Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7733091

ROBUST SPARSE SLICED INVERSE REGRESSION VIA ELASTIC NET

Dhuha Alkim and Ali Alkenani
Department of Statistics, College of Administration and Economics, University of AlQadisiyah, Al Diwaniyah, Iraq, Dhuha.salim21.@qu.edu.iq

Abstract

In regression applications, the sliced inverse regression(SIR) is a method for reducing the dimensions without losing any information about the regression. Although, the SIR has been proven as an efficient method to deal with the high dimensionality problems, but it suffers that it gives directions contains all the original predictors. Many researchers suggested approaches to dealing with this problem by combining variable selection methods with SIR. One of these methods combined the SIR method with Elastic Net penalty(SIR-EN). The SIR –EN is an efficient method without assuming a parametric model. It produces accurate and sparse solutions when the predictors are highly correlated under sufficient dimension reduction settings. However, the SSIR- EN is not robust to outliers because of the method use the loss function which is sensitive to outliers in data. As a result, we suggested RSSIR-EN as a robust version of SSIR-EN for outliers in both the dependent variable and the covariates.

Key words: Dimension reduction, SIR, Robust estimation, Elastic-Net.

1-Introduction

Due to the explosion of large information in the past decades, high-dimensional regression analysis problems appear in several applications. The sufficient dimension reduction(SDR) theory has received great attention in high – dimensional regression. The basic idea of SDR is to exchange X with d-dimensional orthogonal dropping PsX onto S, where d<p and p is a number of covariates, without losing information about the conditional distribution of Y|X and without assuming any parameter pattern. Assume Y is a response variable, $X = \left(X_1, X_2, \dots, X_p\right)^T$ is a predictors vector. SDR aims to find the central subspace $S_{Y|X}$ and $S_{Y|X}$ is the intersection of all subspaces S that achieve $Y \perp\!\!\!\perp X | P_s X$, when $\perp\!\!\!\perp$ it indicates independence. Therefore, $P_\beta X$ extracts that information from X to Y, where β is the basis of $S_{Y|X}$(Cook, 1998).There are several suggested methods to find $S_{Y|X}$. One of the most important of these methods which has proven to be effective in dealing with high dimensions is SIR method (Li, 1991). SIR is one of the useful tools, which help to solve the problem of the "high dimensions". It is applied in different fields, including economics and bioinformatics. The results of SIR are linear sums of all the original variables, which may cause difficulty in interpreting the results of SIR. For this reason, there a need to reduce the number of non-zero coefficients in the SIR directions. Under least squares settings , there are many procedures of regularization methods that have been suggested . For example, Lasso (Tibshirani, 1996), Smoothly Clipped Absolute Deviation (Fan and Li, 2001), Elastic Net (Zou and Hastie, 2005), group lasso (Yuan and Li, 2006), Adaptive Lasso (Zou, 2006), and others.

Under SIR framework, several procedures have been proposed that combine SIR method with the regularization methods. Like, a free-models method for determining the contribution of variables which has been suggested by (Cook, 2004). Also, Lasso is combined with SIR to produce shrinkage estimator of SIR by (Ni, 2005). Sparse SIR (SPSIR) in which that combined lasso with LARS into SIR that suggested by (Li and Nachtsheim, 2006). As well as, a number of SDR methods that integrate with the shrinkage estimator that proposed by (Li, 2007). To improve SIR to work when the covariates are highly correlated and settings $p > n$, where n is a sample size(Li and Yin, 2008) they suggested that regularization SIR method (RSIR), for multiple index models with settings $p > n$. A lasso is combined with SIR that proposed by (Lin, 2019). Many researchers suggested approaches to dealing with this problem by combining variable selection methods with SIR. Alkenani (2021) proposed RSIR-Lasso method that does not have the ability to select groups of highly correlated predictors. Alkenani and Hassel (2020) proposed SIR-EN method which deals with correlated predictors but this method sensitive to outliers and are not robust because the method uses the least squares loss function which is sensitive to outliers in data . It is necessary to deal with this problem and solve this problem. The squared loss criterion is used between the covariates and response. Also, the classical estimates of the sample mean and the sample variance of X is used within the least squares formula. These are all sensitive to outliers and are not robust. In this research, we proposed robust method of SIR method with Elastic Net(EN) by using Tukey biweight criterion instead the squared loss criterion. If the derivative of the loss function is descending, the loss function is robust and insensitive to outliers in X and Y (Rousseeuw and Yohai, 1984). Tukey biweight function has this property.

2- SIR and SSIR Methods

For finding the central subspace $S_{Y|X}$, SIR method is suggested by (Li, 1991). This method requires $Z = \Sigma^{\frac{-1}{2}}\big(X - E(X)\big)$, under the linear condition $E(Z/PgZ) = PgZ$, where $\Sigma_x = Cov(X)$ is a population covariance matrix of X and g is a basis to $S_{Y|Z}$. $S_{Y|Z}$ is the central subspace of regression Y on Z. This condition connects with the inverse regression of Z on Y. The kernel matrix of SIR is M and $M = Cov[E(Z|Y)]$, $span(M) \subseteq S_{Y|Z}$. We took a random sample of size n of (X, Y), which has a joint distribution. Let $\bar{X}$ is the sample mean of X, the sample version of Z is $\hat{Z} = \hat{\Sigma}^{-\frac{1}{2}}(X - \bar{X})$ and $\hat{\Sigma}$ is the estimated covariance matrix of X. Assume h be the number of slices also n_y is a number of observations in yth slice. Let $\widehat{M} = \sum_{y=1}^{h} \hat{f}_y \hat{Z}_y \hat{Z}_y^T$ is an estimator of M, where $\hat{f}_y = n_y|n$ and $\hat{Z}_y$ is the average of Z in slice y. Let $\hat{\delta}_1 > \hat{\delta}_2 > \cdots > \hat{\delta}_p \geq 0$ are the eigenvalues corresponding to the eigenvectors $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_p$ of $\widehat{M}$. If d of $S_{Y|Z}$ is known and $span(\hat{\beta}) = span(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_d)$ is a consistent estimator of $S_{Y|X}$, where $\hat{\beta}_i = \hat{\Sigma}^{\frac{-1}{2}} \hat{v}_i$. The SIR method provides the estimator $span(\hat{\beta})$ of $S_{Y|X}$. Generally, $\hat{\beta} \in \mathbb{R}^{p\times d}$ has nonzero elements, when the number of predictions is huge or when the number of predictions is highly correlated, we only take the important predictions that we need to make 'sufficient predictors' combining the regularizations methods with SIR method is the solution to compress number of the coefficients of $\hat{\beta}$ to 0's. The SIR was formulated by (Cook, 2004) as a regression type minimization problem(least squares problem) as follows :

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

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

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

Where $B\in\mathbb{R}^{p\times d}$. The value of B, which minimizes (2) is $\hat{\beta}$ and $span(\hat{\beta})=span(\hat{\Sigma}^{\frac{-1}{2}}\hat{A})$ is the estimator of $S_{Y|X}$. Ni et al. (2005) suggested shrinkage sliced inverse regression(SSIR) for finding $S_{Y|X}$ as $span(diag(\hat{\alpha})\hat{\beta})$, where the shrinkage indices $\tilde{\alpha}=(\tilde{\alpha}_1,\dots,\tilde{\alpha}_p)^T\in\mathbb{R}^p$ are determined by minimizing

$$\sum_{y=1}^{h}\left\|\hat{f}_y^{\frac{1}{2}}\hat{Z}_y-\hat{\Sigma}^{\frac{1}{2}}diag(\hat{B}\hat{C}_y)\alpha\right\|+\lambda\sum_{i=1}^{p}|\alpha_i|, \quad (3)$$

Where $\hat{B}$ and $\hat{C}=(\hat{C}_1,\dots,\hat{C}_h)$ minimize(2). The minimization of (3) can be done by using a standard Lasso algorithm, let $\tilde{Y}=vec(\hat{f}_1^{\frac{1}{2}}\hat{Z}_1,\dots,\hat{f}_h^{\frac{1}{2}}\hat{Z}_h)\in\mathbb{R}^{ph}$, and

$\tilde{X}=(diag(\hat{B}\hat{C}_1)\hat{\Sigma}^{\frac{1}{2}},\dots,diag(\hat{B}\hat{C}_h)\hat{\Sigma}^{\frac{1}{2}})^T\in\mathbb{R}^{ph\times p}$. Where $vec(.)$ is a matrix operator that it puts the columns of the matrix in the single vector. Also, the vector α is the estimator of the lasso in the regression $\tilde{Y}$ and $\tilde{X}$.

3-1-Robust SSIR –EN

-Methodology

SIR use the classical estimates of the sample mean and the sample covariance. Also, it uses the squared loss between the response variable and the covariates. The classical estimates for the mean and covariance and loss squared criterion are very sensitive to outliers and they are not robust .

Gather et.al, (2002) studied SIR's sensitivity to outliers, also suggested a robust version for SIR. Yohai and Sertter(2005) proposed another a robust version of SIR. Prendergast(2005) studied the influence function of SIR. When the derivative of the loss function is redescending, it is robust and insensitive to outliers in Y and X (Rousseeuw and Yohai, 1984). This property is existed in Tukey's biweight loss function (Tukey, 1960). We exchange the loss squared function with Tukey's biweight function in(2-5), that achieve the robustness against outliers in X and Y. Alkenani (2021) suggested robust shrinkage for SIR through combining Lasso with Tukey biweight criterion for SIR. The drawback of this method is that it does not deal with data in groups and also data with high correlations. For this reason, we propose a robust method for variable selection under SDR settings deals with grouped predictors. The proposed method (RSSIR – EN) is a robust version of SSIR-EN (Alkenani and Hassel,2020).

In this study, we replace the classical estimates of sample mean with a robust estimator such as the median and replace the classical estimates of sample covariance matrix with robust covariance matrix estimator as ball covariance. The estimates of suggested RSSIR-EN can be obtained by minimizing the following .

$$\sum_{y=1}^{h} \rho \left(\frac{\hat{f}_y^{\frac{1}{2}} \widehat{ROZ}_y - \widehat{RO\Sigma}^{\frac{1}{2}} diag(\hat{B}\hat{C}_y)\alpha}{\hat{\sigma}} \right) + \lambda_1 \sum_{i=1}^{p} \alpha_i^2 + \lambda_2 \sum_{i=1}^{p} |\alpha_i|, \qquad (4)$$

The minimizing of (4) contains two parts. The first part is robust SIR by using Tukey's biweight function and the second part is Elastic Net penalty function, where, ρ is Tukey's biweight function .

$\hat{\sigma}$ is a robust estimate of σ and MAD is used as an estimate for σ, where MAD is the median absolute deviation .

$\widehat{ROZ}_y$ is a robust versions of $\hat{Z}_y$.

$\widehat{RO\Sigma}^{\frac{1}{2}}$ is a robust version of $\hat{\Sigma}^{\frac{1}{2}}$.

$\lambda_1, \lambda_2 \geq 0$ is the tuning parameters of EN .

The function of Tukey's biweight is as follows:

$$\rho_c(u) = \begin{Bmatrix} \left(\frac{c^2}{6}\right)\left\{1 - \left[1 - \left(\frac{c}{6}\right)^2\right]^3\right\} & if\ |u| \leq c \\ \frac{c^2}{6} & if\ |u| \leq c \end{Bmatrix} \qquad (5)$$

where c controls the robustness level .

3-2-Robust measures for location and dispersion

SIR method is based on first and second moments estimators of data, which are sensitive to outliers. In this study, we propose a robust version. The median and ball covariance are robust measures to outliers for location and dispersion. Pan et.al(2018) suggested ball covariance (BCov) as a robust measure for dependency between two random vectors as follows;

Let$\{U_k, V_k\}_{k=1}^{n}$ be i.i.d. sample of (U,V). Define $\delta_{ij,k}^{U} = I\{U_k \in \bar{B}_{\xi U}(U_i, U_j)\}$, where $I(.)$ is defined as the indicator function, $\delta_{ij,kl}^{U} = \delta_{ij,k}^{U}\delta_{ij,l}^{U}$ and $\xi_{ij,klst}^{U} = (\delta_{ij,kl}^{U} + \delta_{ij,st}^{U} - \delta_{ij,ks}^{U} - \delta_{ij,lt}^{U})/2$. $\xi_{ij,klst}^{V}$ is definition similar to $\xi_{ij,klst}^{U}$. The ball covariance that defined by; $BCov_n(U,V) = \left(\frac{1}{n^6}\sum_{i,j,k,l,s,t=1}^{n} \xi_{ij,klst}^{U}\xi_{ij,klst}^{V}\right)^{1/2}$

For more details about BCov see (Pan et. al, 2018) and (Zhang and Chen, 2019).

3-3-Selection the tuning parameter λ

There are some information criterion for example, generalized cross validation(GCV) which is proposed by (Ni et.al, 2005), Akaike's information criterion(AIC) which is proposed by (Akaike, 1973), Bayesian information criterion(BIC) which is proposed by (Schwarz, 1978) , Residual information criterion(RIC) which is (Shi and Tsai, 2002) and Robust residual information criterion(RRIC) which is (Alkenani, 2020). These criterion information are proposed to selection λ according to the following formulas ;

$$GCV = \frac{RSS}{n\{1 - p(\lambda)/n\}^2} \qquad (6)$$

$$AIC = n\log(RSS/n) + 2p(\lambda) \qquad (7)$$

$$BIC = n\log(RSS/n) + \log(n)p(\lambda), \qquad (8)$$

$$RIC = \{n - p(\lambda)\}\log(RSS/\{n - p(\lambda)\} + p(\lambda)\{\log(n) - 1\} + \frac{4}{\{n - p(\lambda) - 2\}}, \qquad (9)$$

where $RSS = \sum_{y=1}^{h} \left\| \hat{f}_y^{1/2} \hat{Z}_y - \hat{\Sigma}^{\frac{1}{2}} diag(\hat{B}\hat{C}_y)\alpha \right\|^2$ is the residual sum of squares of lasso fit and $p(\lambda)$ denotes to the number of non-zero coefficients .

The simulation results of (Alkenani, 2020) show that using RRIC for selection λ gives better performance and consistent results for SIR-EN. In this paper, we employed RRIC which is proposed by (Alkenani, 2020) in our simulations, which is as follows:

$$RRIC = \{n - p(\lambda)\} \log (RSS/\{n - p(\lambda)\}) + p(\lambda) \{\log(n) - 1\} + \frac{4}{\{n-p(\lambda)-2}, \quad (10)$$

$$RRSS = \sum_{y=1}^{h} \rho \left(\frac{\hat{f}_y^{1/2} \widehat{ROZ}_y - \widehat{R\ \ \Sigma}^{\frac{1}{2}} diag(\hat{B}\hat{C}_y)\alpha}{\hat{\sigma}} \right) \quad (11)$$

$$\widehat{ROZ}_y = \widehat{BCov}_n^{\frac{-1}{2}} (X - median(X)), \text{ and } \widehat{RO\Sigma}^{\frac{1}{2}} = \widehat{BCov}_n^{\frac{-1}{2}} \quad (12)$$

3-4-Determination of d

For suggested RSSIR-EN, $d = \dim (S_{Y|X})$ is assumed as know, and we need to estimate d through data. Many ways are suggested to determine d. For example, (Li, 1991), (Schott,1994), (Bura and Cook, 2001) and (Cook and Yin, 2001). Zhu et.al(2006) proposed to estimate d via the nonzero eigenvalues number of $Cov[E(X|Y)]$ matrix, or equivalently, number of eigenvalues of $\Omega = Cov[E(Y|X)] + I_p$ that are greater than, where I_p indicates to identity matrix of p.

Let k is the number of $\hat{\delta}_i > 1$, $\hat{\delta}_1, \dots, \hat{\delta}_p$ are the eigenvalues of $\hat{\Omega}$, $\hat{\Omega}$ is the estimated of Ω and $C_n^*$is a constant. Zhu et.al (2006) proposed to estimat d as the follows:

$$\boldsymbol{\hat{d} = arg_{m\in\{0,1,\dots,p-1\}} \max \{\frac{n}{2} \sum_{i=1+\min(k,m)}^{p} (\log(\hat{\delta}_i) + 1 - \hat{\delta}_i) - \frac{C_n^* m(2p-m+1)}{2}} \quad (13)$$

Several forms are suggested for C_n^*. Li and Yin (2008) proposed $C_n^* = \log(n)h/n$ and they employed it simulations.

Alkenani (2020) suggested a robust method to estimated based on(Zhu et. al, 2006) in formula (13).

Under Z- scale and without losing generality of the standardized predictor Z, because of $S_{Y|X} = \Sigma^{\frac{-1}{2}} S_{Y|X}$. Alkenani (2020) estimates d via the eigenvalues number of the robust matrix $Ro\Omega = RoM + I_p$that are greater than one, where RoM is a robust estimate of M the kernel matrix of SIR as follows:

$$\widehat{RoM} = \sum_{y=1}^{h} \hat{f}_y \widehat{RoZ}_y \widehat{RoZ}_y^T, \quad (14)$$

Let k is a number of $\hat{\gamma}_i > 1$, $\hat{\gamma}_1, \dots, \hat{\gamma}_p$ are the eigenvalues of $Ro\Omega$, $\widehat{Ro\Omega}$ is a robust version of $Ro\Omega$. Alkenani (2020) suggested the robust estimator of d as follows:

$$\hat{d} = arg_{m\in\{0,1,\dots,p-1\}} max \left\{ \frac{n}{2} \sum_{i=1+\mathrm{m}\quad(k,m)}^{p} (\log(\hat{\gamma}_i) + 1 - \hat{\gamma}_i) - \frac{C_n^* m(2p-m+1)}{2} \right\}, \quad (15)$$

In the simulation we used formula (15) which is suggested by (Alkenani, 2020).

4-Simulation study

In this part, the main purpose of this section is compare the performance of the proposed method (RSSIR-EN) with RSSIR-Lasso and SSIR-EN methods, in terms the efficiency and variables selection. In all examples, we employed a robust RIC that proposed by (Alkenani, 2021) for the tuning parameter. The R code for SSIR-Lasso is made by (Ni et.al, 2005). The R code for SIR-AL is made by (Alkenani and Salman, 2021). The R code for RSIR-L is made by

(Alkenani,2021) . The R code for SSIR-EN is made by (Alkenani and Hassel, 2020). . The R code for RSSIR-EN is made by (Alkenani and Alkim, 2023). In term of variable selection, the average number of zeros coefficients(Ave0's) is reported. In term of prediction accuracy, the mean squared error (MSE) is reported. Four distributions are assumed for ε and X.

Dist.1. The standard normal distribution N(0,1).

Dist.2. $t_3/\sqrt{3}$, t-distribution with 3 degree of freedom.

Dist.3. $(1-\alpha)N(0,1)+\alpha\, N(0,10^2)$

Dist.4. $(1-\alpha)N(0,1)+\alpha\, U(-50,50)$, $(1-\alpha)$ from standard normal and α from normal with mean 0 and variance 100 and uniform(-50,50).

<u>Example .</u> Let d=1, p=40 and n=50,100 and200. Consider the model,

$$Y = 1 + 2(\theta^T + 3) \times \log(3|\theta^T|) + \varepsilon$$

$$\theta = \left(\underbrace{0,\dots,0}_{10}, \underbrace{2,\dots,2}_{10}, \underbrace{0,\dots,0}_{10}, \underbrace{2,\dots,2}_{10}\right)^T,$$

With pairwise correlation (Xi ,X j) = 0.90 for all i and j.

Table1: The results of example, based on Ave0's, and MSE when n = 50 and =0.05, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	4.487977e-04	1.52
	RSSIR-Lasso	4.809671e-05	3.62
	RSSIR –EN	2.037116e-05	3.99
2	SSIR-EN	0.04487381	1.39
	RSSIR-Lasso	4.824503e-05	3.01
	RSSIR –EN	2.045203e-05	5.04
3	SSIR-EN	0.04483268	1.53
	RSSIR-Lasso	4.7773e-05	3.02
	RSSIR –EN	2.022285e-05	6.75
4	SSIR-EN	0.04484397	1.56
	RSSIR-Lasso	4.768485e-05	3.22
	RSSIR -EN	1.999591e-05	6.32

Table2: The results of example, based on Ave0's, and MSE when n = 50 and =0.10, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	6.4855383e-05	2.57
	RSSIR-Lasso	5.166768e-05	5.03
	RSSIR –EN	1.832232e-05	6.52
2	SSIR-EN	0.04854842	2.46
	RSSIR-Lasso	5.243622e-05	5.02
	RSSIR –EN	1.831266e-05	6.33

3	SSIR-EN	0.04846247	2.44
	RSSIR-Lasso	5.019152e-05	5.02
	RSSIR –EN	1.785424e-05	6.57
4	SSIR-EN	0.04847204	2.19
	RSSIR-Lasso	5.05871e-05	5.10
	RSSIR –EN	1.797277e-05	6.14

Table3: The results of example, based on Ave0's, and MSE when n = 50 and =0.15, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	7.517086e-05	2.58
	RSSIR-Lasso	6.830077e-05	5.04
	RSSIR –EN	7.305918e-06	6.67
2	SSIR-EN	0.07517058	2.04
	RSSIR-Lasso	6.82998e-05	5.02
	RSSIR –EN	7.254978e-06	6.16
3	SSIR-EN	0.07502427	2.48
	RSSIR-Lasso	6.775194e-05	6.02
	RSSIR –EN	7.201068e-06	6.07
4	SSIR-EN	0.07506487	2.94
	RSSIR-Lasso	6.76135e-05	6.02
	RSSIR –EN	7.199168e-06	6.47

Table4: The results of example, based on Ave0's, and MSE when n = 50 and =0.20, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	9.393552e-05	3.50
	RSSIR-Lasso	8.979675e-05	5.76
	RSSIR –EN	5.203075e-06	5.814
2	SSIR-EN	0.09393545	3.48
	RSSIR-Lasso	8.973807e-05	6.26
	RSSIR –EN	5.111469e-06	7.36
3	SSIR-EN	0.09264675	3.37
	RSSIR-Lasso	8.71893e-05	6.02
	RSSIR –EN	5.028723e-06	8.97
4	SSIR-EN	0.09384761	3.49
	RSSIR-Lasso	9.049795e-05	4.04
	RSSIR –EN	4.941705e-06	8.63

Table5:The results of example1, based on Ave0's, and MSE when n = 50 and =0.25, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	5.254388e-05	4.40
	RSSIR-Lasso	4.595704e-05	6.92
	RSSIR –EN	4.261535e-05	7.31
2	SSIR-EN	0.1254414	4.53
	RSSIR-Lasso	0.0001302293	6.01
	RSSIR –EN	4.492947e-06	7.45
3	SSIR-EN	0.1191665	4.47
	RSSIR-Lasso	0.0001157054	6.02
	RSSIR –EN	4.708869e-06	7.06
4	SSIR-EN	0.1254033	5.45
	RSSIR-Lasso	0.0001304494	7.04
	RSSIR –EN	4.488822e-06	7.44

Table6:The results of example, based on Ave0's, and MSE when n = 50 and =0.30, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	1.215631e-06	5.21
	RSSIR-Lasso	1.268868e-06	6.36
	RSSIR –EN	1.115171e-06	6.42
2	SSIR-EN	0.160634	5.48
	RSSIR-Lasso	0.0001462529	6.02
	RSSIR –EN	2.948553e-06	7.73
3	SSIR-EN	0.122353	5.47
	RSSIR-Lasso	0.0001189397	6.03
	RSSIR –EN	4.567956e-06	8.95
4	SSIR-EN	0.1292572	5.53
	RSSIR-Lasso	0.0001287179	7.02
	RSSIR –EN	4.358948e-06	8.52

Table7: The results of example, based on Ave0's, and MSE when n = 50 and =0.35, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	1.437201e-04	5.35
	RSSIR-Lasso	1.328295e-05	6.05
	RSSIR –EN	1.969545e-06	7.27
2	SSIR-EN	0.1437244	5.30

	RSSIR-Lasso	0.000132213	6.03
	RSSIR –EN	3.934988e-06	8.36
3	SSIR-EN	0.1373105	5.51
	RSSIR-Lasso	0.0001336923	7.04
	RSSIR –EN	3.993004e-06	9.01
4	SSIR-EN	0.1436664	6.33
	RSSIR-Lasso	0.0001332097	8.01
	RSSIR –EN	3.880173e-06	9.64

Table8:The results of example, based on Ave0's, and MSE when n = 100 and =0.05, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	2.242592e-05	7.50
	RSSIR-Lasso	2.172927e-05	8.036
	RSSIR –EN	5.463114e-06	8.89
2	SSIR-EN	0.02683092	7.54
	RSSIR-Lasso	2.735996e-05	8.02
	RSSIR –EN	1.121969e-05	9.38
3	SSIR-EN	0.03984808	7.48
	RSSIR-Lasso	4.137197e-05	8.04
	RSSIR –EN	3.213351e-06	10.90
4	SSIR-EN	0.0228326	7.42
	RSSIR-Lasso	2.505456e-05	9.03
	RSSIR –EN	1.722706e-05	10.70

Table9:The results of example, based on Ave0's, and MSE when n = 100 and =0.10, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	5.471879e-05	7.59
	RSSIR-Lasso	5.028328e-05	9.04
	RSSIR –EN	6.315377e-06	10.08
2	SSIR-EN	0.05472697	8.62
	RSSIR-Lasso	5.034745e-05	9.03
	RSSIR –EN	6.38995e-06	10.04
3	SSIR-EN	0.05464707	8.59
	RSSIR-Lasso	5.017851e-05	10.03
	RSSIR –EN	6.064363e-06	10.07
4	SSIR-EN	0.05465397	8.55
	RSSIR-Lasso	5.003533e-05	10.03

	RSSIR –EN	6.061648e-06	11.41

Table10:The results of example, based on Ave0's, and MSE when n = 100 and =0.15, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	8.668329e-05	9.51
	RSSIR-Lasso	8.319031e-05	10.05
	RSSIR –EN	2.55889e-06	10.90
2	SSIR-EN	0.08368376	9.60
	RSSIR-Lasso	8.602455e-05	10.04
	RSSIR –EN	2.634705e-06	11.92
3	SSIR-EN	0.08323136	9.51
	RSSIR-Lasso	8.147524e-05	10.04
	RSSIR –EN	2.540456e-06	12.80
4	SSIR-EN	0.08354529	10.70
	RSSIR-Lasso	8.214336e-05	11.04
	RSSIR –EN	2.551316e-06	12.45

Table11:The results of example, based on Ave0's, and MSE when n = 100 and =0.20, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	1.127955e-04	10.43
	RSSIR-Lasso	1.147732e-05	11.04
	RSSIR –EN	1.852285e-06	12.60
2	SSIR-EN	0.1127933	10.56
	RSSIR-Lasso	0.000114415	11.02
	RSSIR –EN	1.865391e-06	12.14
3	SSIR-EN	0.1102235	10.53
	RSSIR-Lasso	0.0001089213	11.05
	RSSIR –EN	1.682965e-06	12.71
4	SSIR-EN	0.1123473	10.33
	RSSIR-Lasso	0.0001124245	11.02
	RSSIR –EN	1.691537e-06	12.46

Table12:The results of example, based on Ave0's, and MSE when n = 100 and =0.25, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	1.184495e-05	10.60
	RSSIR-Lasso	1.020961e-05	11.04

	RSSIR –EN	1.269786e-06	12.05
2	SSIR-EN	0.1184512	11.57
	RSSIR-Lasso	0.0001218806	12.05
	RSSIR –EN	1.277237e-06	13.38
3	SSIR-EN	0.11415	11.41
	RSSIR-Lasso	0.0001130521	12.03
	RSSIR –EN	1.24263e-06	13.12
4	SSIR-EN	0.1173952	11.37
	RSSIR-Lasso	0.000117889	13.02
	RSSIR –EN	1.225129e-06	13.93

Table13:The results of example, based on Ave0's, and MSE when n = 100 and =0.30, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	1.429102e-05	10.53
	RSSIR-Lasso	1.404676e-05	12.04
	RSSIR –EN	1.005395e-06	13.73
2	SSIR-EN	0.1429153	11.52
	RSSIR-Lasso	0.0001408843	12.04
	RSSIR –EN	1.019099e-06	13.18
3	SSIR-EN	0.1391714	11.55
	RSSIR-Lasso	0.0001337227	13.03
	RSSIR –EN	1.020009e-06	13.88
4	SSIR-EN	0.1428816	11.39
	RSSIR-Lasso	0.0001404269	13.01
	RSSIR –EN	1.003903e-06	14.13

Table14:The results of example, based on Ave0's, and MSE when n = 100 and =0.35, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	1.49855e-04	11.44
	RSSIR-Lasso	1.449111e-04	13.04
	RSSIR –EN	1.062438e-05	13.61
2	SSIR-EN	0.1498605	12.27
	RSSIR-Lasso	0.000142984	13.03
	RSSIR –EN	1.088247e-06	14.14
3	SSIR-EN	0.1496073	12.66
	RSSIR-Lasso	0.0001452582	13.03
	RSSIR –EN	1.014076e-06	14.78

4	SSIR-EN	0.1497792	12.50
	RSSIR-Lasso	0.0001462159	14.02
	RSSIR –EN	1.01036e-06	14.06

Table15:The results of example, based on Ave0's, and MSE when n = 200 and =0.05, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	3.171039e-05	12.39
	RSSIR-Lasso	3.162419e-05	14.03
	RSSIR -EN	5.088275e-06	14.73
2	SSIR-EN	0.03171494	13.42
	RSSIR-Lasso	3.189382e-05	14.02
	RSSIR -EN	5.191862e-06	16.44
3	SSIR-EN	0.03166915	13.38
	RSSIR-Lasso	3.126864e-05	14.02
	RSSIR -EN	4.929933e-06	16.96
4	SSIR-EN	0.0228974	13.53
	RSSIR-Lasso	2.472147e-05	15.04
	RSSIR -EN	8.055795e-06	16.58

Table16:The results of example, based on Ave0's, and MSE when n = 200 and =0.10, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	4.578004e-04	13.47
	RSSIR-Lasso	4.75026e-05	15.03
	RSSIR -EN	1.821982e-05	16.36
2	SSIR-EN	0.04578566	14.56
	RSSIR-Lasso	4.775888e-05	15.03
	RSSIR -EN	1.891925e-06	16.96
3	SSIR-EN	0.04570286	14.49
	RSSIR-Lasso	4.772843e-05	16.04
	RSSIR -EN	1.795677e-06	16.12
4	SSIR-EN	0.04571188	14.60
	RSSIR-Lasso	4.78515e-05	16.06
	RSSIR -EN	1.805547e-06	17.08

Table17:The results of example, based on Ave0's, and MSE when n = 200 and =0.15, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's

1	SSIR-EN	6.662301e-05	14.4
	RSSIR-Lasso	4.697981e-05	16.01
	RSSIR –EN	1.188927e-06	17.75
2	SSIR-EN	0.06662522	15.49
	RSSIR-Lasso	6.687145e-05	17.04
	RSSIR –EN	1.236344e-06	17.96
3	SSIR-EN	0.06649001	15.41
	RSSIR-Lasso	6.761884e-05	17.02
	RSSIR –EN	1.22211e-06	17.44
4	SSIR-EN	0.06652121	15.56
	RSSIR-Lasso	6.790988e-05	17.05
	RSSIR –EN	1.205502e-06	17.15

Table18:The results of example, based on Ave0's, and MSE when n = 200 and =0.20, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	9.373769e-05	14.36
	RSSIR-Lasso	8.388876e-05	16.03
	RSSIR –EN	1.026498e-06	17.13
2	SSIR-EN	0.08373724	14.40
	RSSIR-Lasso	8.373449e-05	16.03
	RSSIR –EN	1.063089e-06	17.90
3	SSIR-EN	0.08225067	14.38
	RSSIR-Lasso	8.186602e-05	16.01
	RSSIR -EN	9.963342e-07	17.06
4	SSIR-EN	0.08357264	14.43
	RSSIR-Lasso	8.476293e-05	17.01
	RSSIR -EN	1.009577e-06	17.54

Table19: The results of example, based on Ave0's, and MSE when n = 200 and =0.25, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	1.336718e-05	14.5
	RSSIR-Lasso	1.265607 e-05	16.06
	RSSIR -EN	1.019899e-06	17.79
2	SSIR-EN	0.1232604	14.37
	RSSIR-Lasso	0.0001305999	16.01
	RSSIR –EN	9.203766e-07	17.63
3	SSIR-EN	0.1220834	15.45

	RSSIR-Lasso	0.0001169491	16.03
	RSSIR –EN	4.841108e-07	17.75
4	SSIR-EN	0.1652626	15.47
	RSSIR-Lasso	0.0001626196	17.03
	RSSIR -EN	5.528127e-07	17.77

Table20:The results of example, based on Ave0's, and MSE when n = 200 and =0.30, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	1.465006e-06	14.49
	RSSIR-Lasso	1.278697e-06	15.03
	RSSIR –EN	4.784754e-07	16.93
2	SSIR-EN	0.1478462	15.31
	RSSIR-Lasso	0.0001443273	16.05
	RSSIR –EN	7.593916e-07	17.13
3	SSIR-EN	0.1511251	15.41
	RSSIR-Lasso	0.0001445022	16.03
	RSSIR –EN	4.93295e-07	17.24
4	SSIR-EN	0.1426464	15.33
	RSSIR-Lasso	0.0001447042	16.03
	RSSIR -EN	6.6451e-07	18.67

Table21:The results of example, based on Ave0's, and MSE when n = 200 and =0.35, for dist3 and dist4.

Dist	Method	MSE	Ave. 0's
1	SSIR-EN	1.341259e-6	15.41
	RSSIR-Lasso	1.259276e-6	16.03
	RSSIR -EN	6.785619e-07	17.48
2	SSIR-EN	0.1341288	15.43
	RSSIR-Lasso	0.0001255179	17.02
	RSSIR -EN	7.097112e-07	18.99
3	SSIR-EN	0.1273917	15.49
	RSSIR-Lasso	0.0001217017	18.01
	RSSIR -EN	8.010773e-07	19.79
4	SSIR-EN	0.1341003	16.51
	RSSIR-Lasso	0.0001287542	18.06
	RSSIR -EN	6.420542e-07	19.41

From the results of tables 1,2,3,…., it can be seen that there is a slight outperform for the suggested approach where it has a lower MSE and it has a bigger values based on Ave.0's. In case of three distributions of x and error, we can note that SIR-EN method was sensitive for the contamination but other methods RSSIR-Lasso and RSSIR-EN were not affected because they have the robustness. Also, we can see that the performance of RSSIR-EN outperformed RSSIR-Lasso method in terms of V.S based on Ave.0's. For the previous example, the MSE values for RSSIR-EN are less than their values for RSSIR-Lasso and SSIR-EN. This means that the suggested RSSIR-EN has the best performance than the rest methods depending on the MSE of simulation studies. It is clear that under various settings, the proposed RSSIR-EN has a good performance in terms of variable selection and estimation accuracy.

5- Boston housing data

This data was collected by (Harrison and Rubinfeld, 1978), the data set includes n = 506 observations and p = 14 predictor, where y is medv (median value of owner occupied homes in $ 1000's). X includes 13 predictors. The predictors are : x1 is (rate of crime), x2 is (proportion of residential land zoned), x3 is (proportion of non-retail business acres), x4 is (the Charles river (= 1 if tract bounds river; 0 otherwise)), x5 is (concentration of nitric oxides), x6 is (average of rooms), x7 is (proportion of owner-occupied units), x8 is (weighted mean of distances), x9 (index of accessibility),x10 is (rate of property tax), x11 (pupil – teacher ratio), x12 is (proportion of black population) and x13 is (lower status). The data set is available and public from R package. The predictors and y are standardized separately for ease of explanation. To verify the performance of the proposed RSSIR-EN.

We made a comparison to evaluate the accuracy of the suggested method RSSIR-EN and SSIR-EN, RSSIR-Lasso methods based on the mean squared error(MSE) and number of zero's coefficient

Table22: The results of Boston housing based on number of zero's and MSE

Method	MSE	Number of zero's
SSIR-EN	0.05335893	9
RSSIR-Lasso	0.01069643	10
RSSIR-EN	0.007849862	11

From the result of table 22, it can be seen that there is a slight outperform for the suggested approach where it has a lower MSE and it has a bigger values based on number of zero's coefficients. We can note that SIR-EN method was sensitive for the contamination but other methods RSSIR-Lasso and RSSIR-EN were not affected because they have the robustness. Also, we can see that the performance of RSSIR-EN outperformed RSSIR-Lasso method in terms of V.S based on number of zero's coefficients. For the Boston housing data, the MSE values for RSSIR-EN are less than their values for RSSIR-Lasso and SSIR-EN. This means that the suggested RSSIR-EN has the best performance than the rest methods depending on the MSE. It is clear that under various settings, the proposed RSSIR-EN has a good performance in terms of variable selection and estimation accuracy.

Table23: The results of Boston housing based on beta

SSIR-EN	RSSIR-Lasso	RSSIR-EN
1.735522	0.0000000	0.00000000
0.000000	0.7516025	0.00000000
0.000000	0.0000000	0.00000000
0.000000	0.0000000	0.00000000
1.504926	0.0000000	0.00000000
0.000000	0.0000000	0.00000000
0.000000	0.0000000	0.00000000
0.000000	0.0000000	0.00000000
0.000000	0.0000000	0.00000000
5.473775	0.0000000	7.76415849
2.399465	1.0064451	0.00000000
0.000000	3.1148360	0.00000000
0.000000	0.0000000	1.83860357

From the correlation matrix in table, it is clear that there are high correlations among the variables. High pairwise correlations are found in (X9,X1),(X10,X1),(X8,X2),(X5,X3),(X7,X3),(X9,X3),(X10,X5),(X13,X5) and others as shown in the following table24;

Table24: The results of Boston housing based on correlation of variables

	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	b	lstat
crim	1	-0.20047	0.406583	-0.05589	0.420972	-0.21925	0.352734	-0.37967	0.625505	0.582764	0.289946	-0.38506	0.455621
zn	-0.20047	1	-0.53383	-0.0427	-0.5166	0.311991	-0.56954	0.664408	-0.31195	-0.31456	-0.39168	0.17552	-0.41299
indus	0.406583	-0.53383	1	0.062938	0.763651	-0.39168	0.644779	-0.70803	0.595129	0.72076	0.383248	-0.35698	0.6038
chas	-0.05589	-0.0427	0.062938	1	0.091203	0.091251	0.086518	-0.09918	-0.00737	-0.03559	-0.12152	0.048788	-0.05393
nox	0.420972	-0.5166	0.763651	0.091203	1	-0.30219	0.73147	-0.76923	0.611441	0.668023	0.188933	-0.38005	0.590879
rm	-0.21925	0.311991	-0.39168	0.091251	-0.30219	1	-0.24026	0.205246	-0.20985	-0.29205	-0.3555	0.128069	-0.61381
age	0.352734	-0.56954	0.644779	0.086518	0.73147	-0.24026	1	-0.74788	0.456022	0.506456	0.261515	-0.27353	0.602339
dis	-0.37967	0.664408	-0.70803	-0.09918	-0.76923	0.205246	-0.74788	1	-0.49459	-0.53443	-0.23247	0.291512	-0.497
rad	0.625505	-0.31195	0.595129	-0.00737	0.611441	-0.20985	0.456022	-0.49459	1	0.910228	0.464741	-0.44441	0.488676
tax	0.582764	-0.31456	0.72076	-0.03559	0.668023	-0.29205	0.506456	-0.53443	0.910228	1	0.460853	-0.44181	0.543993
ptratio	0.289946	-0.39168	0.383248	-0.12152	0.188933	-0.3555	0.261515	-0.23247	0.464741	0.460853	1	-0.17738	0.374044
b	-0.38506	0.17552	-0.35698	0.048788	-0.38005	0.128069	-0.27353	0.291512	-0.44441	-0.44181	-0.17738	1	-0.36609
lstat	0.455621	-0.41299	0.6038	-0.05393	0.590879	-0.61381	0.602339	-0.497	0.488676	0.543993	0.374044	-0.36609	1

As well as testing the presence of outliers through the method (in variables Boston housing data.

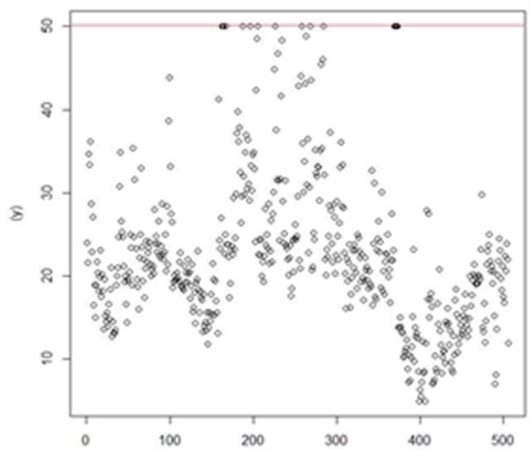

Figure-1: Test for the presence of outliers in Y

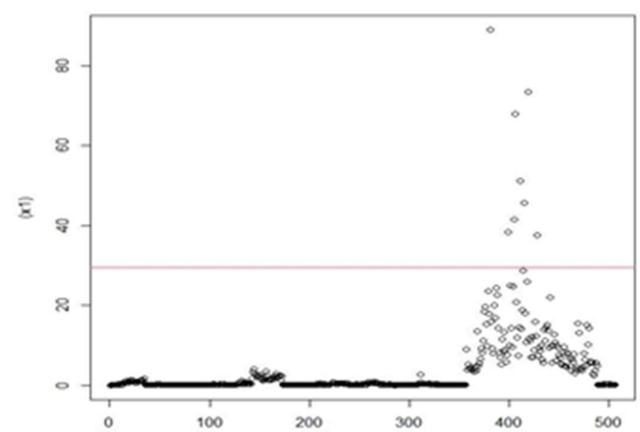

Figure-2: Test for the presence of outliers in X1

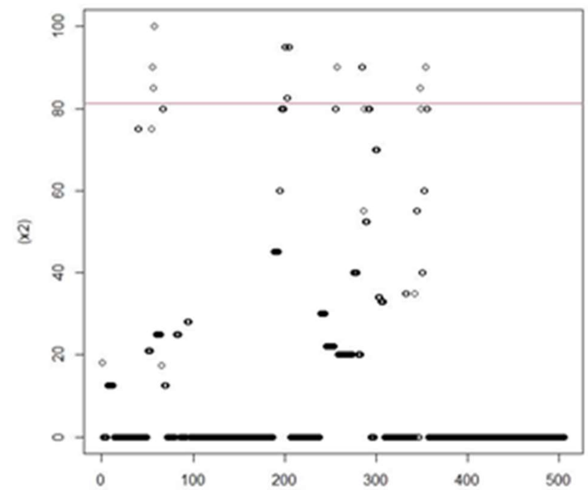

Figure-3: Test for the presence of outliers in X2

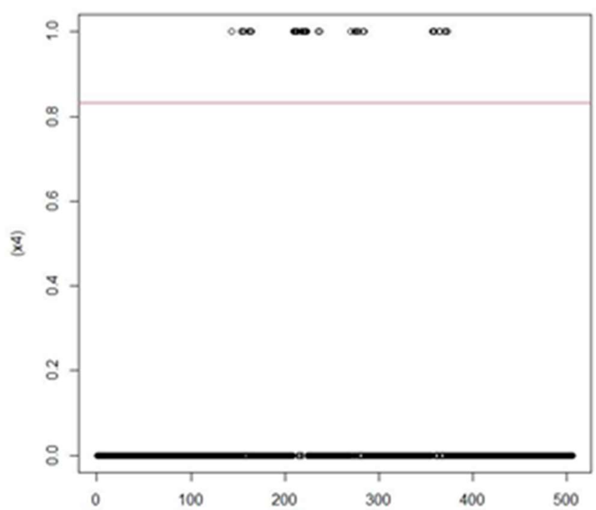

Figure-4: Test for the presence of outliers in X4

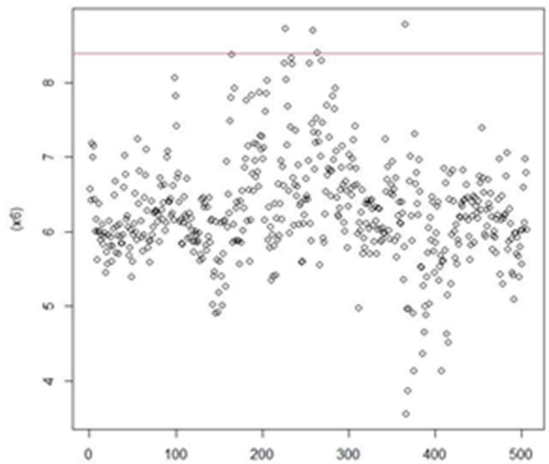

Figure-5: Test for the presence of outliers in X6

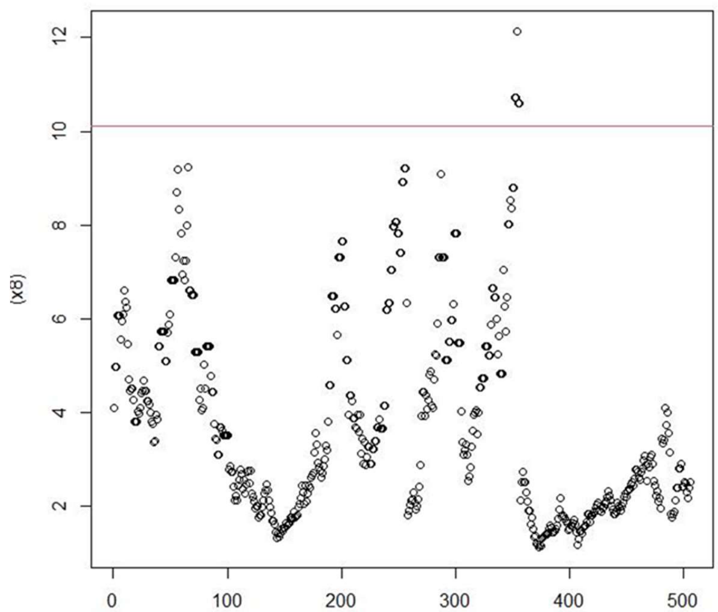

Figure-6: Test for the presence of outliers in X8

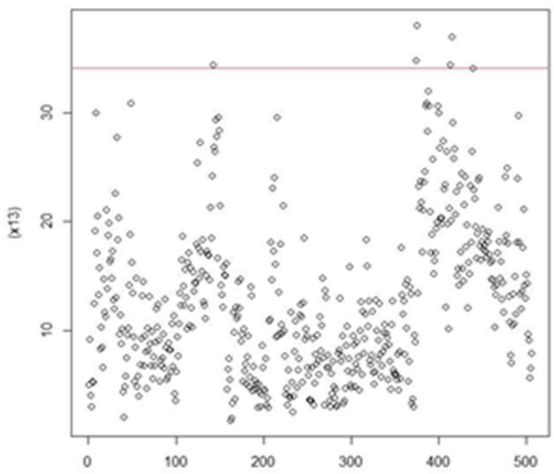

Figure-7: Test for the presence of outliers in X13

6. Conclusion

In this article, RSSIR-EN method is proposed. It is a robust variable selection method under SDR settings. Computationally, the simulations results and the real data analysis show that the RSSIR-EN has better performance than SSIR-EN and RSSIR-Lasso when the outliers exist in Y and X in terms the estimation accuracy and variable selection. Also, the RSSIR-EN gives very close results to SSIR-EN when there are no outliers. Simulations and real data analysis showed that the RSSIR-EN has favorable predictive accuracy.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In second International Symposium on Information Theory. Akademia Kiado, Budapest, pp. 267-281.
- Alkenani, A. (2020). Robust variable selection in sliced inverse regression using Tukey biweight criterion and ball covariance. Journal of Physics Conference Series, 1664, 012034.
- Alkenani, A. (2021). Robust group identification and variable selection in sliced inverse regression using Tukey's biweight criterion and ball covariance. Gazi University Journal of Science 35 (2).
- Alkenani, A. and Hassel, M. (2020). Regularized sliced inverse regression through the elastic net penalty. Journal of Physics Conference Series. Submitted.
- Alkenani, A. and Aljobori, N. (2021). Robust sparse MAVE through elastic net penalty. International journal of Agricultural and Statistical Sciences, Vol.17, Supplement 1, 2039 - 2046.
- Alkenani, A. and Dikheel, T. (2017). Robust Group Identification and Variable Selection in Regression. Journal of Probability and Statistics 2017, Article ID 2170816, 8 pages.

Alkenani, A. and Rahman, E. (2020). Sparse minimum average variance estimation via the adaptive elastic net when the predictors correlated, Journal of Physics Conference Series, 1591, 012041.

 Alkenani, A. and Rahman, E. (2021). Regularized MAVE through the elastic net with correlated predictors, Journal of Physics Conference Series, 1897, 012018.
 Alkenani, A. and Reisan, T (2016). Sparse sliced inverse quantile regression. Journal of Mathematics and Statistics. Volume 12, Issue 3.
 Alkenani, A. and Yu, K. (2013). Sparse MAVE with oracle penalties. Advances and Applications in Statistics 34, 85–105. Bellman, R. E. (1961). Adaptive Control Processes. Princeton University Press, Princeton, New Jersey.
 Bondell, H. D. and Reich, B. J. (2008). Simultaneous regression shrinkage, variable selection and clustering of predictors with OSCAR," Biometrics, 64, 115-123.
 Breiman, L. (1996). Heuristics of instability and stabilization in model selection. The Annals of Statistics, 24(6), 2350–2383.
 Brillinger, D. R. (1983). A generalized linear model with (Gaussian) regression variables. In A Festschrift for Erich L. Lehmann (eds P. J. Bickel, k. A. Doksum and J. L. Hodges, Jr), pp. 97-141 Belmont: Wadsworth.
 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/).
 Chand , S., and Kamal , S .(2011) , “ variable selection by lasso – type method “ . Pakistan Journal of statistics and operation research , pp. 451-464
 Cizek,P. and Hardle, W. (2006). Robust estimation of dimension reduction space. Computational Statistics and data analysis, 51, 545-555.
 Common, P. (1984). Independent component analysis, a new concept?. Signal Processing, 36(3), 287–314.
 Cook, R. (1998). Regression graphics: ideas for studying the regression through graphics. New York, Wily.
 Cook, R. D. and Li, B. (2002). Dimension reduction for the conditional mean in regression. The Annals of Statistics 30, 455–474.
 Cook, R. D. and Weisberg, S. (1991). Discussion of Li (1991). Journal of the American Statistical Association 86, 328–332.
 Desboulets, L. D. D. (2018). A review on variable selection in regression analysis. Econometrics, 6(4), 45
 Donoho, D. L., and Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. Biometrika, 81(3), 425–455.
 Efroymson, M. A. (1960). Multiple regression analysis. Mathematical Methods for Digital Computers, 191–203.
 Efron, B. et al. (2004). Least angle regression. The Annals of Statistics 32, 407–499.
 Fan, J. and Li, R. Z. (2001). Variable selection via non-concave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96,1348–1360.
 Gorsuch, R. L. (1983). Factor Analysis, Hillsdale, New Jersey, L. Erlbaum Associates.
 Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157 – 1182.
 Hassel, M. (2021). Sparse sliced inverse regression via elastic net penalty with an application. Thesis submitted to college of administration and economics. University of Al-Qadisiyah. Iraq.

 Hesterberg, T., Choi, N. H., Meier, L., & Fraley, C. (2008). Least angle and ℓ1 penalized regression: A review. Statistics Surveys, 2, 61-93.
 Härdle, W. and Stoker, T. (1989). Investigating smooth multiple regression by the method of average derivatives. Journal of the American Statistical Association, 84(408), 986–995.
 Horowitz ,J.L., and Lee ,S . (2002) , “ semi-parametric methods in applied econometrics “ . statistical modeling , 2 , 3-22.
 Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics, 58(1–2), 71– 120.
 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.
 Jolliffe, I. T. (2002). Principal components in regression analysis. Principal Component Analysis, 167–198. Kong , E ., Xia , Yi . (2007) , “ variable selection for the single index model “ . Biometrika 94 , pp. 217-229.
 Li, K. (1991). Sliced inverse regression for dimension reduction (with discussion). Journal of the American Statistical Association 86, 316–342.
 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, 1025–1039.
 Li, L. (2007). Sparse sufficient dimension reduction. Biometrika 94, 603– 613.
 Li, L., Cook, R. D. and Nachtsheim, C. J. (2005). Model-free variable selection. Journal of the Royal Statistical Society Series B, 67, 285–299.
 Li, L., Li, B. and Zhu, L.-X. (2010). Groupwise dimension reduction. Journal of the American Statistical Association. 105, 1188–1201.
 Li, L. and Nachtsheim, C. J. (2006). Sparse sliced inverse regression. Technometrics 48, 503–510.
 Li, L. and Yin, X. (2008). Sliced Inverse Regression with regularizations. Biometrics 64, 124–131.
 Lehmann, I. R. (2013). The 3σ-rule for outlier detection from the viewpoint of geodetic adjustment.
 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.
 Meier, L., Van De Geer, S., & Bühlmann, P. (2008). The group lasso for logistic regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(1), 53-71.
 Ni, L. et al. (2005). A note on shrinkage sliced inverse regression.
Biometrika 92, 242–247.
 Powell, J. et al. (1989). Semiparametric estimation of index coefficients. Econometrica: Journal of the Econometric Society, 1403–1430.
 Rousseeuw, P. and Yohai, V. (1984). Robust regression by means of s- estimators. In Robust and Nonlinear Time Series Analysis, pages 256-272.
 Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), 461–464.

- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B, 58, 267–288.
- Tukey, J. W. (1960). A survey of sampling from contaminated distributions. Contributions to Probability and statistics, 2:448-485.
- Wand, M. P. and Jones, M. C. (1995) Kernel Smoothing. Chapman and Hall, London. http://dx.doi.org/10. 1007/978-1-4899-4493-1.
- 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.
- Wang, Q. and Yao, W. (2013). Robust Variable Selection through MAVE. Computational Statistics and Data Analysis 63, 42-49.
- Wang, T. et al. (2013). Penalized minimum average variance estimation. Statist. Sinica 23 543–569.
- Wang, T. et al. (2015). Variable selection and estimation for semi parametric multiple-index models. Bernoulli 21 (1), 242–275.10
- Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. The Annals of Statistics, 35(6), 2654–2690.
- Xia, Y. (2008). A multiple-index model and dimension reduction. Journal of the American Statistical Association, 103(484), 1631– 1640.
- Xia, Y. et al. (2002). An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society Series B 64, 363–410.
- Yin, X. and Cook, R. D. (2005). Direction estimation in single index regressions. Biometrika, 92(2), 371–384. ˇ
- Yin, X. et al. (2008). Successive direction extraction for estimating the central subspace in a multiple-index regression. Journal of Multivariate Analysis, 99(8), 1733–1757.
- Yu, Z. and Zhu, L. (2013). Dimension reduction and predictor selection in semi parametric models. Biometrika, 100, 641-654.
- Zhang, C. H. (2010). Nearly unbiased variable selection under Minimax Concave Penalty. Annals of Statistics 38, 894–942.
- Zhang, J. and Olive, D. J. (2009). Applications of a robust dispersion estimator. Southern Illinois University Carbondale.
- Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101(476), 1418–1429.
- 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.
- Zou, H., and Zhang, H. (2009). On the adaptive elastic-net with a diverging number of parameters. Annals of Statistics, 37(4), 1733.