Republic of Iraq
Ministry of Higher Education
and Scientific Research
University of Al-Qadisiyah
College of Administration and Economics
Statistics Department

Sparse dimension reduction via regularized sliced inverse regression with application

A thesis Submitted to
The Council of the College of Administration and Economics at University of Al-Qadisiyah as partial Fulfillment of the requirements for the M.S.C in Statistics

By

Dheyaa Bayesh Salman

Supervised by

Prof. Dr. Ali J. Kadhim Alkenani

1442 A.H. **2021 A.D**

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

لِيَعْلَمَ أَن قَدْ أَبْلَغُوا رِسَالَاتِ رَبِّهِمْ وَأَحَاطَ بِمَا لَدَيْهِمْ
وَأَحْصَىٰ كُلَّ شَيْءٍ عَدَدًا

صدق الله العظيم

سورة الجن ايه (٢٨)

Supervisor's recommendation

I certify that the thesis entitled (**Sparse dimension reduction via regularized sliced inverse regression with application**) has been under my supervision for the student (**Dheyaa Bayesh Salman**) in the Department of Statistics / College of Administration & Economics / University of Al-Qadisiyah as it is part of the requirements for master's degree in statistics sciences.

Signature:

Supervisor: **Prof. Dr. Ali J. Kadhim Alkenani**

Date:

Recommendation of the head of the Graduate Studies Committee

Based on the available recommendation , I would like to forward this thesis for discussion.

Signature:

Name: Assist. Prof. Dr. Muhannad F. Al-Saadony

Chairman of the Higher Studies Committee in department of statistics

Approval of the linguistics expert

This is to hereby that the entitled (Sparse dimension reduction via regularized sliced inverse regression with application) has been reviewed in terms of stylistics and linguistics (grammar and spelling). Therefore, after the modification of all recommended notes thesis has become free of all linguistics errors and ready to be defended and used as a scientific method to award the degree of master in statistical sciences.

Signature:

Name: Dr. Haider Hadi Muhammad

Date:

(The Decision of the Committee)

University of Al-Qadisiyah

College of Administration and Economics

Statistics department

We are the head and members of the defense committee certify that we have been looked at the thesis entitled (**Sparse dimension reduction via regularized sliced inverse regression with application**) and we have debated the student (**Dheyaa Bayesh Salman**). As a result , the student his defended her thesis and all its content. So that we have found the thesis is worthy to be accepted to award a ( ) master's degree in statistics science.

N.	Name	Scientific Title	Signature	Capacity
1	Abbas Lafta Kneehr	Prof. Dr.		The head of the committee
2	Muhannad F.Al-Saadony	Assist. Prof. Dr		Member
3	Bahr Kadhim Mohammed	M. Dr.		Member
4	Ali J. Kadhim Alkenani	Prof. Dr.		Member and supervisor

Approval of the College Committee

The council of the college of administration and economics at the University of Al-Qadisiyah have been approved on the decision of the defense committee.

Prof. Dr. Sawsan Kareem Hodan

The Dean of the college of administration and economics/Al-Qadisiyah university

DEDICATION

THERE ARE A NUMBER OF PEOPLE WITHOUT WHOM, THIS THESIS MAY NOT HAVE BEEN WRITTEN.

TO THE SUPERVISOR, PROF. DR. ALI AL-KANANI. MY TEACHERS WHO DID NOT SPARE ANY EFFORT IN MY STUDIES. THE DIRECTOR OF MY CIRCLE, MR. ADEL LATIF AND MY FAMILY AND SOME FRIENDS.

ACKNOWLEDGMENTS

"In the name of Allah the Merciful Praise to Allah, Lord of the Worlds"

would like to express my gratitude to my supervisor Prof. Dr Ali Alkenani for his supervision of this work and showed me the right ways to get my destination, through the useful comments and remarks. I am really thankful to you.

My thanks go also to Dr. Tahir Reisan, Dr. Rahim Alhamzawi, Dr. Muhannad Alsaadony, Dr. Ahmed Naeem, Dr. Taha Alshaybawe, Dr. Fadil Alhusseini and Dr. Muhammed Habib, and all the professors for their contribution to my teaching.

My thanks and appreciation to the members of the discussion committee Prof. Dr. Abbas Lafta Kneehr, Assist. Prof. Dr Muhannad F. Al-Saadony and Dr. Bahr Kadhim Mohammed, for their evaluation of my thesis.

Finally, my father, may Allah have mercy on you, and I hope that you are satisfied with me. Mom, I am forever grateful. My sincere gratitude to my brothers, sisters and friends especially (Mustafa, Ismail, Zakaria, Muhammad and all friends).

Last, but by no means least, I would like to thank my children (Noor, Sajjad, Mustafa, Muhammad) for bearing the suffering of separation and default in some matters.

Dheyaa

Abstract

In the last two decades, the sufficient dimension reduction (SDR) theory been introduced by Cook 1998 received considerable interest (Cook, 2009). The idea of sufficient dimension reduction (SDR) is to replace X with a lower-dimensional orthogonal projection P_sX on subspaces S without loss of information about the distribution of $Y|$X and without assuming any specific model. The aim of SDR is the central subspace ($S_{Y|X}$). A lot of methods were done for finding $S_{Y|X}$ such as Principal Hessian Directions (PHD) (Li, 1992) and Sliced Average Variance Estimation (SAVE) (Cook and Weisberg, 1991) as well Sliced Inverse Regression (SIR) (Li, 1991). The SIR is applied in different areas like bioinformatics, marketing, finance and economics.

The problem of the study it that SIR has been proven a powerful for dimension reduction (DR) approach and it is effective in dealing with high-dimensional data (HD) and a sufficient tool for dealing with dimension reduction (DR) in conditional regression (Li and Yin, 2008). However, it produces linear combinations (L.Cs) of all the original predictors. As a result, the interpretation of SIR estimates could be difficult and sometimes misleading.

The objective of our study is to reduce the number of nonzero coefficients in SIR directions for obtaining better interpretability. Through combining one of the regularization methods with the SIR method to produce sparse and accurate estimations, and the SSIR-AL suggested method enables Adaptive Lasso to work with nonlinear and multi-dimensional regression without assuming any specific model.

In this thesis, a variable selection method in the concept of sufficient dimension reduction, called Sparse SIR with Adaptive Lasso penalty

(SSIR-AL), is suggested. The SSIR-AL combining the ideas of Adaptive Lasso with sliced inverse regression (SIR) to obtain a sparse SIR estimator. where leveraging the Adaptive Lasso method, variable selection and parameter estimation are implemented in one process. The effectiveness of SSIR-AL is demonstrated by both simulations and real data analysis. We made a comparison with some methods that used variable selection in the concept of sufficient dimension reduction . These methods are Ni et al. (2005) introduced the shrinkage SIR (SSIR-L) estimator by incorporating Lasso penalty with SIR, as well as with the Li and Yin (2008) method a regularised SIR (RSIR) to enable SIR to work with highly correlated predictions and when $p > n$, where n is the sample size and also with the Lasso-SIR method for the multiple index model within $p > n$ settings that was suggested by Lin et al. (2018).

TABLE OF CONTENTS

Page

The tables

List of Figures

List of abbreviations

SIR	Sliced Inverse Regression
SDR	Sufficient Dimension Reduction
SAVE	Sliced Average Variance Estimation
PHD	Principal Hessian Directions
GR	Graphical Regression
CMS	Central Mean Subspace
IHT	Iterative Hessian Transformation
MAVE	Minimum Average Variance Estimator
DR	Dimension Reduction
QR	Quantile Regression
QMAVE	Quantile MAVE
LQMAVE	Sparse Quantile MAVE with Lasso penalty

SMAVE	Sparse MAVE
Lasso	Least absolute Shrinkage and Selection Operator
SCAD	Smoothly Clipped Absolute Deviation
MCP	Minimax Concave Penalty
SSIR-L	Shrinkage Lasso with SIR
SSIR-AL	Sparse SIR with Adaptive Lasso penalty
OP's	Oracle Properties
SSIR	Sparse Sliced Inverse Regression
SSIRQ	Sparse Sliced Inverse Quantile Regression
C.S	Center Subspaces
LCs	Linear Combinations
V.S	Variable Selection
HD	High-dimensional Data
AIC	Akaike Information Criteria
BIC	Bayesian information Criteria
GLM	General Linear Model
OLS	Ordinary Squares Estimates
MSE	Mean Squares Error
SD	Standard Deviation
MLE	Maximum Likelihood Estimation
Ave 0's	Average number of zero coefficients
C.V	Cross Validation
i.i.d	independent identically distributed

Chapter One

Introduction,

reduce dimensionality methods

and Literature Review

1.1. Introduction:

In some multiple regression applications, the number of predictors have become large, which is why, the analysis of that data has become difficult. For the purpose of dealing with this problem, it is necessary to perform dimensionality shrinkage of data with a few assumptions. when talking about dimensionality reduction, it indicates the fact that there are high dimensions, those dimensions have been referred to as the variables. Increasing the number of these variables in the multiple regression model, means that the model will be more difficult to analysis the data. Thus, a problem will be encountered, known as the curse of dimensions where term curse of dimensions was introduced by Bellman (1961) when data is sparse in multi-dimensional spaces. Also in the case of linear correlation between the high-dimensional data (HD). In which the greater number of the variables make it more difficult predicting a certain quantity. Thise variables might not be all influential or effective, or can be interconnected and thereby, redundant which will require reduction. Which is why, dimensionality reduction process means the conversion of the (HD) into a space of a smaller size. It has a significant impact to solve this problem and by reducing the number of the random variables, in other words, simplifying the understanding of the data only visually or numerically and thus ensuring the integrity of data, moreover, there are other advantages to the reduction of dimensionality where it operates on data compression and reduces the time consuming and there are some methods which do not operate efficiently in the cases of very high dimensions. Thus, it is necessary to work on the reduction of the dimension and making the methods beneficial.

In addition to, it is necessary to deal with the problems of the correlation and can be done by the remove duplicate properties. where

decrease the number of variables in a dataset, without losing much information at the same time as maintaining or improving the performance of the model. The reduction of dimensionality can be performed in two ways the variable selection and variable extraction (Malik and Alkenani, 2019).

1.2.Variable selection (VS):

A method which is extensively utilized in multiple regression and classification for selecting partial sets of the variables, where subgroups of the original set of variables have been specified to obtain the smallest subset which may be utilized for modeling the problem. The procedure of variables selection is helpful for giving a clearer understanding of data through providing information about the important variables as well as their relationship with one another, also, it reduces the cost. In general it has been divided into two types of methods, which are: traditional and regularization method, examples of the traditional methods include stepwise selection (Efroymson, 1960), Akaike information criteria (AIC) (Akaike, 1973) and Bayesian information criteria (BIC) (Schwarz, 1978). If they compared to regularization methods, one can notice the instability and high variance in traditional methods. whereas, the regularization methods that have been first used (for determining the variable) by Donoho and Johnstone (1994). Examples for regularization methods are the Lasso (Tibshirani, 1996), (SCAD) (Fan and Li, 2001), Elastic Net (Zou and Hastie, 2005) and Adaptive Lasso (Zou, 2006) amongst others. Those methods have high stability compared to the traditional methods because the parameters estimation and the variables selection are carried out simultaneously (Alkenani and Yu, 2013).

1.3. Variable extraction:

The variable extraction is the process of transforming (projecting) the variables into a fewer number of new ones. It is sharing the objectives of sub-set selection, the major difference is that the results should be specified with regard to all variables. It is indicating the process to find transformation which is projecting data from original to feature space. This technique attempts to enable picturing of the data through decreasing the p-dimensional predictor vector X dimension without losing information. A lot of variable extraction approaches were suggested to reduce dimensionality, with no loss of information from data. These include principal component analysis (see Jolliffe, 2002; Zhang and Olive, 2009), factor analysis (Gorsuch, 1983), independent component analysis (Common, 1984), canonical correlation analysis (Hotelling, 1936; Fung et al., 2002; Branco et al., 2005; Zhou, 2009; Zhang, 2011; Alkenani and Yu, 2013),single index models (Powell et al., 1989; Härdle and Stoker, 1989; Ichimura, 1993; Delecroix et al. 2003), the Sliced Inverse Regression (SIR) (Li, 1991),the Sliced Average Variance Estimation (SAVE) (Cook and Weisberg, 1991), the Principal Hessian directions (PHd) (Li, 1992), MAVE and the Outer Product of Gradients (OPG) methods (Xia et al., 2002, see also Xia 2007, 2008) and successive direction estimation (Yin and Cook, 2005; Yin et al, 2008), among others.

1.4. Some methods for Variable selection:

Variable selection is important for the construction of the multiple regression model. It operates on improving the prediction of the models, which results in a low cost model (Guyon and Elisseeff, 2003) some of these methods are discussed below:

1.4.1. Traditional Variable selection:

Variable selection methods, such as the stepwise selection (Efroymson ,1960), Akaike information criteria AIC (Akaike, 1973)and Bayesian information criteria BIC (Schwarz, 1978), are time consuming, discrete processes with high variance and suffer from instability (Brieman, 1996).

1.4.1.1. Step wise selection procedure:

This method is a development of the forward selection method, and its basis has been established (Efroymson, 1960) to make it more efficient, and the point of distinction between those two methods is that every independent variable at the end of every step is ascertained through depending upon on the choice of ($F_{partail}$), and re-evaluated once more because there are strong relations between independent variables which have been introduced in the earlier steps.(Jabbar and Alkenani, 2020).

1.4.1.2. Forward selection procedure:

This method is dependent upon starting with no independent variables and the independent variables are selected to be included in the equation one after another based upon the comparison ($F_{partial}$) for every one of the variables with a value ($F_{tabular}$). The maximum value is selected ($F_{tabular}$) which is referred to as the (F_{IN}), for every step and after ensuring that value is higher than (F_{IN}), the variable in question is entered in the equation and the steps keep showing the independent variables one after another to the point of getting to the top ($F_{partial}$) less than (F_{IN}) based on the equation below.

$$F^* = \frac{SSR(x_1)}{\frac{SSE(x_1)}{n-2}} \quad (1.1)$$

where SSR:-represents the deviations shown

SSE:-represents the unclarified deviations

n:- sample size.

1.4.1.3. Backward elimination procedure:

This method begins with adopting all the independent variables in the equation and after that, deleting the variables from the equation one after another based on the value of ($F_{tabular}$) which has been referred to as (F_{IN}) the steps of this process are described below:

Steps 1:- including every independent variable in the equation of regression and calculating the values of ($F_{partial}$) for every one of the variables based on the formula below:-

$$F_{i\ partial} = \frac{SSR\left[\frac{x_i}{all\ other\ explanatory\ variables}\right]}{\frac{SSE(x1,\ldots\ldots,xk)}{n-k-1}} \quad (1.2)$$

Then, selecting a variable which has the minimum value of ($F_{partial}$), and comparing it to (F_{IN}). If it proves that (Fi) < (F_{IN}) the relevant variable will be eliminated from the equation and the second step is moved to the freedom degree of the numerator (1) and denominator (n-k-1).

Step 2 :- every independent variable except the ones that have been eliminated in step (1) are included, for all remaining variables from step 1, the smallest value is selected for ($F_{partial}$) and compared with (F_{IN}) to the freedom degree of numerator (1) and denominator (n-k-2). If ($F_{partial}$) < (F_{IN}), eliminates the variable in question and move to step (3), thereby

this step continue to the point of obtaining the minimal value (F_{IN}) < ($F_{partial}$), and then, the solution is stopped. (Malik and Alkenani, 2019).

1.4.1.4. Akaike Information Criteria(AIC):

Akaike Information Criteria (Akaike, 1973) is a technique based on the comparing the quality of a set of models with each other. AIC is a measure of relative statistical modeling quality. Fundamental aim is to make a distinction between estimated models through the dependence upon its AIC minimal value (Sugiura, 1978), where the model which has the minimal AIC value has been considered the best one. As the criterion basically measures the quantity of lost information, is expressed as:

$$AIC(k)=-2Ln(L)+2k \qquad (1.3)$$

where k: is the number of parameters.

L: is the value of MLE. (Akaike, 1973).

1.4.1.5. Bayesian information criteria (BIC):

Schwarz (1978) suggested the BIC. It is a criterion of selecting model from a limited group of models .It is similar to the AIC, however it differs from it through including the size of the sample where it has been taken into account which makes it more sufficient compared to the AIC (Carlos and Sergioc, 2012). The model that has the minimum value is considered the optimal and it has been expressed by the following equation.

$$BIC(k)=-2Ln(L)+k\,Ln(n), \qquad (1.4)$$

where n: represents the number of observations, k : is the number of parameters, L: is the value of MLE. (Malik and Alkenani, 2019).

1.4.2. Regularization methods:

Regularization techniques can perform the V.S as well, thus, the approaches of regularization can be defined as the method which is used for solving the problem of models complexity. The performance of the generalization is tightly associated with the model of complexity. where the model with high complexity tends to have high variance and low bias. While the low complexity model tends to have low variance and high bias, thus the methods of regularization are often utilized to control the complexity of the model through penalizing the models of higher complexity. First, we use of the methods of regularization for V.S that has been made by Donoho and Johnstone (1994). Regularization approaches can be formed through the addition of penalty term to standard loss functions, as O.L.S loss function. In regularization methods the V.S is implemented with the process of the parameter estimation. Examples of regularization approaches are the Lasso (Tibshirani, 1996), SCAD (Fan and Li, 2001), Elastic Net (Zou and Hastie, 2005), Adaptive Lasso (Zou, 2006), group Lasso (Yuan and Lin, 2006), OSCAR (Bondell and Reich, 2008), MCP (Zhang, 2010) and PACS (Sharma et al., 2013). (Malik and Alkenani, 2019).

1.4.2.1. Lasso:

Lasso is conceptually quite similar to the Ridge regression. Ridge regression adds the summation of the squared coefficients (i.e. the l_2 penalty), but Lasso adds the summation of their absolute values (l_1 penalty).Lasso has been suggested by Tibshirani (1996) for the estimation of parameters and V.S together. It is considered an effective and powerful toll to tackle the HD. In Lasso, the RSS has been minimized subject to $\sum_{k=1}^{p}|\beta_k|$ being less than a constant. Based on this condition, Lasso

shrinks some of the coefficients and eliminates others through zeroing its coefficients. In this approach (Tibshirani, 1996) added a penalty function to the function of the least squares loss as in the equation below:

$$\hat{\beta}(Lasso) = \arg\min \sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda \sum_{k=1}^{p} |\beta_k| \qquad (1.5)$$

$\lambda \geq 0$ controls the strength of penalty, the largest value of λ provides higher shrinkage level (Alkenani and Yu, 2013), the value of (λ) is determined through the Generalized Cross Validation by minimizing

$$\text{GCV} = \frac{RSS}{n\{1-p(\lambda)/\text{n}\}^2},$$

where RSS $= \sum_{i=1}^{n}(y_i - x_i^T\beta)^2$,

$\hat{\beta}_k$ represents OLS estimates, p represent the number of variables $k = 1,\ldots, p$, n represent the sample size i=1,2, ..., n and $p(\lambda)$ the effective number of parameters, larger values of $p(\lambda)$ cause more inflation (penalization).

1.4.2.2. Adaptive Lasso (*A L*):

Lasso has been suggested by Tibshirani (1996) for the estimation of parameters and V.S together. It is considered an effective and powerful to tackle the HD. Fan and Li (2001) had concluded that this method has a bias in the estimation large non-zero coefficients and showed as well that it didn't possess Oracle properties (OP's), and the definition of (OP's): that the method that has this property has the ability to choose the real model with a probability of one magnitude. Zou (2006) suggested a new approach that had been referred to as the Adaptive Lasso, where the concept of this method assigns various adaptive weight values to the variety of the parameters in penalty function l_1, and that results in a reduction in penalty for the parameters which are near the value of zero,

then reduces the bias in estimation of the parameters and enhances the accuracy of the variable selection (Chand and Kamal, 2011), (Xia , Yi , 2007). Zou (2006) showed that the Adaptive Lasso approach is a partial weight approach for l_1 penalty function for the estimation and selection of the parameters simultaneously. He also proved that the Adaptive Lasso method can achieve the (OP's). Adaptive Lasso estimates may be obtained from the equation below:

$$\hat{\beta}(A\,L) = \arg\min \sum_{i=1}^{n}(y_i - x_i^T \beta)^2 + \lambda \sum_{k=1}^{p} \breve{w}_k \,|\beta_k| \qquad (1.6)$$

$\lambda > 0$ is the tuning parameter, p represent the number of variables $k = 1,..., p$, $W_k = (w_1, w_2, ..., w_p)$ adaptive weights, and the (Zou) showed that if the choice of the weights (W_k) has been effective and in a data-driven method, then the Adaptive Lasso approach can achieve (OP's) so that they are executed as if the correct model has been known(Zou, 2006), and calculation is done that the estimated weight values as follows:

$\breve{w}_k = \frac{1}{|\hat{\beta}_{ols}|^{\gamma}}$ is a known weights vector. Using OLS to choose $\breve{w}_k$ where $\hat{\beta}_k$ represents OLS estimates and is a consistent primary estimator $\sqrt{n}$ to β (containing $\sqrt{n}$ affinity ratio), and γ: represents the contraction parameter and its value is greater than zero, most researchers who used the Adaptive Lasso method agree to make a value (γ) equal to one(γ=1).

1.4.2.3. Elastic Net (EN) :

Elastic Net method combined Ridge regression method and the Lasso method, the Ridge is used to deal with the data when the number of variables is greater than the sample size $p > n$, and the Lasso method performs the process of selection the important variables and neglecting the unimportant ones. Zou and Hastie (2005) suggested the Elastic Net. Thus, they dealt with the problems that appear in the work of the Lasso

method, they studied the Lasso and pointed out several problems of the Lasso as follow:

1. In a case where $p > n$ the Lasso selects nearly n variables.

2. In the case where there is a group of the strongly related variables, Lasso will choose only one from this set and ignore the remaining variables. Elastic Net estimates may be obtained as a solution of (Zou and Hastie, 2005):

$$\widehat{\beta}(EN)=\text{argmin}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda_1\sum_{k=1}^{p}\beta_k^2 + \lambda_2\sum_{k=1}^{p}|\beta_k| \quad (1.7)$$

where β_k^2 represents the norm which is related to the Ridge penalty and $|\beta_k|$ is the norm which is associated with Lasso ,and $\lambda_1, \lambda_2 \geq 0$ are the tuning parameters. The Elastic Net penalty is a combination of both penalties.

1.5 . Literature Review for shrinkage SIR:

Sliced inverse regression is an efficient tool for obtaining the SDR(Li and Yin, 2008), nonetheless, all the approaches of the SDR suffers from that every component of the DR is a linear combinations (L.Cs) of predictors, which makes it difficult to explain the resultant estimations. Under the framework of the SDR, related to the studies that used the SIR with some regularization methods to simultaneously obtain parameter estimation and predictors selection, model-selection method for single-index models is suggested by Naik and Tsai (2001). Also, Cook (2004) suggested a method for assessing the contribution of variables. Furthermore, Ni et al. (2005) introduced the shrinkage SIR (SSIR-L) estimator by incorporating Lasso penalty with SIR. Moreover, Li and Nachtsheim (2006) suggested sparse SIR (SPSIR) by merging Lasso and LARS with SIR. Li (2007) merged the regression formulation of some

SDR methods with the ideas of regularization methods. Li (2007) employed this procedure to SIR and some of SDR methods. Li and Yin (2008) suggested a regularised SIR (RSIR) to enable SIR to work with highly correlated predictors and when $p > n$, where n is the sample size. The sliced inverse quantile regression method (SIQR) was suggested by Alkenani and Dikheel (2016). Moreover, the sparse SIQR with Lasso and Adaptive Lasso penalties is suggested in the same paper. The Lasso-SIR method for the multiple index model was suggested under $p > n$ settings by Lin et al.(2018).The authors demonstrated that Lasso-SIR is consistent and achieve the optimal convergence rate. As for some of studies under the framework of the SDR, used MAVE as Wang and Yin (2008) suggested sparse MAVE (SMAVE) approach. Alkenani and Yu (2013) suggested SMAVE with the Adaptive Lasso, SCAD and MCP penalties. Malik and Alkenani, (2019) suggested QR with MAVE (QMAVE) and QMAVE with Lasso penalty (LQMAVE). Jabbar and Alkenani, (2020) suggested SMAVE with the Elastic Net and Adaptive Elastic Net. The contribution in this thesis is a variable selection method under the concept of SDR, called SSIR-AL, is suggested. The SSIR-AL combines the ideas of Adaptive Lasso with SIR to obtain a sparse SIR estimator. The details of SIR will be reported in (Chapter 2). The rest of the thesis is organized as follows: In chapter 2, review of SIR method, SSIR-L were reported and sparse SIR with Adaptive Lasso penalty(SSIR-AL) is suggested. In chapter 3,simulation studies are implemented, and the considered methods were applied with real data. The conclusions and future works were reported in chapter 4.

Chapter Two

Sparse SIR with some Regularization methods

2.1. Introduction:

Dimension reduction method is one of the important methods that the researcher needs in analysis high-dimensional data(HD), especially in recent years. After the development of data collection methods and the development of data storage methods and storage capacity, many dimensional reduction(DR) methods have been suggested, and they can be divided into two types of methods. Classical shorthand methods such as Principal composites analysis (PCA) method, factor analysis, discriminant analysis, and other methods. These methods began to suffer in the analysis of HD. Therefore, Cook 1998 suggested the theory of Sufficient dimensions reduction (SDR) Cook (2009). The well-known approaches of the SDR provide the tool for finding sufficient dimensions with no need for pre specifying an error distribution or a model. Those approaches replace original variables with the linear combinations(LCs) of predictors in which they're low-dimensional. However, the explanation of resulting estimations isn't simple due to the fact that every one of the DR components is a linear combinations of every original predictor.

In this chapter, a summary of the SIR and shrinkage SIR(SSIR-L) is presented in addition to a suggested for the combination of the ideas of shrinkage of Adaptive Lasso with the SIR (SSIR-AL), in order to produce sparse and accurate solutions.

2.2. Sliced Inverse Regression (SIR):

Li (1991) suggested SIR method for estimating the basis of central subspace, denote it by $S_{y|\mathbf{x}}$. Let y is a response variable and $X = (x_1, \dots, x_p)^T$ is a p-dimensional predictor in high dimensional regression model. The basis of this method is to reverse the relationship in the traditional (classical) regression analysis. Regression analysis study the correlation of the dependent variable (y) with the independent variables (X) represented by $E(Y|X)$.While SIR study this relationship through $E(X|Y)$(Li, 1991). In the last two decades, the SDR theory was received considerable interest. The idea of SDR is to replace X with a lower-dimensional orthogonal projection $P_S X$ on subspaces S without loss of information about the distribution of $Y|X$ and without assuming any specific model. The aim of SDR is the central subspace ($S_{Y|X}$). The $S_{Y|X}$ is the intersection of all subspaces S such that.

$$Y \perp\!\!\!\perp X | P_S X \qquad (2.1)$$

where $\perp\!\!\!\perp$ indicates independence. Consequently, $P_\beta X$ extracts all of the information from X about Y, where β is a basis of $S_{Y|X}$, see Cook (2009). A lot of methods were done for finding $S_{Y|X}$ and one of these methods is SIR .(Li, 1991). The SIR is applied in different areas like bioinformatics, marketing, finance and economics.

Li (1991) suggested SIR method for estimating the basis of $S_{y|\mathbf{x}}$. The SIR requires $Z = \sum_x^{-\frac{1}{2}}(X - E(X))$, satisfy the condition $E(Z|P_c Z) = P_c Z$, where $\Sigma_x = Cov(X)$ is the population covariance matrix of X and c is a basis for $S_{Y|Z}$. This condition connects $S_{Y|Z}$ with the inverse regression of Z on Y. The symmetric kernel matrix of SIR is $M = cov\,[E(Z|Y)]$ and $Span(M) \subseteq S_{Y|Z}$. Let a random sample of size n of (X, Y), which has a

joint distribution. Let $\bar{X}$ is the sample mean of X. Also, assume that $\hat{Z} = \hat{\Sigma}^{-\frac{1}{2}} (X - \bar{X})$ is the sample version of Z, where $\hat{\Sigma}$ is the sample covariance matrix of X. Let h is the number of slices and n_y is the number of observations in the yth slice. (Li, 1991). Thus, $\hat{M} = \sum_{y=1}^{h} \hat{f}_y \hat{Z}_y \hat{Z}_y^T$ is the sample version of M, where $\hat{f}_y = n_y/n$ and $\hat{Z}_y$ is the average of Z in the slice y. Let $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_p \geq 0$ are the eigenvalues corresponding to the eigenvectors $\hat{v}_1, \hat{v}_2, \ldots\ldots, \hat{v}_p$ of $\hat{M}$. If the dimension d of $S_{Y|Z}$ is known, $span(\hat{\beta}) = span(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_d)$ is a consistent estimator of $S_{Y|X}$, where $\hat{\beta}_i = \hat{\Sigma}^{-\frac{1}{2}} \hat{v}_i$. The model on which the SIR relies is similar to the semi-parameter regression model:

$$y = f\ (\ \beta_1^T X\ , \beta_2^T X, .\ \ .\ \ .\ , \beta_k^T X, \epsilon\ ) \qquad (2.2)$$

Since that (β's) unknown vector, ϵ random error independent of X and f is an unknown arbitrary function $\mathbb{R}^{k+1}$.

where collecting all sectors information and obtaining the underlying roots. Then the largest of which will be selected to represent the effective dimension reduction (e d r) vectors of the (SIR) respectively. Which represents the new format of data that act as parameters (B_Ks) ,where data is converted to reduced form and replaced with the original data for ease of handling and in this method the (HD)problem is remedy in SIR method .We can summarize the algorithm of SIR as follows:

1- Standardize X via affine transform, for the purpose of getting

$\check{X}_i = \hat{\sum}_{xx}^{-1/2}(x_i - \bar{X})$, (i= 1,. . . ,n) where $\hat{\sum}_{xx}$ represents the covariance matrix of the sample and $\bar{X}$ represents sample's average of X respectively .

2- Dividing the range of y to H of slices, $I_1, \ldots, I_H$; assuming that the proportion of y_i that falls in sliced h be $\hat{p}_h$ that is $\hat{p}_h = (1/n) \sum_{i=1}^{n} \delta_h(yi)$

where $\delta_h(yi)$ can take 2 values either 0 or 1 according to whether yi falls into the $h\,th$ sliced I_h or not .

3- In every one of the slices, sample mean of $\breve{X}_i$ s, is computed, which is denoted by $\hat{m}_h$ (h=1 , . . . , H), such that $\hat{m}_h = (1/\mathrm{n}\,\hat{p}_h)\sum_{yi\in I_h}\breve{X}_i$.

4- Conduct a (weighted) (PCA) for the data $\hat{m}_h$ (h= 1, . . . , H) as: for weighted covariance matrix $\hat{V} = \sum_{h=1}^{H}\hat{p}_h\ \hat{m}_h\ \hat{m}_h{}^T$ then find eigenvalues and eigenvectors for $\hat{V}$.

5- Assuming that K largest eigenvectors be the (row vectors) $\hat{\eta}_k$(k=1,...,K). output

$$\hat{\beta}_k = \hat{\eta}_k\ \hat{\sum}_{xx}^{-1/2}, \qquad (2.3)$$

where (k=1,... ,K), $\eta_k = \beta_k\ \sum_{XX}^{1/2}(\ k = 1, \ldots, K)$ and $\sum_{XX}$ the covariance matrix of X. Steps 2 and 3 produce a crude estimate of the standardised inverse regression curve E(Z | y).(Li, 1991).

2.3. Shrinkage SIR method (SSIR-L):

In the framework of SIR and to obtain parameter estimation and predictor selection simultaneously, model-selection method for single-index models was suggested by Naik and Tsai (2001). Also, Cook (2004) suggested a method for assessing the contribution of variables. Furthermore, Ni et al.(2005) introduced the shrinkage SIR (SSIR-L) estimator by incorporating Lasso penalty with SIR. The SIR provides an estimator $span(\hat{\beta})$ of $S_{Y|X}$. Usually, the elements of $\hat{\beta} \in \mathbb{R}^{p\times d}$ are nonzero. In the construction of 'sufficient predictors', only the important predictors are needed if the number of predictors is large or the predictors are highly-correlated. To this end, a number of authors suggested to employ the variable selection technique with SIR to compress some rows

of $\hat{\beta}$ to 0's. To improve interpretability, Cook (2004) formulated SIR as a regression type optimization problem by minimizing:

$$F(A,C) = \sum_{y=1}^{h} \left\| \hat{f}_y^{1/2} \hat{Z}_y - AC_y \right\|^2, \quad (2.4)$$

over $A \in \mathbb{R}^{p\times d}$ and $C_y \in \mathbb{R}^d$, with $C = (C_1, \ldots, C_h)$. Let $\hat{A}$ and $\hat{C}$ are the values of A and C that minimise F. Then $span(\hat{A})$ equals the space spanned by the d largest eigenvectors of M. By focusing on the coefficients of the X variables, Ni et al.(2005) rewrite F(A, C) as

$$G(B,C) = \sum_{y=1}^{h} \left(\hat{f}_y^{1/2} \hat{\Sigma}^{-\frac{1}{2}} \hat{Z}_y - BC_y \right)^T \hat{\Sigma} \left(\hat{f}_y^{1/2} \hat{\Sigma}^{-\frac{1}{2}} \hat{Z}_y - BC_y \right), \quad (2.5)$$

where $B \in \mathbb{R}^{p\times d}$. The value of B which minimizes (2.5) is exactly $\hat{\beta}$ and $span(\hat{\beta}) = span\left(\hat{\Sigma}^{-\frac{1}{2}} \hat{A}\right)$ is the estimator of $S_{Y|X}$. After that, the authors suggested a shrinkage SIR estimator (SSIR) of $S_{Y|X}$ is $span(diag(\tilde{\alpha})\hat{\beta})$, where the shrinkage indices $\tilde{\alpha} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_p)^T \in \mathbb{R}^p$ are determined by minimizing:

$$\sum_{y=1}^{h} \left\| \hat{f}_y^{1/2} \hat{Z}_y - \hat{\Sigma}^{\frac{1}{2}} diag(\hat{B}\hat{C}_y)\alpha \right\|^2 + \lambda \sum_{i=1}^{p} |\alpha_i| , \quad (2.6)$$

where, $\hat{B}$ and $\hat{C} = \left(\hat{C}_1, \ldots, \hat{C}_h\right)$ minimize (2.5).

By using a standard Lasso algorithm, the minimization problem of (2.6) can be implemented. To be specific, let

$$\tilde{Y} = vec(\hat{f}_1^{1/2} \hat{Z}_1, \ldots, \hat{f}_h^{1/2} \hat{Z}_h) \in \mathbb{R}^{ph}$$

and $\tilde{X} = \left(diag\left(\hat{B}\hat{C}_1\right)\hat{\Sigma}^{\frac{1}{2}}, \ldots, diag\left(\hat{B}\hat{C}_a\right)\hat{\Sigma}^{\frac{1}{2}}\right)^T \in \mathbb{R}^{ph\times p}$,

where $vec(.)$ is a matrix operator that stacks the matrix's columns to a single vector. Then the vector α, is exactly the estimator of Lasso for the regression $\tilde{Y}$ on $\tilde{X}$. (Ni et al., 2005).

For the rest of the methods in which we compared, we used the (RSIR) method suggested by Li and Yin (2008) to enable the SIR approach to work on a large scale based on the least squares formulation.

L2 normalization was introduced, and an alternative least squares algorithm was developed, to enable SIR to work with $p \geq n$ and highly correlative prediction. L1 regulation was also introduced to achieve reduction estimation and predictor selection simultaneously. See (Li and Yin ,2008). As well as the method of Lin et al. (2018) (SIR-L) also suggested a SIR scale approach based on the least squares formulation. To enable SIR to operate with $p \geq n$, L1 regulation was introduced to achieve reduction estimation and predictor selection simultaneously. See (Lin et al. 2018)

2.4.Sparse SIR with Adaptive Lasso penalty (SSIR-AL):

In this thesis, the SSIR-AL is suggested. The SSIR-AL incorporates adaptive Lasso penalty with SIR to produce sparse and accurate solution. In equation (2.6), Lasso penalty was employed with SIR. In spite of the advantages of Lasso, it has some limitations. The Lasso estimator is a biased estimator, which means it is an inconsistent because it penalises all coefficients with the same amount. Consequently, Lasso does not have the property of selecting the predictors with nonzero coefficients with probability equal to one (Fan and Li, 2001). This property is called the oracle property. To settle this limitations, the Adaptive Lasso was suggested by Zou (2006) and the Adaptive Lasso method has been shown to have the property of oracle (Zou, 2006). The author used adaptive weights for penalising different coefficients. The SSIR-AL has advantages over the sparse SIR methods with Lasso in its ability on oracle variable selection under the framework of the SDR. It benefits from the strength of Adaptive Lasso in estimating the parameters and variable selection simultaneously. The superiority of the suggested SSIR-AL method was proved by the results obtained in the simulation

and real data. The suggested SSIR-AL method is obtained by minimize:

$$\sum_{y=1}^{h} \left\| \hat{f}_y^{1/2} \hat{Z}_y - \hat{\Sigma}^{\frac{1}{2}} diag(\hat{B}\hat{C}_y)\alpha \right\|^2 + \lambda \sum_{i=1}^{p} \breve{w}_i \ |\alpha_i| \ , \qquad (2.7)$$

for p represents the number of variables i = 1, . . . , p , where B $\in \mathbb{R}^{p\times d}$ and $C_y \in \mathbb{R}^d$with C = (C_1,...,C_h). The B value is exactly $\hat{\beta}$ and $span(\hat{\beta}) = span\left(\hat{\Sigma}^{-\frac{1}{2}} \hat{A}\right)$ is the estimator of $S_{Y|X}$, shrinkage SIR estimator (SSIR) of $S_{Y|X}$ is $span(diag(\tilde{\alpha})\hat{\beta})$,where the shrinkage indices $\tilde{\alpha} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_p)^T \in \mathbb{R}^p$ are determined by (2.6), h represents the number of slides in model y=(1,...,h), $\breve{w}_i = 1/|\hat{\beta}_{ols}|^{\gamma}$ is a known weights vector. Using regular (OLS) to choose $\breve{w}_i$ where $\hat{\beta}_{ols}$ represents an initial estimate of method OLS , γ represents the contraction parameter and its value is greater than zero, most researchers who used the Adaptive Lasso method agree to make a value (γ) equal to one (γ=1). The λ is tuning parameters and it can be estimated by CV, $\hat{f}_y = n_y/n$ and $\hat{Z}_y$ is the mean value of $\hat{Z}$ in y-th slice, here $\hat{Z} = \hat{\sum}^{-\frac{1}{2}}(X - \bar{X})$.

Chapter Three

Simulation study and Real Data

3.1. Introduction:

In this chapter, what we have presented above in the theoretical side of the thesis was applied in accordance with its goal. In order to achieve this goal it was necessary to do several studies under different (experimental) cases to determine the ability of the suggested method to process high-dimensional data. We will present the results of our method and compare it results with those of the other methods that used SIR with the Lasso method. Since four numbers of explanatory variables (X) were generated (15,24,40,200) to suit the reality of the problem under study, and different (4) of the samples sizes were used (n= 50, 100, 200, 300). Also we have a realistic case study (real data) about diabetes, where we considered glucose reading the dependent variable (y) and explanatory variables (X) it influences on this reading , where was obtained from analysis made by patients visiting the Diabetes Consulting Clinic at Al Zahra Teaching Hospital in Al Kut, Wasit Governorate. In addition to a group of direct questions from the researcher to those people, this work was done on a sample included (n = 82) persons.

3.2. Simulation study:

In this section, five simulation examples were done to compare the performance of the suggested method, SSIR-AL, with methods that combine SIR work with the Lasso method. Ni et al.(2005) (SSIR-L), Li and Yin (2008) Sliced Inverse Regression with Regularizations(RSIR) and Lin et al. (2018) Sparse Sliced Inverse Regression Via Lasso (SIR-L) methods. In these examples various numbers of variables and sample sizes were considered. In all examples, SSIR-AL was computed as a two steps procedure as described in chapter 2. The SSIR-L method were computed using R codes made by Liqiang Ni. The RSIR method were computed using R code made by Lexin Li. The function *LassoSIR* from the R package (LassoSIR) was used to compute SIR-L estimates. For each competitor, the tuning parameter was chosen via cross-validation(CV).To evaluate the efficiency of the variable selection for the suggested method, we report the average number of zeros the coefficients (Ave0's) and mean standard deviation (SD) of the absolute correlation |r| between the estimated predictor $X^T\hat{\beta}_i$ and the true one $X^T\beta_i$ and to evaluate the precision of the estimation we report the mean and SD of the mean squared error (MSE) $E(X^T\hat{\beta}_i - X^T\beta_i)^2$. As shown in the examples below:

Example 1**:** R=500 datasets were generated with size n=50 , 100 and 200 from the model y=$\frac{(X^T\beta_1)}{(0.5+(X^T\beta_2+1.5)^2)}+0.5\ \varepsilon$, where X=$(X_1,\dots,X_{15})^T$, X_i and ε are independent and are identically distributed(i.i.d) from an N(0,1), $\beta_1=(1,0,\dots,0)^T$and $\beta_2=(0,1,0,\dots,0)^T$with $S_{y|\mathbf{x}}=\text{span}(\mathbf{B}_2)$. This means, the model is y= $\frac{X_1}{\{0.5+(X_2+1.5)^2\}}+0.5\ \varepsilon$.

Example 2**:** R=500 data sets were generated with size n=50 and 100 and 200 from the linear model $y = X^T\beta + 2\,\varepsilon$, where X=$(X_1, \dots, X_{40})^T$, $X_i$and ε are i.i.d from an N(0, 1) and $\beta = (2, 2, 2, 0, \dots, 0)^T$with $S_{y|\,\mathbf{x}}$=span($\mathbf{B}_1$). In order to evaluate the performance of the SSIR-AL when the predictors are correlated, we generate X from a N(0, Σ) with $cov(X_i, X_j) = 0.7^{|i-j|}$ for this model. This means, model is $y = 2X_1 + 2X_2 + 2X_3 + 2\,\varepsilon$.

Example 3: R=500 data sets were generated with size n= 100 observations from the model y = sign $(X^T\beta_1)$log $(|X^T\beta_2 + 9|) + 0.5\,\varepsilon$, where X=$(X_1, \dots, X_{24})^T$, $X_i$and ε are i.i.d from an N(0, 1). There are three different forms for β_1 and β_2, namely:

$(1)\beta_1 = (2, 2, 2, 2, 0, \dots, 0)^T$ and $\beta_2 = (0, \dots, 0, 2, 2, 2, 2)^T$.

$(2)\beta_1 = (2, 2, 0, 2, 0,\ 2, 0, \dots, 0)^T$ and β_2 =$(0, \dots, 0, 2, 2, 0, 2, 0, 2)^T$

$(3)\beta_1 = (2, \dots, 2, 0, \dots, 0)^T$ and β_2 =$(0, \dots, 0, 2, \dots, 2)^T$,where each β has 15 elements equal to 2 with $S_{y|\,\mathbf{x}} = \text{span}\,(\mathbf{B}_2)$.

Example 4**:** R=500 data sets were generated with size n=50 and 100 and 200 from the linear model y= $X^T\beta$ +0.9 ε, where X=$(X_1, \dots, X_{24})^T$, $X_i$and ε are i.i.d from an N(0, 1) and $\beta = (1, 1, 1, \dots, 1)^T$with $S_{y|\,\mathbf{x}}$=span($\mathbf{B}_1$).This means, the model is

$y = X_1 + X_2 + \dots + X_{24} + 0.9\,\varepsilon$.

Example 5**:** R=500 data sets were generated with size n=300 from the linear model y= $X^T\beta$ +0.5 ε, where X=$(X_1, \dots, X_{200})^T$, $X_i$and ε are i.i.d from an N(0, 1) and $\beta = (2, 2, 2, 2, 0, \dots, 0)^T$with $S_{y|\,\mathbf{x}}$=span($\mathbf{B}_1$). In order to evaluate the performance of the suggested method when the predictors are correlated, we generate X from a $N(0, \Sigma)$ with $cov(X_i, X_j) = 0.7^{|i-j|}$ for this model.

Table 3.1: Results of simulation for the approaches that were examined depending on example 1.

n=50

		SSIR-AL	SSIR-L	SIR-L	RSIR
$\hat{\beta}_1$	*Ave 0's*	11.98	8.89	11.72	7.90
	Mean |r|	0.9612	0.9311	0.9585	0.9512
	SD|r|	0.0469	0.0522	0.0488	0.0497
	Mean MSE	0.00063	0.00094	0.00080	0.00084
	SD MSE	0.00019	0.00066	0.00025	0.00029
$\hat{\beta}_2$	*Ave 0's*	11.90	5.66	10.75	5.93
	Mean |r|	0.9600	0.8222	0.9514	0.9242
	SD|r|	0.0554	0.1087	0.0655	0.0799
	Mean MSE	0.00082	0.00133	0.00094	0.00098
	SD MSE	0.00088	0.00322	0.00101	0.00178

Example 1, n=100

		SSIR-AL	**SSIR-L**	**SIR-L**	**RSIR**
$\hat{\beta}_1$	*Ave 0's*	12.41	9.93	12.12	12.27
	Mean |r|	0.9682	0.9334	0.9592	0.9555
	SD|r|	0.0452	0.0514	0.0474	0.0486
	Mean MSE	0.00069	0.00084	0.00072	0.00078
	SD MSE	0.00017	0.00062	0.00024	0.00024
$\hat{\beta}_2$	*Ave 0's*	12.33	8.42	11.52	10.88
	Mean |r|	0.9662	0.9191	0.9560	0.9339
	SD|r|	0.0466	0.0590	0.0499	0.0508
	Mean MSE	0.00074	0.00098	0.00078	0.00082
	SD MSE	0.00018	0.00077	0.00027	0.00033

Example 1, n=200

		SSIR- AL	SSIR-L	SIR-L	RSIR
$\hat{\beta}_1$	*Ave 0's*	13.67	10.84	13.24	12.77
	Mean /r/	0.9788	0.9411	0.9754	0.9668
	SD/r/	0.0398	0.0488	0.0411	0.0424
	Mean MSE	0.00044	0.00070	0.00048	0.00052
	SD MSE	0.00015	0.00058	0.00020	0.00042
$\hat{\beta}_2$	*Ave 0's*	13.43	9.48	12.52	11.88
	Mean /r/	0.9692	0.9222	0.9600	0.9543
	SD/r/	0.0412	0.0576	0.0473	0.0481
	Mean MSE	0.00048	0.00077	0.00052	0.00049
	SD MSE	0.00056	0.00082	0.00064	0.00066

According to the value of Ave0's, average and standard deviation of |r| between the $X^T\hat{\beta}_i$ and $X^T\beta_i$ and the SD and mean of MSE. From Table3.1 with n= 50,100 and 200, it has been observed that SSIR-AL shows an efficient performance compared to the other approaches for every considered case. where the accuracy of the results of prediction can be summarized depending on the mean and SD for the MSE, where the small values of these two criteria indicate better parameter estimation. The variables selection depending on (Ave0's) criterion. As for the (Ave0's) criterion, the method its Ave0's value is close to the true number of zeros parameters, which is assumed by the researcher is the best. In the *Mean /r/* criterion, the high criterion value indicates better the performance, while the low values of SD (| r |) indicates better performance. Also, it has been noticed that the value of the Ave0's and *Mean* |r| increases with the increase in the sample size, according to example 1.

Table 3.2: Results of simulation in the case of independent and correlated predictors depending on example 2.

n= 50

		SSIR-AL	SSIR-L	SIR-L	RSIR
Independent Predictors	*Ave 0's*	35.22	20.28	33.92	32.46
	Mean /r/	0.9788	0.9484	0.9742	0.9718
	SD /r/	0.0081	0.0121	0.0090	0.0097
	Mean MSE	0.0178	0.0224	0.0185	0.0194
	SD MSE	0.0467	0.0533	0.0484	0.0499
Correlated Predictors	*Ave 0's*	32.00	19.31	30.46	30.11
	Mean /r/	0.9655	0.9111	0.9642	0.9635
	SD /r/	0.0346	0.1023	0.0398	0.0419
	Mean MSE	0.0122	0.0146	0.0128	0.0129
	SD MSE	0.0513	0.0776	0.0579	0.0580

Example 2, n = 100

		SSIR-AL	SSIR-L	SIR-L	RSIR
Independent Predictors	*Ave 0's*	35.78	34.54	35.36	35.30
	Mean /r/	0.9824	0.9776	0.9810	0.9812
	SD/r/	0.0043	0.0077	0.0049	0.0054
	Mean MSE	0.0052	0.0090	0.0058	0.0061
	SD MSE	0.0064	0.0112	0.0069	0.0073
Correlated Predictors	*Ave 0's*	35.25	32.50	34.50	34.50
	Mean /r/	0.9818	0.9701	0.9811	0.9810
	SD/r/	0.0044	0.0082	0.0052	0.0051
	Mean MSE	0.0057	0.0090	0.0060	0.0063
	SD MSE	0.0069	0.0120	0.0070	0.0071

Example 2, n = 200

		SSIR-AL	SSIR-L	SIR-L	RSIR
Independent Predictors	*Ave 0's*	35.98	34.88	35.54	34.94
	Mean |r|	0.9902	0.9792	0.9898	0.9898
	SD|r|	0.0036	0.0072	0.0039	0.0040
	Mean MSE	0.0044	0.0068	0.0049	0.0047
	SD MSE	0.0060	0.0111	0.0068	0.0069
Correlated Predictors	*Ave 0's*	35.75	33.24	35.15	35.09
	Mean |r|	0.9833	0.9514	0.9821	0.9817
	SD|r|	0.0036	0.0082	0.0041	0.0040
	Mean MSE	0.0049	0.0076	0.0054	0.0055
	SD MSE	0.0063	0.0100	0.0069	0.0071

From Tables 3.2 with n= 50,100 and 200, it is clear that SSIR-AL show a better performance compared to other approaches for all considered case. According to the Ave0's, average and the SD of |r| between $X^T\hat{\beta}_i$ and $X^T\beta_i$ and the SD and mean of MSE. Also, we notice that the value of the Ave0's and *Mean* |r| increases as the sample size increases according to example 2. We find the results of the (SIR-L) method are closest to the results of our suggested method.

Table 3.3: Results of simulation depending on example 3.

Case (1), n=100

		SSIR-AL	SSIR-L	SIR-L	RSIR
$\hat{\beta}_1$	*Ave 0's*	19.37	16.22	18.52	18.25
	Mean |r|	0.9882	0.9717	0.9862	0.9869
	SD|r|	0.0023	0.0072	0.0040	0.0057
	Mean MSE	0.0052	0.0069	0.0060	0.0062
	SD MSE	0.0084	0.0112	0.0092	0.0095
$\hat{\beta}_2$	*Ave 0's*	19.20	16.50	18.76	18.75
	Mean |r|	0.9569	0.8515	0.9399	0.8910
	SD|r|	0.0600	0.1145	0.0933	0.1266
	Mean MSE	0.0054	0.0091	0.0064	0.0085
	SD MSE	0.0069	0.0115	0.0088	0.0092

Example 3 Case (2), n=100

		SSIR-AL	SSIR-L	SIR-L	RSIR
	Ave 0's	18.17	15.26	17.63	16.50
	Mean /r/	0.9925	0.9775	0.9900	0.9898
$\hat{\beta}_1$	*SD/r/*	0.0018	0.0078	0.0030	0.0058
	Mean MSE	0.00072	0.0215	0.0118	0.0137
	SD MSE	0.0126	0.0322	0.0154	0.0177
	Ave 0's	17.75	7.80	16.92	6.00
	Mean /r/	0.8610	0.7722	0.8412	0.7335
$\hat{\beta}_2$	*SD/r/*	0.1212	0.2411	0.1242	0.1478
	Mean MSE	0.0098	0.0213	0.0111	0.0157
	SD MSE	0.0170	0.0427	0.0200	0.0229

Example 3 Case (3) , n=100

		SSIR-AL	SSIR-L	SIR-L	RSIR
	Ave 0's	8.27	6.55	7.80	7.11
	Mean /r/	0.9470	0.9119	0.9448	0.9100
$\hat{\beta}_1$	*SD/r/*	0.0372	0.0550	0.0388	0.0392
	Mean MSE	0.0181	0.0202	0.0192	0.0199
	SD MSE	0.0111	0.0234	0.0145	0.0152
	Ave 0's	8.05	5.94	7.25	6.83
	Mean /r/	0.9500	0.9244	0.9492	0.9378
$\hat{\beta}_2$	*SD/r/*	0.0344	0.0540	0.0358	0.0364
	Mean MSE	0.0246	0.0281	0.0248	0.0256
	SD MSE	0.0342	0.0388	0.0370	0.0384

From Tables 3.3 with n= 100, in this example 3, the sample size is kept constant while the number of parameters is changed; in the third case, we increased the number of parameters. We notice the superiority of our suggested method (SSIR-AL) over the rest methods.

Table 3. 4: Results of simulation depending on example 4.

		SSIR-AL	SSIR-L	SIR-L	RSIR
$n = 50$	*Ave 0's*	0.00	0.00	0.00	0.00
	Mean /r/	0.9600	0.9235	0.9555	0.9415
	SD/r/	0.0190	0.0252	0.0226	0.0245
	Mean MSE	0.0497	0.0536	0.0519	0.0536
	SD MSE	0.0223	0.0495	0.288	0.0324
$n = 100$	*Ave 0's*	0.00	0.00	0.00	0.00
	Mean /r/	0.9875	0.9524	0.9778	0.9722
	SD/r/	0.0035	0.0078	0.0042	00055
	Mean MSE	0.0654	0.0679	0.0665	0.0677
	SD MSE	0.0444	0.0657	0.0484	0.0517
$n = 200$	*Ave 0's*	0.00	0.00	0.00	0.00
	Mean /r/	0.9892	0.9611	0.9820	0.9798
	SD/r/	0.0025	0.0078	0.0032	00039
	Mean MSE	0.0540	0.0682	0.0555	0.0573
	SD MSE	0.0320	0.0460	0.0334	0.0340

From tables 3.4 with n= 50, 100 and 200, all values of (β) are equal to one, and hence exclude (*Ave0's*) criterion because all parameters are given values greater than zero. From the comparison between methods in the example 4. We notice that the suggested method (SSIR-AL) has

preference over the rest of the methods by relying on the average and(SD)for (|r|), as well as on the average and (SD) for (MSE).

Table 3. 5 : Results of simulation depending on the case of independent and correlated predictors on example 5.

			SSIR-AL	SSIR-L	SIR-L	RSIR
Independent	Predictors	*Ave 0's*	193.50	183.55	190.98	185.40
		Mean /r/	0.9978	0.9714	0.9975	0.9744
		SD/r/	0.0019	0.0081	0.0022	0.0078
		Mean MSE	0.0023	0.0029	0.0024	0.0025
		SD MSE	0.0015	0.0019	0.0017	0.0017
Correlated	Predictors	*Ave 0's*	191.88	175.21	189.37	187.20
		Mean /r/	0.9850	0.9665	0.9825	0.9815
		SD/r/	0.0014	0.0044	0.00022	0.0029
		Mean MSE	0.0024	0.0042	0.0031	0.0039
		SD MSE	0.0030	0.0057	0.0044	0.0051

In general, the results of tables 3.1, 3.2, 3.3, 3.4, and 3.5 can be summarized according to the criteria that have been mentioned earlier. First, SSIR-L clearly performs less performance than the rest of the methods. At the same time, the suggested method (SSIR-AL) has higher accuracy compared to all the considered method due to its good results according to the required statistical criteria on which the comparison was made. In general, the SIR-L was a competitor to the SSIR-AL in nearly all examples.

Figure (3.1) shows, in its three sections, the preference of our suggested method over the remaining methods. where the first section represented the superiority of (SSIR-AL) according to the (*Ave0's*) criterion, while the second section showed the preference of (SSIR-AL) as it produced the lowest values of the two criteria (Mean MSE) and (SD MSE) from the rest of the methods, and in the third section it shows that the (SSIR-AL) method has the highest values of (Mean |r|) on the other comparative methods and to both estimators (β_1 and β_2) according to example 1 data.

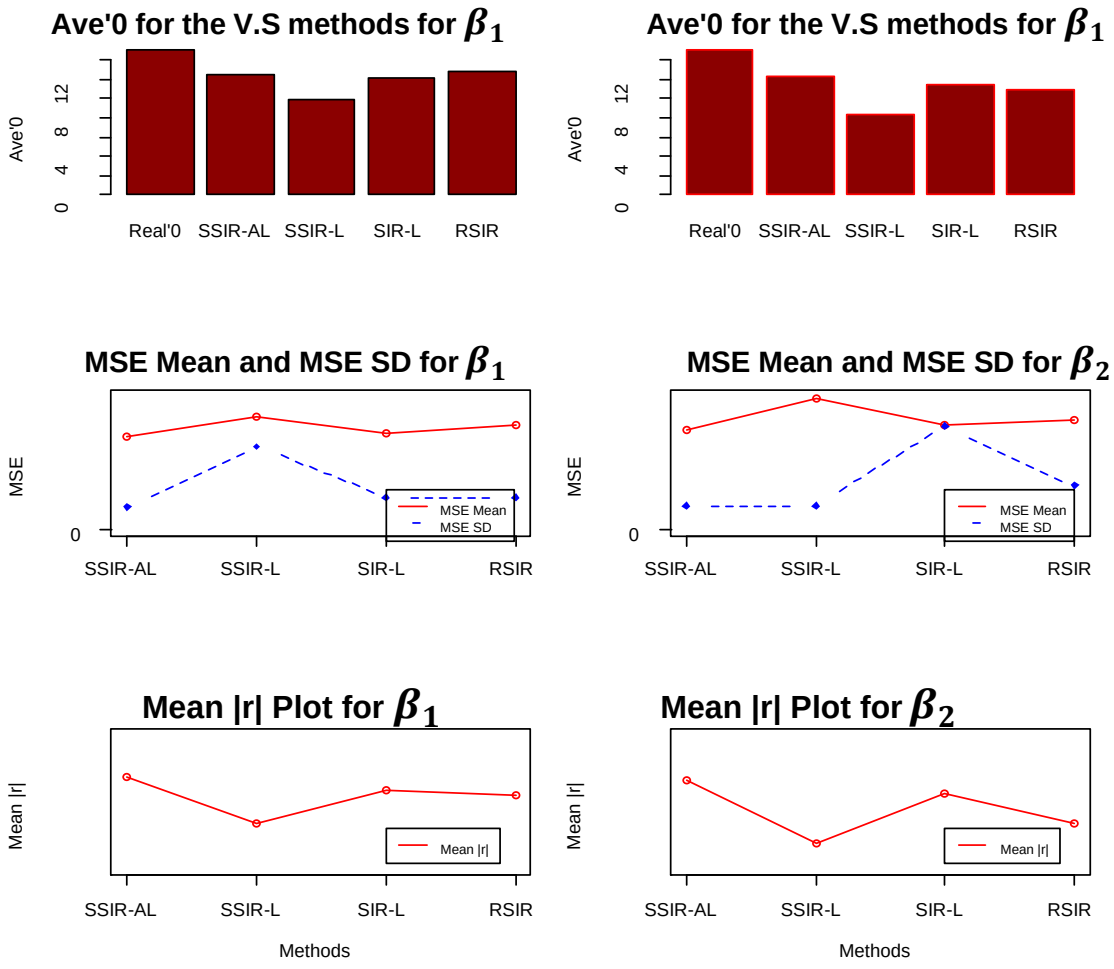

Figure: 3.1 the Ave'0s for V.S methods and Mean MSE , SD MSE and Mean |r| for β_1 and β_2 of the methods studied based on example 1 with n=100.

Figure (3.2) shows the superiority of our suggested method (SSIR-AL) over the rest of the methods we compared according to all the criteria (Ave'0s for V.S methods, Mean MSE , SD MSE and Mean |r| for β_1 and β_2).That were used in the comparison according to the data of example2.

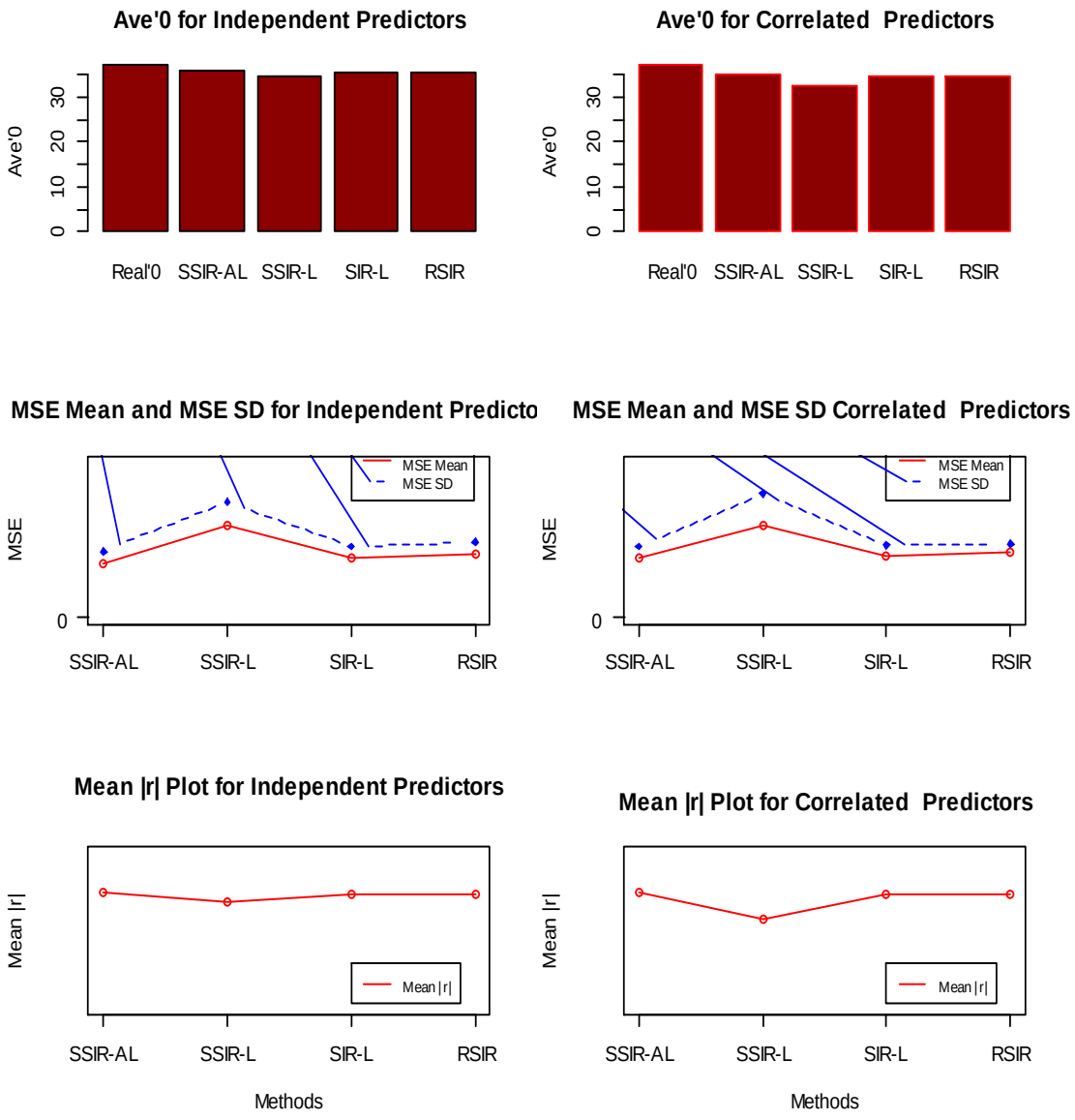

Figure: 3.2 the Ave'0s for V.S methods and Mean MSE , SD MSE and Mean |r| for β_1 and β_2 ,in the case of independent and correlated predictors of the methods studied based on example 2 with n= 100.

Figure (3.3) shows the superiority of our suggested method (SSIR-AL) over the rest of the methods we compared according to all the criteria, that were used in the comparison according to the data of example 3 with case 1 n= 100.

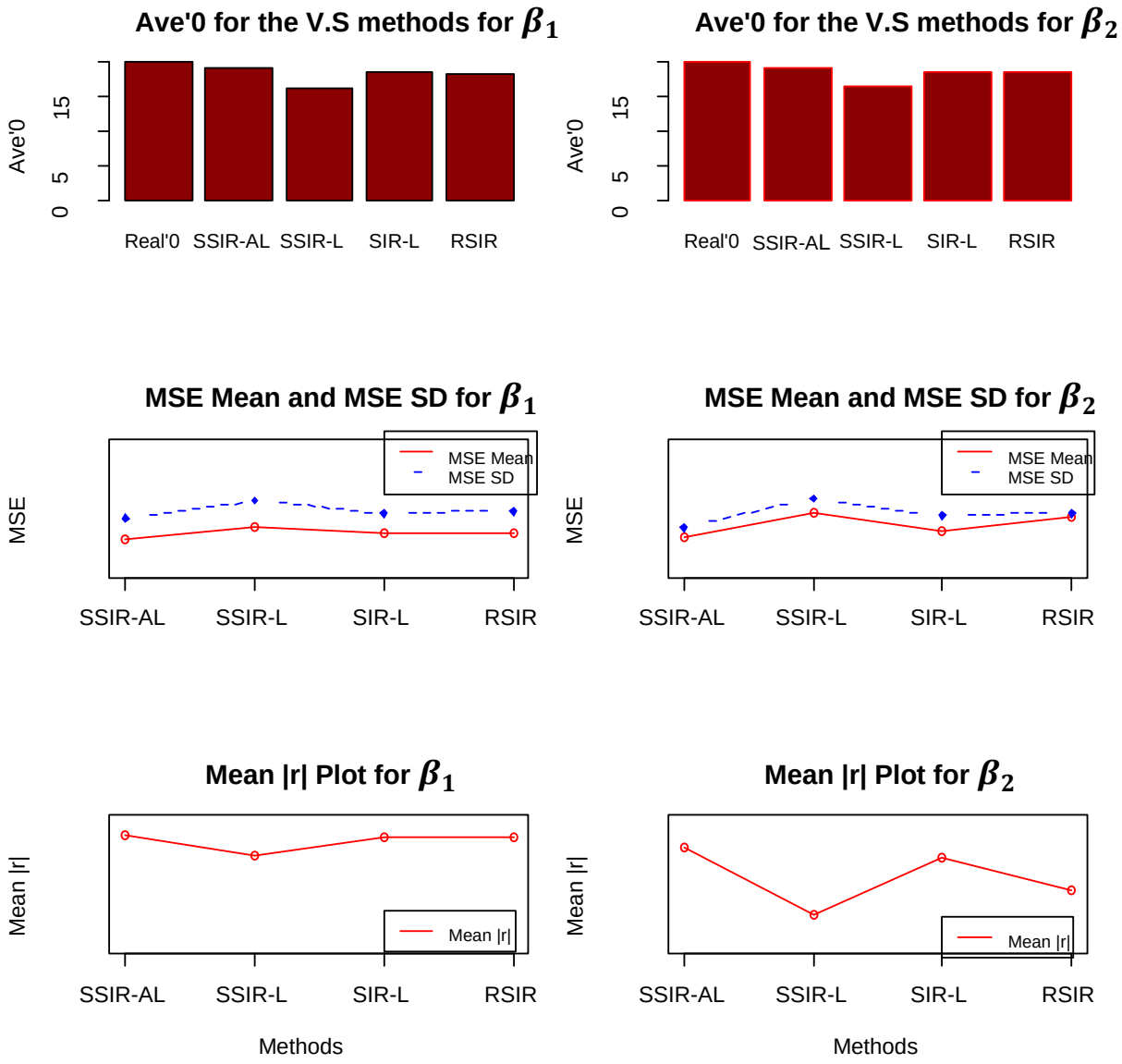

Figure: 3.3 the Ave'0s for V.S methods and Mean MSE , SD MSE and Mean |r| for β_1 and β_2 of the methods studied based on example 3 with case 1 n= 100.

Figure (3.4) shows the superiority of our suggested method (SSIR-AL) over the rest of the methods we compared according to all the criteria (Ave'0s for V.S methods , Mean MSE , SD MSE and Mean |r| for β_1 and β_2). That were used in the comparison according to the data of example 3 with case 2 n= 100.

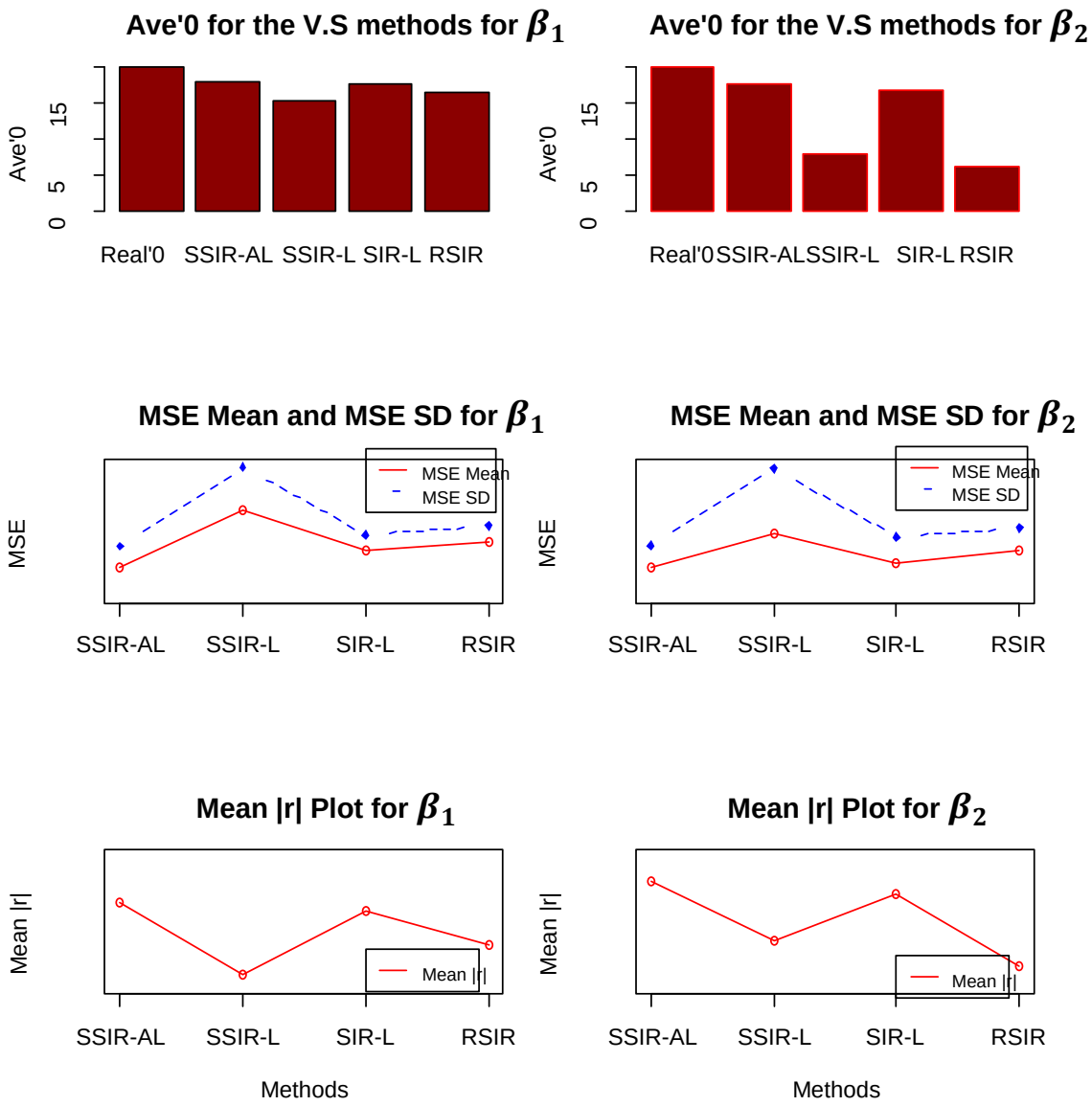

Figure: 3.4 the Ave'0s for V.S methods and Mean MSE , SD MSE and Mean |r| for β_1 and β_2 of the methods studied based on example 3 with case 2, n= 100.

Figures (3.3), (3.4) and (3.5) show the superiority of our suggested method (SSIR-AL) over the rest of the methods we compared according to the criteria that were used in the comparison according to the data of example 3 in all the three cases.

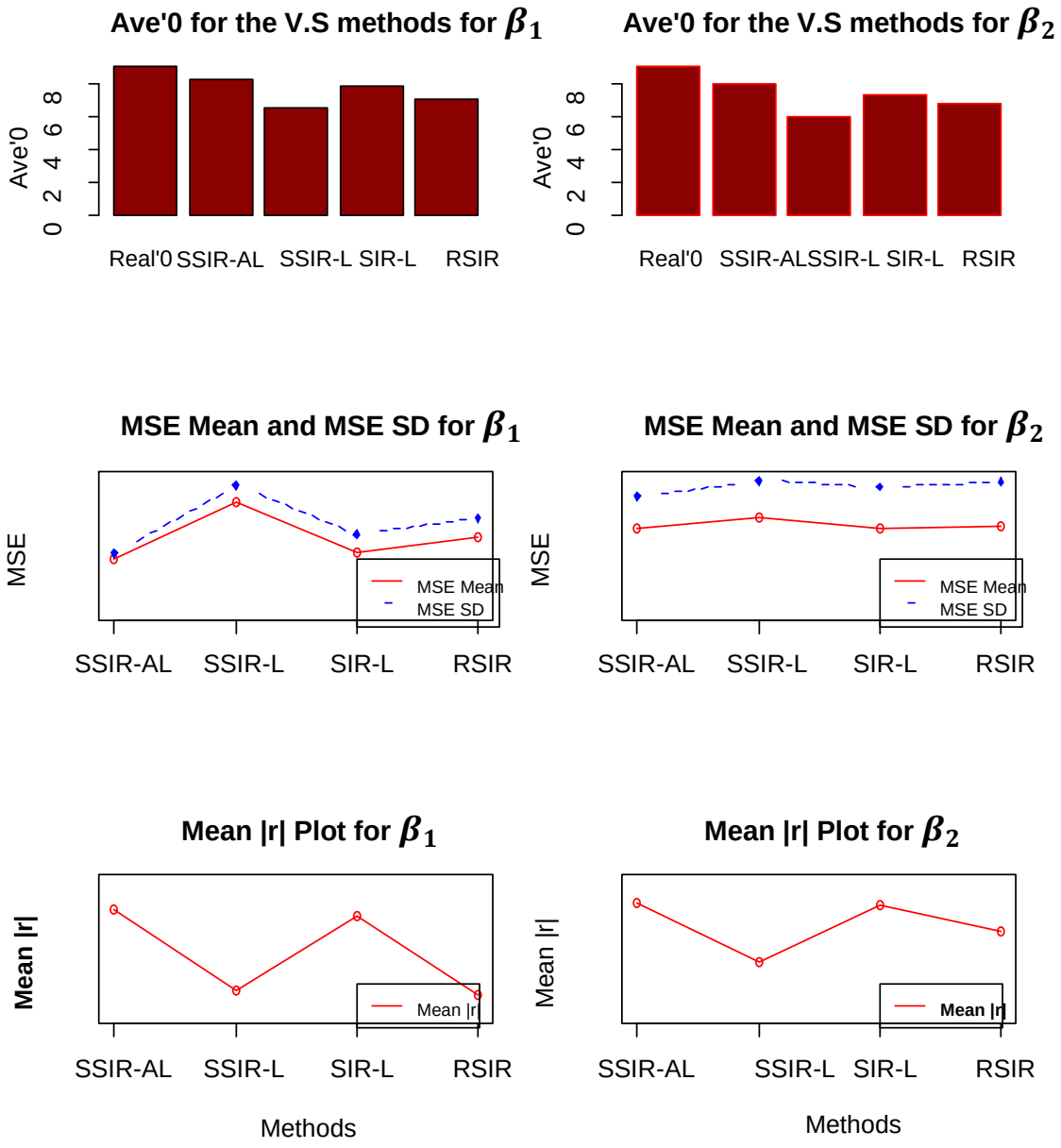

Figure: 3.5 the Ave'0s for V.S methods and Mean MSE , SD MSE and Mean |r| for β_1 and β_2 of the methods studied based on example 3 with case 3, n= 100.

Figure (3.6) according to the data of example 4 with n = 50 and 100 show the superiority of our suggested method (SSIR-AL) over the rest of the methods we compared according to the criteria that were used in the comparison, and in this example criterion (Ave'0s) was excluded because all parameters are given values greater than zero.

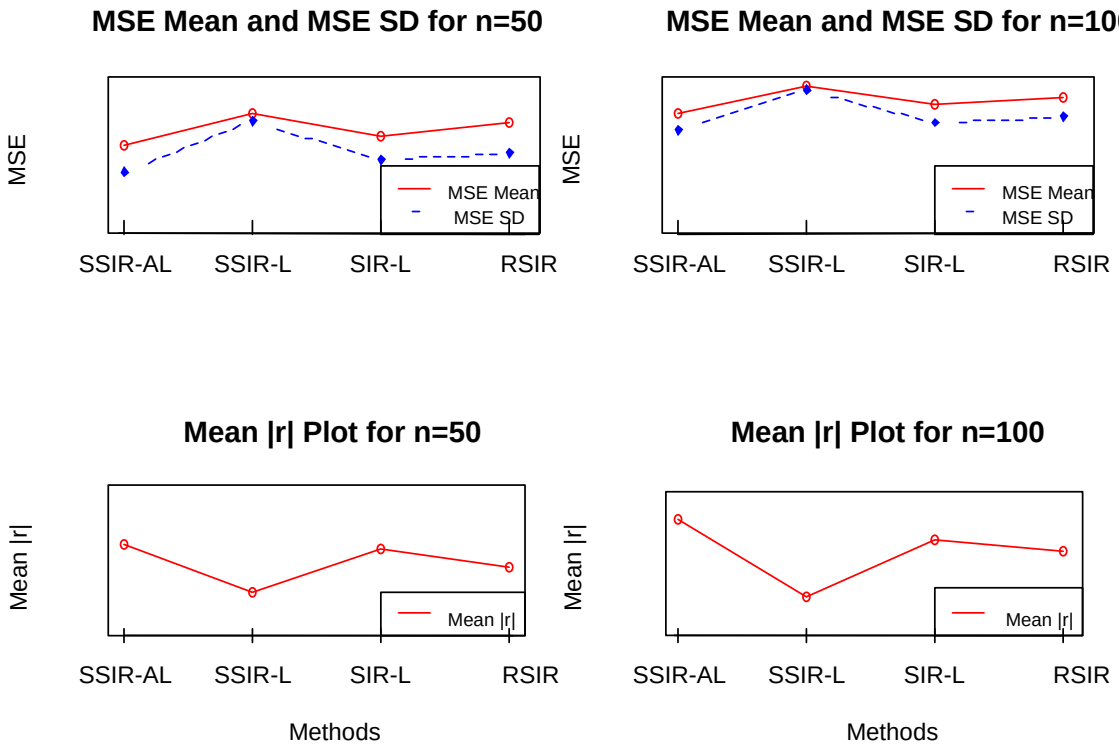

Figure: 3.6 the Mean MSE , SD MSE and Mean |r| for β_1 of the methods studied based on example 4 with n= 50 and 100.

Figure (3.7) shows the superiority of our suggested method (SSIR-AL)over the rest of the methods we compared according to all the criteria that were used in the comparison according to the data of example 5.

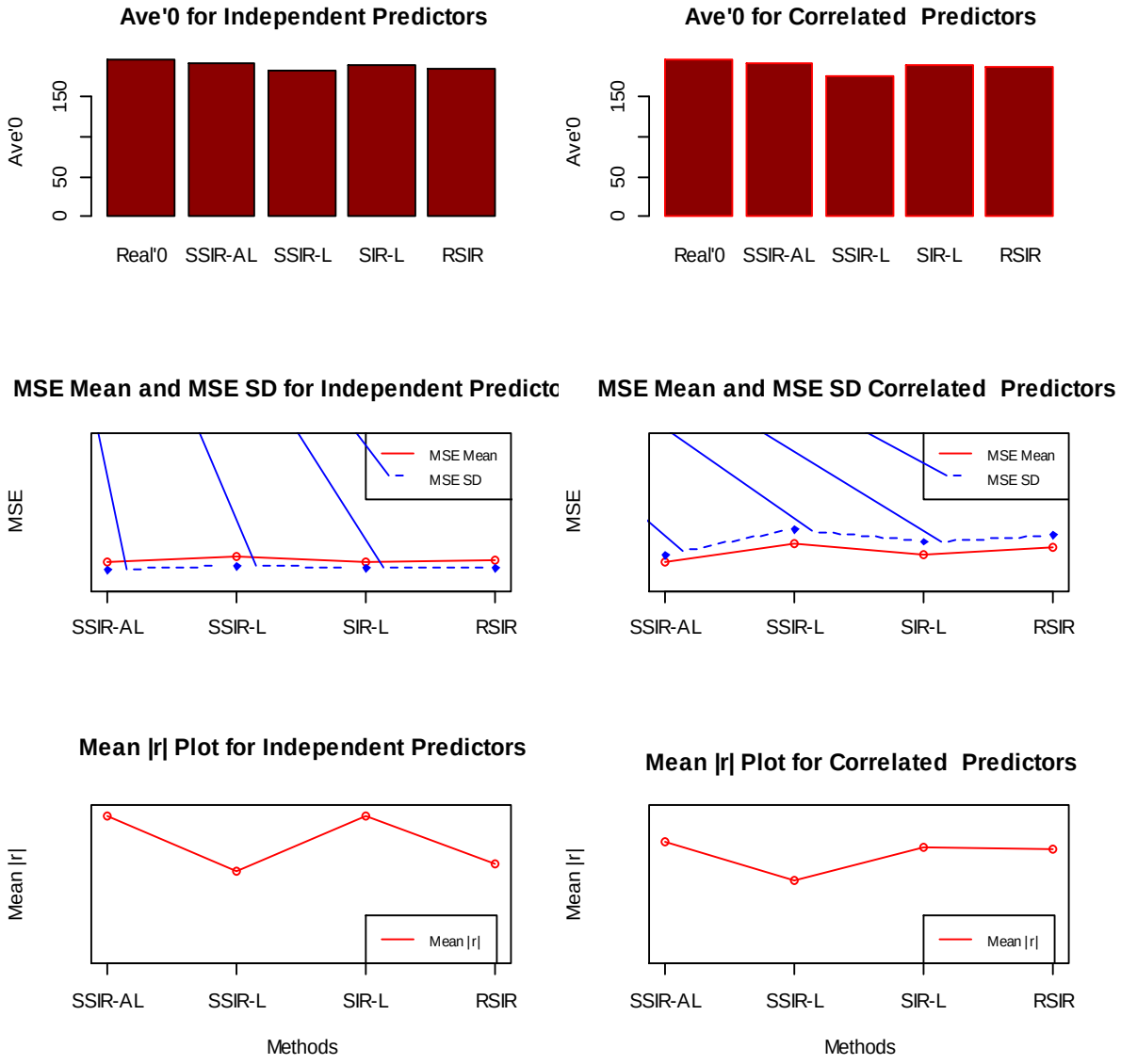

Figure: 3.7 the Ave'0s for V.S methods and Mean MSE , SD MSE and Mean |r| for β_1 and β_2 ,in the case of independent and correlated predictors of the methods studied based on example 5 with n= 300.

3.3. Real Data:

In this section, to check the performance of SSIR-AL, we use real data on diabetes. It is one of the diseases that humanity suffers in general. recently, it has been noticed the increasing of the number of people whom are suffering from this disease in Iraq. We examined a sample of 82 pations who visited the diabetes consultant at Al Zahra Teaching Hospital in Al Kut, in Wasit Governorate, during March 2019. This was done in cooperation with researcher Ahmed Mohamed Oreibi, MSc Microbiology / Clinical Immunology - College of Medicine, University of Wasit. Diabetes can be defined as follows: Even though diabetes is non-infectious, yet it is chronic and can infect most people of different ages. Its causes multiply to several factors, including organic (i.e., the result of a birth defect in the affected person), and genetic factors contribute greatly to the possibility of infection and other factors. Diabetes causes a full or relative shortage of blood insulin, which is released naturally from the pancreas. It makes cells resist the insulin or when the body is unable to utilize the insulin produced properly (insulin is a hormone that transports glucose to carry out metabolism processes that are vital to the energy that the body's systems use to function), and insulin is the hormone that regulates the level of glucose in the blood. The decrease in it, or decreased sensitivity to its work, or both, leads to increased blood sugar levels. This leads to interruptions or impairments in the performance of a few body systems and the occurrence of serious complications with the length of injury time that varies from person to person, such as cardiovascular disease (neuropathy), high blood pressure, high blood lipids, problems of the diabetic foot and other health problems. In addition to delayed healing of wounds, high or low blood sugar also causes urgent symptoms such as frequent urination, fatigue,

severe thirst, dry mouth, severe hunger and tremors, disturbance of vision. We analysis the data according to a multiple regression model were considered the response variable y_i each person's sugar reading; we considered the independent variable X_i it is the result of laboratory tests each person performed according to the analysis requested by the specialist. Data have been analysis with the use of the R-code that has been written by the author. A comparison of the suggested approach to with the existing method has been performed using the real data. After analysis the data, we got the results mentioned in Tables 3.6 and 3.7.

y_i	Reading the degree of sugar
x_1	Age of a person
x_2	Father's age
x_3	Mother's age
x_4	The number of brothers
x_5	Number of sisters
x_6	Body mass index (BMI)
x_7	(High blood pressure) Measuring the person's pressure level
x_8	(Duration) The period of injury of a person measured in months
x_9	(HbA1c%) glycated hemoglobin
x_{10}	(B.urea) The ratio of urea
x_{11}	(S.crea) Creatine ratio
x_{12}	(WBC x 10^3 / µL) white blood cell count
x_{13}	(Neutro x 10^3 / µL) the effective percentage of white blood cells
x_{14}	(Hemoglobin) blood ratio
x_{15}	(PCV%) filled cell volume (blood viscosity)
x_{16}	(PLT x 10^3µL) Platelets
x_{17}	(C3 mg / dl) the proteins that are portion of immune the system.
x_{18}	(C4 mg / dL) the proteins that are portion of immune the system.

We'll explain all the variables in a simple way as follows:

Reading the degree of sugar (y_i):

We read the sugar level by the use of a blood Glucose meter. Then the person is directed to the laboratory to conduct the necessary tests, which the doctor thinks can be actual reasons for reading levels sugar.

Age of the person(x_1):

We ask the person about and calculate his age in years. We also ask him the following direct questions (x_2: Father's age, x_3: Mother's age, x_4: The number of brothers, x_5: Number of sisters).

Body mass index (BMI) (x_6):

BMI can be defined as a value which comes from the person's mass (i.e., weight) and height, that is, the body mass (in kg) is to be divided by the square value of height (in m), and it is represented in kg/m2. The common classification of the range of the BMI values are: (a) underweight (< 18.50kg/m2), (b) normal (between 18.50 and 25), (c) overweight (between 25 and 30), (d) obese (> 30).

Hypertension (x_7):

Commonly, known as high blood pressure (HBP), which is a long-dated medical case. This disease is related to the rise of blood pressure in the arteries. Most commonly, (HBP) does not lead to immediate symptoms, but in the long-term high, blood pressure can be a major cause of many diseases that can be dangerous to human life.

Glycated hemoglobin (**HbA1c**) (**x_8**)**:**

Which can be defined as the form of hemoglobin that is chemically bound to sugar. The majority of monosaccharides, which include galactose, fructose and glucose, are linked to hemoglobin, if present in the human bloodstream, which shows elevated sugar levels in the bloodstream, indicates diabetes. A1C is the center of attention since it can be easily detected**.**

platelet count (**PLT**) (x_9)**:**

PLT is one of the analysis that are performed within a complete CBC blood count to determine the condition of the three blood elements and the titer and ratio of each one of them. The platelets are small cells that contribute mainly to the coagulation process; that is, they stack on top of each other to form the thrombus that stops the Hemorrhagic blood, and many conditions can affect platelet counts, which is reflected in the values and the results of the PLT analysis.

Urea (x_{10}**):**

This is a nitrogen-containing material filtered from blood by kidneys to urine. Disease cases which influence kidney function usually result in high blood urea.

Creatinine (x_{11}**):**

This is a chemical waste that is produced by the muscle metabolism and in small amounts by eating meats. Creatinine is filtered from the blood by a healthy kidney and goes out of the body through urine. In case of kidney failure or if the kidney doesn’t function properly, the creatinine may accumulate in the blood.

White blood cells (WBC) (x_{12}):

This is a cell of blood made by the marrow bones, and it can be found in the tissues of the lymph and the blood. The WBCs are helpful in fighting and resist infections and disease. Thus, White blood cells count as one of the main parts of the immune system.

Neutrophil (x_{13}):

A WBC type, a granulocyte which is filled by microscopic granules, which are small sacs that contain the enzymes which can perform the digestion of the micro-organisms, which have been referred to as polymorph nuclear leukocyte or poly as well.

Hb (hemoglobin) (x_{14}):

This can be defined as iron-containing respiratory pigmentation of the vertebrate red blood cells, which transport the oxygen to tissues following the conversion to the oxygenated form in the lungs or gills and also transport CO2 back to gills and lungs following the delivery of O2.

Duration (x_{15}):

Duration of a person's injury: We ask the person about the period he was found to have diabetes, and the answer is often approximate. We measured it in months in order to ensure the accuracy of the answer.

The packed cell volume (PCV) (x_{16}):

It represents the volume of the ratio of the blood that produces the cells. The ratio is expressed as a percentage or fraction of cells in the blood. when the number of red blood cells rises up or in the case where the total blood volume is decreased as in dehydration, the PVC rises. Decreasing the percentage of PVC less than normal is an indication of the fact that the patient might have anemia.

Complement (C3 and C4) (x_{17} , x_{18}):

This is an important part of the immune system; it consists of proteins, which are measured by a simple blood test that mensuration the grade of C3, C4 in the blood.

Table 3. 6: The adjusted R-square values for the model fit based on the real data

		SSIR-AL	SSIR-L	SIR-L	RSIR
Model Fit	*Linear*	0.94	0.81	0.94	0.82
	Quadratic	0.95	0.90	0.95	0.91
	Cubic	0.95	0.92	0.95	0.92
	Quartic	0.95	0.92	0.95	0.92

Table 3. 6: indicates the adjusted R-squared values that are related to model fit on the basis of using diabetes data. The examined approaches have specified certain non-linear structure that might be approximated through a cubic fit. It has been indicated that the changed value of the adjusted R-square for SSIR-AL approaches has been big in comparison

to the values related to adjusted R-squared for SSIR-L, RSIR .in the linear case. As for the rest of the cases, the difference in the adjusted R-squared degree decreases but remains high for the suggested method SSIR-AL. At the same time, we notice that the degree of adjusted R-squared of SSIR-AL is equal to the method of SIR-L in all cases. Furthermore, adjusted R-squared values for SSIR-L and RSIR have been comparable.

Figure (3.8) shows the value of the criterion (adjusted R-squared), where we notice that its value in the suggested method (SSIR-AL) is greater than the two methods (SSIR-L, RSIR), and is equal to the values of (adjusted R-squared) with the method (SIR-L). This demonstrates the preference for our suggested method (SSIR-AL) using the real data as well.

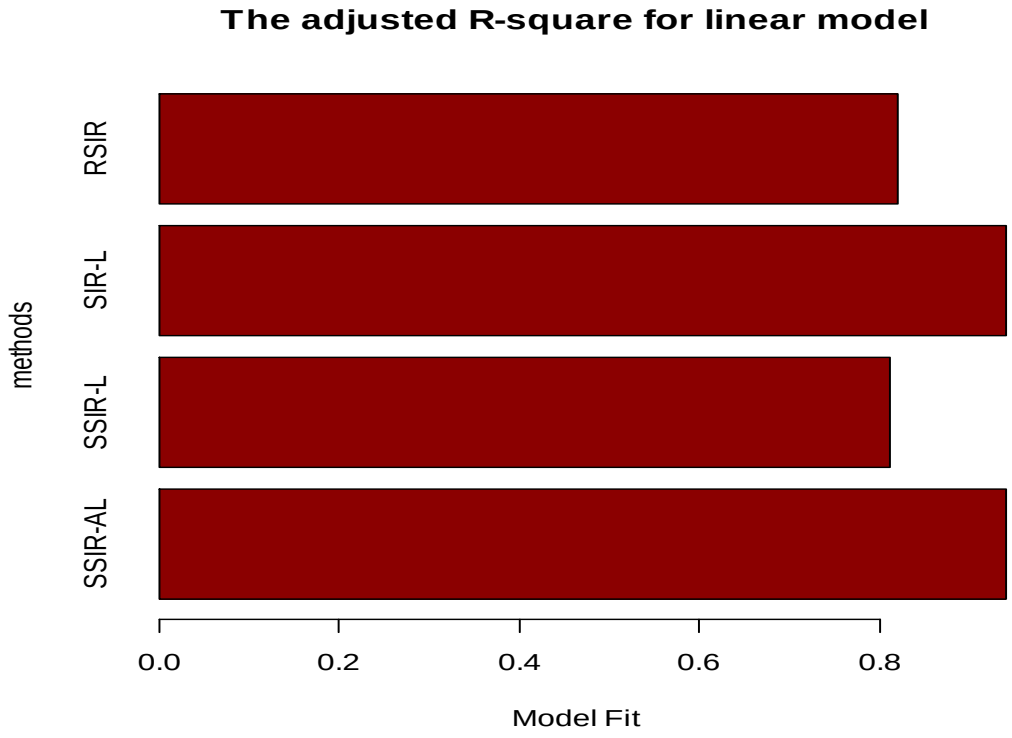

Figure 3. 8: The adjusted R-square for linear model for the considered methods in real data.

Table 3.7: The prediction error of the cubic fit for the studied methods based on the real data.

Methods	Prediction error
SSIR-AL	0.6612
SSIR-L	0.8437
SIR-L	0.6910
RSIR	0.8188

From table 3.7, it has been evident that the SSIR-AL approach had a lower prediction error compared to SSIR-L, SIR-L, and RSIR methods. This means that the suggested SSIR-AL had better performance than the SSIR-L, SIR-L, and RSIR methods.

Figure 3. 9: It shows that our suggested method (SSIR-AL) has the least prediction error for variables selection appropriate from the rest of the studied approaches, depending on the real data. The results of the method (SIR-L) compete with the results of the suggested method (SSIR-AL).

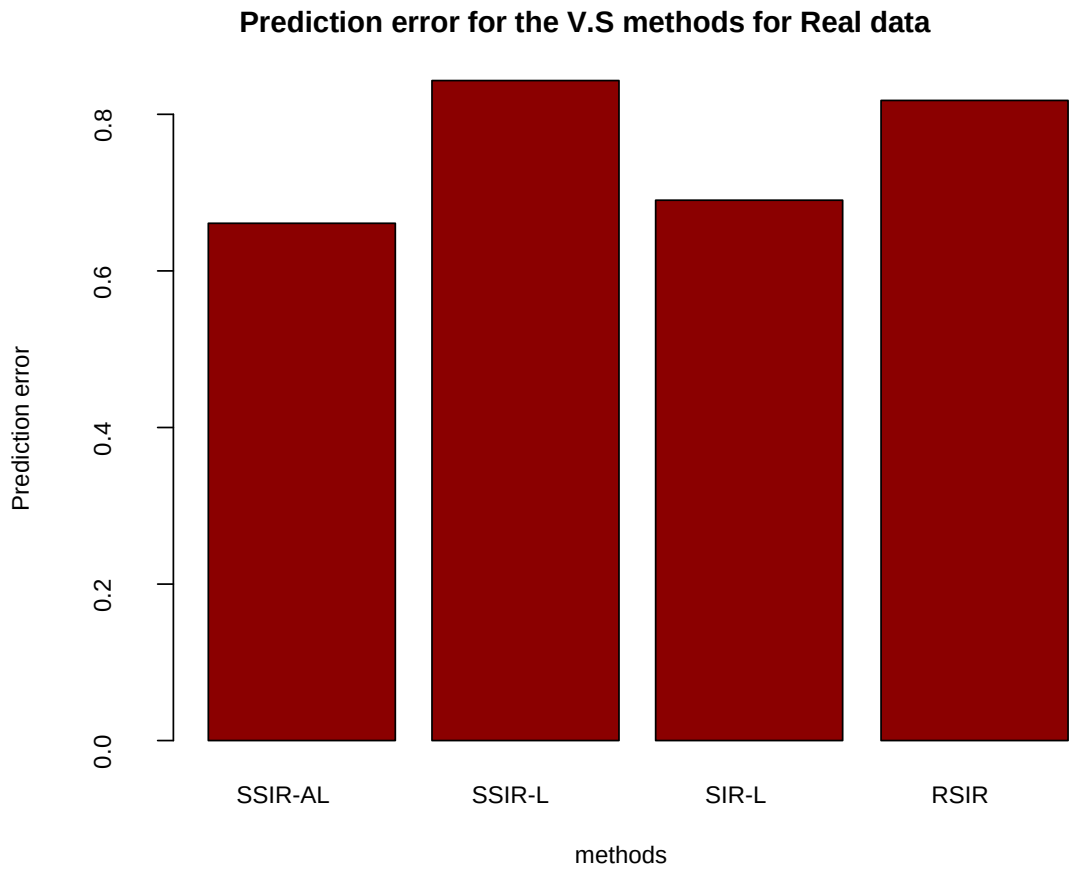

Figure 3. 9: The prediction error of the cubic fit for the studied methods based on the real data.

Chapter Four

Conclusions, Recommendation and Future works and References

4.1. Conclusions:

In this thesis, SSIR-AL is direct incorporate adaptive Lasso into the SIR method. Since SIR can exhaustively estimate $S_{Y|X}$ while Lasso does consistent not oracle variable selection, SSIR-AL can achieve oracle variable selection. SSIR-AL extends adaptive Lasso to the non-linear as well as the multi-dimensional regressions without needing any specific model. Based on the empirical results and simulations, the SSIR-AL has been proven computationally simple to implement and the effective. This work shows that SSIR-AL is more effective than the competitors according to the estimation and the selection accuracies in both simulation and real data.

4.2. Recommendation and Future works:

The suggested approach in this thesis can be extended to other SDR methods, such as PHD (Li, 1992) and SAVE (Cook and Weisberg, 1991). In addition to that, the SSIR-AL may be extended to the models of the binary response. In addition to that, the robust SSIR-AL can be considered as one more potential extension of the suggested approach.

References

1. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267-281.
2. Alkenani, A. and Dikheel, T.R. (2016). Sparse sliced inverse quantile regression. Journal of Mathematics and Statistics. Volume 12, Issue 3.
3. Alkenani, A. and Yu, K. (2013). Sparse MAVE with oracle penalties. Advances and Applications in Statistics 34,pp. 85–105.
4. Bellman, R. E. (1961). Adaptive Control Processes. Princeton University Press, Princeton, New Jersey.
5. Bondell, H. D., & Reich, B. J. (2008). Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. Biometrics, 64(1),pp. 115-123.
6. Branco, J. A., Croux, C., Filzmoser, P. and Olivera, M. R. (2005). Robust canonical correlations: a comparative study. Computational Statistics 20,pp. 203–229.
7. Breiman, L. (1996). Heuristics of instability and stabilization in model selection. The Annals of Statistics 24,pp. 2350-2383.
8. Carlos, A. M. and Sergioc, C. S.(2012). Does BIC Estimate and Forecast Better than AIC?. Available at (https://mpra.ub.uni-muenchen. de/42235/).
9. Chand, S., & Kamal, S. (2011). Variable selection by lasso-type methods. Pakistan Journal of Statistics and Operation Research.
10. Common, P. (1984). Independent Component Analysis, a new concept?. Signal Processing 36,pp. 287–314.

11. Cook, R. D. (2009). Regression graphics: Ideas for studying regressions through graphics (Vol. 482). John Wiley & Sons.

12. Cook, R. (2004). Testing predictor contributions in sufficient dimension reduction. Annals of Statistics 32,pp.1061-92.

13. Cook, R. D. and Weisberg, S. (1991). Discussion of Li (1991). Journal of the American Statistical Association 86,pp. 328–332.

14. Delecroix, M., Härdle, W. and Hristache, M. (2003). Efficient estimation in conditional single-index regression. Journal of Multivariate Analysis 86,pp. 213–226.

15. Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage. Biometrika 81,pp. 425-455.

16. Efroymson, M. A. (1960). Multiple regression analysis. Mathematical methods for digital computers, 191-203.

17. Fan, J. and Li, R. Z. (2001). Variable selection via non-concave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96,pp.1348–1360.

18. Fung, W. K., He, X., Liu, L. and Shi, P. (2002). Dimension reduction based on canonical correlation. Statistic a Sinica 12,pp. 1093–1113.

19. Gorsuch, R. L. (1983). Factor analysis. Hillsdale, New Jersey: Lawrence Erlbaum. Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research 3,pp. 1157–1182.

20. Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research 3,pp. 1157–1182.

21. Härdle, W. and Stoker, T. (1989). Investing smooth multiple regression by the method of average derivatives. Journal of the American Statistical Association 84,pp. 986–995.
22. Hotelling, H. (1936). Relations between two sets of variates. Biometrika 28,pp. 321–377.
23. Ichimura, H. (1993). Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models. Journal of Econometrics 58,pp. 71–120.
24. Jabbar, E.(2020). A non-linear multi-dimensional estimation and variable selection via regularized MAVE method. Thesis submitted to college of administration and economics. University of Al-Qadisiyah. Iraq.
25. Jolliffe, I. T. (2002). Principal component analysis. 2nd ed. Berlin, Germany: Springer Verlag.
26. Li, K. (1991). Sliced inverse regression for dimension reduction (with discussion). Journal of the American Statistical Association 86,pp. 316–342.
27. Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein‟s lemma. Journal of the American Statistical Association 87,pp. 1025–1039.
28. Li, L. (2007). Sparse sufficient dimension reduction. Biometrika 94,pp. 603–613.
29. Li, L. and Nachtsheim, C. J. (2006). Sparse sliced inverse regression. Technometrics 48,pp. 503–510
30. Li, L. and Yin, X. (2008). Sliced Inverse Regression with regularizations. Biometrics 64,pp. 124–131.

31. Lin, Q., Zhao, Z., & Liu, J. S. (2018). On consistency and sparsity for sliced inverse regression in high dimensions. The Annals of Statistics, 46(2), 580-610.

32. Malik, D. (2019). Sparse dimension reduction through penalized quantile MAVE with application. Thesis submitted to college of administration and economics. University of Al-Qadisiyah. Iraq.

33. Naik, P. A., and Tsai, C. L. (2001). Single- index model selections. Biometrika, 88(3), 821-832.

34. Ni, L., Cook, R. D. and Tsai, C. L. (2005). A note on shrinkage sliced inverse regression. Biometrika 92(1), 242–247.

35. Powell, J. L., Stock, J. H., & Stoker, T. M. (1989). Semiparametric estimation of index coefficients. Econometrica: Journal of the Econometric Society, 1403-1430.

36. Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics 6, 461-464.

37. Sharma, D. B. ,Bondell, H. D. and Zhang H. H.(2013).consistent group indentification and variable selection in regression with correlated predictors. Journal of compataional and Graphical Statistics 22, 319-340

38. Sugiura, N. (1978). Further analysts of the data by akaike's information criterion and the finite corrections: Further analysts of the data by akaike's. Communications in Statistics-Theory and Methods, 7(1),pp. 13-26.

39. Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B 58, 267–288.

40. Wang, Q. and Yin, X. (2008). A Nonlinear Multi-Dimensional Variable Selection Method for High Dimensional Data: Sparse MAVE. Computational Statistics and Data Analysis 52, 4512–4520.

41. Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. The Annals of Statistics 35, 2654–2690 .

42. Xia, Y. (2008). A multiple-index model and dimension reduction. Journal of the American Statistical Association 103, 1631–1640.

43. Xia, Y., Tong, H., Li, W. and Zhu, L. (2002). An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society, Series B 64, 363–410.

44. Yin, X. and Cook, R. D. (2005). Direction estimation in single-index regressions. Biometrika 92, 371–384.

45. Yin, X., Li, B. and Cook, R. D. (2008). Successive direction extraction for estimating the central subspace in a multiple-index regression. Journal of Multivariate Analysis 99, 1733–1757.

46. Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49-67.

47. Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. Annals of Statistics 38, 894–942.

48. Zhang, J. (2011). Applications of a Robust Dispersion Estimator. Ph.D. Thesis, Southern Illinois University. Available at (www.math.siu.edu/olive/szhang.pdf).

49. Zhang, J. and Olive, D. J. (2009). Applications of a Robust Dispersion Estimator. Available at (www.math. siu.edu/olive/ pprcovm.pdf).

50. Zhou, J. (2009). Robust dimension reduction based on canonical correlation. Journal of Multivariate Analysis 100, 195–209
51. Zou, H. (2006). The adaptive Lasso and its oracle properties. Journal of the American Statistical Association 101, 1418–142
52. Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society, Series B 67, 301–320.

الملخص :

في العقدين الماضيين ، حظيت نظرية الحد الكافي من الأبعاد (SDR) التي اقترحها كوك 1998 باهتمام كبير (كوك ، 2009). تتمثل فكرة اختزال البعد الكافي (SDR) في استبدال X بإسقاطات متعامدة منخفضة الأبعاد $P_S X$ على الفراغات الفرعية S دون فقدان المعلومات حول توزيع $Y|X$ ودون افتراض أي نموذج محدد. الهدف من SDR هو الفضاء الجزئي المركزي ($S_{Y|X}$). تم عمل الكثير من الطرق لإيجاد ($S_{Y|X}$) وإحدى هذه الطرق هي (الانحدار العكسي الشرائح) (SIR) (Li ,1991). يتم تطبيق SIR في مجالات مختلفة مثل المعلوماتية الحيوية والتسويق والتمويل والاقتصاد.

تكمن مشكلة الدراسة في أن SIR أثبت فعاليته في أسلوب اختزال الأبعاد (DR) وهو فعال في التعامل مع البيانات عالية الأبعاد (HD) وأداة كافية للتعامل مع اختزال البعد (DR) في الانحدار الشرطي (Li and Yin, 2008). ومع ذلك ، فإنه ينتج مجموعات خطية (LCs) لجميع المتنبئين الأصليين. نتيجة لذلك ، قد يكون تفسير تقديرات SIR صعبًا ومضللًا في بعض الأحيان.

الهدف من دراستنا هو تقليل عدد المعاملات غير الصفرية في اتجاهات SIR للحصول على تفسير أفضل. من خلال دمج إحدى طرق "Regularization" مع طريقة SIR لإنتاج تقديرات متفرقة ودقيقة اي (يتم تصفير المتغيرات الغير مهمة), وتمكن الطريقة المقترحة "SSIR-AL" طريقة Adaptive Lasso من العمل مع الانحدار غير الخطي ومتعدد الأبعاد دون افتراض أي نموذج محدد.

في هذه الأطروحة ، تم اقتراح طريقة اختيار المتغيرات في مفهوم تقليل البعد الكافي ، تسمى "Sparse SIR with Adaptive Lasso penalty (SSIR-AL)" ان الاقتراح

"SSIR-AL" يجمع بين أفكار طريقة Adaptive Lasso مع انحدار العكسي الشرائح (SIR) للحصول على مقدر SIR متفرق. حيث يتم الاستفادة من طريقة Adaptive Lasso، في اختيار المتغيرات وتقدير المعلمات انيا في نفس الوقت.

يتم إثبات فعالية "SSIR-AL" من خلال كل من المحاكاة وتحليل البيانات الحقيقية. قمنا بإجراء مقارنة مع بعض الطرق التي استخدمت اختيار المتغيرات في مفهوم تقليل البعد الكافي .وهذه الطرق هي طريقة Ni واخرون (2005) مقدر الانكماش (shrinkage SIR)

(SSIR-L)من خلال دمج طريقة "Lasso penalty" مع "SIR" ومع طريقة Li and Yin (2008) (RSIR) "Regularised SIR " لتمكين SIR من العمل مع تنبؤات شديدة الارتباط وعندما يكون حجم العينة اقل من عدد المتغيرات المستقلة وكذلك مع طريقة -Lasso SIR لنموذج الفهرس المتعدد (multiple index) حيث تم اقتراحها ضمن إعدادات يكون حجم العينة اقل من عدد المتغيرات المستقلة بواسطة Lin واخرون (2018).

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة القادسية /كلية الإدارة والاقتصاد
قسم الاحصاء

اختزال الأبعاد الصفرية عن طريق الانحدار العكسي الجزائي مع التطبيق

رسالة مقدمة الى
مجلس كلية الإدارة والاقتصاد - جامعة القادسية
وهي جزء من متطلبات نيل شهادة الماجستير
في الإحصاء

من قبل
ضياء بايش سلمان

بإشراف
أ.د. علي جواد كاظم الكناني

١٤٤٢هـ ٢٠٢١ م