



HAL
open science

A new approach on recursive and non-recursive SIR methods

Bernard Bercu, Thi Mong Ngoc Nguyen, Jérôme Saracco

► **To cite this version:**

Bernard Bercu, Thi Mong Ngoc Nguyen, Jérôme Saracco. A new approach on recursive and non-recursive SIR methods. Journal of the Korean Statistical Society, 2012, 41, pp.17-36. 10.1016/j.jkss.2011.05.005 . hal-00642653

HAL Id: hal-00642653

<https://hal.science/hal-00642653>

Submitted on 18 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A new approach on recursive and non recursive SIR methods

Bernard BERCU^{1,2}, Thi Mong Ngoc NGUYEN^{1,3}, Jérôme SARACCO^{1,3,4}

¹ Institut de Mathématiques de Bordeaux, UMR CNRS 5251, Université de Bordeaux,
351 cours de la libération, 33405 Talence Cedex, France.

² INRIA Bordeaux Sud-Ouest, ALEA team, France.

³ INRIA Bordeaux Sud-Ouest, CQFD team, France.

⁴ Institut Polytechnique de Bordeaux, 1 avenue du Dr Albert Schweitzer, 33402 Talence Cedex, France.

e-mail : {Bernard.Bercu,Thi.Mong.Ngoc.Nguyen,Jerome.Saracco}@math.u-bordeaux1.fr

Abstract

We consider a semiparametric single index regression model involving a p -dimensional quantitative covariable x and a real dependent variable y . A dimension reduction is included in this model via an index $x'\beta$. Sliced inverse regression (SIR) is a well-known method to estimate the direction of the Euclidean parameter β which is based on a “slicing step” of y in the population and sample versions. The goal of this paper is twofold. On the one hand, we focus on a recursive version of SIR which is also suitable for multiple indices model. On the other one, we propose a new method called SIRoneslice when the regression model is a single index model. The SIRoneslice estimator of the direction of β is based on the use of only one “optimal” slice chosen among the H slices. Then, we provide its recursive version. We give an asymptotic result for SIRoneslice approach. Simulation study shows good numerical performances of SIRoneslice method and clearly exhibits the main advantage of using recursive versions of the SIR and SIRoneslice methods from a computational times point of view. A real data set is also used to illustrate the approach. Some extensions are discussed in concluding remarks. The proposed methods and criterion have been implemented in R and the corresponding codes are available from the authors.

Keywords: *recursive estimation, semiparametric regression model, sliced inverse regression (SIR).*

1 Introduction

In statistical applications, high-dimensional data became common. In a regression framework, let $y \in \mathbb{R}$ be the response variable and $x \in \mathbb{R}^p$ be the regressor. In order to study the relationship between y and x , it is usual to impose assumptions for a specific structure on the mean regression function of y on x such as linearity or additivity. To cope with dimensionality, in a dimension reduction setting, many authors suppose that x

can be replaced by a linear combination of its components, $\beta'x$, without losing information on the conditional distribution of y given x . This assumption can be expressed as

$$y \perp x \mid \beta'x \tag{1}$$

where the notation $u_1 \perp u_2 \mid u_3$ means that the random variable u_1 is independent of the random variable u_2 given any values for the random variable u_3 . An example where (1) holds is the following single index model with an additive error:

$$y = f(\beta'x) + \varepsilon, \tag{2}$$

where $\varepsilon \perp x$, the distribution of ε is arbitrary and unknown, and f is an unknown real-valued function. In this model, sufficient dimension reduction of the regression is achieved and leads to summary plot of y versus $\beta'x$ which provides graphical modelling information. Then, it is possible to nonparametrically estimate the link function f on the reduced data, more efficiently in the sense that the curse of dimensionality has been overcome. Finally, when (1) holds, it straightforwardly also holds for any vector collinear to β . Let us call $E := \text{span}(\beta)$ the dimension reduction subspace, also named effective dimension reduction (EDR) space following Duan and Li (1991) in their original presentation of sliced inverse regression (SIR).

Li (1991) consider a multiple indices regression model. The Euclidean parameter β is a $p \times K$ matrix: $\beta = [\beta_1, \dots, \beta_K]$ where the vectors β_k are assumed linearly independent. The EDR space is now the K -dimensional linear subspace of \mathbb{R}^p spanned by the β_k 's.

Methods based on the use of inverse regression are available in literature for estimating the EDR space. In order for inverse regression to be useful in estimating the EDR space, some of them, like SIR and related methods (for instance principal Hessian direction introduced by Li (1992) or sliced average variance estimation discussed by Cook (2000)) place certain conditions on the marginal distribution of the covariable x . In this paper, we will focus on SIR. Let us then now recall the theory of SIR and its necessary assumption on x . The SIR approach relies on the following assumption often called linearity condition (LC):

$$\text{For all } b \in \mathbb{R}^p, \mathbb{E}[b'x \mid \beta'x] \text{ is linear in } x'\beta. \tag{3}$$

The LC is required to hold only for the true Euclidean parameter β . Since β is unknown, in practice it is not possible to verify a priori this assumption. Hence, we can assume that LC holds for all possible β , which is equivalent to elliptical symmetry of the distribution of x . Recall that the well-known multivariate normal distribution is an example of such a distribution. Finally Hall and Li (1993) mentioned that the LC is not a severe restriction because this LC holds to a good approximation in many problems as the dimension p of the predictors increases, see also Chen and Li (1998) or Cook and Ni (2005) for interesting discussions on the LC.

Let us consider a monotone transformation T . Under model (1) and LC, Duan and Li (1991) showed that the centered inverse regression curve satisfies:

$$\mathbb{E}[x | T(y)] - \mu \in \text{Span}(\Sigma\beta), \quad (4)$$

where $\mu := \mathbb{E}[x]$ and $\Sigma := \mathbb{V}(x)$. This result implies that the space spanned by the centered inverse curve, $\{\mathbb{E}[x | T(y)] - \mathbb{E}[x] : y \in \mathcal{Y}\}$ where \mathcal{Y} is the support of response variable y , is a subspace of the EDR space but it does not guarantee equality. For instance, pathological model has been identified in the literature and are called symmetric dependent model: that is model for which the centered inverse regression curve is degenerated.

From (4), the centered inverse regression curve can be used to recover the EDR space if the model is not pathological. Indeed, a direct consequence of this result is that the covariance matrix of this curve,

$$\Gamma := \mathbb{V}(\mathbb{E}[x | T(y)]),$$

is degenerate in any direction Σ -orthogonal to β (i.e. to the β_k 's for a multiple indices model). Therefore, the eigenvectors associated with the non null eigenvalues of $\Sigma^{-1}\Gamma$ are EDR directions, which means that they span the EDR space E .

In the slicing step of SIR, the range of y is partitioned into H slices $\{s_1, \dots, s_H\}$. With such slicing, the covariance matrix Γ can be straightforwardly written this way:

$$\Gamma := \sum_{h=1}^H p_h (m_h - \mu)(m_h - \mu)'$$

where $p_h = P(y \in s_h)$ and $m_h = \mathbb{E}[x | y \in s_h]$. Let us consider a random sample $\{(x_i, y_i), i = 1, \dots, n\}$ generated from model (1). By substituting empirical versions of μ , Σ , p_h and m_h for their theoretical counterparts, we obtain an estimated basis of E spanned by the eigenvector \hat{b}_{SIR} associated with the largest eigenvalue of the estimate $\hat{\Sigma}_n^{-1}\hat{\Gamma}_n$ of $\Sigma^{-1}\Gamma$ where

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)' \quad \text{and} \quad \hat{\Gamma}_n = \sum_{h=1}^H \hat{p}_{h,n} (\hat{m}_{h,n} - \bar{x}_n)(\hat{m}_{h,n} - \bar{x}_n)',$$

with

$$\bar{x}_n = \sum_{i=1}^n x_i/n, \quad \hat{p}_{h,n} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{[y_i \in s_h]} = \frac{\hat{n}_{h,n}}{n} \quad \text{and} \quad \hat{m}_{h,n} = \frac{1}{\hat{n}_{h,n}} \sum_{i \in s_h} x_i,$$

the notation $\mathbb{I}_{[\cdot]}$ designating the indicator function. This approach is the one proposed by Duan and Li (1991) and Li (1991) when they initially introduced SIR approach. The SIR method has been extensively studied by several authors, see for instance Hsing and Carroll (1992), Zhu and Ng (1995), Saracco (1997), Chen and Li (1998), Cook and Ni (2005) among others.

The goal of this paper is twofold. On the one hand, we focus on recursive version of SIR which is suitable for multiple indices model and has never been introduced in the previous literature. The recursive approach relies

on the use of the estimator defined on the $n - 1$ first observations (x_i, y_i) and the new observation (x_n, y_n) . For instance, we can illustrate this recursive framework with the following practical example where the data do not arrive at the same time and thus the recursive version of SIR is useful. Let us consider a database of costumers (of a website for example) which increases as time goes by. The statistician uses this database in order to infer the association between a dependent variable y (the amount of expenses for example) and a p -dimensional vector x of covariables (some numerical informations given by the consumer when he creates an account on this website for instance). For each new costumer, the recursive SIR approach does not require the use of the entire database (which can be huge) to update the relationship between y and x (that is the EDR direction) and the time-saving will be important from a computational point of view. On the other hand, we propose a new method called SIRoneslice when the regression model is a single index model. We define in Section 2 the SIRoneslice estimator of the direction of β based on the use of only one “optimal” slice chosen among the H slices. Then, we provide in Section 3 the recursive versions of the SIR and SIRoneslice procedures and we also establish an asymptotic result for the SIRoneslice estimator. The proof is postponed in the Appendix. In Section 4, a simulation study shows very good numerical performances of SIRoneslice method and clearly exhibits the main advantage of using recursive versions of the SIR and SIRoneslice methods from a computational times point of view. A real data set is used to illustrate the proposed approach. All the methods have been implemented in R. Finally concluding remarks are given in Section 5.

2 A “one slice-based SIR” approach

Under model (1) and the LC, $\mathbb{E}[x | T(y)]$ can be written as, for each slice h ,

$$\mathbb{E}[x | y \in s_h] = \mu + k_h \Sigma \beta, \quad (5)$$

where

$$k_h = \frac{\mathbb{E}[(x - \mu)' \beta | y \in s_h]}{\beta' \Sigma \beta}, \quad (6)$$

details are given in Appendix A.

For all $h = 1, \dots, H$, let

$$z_h := \mathbb{E}[x | y \in s_h] - \mu = m_h - \mu.$$

We have respectively from (5) and (6)

$$z_h = k_h \Sigma \beta \quad \text{and} \quad k_h = \frac{z_h' \beta}{\beta' \Sigma \beta}. \quad (7)$$

Then, z_h falls along a line spanned by $\Sigma \beta$ if k_h is non null. From (7), if the scalar k_h is non null, we can estimate

an EDR direction from a random sample $\{(x_i, y_i), i = 1, \dots, n\}$, using the following estimator:

$$\hat{b}_{h,n} := \hat{\Sigma}_n^{-1} \hat{z}_{h,n},$$

with $\hat{z}_{h,n} = \hat{m}_{h,n} - \bar{x}_n$. If k_h is non null, $\hat{b}_{h,n}$ is a \sqrt{n} -consistent estimator of the EDR direction since it can easily be shown that $\hat{b}_{h,n}$ converges to $k_h \beta$ at root n rate.

Note that, contrary to the usual SIR approach, we do not use in this SIRoneslice approach the information from all the H slices. We only focus on one slice \tilde{h} for which $k_{\tilde{h}}$ is non null. In some situations, for instance when the regression model is partially symmetric, the estimator $\hat{b}_{\tilde{h},n}$ obtained for the “best” slice \tilde{h} performs better than \hat{b}_{SIR} . This point is illustrated in the simulation study described in Section 4. One choice of this best slice will be discussed hereafter.

In the following, let us assume that β is such that $\|\beta\|_{\Sigma} = 1$. Then the term k_h is given by $k_h = z'_h \beta$ but depends on the unknown index parameter β . However, from (7), we have $\beta = \frac{1}{k_h} \Sigma^{-1} z_h$, and using $\|\beta\|_{\Sigma} = 1$, we get $\frac{1}{(k_h)^2} z'_h \Sigma^{-1} \Sigma \Sigma^{-1} z_h = 1$ and then

$$(k_h)^2 = \|z_h\|_{\Sigma^{-1}}^2,$$

which can be easily estimated from the sample data. From this result, we can propose an “optimal” slice for the SIRoneslice approach defined by:

$$h^o = \arg \max_h (k_h)^2, \quad (8)$$

and the corresponding population version of the estimator is then

$$\beta_{h^o} := \frac{\Sigma^{-1} z_{h^o}}{\|z_{h^o}\|_{\Sigma^{-1}}}.$$

From a computational point of view, it is easy to estimate each k_h from a sample by substituting z_h and Σ by their sample versions $\hat{z}_{h,n}$ and $\hat{\Sigma}_n$. Then we find that

$$(\hat{k}_{h,n})^2 = \|\hat{z}_{h,n}\|_{\hat{\Sigma}_n^{-1}}^2.$$

From these values, we can define

$$\hat{h}_n^o = \arg \max_h (\hat{k}_{h,n})^2. \quad (9)$$

With this choice of slice, the corresponding $\hat{\Sigma}_n$ -normalized estimator is

$$\hat{\beta}_{\hat{h}_n^o,n} := \frac{\hat{\Sigma}_n^{-1} \hat{z}_{\hat{h}_n^o,n}}{\|\hat{z}_{\hat{h}_n^o,n}\|_{\hat{\Sigma}_n^{-1}}}. \quad (10)$$

By the law of large numbers, it can be shown that $\hat{\beta}_{\hat{h}_n^o,n}$ converges almost surely to $\pm \beta$.

Straightforwardly, an alternative estimator of the direction of β , which is not $\hat{\Sigma}_n$ -normalized, is given by

$$\hat{b}_{\hat{h}_n^o,n} := \hat{\Sigma}_n^{-1} \hat{z}_{\hat{h}_n^o,n}. \quad (11)$$

In Theorem 3.1, we show that this estimator converges almost surely at $\sqrt{\log(\log n)/n}$ rate to

$$b_{h^\circ} := \Sigma^{-1} z_{h^\circ} \quad (12)$$

which is collinear to β under model (1) and the LC.

3 Recursive versions of SIR and “one slice-based SIR” approaches

We first present in Section 3.1 the recursive versions of the SIR methods and we give recursive expressions of \bar{x}_n , $\hat{\Sigma}_n$, $\hat{\Sigma}_n^{-1}$, $\hat{p}_{h,n}$ and $\hat{z}_{h,n}$. Then, we define in Section 3.2 the recursive SIR estimator of the EDR direction as the major eigenvector of the recursive expression of the matrix of interest $\hat{\Sigma}_n^{-1} \hat{\Gamma}_n$. Then, we propose in Section 3.3 the recursive version of the SIRoneslice estimator. The almost surely convergence at $\sqrt{\log(\log n)/n}$ rate of the SIRoneslice estimator is obtained in Section 3.4.

3.1 Description of the recursive approach

Let us consider that the sample $\{(x_i, y_i), i = 1, \dots, n\}$ is splitted into two subsets: the subsample of the first $(n - 1)$ observations $\{(x_i, y_i), i = 1, \dots, n - 1\}$ and the new observation (x_n, y_n) .

Let us first give the recursive form of the estimators \bar{x}_n , $\hat{\Sigma}_n$ and $\hat{\Sigma}_n^{-1}$ of μ , Σ and Σ^{-1} .

The recursive expression of the empirical mean \bar{x}_n is the following:

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = \frac{n-1}{n} \bar{x}_{n-1} + \frac{1}{n} x_n = \bar{x}_{n-1} + \frac{1}{n} \Phi_n \quad (13)$$

where $\Phi_n = x_n - \bar{x}_{n-1}$.

In the first term, we can observe the presence of \bar{x}_{n-1} , the empirical mean of the first $(n - 1)$ observations, and in the second one, the presence of the n th observation x_n . Similarly, the recursive form of the empirical covariance matrix $\hat{\Sigma}_n$ is given by:

$$\begin{aligned} \hat{\Sigma}_n &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)' \\ &= \frac{n-1}{n} \hat{\Sigma}_{n-1} + \frac{n-1}{n^2} (x_n - \bar{x}_{n-1})(x_n - \bar{x}_{n-1})' \\ &= \frac{n-1}{n} \hat{\Sigma}_{n-1} + \frac{n-1}{n^2} \Phi_n \Phi_n'. \end{aligned} \quad (14)$$

It can be shown, via Riccati equation, that the inverse of $\hat{\Sigma}_n$ has the following recursive expression:

$$\hat{\Sigma}_n^{-1} = \frac{n}{n-1} \hat{\Sigma}_{n-1}^{-1} - \frac{n}{(n-1)(n+\rho_n)} \hat{\Sigma}_{n-1}^{-1} \Phi_n \Phi_n' \hat{\Sigma}_{n-1}^{-1} \quad (15)$$

where $\rho_n = \Phi_n' \hat{\Sigma}_{n-1}^{-1} \Phi_n$.

Let us now give the recursive form of the estimators $\hat{p}_{h,n}$ and $\hat{m}_{h,n}$ of p_h and m_h . Denote by h^* the slice containing y_n , the n th observation of y . We can now give the recursive expression of the estimator $\hat{p}_{h,n}$ of p_h :

$$\hat{p}_{h,n} = \begin{cases} \frac{n-1}{n}\hat{p}_{h^*,n-1} + \frac{1}{n} & \text{if } h = h^*, \\ \frac{n-1}{n}\hat{p}_{h,n-1} & \text{otherwise.} \end{cases} \quad (16)$$

The recursive form of $\hat{m}_{h,n}$ is given by:

$$\hat{m}_{h,n} = \begin{cases} \hat{m}_{h^*,n-1} + \frac{1}{\hat{n}_{h^*,n-1}+1}\Phi_{h^*,n} & \text{if } h = h^*, \\ \hat{m}_{h,n-1} & \text{otherwise,} \end{cases} \quad (17)$$

where $\Phi_{h^*,n} = x_n - \hat{m}_{h^*,n-1}$.

From (13) and (17), we obtain the recursive form of $\hat{z}_{h,n}$:

$$\hat{z}_{h,n} = \begin{cases} \hat{z}_{h^*,n-1} - \frac{1}{n}\Phi_n + \frac{1}{\hat{n}_{h^*,n-1}+1}\Phi_{h^*,n} & \text{if } h = h^*, \\ \hat{z}_{h,n-1} - \frac{1}{n}\Phi_n & \text{otherwise.} \end{cases} \quad (18)$$

3.2 Recursive SIR

Using the recursive forms of \bar{x}_n , $\hat{p}_{h,n}$, $\hat{\Sigma}_n^{-1}$ and $\hat{z}_{h,n}$, we describe the recursive form of the matrix interest $\hat{\Sigma}_n^{-1}\hat{\Gamma}_n$ so that determine the eigenvector associated with the largest eigenvalue of this matrix.

In writing the form $\hat{\Gamma}_n$

$$\hat{\Gamma}_n = \sum_{h \neq h^*} \hat{p}_{h,n}\hat{z}_{h,n}\hat{z}'_{h,n} + \hat{p}_{h^*,n}\hat{z}_{h^*,n}\hat{z}'_{h^*,n},$$

we can deduce from (16) and (18) the following recursive form for $\hat{\Gamma}_n$:

$$\begin{aligned} \hat{\Gamma}_n &= \frac{n-1}{n}\hat{\Gamma}_{n-1} - \frac{n-1}{n^2} \sum_{h=1}^H \hat{p}_{h,n-1} (\hat{z}_{h,n-1}\Phi'_n + \Phi_n\hat{z}'_{h,n-1}) \\ &\quad + \frac{n-1}{n^3}\Phi_n\Phi'_n + \frac{n-1}{n}\hat{p}_{h^*,n-1}A_{h^*,n} + \frac{1}{n}B_{h^*,n}B'_{h^*,n} \end{aligned} \quad (19)$$

with

$$\begin{aligned} A_{h^*,n} &= \frac{1}{\hat{n}_{h^*,n-1}+1} (\hat{z}_{h^*,n-1}\Phi'_{h^*,n} + \Phi_{h^*,n}\hat{z}'_{h^*,n-1}) \\ &\quad - \frac{1}{n(\hat{n}_{h^*,n-1}+1)} (\Phi_n\Phi'_{h^*,n} + \Phi_{h^*,n}\Phi'_n) + \frac{1}{(\hat{n}_{h^*,n-1}+1)^2}\Phi_{h^*,n}\Phi'_{h^*,n} \end{aligned}$$

and $B_{h^*,n} = \hat{z}_{h^*,n-1} - \frac{1}{n}\Phi_n + \frac{1}{\hat{n}_{h^*,n-1}+1}\Phi_{h^*,n}$.

From (15) and (19), we obtain the recursive form of the matrix interest $\hat{\Sigma}_n^{-1}\hat{\Gamma}_n$:

$$\hat{\Sigma}_n^{-1}\hat{\Gamma}_n = \hat{\Sigma}_{n-1}^{-1}\hat{\Gamma}_{n-1} + \hat{\Sigma}_{n-1}^{-1}C_{h^*,n} - \frac{1}{n+\rho_n}\hat{\Sigma}_{n-1}^{-1}\Phi_n\Phi'_n\hat{\Sigma}_{n-1}^{-1} [\hat{\Gamma}_{n-1} + C_{h^*,n}] \quad (20)$$

where

$$\begin{aligned}
C_{h^*,n} &= -\frac{1}{n} \sum_{h=1}^H \hat{p}_{h,n-1} (\hat{z}_{h,n-1} \Phi'_n + \Phi_n \hat{z}'_{h,n-1}) \\
&\quad + \frac{1}{n^2} \Phi_n \Phi'_n + \hat{p}_{h^*,n-1} A_{h^*,n} + \frac{1}{n-1} B_{h^*,n} B'_{h^*,n}.
\end{aligned}$$

Finally, for a multiple indices model, we obtain an estimated basis of E spanned by the eigenvectors associated with the K largest eigenvalues of the estimate $\widehat{\Sigma}_n^{-1} \widehat{\Gamma}_n$ of $\Sigma^{-1} \Gamma$.

3.3 A recursive version of the SIRoneslice approach

We gave in (10) and (11) the non-recursive expression of the estimators $\hat{\beta}_{\hat{h}_n^o,n}$ and $\hat{b}_{\hat{h}_n^o,n}$. From (15) and (18), it is possible to obtain their corresponding recursive expression. We only detail in this paper the recursive expression of $\hat{b}_{\hat{h}_n^o,n} = \widehat{\Sigma}_n^{-1} \hat{z}_{\hat{h}_n^o,n}$:

$$\begin{aligned}
\hat{b}_{\hat{h}_n^o,n} &= \frac{n}{n-1} \widehat{\Sigma}_{n-1}^{-1} \hat{z}_{\hat{h}_n^o,n-1} - \frac{1}{n-1} \widehat{\Sigma}_{n-1}^{-1} \Phi_n - \frac{1}{(n-1)(n+\rho_n)} \widehat{\Sigma}_{n-1}^{-1} \Phi_n \Phi'_n \widehat{\Sigma}_{n-1}^{-1} (n \hat{z}_{\hat{h}_n^o,n-1} - \Phi_n) \\
&\quad + \frac{1}{n-1} \Phi'_n \widehat{\Sigma}_{n-1}^{-1} \left(\widehat{\Sigma}_{n-1}^{-1} - \frac{1}{(n+\rho_n)} \widehat{\Sigma}_{n-1}^{-1} \Phi_n \Phi'_n \widehat{\Sigma}_{n-1}^{-1} \right) \mathbb{I}_{[h^*=\hat{h}_n^o]},
\end{aligned}$$

where h^* still designates the slice containing the n th observation. In a care of simplicity, we do not provide the recursive expression of $\hat{\beta}_{\hat{h}_n^o,n}$ since this expression is a cumbersome and very complicated expression.

3.4 An asymptotic result for SIRoneslice

The asymptotic result needs the following additional assumptions:

- (A₁) The observations $(x_i, y_i), i = 1, \dots, n$, are sampled independently from the model (2).
- (A₂) The support of y is partitioned into H fixed slices s_1, \dots, s_H such that $p_h \neq 0, \forall h = 1, \dots, H$.
- (A₃) $\exists ! h^o$ such that $k_{h^o} > k_h$.

In the following, let $\| \cdot \|$ denote the usual Euclidean norm.

Theorem 3.1 *Under model (1), LC and assumptions (A₁), (A₂) and (A₃), we have for n large enough*

$$\| \hat{b}_{\hat{h}_n^o,n} - b_{h^o} \| = \mathcal{O} \left(\sqrt{\frac{\log(\log n)}{n}} \right) \quad a.s.$$

where b_{h^o} , defined in (12), is collinear to β .

The proof of this Theorem is given in Appendix B. Note that in the asymptotic context of this theorem, the selected slice \hat{h}_n^o converges to the optimal ‘‘theoretical’’ slice h^o (which is fixed and unknown). With finite samples, \hat{h}_n^o may slightly vary but will provide an optimal ‘‘empirical’’ slice.

4 Simulation study and real data application

In this section, in order to compare the numerical performance of SIRoneslice versus SIR, we will only focus on single index regression model. Nevertheless let us recall that the recursive version of SIR can be straightforwardly applied to multiple indices regression model.

In the simulation study, we will consider the four methods described in the previous sections: SIR, SIRoneslice, and their recursive versions (recursive SIR and recursive SIRoneslice). In Section 4.1, we describe the simulated model which will be used in the simulation study. We compare in Section 4.2 the computational times of the four methods and we will observe that the recursive versions are the fastest ones. Let us recall here that SIR and recursive SIR (resp. SIRoneslice and recursive SIRoneslice) applied on the same data $\{(x_i, y_i), i = 1, \dots, n\}$ provide the same estimations, only the manner to calculate the estimate differs (recursive version or non recursive one). For this reason, we only compare the quality of the estimations obtained from SIR and SIRoneslice in Section 4.3. We also provide in this subsection a naive bootstrap criterion in order to select the number H of slices for SIRoneslice. In Section 4.4, we exhibit the evolution of the quality of the recursive SIRoneslice estimator according to the sample size n .

Finally a real data application is postponed in Section 4.5 in order to show the predictive performance of SIRoneslice versus SIR.

4.1 Presentation of the simulated model

In the simulation study, we consider the following regression model in order to generate simulated datasets:

$$y = (x'\beta)^2 \exp(x'\beta/A) + \varepsilon, \quad (21)$$

where x follows a p -dimensional centered normal distribution with covariance matrix Σ defined below, ε follows the normal distribution $\mathcal{N}(0, \sigma)$ with $\sigma = 1.5$, and $\beta = (1, -1, 2, -2, 0, \dots, 0)'$. The covariance matrix Σ is chosen as follows: $\Sigma = \Lambda'\Lambda + 0.5I_p$ where the p^2 components of the $p \times p$ matrix Λ have been generated from the uniform distribution on $[-2, 2]$, the second term $0.5I_p$ allows to avoid numerical inversion problem for Σ . In this simulation study, for each value p of the dimension, we generate a covariance matrix Σ which will be used for each replication. The parameter A has an influence on the form on the dependence between the index $x'\beta$ and y .

- When the value of A is small (for instance $A = 1$), the model presents “no symmetric dependence” since the exponential part in (21) is predominant. Since each slice is informative on the direction of β , SIR which is based on the H slices will performs pretty well to recover the EDR direction as well as the proposed SIRoneslice which is based on only one (“optimal”) slice.

- When the value of A is medium (for instance $A = 2.5$), the model presents a “moderate symmetric dependence”, the influence of the exponential part tends to disappear in favor of the squared polynomial part. In this case, SIR may not perform well contrary to the SIRoneslice which can recover information on the EDR direction from the “best” selected slice which does not potentially suffer of symmetric dependence.
- When the value of A is high (for instance $A = 5$), the model presents a “strong symmetric dependence” since the squared polynomial part in (21) is predominant. Each slice suffers of symmetry dependence and then the two approaches, SIR and SIRoneslice, will not properly recover the EDR direction, however SIRoneslice may possibly provide suitable estimation obtained from an “optimal” selected slice which escape from symmetric dependence if the number of slices is large enough.

These comments will be illustrated in Section 4.3. From this model, we will generate $n \times p$ datasets for various values of A ($=1, 2.5$ and 5), p ($=5, 10, 15$ and 20) and n ($=300, 600, 900$ and 1200). For each simulated sample, we estimate the direction of β with SIR, recursive SIR, SIRoneslice and/or recursive SIRoneslice. Let us denote by \hat{b} the corresponding estimated EDR direction. Since only the direction of β is identifiable, we use the following efficiency measure in order to evaluate the numerical quality of the estimator:

$$\cos^2(\hat{b}, \beta) = \frac{(\hat{b}'\Sigma\beta)^2}{(\hat{b}'\Sigma\hat{b})(\beta'\Sigma\beta)}.$$

The closer this squared cosine of the angle between \hat{b} and β is to one, the better is the estimation.

All the methods (SIR, SIRoneslice, recursive SIR and recursive SIRoneslice) have been implemented in R. The simulation study has been made with this software. The corresponding codes are available from the authors.

4.2 Comparison of computational times between non recursive and recursive approaches of SIR and SIRoneslice

In this part of the simulation study, we only focus on the computational times of the four estimators SIR, recursive SIR, SIRoneslice and recursive SIRoneslice. For each method and for a given simulated sample of size N , we measure in seconds the computational time needed to calculate the corresponding estimators \hat{b}_n of the EDR direction for n going from $N_0 = 30$ in N , where \hat{b}_n is the estimator only based on the first n observations of the sample. More precisely, the computational time is the global time needed to calculate the $N - N_0 + 1$ estimators: $\hat{b}_{N_0}, \hat{b}_{N_0+1}, \dots, \hat{b}_{N-1}$ and \hat{b}_N .

For various values of p and N , we generate $\mathcal{B} = 100$ replicated samples from model (21) with $A = 2.5$. Then, for each method and each simulated sample, we estimate the EDR direction for various values of H .

In Table 1, we set the number H of slices to 10 and we give the means of computational times and the corresponding standard deviations evaluated on the $\mathcal{B} = 100$ replicated samples for different values of p ($=5, 10,$

15, 20) and N (=300, 600, 900 and 1200). In Table 2, we set the size N to 900 and we give these means and the corresponding standard deviations evaluated on the $\mathcal{B} = 100$ replicated samples for different values of p (=5, 10, 15, 20) and H (=5, 10, 15 and 20). From the reading of these tables, one can exhibit the following comments.

- From Table 1, one can observe that SIR and Sironeslice provide nearly the same computational times in mean with a slight advantage for SIRoneslice which does not require an eigenvalue decomposition. Not surprisingly, the larger are the dimension p or the size N , the larger is the mean of computational times. The recursive versions of SIR and Sironeslice provide very smaller mean of computational times in comparison with their non recursive versions. Moreover the recursive SIRoneslice method is clearly the fastest one: recursive SIRoneslice seems to be twice faster than recursive SIR. One explanation of this gain in term of computational time is certainly due to the recursive calculation of the $p \times p$ matrix inverse $\hat{\Sigma}_n$. Note that when the dimension p increases, the computational time in mean of recursive methods only slightly increases contrary to the non recursive ones: for instance, the computational time is multiplied by around 2 from $p = 5$ to $p = 20$ for the non recursive approaches, whereas it is multiplied by around 1.4 (resp. 1.1) for recursive SIR (resp. recursive SIRoneslice). For all the methods, the computational time increases at the same rate according to size N .
- From Table 2, the main information is that the computational times in mean are penalized by the number H of slices. The recursive approaches are always widely the fastest ones. One can however mention that for SIR and recursive SIR, the computational times are multiplied by around 2 from $H = 5$ to $H = 20$ for all values of p while these computational times are multiplied by around 3 for SIRoneslice and recursive SIRoneslice. This phenomenon is explained by the cost of the research of the “optimal” slice which spends more times for large value of H . Nevertheless, the advantage of the recursive approaches remain very important in terms of computational times: for instance when $p = 20$ and $H = 20$, the recursive SIRoneslice is still around 12 times (resp. around 1.8 times) faster than SIR (resp. recursive SIR).

4.3 SIRoneslice versus SIR

In this part of this simulation, the goal is to study the numerical behavior of SIRoneslice and to compare it with SIR. To do this, we generate $\mathcal{B} = 500$ datasets of size $n = 300$ from model (21) with $p = 10$ and for different values of A (=1, 2.5, 5). For each simulated sample, we estimate the EDR direction with the non recursive versions of SIR and SIRoneslice. We first illustrate the efficiency of the choice of the “optimal” slice \hat{h}_n^o by our proposed criterion on an example when $H = 5$ (arbitrary fixed) and $A = 1$. Then we consider several values of H in order to evaluate the possible influence of this tuning parameter on the quality of the estimators for

		$p = 5$	$p = 10$	$p = 15$	$p = 20$
$N = 300$	SIR	1.790 (0.055)	2.318 (0.012)	2.881 (0.016)	3.486 (0.011)
	recursive SIR	0.502 (0.008)	0.547 (0.009)	0.608 (0.010)	0.708 (0.011)
	SIRoneslice	1.671 (0.007)	2.113 (0.008)	2.608 (0.010)	3.126 (0.016)
	recursive SIRoneslice	0.255 (0.010)	0.260 (0.006)	0.262 (0.002)	0.282 (0.005)
$N = 600$	SIR	3.908 (0.043)	5.099 (0.017)	6.366 (0.009)	7.740 (0.014)
	recursive SIR	1.044 (0.017)	1.147 (0.019)	1.258 (0.014)	1.478 (0.018)
	SIRoneslice	3.580 (0.016)	4.683 (0.014)	5.795 (0.011)	6.976 (0.016)
	recursive SIRoneslice	0.526 (0.006)	0.536 (0.004)	0.546 (0.006)	0.580 (0.002)
$N = 900$	SIR	6.120 (0.009)	8.128 (0.022)	10.235 (0.030)	12.523 (0.030)
	recursive SIR	1.594 (0.015)	1.735 (0.017)	1.921 (0.027)	2.245 (0.021)
	SIRoneslice	5.665 (0.028)	7.487 (0.016)	9.358 (0.015)	11.327 (0.013)
	recursive SIRoneslice	0.796 (0.007)	0.816 (0.006)	0.826 (0.003)	0.881 (0.004)
$N = 1200$	SIR	8.522 (0.014)	11.380 (0.014)	14.390 (0.058)	17.901 (0.045)
	recursive SIR	2.137 (0.018)	2.325 (0.029)	2.564 (0.019)	2.952 (0.024)
	SIRoneslice	7.909 (0.009)	10.537 (0.016)	13.200 (0.013)	16.253 (0.046)
	recursive SIRoneslice	1.071 (0.005)	1.098 (0.010)	1.113 (0.005)	1.187 (0.008)

Table 1: Computational times (in seconds) for calculating estimators \hat{b}_n of the direction of β (for n going from $N_0 = 30$ to N) with SIR, recursive SIR, SIRoneslice and recursive SIRoneslice methods (with $H = 10$): mean and standard deviation in parentheses, calculated on $\mathcal{B} = 100$ replicated samples from the model (21) with $A = 2.5$ and different values of p and N

		$p = 5$	$p = 10$	$p = 15$	$p = 20$
$H = 5$	SIR	4.888 (0.053)	5.824 (0.152)	7.365 (0.060)	9.063 (0.014)
	recursive SIR	1.421 (0.013)	1.554 (0.017)	1.750 (0.018)	2.077 (0.020)
	SIRoneslice	3.534 (0.010)	4.686 (0.007)	5.976 (0.010)	7.318 (0.009)
	recursive SIRoneslice	0.481 (0.001)	0.496 (0.005)	0.504 (0.002)	0.540 (0.005)
$H = 10$	SIR	6.120 (0.009)	8.128 (0.022)	10.235 (0.030)	12.523 (0.030)
	recursive SIR	1.594 (0.015)	1.735 (0.017)	1.921 (0.027)	2.245 (0.021)
	SIRoneslice	5.665 (0.028)	7.487 (0.016)	9.358 (0.015)	11.327 (0.013)
	recursive SIRoneslice	0.796 (0.007)	0.816 (0.006)	0.826 (0.003)	0.881 (0.004)
$H = 15$	SIR	7.489 (0.007)	10.081 (0.022)	12.832 (0.054)	15.697 (0.096)
	recursive SIR	1.766 (0.016)	1.918 (0.019)	2.319 (0.023)	2.470 (0.006)
	SIRoneslice	7.594 (0.050)	9.992 (0.047)	12.457 (0.043)	15.031 (0.091)
	recursive SIRoneslice	1.092 (0.009)	1.117 (0.006)	1.135 (0.008)	1.207 (0.005)
$H = 20$	SIR	9.046 (0.014)	12.251 (0.028)	15.646 (0.032)	19.231 (0.061)
	recursive SIR	1.922 (0.014)	2.082 (0.021)	2.325 (0.021)	2.728 (0.011)
	SIRoneslice	9.628 (0.013)	12.660 (0.042)	15.760 (0.013)	19.006 (0.083)
	recursive SIRoneslice	1.394 (0.005)	1.432 (0.006)	1.459 (0.009)	1.544 (0.007)

Table 2: Computational times (in seconds) for calculating estimators \hat{b}_n of the direction of β (for n going from $N_0 = 30$ to N) with SIR, recursive SIR, SIRoneslice and recursive SIRoneslice methods for different values of H : mean and standard deviation in parentheses, calculated on $\mathcal{B} = 100$ replicated samples from the model (21) with $A = 2.5$, $N = 900$ and different values of p

various values of A . From the obtained results, the parameter H needs to be properly chosen and therefore we provide a naive bootstrap choice for the number of slices for SIRoneslice. Finally we provide simulation results of comparison of SIRoneslice versus SIR when the tuning parameter H is chosen by the bootstrap criterion.

Illustration of the efficiency of the choice of \hat{h}_n^o for SIRoneslice. We only present here an example results obtained from $\mathcal{B} = 500$ replicated samples from the model (21) with $A = 1$ and $n = 300$. For SIRoneslice, we set $H = 5$ slices. We plot on Figure 1(a) the boxplots of the $\mathcal{B} = 500$ values of $\hat{k}_{h,n}^2$ for $h = 1, \dots, 5$. Clearly, for all the \mathcal{B} simulated samples, the value of $\hat{h}_n^o = \arg \max_n (\hat{k}_{h,n}^2)$ is 5. In Figure 1(b), we represent the boxplots of the square cosines $\cos^2(\hat{b}_{h,n}, \beta)$ for $h = 1, \dots, 5$. When $h = \hat{h}_n^o = 5$, the corresponding boxplot is very close to one contrary to those associated with the other values of h . Hence, one can observe that the proposed choice for the “optimal” slice seems to be efficient on this example. We do not exhibit other examples (with different values for A and H) since they provide the same conclusions.

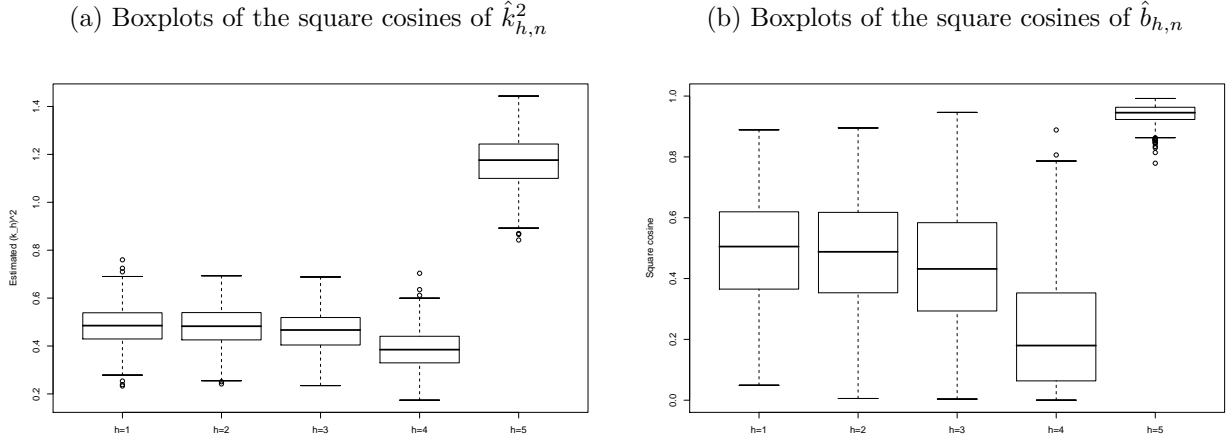


Figure 1: Illustration of the efficiency of the choice of \hat{h}_n^o for SIRoneslice using model (21) with $A = 1$, $n = 300$, $p = 10$ and for a fixed number of slices $H = 5$

Influence of the number H of slices on SIR and SIRoneslice estimates. We consider here $\mathcal{B} = 500$ replicated samples of size $n = 300$ from the model (21) with $p = 10$ and $A = 1, 2.5$ or 5 . For each simulated sample, we estimate the EDR direction with SIRoneslice and SIR for various values of H ($=3, 4, \dots, 10, 15, 20, 25, 30$). For each H , the chosen slice for the SIRoneslice method is selected by the proposed criterion: the slice \hat{h}_n^o is used for estimating the EDR direction. For each value of A , we calculate the mean of the quality measures of each estimator over the $\mathcal{B} = 500$ samples, see Tables 3-5. Moreover we indicate in these tables the percent of times when SIRoneslice dominates SIR. From the reading of Tables 3-5, one can observe

that:

- When the model presents “no symmetric dependence” ($A = 1$), SIRoneslice and SIR methods does not seem to be very sensible to the number H of slices. This is particularly true for SIR and this has been already mentioned by several authors. Even if the means of the square cosines obtained with SIRoneslice are greater than 0.89 for most values of H , a good choice for H (here between 6 and 10) will provide in mean square cosine greater than 0.94.

H	SIRoneslice	SIR	Percent of “SIRoneslice>SIR”
3	0.890	0.888	57.8
4	0.924	0.921	55.0
5	0.935	0.934	52.2
6	0.947	0.946	51.2
7	0.946	0.947	47.2
8	0.946	0.950	43.4
9	0.948	0.952	37.6
10	0.944	0.953	32.8
15	0.925	0.953	17.4
20	0.903	0.952	11.0
25	0.886	0.951	8.0
30	0.866	0.949	4.8

Table 3: Means of quality measures obtained with SIRoneslice and SIR over $\mathcal{B} = 500$ samples generated from model (21) with $A = 1$, $n = 300$, $p = 10$ for different values of H

With this kind of non symmetric dependent model, all slices bring information on the EDR direction. When the number H of slices becomes large, the number of observations in each slice decreases and straightforwardly the quality measures of the SIRoneslice estimates (based only on one slice) will be penalized whereas the SIR estimates still provide high quality measures. Finally, note that for values of H lower than 7 (that is for moderate values of H according to sample size n), half the time SIRoneslice performs better than SIR.

- When the model present a moderate symmetric dependence ($A = 2.5$) or a strong symmetric dependence ($A = 5$), the two methods are sensitive to the choice of the parameter H . In Table 5, for instance the mean of square cosine varies from 0.286 (when $H = 3$) to 0.714 (when $H = 30$) for SIRoneslice. Therefore, a good choice of H appears to be important from a computational point of view. We propose in the next paragraph a naive bootstrap choice for H which provides suitable results.

The symmetric dependence of the model leads to non informative slices to retrieve the EDR direction. When the number of slices is large, we have more chance to obtain an informative slice which is not affected by the symmetric dependence. Note that when $A = 2.5$ or 5, most of the time SIRoneslice performs better

than SIR (with percents around 75% for $A = 5$ and $H \geq 8$) and provides quality measures close to 0.71 (in mean, with $H = 30$) versus 0.57 for SIR.

H	SIRoneslice	SIR	Percent of “SIRoneslice>SIR”
3	0.610	0.605	56.0
4	0.704	0.682	64.2
5	0.781	0.755	75.6
6	0.804	0.774	74.8
7	0.822	0.789	73.6
8	0.846	0.814	72.2
9	0.865	0.841	69.8
10	0.877	0.857	70.6
15	0.891	0.873	61.2
20	0.885	0.870	59.0
25	0.876	0.870	48.6
30	0.862	0.862	47.2

Table 4: Means of quality measures obtained with SIRoneslice and SIR over $\mathcal{B} = 500$ samples generated from model (21) with $A = 2.5$, $n = 300$, $p = 10$ for different values of H

H	SIRoneslice	SIR	Percent of “SIRoneslice>SIR”
3	0.286	0.287	47.0
4	0.382	0.357	56.0
5	0.418	0.374	65.4
6	0.487	0.441	66.0
7	0.533	0.477	71.0
8	0.593	0.504	78.0
9	0.641	0.540	76.0
10	0.566	0.487	72.4
15	0.661	0.536	78.8
20	0.687	0.523	77.4
25	0.701	0.556	79.8
30	0.714	0.574	76.6

Table 5: Means of quality measures obtained with SIRoneslice and SIR over $\mathcal{B} = 500$ samples generated from model (21) with $A = 5$, $n = 300$, $p = 10$ for different values of H

A naive bootstrap choice for the number H of slices for SIRoneslice. We propose hereafter a criterion in order to choose an “optimal” number H of slices for SIRoneslice. This criterion could also be used for SIR. The idea is to provide a value of H such that the quality measure of the estimator remains stable over bootstrap replications of the available sample $s^{(r)} = \{(x_i, y_i), i = 1, \dots, n\}$.

Let B^* be the number of bootstrap replications. For $r = 1, \dots, B^*$, let us consider $\{(x_i^{(r)}, y_i^{(r)}), i = 1, \dots, n\}$ a non parametric bootstrap replication. According to Efron (1982) naive bootstrap estimate of the mean of the

expectation of the quality measure $\cos^2(\hat{b}_{\hat{h}_n^{(r)},n}, \beta)$ is defined by:

$$\hat{g}_n(H) = \frac{1}{B^*} \sum_{r=1}^{B^*} \hat{q}_n^{(r)}$$

with

$$\hat{q}_n^{(r)} = \frac{\left(\hat{b}_{\hat{h}_n^{(r)},n}^{(r)'} \hat{\Sigma}_n \hat{b}_{\hat{h}_n^{(r)},n} \right)^2}{\left(\hat{b}_{\hat{h}_n^{(r)},n}^{(r)'} \hat{\Sigma}_n \hat{b}_{\hat{h}_n^{(r)},n}^{(r)} \right) \left(\hat{b}_{\hat{h}_n^{(r)},n}' \hat{\Sigma}_n \hat{b}_{\hat{h}_n^{(r)},n} \right)}$$

where $\hat{b}_{\hat{h}_n^{(r)},n}^{(r)}$ is the SIRoneslice estimator based on the bootstrap replication sample $s^{(r)}$. In practice this criterion will be computed for values of H from H_{\min} to H_{\max} with H_{\min} and H_{\max} chosen by the user. An optimal number of slices can then be defined as: $\hat{H}_n = \arg \max_H \hat{g}_n(H)$.

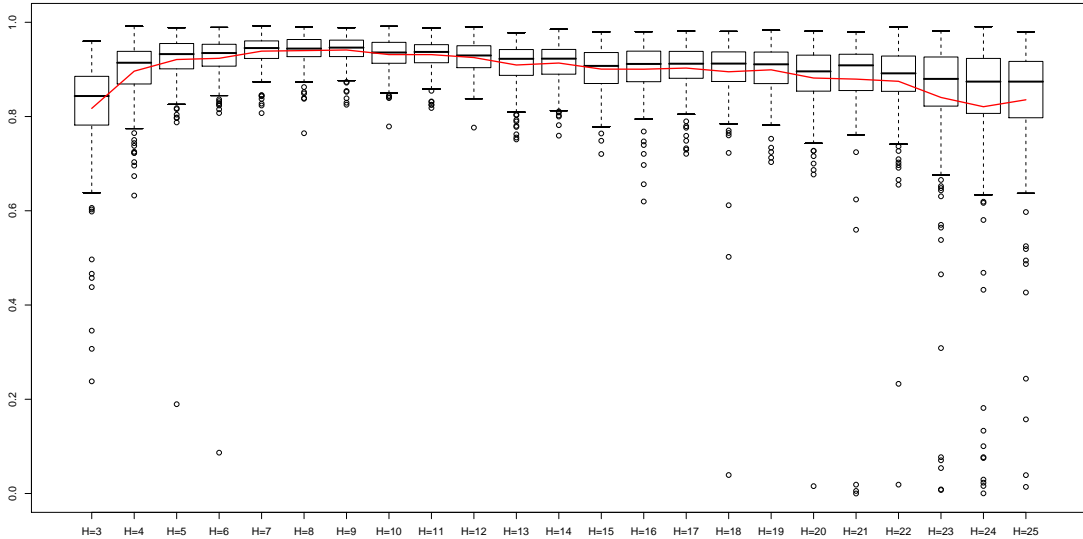


Figure 2: An example of the bootstrap criterion on a simulated sample from model (21) with $n = 300$, $p = 10$ and $A = 2.5$

To illustrate this numerical behavior of this proposed criterion, we first consider a simulated sample of size $n = 300$ from model (21) with $A = 2.5$ and $p = 10$. We set $B^* = 200$ and choose $H_{\min} = 3$ and $H_{\max} = 25$. In Figure 2, for each value of H in $[H_{\min}; H_{\max}]$ the boxplot of the values $\hat{q}_n^{(r)}$, $r = 1, \dots, B^*$ are plotted. One can see that the dispersion of the $\hat{q}_n^{(r)}$'s are large for too small or too large values of H . The solid line links the points $(H, \hat{g}_n(H))$ in order to easily see the values of the criterion. One can observe that the criterion chooses in this example $\hat{H}_n = 7$. Note that the corresponding boxplot of the $\hat{q}_n^{(r)}$'s shows a small dispersion of these values. Let us finally mention that our R code provides the numerical values of $\hat{g}_n(H)$ for all H and the optimal value \hat{H}_n if the user is not interested in graphical representation of the criterion.

In order to show the efficiency of the criterion, we generate 50 samples of size $n = 300$ from model (21) with $p = 10$ and $A = 2.5$. We keep the same parameter values for the criterion: $B^* = 200$, $H_{\min} = 3$ and $H_{\max} = 25$. For each simulated sample, we estimate the EDR direction with SIRoneslice using the optimal value \hat{H}_n obtained with the bootstrap criterion. We also estimate this direction for all values of H in $[H_{\min}; H_{\max}]$, and we select the value of H which provides the maximum value of the quality measure. In Figure 3, we plot the boxplot of the square cosines obtained with SIRoneslice based on \hat{H}_n slices (named “Bootstrap criterion” in the graphic), the boxplot of the “best” square cosines as defined above (named “with best value for H ” in the graphic) and the boxplot of the square cosines obtained with SIRoneslice based on all possible values of H in $[H_{\min}; H_{\max}]$ (named “for various values of H ” in the graphic). Let us mention that the selection of the “best” number H of slices can only be done in a simulation study when the true direction β is known whereas our bootstrap criterion can always be used in practice. One can observe that the quality measures obtained with the proposed bootstrap criterion are very close to the “best” ones. Note that if the choice of H is arbitrary made by the user, the quality of the corresponding estimated direction could behave worse, see the large dispersion of the corresponding boxplot. From these simulation results, the proposed naive bootstrap criterion to select H seems to be useful in practice and we use it in the sequel.

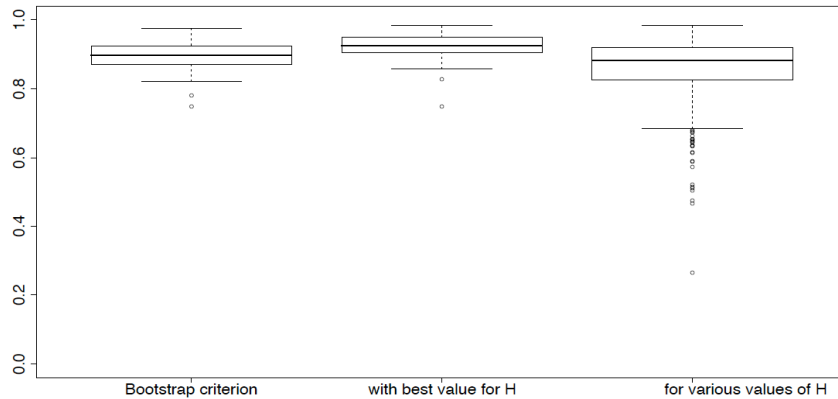


Figure 3: Illustration of the efficiency of the proposed bootstrap criterion via a simulation study using model (21) with $n = 300$, $p = 10$ and $A = 2.5$

SIRoneslice versus SIR. We consider $\mathcal{B} = 500$ replicated samples of size $n = 300$ from the model (21) with $p = 10$ and $A = 1, 2.5$ or 5 . For each simulated sample, we estimate the EDR direction with SIR and SIRoneslice. Note that for the two methods, the number H of slices has been chosen by the proposed bootstrap criterion introduced in the previous paragraph (with $H_{\min} = 3$ and $H_{\max} = 25$).

For each value of A , we plot in Figure 4(a), (c) and (e) the boxplots of the $\mathcal{B} = 500$ values of the corresponding square cosines of SIR and SIRoneslice estimates. In addition, we represent, in Figure 4(b), (d) and (f), the scatter plots of the square cosines of the SIR estimates versus those of SIRoneslice estimates. From this figure, one can point out the following comments.

- When there is no symmetric dependence ($A = 1$), the two methods provide very good estimations with square cosines greater than 0.8. Moreover the quality measures for SIRoneslice and SIR are very close.
- When there is moderate symmetric dependence ($A = 2.5$), SIRoneslice seems to be better than SIR. We clearly see on the scatter plot that the great majority of the point are over the first bisecting line.
- When there is strong symmetric dependence ($A = 5$) the quality measures of the SIRoneslice estimates are largely better than those obtained with SIR. The corresponding scatter plot confirm a uniformly advantage in favor of SIRoneslice. An intuitive explanation of the fact that for a large symmetric dependence SIRoneslice performs most of the time better than SIR is the following. SIRoneslice is often able to find a slice such that z_h is non null and thus the corresponding estimate performs well, while SIR always uses all the slices and the corresponding estimate is then disturbed by the numerous “non-informative” slices in this case.

Dealing with high-dimensional x . We also evaluate the numerical performance of SIRoneslice on high-dimensional data ($p = 50, 100, 150$ and 200). We generate $\mathcal{B} = 500$ simulated samples of size $n = 1000$ from model (21) for three different values of A ($=1, 2.5$ and 5). We present in Figure 5 all the boxplots of square cosines obtained with SIRoneslice and we compare them with the square cosines obtained with classical SIR approach on the same dataset. As in the previous simulation, the number H of slices has been chosen by the proposed bootstrap criterion with $H_{\min} = 3$ and $H_{\max} = 25$ for SIR or SIRoneslice. One can observe that the performance of SIRoneslice and SIR are very similar when $A = 1$ (no symmetric dependence). When the model present a moderate symmetric dependence ($A = 2.5$), SIRoneslice seems to be slightly better than SIR. When $A = 5$ (strong symmetric dependence), SIRoneslice clearly outperforms SIR. Not surprisingly, the quality measure decreases as the dimension p increases and as the symmetric dependence increases.

4.4 Recursive SIRoneslice simulation results

We consider here a samples of size $n = 1200$ from model (21) with $p = 10$ and $A = 1, 2.5$ or 5 . We arbitrary set $H = 15$ slices. For each simulated sample, we estimate for $n = N_0 = 60$ until $N = 1200$, the EDR direction with the recursive SIRoneslice method and we calculate the corresponding quality measures $\cos^2(\hat{b}_{\hat{h}_n^o, n}, \beta)$.

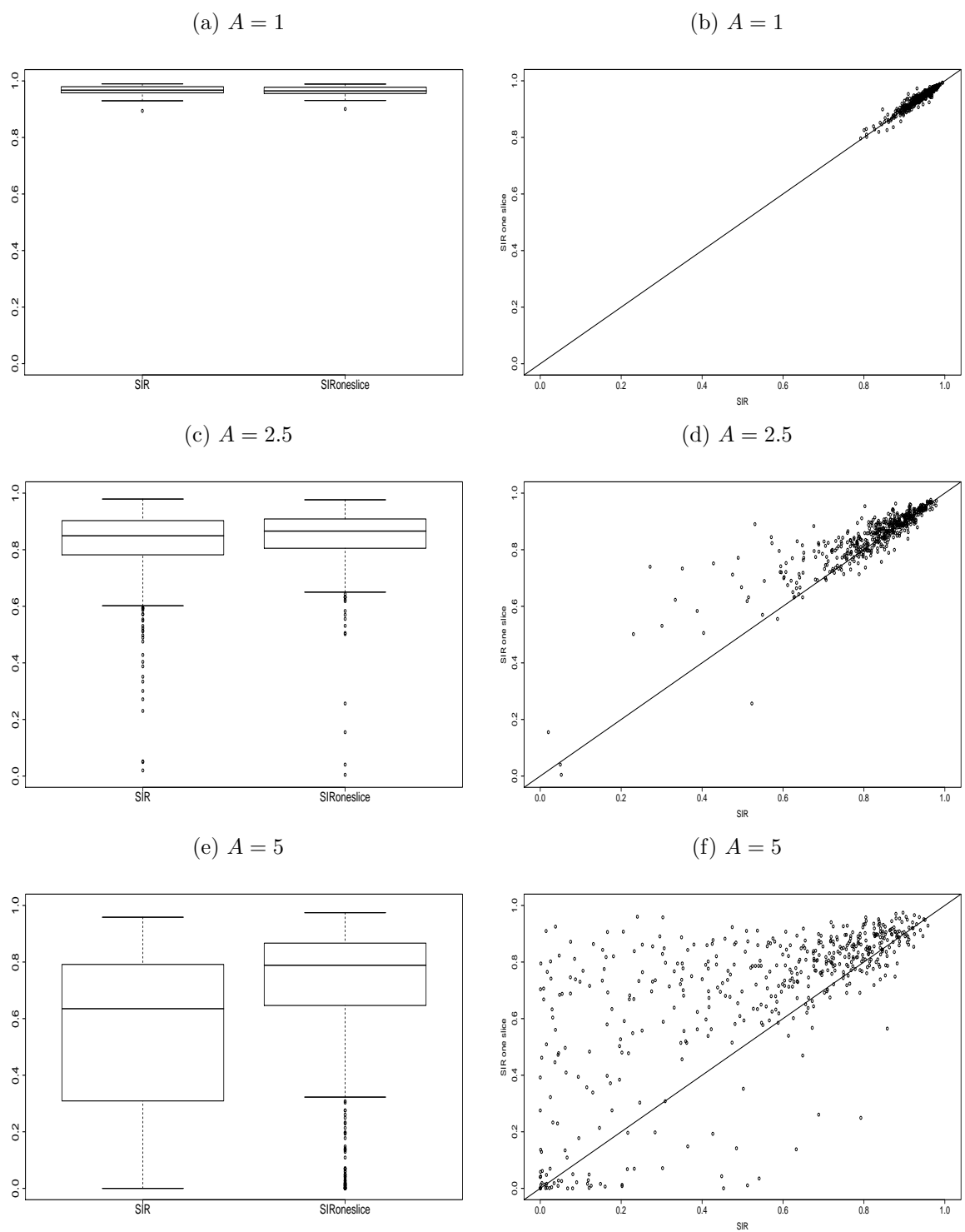


Figure 4: Simulation results obtained with SIR and SIRoneslice for $\mathcal{B} = 500$ samples generated from model (21) with $n = 300$, $p = 10$ and different values of A : boxplots and scatter plots of square cosines

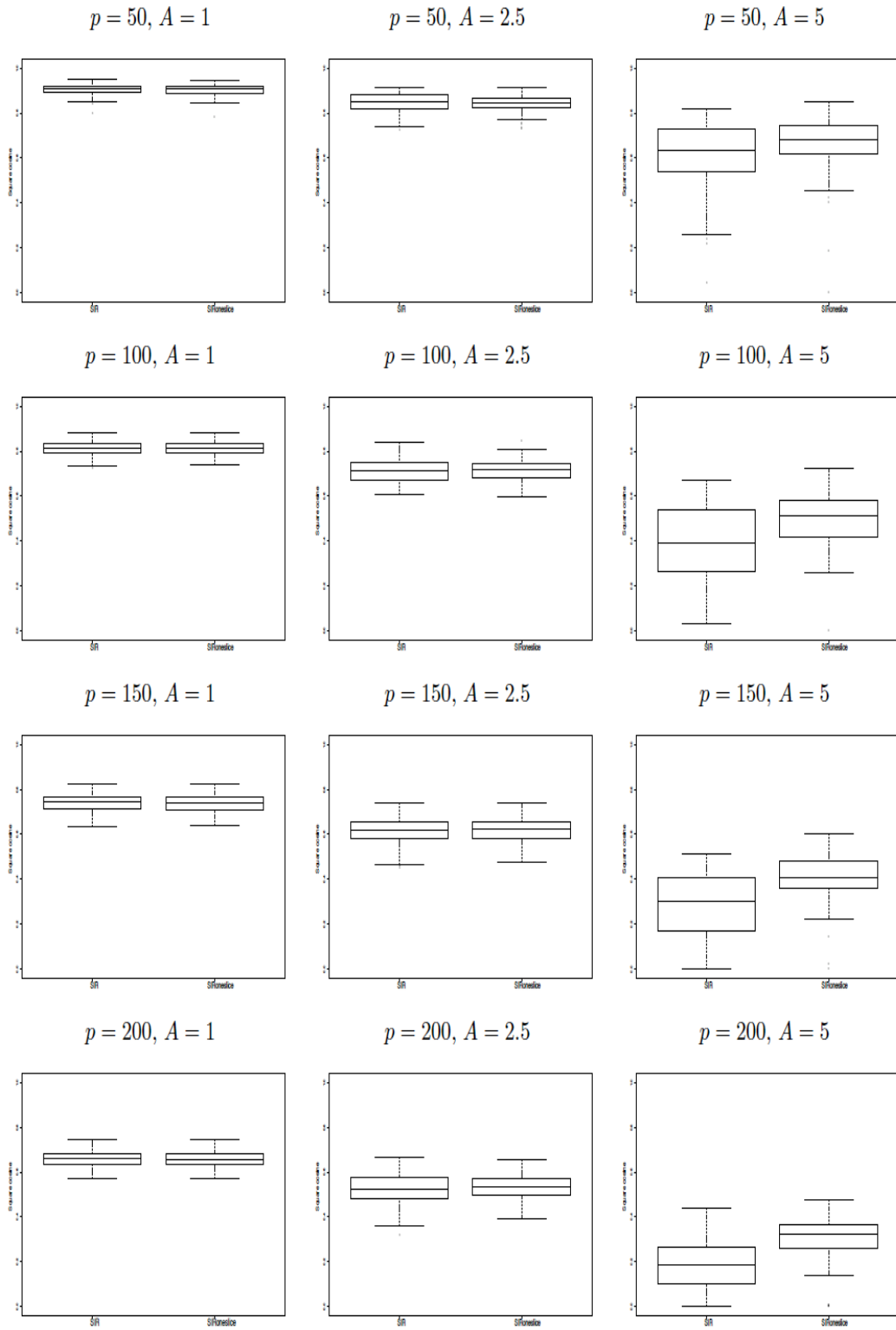


Figure 5: Boxplots of square cosines obtained with SIR and SIRoneslice for $\mathcal{B} = 100$ samples generated from model (21) with high dimensions p of x , different values of A and $n = 1000$

In Figure 6, we represent the evolution of the quality measures $\cos^2(\hat{b}_{\hat{h}_n^o, n}, \beta)$ according to n . Not surprisingly, we clearly observe that the more the sample size n is important, the greater is the quality measure. As previously mentioned, the parameter A have influence on the dependence between the index $x'\beta$ and y in (21) and then effect the quality estimator $\hat{b}_{\hat{h}_n^o, n}$. For instance, in Figure 6, the estimations are clearly better for small sample size when the value of A is small.

In Figure 7, we provide, for some various values of N ($= 100, 200, 300, 600, 900$ and 1200), the boxplots of $\cos^2(\hat{b}_{\hat{h}_n^o, n}, \beta)$ calculated on $\mathcal{B} = 500$ samples simulated from model (21) with $p = 10$ and $A = 2.5$. We observe the same kind of phenomenon as those described in the corresponding case on Figure 6. Since the sample size n is sizeable, the quality of estimations are very good. For example, the boxplots of the $\cos^2(\hat{b}_{\hat{h}_n^o, n}, \beta)$'s are very concentrated around the value 0.95 when $n \geq 600$.

4.5 Real data application

We illustrate our approach on a real data set. We consider the horse mussel data which is prevalent in the literature (see for instance Cook, 1998 or Cook and Weisberg, 1999). The observations correspond to $n = 82$ horse mussels captured in the Malborough Sounds at the Northeast of New Zealand's South Island. These data are available thanks the R package "dr". Following the results obtained by Cook (1998) which used SIR and found an unique EDR direction (via a test procedure), we apply SIRoneslice with the muscle mass (M) in g as response variable and 4 regressors (H, $W^{0.36}$, $S^{0.11}$ and L) where H is the shell height in mm, W is the shell width in mm, S is the shell mass in g and L the shell length in mm. Note that the dependent variable is slightly transformed as follows $y = y + \epsilon, \epsilon \sim \mathcal{N}(0, 0.01^2)$ in order to improve the slicing step of SIR and SIRoneslice. The optimal number of slices has been determined via our bootstrap approach for SIRoneslice and SIR (see Figure 8(a) and (b)). One can observe in Figure 8(b) that SIR is insensitive to the choice of H and we chose $H = 6$ in order to estimate the EDR direction. For SIRoneslice, the bootstrap criterion in Figure 8(a) indicates that a good choice for H can be 2 or 7. We compute the corresponding EDR direction for these two values of H and the square cosine between these two directions is around 0.95. We plot in Figure 8(c) the scatter plot of the dependent variable versus the estimated EDR indices (obtained with SIRoneslice based on $H = 7$). We add on this plot the kernel estimator of Y given the estimated index, the bandwidth is obtained by cross validation. A strong decreasing link between the muscle mass and the index clearly appears in this graphic. Note that the square cosine between this EDR directions calculated with SIRoneslice and SIR is around 0.98.

Finally we compare the prediction reached on test samples with SIR and SIRoneslice using the following algorithm.

- *Step (i)*. We split the data into two subsets: $S_J = \{(y_j, \mathbf{x}'_j), j \in J\}$ the training sample containing almost

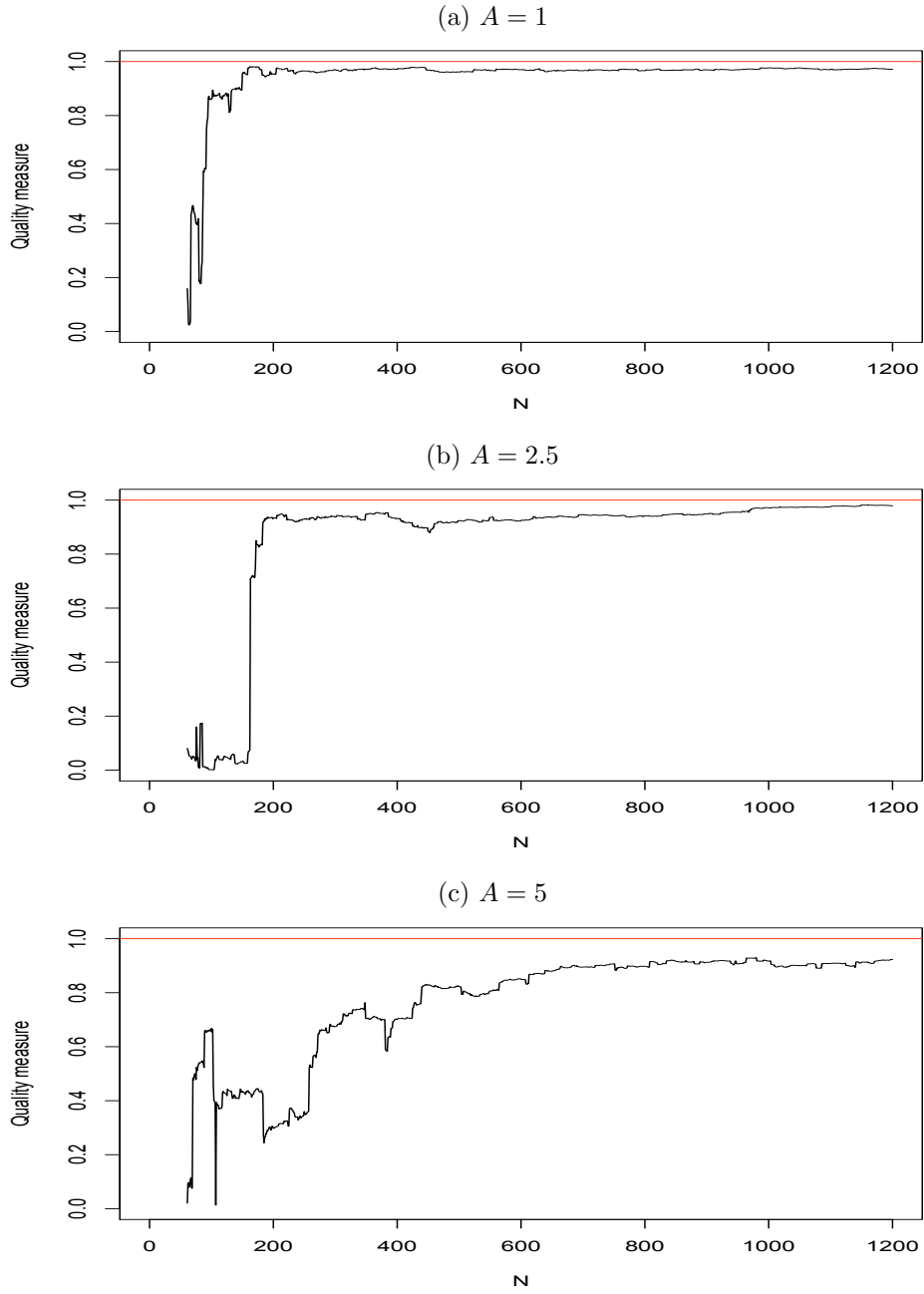


Figure 6: Evolution, according to n , of the quality measures $\cos^2(\hat{b}_{\hat{h}_n^o, n}, \beta)$ obtained with recursive SIRoneslice for three simulated samples from model (21) with $p = 10$ and various values of A .

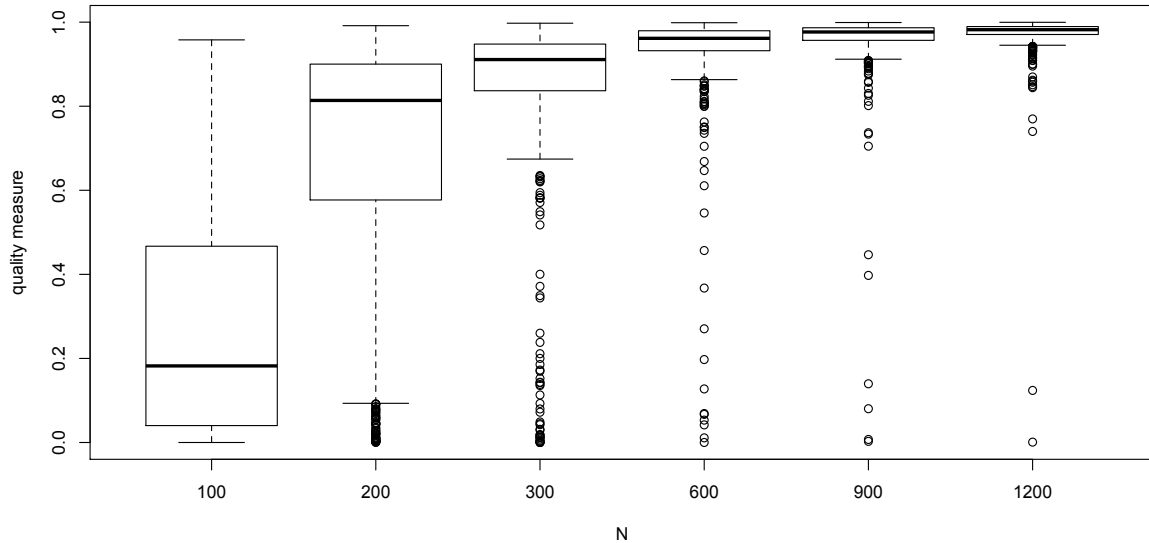


Figure 7: Simulation results obtained with recursive SIRoneslice for $\mathcal{B} = 500$ samples generated from model (21) with $p = 10$ and $A = 2.5$: boxplots of square cosines for various values of n

75% of the total number of observations, and $S_I = \{(y_i, \mathbf{x}'_i), i \in I\}$ the test sample of the remaining observations. Let $n_J = \text{card}(J) = 62$ and $n_I = \text{card}(I) = 20$.

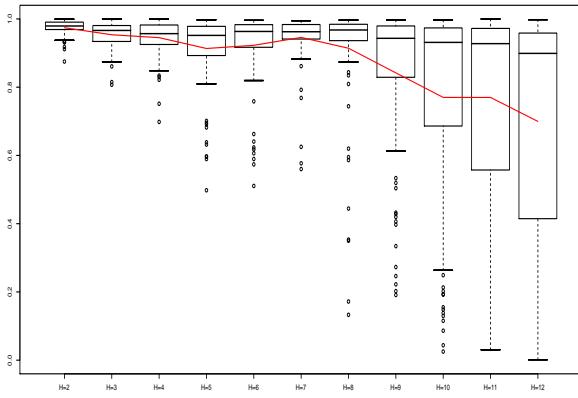
- *Step (ii)*. We use the training sample S_J to compute the estimated EDR direction with SIR, denoted \hat{b}_{SIR} , and with SIRoneslice, denoted $\hat{b}_{\text{SIRoneslice}}$.
- *Step (iii)*. We compute the kernel estimate \hat{y}_i of $\mathbb{E}(y|\mathbf{x}'_i\hat{b})$ for $i \in I$ using the sample $\{(y_j, \mathbf{x}'_j\hat{b}), j \in J\}$ where \hat{b} is either $\hat{b}_{\text{SIRoneslice}}$ or \hat{b}_{SIR} . Thus we get for $i \in I$, $\hat{y}_{i,\text{SIR}}$ for SIR and $\hat{y}_{i,\text{SIRoneslice}}$ for SIRoneslice.
- *Step (iv)*. We compute the Mean Square Relative Error (MSRE) for both SIR and SIRoneslice estimates as follows:

$$\text{MSRE} = \frac{1}{n_I} \sum_{i \in S_I} \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2,$$

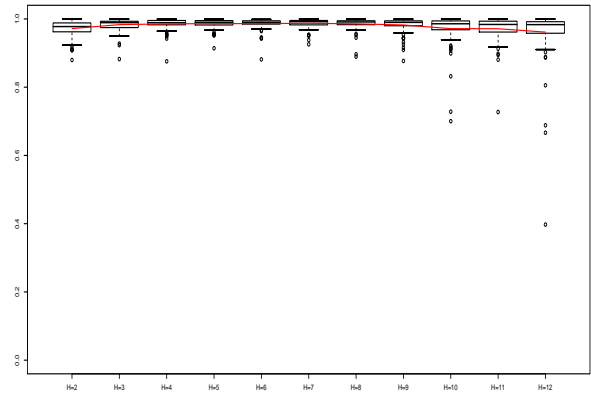
where \hat{y}_i stands for $\hat{y}_{i,\text{SIR}}$ or $\hat{y}_{i,\text{SIRoneslice}}$.

This algorithm is repeated $N = 1000$ times. Figure 8(d) shows the boxplots of the MSRE values obtained with SIRoneslice and SIR. SIRoneslice seems to be more efficient than SIR: the range of the boxplot is smaller with SIRoneslice. Note that nevertheless the median of the MSRE obtained with SIR is lower to the SIRoneslice one.

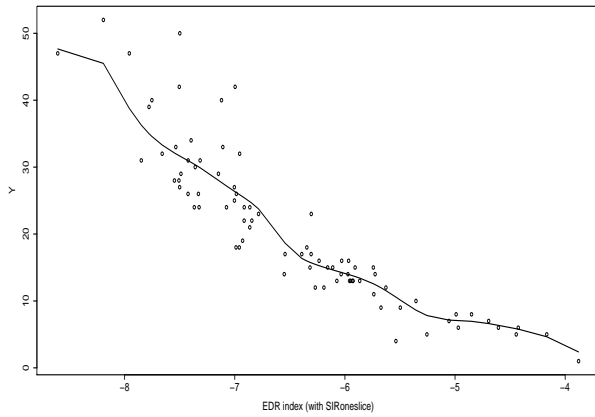
(a) Bootstrap criterion for SIRoneslice



(b) Bootstrap criterion for SIR



(c) Scatter plot for SIRoneslice



(d) Boxplots of MSRE for SIRoneslice and SIR

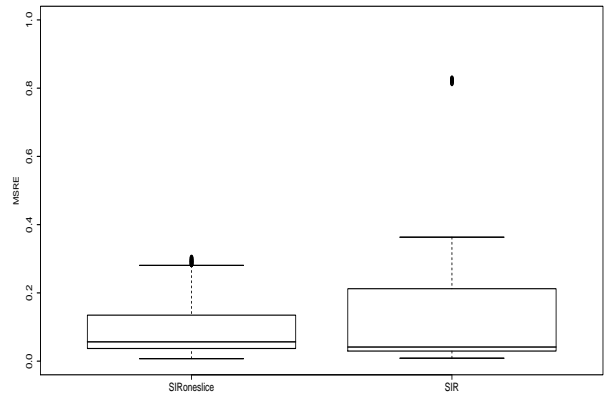


Figure 8: Application on mussels data

5 Concluding remarks

In this paper, we first proposed an estimator of the direction of β based on the use of only one “optimal” slice chosen among the H slices of the slicing step. We called this method SIRoneslice. We also proposed recursive versions of the SIR and SIRoneslice estimators. We showed the almost surely convergence together with a rate of convergence for the SIRoneslice estimator. Finally, we illustrated on simulation the good numerical performances of our SIRoneslice procedure. From a practical point of view, we proposed a naive bootstrap criterion in order to choose the number H of slices. We clearly exhibited the main advantage of using recursive versions of the SIR and SIRoneslice methods from a computational times point of view. The SIRoneslice, recursive SIR and recursive SIRoneslice methods have been implemented in R and the corresponding codes are available from the authors.

Interesting extensions of this work concern the development of recursive versions for other dimension reduction approaches such as SAVE (see for instance Cook, 2000), pHd (see for example Li, 1992) or SIR_α (see for instance Li, 1991, or Gannoun and Saracco, 2003). Moreover, in order to save computational time when the dependent variable y is multivariate, recursive estimators could be also introduced in multivariate SIR methods (see for example Li et al., 2003, Saracco, 2005 or Barreda et al., 2007).

Appendix A

From the LC, we have

$$\mathbb{E}[x \mid x'\beta] = \mu + \frac{\Sigma\beta\beta'(x - \mu)}{\beta'\Sigma\beta}.$$

From model (1) we get

$$\begin{aligned} \mathbb{E}[x \mid T(y)] &= \mathbb{E}\{\mathbb{E}[x \mid x'\beta, T(y)] \mid T(y)\} \\ &= \mathbb{E}\{\mathbb{E}[x \mid x'\beta] \mid T(y)\} \end{aligned}$$

We straightforwardly deduce that

$$\mathbb{E}[x \mid T(y)] = \mu + \frac{\mathbb{E}[\beta'(x - \mu) \mid T(y)]}{\beta'\Sigma\beta}\Sigma\beta.$$

Hence, when T is a slicing, we obtain for the h th slice:

$$m_h := \mathbb{E}[x \mid y \in s_h] = \mu + k_h\Sigma\beta$$

where $k_h = \frac{\mathbb{E}[\beta'(x - \mu) \mid y \in s_h]}{\beta'\Sigma\beta}$.

Appendix B: proof of Theorem 3.1

For $h = 1, \dots, H$, we have $\hat{b}_{h,n} - b_h = (\hat{\Sigma}_n^{-1} - \Sigma^{-1})\hat{z}_{h,n} + \Sigma^{-1}(\hat{z}_{h,n} - z_h)$ where $b_h = \Sigma^{-1}z_h$ and $\hat{b}_{h,n} = \hat{\Sigma}_n^{-1}\hat{z}_{h,n}$. Then, we can deduce the following inequality:

$$\|\hat{b}_{h,n} - b_h\|^2 \leq 2\|\hat{z}_{h,n}\|^2\|\hat{\Sigma}_n^{-1} - \Sigma^{-1}\|^2 + 2\|\Sigma^{-1}\|^2\|\hat{z}_{h,n} - z_h\|^2. \quad (22)$$

In the following proof, we first specify the rate of convergence of $\hat{z}_{h,n}$ based on the rates of convergence of \bar{x}_n and $\hat{m}_{h,n}$. Then, applying Riccati equation for the inverse matrix, we obtain the rate of convergence of $\hat{\Sigma}_n^{-1}$. Finally, we deduce the rate of convergence of $\hat{b}_{h,n}$.

For a matrix M , we denote by $\lambda_{max}(M)$ (resp. $\lambda_{min}(M)$) its largest (resp. smallest) eigenvalue.

Step 1: Study of the rate of convergence of $(\hat{z}_{h,n})$

For $h = 1, \dots, H$, we have $\hat{z}_{h,n} - z_h = (\hat{m}_{h,n} - m_h) - (\bar{x}_n - \mu)$. We clearly have

$$\|\hat{z}_{h,n} - z_h\|^2 \leq 2\|\hat{m}_{h,n} - m_h\|^2 + 2\|\bar{x}_n - \mu\|^2. \quad (23)$$

Study of the rate of convergence of \bar{x}_n

Under assumption (A_1) , from the Hartman-Wintner law of the iterated logarithm (see Theorem 3.2.9, page 136 of Stout,1974) we have

$$\limsup \frac{\|S_n - n\mu\|^2}{2n \log(\log n)} \leq \lambda_{max}(\Sigma) \quad \text{a.s.}$$

where $S_n = \sum_{i=1}^n x_i$. It follows that

$$\limsup \left(\frac{n}{2 \log(\log n)} \right) \|\bar{x}_n - \mu\|^2 \leq \lambda_{max}(\Sigma) \quad \text{a.s.}$$

leading to

$$\|\bar{x}_n - \mu\|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (24)$$

Study of the rate of convergence of $\hat{m}_{h,n}$

We have

$$\hat{m}_{h,n} - m_h = \hat{m}_{h,n} - \frac{n}{\hat{n}_{h,n}} \mathbb{E}[x \mathbb{I}_{[y \in s_h]}] + \frac{n}{\hat{n}_{h,n}} \mathbb{E}[x \mathbb{I}_{[y \in s_h]}] - m_h$$

with

$$\hat{m}_{h,n} = \frac{1}{\hat{n}_{h,n}} \sum_{i=1}^n x_i \mathbb{I}_{[y_i \in s_h]} \quad \text{and} \quad \hat{n}_{h,n} = \sum_{i=1}^n \mathbb{I}_{[y_i \in s_h]}$$

Similarly to the study of \bar{x}_n , we get

$$\|\hat{m}_{h,n} - m_h\|^2 \leq 2 \left\| \hat{m}_{h,n} - \frac{n}{\hat{n}_{h,n}} \mathbb{E}[x \mathbb{I}_{[y \in s_h]}] \right\|^2 + 2 \left\| \frac{n}{\hat{n}_{h,n}} \mathbb{E}[x \mathbb{I}_{[y \in s_h]}] - m_h \right\|^2. \quad (25)$$

Let us focus on the first term. We have

$$\hat{n}_{h,n} \hat{m}_{h,n} = \sum_{i=1}^n x_i \mathbb{I}_{[y_i \in s_h]} = \sum_{i=1}^n \varepsilon_i(h) + n \mathbb{E}[x \mathbb{I}_{[y \in s_h]}]$$

where $\varepsilon_i(h) = x_i \mathbb{I}_{[y_i \in s_h]} - \mathbb{E}[x \mathbb{I}_{[y_i \in s_h]}] = x_i \mathbb{I}_{[y_i \in s_h]} - \mathbb{E}[x \mathbb{I}_{[y \in s_h]}]$. Let $M_n(h) := \sum_{i=1}^n \varepsilon_i(h)$. The sequence $(\varepsilon_i(h))$ is a sequence of independent random variables with $\mathbb{E}[\varepsilon_i(h)] = 0$ et $\mathbb{E}[\varepsilon_i(h) \varepsilon_i'(h)] \leq \Sigma + \mu \mu'$. We deduce again from the Hartman-Wintner law of the iterated logarithm that

$$\|M_n(h)\|^2 = \mathcal{O}\left(n \log(\log n)\right) \quad \text{a.s.}$$

It follows that

$$\left\| \hat{n}_{h,n} \hat{m}_{h,n} - n \mathbb{E}[x \mathbb{I}_{[y \in s_h]}] \right\|^2 = \mathcal{O}\left(n \log(\log n)\right) \quad \text{a.s.}$$

Since we have $\frac{\hat{n}_{h,n}}{n} \xrightarrow{\text{a.s.}} p_h$ where $p_h = \mathbb{P}(y \in s_h) > 0$ (see assumption (A_2)), we finally obtain

$$\left\| \hat{m}_{h,n} - \frac{n}{\hat{n}_{h,n}} \mathbb{E}[x \mathbb{I}_{[y \in s_h]}] \right\|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (26)$$

Let us now study the second term. Since $m_h p_h = \mathbb{E}[x \mathbb{I}_{[y \in s_h]}]$, we have

$$\frac{n}{\hat{n}_{h,n}} \mathbb{E}[x \mathbb{I}_{[y \in s_h]}] - m_h = \frac{n}{\hat{n}_{h,n}} m_h p_h - m_h = \frac{m_h}{\hat{n}_{h,n}} \left(n p_h - \hat{n}_{h,n} \right). \quad (27)$$

Moreover, as $\hat{n}_{h,n} = \sum_{i=1}^n Z_i(h) + n p_h$ with $Z_i(h) = \mathbb{I}_{[y_i \in s_h]} - \mathbb{E}[\mathbb{I}_{[y_i \in s_h]}]$, we have

$$\hat{n}_{h,n} - n p_h = M_n = \sum_{i=1}^n Z_i(h).$$

By the Hartman-Wintner law of the iterated logarithm, we get

$$\left\| \hat{n}_{h,n} - n p_h \right\|^2 = \mathcal{O}\left(n \log(\log n)\right) \quad \text{a.s.}$$

Then, using the fact that $\hat{n}_{h,n} \sim p_h n$, we deduce from (27) that

$$\left\| \frac{n}{\hat{n}_{h,n}} \mathbb{E}[x \mathbb{I}_{[y \in s_h]}] - m_h \right\|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (28)$$

Finally, using the rates obtained in (26) and (28), we deduce from (25) that

$$\| \hat{m}_{h,n} - m_h \|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (29)$$

Study of the rate of convergence of $\hat{z}_{h,n}$

Using (24) and (29), we obtain from (23) that

$$\|\hat{z}_{h,n} - z_h\|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (30)$$

Step 2: Study of the rate of convergence of $(\hat{\Sigma}_n^{-1})$

According to the Riccati equation for the matrix inverse (see for instance Duflo (1997), page 96), we can write $\hat{\Sigma}_n^{-1}$ as

$$\hat{\Sigma}_n^{-1} = \Sigma^{-1} - \Sigma^{-1}(\hat{\Sigma}_n - \Sigma)\Sigma^{-1} + R_n$$

where $R_n = \Sigma^{-1}(\Sigma - \hat{\Sigma}_n)\hat{\Sigma}_n^{-1}(\Sigma - \hat{\Sigma}_n)\Sigma^{-1}$. Straightforwardly, we get

$$\|\hat{\Sigma}_n^{-1} - \Sigma^{-1}\|^2 \leq 2\|R_n\|^2 + 2\|\Sigma^{-1}(\hat{\Sigma}_n - \Sigma)\Sigma^{-1}\|^2. \quad (31)$$

Study of the rate of convergence of $\hat{\Sigma}_n$

Since $\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)'$ $= \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)' - (\bar{x}_n - \mu)(\bar{x}_n - \mu)'$, we can write $\hat{\Sigma}_n - \Sigma = \frac{1}{n} \sum_{i=1}^n e_i - (\bar{x}_n - \mu)(\bar{x}_n - \mu)'$, where $e_i = (x_i - \mu)(x_i - \mu)' - \Sigma$. Therefore, we obtain

$$\|\hat{\Sigma}_n - \Sigma\|^2 \leq \frac{2}{n^2} \left\| \sum_{i=1}^n e_i \right\|^2 + 2\|\bar{x}_n - \mu\|^4. \quad (32)$$

Let $M_N(u) = u' \left(\sum_{i=1}^n e_i \right) u = \sum_{i=1}^n u' e_i u = \sum_{i=1}^n e_i(u)$, where $e_i(u) = u' e_i u = u'(x_i - \mu)(x_i - \mu)' u - u' \Sigma u$.

We have $\mathbb{E}[e_i(u)] = \mathbb{E}[(u'(x_i - \mu)(x_i - \mu)' u - u' \Sigma u)] = u' \Sigma u - u' \Sigma u = 0$ and

$$\begin{aligned} \mathbb{E}[(e_i(u))^2] &= \mathbb{E}[(u'(x_i - \mu)(x_i - \mu)' u - u' \Sigma u)^2] \\ &= \mathbb{E}[u'(x_i - \mu)(x_i - \mu)' u u'(x_i - \mu)(x_i - \mu)' u] - (u' \Sigma u)^2 \\ &\leq \mathbb{E}[(u'(x_i - \mu)(x_i - \mu)' u)^2] \\ &\leq \mathbb{E}[\|x_i - \mu\|^4] \times \|u\|^4. \end{aligned}$$

We can deduce again from the Hartman-Wintner law of the iterated logarithm that

$$\limsup \frac{\left\| \sum_{i=1}^n e_i(u) \right\|^2}{2n \log(\log n)} \leq \tau^4 \times \|u\|^4 \quad \text{a.s.}$$

where $\tau^4 = \mathbb{E}[\|x_i - \mu\|^4]$. It follows that

$$\left\| \sum_{i=1}^n e_i \right\|^2 = \mathcal{O}\left(n \log(\log n)\right) \quad \text{a.s.} \quad (33)$$

From the rates obtained in (24) and (33), we deduce from (32) that

$$\begin{aligned} \|\hat{\Sigma}_n - \Sigma\|^2 &= \mathcal{O}\left(\frac{1}{n^2} n \log(\log n)\right) + \mathcal{O}\left(\left(\frac{\log(\log n)}{n}\right)^2\right) \quad \text{a.s.} \\ &= \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \end{aligned} \quad (34)$$

Study of the rate of convergence of R_n

We deduce from the expression of R_n that

$$\begin{aligned} R_n &\leq \lambda_{max}(\widehat{\Sigma}_n^{-1})\Sigma^{-1}(\Sigma - \widehat{\Sigma}_n)^2\Sigma^{-1} \\ &\leq \frac{1}{\lambda_{min}(\widehat{\Sigma}_n)}\Sigma^{-1}(\Sigma - \widehat{\Sigma}_n)^2\Sigma^{-1}. \end{aligned}$$

Then, we have

$$\begin{aligned} \|R_n\| &\leq \frac{\lambda_{max}(\Sigma^{-2})\|\widehat{\Sigma}_n - \Sigma\|^2}{\lambda_{min}(\widehat{\Sigma}_n)} \\ \|R_n\| &\leq \frac{\|\widehat{\Sigma}_n - \Sigma\|^2}{\lambda_{min}(\widehat{\Sigma}_n)\lambda_{min}(\Sigma^2)}. \end{aligned}$$

From (34), we find that

$$\|R_n\|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n^2}\right) \quad \text{a.s.} \quad (35)$$

Study of the rate of convergence of $\widehat{\Sigma}_n^{-1}$

Using (34) and (35), we straightforwardly deduce from (31) that

$$\|\widehat{\Sigma}_n^{-1} - \Sigma^{-1}\|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (36)$$

Setp 3: Study of the rate of convergence of $\hat{b}_{\hat{h}_n^o, n}$

Finally, from the rates of convergence obtained in (30) et (36), we obtain from (22) that, $\forall h = 1, \dots, H$, we have

$$\|\hat{b}_{h, n} - b_h\|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (37)$$

Obviously, this result still remains true for $h = h^o$. In addition, since we have $\forall h, \hat{k}_{h, n} \rightarrow k_h$, then, under (A_3) , $\exists n_o$ such that $\forall n > n_o, (\hat{k}_{h^o, n})^2 > (\hat{k}_{h, n})^2$. According to the definitions of h^o and \hat{h}_n^o given in (8) and (9), we have $\hat{h}_n^o \rightarrow h^o$. From this result and (37), we obtain that

$$\|\hat{b}_{\hat{h}_n^o, n} - b_{h^o}\|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.}$$

This completes the proof of the Theorem 3.1.

References

- Barreda, L., Gannoun, A. and Saracco, J. (2007). Some extensions of multivariate sliced inverse regression. *J. Stat. Comput. Simul.*, **77**(1-2), 1-17.
- Chen, C.-H., Li, K.-C. (1998). Can SIR be as popular as multiple linear regression? *Statist. Sinica*, **8**(2), 289-316.
- Cook, R.D. (1998). *Regression graphics. Ideas for studying regressions through graphics*. Wiley Series in Probability and Statistics. New York, NY: Wiley.

- Cook, R. D. (2000). SAVE: A method for dimension reduction and graphics in regression. *Comm. Statist. Theory Methods*, **29**, 2109-2121.
- Cook, R. D. and Ni, L. (2005). Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. *Journal of American Statistical Association*, **100**, 410-418.
- Cook, R.D. and Weisberg, S. (1999). *Applied Statistics Including Computing and Graphics*. New York: Wiley.
- Duan, N. and Li, K. C. (1991). Slicing regression: a link-free regression method. *The Annals of Statistics*, **19**, 505-530.
- Duflo, M. (1997). *Random Iterative Models*, Springer-Verlag, Berlin.
- 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.
- Gannoun, A., Saracco, J. (2003). An asymptotic theory for SIR_α method. *Statistica Sinica*, **13**, 297-310.
- Hall, P. and Li, K.-C. (1993). On almost linearity of low-dimensional projections from high-dimensional data. *Ann. Statist.*, **21**(2), 867-889.
- Hsing, T. and Carroll, R. J. (1992). An asymptotic theory for Sliced Inverse regression. *The Annals of Statistics*, **20**, 1040-1061.
- Li, K. C. (1991). Sliced inverse regression for dimension reduction, with dicussion. *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. *J. Amer. Statist. Assoc.*, **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**, 2141-2717.
- Saracco, J. (2005). Asymptotics for pooled marginal slicing estimator based on SIR a approach. *Journal of multivariate Analysis*, **96**, 117-135.
- Stout, W. F. (1974). *Almost sure convergence*. Probability and Mathematical Statistics, Vol. 24. Academic Press , New York-London, 1974.
- Zhu, L. X. and Ng, K. W. (1995). Asymptotics of sliced inverse regression. *The Annals of Statistics*, **24**, 1053-1068.