Robust variable selection in sliced inverse regression using Tukey's biweight criterion and ball covariance

Ali Alkenani

Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Al Diwaniyah, Iraq.
ali.alkenani@qu.edu.iq

Abstract

The shrinkage sliced inverse (SSIR) is a variable selection method under the settings of sufficient dimension reduction (SDR) theory. The SSIR merges the ideas of Lasso shrinkage and sliced inverse regression (SIR) to obtain sparse and accurate solutions. However, the dependency of SSIR on squared loss function and classical estimates for location and dispersion measures make it very sensitive to outliers. In this paper, a robust variable selection method based on SSIR, which is called RSSIR, is proposed. The squared loss is replaced by Tukey's biweight criterion. Also, the classical estimates of the mean and covariance matrix are replaced with the median and ball covariance, which are robust measures for location and dispersion, respectively. In both the response and covariates, the proposed RSSIR is resistant to outliers. In addition, a robust version of the residual information criterion (RIC) is proposed to select the regularisation parameter. Depending on the results of simulations and real data analysis, very reliable results are achieved through RSSIR. In the presence of outliers, the performance of RSSIR is significantly better than the performance of SSIR and other existing methods.

Keywords: Robust sufficient dimension reduction, Lasso, SIR, Tukey's biweight criterion, Ball correlation.

1. Introduction

In high-dimensional regressions problems, a significant interest to sufficient dimension reduction (SDR) is given. Let $Y$ and $X = (x_1, ..., x_p)^T$ are the response variable and a $p$-dimensional covariates vector, respectively. The idea of SDR is to replace $X$ with $d$-dimensional orthogonal projection $P_X X$ on to $S$, where $d < p$, without loss of information about the conditional distribution of $Y|X$ and without assuming any parametric model. The goal of SDR is how to find the central subspace $(S_{Y|X})$. The $S_{Y|X}$ is the intersection of all subspaces $S$ such as $Y \perp X|P_X X$, where $\perp$
refers to the independency. Consequently, $P_{\beta}X$ extracts the information from $X$ about $Y$, where $\beta$ is a basis of $S_{Y/X}$, see Cook (1998).

There are a number of methods to estimate $S_{Y/X}$. One of the most widely spread SDR methods is SIR (Li, 1991). It is employed in different fields such as bioinformatics, marketing and economics. However, the outcomes of SIR are linear combinations of all the original covariates, which may cause difficulties in interpreting the results of SIR. For this reason, there is a need to decrease the number of nonzero coefficients in SIR directions.

Under least squares settings, a number of regularisation procedures is proposed. For example, the Lasso (Tibshirani, 1996), SCAD (Fan and Li, 2001), Elastic Net (Zou and Hastie, 2005), group Lasso (Yuan and Lin, 2006), adaptive Lasso (Zou, 2006), OSCAR (Bondell and Reich, 2008), MCP (Zhang, 2010) and PACS (Sharma et al., 2013).

Under SIR framework, many procedures are proposed to combine the ideas of SIR and regularisation methods. For example, Cook (2004) proposed a model-free method for determining the variables contribution. Also, Ni et al. (2005) incorporate Lasso with SIR in order to produce shrinkage SIR (SSIR) estimator. Moreover, sparse SIR (SPSIR) is proposed by Li and Nachtsheim (2006) through incorporating Lasso and LARS into SIR. Furthermore, a number of SDR methods are combined with shrinkage estimation ideas by Li (2007). To enable SIR to work under highly correlated covariates and $p > n$ settings where $n$ is the sample size, Li and Yin (2008) proposed a regularised SIR (RSIR) method. For multiple index model and under $p > n$ settings, Lasso-SIR is proposed by Lin et al. (2019).

The above mentioned regularised SIR methods improve the interpret of SIR. However, the squared loss criterion is used between the covariates and the response. Also, the classical estimates of the sample mean and sample covariance matrix of $X$ are employed inside the least square formulation. It is known that the least-squares criterion and the classical estimates of the sample mean and sample covariance matrix of $X$ are not robust and sensitive to outliers.

The above mentioned limitation of the existing regularised SIR methods motivates us to propose a robust version of SSIR which is called RSSIR. Tukey's biweight criterion is used instead of squared loss. Furthermore, the classical estimate of the mean and covariance matrix are replaced with the median and ball covariance, which are robust measures for location and dispersion, respectively. If the derivative of the loss function is redescending, the loss function will be robust and no sensitive to outliers in both $Y$ and $X$ (Rousseeuw and Yohai, 1984). Tukey's biweight loss function has this property (Tukey, 1960). For this reason, the proposed RSSIR is resistant to outliers in both $Y$ and $X$.

The rest of this paper is organised as the following. In Section 2, a short review of SIR and SSIR are given. RSSIR method and robust RIC criteria for selection the tuning parameter are proposed in Sections 3. Simulation studies are implemented in Section 4. In Section 5, the methods under consideration are applied to real data. Finally, the conclusions are referred to in Section 6.
2. SIR and SSIR

For finding $S_{Y|X}$, SIR method is proposed by Li (1991). It requires $Z = \Sigma^{-\frac{1}{2}} (X - E(X))$, under the condition $E(Z|P_Z) = P_Z$, where $\Sigma = Cov(X)$ is the population covariance matrix of $X$ and $s$ is a basis for $S_{Y|Z}$ which is the central subspace of regression $Y$ on $Z$. This condition connects $S_{Y|Z}$ with the inverse regression of $Z$ on $Y$. The SIR's kernel matrix is $M = Cov \{E(Z|Y)\}$ and $\text{Span}(M) \subseteq S_{Y|Z}$.

Let a random sample of size $n$ of $(X, Y)$, which has a joint distribution. Let $\bar{X}$ is the sample mean of $X$ and $\hat{Z} = \Sigma^{-\frac{1}{2}} (X - \bar{X})$ is the sample version of $Z$, where $\Sigma$ is the sample covariance matrix of $X$. Let $h$ is the number of slices and $n_y$ is the number of observations in the $y$th slice. Thus, $\hat{M} = \sum_{y=1}^{h} \hat{f}_y \hat{Z}_y Z^T_y$ is the estimate of $M$, where $\hat{f}_y = n_y/n$ and $\hat{Z}_y$ is the average of $Z$ in the slice $y$. Let $\delta_1 > \delta_2 > \cdots > \delta_p \geq 0$ are the eigenvalues corresponding to the eigenvectors $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_p$ of $\hat{M}$. If the dimension $d$ of $S_{Y|Z}$ is known, $\text{span}(\hat{\beta}) = \text{span}(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_d)$ is a consistent estimator of $S_{Y|X}$, where $\hat{\beta}_i = \hat{\Sigma}^{-\frac{1}{2}} \hat{v}_i$.

The SIR gives an estimator $\text{span}(\hat{\beta})$ of $S_{Y|X}$. Generally, $\hat{\beta} \in \mathbb{R}^{p \times d}$ has nonzero elements. If the covariates number is huge, only the important covariates are employed to construct the ‘sufficient covariates’. Consequently, there is need to merge the variable selection procedures with SIR to set some elements of $\hat{\beta}$ to 0’s.

Cook (2004) formulated SIR as a regression type minimisation problem as follows

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

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

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

where $B \in \mathbb{R}^{p \times d}$. The value of $B$, which minimises (2) is $\hat{\beta}$ and $\text{span}(\hat{\beta}) = \text{span} \left( \hat{\Sigma}^{-\frac{1}{2}} \hat{A} \right)$ is the estimator of $S_{Y|X}$. After that, Ni et al. (2005) proposed SSIR estimator of $S_{Y|X}$ as a $\text{span}(\text{diag}(\hat{\alpha}) \hat{\beta})$, where the shrinkage indices $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_p)^T \in \mathbb{R}^p$ are determined by minimising

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

where, $\hat{B}$ and $\hat{C}$ are minimise (2).

By using a standard Lasso algorithm, the minimisation of (3) can be done. Let $\hat{\bar{Y}} = \text{vec}(\hat{f}_1^{1/2} \hat{Z}_1, \ldots, \hat{f}_h^{1/2} \hat{Z}_h) \in \mathbb{R}^{ph}$ and $\hat{\bar{X}} = \left( \text{diag}(\hat{B} \hat{C}_1) \hat{\Sigma}^{1/2}, \ldots, \text{diag}(\hat{B} \hat{C}_h) \hat{\Sigma}^{1/2} \right)^T \in \mathbb{R}^{ph \times p}$,
where \( \text{vec}(.) \) is a matrix operator that puts the columns of matrix in a single vector. Then the vector \( \alpha \), is the Lasso estimator of regression \( \hat{Y} \) on \( X \).

### 3. Robust SSIR (RSSIR)

#### 3.1 Methodology of RSSIR

Ni et al. (2005) proposed SSIR method as a combination of Lasso and SIR. It is well known that SIR method based on the classical estimates of the first and second moments of the data, which are sensitive to outliers. Gather et al. (2002) study SIR's sensitivity to outliers and propose a robust version of SIR. Also, another robust version of SIR is proposed by Yohai and Sertter (2005). Moreover, Prendergast (2005) studies the influence of function on SIR. Although the good performance of SSIR under normal errors has shown by Ni et al. (2005), SSIR inherits the sensitivity to outliers from its components. The high sensitivity to outliers is the main drawback of SIR where a single outlier can alter the good performance of SSIR estimate totally. This motivates us to propose RSSIR in this paper.

Note that, in (3), the squared loss is used between the covariates and the response. Also, the classical estimates of the sample mean and sample covariance matrix are employed inside the squared loss. It is well known that the least-squares criterion and the classical estimates of the sample mean and sample covariance matrix are sensitive and not robust to outliers (Alkenani and Dikheel, 2017).

If the derivative of the loss function is redescending, it will be robust and insensitive to outliers in both \( Y \) and \( X \) (Rousseeuw and Yohai, 1984). Tukey's biweight loss function has this property (Tukey, 1960). In this paper, the squared loss in (3) is replaced with Tukey's biweight loss function to achieve the robustness against outliers in both \( Y \) and \( X \) and select the informative covariates robustly. Furthermore, the classical estimator of the sample mean is replaced with the median as a robust estimator. Moreover, the classical estimator of sample covariance matrix is replaced with robust covariance matrix estimator such as ball covariance. The estimates of proposed RSSIR can be obtained by minimising the following:

\[
\sum_{y=1}^{h} \rho \left( \frac{\hat{y}_{y}^{1/2} \hat{R} \hat{Z}_{y} - \hat{R} \hat{\sigma}^{\frac{1}{2}} \text{diag}(\hat{\beta}) \alpha}{\hat{\sigma}} \right) + \lambda \sum_{i=1}^{p} |\alpha_i|, \quad (4)
\]

where, \( \lambda \geq 0 \) is the tuning parameter. \( \hat{R} \hat{Z}_{y} \) and \( \hat{R} \hat{\sigma}^{\frac{1}{2}} \) are robust versions of \( \hat{Z}_{y} \) and \( \hat{\sigma}^{\frac{1}{2}} \), respectively. Also, \( \rho \) is Tukey's biweight function and \( \hat{\sigma} \) is a robust estimate of \( \sigma \). In this article, \( \sigma \) is estimated by the median absolute deviation (MAD) of the residuals. Tukey's biweight function is as follows:
\[ \rho_c(u) = \left\{ \begin{array}{ll} \left( \frac{c^2}{6} \right) \left\{ 1 - \left[ 1 - \left( \frac{u}{c} \right)^2 \right] \right\}^3 & \text{if } |u| \leq c \\ \frac{c^2}{6} & \text{if } |u| \leq c \end{array} \right. \] (5)

where \( c \) controls the level of robustness. To achieve 95% asymptotic efficiency at the standard normal distribution, the suggested value of \( c \) is 4.685.

3.2 Robust measures for location and dispersion

SIR method is based on the classical estimates of the first and second moments of the data which are sensitive to outliers. In this article, we employ robust versions of the mean and covariance matrix instead of the classical estimates. The median and ball covariance were used as robust measures for location and dispersion, respectively. Pan et al. (2018) proposed Ball covariance as a robust measure of the dependence between two random vectors as follows:

Let \( \{U_k, V_k\}_{k=1}^n \) be i.i.d. sample of \((U, V)\). Define \( \delta_{ij,k}^U = I[U_k \in \tilde{B}_{\xi V}(U_i, U_j)] \), where \( I(.) \) denotes the indicator function, \( \delta_{ij,k}^V = \delta_{ij,k}^U \delta_{ij,N}^U \) and \( \xi_{ij,kst}^U = (\delta_{ij,k}^U + \delta_{ij,kst}^U - \delta_{ij,ksi}^U - \delta_{ij,js}^U)/2 \). \( \xi_{ij,kst}^V \) is defined in the same way as that of \( \xi_{ij,kst}^U \). The empirical ball covariance is then defined by

\[ \text{BCov}_n(U, V) = \left( \frac{1}{n^6} \sum_{i,j,k,l,s,t=1}^n \xi_{ij,kst}^U \xi_{ij,kst}^V \right)^{1/2} \] (6)

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

3.3. Choosing the tuning parameter \( \lambda \)

Ni et al. (2005) employed generalised cross validation(GCV), Akaike’s information criterion (AIC) (Akaike, 1973), the Bayesian information criterion (BIC) (Schwarz, 1978), and RIC (Shi and Tsai, 2002) for selecting \( \lambda \) according to the following formulas:

\[ \text{GCV} = \frac{\text{RSS}}{n(1-p(\lambda)/n)^2} \] (7)

\[ \text{AIC} = n \log(\text{RSS}/n) + 2p(\lambda), \] (8)

\[ \text{BIC} = n \log(\text{RSS}/n) + \log(n) p(\lambda), \] (9)

\[ \text{RIC} = \{n - p(\lambda)\} \log(\text{RSS}/\{n - p(\lambda)\}) + p(\lambda)(\log(n) - 1) + \frac{4}{\{n - p(\lambda) - 2\}}, \] (10)

where \( \text{RSS} = \sum_{y=1}^b \left\| \hat{\beta}_y^{1/2} \tilde{\epsilon}_y - \hat{\beta}_y^{1/2} \text{diag}(\tilde{\beta}_y) \alpha \right\|_2^2 \) is the residual sum of squares of Lasso fit, and \( p(\lambda) \) is the number of non-zero coefficients.
The simulation results of Ni et al. (2005) show that using RIC for selection \( \lambda \) gives better performance and stable results for SSIR. In this paper, a robust version of RIC, which is called RRIC is proposed as follows:

\[
\text{RRIC} = \{ -\delta \log \left( \frac{1}{n-p(\lambda)} \right) + p(\lambda) \{ \log(n) - 1 \} + \frac{4}{n-p(\lambda)-2} \}.
\]

(11)

where

\[
\text{RRSS} = \sum_{y=1}^{h} \rho \left( \frac{f_{y}^{1/2} \mathcal{R}^{2} - \mathcal{R} \Sigma Z_{y} \mathcal{R} \Sigma Z_{y}}{\sigma} \right)
\]

(12)

\[
\mathcal{R}Z_{y} = \text{BCov}_{n}^{-1/2} (X - \text{median}(X)) \quad \text{and} \quad \mathcal{R}^{2} = \text{BCov}_{n}^{-1/2}
\]

(13)

3.4. Determination of \( \delta \)

For proposed RSSIR, \( d = \text{dim}(S_{Y|X}) \) is assumed as known. Practically, \( d \) is estimated through data. To estimate \( d \), many methods are suggested. See, for example, Li (1991), Schott (1994), Bura and Cook (2001), Cook and Yin (2001) and Zhu et al. (2006). Zhu et al. (2006) suggested to estimate \( \delta \) via the nonzero eigenvalues number of the matrix \( \text{Cov}[E(X|Y)] \), or equivalently, eigenvalues number of \( \Omega = \text{Cov}[E(X|Y)] + I_{p} \) that are greater than one, where \( I_{p} \) is a \( p \)-dimensional identity matrix.

Let \( \delta_{1}, ..., \delta_{p} \) be the eigenvalues of \( \bar{\Omega} \), which is the estimate of \( \Omega \), \( k \) is the number of \( \delta_{i} > 1 \), and \( C_{n}^{*} \) is a penalty constant. Zhu et al. (2006) suggested to estimate \( d \) as follows:

\[
\hat{d} = \arg \max_{m} \left\{ \frac{n}{2} \sum_{i=1}^{p} \left( \log(\delta_{i}) + 1 - \delta_{i} \right) - C_{n}^{*} m (2p - m + 1) \right\}
\]

(14)

Different forms are proposed for \( C_{n}^{*} \) by the researchers. For example, \( C_{n}^{*} = \log(n) \frac{h}{n} \) is proposed by Li and Yin (2008) and they employ it in their simulations.

Alkenani (2020) proposed a robust version of Zhu et al. (2006) estimator of \( d \) in (14) as follows:

For the standardised predictor \( Z \), there is no loss of generality of working in \( Z \)-scale, because of \( S_{Y|X} = \Sigma^{-1/2} S_{Y|Z} \). Alkenani (2020) estimates \( d \) through the eigenvalues number of the robust matrix \( \mathcal{R} \Omega = \mathcal{R} M + I_{p} \) that are greater than one, where \( \mathcal{R} M \) is a robust version of the kernel matrix \( M \) of SIR. The robust sample estimate of \( M \) is

\[
\mathcal{R} M = \sum_{y=1}^{h} \tilde{f}_{y} \mathcal{R} Z_{y} \mathcal{R} Z_{y}^{T}
\]

(15)
Let $\gamma_1, \ldots, \gamma_p$ are the eigenvalues of $R\Omega$ which is the robust sample estimate of $R\Omega$, $k$ is the number of $\gamma_i > 1$. The robust estimator of $d$ is proposed by Alkenani (2020) as follows:

$$
\hat{d} = \arg \max_{m \in \{0, 1, \ldots, p - 1\}} \left\{ \frac{1}{2} \sum_{i=1+\min(k, m)}^{n} (\log(\gamma_i) + 1 - \gamma_i) - \frac{C_n m (2p - m + 1)}{2} \right\}
$$

(16)

In our simulations, we employ formula (16) that is proposed by Alkenani (2020) to estimate $d$.

4. Simulation study

In this section, the proposed RSSIR method is compared with SSIR (Ni et al., 2005) and RSMAVE (Yao and Wang, 2013) through selected examples.

The prediction accuracy for the compared methods is measured through the trace correlation $r^*$ (Zhu and Zeng, 2006). Let $S(A)$ and $S(B)$ are the column space spanned by two $p \times d$ matrices of full column rank. Let $P_A = A(A^T A)^{-1} A^T$ and $P_B = B(B^T B)^{-1} B^T$ are the projection matrices onto $S(A)$ and $S(B)$, respectively. Thus, $r^* = \frac{\frac{1}{d} \text{tr}(P_A P_B)}{\sqrt{d}}$, where, $0 \leq r^* \leq 1$. If $r^*$ is close to 1, and $S(A)$ is close to $S(B)$.

The ability of the methods in terms of variable selection is evaluated via the true and false positive rates TPR and FPR, respectively. TPR denotes the ratio of covariates number correctly identified as active to the number of active covariates. FPR is the ratio of covariates number falsely identified as active to the number of inactive covariates. In terms of variable selection, the best method is the method which has closer TPR to 1 and closer FPR to 0. The optimal $\lambda$ in RSSIR is selected by using the robust version of RIC in (11) and the dimension $d$ is estimated using (16).

In order to check the performance of the proposed method in terms the direction estimation and variable selection, we generate the data from the following model:

$$
y = \frac{\beta_1^T x}{0.5 + (1.5 + \beta_2^T x)^2} + \epsilon, \quad (17)
$$

where,

$\beta_1 = (1, 0, \ldots, 0)^T$, $\beta_2 = (0, 1, 0, \ldots, 0)^T$, and $x \in \mathbb{R}^d$ with $d = 2$. The settings for $x$ are as follow:

(a) $x$ is uncorrelated and the correlation $r_{ij} = 0$ for all values of $i$ and $j$ (b) $x$ is correlated and $r_{ij} = 0.5^{|i-j|}$.

The distributions of $x$ and $\epsilon$ are as follow:

1. $N(0, 1)$, the standard normal.
2. $t_3/\sqrt{3}$, t-distribution with degree of freedom is 3.
3. $0.95 N(0, 1) + 0.05 N(0, 10^2)$.
4. $0.95 \mathcal{N}(0,1) + 0.05\mathcal{U}(-50, 50)$, the standard normal were contaminated with 5% uniform distribution.

The EDR directions are estimated by SSIR (Ni et al., 2005), RSMAVE (Yao and Wang, 2013) and the proposed RSSIR methods. We generate 200 datasets with different sample size $n = 100, 200, 400$. The results of comparison among the methods are reported in Table 1 and 2. The estimation accuracy are assessed via the mean and standard error of $r^*$. Also, the ability of the proposed method in terms of variable selection are checked through TPR and FPR.

Table 1. Results of $r^*$, TPR, and FPR for the considered methods for the four different error distributions in case of the predictors were independent.

| Dist. | $n$ | Criterion | SSIR Mean (s.e) | RSMAVE Mean (s.e) | RSSIR where $\lambda$ is selected via RIC Mean (s.e) | RSSIR where $\lambda$ is selected via RRIC Mean (s.e) |
|-------|-----|-----------|-----------------|-------------------|--------------------------------------------------|--------------------------------------------------|
| Dist.1 | 100 | The trace r Mean (s.e) | 0.940(0.132) | 0.936(0.158) | 0.930(0.160) | 0.932(0.158) |
|       |     | TPR       | 0.939          | 0.916            | 0.913             | 0.914             |
|       |     | FPR       | 0.089          | 0.109            | 0.132             | 0.129             |
|       | 200 | The trace r Mean (s.e) | 0.990(0.083) | 0.989(0.090) | 0.981(0.096) | 0.984(0.092) |
|       |     | TPR       | 0.970          | 0.966            | 0.961             | 0.964             |
|       |     | FPR       | 0.059          | 0.070            | 0.076             | 0.073             |
|       | 400 | The trace r Mean (s.e) | 0.999(0.011) | 0.996(0.012) | 0.990(0.015) | 0.993(0.014) |
|       |     | TPR       | 1.000          | 1.000            | 1.000             | 1.000             |
|       |     | FPR       | 0.029          | 0.035            | 0.044             | 0.040             |
| Dist.2 | 100 | The trace r Mean (s.e) | 0.841(0.156) | 0.900(0.135) | 0.915(0.119) | 0.919(0.117) |
|       |     | TPR       | 0.856          | 0.892            | 0.910             | 0.915             |
|       |     | FPR       | 0.194          | 0.163            | 0.148             | 0.144             |
|       | 200 | The trace r Mean (s.e) | 0.925(0.098) | 0.964(0.095) | 0.978(0.082) | 0.983(0.079) |
|       |     | TPR       | 0.940          | 0.942            | 0.951             | 0.959             |
|       |     | FPR       | 0.110          | 0.098            | 0.080             | 0.077             |
|       | 400 | The trace r Mean (s.e) | 0.947(0.049) | 0.977(0.025) | 0.986(0.017) | 0.989(0.015) |
|       |     | TPR       | 0.960          | 0.996            | 1.000             | 1.000             |
|       |     | FPR       | 0.089          | 0.076            | 0.055             | 0.051             |
| Dist.3 | 100 | The trace r Mean (s.e) | 0.650(0.241) | 0.835(0.159) | 0.886(0.132) | 0.890(0.129) |
|       |     | TPR       | 0.727          | 0.794            | 0.844             | 0.849             |
|       |     | FPR       | 0.412          | 0.168            | 0.140             | 0.136             |
|       | 200 | The trace r Mean (s.e) | 0.678(0.214) | 0.919(0.126) | 0.959(0.107) | 0.964(0.103) |
| Dist.1 | η | Criterion | SSIR | RSMAVE | RSSIR where λ is selected via RIC | RSSIR where λ is selected via RRIC |
|-------|---|-----------|------|--------|-------------------------------|-------------------------------|
| 100   | The trace r Mean (s.e) | 0.928(0.138) | 0.923(0.165) | 0.919(0.167) | 0.921(0.165) |
|       | TPR | 0.932 | 0.911 | 0.908 | 0.910 |
|       | FPR | 0.097 | 0.116 | 0.139 | 0.132 |
| 200   | The trace r Mean (s.e) | 0.979(0.091) | 0.976(0.097) | 0.970(0.101) | 0.973(0.099) |
|       | TPR | 0.964 | 0.966 | 0.961 | 0.964 |
|       | FPR | 0.066 | 0.077 | 0.083 | 0.081 |
| 400   | The trace r Mean (s.e) | 0.987(0.018) | 0.983(0.019) | 0.979(0.021) | 0.981(0.020) |
|       | TPR | 0.994 | 0.995 | 0.994 | 0.995 |
|       | FPR | 0.037 | 0.041 | 0.050 | 0.048 |
| Dist.2 | η | Criterion | SSIR | RSMAVE | RSSIR where λ is selected via RIC | RSSIR where λ is selected via RRIC |
| 100   | The trace r Mean (s.e) | 0.827(0.161) | 0.887(0.141) | 0.902(0.124) | 0.905(0.120) |
|       | TPR | 0.848 | 0.886 | 0.904 | 0.907 |
|       | FPR | 0.202 | 0.170 | 0.155 | 0.151 |
| 200   | The trace r Mean (s.e) | 0.911(0.098) | 0.951(0.102) | 0.965(0.087) | 0.969(0.083) |
|       | TPR | 0.932 | 0.937 | 0.945 | 0.949 |

Table 2. Results of $r^*$, TPR, and FPR for the considered methods for the four different error distributions in case of the predictors were correlated.
From Tables 1 and 2, we have the following observations:

1. In case of Dist.1, the very good behavior of SSIR is clear and its performance exceed the performance of competitors.
2. In case of Dist.2 or Dist.3 or Dist.4, the behavior of SSIR was strongly affected while the stability in the good performance of RSSIR and RSMAVE was clear. Also, the preference is for RSSIR with different sample sizes.
3. Under different settings, the very good performance of RSSIR where $\lambda$ is selected via RRIC is clear. It is stable and better than the performance of RSSIR where $\lambda$ is selected via RIC in terms of estimation accuracy and variable selection.
4. The variations in the trace r, TPR and FPR values for RSSIR estimates are close under different settings. The case is different for SSIR estimates where the variations according to the comparative criteria values are big under the different considered settings.

5. Analysis of real data

In this section, the studied methods are applied to real data. The Boston housing data (Harrison and Rubinfeld, 1978) are analysed by using the studied methods. Y is centered and the covariates are standardised.
SSIR (Ni et al., 2013) identified \(d = 1\) direction with 11 significant predictors. To check the performance of the proposed RSSIR, this data are re-analysed by adding some outliers in Y and X. A percentage of 5% contaminated observations and a single outlier are added to the data. We increased the value Y to \(Y + c\) and X to \(X + c\), the results from \(c = 10\) and \(20\) are reported.

In Table 3, the number of selected variables by SSIR and RSSIR are compared. Also, to check the accuracy of estimation, we reported \(corr(\hat{\beta}, \hat{\beta}_{SSIR})\) which is the correlation between \(\hat{\beta}\) and \(\hat{\beta}_{SSIR,0}\) from SSIR without outliers.

Boston housing data

The data are collected by Harrison and Rubinfeld (1978). The data consist of \(n = 506\) and \(p = 14\); Y is medv, the median value of owner-occupied homes in $1000's. X contains 13 statistical measurements on the 506 census tracts in suburban Boston. Predictors include rate of crime \(x_1\), proportion of residential land zoned \(x_2\), proportion of non-retail business acres \(x_3\), the Charles River \((=1\) if tract bounds river; 0 otherwise) \(x_4\), concentration of nitric oxides \(x_5\), average of rooms \(x_6\), proportion of owner-occupied units \(x_7\), weighted mean of distances \(x_8\), index of accessibility \(x_9\), rate of property tax \(x_{10}\), pupil-teacher ratio \(x_{11}\), proportion of black population \(x_{12}\), and lower status \(x_{13}\). The data set is available publicly from R.

Table 3. Results of number of selected variables and \(corr(\hat{\beta}, \hat{\beta}_{SSIR})\) for SSIR and RSSIR.

| Outliers                   | Number of selected variables(NSV) | \(corr(\hat{\beta}, \hat{\beta}_{SSIR})\) |
|----------------------------|-----------------------------------|------------------------------------------|
|                            | SSIR     | RSSIR | SSIR | RSSIR        |
| No outlier                 | 11       | 11     | 1    | 0.9965       |
| Single outlier \((c = 10)\)| 13       | 11     | 0.9017 | 0.9949   |
| Single outlier \((c = 20)\)| 13       | 11     | 0.8006 | 0.9932   |
| 5% outliers \((c = 10)\)   | 13       | 11     | 0.3359 | 0.9912   |
| 5% outliers \((c = 20)\)   | 13       | 11     | 0.0838 | 0.9885   |

Based on \(corr(\hat{\beta}, \hat{\beta}_{SSIR,0})\) and NSV results which are reported in Table 3, we notice the following:

1. In case of no contamination, it is clear that the behavior of RSSIR method and SSIR method is approximately similar in terms of the correlations and the number of selected variables.
2. In case of contamination, it can be seen that the outliers strongly affected the performance of SSIR especially when the percentage of outlier is 5%. On one side, the sensitivity of SSIR method to outliers is very clear, and the results confirm this behavior. On the other side, the performance of RSSIR method is very good and the method gives very stable results for all contamination cases in terms of the correlations and NSV.

In summary, the results of real data analysis confirm the very good behavior of the proposed RSSIR method.
6. Conclusion

In this article, RSSIR 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 has better performance than SSIR when the outliers exist in $Y$ and $X$ in terms the estimation accuracy and variable selection. Also, the RSSIR gives very close results to SSIR when there are no outliers. Simulations and real data analysis showed that the RSSIR has favorable predictive accuracy. In addition, a robust version of RIC which is called RRIC is proposed to select the tuning parameter $\lambda$ robustly. The simulation results show that RSSIR with selected $\lambda$ via RRIC gives better results than RSSIR with selected $\lambda$ via RIC.

References

1- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In 2nd International Symposium on Information Theory, Ed. B. N. Petrov and F. Csaki, pp. 267–81. Budapest: Akademia Kiado.
2- Alkenani, A. (2020). Robust group identification and variable selection in sliced inverse regression using Tukey's biweight criterion and ball covariance. Gazi University Journal of Science. Accepted.
3- 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.
4- Bondell, H. D. and Reich, B. J. (2008). Simultaneous regression shrinkage, variable selection and clustering of predictors with OSCAR. Biometrics 64, 115-123.
5- Bura, E. and Cook, R. D. (2001). Extending sliced inverse regression: The weighted chi-squared test. Journal of the American Statistical Association 96, 996–1003.
6- Cook, R. (1998). Regression graphics: ideas for studying the regression through graphics. New York, Wily.
7- CooK, R. (2004). Testing predictor contributions in sufficient dimension reduction. Annals of Statistics 32, 1061–92.
8- Cook, R. D. and Yin, X. (2001). Dimension reduction and visualization in discriminant analysis. Australian and New Zealand Journal of Statistics 43, 147–177.
9- 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.
10- Gather, U., Hilker, T. and Becker C. (2002). A note on outlier sensitivity of sliced inverse regression. Statistics 36, 271–281.
11- Harrison, D. and Rubinfeld, D. (1978). Hedonic Housing Prices and the Demand for Clean Air. Journal of Environmental Economics and Management, 5, 81- 102.
12- Li, L. (2007). Sparse sufficient dimension reduction. Biometrika 94, 603–613.
14- Li, K. (1991). Sliced inverse regression for dimension reduction (with discussion). Journal of the American Statistical Association 86, 316–342.
15- Li, L. and Nachtsheim, C. J. (2006). Sparse sliced inverse regression. Technometrics 48, 503–510.
16- Li, L. and Yin, X. (2008). Sliced Inverse Regression with regularizations. Biometrics 64, 124–131.
17- Lin, Q., Zhao, Z. and Liu J. (2019). Sparse sliced inverse regression via lasso. Journal of the American Statistical Association 114, 1726-1739.
18- Ni, L., Cook, R. D. and Tsai, C. L. (2005). A note on shrinkage sliced inverse regression. Biometrika 92, 242–247.
19- Pan, W., Wang, X., Xiao, W., Zhu, H. (2018). A generic sure independence screening procedure. J. Amer. Statist. Assoc. 1-29, (just-accepted).
20- Prendergast, L. (2005). Influence functions for sliced inverse regression. Scand J Stat 32:385–404.
21- Rousseeuw, P. and Yohai, V. (1984). Robust regression by means of s-estimators. In Robust and Nonlinear Time Series Analysis, pages 256-272.
22- Schott, J. R. (1994). Determining the dimensionality in sliced inverse regression. Journal of the American Statistical Association 89, 141–148.
23- Sharma, D. B., Bondell, H. D. and Zhang, H. H. (2013). Consistent group identification and variable selection in regression with correlated predictors. Journal of Computational and Graphical Statistics 22, 319-340.
24- Shi, P. and Tsai, C.L. (2002). Regression model selection-a residual likelihood approach. J. R. Statist. Soc. B 64, 237–52.
25- Shwarz, G. (1978). Estimation the dimension of a model. Ann. Statist. 6, 461–4.
26- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B 58, 267–288.
27- Tukey, J. W. (1960). A survey of sampling from contaminated distributions. Contributions to Probability and Statistics, 2:448-485.
28- Yao, W. and Wang, Q. (2013). Robust variable selection through MAVE. Computational Statistics and Data Analysis, 63, 42-49.
29- Yohai, V. and Sertter, M. (2005) A robust proposal for sliced inverse regression. In: International conference on robust statistics, abstract
30- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society B, 68, 49-67.
31- Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. Annals of Statistics 38, 894–942.
32- Zhang, J and Chen, X. (2019). Robust sufficient dimension reduction via ball covariance. Computational Statistics and Data Analysis 140, 144–154.
33- Zhu, L., Miao, B., and Peng, H. (2006). On sliced inverse regression with large dimensional covariates. Journal of American Statistical Association 101, 630–643.
34- Zhu, Y. and Zeng, P. (2006). Fourier methods for estimating the central subspace and the central mean subspace in regression. Journal of the American Statistical Association 101, 1638–1651.
35- Zou, H. (2006). The adaptive Lasso and its oracle properties. Journal of the American Statistical Association 101, 1418–1429.
36- 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.