# A Robustified Version of Sliced Inverse Regression

Ursula Gather * Torsten Hilker† Claudia Becker‡

**Abstract**

Sliced Inverse Regression (SIR) (Li, 1991) is a method for dimension reduction in (non-parametric) regression models, based on the idea of using the information contained in the inverse regression curve. In the various steps of the SIR procedure, classical statistical estimators are used. Thus, the resulting method is nonrobust, and an immediate possibility to robustify SIR is to replace the classical estimators by robust ones. This leads to procedures which maintain the clever estimation scheme of the original SIR method but can cope better with outliers in the regressor space. We present such robustified versions of SIR and compare them with the original procedure and among each other under different model assumptions.

*AMS 1991 Subject Classifications*: Primary 62F35; secondary 62H12.
*Key words and phrases*: Dimension reduction; Outliers; Robustness.

## 1 Introduction

In modern statistical practice, we often have to cope with data of very high dimension which have to be analysed. The well-known "curse of dimensionality" has become a real and severe problem in data analysis. While the theoretical concepts of multivariate statistics do not cause any trouble in high dimensions, most of the multivariate methods are not applicable for high-dimensional data, because real life samples are never large enough to fill the high-dimensional space densely enough with observations (Bellman, 1961; Friedman, 1994). For the case of nonparametric regression problems with high-dimensional regressor spaces, Li (1991) proposed a technique for reducing the dimension of the regressor space in a first step in order to make usual nonparametric regression methods applicable to the data. The model considered by Li,

$$Y = f(\boldsymbol{\beta}_1^T \boldsymbol{X}, \ldots, \boldsymbol{\beta}_K^T \boldsymbol{X}, \varepsilon), \tag{1}$$

*Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany, (gather@statistik.uni-dortmund.de).

†Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany, (tumhilker@aol.com).

‡Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany, (cbecker@statistik.uni-dortmund.de).

assumes that the real-valued response $Y$ depends on the explanatory variables $X_1, \ldots, X_d \in I\!R$ only via some number $K$ of linear combinations $\boldsymbol{\beta}_1^T \boldsymbol{X}, \ldots, \boldsymbol{\beta}_K^T \boldsymbol{X}$. If $K \ll d$, a dimension reduction from $d$ to $K$ is achieved. Here, $\boldsymbol{\beta}_i \in I\!R^d$, $i = 1, \ldots, K$, $\boldsymbol{X} = (X_1, \ldots, X_d)^T$, $\boldsymbol{X}, \varepsilon$ are stochastically independent, and $f : I\!R^d \to I\!R$ is an unknown link function.

The unknown directions $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ are called effective dimension reduction (edr) directions. Their span is the so-called edr space $\mathcal{B}$ which has to be estimated from the data. Here, we assume that $\mathcal{B}$ has the smallest dimension of all possible dimension reduction spaces for the data, and that $\mathcal{B}$ equals the central dimension reduction subspace in the sense of Cook (1994, 1996, 1998a,b).

Li (1991) shows that under certain conditions on the distribution of $\boldsymbol{X}$, the centered inverse regression curve $E(\boldsymbol{X}|Y) - E(\boldsymbol{X})$ falls almost surely into a subspace spanned by $\boldsymbol{\Sigma}\boldsymbol{\beta}_1, \ldots, \boldsymbol{\Sigma}\boldsymbol{\beta}_K$, where $\boldsymbol{\Sigma} = Cov(\boldsymbol{X})$. Thus, an estimate of the inverse regression curve can be used to find the subspace of interest. This is done by the sliced inverse regression (SIR) procedure.

Information on the performance of SIR and on modifications of SIR can be found in a wide variety of articles, for example Aragon and Saracco (1997), Bura (1997), Carroll and Li (1992), Chen and Li (1998), Cook (1998a,b), Cook and Nachtsheim (1994), Cook and Weisberg (1991), Ferré (1997, 1998), Härdle and Tsybakov (1991), Hall and Li (1993), Hsing and Carroll (1992), Kötter (1996), Li (1992), Saracco (1997), Schott (1994), Zhu and Fang (1996).

In Section 2, we shortly describe the sliced inverse regression procedure of Li (1991). The estimation scheme of SIR uses estimators which are known to be extremely nonrobust (sample mean and covariance, classical principal component analysis). Hence, some bad points in the data could be really dangerous for the results of SIR. A detailed analysis of the effect of outliers on sliced inverse regression is given in Gather et al. (1999). A straightforward proposal for a robustified SIR-type method is to replace the nonrobust components of SIR by some suitably chosen robust alternatives while maintaining the clever estimation scheme. Such a procedure, which we call here Dimension Adjustment MEthod (DAME) is introduced in Section 3. The final section presents the results of a simulation study comparing the performance of SIR and some versions of DAME.

# 2 Sliced Inverse Regression (SIR)

Under the assumptions of model (1), the SIR procedure for a given dataset $(y_i, \boldsymbol{x}_i^T)$, $i = 1, \ldots, n$, $\boldsymbol{x}_i \in I\!R^d$,$y_i \in I\!R$, consists of the following five steps:

1. Standardization: $\boldsymbol{z}_i = \widehat{\boldsymbol{\Sigma}}^{-1/2}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})$, $i = 1, \ldots, n$, where $\widehat{\boldsymbol{\Sigma}} = \sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T / n$, $\overline{\boldsymbol{x}} = \sum_{i=1}^n \boldsymbol{x}_i / n$.

2. Slicing: Distribute $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ into $H$ slices $S_h$, $h = 1, \ldots, H$, according to the order of the respective values of $y_1, \ldots, y_n$; let $n_h$ be the number of observations in slice $S_h$.

3. Calculating slice means: $\widehat{\boldsymbol{m}}_h = \sum_{S_h} \boldsymbol{z}_i / n_h$, $h = 1, \ldots, H$.

4. (Weighted) principal component analysis for the slice means: $\widehat{\boldsymbol{SIR}} = \sum_{h=1}^{H} n_h \widehat{\boldsymbol{m}}_h \widehat{\boldsymbol{m}}_h^T / n$ with eigenvalues $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_d$ and respective normalized eigenvectors $\widehat{\boldsymbol{\eta}}_1, \ldots, \widehat{\boldsymbol{\eta}}_d$.

5. Estimation of the dimension reducing directions $\boldsymbol{\beta}_k$: $\widehat{\boldsymbol{\beta}}_k = \widehat{\boldsymbol{\Sigma}}^{-1/2} \widehat{\boldsymbol{\eta}}_k$, $k = 1, \ldots, K$.

In the second step of the SIR procedure, the number $H$ of slices is chosen in advance. Up to now, there do not exist general recomendations of how to choose $H$. Chen and Li (1998) mention that the number of slices does not have great influence on the SIR estimates of the dimension reducing directions. In our simulation study (see Section 5), we choose various numbers of slices, and the results indeed do not differ too much for the different choices of $H$. In the last step of SIR, the eigenvectors of $\widehat{\boldsymbol{SIR}}$ corresponding to the $K$ largest eigenvalues yield the estimated directions $\widehat{\boldsymbol{\beta}}_k$, where $K$ is first assumed to be known. We restrict ourselves to this case here, too. The estimation of $K$ is treated in e.g. Cook and Weisberg (1994, pp. 124–125), Ferré (1998), Schott (1994), Sheather and McKean (1997), or Velilla (1998). Its robustness will be investigated separately and is partially under current research. More details and comments on the estimation scheme and the relevance of SIR can be found in Li (1991), Chen and Li (1998), and Cook (1998b).

# 3 SIR-Type Procedures

We introduce here the general concept of a SIR-type procedure, using a location estimator $\boldsymbol{T}_1$ and a positive definite estimator $\boldsymbol{C}_1$ of scatter in the standardization step and a location estimator $\boldsymbol{T}_2$ and a positive semidefinite estimator $\boldsymbol{C}_2$ of scatter in the steps 3 and 4, respectively. Here, $\boldsymbol{C}_2$ is applied to a set $\widehat{M}$ containing the estimated locations $\boldsymbol{T}_2$(slice $h$) from step 3, each with suitably chosen multiplicity $w_h$,$h = 1, \ldots, H$. We call such procedures SIR-type methods. Given a sample $(y_i, \boldsymbol{x}_i^T)^T$, $i = 1, \ldots, n$, of $n$ observations of $(Y, \boldsymbol{X}^T)^T$ we denote the covariance estimator computed in the fourth step of this generalized method by $\boldsymbol{GSIR}((y_i, \boldsymbol{x}_i^T)_n^T)$.

As for many other multivariate statistical procedures,we are interested in obtaining an affine equivariant method.

**Definition 1** *A SIR-type procedure is affine equivariant if for each nonsingular matrix* $\boldsymbol{A} \in \mathbb{R}^{d\times d}$*, for each* $\boldsymbol{v} \in \mathbb{R}^d$*, for each* $a \in \mathbb{R}$*,* $a \neq 0$*, for each* $b \in \mathbb{R}$ *and for each sample* $(y_i, \boldsymbol{x}_i^T)_n^T$ *we have the following property: If* $\widehat{\boldsymbol{\eta}}_i$ *is an eigenvector of* $\boldsymbol{GSIR}((y_i, \boldsymbol{x}_i^T)_n^T)$ *with corresponding eigenvalue* $\widehat{\lambda}_i$*, then*

$$\boldsymbol{C}_{1,n}^{1/2}(\boldsymbol{A}(\boldsymbol{x})_n + \boldsymbol{v})(\boldsymbol{A}')^{-1}\boldsymbol{C}_{1,n}^{-1/2}((\boldsymbol{x})_n)\widehat{\boldsymbol{\eta}}_i$$

*is an eigenvector of* $\boldsymbol{GSIR}((\boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{v}, a y_i + b)_n)$ *with corresponding eigenvalue* $\widehat{\lambda}_i$*,* $i = 1, \ldots, d$*. Here,* $(\boldsymbol{x})_n = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$*.*

It is not necessary that the estimators $\boldsymbol{T}_1, \boldsymbol{C}_1, \boldsymbol{T}_2, \boldsymbol{C}_2$ are all affine equivariant in order to get a SIR-type procedure with this property.

**Lemma 1** *A SIR-type procedure is affine equivariant, if $\boldsymbol{C}_1$ is affine equivariant, if $\boldsymbol{T}_2$ and $\boldsymbol{C}_2$ are orthogonal equivariant and if*

$$n_h = n_{H-h+1} \text{ and } w_h = w_{H-h+1}, \quad h = 1, \ldots, \lfloor H/2 \rfloor,$$

*where $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to $x \in \mathbb{R}$.*

The choice of $\boldsymbol{T}_1$ does not influence the results of the estimation scheme such that, apart from numerical considerations, $\boldsymbol{T}_1$ can be chosen arbitrarily. From the result of Lemma 1, we see that a robustified version of SIR which maintains the affine equivariance of the original method, should consist of a robust affine equivariant covariance estimator in the first step and of robust orthogonal equivariant location and scatter estimates in steps 3 and 4.

# 4 Dimension Adjustment Method (DAME)

As mentioned above, a straightforward approach to robustify SIR is to maintain Li's basic idea, and to replace all classical estimators by robust alternatives with the respective equivariance properties, for example by estimators with high breakdown points as discussed in Becker and Gather (1999a,b) in the context of outlier identification procedures.

The dimension adjustment method (DAME) then proceeds in five steps similar to SIR. The estimators used in the respective steps are described in more detail below.

1. Standardization: $\boldsymbol{z}_i = \boldsymbol{C}_1^{-1/2}(\boldsymbol{x}_i - \boldsymbol{T}_1)$, $i = 1, \ldots, n$, where $\boldsymbol{T}_1, \boldsymbol{C}_1$ are S-estimators of location and covariance.

2. Slicing: Distribute $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ evenly into $H$ slices $S_h$, $h = 1, \ldots, H$, according to the order of the respective values of $y_1, \ldots, y_n$.

3. Calculating slice locations: $\widehat{\boldsymbol{m}}_h = \boldsymbol{T}_{2,h}$, $h = 1, \ldots, H$, where $\boldsymbol{T}_{2,h}$ denotes the $L_1$- or spatial median within the $h$th slice.

4. (Weighted) principal component analysis for the slice locations: Let the set $\widehat{M}$ contain all estimated locations $\widehat{\boldsymbol{m}}_h$;
   $\widehat{\boldsymbol{DAME}} = \boldsymbol{C}_2(\widehat{M})$ with eigenvalues $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_d$ and respective normalized eigenvectors $\widehat{\boldsymbol{\eta}}_1, \ldots, \widehat{\boldsymbol{\eta}}_d$. Here, $\boldsymbol{C}_2$ is a projection pursuit covariance estimator based on the univariate robust scale statistic $RCQ_\alpha$.

5. Estimation of the dimension reducing directions $\boldsymbol{\beta}_k$: $\widehat{\boldsymbol{\beta}}_k = (\boldsymbol{C}_{1,n}^{-1}\widehat{\boldsymbol{\eta}}_k)/\sqrt{\widehat{\boldsymbol{\eta}}_k^T \boldsymbol{C}_{1,n}^{-1}\widehat{\boldsymbol{\eta}}_k}$, $k = 1, \ldots, K$.

In the standardization step of DAME, we take S-estimators as defined in Rousseeuw and Yohai (1984) and Lopuhaä (1989). They are affine equivariant with high finite-sample breakdown points, and still can be calculated successfully for high-dimensional datasets. The $\rho$-function is chosen to be the modified biweight as introduced by Rocke (1996), to achieve an improved performance in high dimensions. The $\rho$-function, which is called translated biweight, thus is given by

$$\rho(d) = \begin{cases} d^2/2 & , \quad 0 \le d \le M, \\ M^2/2 - M^2(M^4 - 5M^2c_0^2 + 15c_0^4)/(30c_0^4) & \\ + d^2\,(1/2 + M^4/(2c_0^4) - M^2/(c_0^2)) & \\ + d^3\,(4M/(3c_0^2) - 4M^3/(3c_0^4)) & \\ + d^4\,(3M^2/(2c_0^4) - 1/(2c_0^2)) & \\ - d^5\,(4M/(5c_0^4)) + d^6/(6c_0^4) & , \quad M \le d \le M + c_0, \\ M^2/2 + c_0(5c_0 + 16M)/30 & , \quad d > M + c_0. \end{cases}$$

The two parameters of this function are chosen to achieve maximal finite-sample breakdown points of the estimators and an asymptotic rejection probability of 0.01. For a detailed discussion see Rocke (1996).

In the third step of the procedure, we need only a rough estimate of the inverse regression curve. We choose the $L_1$-median or spatial median as a location estimator which is orthogonal equivariant and possesses the highest possible finite-sample breakdown point in this class of estimators (Lopuhaä and Rousseeuw, 1991). Small (1990) discusses properties of the $L_1$-median and other variants of multivariate medians. Other generalizations of the univariate median to the multivariate case, which could also be used in this step of the procedure, can be found in work on data depth (Liu, 1990) and on the bagplot (Rousseeuw and Ruts, 1997).

In step 4, we adopt the proposal of Li and Chen (1985), to calculate a robust covariance estimator based on the projection pursuit principle. Take a sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ in $\mathbb{R}^d$ and consider projections $\boldsymbol{a}^T\boldsymbol{x}_1, \ldots, \boldsymbol{a}^T\boldsymbol{x}_n$, where $\boldsymbol{a} \in S(d) = \{\boldsymbol{a} \in \mathbb{R}^d : \boldsymbol{a}^T\boldsymbol{a} = 1\}$. Determine the direction $\boldsymbol{a}_1$ such that a robust univariate scale estimator applied to $\boldsymbol{a}_1^T\boldsymbol{x}_1, \ldots, \boldsymbol{a}_1^T\boldsymbol{x}_n$ is maximized. This direction $\boldsymbol{a}_1$ is called the first projection pursuit principal component. The value of the scale estimator applied to the projected sample gives the corresponding first eigenvalue. From the successively determined principal components and eigenvalues, a covariance estimator can be constructed, where like in the usual principal component analysis, the searching directions $\boldsymbol{a}$ for each new principal component are restricted to those orthogonal to all previously found components. We use two univariate robust scale statistics here. One is the usual MAD, the second one is a modification of a proposal of Rousseeuw and Croux (1993) closely related to the MAD. It is defined by

$$RCQ_\alpha(u_1, \ldots, u_n) := \{|u_i - u_j| : 1 \le i < j \le n\}_{(k)},$$

where $u_1, \ldots, u_n$ denotes a univariate sample, and $k = (\lfloor n\alpha \rfloor + 1)\lfloor n\alpha \rfloor/2$, $\alpha \in [0.5, 1)$. Thus, the estimator equals the $k$-th order statistic of the absolute differences $|u_i - u_j|$, $1 \le i < j \le n$.

Of course, many more versions of robust SIR-type procedures can be suggested. For example, other projection based robust scatter estimates can be used in the fourth step, like the Stahel-Donoho estimator (Donoho, 1982, Stahel, 1981, Tyler, 1994, Gather and Hilker, 1997). In the following section, a simulation study gives a first impression of the performance of some versions of DAME as compared to SIR. Theoretical considerations are needed to supplement these first simulation results. Robustness criteria like breakdown properties or the bias of a SIR-type procedure under outlier situations should be taken into account (see also Davies and Gather (1993), Gather and Becker (1997), Becker and Gather (1999a, b)), since it is well known that a bounded bias under outliers, which is strongly related to a high breakdown point, and efficiency may be contradictory aims. As already discussed in Sheather and McKean (1997), Cook and Weisberg (1994, pp. 124–125), determining properly the dimension $K$ of the edr space can be difficult when there are outliers in the data. In this article, we assume $K$ to be known, but in practical situations, $K$ has to be estimated in an appropriate robust way, too. Research on this problem is currently done.

# 5 Simulations

In order to investigate the performance of DAME, we show here the results of a simulation study. An important aspect concerning the application of robust procedures is the availability of algorithms to compute the estimates. For the above DAME procedure, algorithms for the various steps of the procedure have to be chosen.

For the computation of the S-estimators in the first step of DAME we use the hybrid algorithm described by Rocke and Woodruff (1996). This algorithm is based upon partitioning the data into cells. Within each cell, MCD-estimates (Rousseeuw, 1985) are used as starting points for an iterative computation of the desired estimates. Rocke and Woodruff (1996) demonstrate that partitioning the data leads to improved estimates especially in high-dimensional situations.

In order to compute the $L_1$-median in the third step of DAME, the well-known algorithm of Bedall and Zimmermann (1979) is applied.

Finally, for the projection pursuit estimator in the fourth step of DAME an algorithm proposed by Croux and Ruiz-Gazen (1996) is used. This algorithm limits the search for the $k$-th principal component to some specific directions which are chosen depending on the data points and the $k-1$ principal components which have already been found. Of course, the original optimization problem is not solved exactly in this way, but the algorithm is easy to implement and it seems that the results are sufficiently precise. In our study we use $RCQ_{0.5}$ as well as the MAD as univariate scale estimators.

We compare the SIR-type prodecures under two models of type (1), which will be contaminated:

$$\begin{aligned} Y &= X_1 + 0.1 \cdot \varepsilon, \\ Y &= \frac{X_1}{0.5 + (1.5 + X_2)^2} + 0.1 \cdot \varepsilon. \end{aligned}$$

The first model corresponds to the situation of $K = 1$ with $\boldsymbol{\beta}_1 = (1, 0, \ldots, 0)^T$. In the second model we have $K = 2$ and the edr space is given by span$[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2]$ with $\boldsymbol{\beta}_1 = (1, 0, \ldots, 0)^T$ and $\boldsymbol{\beta}_2 = (0, 1, 0, \ldots, 0)^T$.

In each case we generate $n = 300$ data points. The dimension $d$ equals 10; $X_1, \ldots, X_{10}$ and $\varepsilon$ are i.i.d. random variables from a standard normal distribution.

We contaminate the samples by adding $2\sqrt{\chi^2_{d;0.999}} \cdot \boldsymbol{\beta}_1$ to 30 randomly chosen data points, which means that 10% of the observations in our samples are replaced by large outliers (see also Rocke, 1996).

For the numbers of slices we use the different values $H = 5, 10, 30, 60$.

The results are reported, using as performance criterion $|corr(\boldsymbol{\beta}_1^T \boldsymbol{x}, \widehat{\boldsymbol{\beta}}_1^T \boldsymbol{x})|$, the absolute value of the correlation between $\boldsymbol{\beta}_1^T \boldsymbol{x}$ and the estimated variable $\widehat{\boldsymbol{\beta}}_1^T \boldsymbol{x}$ for the first model. For the two component model we use the canonical correlations between $\boldsymbol{\beta}_1^T \boldsymbol{x}, \boldsymbol{\beta}_2^T \boldsymbol{x}$ and $\widehat{\boldsymbol{\beta}}_1^T \boldsymbol{x}, \widehat{\boldsymbol{\beta}}_2^T \boldsymbol{x}$. This means that in each case we look at the angle between the model edr's and the estimated edr's. The results of 500 replications are visualized graphically using boxplots.

Figure 1 shows the results for the first model.

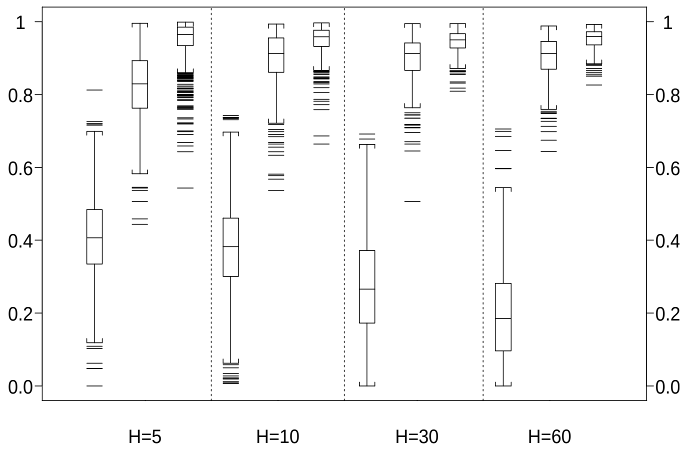


Figure 1: *Simulation results for model 1: Boxplots of correlations between true and estimated edr direction, for SIR and two versions of DAME.*

For each value of $H$, three boxplots are shown. The first one presents the results for SIR, the second and the third boxplot show the results for DAME using the MAD and $RCQ_{0.5}$ as univariate scale estimators in the fourth step, respectively.

First of all, we note that the results of SIR are rather bad in the simulated outlier situations. Due to the outliers in the data, SIR is not able to give good estimates of the edr direction $\boldsymbol{\beta}_1$. In contrast to this, DAME is more successful. Especially with the use of $RCQ_{0.5}$ as univariate scale estimator in the fourth step we get good results.

For the second model, the results are very similar which is reflected by Figure 2.

Here, the first and the second boxplot show the results of SIR, the third and fourth boxplot represent DAME using the MAD and the last two boxplots give the results for DAME using $RCQ_{0.5}$. We see that the first canonical correlation achieves high values for the dimension reduction based upon classical SIR. This means that "a part of the edr space" is estimated very well by SIR and that the outliers do not influence this. But on the other hand, the second canonical correlation gives low values, thus the SIR procedure does not find

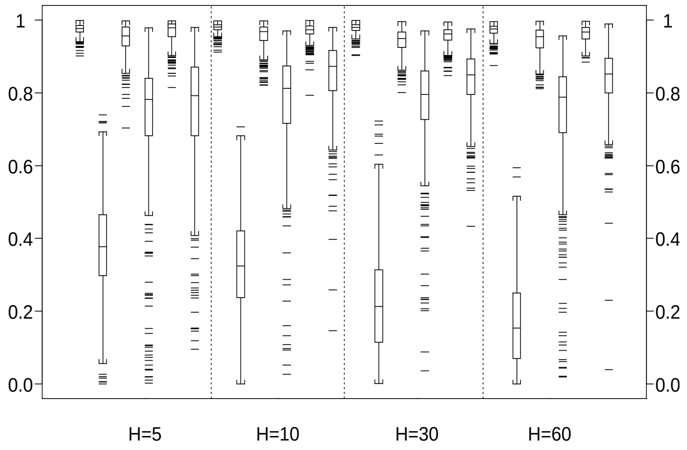


Figure 2: *Simulation results for model 2: Boxplots of canonical correlations between true and estimated edr directions, for SIR and two versions of DAME.*

the complete edr space. The robust DAME procedures again yield better results. Moreover, the use of $RCQ_{0.5}$ instead of the MAD seems to lead to improved results in most cases.

Summing up it may be said that the effects of outliers on SIR turned out dramatic in our simulations. The results of SIR can become totally unreliable if there are a few bad outliers in the data. Robustified SIR-methods, such as DAME proposed above, can be used to overcome these problems.

# 6 Acknowledgements

The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of Complexity in Multivariate Data Structures") is gratefully acknowledged.

# References

[1] Aragon, Y., Saracco, J. (1997). Sliced Inverse Regression (SIR): An Appraisal of Small Sample Alternatives to Slicing. *Comput. Statist.* **12**, 109–130.

[2] Becker, C., Gather, U. (1999a). The Masking Breakdown Point of Multivariate Outlier Identification Rules. *J. Amer. Statist. Assoc.* **94**, 947–955.

[3] Becker, C., Gather, U. (1999b). The Largest Nonidentifiable Outlier: A Comparison of Multivariate Simultaneous Outlier Identification Rules. *Submitted for Publication.*

[4] Bedall, F. K., Zimmermann, H. (1979). Algorithm AS 143: The Mediancentre. *Appl. Statist.* **23**, 325–328.

[5] Bellman, R.E. (1961). *Adaptive Control Processes.* Princeton University Press.

[6] Bura, E. (1997). Dimension Reduction via Parametric Inverse Regression. *$L_1$-Statistical Procedures and Related Topics*, IMS Lecture Notes **31**, 215–228.

[7] Carroll, R. J., Li, K.-C. (1992). Measurement Error Regression With Unknown Link: Dimension Reduction and Data Visualization. *J. Amer. Statist. Assoc.* **87**, 1040–1050.

[8] Chen, C.-H., Li, K.-C. (1998). Can SIR be as Popular as Multiple Linear Regression? *Statist. Sinica* **8**, 289–316.

[9] Cook, R.D. (1994). On the Interpretation of Regression Plots. *J. Amer. Statist. Assoc.*, **89**, 177–189.

[10] Cook, R.D. (1996). Graphics for Regressions With a Binary Response. *J. Amer. Statist. Assoc.*, **91**, 983–992.

[11] Cook, R.D. (1998a). Principal Hessian Directions Revisited. *J. Amer. Statist. Assoc.*, **93**, 84–100.

[12] Cook, R.D. (1998b), *Regression Graphics: Ideas for Studying Regressions through Graphics*, Wiley, New York.

[13] Cook, R. D., Nachtsheim, C. J. (1994). Reweighting to Achieve Elliptically Contoured Covariates in Regression. *J. Amer. Statist. Assoc.* **89**, 592–599.

[14] Cook, R. D., Weisberg, S. (1991). Comment on "Sliced Inverse Regression for Dimension Reduction" by K.C. Li. *J. Amer. Statist. Assoc.* **86**, 328–332.

[15] Cook, R. D., Weisberg, S. (1994). *An Introduction to Regression Graphics.* Wiley, New York.

[16] Croux, C., Ruiz-Gazen, A. (1996). A Fast Algorithm for Robust Principal Components Based on Projection Pursuit. In: A. Prat (ed.), *Compstat, Proceedings in Computational Statistics*, 12th Symposium, Barcelona 1996, Physika–Verlag, Heidelberg, 211–216.

[17] Davies, P.L., Gather, U. (1993). The Identification of Multiple Outliers. Invited Paper With Discussion and Rejoinder. *J. Amer. Statist. Assoc.* **88**, 782–801.

[18] Donoho, D.L. (1982). Breakdown Properties of Multivariate Location Estimators. *Ph.D. Qualifying Paper*, Department of Statistics, Harvard University.

[19] Ferré, L. (1997). Dimension Choice for Sliced Inverse Regression Based on Ranks. *Student* **2**, 95–108.

[20] Ferré, L. (1998). Determining the Dimension in Sliced Inverse Regression and Related Methods. *J. Amer. Statist. Assoc.* **93**, 132–140.

[21] Friedman, J.H. (1994). An Overview of Predictive Learning and Function Approximation. In: Cherkassky, V., Friedman, J.H., Wechsler, H. (eds.), *From Statistics to Neural Networks. Theory and Pattern Recognition Applications*, Springer, Berlin, 1–61.

[22] Gather, U., Becker, C. (1997). Outlier Identification and Robust Methods. In: Maddala, G.S., Rao, C.R. (eds.), *Handbook of Statistics, Vol. 15: Robust Inference*,Elsevier, Amsterdam, 123–143.

[23] Gather, U., Hilker, T. (1997). A Note on Tyler's Modification of the MAD for the Stahel-Donoho Estimator. *Ann. Statist.* **25**, 2024–2026.

[24] Gather, U., Hilker, T., Becker, C. (1999). A Note on Outlier Sensitivity of Sliced Inverse Regression. *Preprint*, Department of Statistics, University of Dortmund.

[25] Härdle, W., Tsybakov, A. B. (1991). Comment on "Sliced Inverse Regression for Dimension Reduction" by K.-C. Li. *J. Amer. Statist. Assoc.* **86**, 333–335.

[26] Härdle, W., Klinke, S., Turlach, B. A. (1995). *XploRe: An Interactive Computing Environment.* Springer, New York.

[27] Hall, P., Li, K.-C. (1993). On Almost Linearity of Low Dimensional Projections from High Dimensional Data. *Ann. Statist.* **21**, 867–889.

[28] Hsing, T., Carroll, R. J. (1992). An Asymptotic Theory for Sliced Inverse Regression. *Ann. Statist.* **20**, 1040–1061.

[29] Kötter, T. (1996). An Asymptotic Result for Sliced Inverse Regression. *Comput. Statist.* **11**, 113–136.

[30] Li, K.-C. (1991). Sliced Inverse Regression for Dimension Reduction (with discussion). *J. Amer. Statist. Assoc.* **86**, 316–342.

[31] Li, K.-C. (1992). On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma. *J. Amer. Statist. Assoc.* **87**, 1025–1039.

[32] Li, G., Chen, Z. (1985). Projection–Pursuit Approach to Robust Dispersion Matrices and Principal Components: Primary Theory and Monte Carlo. *J. Amer. Statist. Assoc.* **80**, 759–766. Correction: *J. Amer. Statist. Assoc.* **80**, 1084.

[33] Liu, R. Y. (1990). On a Notion of Data Depth Based on Random Simplices. *Ann. Statist.* **18**, 405–414.

[34] Lopuhaä, H. P. (1989). On the Relation between S–Estimators and M–Estimators of Multivariate Location and Covariance. *Ann. Statist.* **17**, 1662–1683.

[35] Lopuhaä, H. P., Rousseeuw, P. J. (1991). Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices. *Ann. Statist.* **19**, 229–248.

[36] Rocke, D. M. (1996). Robustness Properties of S–Estimators of Multivariate Location and Shape in High Dimension. *Ann. Statist.* **24**, 1327–1345.

[37] Rocke, D. M., Woodruff, D. L. (1996). Identification of Outliers in Multivariate Data. *J. Amer. Statist. Assoc.* **91**, 1047–1061.

[38] Rousseeuw, P.J. (1985). Multivariate Estimation with High Breakdown Point. In: W. Grossmann, G. Pflug, I. Vincze and W. Wertz (eds.), *Mathematical Statistics and Applications*, Reidel, Dordrecht, 283–297.

[39] Rousseeuw, P. J., Croux, C. (1993). Alternatives to the Median Absolute Deviation. *J. Amer. Statist. Assoc.* **88**, 1273–1283.

[40] Rousseeuw, P. J., Ruts, I. (1999). Boxplot, Bivariate. In: S. Kotz, C.R. Read and D.L. Banks (eds.), *Encyclopedia of Statistical Sciences, Update Volume 3*, 55–60.

[41] Rousseeuw, P. J., Yohai, V. (1984). Robust Regression by Means of S-Estimators. In: *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statistics, Volume 26*, Springer, New York, 256–272.

[42] Saracco, J. (1997). An Asymptotic Theory for Sliced Inverse Regression. *Comm. Statist. – Theory Methods*, **26**, 2141–2171.

[43] Schott, J. R. (1994). Determining the Dimensionality in Sliced Inverse Regression. *J. Amer. Statist. Assoc.* **89**, 141–148.

[44] Sheather, S. J., McKean, J. W. (1997). A Comparison of Procedures based on Inverse Regression. *$L_1$-Statistical Procedures and Related Topics*, IMS Lecture Notes **31**, 271–278.

[45] Small, C. G. (1990). A Survey of Multidimensional Medians. *Internat. Statist. Review* **58**, 263–277.

[46] Stahel, W.A. (1981). Breakdown of Covariance Estimators. *Research Report 31*, Fachgruppe für Statistik, ETH Zürich.

[47] Tyler, D.E. (1994). Finite Sample Breakdown Points of Projection Based Multivariate Location and Scatter Statistics. *Ann. Statist.* **22**, 1024–1044.

[48] Velilla, S. (1998). Assessing the Number of Linear Components in a General Regression Problem. *J. Amer. Statist. Assoc.* **93**, 1088–1098.

[49] Zhu, L.-X., Fang, K.-T. (1996). Asymptotics for Kernel Estimate of Sliced Inverse Regression. *Ann. Statist.* **24**, 1053–1068.