Application of the Bootstrap Approach to the Choice of Dimension and the alpha Parameter in the SIRa Method alpha

Benoit Liquet, Jérôme Saracco

Application of the bootstrap approach to the choice of dimension and the α parameter in the SIR$_\alpha$ method

Benoît Liquet[1] and Jérôme Saracco[2,3]

[1] INSERM U875, ISPED

Université Victor Segalen Bordeaux 2

146 rue Leo Saignat, 33076 Bordeaux Cedex, France

e-mail: benoit.liquet@isped.u-bordeaux2.fr

[2] GREThA, UMR CNRS 5113

Université Montesquieu - Bordeaux IV

Avenue Léon Duguit, 33608 Pessac Cedex, France

e-mail: jerome.saracco@u-bordeaux4.fr

Université Bordeaux 1

[3] Institut de Mathématiques de Bordeaux, UMR CNRS 5251

351 cours de la libération, 33405 Talence Cedex, France

e-mail: jerome.saracco@math.u-bordeaux1.fr

ABSTRACT

To reduce the dimensionality of regression problems, sliced inverse regression approaches make it possible to determine linear combinations of a set of explanatory variables $\mathbf{X}$ related to the response variable Y in general semiparametric regression context. From a practical point of view, the determination of a suitable dimension (number of the linear combination of $\mathbf{X}$) is important. In the literature, statistical tests based on the nullity of some eigenvalues have been proposed. Another approach is to consider the quality of the estimation of the effective dimension reduction (EDR) space. The square trace correlation between the true EDR space and its estimate can be used as goodness of estimation. In this paper, we focus on the SIR$_\alpha$ method and propose a naïve bootstrap estimation of the square trace correlation criterion. Moreover, this criterion could also select the α parameter in the SIR$_\alpha$ method. We indicate how it can be used in practice. A simulation study is performed to illustrate the behaviour of this approach.

Keywords: Bootstrap; Dimension Reduction; Sliced Inverse Regression.

1 Introduction

Conventional parametric or nonparametric methods of curve fitting estimate the unknown relationship between an explanatory variable and a response variable. These methods work well when the dimensionality of the data is low. In high dimensions, all data are sparse. Thus the number of observations available to give information about the local behaviour of the regression function becomes very small with large dimensions. Moreover, the complexity of the possible underlying structure increases more than exponentially fast as the dimension increases, so enormous numbers of data are needed in order to decisively select one parametric model over another. This is the so-called *curse of dimensionality* which quickly defeats parametric or nonparametric regression analysis for fitting usefully predictive models. In the statistical literature, various strategies exist to challenge its effect. One of them is to consider dimension reduction models.

In the following, we consider a high-dimensional data set with one response variable Y and many predictor variables $X_1, \ldots, X_p$. Let $\mathbf{X} = (X_1, \ldots, X_p)^T$ denote the p-dimensional column vector of explanatory variables. Parametric approaches such as linear and polynomial regression, or more generally, transformation or generalized linear models, are examples of dimension-reduction methods. Other methods aim to relieve the problem of the curse of dimensionality without requiring assumptions as strong as those of the parametric ones. Such methods are nonparametric in nature with a parametric part and are called semiparametric. Many dimension-reduction tools are available. There are essentially based on two families of dimension-reduction models. In the first, the effects of the p predictor variables are assumed to be additive: the generalized additive models belong to this family, see for instance Hastie and Tibshirani (1986, 1990), Stone (1985, 1986) or Friedman and Silverman (1989). Among the well-known methods, there is the ACE (Alternating Conditional Expectation) introduced in the Brieman and Friedman (1985) algorithm and the MARS (Multivariate Adaptative Regression Splines) method described by Friedman (1991). The second family assumes that the effects of the explanatory variable $\mathbf{X}$ can be captured in a K (with $K \leq p$) dimensional projection subspace, $(\mathbf{X}^T\beta_1, \ldots, \mathbf{X}^T\beta_K)$ where $\beta_1, \ldots, \beta_K \in \Re^p$. The popular PPR (Projection Pursuit Regression) approach belongs to those projection-based models. It was introduced by Friedman and Stuetzle (1981) and has been discussed by many authors, see for instance Chen (1991) or Hall (1989) among others. However, most of the dimension reduction methods mentioned so far involve a backfitting (iterative) algorithm which needs a smoothing

curve-fitting at each step. These methods generally are computing-intensive and may not converge. Several methods which make it possible to estimate the directions β_k first, without a smoothing step of the link functions, have been proposed by Duan and Li (1991) and Li (1991, 1992). These are the SIR (Sliced Inverse Regression) and PHD (Principal Hessian Direction) methods. They are computationally simple, since they require only simple averaging and an eigenvalue decomposition. In this paper, we will concentrate on the so-called SIR methods and more particulary on the SIR_α method.

The dimension-reduction assumption is the following: $Y \perp X|\mathbf{X}^T\beta_1, \ldots, \mathbf{X}^T\beta_K$ where the notation $U \perp V|W$ means that the random variable U and V are independent conditionally to the random variable W. The underlying semiparametric regression model can be written as :

$$Y = f(\mathbf{X}^T\beta_1, \ldots, \mathbf{X}^T\beta_K, \epsilon), \tag{1}$$

where the response variable Y is associated with the p-dimensional regressor $\mathbf{X}$ only through the linear combinations $\mathbf{X}^T\beta_k$ and ϵ is a random error term independent of $\mathbf{X}$. No assumption is made about the functional form of the unknown link function f or the distribution of ϵ. Since no structural conditions on f are imposed, the vectors β_k are not identifiable, unlike the linear subspace spanned by the β_k's. Li (1991) used this model to introduce the notion of EDR (Effective Dimension Reduction) space, namely the linear subspace E of $\mathbb{R}^p$ spanned by the β_k's. Any vector in E is called an EDR direction. This semiparametric regression model appears to be a reasonable compromise between fully parametric and fully nonparametric modeling. Two estimation problems in (1) clearly arise. The first is to recover the EDR space. The second consists in estimating the link function f. Since $\{(\mathbf{X}^T\beta_1, \ldots, \mathbf{X}^T\beta_K) : \mathbf{X} \in \mathbb{R}^p\}$ theoretically captures all the information on the distribution of Y given X and can be estimated first, one can then easily apply standard nonparametric smoothing techniques (such as kernel, spline or wavelet regression) to estimate the link function f. The goal of dimension reduction is thus achieved.

In this paper our main objective concerns dimension K of the EDR space. Obviously, in most applications, dimension K is unknown and hence must be estimated from the data $\{(\mathbf{X}_i, Y_i),\ i = 1, \ldots, n\}$ with $p < n$. Several approaches have been proposed in the literature. For the original approach introduced by Li (1991), named SIR-I in the following, and for SIR-II (see Scott, 1994), the choice of dimension K is based on statistical nested test procedures. However, from a practical point of view, such test procedures present two main drawbacks. First, the validity of these test

approaches rely on normality distribution or elliptically symmetric distribution assumptions for $\mathbf{X}$. Secondly, the nested test procedures do not allow to control the overall level since knowledge of the level at each step does not imply knowledge of the overall level because of the problem of multiplicity of tests. Note that Bai and He (2004) studied the limiting distribution of the test statistic for dimensionality without the normality assumption, and obtained a necessary and sufficient condition for the chi-square limiting distribution to hold. However, this approach focused on a variant of SIR-I, called CANCOR and introduced by Fung et al. (2002).

There is an alternative to the test approach: this kind of approach is based on the trace criterion and was firstly developed by Ferré (1997,1998) in the Sliced Inverse Regression and Principal Hessian direction context. In this paper, as in Ferré (1998), we prefer a model selection approach which does not require the estimation of the link function f (depending itself on the selected dimension). Thus, while a criterion based on the prediction of Y would be optimal, we only focus on the projections to choose the dimension on the EDR space, without including the estimation of f.

We consider a sample $s = \{(\mathbf{X}_i, Y_i), i = 1, \ldots, n\}$ from the model (1). Let $\Sigma = \mathbb{V}(\mathbf{X})$ be the covariance matrix of $\mathbf{X}$, and let $\widehat{\Sigma}$ be the empirical covariance matrix of the $\mathbf{X}_i$'s. Let $B = [\beta_1, \ldots, \beta_K]$ denote the $p \times K$ matrix of the true β_k's assumed to be linearly independent. Clearly, when the true dimension is K, we have $\mathrm{Span}(B) = E$. Let $\widehat{B} = [\hat{b}_1, \ldots, \hat{b}_K]$ denote the corresponding estimated EDR directions matrix. Let us define $\widehat{E} = \mathrm{Span}(\widehat{B})$ the estimated EDR space. To study the closeness between two subspaces, the square trace correlation can be naturally used. The corresponding risk function is defined by considering the following expectation:

$$R_k = \mathbb{E}\left[\mathrm{Trace}(P_k \widehat{P}_k)\right]/k, \tag{2}$$

where P_k denotes the Σ-orthogonal projector onto the space spanned by the first k vector β_l of B and $\widehat{P}_k$ is the $\widehat{\Sigma}$-orthogonal projector onto the space spanned the first k vector $\hat{b}_l$ of $\widehat{B}$. More precisely, let $B_k = [\beta_1, \ldots, \beta_k]$ and $\widehat{B}_k = [\hat{b}_1, \ldots, \hat{b}_k]$, then $P_k = B_k(B_k^T \Sigma B_k)^{-1} B_k^T \Sigma$ and $\widehat{P}_k = \widehat{B}_k(\widehat{B}_k^T \widehat{\Sigma} \widehat{B}_k)^{-1} \widehat{B}_k^T \widehat{\Sigma}$. Note that R_k is defined for any dimension k lower than or equal to K.

A value of R_k close to one indicates that the set of the k estimated linear combinations of $\mathbf{X}$ is close to the ideal set. So in terms of dimensionality, k is a feasible solution. On the other hand, a value of R_k perceptibly different from 1 means that this estimated set is slightly different from the ideal one, so the solution for the dimension is greater than k. Since R_K will converge to one as n tends to infinity (for the true dimension K), then, for a fixed n, a reasonable way to

assess whether an EDR direction is available is given by looking at how much R_k departs from one. From a computational point of view, we require consistent estimates $\widehat{R}_k$ of R_k, so the feasible solution for the dimension can be obtained by computing the values of $\widehat{R}_k$ for $k = 1$ to p and observing how much it departs from one. In Ferré (1997, 1998), consistent estimators of R_k were proposed. They were obtained from asymptotic expansions of the loss function R_k which require the normality or the elliptically symmetric distribution of $\mathbf{X}$. Moreover, the estimate is not easily computable.To avoid complex computing, Ferré (1997) suggested considering the SIR approach of Hsing and Carroll (1992) based on slices of cardinal 2 and to use a Jack-Knife (JK) method to estimate R_k in this special case: the corresponding estimated loss function measuring how far the estimated subspace obtained by deleting two observations (that is a slice) differs from the sample estimated over the whole sample. However, Ferré mentioned that the Hsing and Carroll estimators, which use slices with two observations, are unable to capture some relevant information, see also Aragon and Saracco (1997) for a simulation study exhibiting the undersmoothing effect of these estimators. Hence, the efficiency of the corresponding JK criterion to choose dimension K is not really clear from a practical point of view.

In this paper, we use this criterion to determine a suitable dimension K using the SIR$_\alpha$ approach. The SIR$_\alpha$ method (described in the next section) depends on a parameter α which controls the combination of the SIR-I and the SIR-II methods. Therefore, the SIR$_\alpha$ approach will be used to define $\widehat{P}_k$, which depends on the choice of α and will be henceforth denoted $\widehat{P}_{k,\alpha}$. The practical choice of α can be based on the test approach proposed by Saracco (2001) which does not require estimation of the link function. Two cross-validation criteria were also developed by Gannoun and Saracco (2003b) to select the parameter α, but these criteria require kernel smoothing estimation of the link function.

To determine the suitable α parameter and the suitable dimension K, we use the criterion defined in (2), which now becomes $R_{k,\alpha}$. Criterion $R_{k,\alpha}$ is precisely defined in (5). To compute estimates for this risk function, computer intensive methods can be used. Note that $R_{k,\alpha}$ is symmetrically defined since its value is invariant under any permutation of the observations $(\mathbf{X_i}, Y_i)$ of s. We propose here to use a well-known resampling method, the bootstrap, which is a natural candidate for estimating $R_{k,\alpha}$. In Section 2, we give an overview of the SIR$_\alpha$ approach. The bootstrap estimate of $R_{k,\alpha}$ is described in Section 3. Section 4 is devoted to a simulation study which was conducted in order to show the efficiency of the proposed approach. A comparison with existing

methods is also provided in Section 4.4. Finally, some concluding remarks are given in Section 5.

2 Overview of the SIR$_\alpha$ approach

We give an overview of the univariate SIR approaches (that is when $Y \in \Re$). While there are several possible variations, the basic principle of SIR methods (SIR-I, SIR-II or SIR$_\alpha$) is to reverse the role of Y and $\mathbf{X}$. Instead of regressing the univariate response variable Y on the multivariate covariable $\mathbf{X}$, the explanatory variable $\mathbf{X}$ is regressed on the dependent variable Y. The SIR-I estimates, based on the first moment $\mathbb{E}(\mathbf{X}|Y)$, were introduced by Duan and Li (1991) and Li (1991), and have been studied extensively by several authors: see for instance Carroll and Li (1992), Hsing and Carroll (1992), Zhu and Ng (1995), or Aragon and Saracco (1997). The estimation scheme and the application of SIR-I have been discussed in detail by Chen and Li (1998). Asymptotic properties of SIR have been investigated in several articles with emphasis on the convergence and the asymptotic distribution of the estimator of the EDR space, see for instance Hsing and Carroll (1992), Koetter (1996), Zhu and Fang (1996), Saracco (1997), Hsing (1999), among others.

However, this approach is "blind" for symmetric dependencies (see Cook and Weisberg (1991) or Kötter (2000)). Therefore, SIR-II estimates based on the inverse conditional second moment $\mathbb{V}(\mathbf{X}|Y)$ have been suggested, see for instance Li (1991), Cook and Weisberg (1991), Kötter (2000) or Yin and Seymour (2005). Hence these two approaches concentrate on the use of the inverse conditional moments $\mathbb{E}(\mathbf{X}|Y)$ or $\mathbb{V}(\mathbf{X}|Y)$ to find the EDR space. To increase the chance of discovering all the EDR directions, the idea of the SIR$_\alpha$ method is to conjugate the information provided by SIR-I and SIR-II methods: if an EDR direction can only be marginally detected by SIR-I or SIR-II, SIR$_\alpha$ considers a mixture of these two methods (see the matrix M_α defined below) which may provide all the EDR directions.

Let us now recall the geometric properties of model (1). In SIR$_\alpha$ approach, Li (1991) considered, for $\alpha \in [0,1]$, the eigen-decomposition of the matrix

$$\Sigma^{-1} M_\alpha$$

where $M_\alpha = (1-\alpha) M_I \Sigma^{-1} M_I + \alpha M_{II}$. The matrices M_I and M_{II} are respectively the matrices used in the usual SIR-I and SIR-II approaches. They are defined as follows: $M_I = \mathbb{V}(\mathbb{E}(\mathbf{X}|T(Y)))$ and $M_{II} = \mathbb{E}\left\{ \left(\mathbb{V}(\mathbf{X}|T(Y)) - \mathbb{E}(\mathbb{V}(\mathbf{X}|T(Y)))\right) \Sigma^{-1} \left(\mathbb{V}(\mathbf{X}|T(Y)) - \mathbb{E}(\mathbb{V}(\mathbf{X}|T(Y)))\right)^T \right\}$ where T denotes a

monotonic transformation of Y. Specific transformations T are ordinarily used in order to simplify the expression of the matrices M_I and M_{II}. The most usual one is the slicing function (defined below and used in the rest of the paper). Note that the monotonicity of T is necessary in order to avoid pathological case of SIR-I only due to a bad choice of T. It can be shown that, under the linearity condition (3) and the constant variance assumption defined in Remark 2, and for any monotonic transformation T, the eigenvectors associated with the largest K eigenvalues of $\Sigma^{-1}M_\alpha$ are some EDR directions. Note also that, when $\alpha = 0$ (resp. $\alpha = 1$), SIR_α is equivalent to SIR-I (resp. SIR-II).

Li (1991) proposed a transformation T, called a slicing, which categorizes the response Y into a new response with $H > K$ levels. The support of Y is partitioned into H non-overlapping slices $s_1, \ldots, s_h, \ldots, s_H$. With such a transformation T, the matrices of interest are now written as

$$M_I = \sum_{h=1}^{H} p_h (m_h - \mu)(m_h - \mu)^T \quad \text{and} \quad M_{II} = \sum_{h=1}^{H} p_h \left(V_h - \overline{V}\right) \Sigma^{-1} \left(V_h - \overline{V}\right),$$

where $\mu = \mathbb{E}(\mathbf{X})$, $p_h = P(Y \in s_h)$, $m_h = \mathbb{E}(\mathbf{X}|Y \in s_h)$, $V_h = \mathbb{V}(\mathbf{X}|Y \in s_h)$ and $\overline{V} = \sum_{h=1}^{H} p_h V_h$.

Therefore, it is straightforward to estimate these matrices by substituting empirical versions of the moments for their theoretical counterparts, and therefore to obtain the estimation of the EDR directions. Each estimated EDR direction converges to an EDR direction at rate $\sqrt{n}$, see for instance Li (1991) or Saracco (2001). Asymptotic normality of the SIR_α estimates has been studied by Gannoun and Saracco (2003a).

Remark 1. The practical choice of the slicing function T is discussed in Li (1991), Kötter (2000) and Saracco (2001). Note that the user has to fix the slicing strategy and the number H of slices, then observations are assigned to slices by value. The SIR theory makes no assumption about the slicing strategy. In practice, there are naturally two possibilities: to fix the width of the slices or to fix the number of observations per slice. In their investigation of SIR-I, various researchers have preferred the second approach. From the sampling point of view, the slices are such that the number of observations in each slice is as close to each other as possible. To avoid artificial reduction of dimension, H must be greater than K. Also, in order to have at least two cases in each slice, H must be less than $[n/2]$ where $[a]$ denotes the integer part of a. Li (1991) noticed that the choice of the slicing is less crucial than the choice of a bandwidth, as in kernel-based methods. Simulation studies

(with $p = 5$ and 10) show that the influence of the "slicing parameter" is small when the sample size is greater than 100. Note that, in order to avoid the choice of a slicing (the transformation T is here the identity function and not the slicing function), the kernel-based estimate of SIR-I has been investigated, see Zhu and Fang (1996) or Aragon and Saracco (1997). However, these methods are hard to implement with regard to basic slicing one and are computationally slow. Moreover, Bura (1997) and Bura and Cook (2001) proposed a parametric version of SIR-I.

Remark 2. Note that two crucial conditions for the theoretical succes of SIR_α methods are the following: a linearity condition

$$\mathbb{E}(b^T\mathbf{X}|\beta_1^T\mathbf{X},\ldots,\beta_K^T\mathbf{X}) \text{ is linear in } \beta_1^T\mathbf{X},\ldots,\beta_K^T\mathbf{X};\ \forall b\in\Re^p, \tag{3}$$

and a constant variance condition

$$\mathbb{V}(\mathbf{X}|\beta_1^T\mathbf{X},\ldots,\beta_K^T\mathbf{X}) \text{ is non-random.} \tag{4}$$

Condition (3) is satisfied when $\mathbf{X}$ has an elliptically symmetric distribution and (4) is satisfied when $\mathbf{X}$ follows a multivariate normal distribution (which is an elliptically one): in this case, the linearity condition is satisfied for any directions, even if they may not be in the EDR space. Note that it is not possible to verify (3) since this involves the unknown directions β_k, contrary to ellipticity or normality condition which can be tested. Using the bayesian argument of Hall and Li (1993), we can infer that (3) holds approximately for many high dimensional data sets. An interesting and detailed discussion on the linearity condition can be found in Chen and Li (1998). Note that when the distribution of $\mathbf{X}$ is far from being multinormal (or from conditions (3) and (4)), the dimension reduction may not be possible with SIR approaches: that is the choice of K (based on the eigenvalues or on R_k) will give $K = p$.

3 Bootstrap estimate of the risk function $R_{k,\alpha}$

The risk function is defined by:

$$R_{k,\alpha} = \mathbb{E}\left[\text{Trace}(P_k\widehat{P}_{k,\alpha})\right]/k, \quad \forall k = 1\ldots,K \tag{5}$$

where $\widehat{P}_{k,\alpha}$ is the $\widehat{\Sigma}$-orthogonal projector onto the space spanned by the first k vector $\hat{b}_{l,\alpha}$ of $\widehat{B}_\alpha = [\hat{b}_{1,\alpha},\ldots,\hat{b}_{K,\alpha}]$, with $\hat{b}_{l,\alpha}$ the eigenvector associated to the lth greater eigenvalues of $\widehat{\Sigma}^{-1}\widehat{M}_\alpha$ the empirical version of $\Sigma^{-1}M_\alpha$ which is obtained from the sample $s = \{(\mathbf{X}_i, Y_i), i = 1,\ldots,n\}$.

Let $\mathcal{B}$ be the number of bootstrap replications and let $s^{(b)} = \left\{(\mathbf{X}_i^{(b)}, Y_i^{(b)}), i = 1, \ldots, n\right\}$ a non-parametric bootstrap sample replication. According to Efron (1982), a naïve bootstrap estimate of the mean square risk function is defined by:

$$\widehat{R}_{k,\alpha} = \frac{1}{\mathcal{B}} \sum_{b=1}^{\mathcal{B}} \widehat{R}_{k,\alpha}^{(b)} \tag{6}$$

where $\widehat{R}_{k,\alpha}^{(b)} = \text{Trace}\left(\widehat{P}_{k,\alpha}\widehat{P}_{k,\alpha}^{(b)}\right)/k$ and $\widehat{P}_{k,\alpha}^{(b)}$ is the projector onto the subspace spanned by the first k eigenvectors of the matrix of interest, which is obtained from the bootstrap replication sample $s^{(b)}$. Note that the practical criterion $\widehat{R}_{k,\alpha}$ could be computed for all $k = 1 \ldots, p$ whereas the $R_{k,\alpha}$ is only defined for $k = 1, \ldots, K$.

The aim is to find the dimension k of the model and to have a practical choice of α in the SIR_α method thanks to the bootstrap estimate. The proposed method consists in evaluating the $\widehat{R}_{k,\alpha}$ for all $(k, \alpha) \in \{1, \ldots, p\} \times [0, 1]$ and then in observing how much it departs from one. The best choice will be a couple $(\hat{K}, \hat{\alpha})$ which gives a value of $\widehat{R}_{k,\alpha}$ close to one, such that $\hat{K} << p$. In practice, there is no objective criteria in order to establish when a departure from one is close, but a visual expertise of the plot of the $\hat{R}_{k,\alpha}$ versus k and α allows us to choose the best couple. This point will be illustrated with simulations in the next section. Note that we only use N_α values in the interval $[0, 1]$ for α: we choose $\hat{\alpha}$ in the set $S_{N_\alpha} = \{\alpha_j = j/N_\alpha,\ j = 0, 1, \ldots, N_\alpha - 1\}$.

4 Simulation study

To evaluate the performance of the proposed method, we generate simulated data from the following regression model:

$$Y = (\mathbf{X}^T\beta_1)^2 \exp(\mathbf{X}^T\beta_1/\theta) + \gamma(\mathbf{X}^T\beta_2)^2 \exp(\mathbf{X}^T\beta_2/\theta) + \epsilon, \tag{7}$$

where $\mathbf{X}$ follows a p-dimensional standardized normal distribution and ϵ is standard normally distributed. We take $\beta_1 = (1, 1, 1, 0, \ldots, 0)^T$ and $\beta_2 = (0, \ldots, 0, 1, 1, 1)^T$. To visualize the data, we simulate samples of $n = 100$ data points from this model for different levels of θ: 1, 5 and 100, and for $\gamma = 0$. Figure 1 shows the plots of the response variable Y versus the index $\mathbf{X}^T\beta_1$ for these different values of θ. Clearly, the parameter θ has an influence on the form of the dependence between the index $\mathbf{X}^T\beta_1$ and Y. For $\gamma = 0$, when the value of θ is large (resp. small or medium), model (8)

is a symmetric (resp. non symmetric or partially symmetric) dependent model. Therefore, the parameter θ affects the choice of α in the SIR$_\alpha$ method.

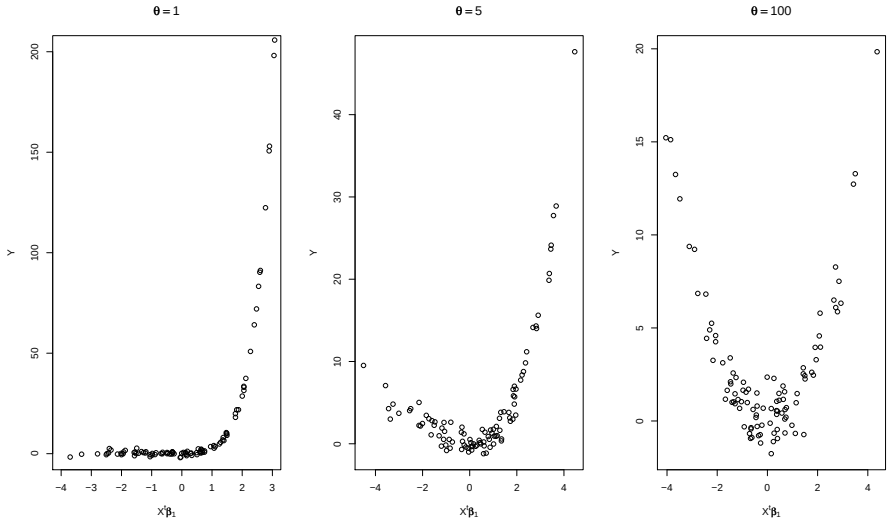

Figure 1: Plots of Y versus the index $\mathbf{X}^T\beta_1$ for different values of θ with $\gamma = 0$.

The parameter γ clearly has an influence on the choice of the dimension K: when $\gamma = 0$, only one direction is used in model (8), whereas if γ is nonnull (for instance if γ is fixed at 1), we have two directions.

The performance of the proposed method will be evaluated for different sample size ($n = 300$ or 500), various dimensions of the explanatory variable ($p = 5$ or 10), several choices of θ (=1, 5 or 100) and γ (=0 or 1). The number of slices used in the slicing step is given by the expression $H = \max(\sqrt{n}, p)$. The number $\mathcal{B}$ of bootstrap replications is chosen to be equal to 50. In these simulations, the parameter α varies in S_{N_α} for $N_\alpha = 11$.

In the next subsection, we detail our method applied on some simulated samples. Then, in subsection 4.2, we comment a complete simulation study. In subsection 4.3, we evaluate the robustness of the approach when $\mathbf{X}$ does not have a multivariate normal distribution. Simulations were performed with R. All the source codes are available from the authors by e-mail.

4.1 Simulated example

We consider here several samples corresponding to specific choices for the couple of parameters (k, α).

Symmetric dependent model ($\theta = 100$) and one direction ($\gamma = 0$). We generate a sample data of size $n = 300$ from the model (8) with $p = 5$. For each value of $\alpha \in S_{N_\alpha}$ and $k = 1, \ldots, 5$, we compute $\widehat{R}_{k,\alpha}$. For simplicity, we first investigate the effect of the parameter α on $\widehat{R}_{k,\alpha}$ for fixed values of k, then we investigate the behavior of k on $\widehat{R}_{k,\alpha}$ for fixed values of α. Finally, the influence of the couple (k, α) is simultaneously studied.

- *Influence of α on SIR$_\alpha$ estimates.* Figure 2 gives the boxplots of the $\widehat{R}^{(b)}_{k,\alpha}$ and the plots of the $\hat{R}_{k,\alpha}$ versus the values of α for several values of k (1, 2 and 3). Note that for $\alpha = 0$, the SIR$_\alpha$ method fails ($\widehat{R}_{k,\alpha} \simeq 0.6$ for every values of k). When $\alpha > 0$, $\hat{R}_{k,\alpha}$ provides greater values which are close to one for $k = 1$, whereas the $\hat{R}_{k,\alpha}$'s values are always lower than 0.8 when $k = 2$ or 3.

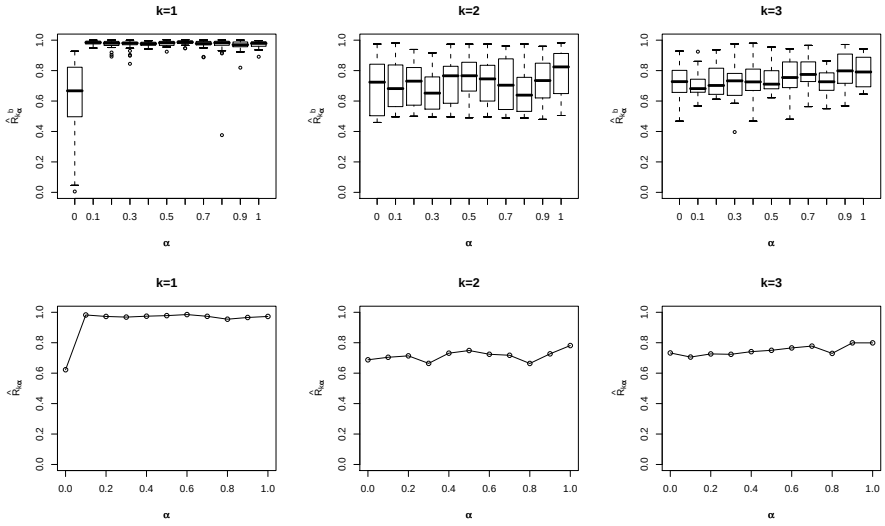

Figure 2: Boxplots of the $\widehat{R}^{(b)}_{k,\alpha}$'s values (above) and plots of $\hat{R}_{k,\alpha}$ (below) versus α, for $k = 1$, 2 and 3.

- *Choice of dimension K in the SIR$_\alpha$ method.* Figure 3 gives the boxplots of the $\widehat{R}^{(b)}_{k,\alpha}$'s and the plots of $\hat{R}_{k,\alpha}$ versus the values of k for several values of α (0, 0.1 and 1). Note that for $\alpha = 0.1$ and $\alpha = 1$, the $\hat{R}_{k,\alpha}$'s are close to 1 when $k = 1$, and they decrease for $k = 2$ before slowly increasing again.

- *Choice of dimension K and α in the SIR$_\alpha$ method.* The 3D-graphic in Figure 4 exhibits the $\hat{R}_{k,\alpha}$ values versus α and k. Note that the plots at the bottom of Figure 2 (resp. 3) are contained in the 3D-plot of Figure 4 if we focus only on the α-axis for a fixed k (resp. on the k-axis for a fixed α). In view of these figures, we clearly choose the dimension $\hat{K} = 1$ and $\hat{\alpha} = 0.1$.

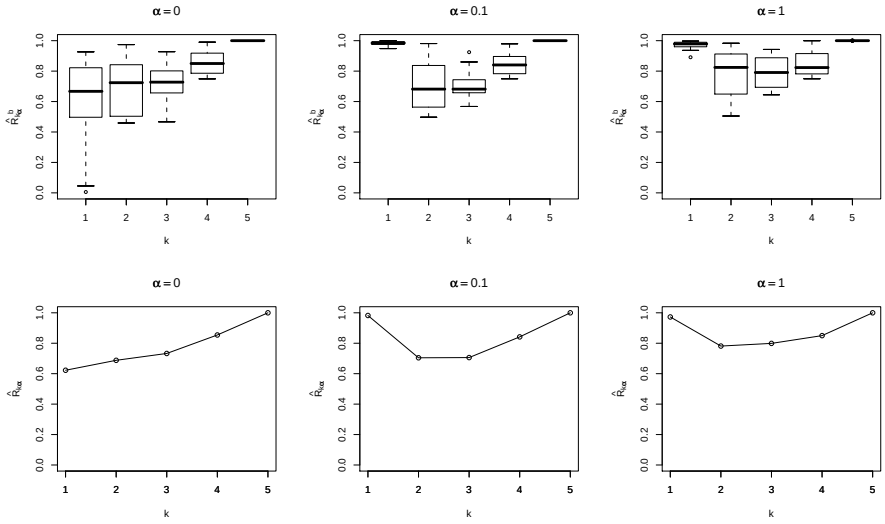

Figure 3: Boxplots of the $\widehat{R}_{k,\alpha}^{(b)}$'s values (above) and plots of $\hat{R}_{k,\alpha}$ (below) versus k, for $\alpha = 0$, 0.1 and 1.

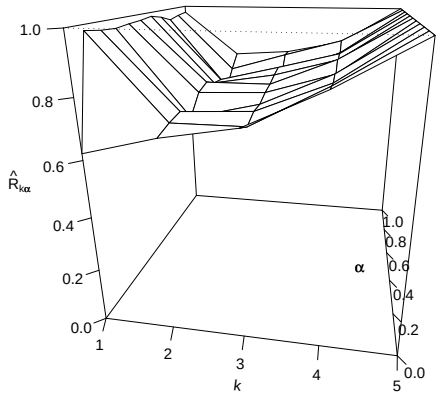

Figure 4: Plot of $\hat{R}_{k,\alpha}$ versus α and k.

Symmetric dependent model ($\theta = 100$) and two directions ($\gamma = 1$). First, we generate a sample data of size $n = 500$ from the model (8) with $p = 5$. For each value of α and $k = 1, \ldots, 5$, we compute $\widehat{R}_{k,\alpha}$.

- *Influence of α on SIR$_\alpha$ estimates.* Figure 5 gives the boxplots of the $\widehat{R}_{k,\alpha}^{(b)}$'s and the plots of $\hat{R}_{k,\alpha}$ versus the values of α for several values of k (1, 2 and 3). Note that for $\alpha = 0$, the SIR$_\alpha$ method fails ($\widehat{R}_{k,\alpha} = 0.68$). For $\alpha > 0$, $\hat{R}_{k,\alpha}$ gives greater values which are close to one for $k = 1$ and 2. When $k = 3$, the values of $\widehat{R}_{k,\alpha}$ are significantly lower than the previous ones.

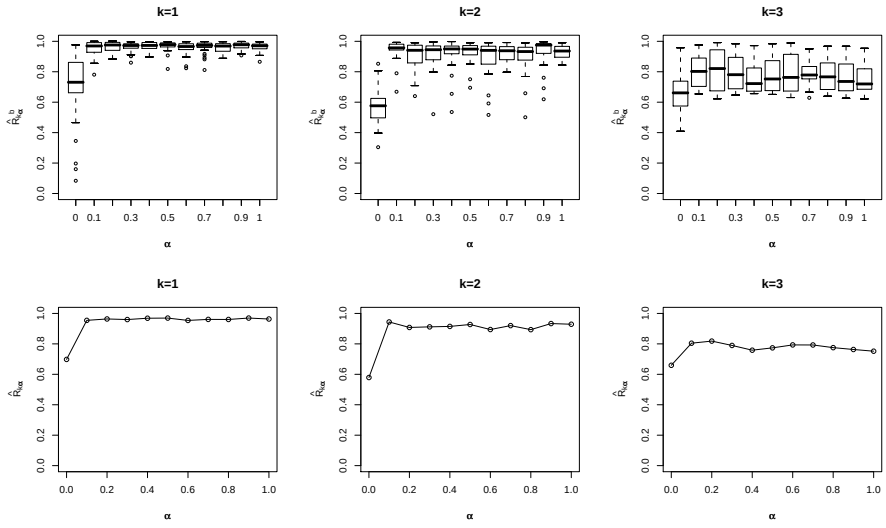

Figure 5: Boxplots of the $\widehat{R}_{k,\alpha}^{(b)}$'s values (above) and plots of $\hat{R}_{k,\alpha}$ (below) versus α, for $k = 1$, 2 and 3.

- *Choice of dimension K in the SIR$_\alpha$ method.* Figure 6 gives the boxplots of the $\widehat{R}_{k,\alpha}^{(b)}$'s and the plots of $\hat{R}_{k,\alpha}$ versus the values of k for several values of α (0, 0.1 and 1).Note that for $\alpha = 0.1$ and $\alpha = 1$, the $\hat{R}_{k,\alpha}$'s are close to 1 when $k = 1$ and 2, then they decrease for $k = 3$ and increase after.

- *Choice of dimension K and α in the SIR$_\alpha$ method.* The 3-D graphic in Figure 7 shows the $\hat{R}_{k,\alpha}$'s values versus α and k. In view of Figures 5, 6 and 7, we can choose $\hat{K} = 2$ and $\hat{\alpha} = 0.1$. In this case, we observe a plateau in Figure 7 for $\alpha = 0.1$. The explanation of this phenomenon is that the estimated eigenvectors $\hat{b}_1$ (resp. $\hat{b}_2$) and $\hat{b}_1^{(b)}$, corresponding to the $s^{(b)}$ samples, (resp. $\hat{b}_2^{(b)}$) are colinear (i.e. each direction appears to be individually identifiable); then the quality of the estimated EDR space for $k = 1$ and $k = 2$ is good (close to one). Nevertheless, if we only keep the first direction, we lose information provided by the second one (which is in the true two-dimensional

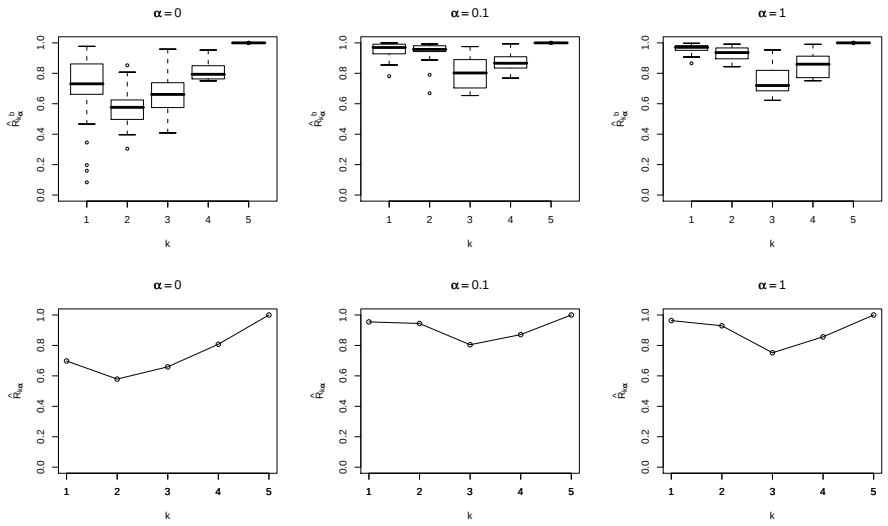

Figure 6: Boxplots of the $\widehat{R}_{k,\alpha}^{(b)}$'s values (above) and plots of $\hat{R}_{k,\alpha}$ (below) versus k for $\alpha = 0$, 0.1 and 1.

EDR space).

4.2 Results of the simulation study

For each combination of θ (=1, 5 or 100), γ (=0 or 1), p (=5 or 10) and n(=300 or 500), we generate $N = 500$ samples $\{(\mathbf{X}_i, Y_i),\ i = 1, \ldots, n\}$ from model (8). The EDR space has been estimated by the SIR$_\alpha$ method. For each sample, we evaluate the corresponding values of $\widehat{R}_{\alpha,k}$ for $k = 1, \ldots, p$ and α varying in S_{N_α}. For concision, we show here only few results for several combinations of θ, γ, p and n.

Single index ($\gamma = 0$) model. Table 1 shows the means of the $\widehat{R}_{\alpha,k}$'s over the $N = 500$ replications (denoted by $\overline{\widehat{R}}_{k,\alpha}$) for $n = 300$ and $p = 5$. When there is no symmetric dependence ($\theta = 1$) in the model, the values 0 or 0.1 for α provide $\widehat{R}_{\alpha,k}$ values close to 1 only for $k = 1$. When $\theta = 5$ or 100 (symmetric dependent model), the SIR-I ($\alpha = 0$) fails and we need to take $\hat{\alpha} > 0$ in order to get a $\widehat{R}_{\alpha,k}$ value close to 1 for $k = 1$. When $k > 1$ and $\alpha > 0$, the averages of the $R_{\alpha,k}$'s values are always smaller than those obtained with $k = 1$.

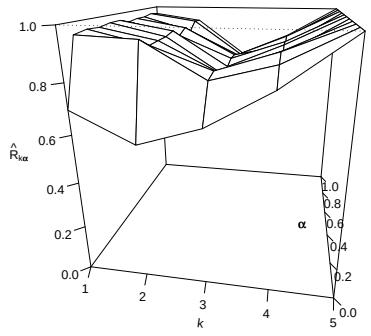

Figure 7: Plot of $\hat{R}_{k,\alpha}$ versus α and k.

$\overline{\widehat{R}}_{k,\alpha}$	$\theta = 1$				$\theta = 5$				$\theta = 100$			
α	$k=1$	$k=2$	$k=3$	$k=4$	$k=1$	$k=2$	$k=3$	$k=4$	$k=1$	$k=2$	$k=3$	$k=4$
0	0.989	0.706	0.743	0.843	0.922	0.692	0.738	0.843	0.393	0.554	0.694	0.831
0.1	0.990	0.710	0.748	0.851	0.972	0.697	0.746	0.850	0.971	0.697	0.742	0.850
0.2	0.989	0.713	0.750	0.851	0.965	0.703	0.749	0.849	0.971	0.697	0.743	0.850
0.3	0.977	0.711	0.752	0.851	0.939	0.697	0.746	0.851	0.968	0.700	0.743	0.849
0.4	0.890	0.706	0.749	0.852	0.898	0.691	0.744	0.850	0.967	0.700	0.742	0.849
0.5	0.548	0.669	0.745	0.849	0.844	0.683	0.742	0.848	0.966	0.697	0.743	0.849
0.6	0.250	0.537	0.717	0.847	0.798	0.674	0.740	0.848	0.964	0.698	0.744	0.849
0.7	0.355	0.494	0.659	0.830	0.763	0.666	0.738	0.848	0.964	0.696	0.743	0.850
0.8	0.409	0.580	0.703	0.825	0.731	0.656	0.736	0.847	0.963	0.697	0.744	0.849
0.9	0.419	0.608	0.763	0.903	0.714	0.651	0.734	0.847	0.962	0.698	0.742	0.848
1	0.420	0.613	0.780	0.957	0.698	0.642	0.730	0.846	0.962	0.696	0.742	0.848

Table 1: Means of the $\widehat{R}_{k,\alpha}$'s values over the $N = 500$ replications, denoted by $\overline{\widehat{R}}_{k,\alpha}$, for different values of θ when $\gamma = 0$, $n = 300$ and $p = 5$.

Model with two directions ($\gamma = 1$). Let us consider the $N = 500$ samples of size $n = 500$ for $p = 10$ and $\theta = 100$. Figure 8 shows the boxplots of the $\widehat{R}_{k,\alpha}$ over the $N = 500$ replications and the plots of the mean of the $\hat{R}_{k,\alpha}$'s versus the values of α, for several values of k (1, 2 and 3). Note that for $\alpha = 0$, the SIR$_\alpha$ method fails (with $\overline{\widehat{R}}_{k,\alpha} \leq 0.5$). For $\alpha > 0$, the $\hat{R}_{k,\alpha}$'s provide greater values that are close to one for $k = 2$. Figure 9 shows the boxplots of the $\widehat{R}_{k,\alpha}$'s values over the $N = 500$

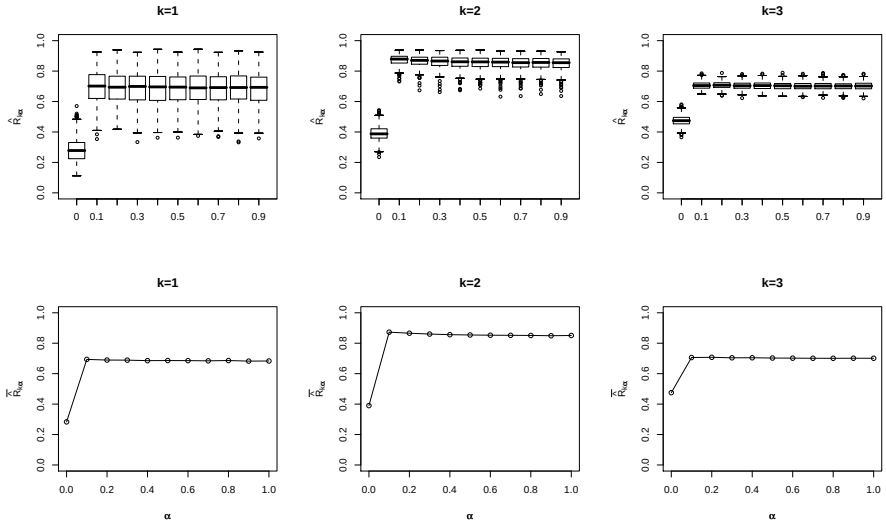

Figure 8: Boxplots of the $\widehat{R}_{k,\alpha}$'s values (above) and plots of $\overline{\widehat{R}}_{k,\alpha}$, mean of the $\widehat{R}_{k,\alpha}$'s, (below) versus α, for $k = 1$, 2 and 3.

replications and the plots of the mean of the $\hat{R}_{k,\alpha}$'s versus the values of k, for several values of α (0, 0.1 and 1). Note that for $\alpha = 0.1$ and 1, the $\hat{R}_{k,\alpha}$'s are close to 0.8 when $k = 1$, that they increase to 0.9 for $k = 2$, and then decrease for $k = 3$ and 4, before increasing slowly thereafter. Figure 10(a) represents the mean of the $\hat{R}_{k,\alpha}$'s values over the $N = 500$ replications versus α and k. In view of Figures 8, 9 and 10(a), the best choice (in mean) is the dimension $\hat{K} = 2$ and an $\hat{\alpha} > 0.1$.

We also show the results when $p = 5$ (with the same simulation parameters: $\gamma = 1$, $\theta = 100$, $n = 500$). Figure 10(b) shows the mean of the $\hat{R}_{k,\alpha}$'s values over the $N = 500$ replications versus α and k. Note that for $\alpha = 0$, the SIR$_\alpha$ method fails. For $\alpha > 0$, $\hat{R}_{k,\alpha}$ gives greater values that are close to one for $k = 1$ and 2. When $k = 3$, the values of $\widehat{R}_{k,\alpha}$ are significantly lower than the previous ones. In this case, we observe the same plateau phenomenon as described in the second example of section 4.1. The two directions appear to be individually identifiable, and a

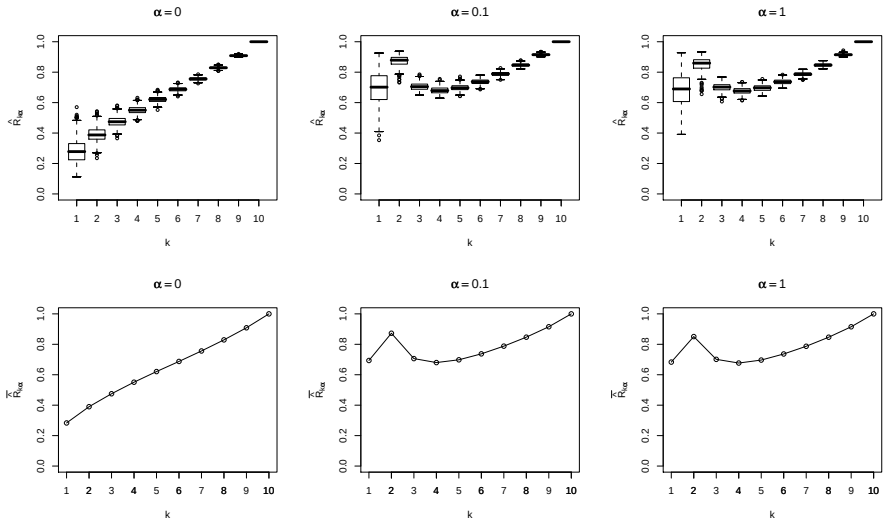

Figure 9: Boxplots of the $\widehat{R}_{k,\alpha}$'s (above) and plots of $\overline{\widehat{R}}_{k,\alpha}$ (below) versus k for $\alpha = 0$, 0.1 and 1.

(a) $p = 10$ (b) $p = 5$

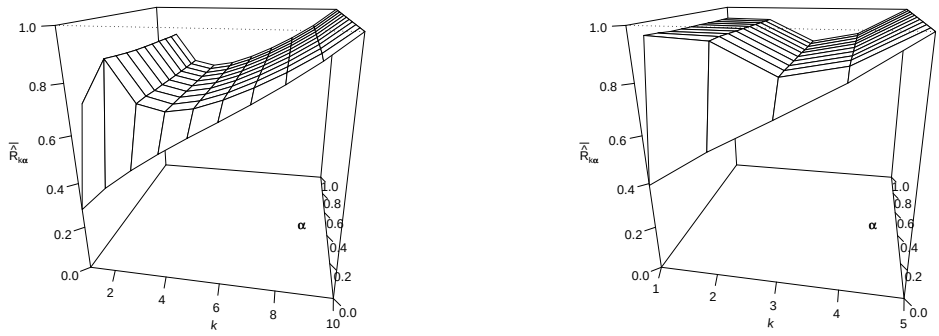

Figure 10: Plot of $\overline{\widehat{R}}_{k,\alpha}$ versus α and k.

straightforward consequence is that the quality of the estimated EDR space for $k = 1$ and $k = 2$ is good (close to one). Nevertheless, in order to keep all the EDR space information, we need to consider the two-dimensional subspace. Note that in the previous case (see Figure 10(a)), the first two directions were not individually identifiable, contrary to the two-dimensional EDR space; therefore the quality of the criterion for $k = 1$ is poor while it is close to one when $k = 2$. Hence, the best choice in mean is the dimension $\hat{K} = 2$ and an $\hat{\alpha} > 0.1$.

4.3 Case of a non-multivariate X

We also consider the case when $\mathbf{X}$ does not come from a multivariate normal distribution in order to obtain some idea of the robustness of our approach: the distribution of $\mathbf{X}$ is $\frac{1}{2}\mathcal{N}_p(\mu_2, \Sigma_2) + \frac{1}{2}\mathcal{N}_p(\mu_1, \Sigma_1)$ where $\mu_1 = (-2, \ldots, -2)^T$, $\Sigma_1 = 0.5I_p$, $\mu_2 = (2, \ldots, 2)^T$, $\Sigma_2 = I_p$. We present the results for the case where $n = 500$, $p = 10$, $\gamma = 1$ and $\theta = 100$. Figure 11 shows the boxplots

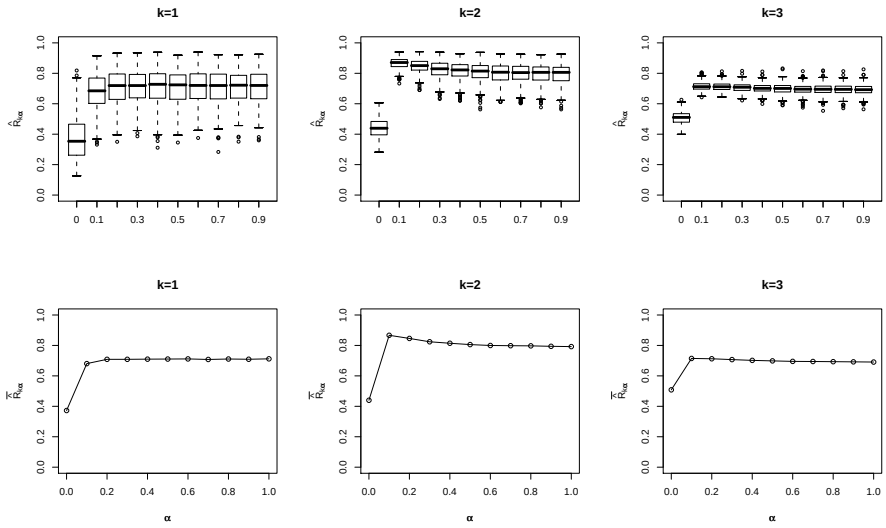

Figure 11: Boxplots of the $\widehat{R}_{k,\alpha}$'s values (above) and plots of $\overline{\widehat{R}}_{k,\alpha}$ (below) versus α, for $k = 1$, 2 and 3.

of the $\widehat{R}_{k,\alpha}$'s values over the $N = 500$ replications and the plots of the mean of the $\hat{R}_{k,\alpha}$'s versus the values of α for several values of k (1, 2 and 3). Note that for $\alpha = 0$, the SIR_α method fails ($\bar{\widehat{R}}_{k,\alpha} \simeq 0.4$). When $k = 2$, the $\hat{R}_{k,\alpha}$'s provide greater values for $\alpha \geq 0.1$. Figure 12 shows the boxplots of the $\widehat{R}_{k,\alpha}$'s values over all the replications and the plots of the mean of the $\hat{R}_{k,\alpha}$'s values versus the values of k for several values of α (0, 0.1 and 1). Note that for $\alpha = 0.1$, the $\hat{R}_{k,\alpha}$'s are

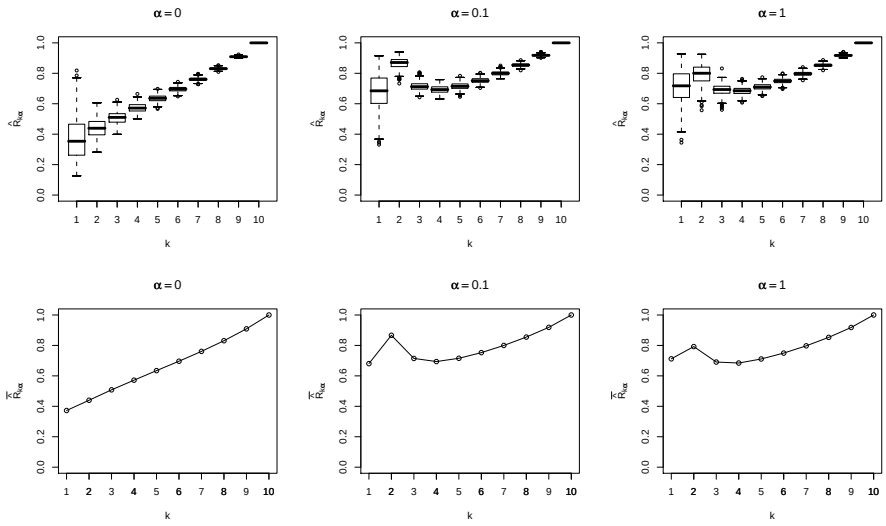

Figure 12: Boxplots of the $\widehat{R}_{k,\alpha}$'s values (above) and plots of $\overline{\widehat{R}}_{k,\alpha}$ (below) versus k for $\alpha = 0$, 0.1 and 1.

close to 1 when $k = 2$, then they decrease until $k = 4$, and that they slowly increase thereafter. Figure 13 represents the mean of the $\hat{R}_{k,\alpha}$'s values over all the replications versus α and k. In view of figures 11, 12 and 13, the best choice in mean is the dimension $\hat{K} = 2$ and an $\hat{\alpha} = 0.1$.

4.4 Comparison with other methods

In this subsection, we compare the proposed bootstrap method with the nested tests procedure introduced by Li (1991) and with the method developed by Ferré (1998), which is based on an estimation of the risk function R_k thanks to asymptotic expansions. These two methods only focus on the SIR-I approach (that is when $\alpha = 0$ in SIR_α). Therefore, the choice of the parameter α is not necessary and is fixed at zero in our approach. This is a first limitation of these two existing methods for determining dimension K. In the following,we recall these two methods. Then we describe the simulated model. Finally we comment the results of this simulation study.

The nested tests procedure. Li (1991) suggested evaluating K by successively testing the nullity of the $(p-k)$ smallest eigenvalues, starting at $k = 0$. He proposed a chi-squared test based on the mean of the smallest $(p-k)$ eigenvalues of the matrix M_I, denoted by $\bar{\lambda}_{(p-k)}$. Theoretically, for the true dimension K and if $\mathbf{X}$ is normally distributed, Li (1991) showed that $n(p-K)\bar{\lambda}_{(p-K)}$

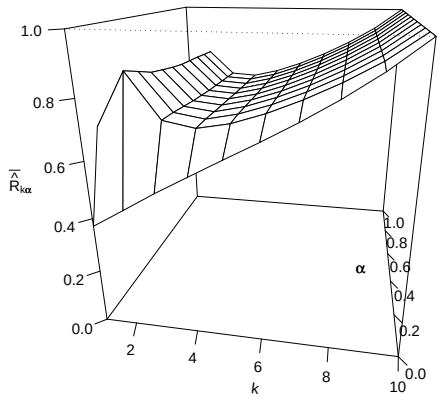

Figure 13: Plot of $\overline{\widehat{R}}_{k,\alpha}$ versus α and k.

follows a χ^2 distribution with $(p-K)(H-K-1)$ degrees of freedom asymptotically. From this result, if the rescaled $\bar{\lambda}_{(p-k)}$ is larger than the corresponding χ^2 value (say the 95th percentile), we may infer that there are at least $k+1$ (significant) indices in the model.

Ferré's method. Ferré (1998) proposed to estimate the risk function R_k defined in (2). Under technical assumptions and when $\mathbf{X}$ follows an elliptically distribution, Ferré showed that:

$$R_k = 1 - \frac{1}{nk}\left[\sum_{i=1}^{k}\sum_{j=k+1}^{K}\frac{(\lambda_i+\lambda_j+(-1+\kappa)\lambda_i\lambda_j)}{(\lambda_i-\lambda_j)^2} + (p-K)\sum_{i=1}^{q}\frac{1}{\lambda_i}\right] + O(n^{-3/2})$$

for $k = 1$ to K where κ is the usual kurtosis parameter. From this asymptotic expansion, an estimator $\hat{R}_k$ is defined by substituting the eigenvalues and κ by their estimates. Nevertheless, this criterion depends on the true dimension K (unknown). Ferré proposed to compute $\hat{R}_k$ (in fact a best notation should be here $\hat{R}_{k,K}$) for $K = 1$ to p and $k = 1$ to p, and he mentioned that "the decision rule deduces from the observation of the p curves obtained by plotting $\hat{R}_{k,K}$ versus k, for $k = 1$ to p and for $K = 1$ to p". In practice, it is not really easy to determine the dimension from this graphical aspect (in particular when p is large).

Simulated model. We generate simulated data from the following regression model (in which the true dimension K is equal to 2):

$$Y = (\mathbf{X}'\beta_1)\exp(\mathbf{X}'\beta_2) + \epsilon, \tag{8}$$

where $\mathbf{X}$ follows a 10-dimensional standardized normal distribution and ϵ is normally distributed with σ^2. We take $\beta_1 = (1, 1, 0, \ldots, 0)'$ and $\beta_2 = (0, 0, 0, 1, 1, 0, \ldots, 0)'$. Note that we consider here the multinormal case corresponding with the theoretical assumption of the nested tests procedure and Ferré method. Moreover, this model is not a symmetric dependent model, then SIR-I should provide good performance.

The performances of the different methods are evaluated for different sample sizes ($n = 100$, 400 or 1000) and several choices of σ^2 ($= 0.01$ or 1). For each value of (n, σ^2), $N = 500$ samples have been generated. For each sample, we determined the dimension with the three methods (nested tests, Ferré's approach, bootstrap approach).

Results and comments. Using our bootstrap approach, Figure 14 shows the boxplots of the $\hat{R}_{k,\alpha}$'s values over the $N = 500$ replications and the plots of the mean of the $\hat{R}_{k,\alpha}$'s versus the values of k, in the case where $\sigma^2 = 1$, $n = 100$ and $n = 400$. Note that for $n = 100$, the $\hat{R}_{k,\alpha}$'s are close to 0.7 when $k = 1$, increase to 0.8 for $k = 2$, and then decrease for $k = 3$ and 4, before increasing slowly thereafter. In view of this figure, the best choice (in mean) is the dimension $\hat{K} = 2$. Similar (and best) results have been observed for $n = 400$ and $\sigma^2 = 1$ (see Figure 14). When $\sigma^2 = 0.01$ and $n = 100$, 400 or 1000, we obtained similar graphics which are not shown in the paper.

Let us now comment the results from the nested tests procedure. Figure 15 shows the boxplots of the p-values of the nested tests for all values of (n, σ^2). On these graphics, if a p-value is lower than 5% for a value of k and is greater than 5% for $k+1$, the dimension $k+1$ is chosen. Note that the only case where this procedure had good performance (that is determine the true dimension) is for large sample size ($n = 1000$) and small variance ($\sigma^2 = 0.01$). The dimension $\hat{K} = 2$ has been always chosen. In the other cases, the method failed. For instance, for medium sample size ($n = 400$) and small variance ($\sigma^2 = 0.01$), the procedure chose $\hat{K} = 2$ for 23 simulated samples, and $\hat{K} = 1$ for 477 samples. When the variance is $\sigma^2 = 1$, the procedure never find the true dimension ($\hat{K} = 1$ for $n = 1000$, no dimension reduction for $n = 100$).

Concerning Ferré approach, the estimate $\hat{R}_{k,K}$ of R_K only gave reasonable values (in [0,1]) for $K = 1$ and for the true dimension $K = 2$. In the other cases, the values for $\hat{R}_{k,K}$ are always (in

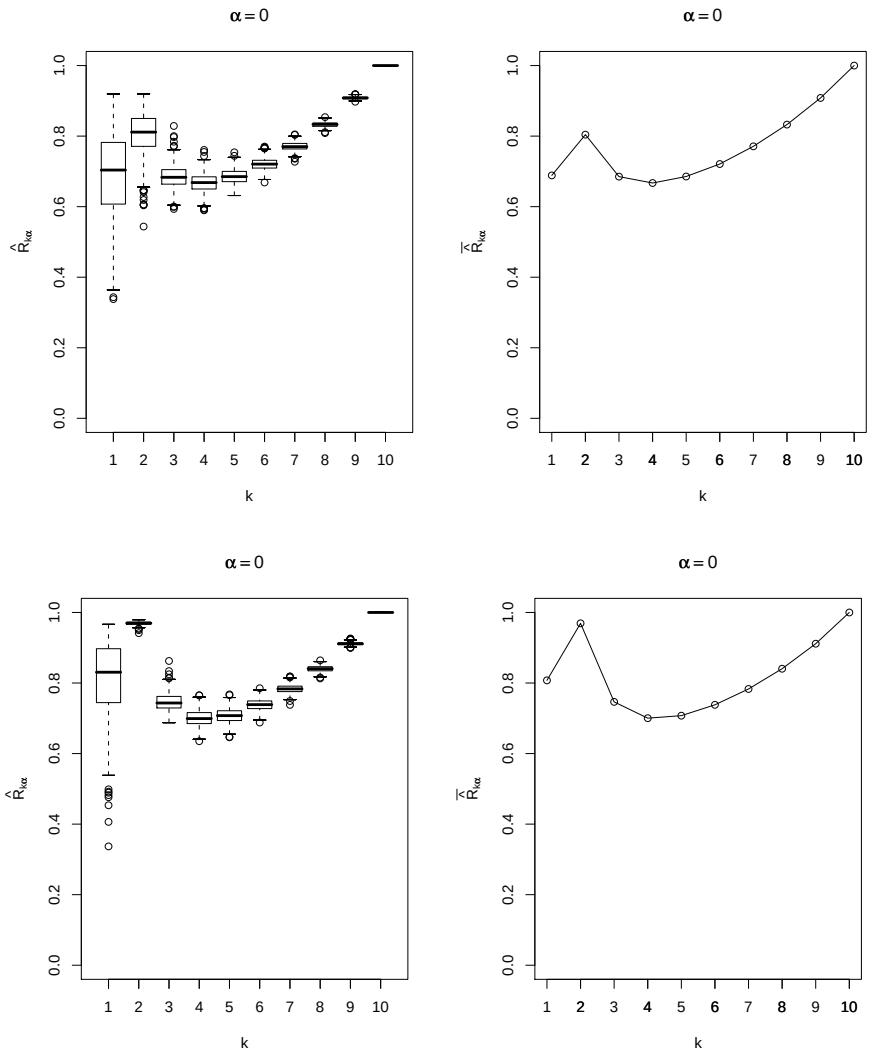

Figure 14: Boxplots of the $\widehat{R}_{k,\alpha}$ (on the left handside) and plots of the mean of $\hat{R}_{k,\alpha}$'s (on the right handside) versus k for $\alpha = 0$ with $\sigma^2 = 1$, $n = 100$ (above) and $n = 400$ (below)

mean over the N replications) negative values. This phenomenon is particularly obvious for small sample size and large variance. This problem has been already mentioned by Ferré and Yao (1999) for a SIR-II version of this approach. An explanation of these "bad" values can be the proximity of the eigenvalues. From these remarks, it seems difficult to use this method to determine dimension, even if $\hat{R}_{k,K}$ appears to be well estimated for the true dimension (which is unknown in practice).

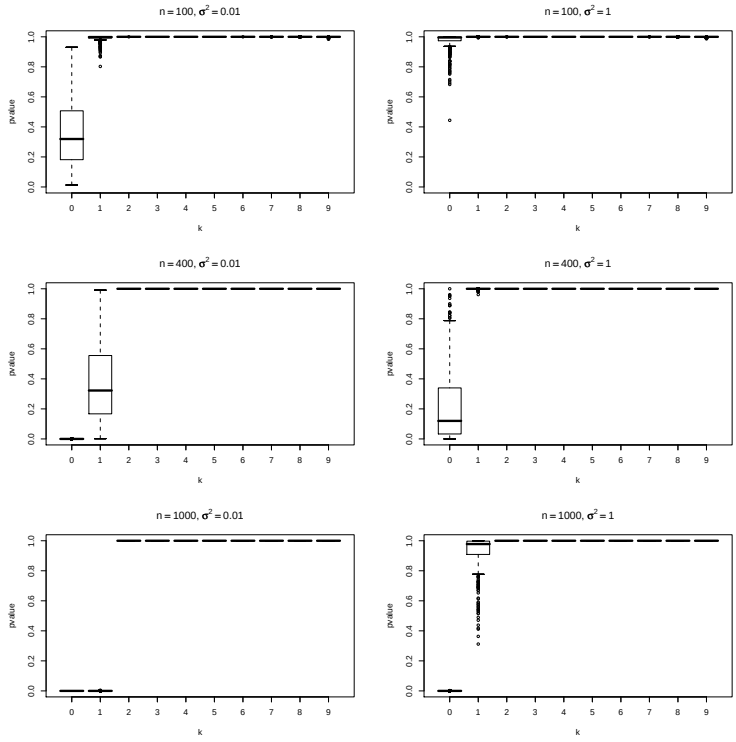

Figure 15: Boxplots of the p-values of the nested tests for different values of (n, σ^2)

Final comments. Note that our bootstrap approach always provides a good decision concerning the choice of dimension, contrary to the other ones. Moreover, this approach also makes it possible the choice of α (which was fixed, $\alpha = 0$, in the previous comparison) for all kinds of models (especially for symmetric dependent models).

The nested tests procedure and the Ferré approach rely on multinormal or elliptical assumption for the distribution of $\mathbf{X}$. In the previous subsection, we show that our method still provides good performance in non elliptical case.

5 Concluding remarks

This article proposes a practical criterion based on bootstrap to choose the dimension and α in the SIR_α method. The simulation study demonstrates good numerical behaviour of this approach, contrary to existing methods (which only focus on the choice of dimension K). We underline the pleasant graphical aspect of the criterion which allows the practitioner to choose both K and α.

The method has been implemented in R, and the source codes are available from the author. They provide the three kinds of graphics in order to visualize the quality of the estimated EDR space when k (resp. α) varies for a fixed α (resp. k), and they give the 3D-plot of this quality when the couple (k, α) varies on a grid. We used a rough grid in our simulation study, but for a dataset, we recommend working with a thin grid.

Importantly the choice of parameter α seems to be less sensitive than the choice of dimension K. More precisely, for the choice of α, there are generally two scenarios: the case where $\alpha = 0$ and the case where $\alpha > 0$ (when a symmetric dependence occurs in the model). The 3D-graphics in Section 4 show that for each fixed k, the profiles of the criterion are nearly similar when $\alpha > 0$.

Finally, this approach can be extended to the multivariate SIR_α method named pooled marginal slicing PMS_α, see Saracco (2005) or Barreda et al. (2007).

References

Aragon, Y. and Saracco, J. (1997). Sliced Inverse Regression (SIR): an appraisal of small sample alternatives to slicing. *Computational Statistics*, **12**, 109-130.

Bai, Z. D. and He, X. (2004). A chi-square test for dimensionality for non-Gaussian data. *Journal of Multivariate Analysis*, **88**, 109-117.

Barreda, L., Gannoun, A. and Saracco, J. (2007). Some extensions of multivariate sliced inverse regression. *Journal of Statistical Computation and Simulation*, **77**, 1-17.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). *Journal of the American Statistical Association*, **80**, 580-619.

Bura, E. (1997). Dimension reduction via parametric inverse regression. *L_1-statistical procedures and related topics (Neuchel, 1997), IMS Lecture Notes Monogr. Ser.*, **31**, 215-228.

Bura, E. and Cook, R. D. (2001). Estimating the structural dimension of regressions via parametric inverse regression. *Journal of the Royal Statistical Society, Series B*, **63**, 393-410.

Carroll, R. J. and Li, K. C.(1992). Measurement error regression with unknown link: dimension reduction and data visualization. *Journal of the American Statistical Association*, **87**, 1040-1050.

Chen, H. (1991). Estimation of a projection-pursuit type regression model. *Annals of Statistics*, **19**, 142-157.

Chen, C. H. and Li, K. C. (1998). Can SIR be as popular as multiple linear regression? *Statistica Sinica*, **8**, 289-316.

Cook, R. D. and Weisberg, S. (1991). Discussion of "Sliced inverse regression". *Journal of the American Statistical Association*, **86**, 328-332.

Duan, N. and Li, K. C. (1991). Slicing regression: a link-free regression method. *The Annals of Statistics*, **19**, 505-530.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans.* CBMS-NSF Regional Conference Series in Applied Mathematics, 38. Society for Industrial and Applied Mathematics (SIAM), Philadelphia.

Ferré, L. (1997). Dimension choice for sliced inverse regression based on ranks. *Student*, **2**, 95-108.

Ferré, L. (1998). Determining the dimension in Sliced Inverse Regression and related methods. *Journal of the American Statistical Association*, **93**, 132-140.

Ferré, L. and Yao, A. F. (1999). Un critere de choix de la dimension dans la méthode SIR-II. *Revue de Statistique Appliquée*, **47**, 33-46.

Friedman, J. H. (1991). Multivariate adaptative regression splines (with discussion). *Annals of Statistics*, **19**, 1-141.

Friedman, J. H. and Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modeling (with discussion). *Technometrics*, **31**, 3-39.

Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association*, **76**, 817-823.

Fung, W. K., He,, X., Liu, L. and Shi, P. D. Dimension reduction based on canonical correlation. *Statistica Sinica*, **12**, 1093-1114.

Gannoun, A. and Saracco, J. (2003a). An asymptotic theory for SIR_α method. *Statistica Sinica*, **13**, 297-310.

Gannoun, A. and Saracco, J. (2003b). Two Cross Validation Criteria for SIR_α and PSIR_α methods in view of prediction. *Computational Statistics*, **18**, 585-603.

Hall, P. (1989). On projection pursuit regression. *Annals of Statistics*, **17**, 573-588.

Hall, P. and Li, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. *The Annals of Statistics*, **21**, 867-889.

Hastie, T. J. and Tibshirani, R. J. (1986). Generalized additive models. *Statistical Science*, **1**, 297-318.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models.* Chapman and Hall, London.

Hsing, T. (1999). Nearest neighbor inverse regression. *The Annals of Statistics*, **27**, 697-731.

Hsing, T. and Carroll, R. J. (1992). An asympotic theory for Sliced Inverse regression. *The Annals of Statistics*, **20**, 1040-1061.

Kötter, T. (1996). An asymptotic result for Sliced Inverse Regression. *Computational Statistics*, **11**, 113-136.

Kötter, T. (2000). Sliced Inverse Regression. In *Smoothing and Regression. Approaches, Computation, and Application* (Edited by M. G. Schimek), 497-512. Wiley, New York.

Li, K. C. (1991). Sliced inverse regression for dimension reduction, with discussion. *Journal of the American Statistical Association*, **86**, 316-342.

Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma. *Journal of the American Statistical Association*, **87**, 1025-1039.

Li, K. C., Aragon, Y., Shedden, K. and Thomas Agnan, C. (2003). Dimension reduction for multivariate response data. *Journal of the American Statistical Association*, **98**, 99-109.

Saracco, J. (1997). An asymptotic theory for Sliced Inverse Regression. *Communications in Statistics - Theory and methods*, **26**, 2141-2171.

Saracco, J. (2001). Pooled Slicing methods versus Slicing methods. *Communications in Statistics - Simulation and Computation*, **30**, 489-511.

Saracco, J. (2005). Asymptotics for pooled marginal slicing estimator based on SIR_α approach. *Journal of Multivariate Analysis*, **96**, 117-135.

Schott, J. R. (1994). Determining the dimensionality in Sliced Inverse Regression. *Journal of the American Statistical Association*, **89**, 141-148.

Stone, C. J. H. (1985). Additive regression and other nonparametric models. *Annals of Statistics*, **13**, 689-705.

Stone, C. J. H. (1986). The dimensionality reduction principle for generalized additive models. *Annals of Statistics*, **14**, 590-606.

Yin, X. and Seymour, L. (2005). Asymptotic distributions for dimension reduction in the SIR-II method. *Statistica Sinica*, **15**, 1069-1079.

Zhu, L. X. and Ng, K. W. (1995). Asymptotics of sliced inverse regression. *Statistica Sinica*, **5**, 727-736.

Zhu, L. X. and Fang, K. T. (1996). Asymptotics for kernel estimate of Sliced Inverse Regression. *The Annals of Statistics*, **24**, 1053-1068.