5cm} j = 0, 1, 2, \ldots, k$$  \hspace{1cm} (4)

The residuals calculated from $\varepsilon$ sample size $n$ may be defined the OLS criterion from Eq. (4) as follow:

$$MSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$  \hspace{1cm} (5)

where $(y_i; \hat{y}_i)$ are the observed and the estimated response variable for subject $i$ respectively, as in Eq. (4). The minimization is performed with respect to the $(p + 1)$ parameters $[\beta_0; \beta_1; \ldots; \beta_p]$ with a constrain that $n > (p + 1)$ this means the sample size is larger than the number of parameters to be estimated from that sample [4].

There are other important assumptions to obtain a valid estimation of MLR model as in Eq. (4). These assumptions are the error term follows a normal distribution and the error process $u_i$ independent of all explanatory variables where [14]

$$u_i \sim N(0, \sigma^2), \hspace{0.5cm} \text{and} \hspace{0.5cm} \mathbb{E}(u_i | X_i) = 0; \hspace{0.5cm} \text{Cov}(u_i, u_j) = \sigma^2 I_n.$$  \hspace{1cm} (5)

The other important assumption is $n > p$ which is our concern here because the size of the sample must be large to offer enough power for the test. Meanwhile, if the number of observations $n$ gets closer to or less than $p$, the number of covariates then there is more variability in the OLS fit [17]. Also, the matrix ($X'X$) is not invertible in this case therefore there is no unique solution for OLS regression [18]. Consequently, there may be some irrelevant variables included in the MLR model.

Hence, in the case of high dimensional data, when the sample size ($n$) becomes closer to or less than the number of explanatory variables ($p$). One of four groups of estimation methods can be used based on the validity of the assumptions of the MLR as mentioned above. The important question is which estimation method is appropriate to estimate the unknown parameters ($\beta$’s) of this model in this case? Consequently, the goal of this paper is to compare the performances of two estimation methods OLS and LASSO with KR method as a proposed method to estimate the parameters of MLR model in this case. The following subsections are consisting of a brief concepts for these methods.

2.1 Regularization method (LASSO)

When the sample size ($n$) in the model may be less than or equal to the number of explanatory variables ($p$), Least Absolute Shrinkage and Selection Operator (LASSO) is the appropriate method for estimating the parameters of MLR models [19]. Also, there are two important considerations when fitting high-dimensional data: model sparsity and prediction ability. Therefore, LASSO is defined as “a regularization methods for simultaneous estimation and variable selection” [20]. LASSO deals with many predictors may be $p \geq n$, and with an ill-conditioned model matrix $X$ (i.e. $X'X$ is not invertible or near singular). However, LASSO is differing from other methods, because it is making the interpretation of the statistical model more plausible by sets many coefficient estimates exactly to zero. Regularization techniques work by introducing a penalty to the OLS estimator as in Eq. (5). Then, LASSO method of estimation is introduced by [21] and is defined as follows:

$$L(\beta, \lambda) = ||Y - X\beta||^2 + \lambda||\beta||_1$$

$$= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{k} |\hat{\beta}_j|$$

$$\hat{\beta}_{l_{asso}} = \arg\min_{\beta} \{||Y - X\beta||^2 + \lambda||\beta||_1\}$$  \hspace{1cm} (7)
The notations in Eq. (7) are defined as in Eq. 1 where the estimated OLS model of MLR contains the coefficients of interest [20]. Eq. (7) can be solved by using quadratic programming techniques such as a coordinate gradient descent algorithm [22], where it becomes is:

\[ \hat{\beta}_{\text{Lasso}} = \arg\min_{\beta} \{ (y_i - \hat{y}_i)^2 \} \]  

Subject to \( \sum_{j=1}^{k} |\hat{\beta}_j| \leq s \)

the parameter \( \lambda \geq 0 \) represents the degree of the coefficients of the model which have small weighted or removed from the model. Therefore, larger \( \lambda \) values means greater shrinkage, and LASSO is converted to the OLS estimator when \( \lambda = 0 \). The optimal value of \( \lambda \) is estimated by using Jackknife cross-validation, which is described in [21]. LASSO method gives optimal values of the \( \hat{\beta}_j \) depending on the importance of the variable; where the greatest importance explanatory variables receive higher values, and the smallest importance are allocated coefficients at or near 0 [11]. [19] stated that LASSO method has good experimental and theoretical characteristics for estimation and variable selection. In spite of the LASSO has shown success in many situations, it has some limitations [21] [23]. Therefore we propose KR method to overcome these limitations.

### 2.2 Kernel Regression (KR) Model

Kernel regression (KR) model has been considered as one of the efficient nonparametric regression models, it is proposed by Nadaraya and Watson at 1964. KR method is particularly powerful in high-dimensional and nonlinear settings [24]. In practice, KR method regresses the dependent variables \( Y \) onto the similarity of independent variables measured through the kernel function. KR model is weighted average estimators that use kernel functions as weights. Suppose we have the sample of observations \((y_i, x_i)\) \( i = 1, \ldots, n \), then, the Kernel estimation of \( f(Y) \) based on the sample is [25]:

\[ f_n(Y) = \frac{1}{n} \sum_{i=1}^{n} K(Y - y_i|x_i) \]  

where \( K(Y - y|x) \) is one of Kernel function given \( x \) observation. Two choices must be made (the kernel function \( K(.) \) and the smoothing parameter \( h \)) when working with a kernel estimator. On the other hand, given a choice of kernel \( K(\cdot) \), and a smoothing parameter \( h \), Kernel regression model for one dimension is [26]:

\[ \hat{y}(y|x) = \sum_{i=1}^{n} w(x, x_i) y_i \]  

where \( w(x, x_i) \) is the weight of Kernel. Because these weights are smoothly varying with \( x \), the kernel regression estimator \( \hat{y}(y|x) \) itself is also smoothly varying with \( x \). The selection of \( K(.) \) can be easily changed to support of the density to be estimated. Also, the selection of appropriate smoothing parameter \( h \) is very important, because of the effect of the \( h \) on the shape of the corresponding estimator. When \( h \) value is small, we will obtain an undersmoothed estimator, with high variability. Whereas, if the value of \( h \) is large, the resulting estimator will be oversmoothed. In practice, we tend to select \( h \) by one of two methods; guessing it or using cross-validation method.

The multiple nonparametric regression (MNR) model is written as [27]:

\[ y_i = f(x_i^T) + \varepsilon_i \]  

where the function \( f \) is left unspecified. Moreover, the object of nonparametric regression is to estimate the regression function \( f(\cdot) \) directly, rather than to estimate parameters. It is difficult to fit the MNR model when there are many predictors, several models have been developed one of them is the additive nonparametric regression (ANR) model as:

\[ y_i = \alpha + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_p(x_{ip}) + \varepsilon_i \]
where \( f_j (\cdot) \) is representing the partial-regression functions which are assumed smooth, and to be estimated from the data. These partial-regression functions \( f_j (\cdot) \) are fitted using one of the simple-regression smoother, such as local polynomial regression.

Therefore, the first step to fit this model is to define a multiple neighborhood around the original point \((x_j^2, x_{01}, x_{02}, \ldots, x_{0p})\) and the second step is to calculate the weight of these points by using the scaled Euclidean distances as a default in LOESS function, where,
\[
D(x_i, x_0) = \left( \sum_{i=1}^{p} (x_{ij} - z_{ij})^2 \right),
\]
where \( z_{ij} = (x_{ij} - \bar{x}_j)/S_j \), and \( \bar{x}_j \) and \( S_j \) are the mean and the standard deviation of the \( jth \) predictor. Then, the scaled distances is:
\[
w_i = W \left( \frac{D(x_i, x_0)}{b} \right)
\]
where \( W (\cdot) \) is an appropriate weight function, such as the tricube and Epanechnikov and others. The third step is constructing a weighted polynomial regression of \( y \) on the \( x's \); such as a local linear fit takes the following form:
\[
y_i = \alpha + f_1 (x_{i1} - x_{01}) + f_2 (x_{i2} - x_{02}) + \cdots + f_p (x_{ip} - x_{0p}) + e_i
\]
Then the estimated value at \( x_0 \) is then simply by \( \hat{y} = \alpha \). Performance of kernel is measured by many criteria among them MSE and RMSE as presented in the next subsection.

2.3 Assessing Criteria of the Estimation Methods.
There are some useful measures (criteria) to assess the adequacy of the model to the data when we have several alternative methods to estimate the coefficients of the model. The average of mean square error (AMSE)is the first criterion of goodness of fit which is used to select the most appropriate model [28]:
\[
AMSE_m = \frac{1}{m} \left( \sum_{i=1}^{n} (Y_i - \hat{y}_i)^2 \right), \quad m = 1, 2, \ldots, M
\]
Since we repeat the dataset for \( M \) times in the simulation study then we will use the average mean square error as:
\[
AMSE = 1 / M (\sum_{m=1}^{M} \text{AMSE}_m)
\]
where \( M \) is the number of replications. At most, the number of observations (\( n \)) is less than or equal the number of explanatory variables in the case of high dimensional data. Therefore, the value of AMSE is larger than its value in the other case of fitting MLR model, meanwhile, MSE for the KR model is [29]:
\[
MSE(x) = \text{var} \{ \hat{g}_n(x) \} + \text{bias}^2 \{ \hat{g}_n(x) \}, \quad x \in N
\]
We will use these criteria to assess the goodness of fit the estimation method as in the following section.

3. Applying Estimation Methods
The proposed method and other estimation methods are applied using simulation study and analyzing the real dataset to compare the performance of them. The results are as in the following subsections.

3.1 Simulation Study
The first variable generated is the dependent variable by using the following model:
\[
y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i
\]
The regression coefficients (\( \beta_j \)) are assumed the first five parameters of the first ten equal to one and the rest five equal to zero and the first five of the second ten parameters are equal to one and the rest five equal to zero. Three levels of \( \sigma^2 \) are used (1, 5, and 9). The first ten explanatory variables (\( x_j \)) are sampled independently from uniform distribution U(0,1) and the second ten explanatory variables from standard normal distribution N(0,1). The data sets are generated under the sample sizes (15, 20
and 25) and the experiment was repeated 1000 times. The program code for the data simulation was adapted in SAS and XLSTAT software. Consequently, the values of the Root average of mean square error (RAMSE) for different methods, different sample sizes and different values of variance of the error are as in Table (1):

Table (1): RAMSE for different methods of estimation, different sample sizes and different values of variance of error

| criteria | Method | n=15 | n=20 | n=30 |
|----------|--------|------|------|------|
|          | = 1    | = 5  | = 9  | = 1  | = 5  | = 9  | = 1  | = 5  | = 9  |
| RAMSE    | OLS    | ***  | ***  | ***  | ***  | ***  | ***  | 1.885| 2.038| 2.227|
|          | LASSO  | ***  | ***  | ***  | ***  | ***  | ***  | 1.240| 2.116| 2.281|
|          | KR     | 4.247| 5.321| 6.119| 5.903| 6.743| 7.033| 6.598| 6.769| 7.297|

(***: The models when \( n \leq p \) are not fitted because model is not full rank.)

Meanwhile, the results of AMSE for different methods, different sample sizes and different values of variance of error are as in Table (2).

Table (2): AMSE for different methods of estimation and different sample sizes and different values of variance of error

| criteria | Method | n=15 | n=20 | n=30 |
|----------|--------|------|------|------|
|          | = 1    | = 5  | = 9  | = 1  | = 5  | = 9  | = 1  | = 5  | = 9  |
| AMSE     | OLS    | ***  | ***  | ***  | ***  | ***  | ***  | 1.4126| 4.1538| 4.9602|
|          | LASSO  | ***  | ***  | ***  | ***  | ***  | ***  | 1.5363| 4.4778| 5.2005|
|          | KR     | 20.26| 28.316| 37.441| 34.841| 45.473| 49.461| 43.527| 45.82| 53.244|

(***: The models when \( n \leq p \) are not fitted because model is not full rank.)

3.2 Analyzing of Real Data

Almost newborns come with healthy weight. Most of them born after 37 or 40 weeks weigh between 2.5 Kg and 4.0 Kg. Newborns who are lighter or heavier than the average baby are usually fine. Therefore, a simple random samples with sizes (15, 20, 30) consists of all mothers who visit the primary health care center in Babylon province in year 2018, are drawn to compare the performance of the estimation methods presented above. The weight (Kg) of newborn children is a response variable \( Y \) which is considered as an indicator of healthy generation, while the risk factors are as following:

- mother age (year);
- Age at marriage (year);
- Educational attainment of mother;
- Educational attainment of Husband;
- Weight of mother (Kg), Contraception Using; Mother's smoking (yes; No);
- Age of Husband (year);
- Job of Husband (yes, No);
- Period of Marriage (year);
- Number of born children (#);
- Period of exercise per week (hour);
- Thyroid disease (yes, no);
- Mother's sleeping per day (hour);
- Taking medications (yes, No), Breastfeeding Duration (Month), Mother's job (yes, No);
- number of dead children (#).

The weight of newborns (Kg) as the dependent variable. When we apply OLS, LASSO and KR methods to fit the MLR model for this data set then the results of assessing criteria are as in Table (3):

Table (3) values of assessing criteria for different samples and different methods

| Method | criteria | n=15 | n=20 | n=30 |
|--------|----------|------|------|------|
|        | MSE      | RMSE | MSE  | RMSE | MSE  | RMSE |
| OLS    | ***      | ***  | ***  | ***  | 0.262| 0.512|
|        | LASSO    | ***  | ***  | ***  | 0.129| 0.3595|
|        | KR       | 4.554| 2.134| 0.67 | 1.252| 2.292|

(***: The models when \( n \leq p \) are not fitted because model is not full rank.)
4. Discussing the Results

computational and statistical challenges have been introduced by high dimensional data especially in estimating MLR models. High dimensionality brings noise accretion, false correlations and related heterogeneity. These problems of high dimensional data make traditional statistical methods invalid. The results in tables 1 and 2 show that the OLS regression and LASSO cannot fit the MLR model when the sample size \( n = 15 \) \( \leq 20 \) is smaller than or equal to the number of explanatory variables \( p = 20 \) (i.e. \( n \leq p \)) because the models are not full rank, OLS solutions for the parameters are not unique and some statistics will be misleading. Meanwhile the KR method gives an accepted criteria of RAMSE and AMSE in these cases. But when the sample size become \( n = 30 \) with the same number of explanatory variables the results show an improvement of the results of traditional method OLS to give a less values of the criteria. These results show that KR is outperformance of OLS and LASSO methods to fit the MLR with high dimensional data. Also, analyzing real dataset confirms that KR gives better results than other methods with high dimensional data when \( (n = 15 \) \( \leq 20 \) and the number of covariates equal to \( (p = 18) \). Finally all the results of simulation study and analyzing real data show that KR is powerful than other method in estimating MLR with high dimensional data.

5. Conclusions

The above results show that KR method is better than traditional method (OLS) and the Least LASSO to estimate the predicted values of response variable of MLR model with high dimensional data because it has smallest values of RAMSE and AMSE and gives a predicted values of response variable; mean while the other method cannot fit the MLR model when the sample size less than or equal to the number of explanatory variables. Also, the results from analyzing the real dataset are confirming the results of simulation study that KR is the preferred method than others. For the future works may be compare the performance of KR method with the Principal Component and Ridge regression methods to estimate the coefficients of MLR with high dimensional data. Also, may be compare the performance of KR method with other nonparametric methods (e.g. spline method).

References

[1] Fujiwara M, Minamidani T, Nagai I and Wakaki, H 2012 Principal Components Regression by using Generalized Principal Components Analysis, retrieved on 10-6-2019 from: www.math.sci.hiroshima-u.ac.jp/stat/TR/TR12/TR12-09.pdf
[2] Hastie T, Tibshirani R and Friedman, J 2009 The elements of statistical learning: Data mining, inference, and prediction. Springer-Verlag, New York.
[3] Finch W H, Hernandez Finch M E and Moss L 2014 Dimension reduction regression techniques for high dimensional data, General Linear Model Journal 40(2) 1-15.
[4] Finch W H and Hernandez Finch M E 2017 Multivariate Regression with Small Samples: A Comparison of Estimation Methods. General Linear Model Journal 43(1) 16-30.
[5] Lee H, Park Y M and Lee S 2015 Principal Component Regression by Principal Component Selection. Communications for Statistical Applications and Methods 22(2) 173–180.
[6] Cankaya S, Kaylaap GT, Sangun L, Tahtali Y and Akar M 2006 A Comparative Study of Estimation Methods for Parameters in Multiple Linear Regression Model. J. Appl. Anim. Res 29 43-47.
[7] Alxepoulos E C 2010) Introduction to Multivariate Regression Analysis, Hippokratia Journal 14(1) 23-28.
[8] Alma Ö G 2011 Comparison of Robust Regression Methods in Linear Regression, Int. J. Contemp. Math. Sciences 6(9) 409 – 421
[9] Stoica F, and Boitor C G 2014 Using the Breeder GA to Optimize a Multiple Regression Analysis Model used in Prediction of the Mesiodistal Width of Unerupted Teeth, Int. J. Comput. Commun., 9(1) 62-70.
[10] Sapra S 2016 Econometric Modeling with High-dimensional Data in Business and Economics, *The MT International Conference on Business Research, the macro Conference Proceedings, Milano*, 62-69.

[11] Finch W H and Hernandez Finch M E 2016 Regularization Methods for Fitting Linear Models with Small Sample Sizes: Fitting the Lasso Estimator using R. Practical Assessment, Research & Evaluation, 21(7) 1-13.

[12] Sirimongkolkasem T and Drikvandi R 2019 On Regularization Methods for Analysis of High Dimensional Data. *Annals of Data Science*, 6(4) 737–763.

[13] Fan J, Han F and Liu H 2014 Challenges of Big Data analysis, *National Science Review*, 1: 293–314.

[14] Ray S 2015 7 types of regression should be Know. retrieved on 15-5-2019 from, https://www.analyticsvidhya.com/.

[15] Januaviani T MA, Gusriani N, Joebaedi K, Sukono, Subiyanto and Bon AT 2019 The LASSO (Least Absolute Shrinkage and Selection Operator) Method to Predict Indonesian Foreign Exchange Deposit Data, *Proceedings of the International Conference on Industrial Engineering and Operations Management Bangkok, Thailand*, March 5-7.

[16] AlNasser H 2017 On Ridge Regression and Least Absolute Shrinkage and Selection Operator, *M.SC. thesis, University of Victoria*, USA.

[17] James G, Witten D, Hastie T and Tibshirani R 2013 *An Introduction to Statistical Learning: With Applications in R*, Springer Publishing Company, New York.

[18] Hastie T, Tibshirani R, and Wainwright M 2015 *Statistical Learning with Sparsity: The Lasso and Generalizations*, Taylor Francis Group, Florida, USA.

[19] Jankova J, Shah R D, Buhlmann P and Samworth R J 2019 Goodness-of-Fit testing in high-dimensional generalized linear models, retrieved at 5th Feb 2020 from https://arxiv.org/pdf/1908.03606.pdf.

[20] Lambert-Lacroix S and Zwald I 2011 Robust regression through the Huber’s criterion and adaptive lasso penalty, *Electronic Journal of Statistics*, 5, 1015–1053.

[21] Tibshirani R 1996 Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B.*, 58, 267-288.

[22] Lopes H F and Marques P 2018 Multiple linear regression, retrieved at 11th Jan 2020, http://hedibert.org/wp-content/uploads/2018/05/pdf.

[23] Zou H and Hastie T 2005 Regularization and variable selection via the elastic net, *J. R. Statist. Soc. B*, 67(2), 301–320

[24] Du P, Parmeter C F and Racine J S 2013 Nonparametric Kernel Regression with Multiple Predictors and Multiple Shape Constraints, *Statistica Sinica* 23, 1347-1371.

[25] Marsh L C and Mukhopadhyay K 2014 Count Regression Model Using Kernel Regression, *paper of conference, Researchgate, uploaded by Lawrence C. Marsh on 31 May 2014*

[26] Hardle W 1994 *Applied Nonparametric Regression*, Institute for Statistics and Econometrics, Berlin.

[27] Fox J 2000 *Multiple and Generalized Nonparametric Regression*, Thousand Oaks CA: Sage.

[28] Sprent P and Seemont N C 2001 *Applied nonparametric statistical methods*, 3rd ed., Chapman & Hall/CRC, UK.

[29] Kokonendji C, Kiesse TS, Clarice G B and Demétrio, C G B 2009 Appropriate kernel regression on a count explanatory variable and applications, *Advances and Applications in Statistics*, 12(1), 99-125.