Robust Sliced Inverse Regression

Asst. Prof. Tahir Reisan Dikheel (Ph.D.)
Statistics dept.
College of Administration and Economics
University of Al-Qadisiya
tahir_reisan@yahoo.com

الانحدار المعكوس المُجَزّأ الحصين

أ. م. د. طاهر ريسان دخيل
كلية الادارة والاقتصاد / جامعة القادسية

تاريخ استلام البحث: 2013/12/31   تاريخ قبول النشر: 2014/1/29

المستخلص

في هذا البحث تم اقتراح طريقة لجعل مقدرات اتجاهات تخفيض البعد المؤثر (اتجاهات E.D.R.) حصينا في طريقة الانحدار المعكوس المجزأ SIR، من خلال تقدير مصفوفة التباين المشترك التي تعتمد عليها الطريقة بشكل حصين باستخدام طريقتي FCH و RFCH الحصينتين وسميت الطريقتين المقترحتين FCH-SIR و RFCH-SIR ، وقد تم تلويث البيانات المستخدمة بنوعين من القيم الشاذة وهي ACN و SCN وبنسب تلويث واحجام عينات مختلفة . وتم التوصل من خلال تجارب المحاكاة والبيانات الحقيقية الى استنتاجات بينت ان الطريقتين المقترحتين في هذا البحث اعطت نتائج افضل مقارنة بطريقة SIR الاعتيادية اعتمادا على معيار MSE اساسا للمقارنة.

Abstract

In this paper, two methods were suggested to make the estimations of Effective Dimension Reduction directions (E.D.R.-directions) robust in sliced inverse regression (SIR), through the robust estimate of the matrix of covariance, which depends upon the method, by using fast consistent high breakdown (FCH) and reweighted fast consistent high breakdown (RFCH) methods, we called the proposed methods (FCH-SIR) and (RFCH-SIR). Data has been contaminating by two types of outliers values which are asymmetric contamination (ACN) and symmetric contamination (SCN), and different contaminating ratios and sample sizes. Have been reached, through simulation experiments and real data. Conclusions showed that the two proposed methods in this paper gave better results compared to the ordinary SIR depending on the mean square errors (MSE) criterion for comparison.

Keywords: SIR; FCH; RFCH; ACN; SCN; inverse regression; curse of dimensionality.

Introduction

Recent developments in data gathering and storage capacities have resulted in huge amounts of multivariate data being collected at a rapid rate. For such large amounts of data, the problem of the "Curse of Dimensionality" poses a challenge to most statistical methods. The "Curse of Dimensionality" means that the increasing of the sparsity will be exponential given a fixed amount of data points. This problem causes the multiple linear regression methods and other standard statistical methods fail in high dimensional (HD) data. The operation of reducing the number of random variables with as little loss of information as possible is one of the main solutions for the "Curse of Dimensionality". The main two ways to achieve this aim are the subset selection and the feature extraction. The subset selection is the process of selecting a subset of the important variables and the feature extraction is the process of transforming (projecting) the variables into a fewer number of new ones. Improving the performance of the model's prediction, providing faster and lower cost models and giving a good understanding of the dataset are the central aim of subset selection (Guyon and Elisseeff, 2003). Feature extraction shares the objective of subset selection, with the difference that the results must be explained in terms of all of the variables. It denotes the process of finding the transformation that projects the data from the original space to the feature space. Our focus in this article on Feature extraction methods, a vast number of feature extraction techniques have emerged in the literature for reducing the dimensionality, without the loss of as much information as possible from the data. These include principal component analysis (see Jolliffe, 2002; Zhang and Olive, 2009), factor analysis (see Gorsuch, 1983), independent component analysis (Comon, 1994), 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), the minimum average variance estimator (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. It is well known that the majority of the above methods are not robust and sensitive to the outliers. In this article, the robustness of sliced inverse regression method has been studied and two robust versions of sliced inversed regression have been proposed.

The remainder of this article is arranged as follows. In Section 2, a brief review of inverse regression is given. Sliced Inverse Regression is reviewed in Section 3. FCH and RFCH estimators are introduced in Sections 4 and 5, respectively. Simulation studies are conducted in Section 6. Practical studies are showed in section 7. Finally, the conclusions are summarized in Section 8.

1. Inverse regression

In this section, we will discuss the main idea that the inverse regression depends on. It's well known that the ordinary regression deals with the regressing of x's on the response y, in other word $E[y|\mathbf{x}]$. However, if the number of explanatory variables is large, the curse of dimensionality problem appears.

Let we have one dimension x, the properties of regression function can be found from data separation because it will be sufficient. However, when the explanatory variables increase, we get a huge amount of data. This leads to increase the sparsity of the data and the standard statistical methods are breakdown. To tackle this problem, the inverse regression is used to deal with the regressor of $\mathbf{x}$ given y instead of y given $\mathbf{x}$. In another words, the inverse regression deals with $E[\mathbf{x}|y]$ in order to get regression model with one dimension.

Consider the trajectory of the inverse regression curve $E[\mathbf{x}|y]$ as y varies. The center of this curve is located at $E\big[E[\mathbf{x}|y]\big] = E[\mathbf{x}]$. In general, it is a p-dimensional curve in R^P. We shall see that it lies on a K-dimensional subspace, $K<P$, if the following condition is satisfied:

For any b in R^P, the conditional expectation $E[b\mathbf{x}|\beta_1\mathbf{x}, \beta_2\mathbf{x}, \dots, \beta_K\mathbf{x}]$ is linear in $\beta_1\mathbf{x}, \beta_2\mathbf{x}, \dots, \beta_K\mathbf{x}$; that is, for some constants $c_0, c_1, \dots, c_K$,

$$E[b\mathbf{x}|\beta_1\mathbf{x}, \beta_2\mathbf{x}, \dots, \beta_K\mathbf{x}] = c_0 + c_1\beta_1\mathbf{x} + c_2\beta_2\mathbf{x} + \cdots + c_K\beta_K\mathbf{x} \qquad (1)$$

Li (1991) stated that this condition is satisfied when the distribution of x is elliptically symmetric (e.g., the normal distribution).

2. Sliced Inverse Regression (SIR)

The sufficient dimension reduction (SDR) theory (Cook, 1998) has been introduced to minimize the high dimensionality of the predictors, while keeping the regression information and making few assumptions. For regression models, assume y is a scalar response variable and $\mathbf{x} = (\mathrm{x}_1, \dots, \mathrm{x}_p)^T$ is a $p \times 1$ predictor vector. The SDR investigates a $p \times k$ matrix $\mathbf{B}$, k is unknown and assumed to be less than p, such that $y \perp \mathbf{x}|\mathbf{x}^T\mathbf{B}$, where $\perp$ indicates statistical independence. The column space spanned by $\mathbf{B}$ is called the dimension reduction subspace. The intersection of all of the dimension reduction subspaces is called the central subspace if it is a dimension reduction subspace, which is denoted by $S_{y|\mathbf{x}}$. Finding a $S_{y|\mathbf{x}}$ is an essential goal in SDR because $S_{y|\mathbf{x}}$ contains all of the regression information of y, given $\mathbf{x}$. The dimension (k) of $S_{y|\mathbf{x}}$

is called the structural dimension (Yu and Zhu, 2013). Knowledge of $S_{y|\mathbf{x}}$ is beneficial to answer the question, "how does the distribution of $y|\mathbf{x}$ alter with the value of $\mathbf{x}$?". Various approaches have been proposed to estimate $S_{y|\mathbf{x}}$. For example, the Sliced inverse regression method (Li, 1991), sliced average variance estimator method (Cook and Weisberg, 1991), pHd method (Li, 1992), see Cook (1998) for more details.

Sliced inverse regression method (SIR), which proposed by Li (1991), is one of the methods used to reduce the dimension of the explanatory matrix without need to a complicated model fitting process.

The idea is to find a smooth regression function that operates on a variable set of projections. Given a response variable y and a (random) vector $X \in R^p$ of explanatory variables, SIR is based on the model:

$$y = f(\boldsymbol{\beta}_1\mathbf{x}, \boldsymbol{\beta}_2\mathbf{x}, \dots, \boldsymbol{\beta}_k\mathbf{x}, \epsilon), \qquad (2)$$

where $\beta_j, j = 1,2,\dots,k$ are unknown projection vectors, f is an unknown function, and ϵ is the noise random variable, ϵ and X are independent and $E(\epsilon|X) = 0$.

Model (2) describes the situation where the response variable y depends on the p-dimensional random variable x only through a k-dimensional subspace. Thus we can estimate $\beta_j, j = 1,2,\dots,k$ more efficient than if we use all variables.

SIR tries to find this k-dimensional subspace of R^p which under the model (2) carries the essential information of the regression between y and x, by computing the inverse regression (IR) curve. That means instead of looking for E[y|x], we investigate E[x|y], a curve in R^p consisting of p one-dimensional regressions. SIR also focuses on small k, so that nonparametric methods can be applied for the estimation of f. A direct application of nonparametric smoothing to x is for high dimension P generally not possible due to the sparseness of the observations, or the curse of dimensionality.

Li (1991) suggested an algorithm depends on dividing the response variable y into several parts, and then dividing the overall data into parts according to y partition. The mean of these parts are calculated, after that the principle component analysis is applied on it to identify the more important subspace.

The following simple algorithm implements SIR:

a) Standardize X by an affine transformation to get

$$\tilde{x} = \hat{\Sigma}_{xx}^{-\frac{1}{2}}(x_i - \bar{x}), (i = 1,2,\dots,n), \qquad (3)$$

where $\bar{x}$, $\hat{\Sigma}_{xx}$ denote respectively the sample mean and the sample covariance matrix.

b) Divided the range of y into H slices denoted by $I_1, I_2, \dots, I_H$. Let

$$\hat{P}_h = \left(\frac{1}{n}\right)\sum_{i=1}^{n} \delta_h\,(y_i) \qquad (4)$$

be the proportion of $y_i's$ that fall in the slice $I_h,\ h = 1,2,\dots,H$.

c) Within each slice I_h the sample mean vector of the $\tilde{x}_i's$ is computed

$$\hat{m}_h = \left(\frac{1}{n\hat{P}_h}\right)\sum_{y_i \in I_h} \tilde{x}_i \qquad (5).$$

d) Conduct a principle component analysis for the data $\hat{m}_h$, in the following way from the weighted covariance matrix

$$\hat{V} = \sum_{h=1}^{H} \hat{P}_h \hat{m}_h \hat{m}_h' \qquad (6)$$

and then find the eigenvalues and eigenvectors.

e) Denote the eigenvectors associated with the largest k eigenvalues by $\hat{\eta}_j\ , j = 1,2,\dots,K$ and compute

$$\hat{\boldsymbol{\beta}}_j = \hat{\eta}_j \hat{\Sigma}_{xx}^{-1/2} \qquad (7)$$

to transform back to the original scale.

$\hat{\beta}_j$ represents the estimated effective dimension reduction directions (E.D.R.-directions), and the estimated effective dimension reduction space (E.D.R.-space) $(\hat{B})$ is then calculated from $\hat{\beta}_j$.

Gather et. (2001, 2002) show that the SIR is sensitive to certain types of data contamination which may influence the subspace estimate, due to the classical estimators included in the SIR computations. The authors give examples indicate that in the presence of only a small amount of contamination, SIR can return bad estimates. They proposed to substitute the non-robust estimators with robust estimators in order to obtain robust sliced inverse regression.

An apparent procedure to make SIR more robust is to estimate a sample covariance matrix using methods that can account for outliers. There are many estimators for robust multivariate location and dispersion (RMLD). The minimum covariance determinant (MCD) estimator is the fastest estimator of the RMLD that has been shown to be both consistent and having a high breakdown point. It has less complexity from other estimators (see Bernholt and Fischer, 2004). The complexity of the minimum volume ellipsoid (MVE) is far higher and there may be no known method for computing the projection based, constrained M, M-estimate of the scale of the residuals and the M-estimate of the parameters and Stahel–Donoho estimators (Olive and Hawkins, 2010). Since the mentioned estimators are computationally time consuming, these estimators have been replaced by practical estimators which strike a balance between accuracy and computing cost. However, none of the workable estimators have been proved to be consistent and having a high breakdown point. For example, the fast minimum covariance determinant (FMCD) estimator, which is given in (Rousseeuw and Van Driessen, 1999), is used to replace the MCD estimator. The robust multivariate techniques that claim to use the impractical MCD estimator actually use Rousseeuw and Van Driessen (1999) FMCD estimator.

Olive and Hawkins (2010) showed that the FMCD estimator is not a high breakdown estimator. The authors proposed practical consistent, outlier resistant estimators for multivariate location and dispersion. They suggested the fast consistent high breakdown FCH, RFCH and RMVN estimators. In this article, the FCH and reweighted fast consistent high breakdown RFCH estimators are employed to obtain robust sliced inverse regression methods.

3. FCH estimator

The fast consistent high breakdown FCH estimator was suggested in (2010) by Olive and Hawkins. The $\sqrt{n}$ consistent Devlin, Gnanadesikan and Kettenring estimator DGK estimator in (Devlin et al., 1981) and the high breakdown median ball (MB) estimator in (Olive, 2004) is used by the FCH estimator as attractors. The robust estimator uses an attractor that is one of the trial fits. Therefore if the robust estimator draws κ elemental sets and then refines them with concentration, then the κ refined elemental sets are the attractors. Also a location criterion is used by the FCH estimator to choose the attractors.

If DGK location estimator $T_{\kappa,D}$ has a greater Euclidean distance from median MED(X) than 50% of the data, where MED(X) is the coordinate-wise median, then the MB attractor is used by FCH. The smallest determinant of the attractor is used only by the FCH estimator if

$$\left\|T_{\kappa,D} - \mathrm{MED}(X)\right\| \leq \mathrm{MED}\left(D_i\left(\mathrm{MED}(X), I_p\right)\right) \qquad (8),$$

where $D_i\left(\mathrm{MED}(X), I_p\right)$ is the Euclidean distance from $\mathrm{MED}(X)$ and I_p is $p \times p$ identity matrix. Here $\|.\|$ refers to the Euclidean distance.

Let (T_A , C_A) be the attractor that is used, where T_A and C_A are the location and dispersion estimators, respectively. Then, the estimator (T_F , C_F) takes $T_F = T_A$ and

$$C_F = \frac{\mathrm{MED}\left(D_i^2(T_A, C_A)\right)}{\chi^2_{p,0.5}} C_A \qquad (9),$$

where $D_i^2(T_A, C_A)$ is the ith squared sample Mahalanobis distance, which takes the form $D_i^2(T_A, C_A) = \left(x_i - T_A(X)\right)' C_A^{-1}(X)(x_i - T_A(X))$ for each observation, the $\chi^2_{p,0.5}$ is the 0.50th percentile of a chi-squared distribution and F is the FCH estimator. Olive and Hawkins (2010) showed that the FCH estimator is a high breakdown estimator and C_F is non-singular, even with up to nearly 50% outliers.

4. RFCH estimator

Olive and Hawkins (2010) used two standard reweighting steps to produce the reweighted fast consistent high breakdown RFCH estimator. Let $(\hat{\mu}_1 , \tilde{\Sigma}_1)$ be the traditional estimator computed to n_1 cases with $D_i^2(T_{FCH}, C_{FCH}) \leq \chi^2_{p,0.975}$ and let

$$\hat{\Sigma}_1 = \frac{MED\left(D_i^2(\hat{\mu}_1,\tilde{\Sigma}_1)\right)}{\chi^2_{p,0.5}}\tilde{\Sigma}_1 \qquad (10).$$

Then, let $\left(T_{RFCH}, \tilde{\Sigma}_2\right)$ be the traditional estimator computed to the cases with $D_i^2\left(\hat{\mu}_1, \hat{\Sigma}_1\right) \leq \chi^2_{p,0.975}$ and let

$$C_{RFCH} = \frac{MED\left(D_i^2(T_{RFCH},\tilde{\Sigma}_2)\right)}{\chi^2_{p,0.5}}\tilde{\Sigma}_2 \qquad (11),$$

Olive and Hawkins (2010) showed that the RFCH is also a $\sqrt{n}$ consistent estimator.

5. Simulation

In this section, many simulations have been implemented in order to check the behavior of the suggested methods to estimate the E.D.R.-directions. The following methods have been considered:

SIR is the classical sliced inverse regression.

FCH-SIR is robust sliced inverse regression based on robust covariance matrix estimated by fast consistent high breakdown point estimator (FCH).

RFCH-SIR is robust sliced inverse regression based on robust covariance matrix estimated by reweighted fast consistent high breakdown point estimator (RFCH).

The data has been generated from the following sampling distributions:

1) Normal distribution (NOR), $N_p(0, \mathbf{\Sigma})$.
2) Asymmetric contamination (ACN), where 95% and 90% of the observations have been generated from $N_p(0, \mathbf{\Sigma})$ and 5% and 10%

of the observations equals the point $tr(\mathbf{\Sigma})\mathbf{1}^T$ (where $tr(\mathbf{\Sigma})$ is the trace of $\mathbf{\Sigma}$), respectively.

3) Symmetric contamination (SCN), where 95% and 90% of the observations have been generated from $N_p(0,\mathbf{\Sigma})$ and 5% and 10% have been generated from $N_p(10,9\mathbf{\Sigma})$, respectively.

Some typical examples are given below:

Simulate 1: R=2000 datasets were generated with sizes n= (50, 100, 200, 500, 1000) from the model $\mathrm{y} = 0.5\mathrm{x}_1 + 0.5\mathrm{x}_2 + 0.5\mathrm{x}_3 + 0.5\mathrm{x}_4 + 0\mathrm{x}_5 + \epsilon$, where x_i and ϵ are independent and are identically distributed from an $N(0,1)$.

Simulate 2: R=2000 datasets were generated with sizes n= (50, 100, 200, 500, 1000) from the model $\mathrm{y} = 0.5\mathrm{x}_1 + 0.5\mathrm{x}_2 + 0.5\mathrm{x}_3 + 0.5\mathrm{x}_4 + 0\mathrm{x}_5 + \epsilon$, where x_i generated from ACN data and ϵ from an $\mathrm{N}(0,1)$.

Simulate 3: R=2000 datasets were generated with sizes n= (50, 100, 200, 500, 1000) from the model $= 0.5\mathrm{x}_1 + 0.5\mathrm{x}_2 + 0.5\mathrm{x}_3 + 0.5\mathrm{x}_4 + 0\mathrm{x}_5 + \epsilon$, where x_i generated from SCN data and ϵ from an $\mathrm{N}(0,1)$.

To evaluate the precision of the simulation, the mean squared error (MSE) has been computed.

Table1.
E.D.R.-directions (β_j) estimated and their MSE for
data were generated from normal distribution, $N_p(0,\mathbf{\Sigma})$.

		E.D.R-directions			MSE		
		RFCH-SIR	FCH-SIR	SIR	RFCH-SIR	FCH-SIR	SIR
n=50	β_1	0.47	0.43	0.49	0.326	0.918	0.128
	β_2	0.47	0.39	0.49	0.355	0.916	0.118
	β_3	0.47	0.44	0.48	0.327	0.922	0.114
	β_4	0.48	0.44	0.50	0.338	0.897	0.124
	β_5	0.12	0.23	0.08	0.316	0.843	0.096
n=100	β_1	0.50	0.48	0.50	0.088	0.520	0.055
	β_2	0.49	0.47	0.49	0.089	0.532	0.051
	β_3	0.49	0.44	0.50	0.093	0.560	0.052
	β_4	0.49	0.43	0.49	0.086	0.540	0.058
	β_5	0.07	0.18	0.05	0.079	0.448	0.046
n=200	β_1	0.49	0.46	0.49	0.041	0.303	0.026
	β_2	0.50	0.50	0.51	0.038	0.324	0.026
	β_3	0.49	0.48	0.49	0.035	0.298	0.025
	β_4	0.50	0.45	0.50	0.038	0.323	0.024
	β_5	0.04	0.12	0.04	0.029	0.228	0.021
	β_1	0.50	0.48	0.50	0.014	0.140	0.010
	β_2	0.50	0.48	0.50	0.013	0.156	0.010

n=500	β_3	0.50	0.49	0.50	0.013	0.139	0.010
	β_4	0.50	0.51	0.50	0.014	0.129	0.010
	β_5	0.03	0.08	0.02	0.011	0.115	0.008
n=1000	β_1	0.50	0.49	0.50	0.007	0.075	0.005
	β_2	0.50	0.50	0.50	0.007	0.077	0.005
	β_3	0.50	0.50	0.50	0.007	0.069	0.005
	β_4	0.50	0.49	0.50	0.007	0.075	0.005
	β_5	0.02	0.06	0.01	0.006	0.059	0.004

Table 2.

E.D.R.-directions (β_j) estimated and their MSE for data were generated with (ACN), 95% of the observations from $N_p(0, \boldsymbol{\Sigma})$ and 5% of the observations equals the point $tr(\boldsymbol{\Sigma})\mathbf{1}^T$.

		E.D.R.-directions			MSE		
		RFCH-SIR	FCH-SIR	SIR	RFCH-SIR	FCH-SIR	SIR
n=50	β_1	0.44	0.41	0.38	0.032	0.086	0.140
	β_2	0.45	0.41	0.39	0.027	0.078	0.112
	β_3	0.45	0.43	0.39	0.022	0.056	0.122
	β_4	0.45	0.40	0.39	0.028	0.100	0.121
	β_5	0.30	0.31	0.45	0.890	0.969	1.982
n=100	β_1	0.47	0.43	0.41	0.011	0.043	0.074
	β_2	0.47	0.44	0.42	0.008	0.034	0.067
	β_3	0.47	0.43	0.41	0.008	0.046	0.078
	β_4	0.47	0.44	0.41	0.009	0.037	0.086
	β_5	0.27	0.29	0.46	0.734	0.828	2.085
n=200	β_1	0.48	0.46	0.43	0.006	0.020	0.052
	β_2	0.47	0.45	0.42	0.007	0.022	0.059
	β_3	0.48	0.46	0.43	0.005	0.017	0.048
	β_4	0.48	0.46	0.43	0.006	0.018	0.053
	β_5	0.26	0.28	0.47	0.692	0.791	2.178
n=500	β_1	0.48	0.47	0.44	0.004	0.009	0.041
	β_2	0.48	0.47	0.44	0.004	0.010	0.040
	β_3	0.48	0.47	0.44	0.004	0.010	0.042
	β_4	0.48	0.47	0.43	0.005	0.008	0.043
	β_5	0.27	0.28	0.47	0.735	0.786	2.201
n=1000	β_1	0.48	0.47	0.44	0.004	0.007	0.037
	β_2	0.48	0.47	0.44	0.004	0.007	0.039
	β_3	0.48	0.48	0.44	0.004	0.006	0.037
	β_4	0.48	0.48	0.44	0.004	0.006	0.039
	β_5	0.27	0.28	0.47	0.752	0.780	2.204

Table 3.
E.D.R.-directions (β_j) estimated and their MSE for data were generated with (ACN), 90% of the observations from $N_p(0,\mathbf{\Sigma})$ and 10% of the observations equals the point $tr(\mathbf{\Sigma})\mathbf{1}^T$.

		E.D.R.-directions			MSE		
		RFCH-SIR	FCH-SIR	SIR	RFCH-SIR	FCH-SIR	SIR
n=50	β_1	0.44	0.41	0.40	0.041	0.088	0.104
	β_2	0.43	0.41	0.40	0.043	0.087	0.099
	β_3	0.43	0.40	0.39	0.045	0.110	0.112
	β_4	0.44	0.41	0.40	0.042	0.089	0.099
	β_5	0.34	0.36	0.40	1.149	1.329	1.565
n=100	β_1	0.45	0.42	0.42	0.021	0.064	0.066
	β_2	0.45	0.43	0.42	0.023	0.056	0.064
	β_3	0.46	0.43	0.42	0.018	0.052	0.065
	β_4	0.45	0.42	0.42	0.024	0.057	0.061
	β_5	0.34	0.36	0.41	1.172	1.331	1.651
n=200	β_1	0.46	0.44	0.43	0.018	0.031	0.045
	β_2	0.46	0.44	0.44	0.017	0.036	0.038
	β_3	0.46	0.44	0.44	0.014	0.032	0.038
	β_4	0.46	0.44	0.44	0.018	0.037	0.038
	β_5	0.35	0.37	0.42	1.217	1.361	1.728
n=500	β_1	0.46	0.45	0.45	0.014	0.024	0.030
	β_2	0.46	0.45	0.45	0.014	0.022	0.028
	β_3	0.46	0.45	0.44	0.013	0.021	0.031
	β_4	0.46	0.46	0.45	0.013	0.019	0.029
	β_5	0.36	0.37	0.42	1.296	1.353	1.804
n=1000	β_1	0.46	0.46	0.45	0.013	0.017	0.027
	β_2	0.46	0.45	0.45	0.014	0.022	0.025
	β_3	0.46	0.46	0.45	0.013	0.016	0.027
	β_4	0.46	0.46	0.45	0.013	0.014	0.027
	β_5	0.37	0.37	0.43	1.340	1.365	1.839

Table 4

E.D.R.-directions (β_j) estimated and their MSE for data were generated with (SCN), 95% of the observations from $N_p(0,\mathbf{\Sigma})$ and 5% from $N_p(10,9\mathbf{\Sigma})$.

		E.D.R.-directions			MSE		
		RFCH-SIR	FCH-SIR	SIR	RFCH-SIR	FCH-SIR	SIR
n=50	β_1	0.44	0.40	0.37	0.432	0.923	1.003
	β_2	0.44	0.42	0.37	0.425	0.891	0.969
	β_3	0.44	0.40	0.38	0.417	0.849	1.008
	β_4	0.44	0.40	0.35	0.404	0.845	1.027
	β_5	0.31	0.32	0.46	1.327	1.495	2.568
n=100	β_1	0.46	0.43	0.38	0.164	0.644	0.814
	β_2	0.46	0.42	0.36	0.175	0.605	0.836
	β_3	0.46	0.42	0.36	0.173	0.616	0.837
	β_4	0.46	0.42	0.36	0.172	0.621	0.837
	β_5	0.31	0.33	0.52	1.254	1.349	3.013
n=200	β_1	0.46	0.45	0.37	0.082	0.356	0.547
	β_2	0.46	0.44	0.38	0.088	0.371	0.563
	β_3	0.46	0.44	0.38	0.083	0.363	0.570
	β_4	0.46	0.44	0.37	0.083	0.364	0.535
	β_5	0.33	0.35	0.56	1.271	1.340	3.258
n=500	β_1	0.47	0.46	0.39	0.035	0.160	0.294
	β_2	0.47	0.45	0.39	0.037	0.160	0.277
	β_3	0.46	0.46	0.39	0.036	0.162	0.292
	β_4	0.47	0.45	0.39	0.037	0.163	0.286
	β_5	0.34	0.35	0.59	1.248	1.266	3.459
n=1000	β_1	0.47	0.46	0.39	0.023	0.089	0.192
	β_2	0.47	0.46	0.39	0.023	0.094	0.181
	β_3	0.47	0.46	0.39	0.023	0.093	0.183
	β_4	0.47	0.46	0.40	0.023	0.093	0.193
	β_5	0.35	0.35	0.59	1.254	1.276	3.524

Table 5
E.D.R.-directions (β_j) estimated and their MSE for data were generated with (SCN), 90% of the observations from $N_p(0,\mathbf{\Sigma})$ and 10% from $N_p(10,9\mathbf{\Sigma})$.

		E.D.R.-directions			MSE		
		RFCH-SIR	FCH-SIR	SIR	RFCH-SIR	FCH-SIR	SIR
n=50	β_1	0.43	0.39	0.38	0.428	0.864	1.033
	β_2	0.42	0.39	0.38	0.420	0.866	1.064
	β_3	0.43	0.40	0.38	0.441	0.843	1.078
	β_4	0.43	0.40	0.38	0.416	0.882	1.026
	β_5	0.36	0.38	0.39	1.718	1.807	1.972
n=100	β_1	0.44	0.42	0.40	0.195	0.619	0.844
	β_2	0.45	0.41	0.40	0.205	0.648	0.833
	β_3	0.44	0.41	0.41	0.203	0.602	0.833
	β_4	0.44	0.41	0.41	0.196	0.599	0.820
	β_5	0.37	0.38	0.40	1.665	1.735	1.765
n=200	β_1	0.45	0.43	0.43	0.098	0.400	0.503
	β_2	0.45	0.43	0.42	0.092	0.360	0.471
	β_3	0.45	0.43	0.42	0.100	0.378	0.493
	β_4	0.45	0.43	0.43	0.098	0.351	0.513
	β_5	0.37	0.38	0.40	1.639	1.672	1.689
n=500	β_1	0.45	0.44	0.45	0.047	0.170	0.210
	β_2	0.45	0.44	0.44	0.052	0.176	0.232
	β_3	0.45	0.45	0.44	0.047	0.178	0.221
	β_4	0.45	0.44	0.45	0.047	0.166	0.223
	β_5	0.39	0.40	0.41	1.549	1.682	1.689
n=1000	β_1	0.46	0.45	0.46	0.032	0.093	0.119
	β_2	0.45	0.45	0.45	0.032	0.096	0.120
	β_3	0.46	0.45	0.45	0.031	0.096	0.118
	β_4	0.46	0.45	0.45	0.032	0.099	0.116
	β_5	0.38	0.40	0.41	1.509	1.674	1.687

According to the MSE, From Tables 1, 2, 3, 4 and 5, it can be seen that the FCH-SIR and RFCH-SIR show a better performance than the SIR method for the majority of the cases under consideration. The proposed methods produce a lower MSE than SIR method. Moreover, the variations in the FCH-SIR and RFCH-SIR estimates are approximately similar in the majority of cases and are less than the variations in the estimate of the SIR method.

6. Body data

To illustrate the performance of our methods, we consider the body data (Heinz et al., 2003). This study had a total of 507 individuals, 247 men and 260 women, and it is available in the R package “Brq”

(Alhamzawi, 2012). The response variable is the weight in (kg) and there are sixteen predictors. These predictors are the BitrSk: Bitrochanteric diameter (cm) x(1), CheDeSk: Chest depth between spine and sternum at nipple level, mid-expiration (cm) x(2), CheDiSk: Chest diameter at nipple level, mid-expiration (cm)x(3), ElbowSk: Elbow diameter, sum of two elbows (cm) x(4), WristSk: Wrist diameter, sum of two wrists (cm) x(5), KneeSk: Knee diameter, sum of two knees (cm) x(6), AnkleSk: Ankle diameter, sum of two ankles (cm) x(7), ChestGi: Chest girth, nipple line in males and just above breast tissue in females, mid- expiration (cm) x(8), WaistGi: Waist girth, narrowest part of torso below the rib cage, average of contracted and relaxed position (cm) x(9), HipGi: Hip girth at level of bitrochanteric diameter (cm) x(10), ThighGi: Thigh girth below gluteal fold, average of right and left girths (cm) x(11), BicepGi: Bicep girth, flexed, average of right and left girths (cm) x(12), ForeaGi: Forearm girth, extended, palm up, average of right and left girths (cm) x(13), CalfGi: Calf maximum girth, average of right and left girths (cm) x(14), WristGi: Wrist minimum girth, average of right and left girths (cm) x(15) and Height: (cm) x(16). We estimate an E.D.R.-directions of the model between the response weight and the 20 independent variables. Finally, the comparison is mead between the methods SIR, FCH-SIR and RFCH-SIR. The table 6 below gives the results about the 16^{th} E.D.R.-directions and MSE for each method:

Table 6
E.D.R.-directions (β_j) estimated and MSE for body data

E.D.R.-directions	SIR	FCH-SIR	RFCH-SIR
β_1	0.121	0.102	-0.037
β_2	-0.164	-0.352	-0.294
β_3	-0.037	0.061	0.077
β_4	-0.253	0.459	0.119
β_5	0.246	-0.405	-0.688
β_6	-0.533	-0.484	0.323
β_7	-0.225	0.342	0.109
β_8	-0.089	-0.092	0.005
β_9	-0.243	-0.117	-0.293
β_{10}	-0.104	-0.188	-0.181
β_{11}	-0.256	0.192	0.088
β_{12}	-0.191	-0.002	0.110
β_{13}	-0.292	0.142	-0.096
β_{14}	-0.328	-0.125	-0.360
β_{15}	-0.269	0.075	-0.159
β_{16}	-0.230	0.022	-0.012
MSE	421.58	155.15	91.39

We find that the same results in simulation study are extended to practical study. i.e., the FCH-SIR and RFCH-SIR show a better performance than the SIR method.

7. Conclusions

In this article, the FCH-SIR and RFCH-SIR robust methods have been proposed. The effectiveness of the proposed methods is explained via many simulation studies and body data application. From the simulation study and the body data, it can be concluded that the proposed methods perform well in comparison to the SIR method. We believe that the proposed methods would supply helpful robust dimension reduction tools.

References

Alhamzawi, R. (2012). Brq: R package version 1.0.

Alkenani, A. and Yu, K. (2013). A Comparative Study for Robust Canonical Correlation Methods. Journal of Statistical Computation and Simulation 83, 690–718.

Alkenani, A. and Yu, K. (2013). Penalized single-index quantile regression. International Journal of Statistics and Probability 2, 12–30.

Bernholt, T. and Fischer, P. (2004). The Complexity of Computing the MCD-Estimator. Theoretical Computer Science 326, 383–398.

Branco, J. A., Croux, C., Filzmoser, P. and Olivera, M. R. (2005). Robust canonical correlations: a comparative study. Computational Statistics 20, 203–229.

Common, P. (1984). Independent Component Analysis, a new concept?. Signal Processing 36, 287–314.

Cook, R. (1998). Regression graphics: ideas for studying the regression through graphics. New York, Wily.

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

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

Devlin, S. J., Gnanadesikan, R. and Kettenring, J. R. (1981). Robust estimation of dispersion matrices and principal components. Journal of the American Statistical Association 76, 354–362.

FAN, J. and Gijbels, I. (1996). Local Polynomial Modelling and Its Application. Chapman & Hall/CRC.

Gather, U., Hilker, T. and Becker, C. (2001). A Robustified Version of Slice Inverse Regression, in Statistics in Genetics and in the Environmental Sciences. Edt's. L. T. Fernholz, S. Morgenthaler, W. Stahel. Birkh¨auser, Basel, 147-157.

Gather, U., Hilker, T. and Becker, C. (2002). A Note On Outlier Sensitivity of Sliced Inverse Regression. Statistics. 13, 271-281.

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, 1157–1182.

Härdle, W. and Stoker, T. (1989). Investing smooth multiple regression by the method of average derivatives. Journal of the American Statistical Association 84, 986–995.

Heinz, G., Peterson, L.J., Johnson, R.W. and Kerk, C.J. (2003). Exploring Relationships in Body

Dimensions. Journal of Statistics Education , 11.

Hotelling, H. (1936). Relations between two sets of variates. Biometrika 28, 321–377.

Ichimura, H. (1993). Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models. Journal of Econometrics 58, 71–120.

Jolliffe, I. T. (2002). Principal component analysis. 2nd ed. Berlin, Germany: Springer Verlag.

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.

Olive, D. J. and Hawkins, D. M. (2010). Robust Multivariate Location and Dispersion. Available at (www.math.siu.edu/olive/pphbmld.pdf).

Powell, J. L., Stock, J. M. and Stoker, T. M. (1989). Semi-parametric estimation of index coefficients. Econometrica 57, 1403–1430.

Rousseeuw, P. J. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. Technometrics 41, 212–223.

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

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

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.

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

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.

Yu, Z. and Zhu, L. (2013). Dimension reduction and predictor selection in semiparametric models. Biometrika, forthcoming.

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

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

Zhou, J. (2009). Robust dimension reduction based on canonical correlation. Journal of Multivariate Analysis 100, 195–209.