Pooled marginal slicing approach via SIR a with discrete covariables

Benoit Liquet, Jérôme Saracco

Pooled Marginal Slicing approach via SIR_α with discrete covariables

Benoît Liquet[1] and Jérôme Saracco[2]

[1] LabSAD
BSHM, 1251 avenue centrale, BP 47,
38040 Grenoble Cedex 09, France
e-mail: benoit.liquet@upmf-grenoble.fr

[2] Institut de Mathématiques de Bourgogne,
UMR CNRS 5584,
Université de Bourgogne,
9 avenue Alain Savary, 21 078 Dijon Cedex, France
e-mail: Jerome.Saracco@u-bourgogne.fr

Summary

In this paper, we consider a semiparametric regression model involving both p-dimensional quantitative covariable X and categorical predictor Z, and including a dimension reduction of X via K indices $X'\beta_k$. The dependent variable Y can be real or q-dimensional. We propose an approach based on SIR_α and Pooled Marginal Slicing methods in order to estimate the space spanned by the β_k's. We establish $\sqrt{n}$-consistency of the proposed estimator. Simulation studies show the numerical qualities of our estimator.

Keywords: Discrete Predictor; Semiparametric Regression Model; Sliced Inverse Regression (SIR); Pooled Marginal Slicing (PMS).

[2]Corresponding author.

1 Introduction

Regression analysis studies the relationship between a covariable X and a response variable Y. In parametric regression, the link function is a simple algebraic function of X, and least squares or maximum likelihood methods (among others) can be applied in order to find the best global fit. In nonparametric regression, the class of fitted function is enlarged in order to obtain greater flexibility via sophisticated smoothing procedures (such kernel or smoothing splines methods). However as the dimension p of the covariable X becomes large, increased difficulties in modeling are often encountered. This is the well-known curse of dimensionality. In this framework of high dimensional regression, Li (1991) proposed the following semiparametric dimension reduction model:

$$Y = g(\beta_1' X, \ldots, \beta_K' X, \varepsilon), \tag{1}$$

where the univariate response variable Y is associated with the p-dimensional regressor X (with expectation $\mathbf{E}[X] = \mu$ and covariance matrix $\mathbf{V}(X) = \Sigma$) only through the reduced K-dimensional variable $(\beta_1' X, \ldots, \beta_K' X)$ with $K < p$. The error term ε is independent of X. The link function g and the β-vectors are unknown. We are interested in finding the linear subspace spanned by the K unknown β-vector, called the effective dimension reduction (e.d.r.) space .

Li (1991) introduced Sliced Inverse Regression (named SIR-I in the following) which is a computationally simple and fast method to estimate the e.d.r. space without assuming the functional form of g and the distribution of ε. All the SIR approaches are based on some properties of the conditional distribution of X given Y. The SIR-I method exploits a property of the first inverse moment $\mathbf{E}(X|Y)$; see for instance Duan and Li (1991), Carroll and Li (1992), Hsing and Carroll (1992), Zhu and Ng (1995), Kötter (1996), Saracco (1997, 1999), Aragon and Saracco (1997), Bura and Cook (2001) or Gather et al. (2002) among others. The SIR-II method relies on a property ot the second inverse moment $\mathbf{V}(X|Y)$, which is not blind for "symmetric dependencies"; see for instance Li (1991) or Cook and Weisberg (1991). In order to conjugate information from SIR-I and SIR-II approaches, Li (1991) proposed the SIR_α method which is a suitable combinaison of the methods.

Several authors (see Aragon, 1997, Li et al., 2003) extented the univariate model (1) to a multivariate response variable: Y is assumed to be q-dimensional with $q > 1$, the corresponding link function is then $\Re^q$-valued. In this multivariate framework, the associated dimension reduction model can be written as follows:

$$Y = g(\beta_1' X, \ldots, \beta_K' X, \varepsilon) = \begin{cases} g_1(\beta_1' X, \ldots, \beta_K' X, \varepsilon_1) \\ \vdots \\ g_q(\beta_1' X, \ldots, \beta_K' X, \varepsilon_q) \end{cases} \tag{2}$$

where the error terms ε_j are assumed independent of X and the link functions g_j's are unknown real-valued functions. Clearly, as in (1), only the e.d.r. space is identifiable. A few methods based on SIR-I approach have been developed in this multivariate context; they are named Complete Slicing, Marginal Slicing, Pooled Marginal Slicing and Alternating SIR methods. Barreda et al. (2003) and Saracco (2005) focused on some extensions of the existing multivariate SIR approaches by using the SIR_α method.

Another extension of model (1) is to incorporate a qualitative or categorical predictor Z in addition to the quantitative covariable X. Many covariates (often called factors) are qualitative in the nature such as gender, treatment, type of population, ... Generally, the categorical predictor Z can be viewed as a classification variable with L "levels" which identifies a number of subpopulations. Thus to introduce this qualitative predictor, we consider the following model: for $l = 1, \ldots, L$,

$$Y = g^{(l)}(\beta_1' X, \ldots, \beta_K' X, \varepsilon) \quad \text{when} \quad Z = l. \tag{3}$$

For each subpopulation l, Y is related to the p-dimensional quantitative regressor X only through the K indices $\beta_k' X$. The categorical variable Z is not assumed to be independent of X. It affects the conditional distribution of X given Z as follows: $\mathbf{E}(X|Z = l) = \mu^{(l)}$ and $\mathbf{V}(X|Z = l) = \Sigma^{(l)}$ for $l = 1, \ldots, L$. It also influences the dependency between Y and the indices $\beta_k' X$ via the different link function $g^{(l)}$ associated with each subpopulation l. In other words, a statement equivalent to (3) is that Y and (X, Z) are independent conditionally on $(\beta_1' X, \ldots, \beta_K' X, Z)$.

In a similar dimension reduction model context with binary regressor, Carroll and Li (1995) presented a new look at treatment comparisons. They considered the covariable Z as the treatment indicator and the proposed model is $Y = g(\beta' X + \theta Z, \varepsilon)$. Estimates of β and θ are obtained without assuming a functional form for g. Their method is based on the use of SIR-I in order to estimate the direction of β (e.d.r. directions estimated from each subpopulation are combined in order to obtain a final direction e.d.r.), followed by a partial-inverse mean matching method for estimating the treatment effect θ.

When the number L of levels for Z is greater than two, Chiaromonte et al. (2002) considered a similar context to (3) and they presented a *partial* dimension reduction of X, for the regression of Y on (X, Z). They mentioned that this approach need not coincide with *marginal* dimension reduction for the regression of Y on X, nor with *conditional* dimension reduction for the regression of Y on X within the subpopulations indentified by Z. Assuming the simplifying hypothesis that the predictors' covariance structure is the same across subpopulations ($\Sigma^{(l)} = \Sigma^\star,\ l = 1, \ldots, L$), they introduced a corresponding estimation method of the e.d.r. space, based on SIR-I technique and named Partial Sliced Inverse Regression.

In this paper, we consider an extension of model (3) in which Y is assumed to be q-dimensional ($q > 1$), the link functions $g^{(l)}$ are now $\Re^q$-valued as the

function g in model (2). In order to estimate the e.d.r. space, we propose a new method based on the Pooled Marginal Slicing approach via SIR_α method. In section 2, we give an overview of the univariate SIR_α method and describe the corresponding Pooled Marginal Slicing (PMS_α) estimator. In Section 3, we exhibit two estimators, the first one in the homoscedastic case and the second one in the heteroscedastic context. We also establish asymptotic properties for these estimators. Some simulations are presented in Section 4 in order to show the numerical qualities of our estimators. Concluding remarks are given in section 5.

2 The PMS_α estimator without discrete predictors

First, we recall some properties of the univariate SIR_α method. Then, we introduce the corresponding Pooled Marginal Slicing method, named PMS_α hereafter.

2.1 Overview of the univariate SIR_α method

We consider the framework of model (1), that is when $q = 1$ and without discrete predictors. We give an overview of the univariate SIR_α approach. While there are several possible variations, the basic principle of the SIR methods (SIR-I, SIR-II or SIR_α) is to reverse the role of Y and X. Instead of regressing the univariate Y on the multivariate X, the covariable X is regressed on the response variable Y. The SIR-I estimates based on the first moment $\mathbf{E}(X|Y)$ have been studied extensively; see for instance Duan and Li (1991), Li (1991), Carroll and Li (1992), Hsing and Carroll (1992), Zhu and Ng (1995), Kötter (1996), Saracco (1997, 1999), Aragon and Saracco (1997). But this approach is "blind" for symmetric dependencies (see Cook and Weisberg (1991) or Kötter (2000)). Then, SIR-II estimates based on the inverse conditional second moment $\mathbf{V}(X|Y)$ have been suggested; see for instance Li (1991), Cook and Weisberg (1991) or Kötter (2000). Hence these two approaches concentrate on the use of the inverse conditional moments $\mathbf{E}(X|Y)$ or $\mathbf{V}(X|Y)$ to find the e.d.r. space. For increasing the chance of discovering all the e.d.r. directions, the idea of the SIR_α method is to conjugate these informations: if an e.d.r. direction can only be marginally detected by SIR-I or SIR-II, a suitable combination of these two methods may sharpen the result.

Let us now recall the geometric properties of the model (1). In order to conjugate information from the SIR-I and SIR-II approaches, Li (1991) considered, for $\alpha \in [0, 1]$, the eigen-decomposition of the matrix $\Sigma^{-1}M_\alpha$ where $\Sigma = \mathbf{V}(X)$ and $M_\alpha = (1-\alpha)M_I\Sigma^{-1}M_I + \alpha M_{II}$. 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: for a monotonic transformation T of Y,

$M_I = \mathbf{V}(\mathbf{E}(X|T(Y)))$ and $M_{II} = \mathbf{E}\left\{\left(\widetilde{V}_T - \mathbf{E}(\widetilde{V}_T)\right)\Sigma^{-1}\left(\widetilde{V}_T - \mathbf{E}(\widetilde{V}_T)\right)'\right\}$, where $\widetilde{V}_T = \mathbf{V}(X|T(Y))$. It can be shown that the eigenvectors associated with the largest K eigenvalues of $\Sigma^{-1}M_\alpha$ are some e.d.r. directions. Let us remark 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 transformation T, the matrices of interest are now written as $M_I = \sum_{h=1}^{H} p_h(m_h - \mu)(m_h - \mu)'$ and $M_{II} = \sum_{h=1}^{H} p_h\left(V_h - \overline{V}\right)\Sigma^{-1}\left(V_h - \overline{V}\right)$, where $p_h = P(Y \in s_h)$, $m_h = \mathbf{E}(X|Y \in s_h)$, $V_h = \mathbf{V}(X|Y \in s_h)$ and $\overline{V} = \sum_{h=1}^{H} p_h V_h$.

So, 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 e.d.r. directions. Each estimated e.d.r. direction converges to an e.d.r. direction at rate $\sqrt{n}$ when the corresponding eigenvalues are assumed to be distinct, see for instance Li (1991) or Saracco (2001). Asymptotic normality of the SIR_α estimates has been studied by Gannoun and Saracco (2003a).

Remarks. 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. In order to avoid the choice of a slicing, 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 and Cook (2001) proposed a parametric version of SIR-I. Determining the number K (of indices) is considered by Li (1991), Schott (1994) and Ferré (1998), for the SIR-I method.

The practical choice of α can be based on the test approach proposed by Saracco (2001), which does not require the estimation of the link function. Two cross-validation criteria have been also developed by Gannoun and Saracco (2003b) to select the parameter α, these criteria require the kernel smoothing estimation of the link function.

Note that one crucial condition for the success of SIR-I method is:

$$\mathbf{E}(b'X|\beta_1'X, \ldots, \beta_K'X) \text{ is linear for any } b. \tag{4}$$

This assumption will hold if the distribution of X is elliptically symmetric. The classical example of an elliptically symmetric distribution is the multivariate normal distribution. It does not seem possible to verify (4), this involves the unknown directions of main interest as a start. As Li (1991)

pointed out, this linearity condition is not a severe restriction. Using a bayesian argument of Hall and Li (1993), we can infer that (4) holds approximatively for many high dimensional data sets.

For SIR-II and SIR$_\alpha$ methods, X is generally assumed to have a multivariate normal distribution. Alternatively, note that another commonly used assumptions in the literature are as follows: (i) the condition of the elliptical symmetry made under the design condition (4), and (ii) the conditional variance $\mathbf{V}(X|\beta_1'X,\ldots,\beta_K'X)$ is non-random.

2.2 Pooled Marginal Slicing estimator based on SIR$_\alpha$

We consider here the multivariate framework of model (2) without discrete predictors. We present a short description of the Pooled Marginal Slicing based on the SIR$_\alpha$ approach; see Saracco (2005) for details. Let Y^j denote the jth component of Y.

The idea of the Pooled Marginal Slicing method is to consider the q univariate SIR$_\alpha$ methods of each component Y^j of Y on X (based on a specific slicing T_j) and to combine the corresponding M_α matrices (denoted by $M_{\alpha_j}^{(j)}$) in the following pooling:

$$M_{\alpha,P} = \sum_{j=1}^{q} w_j M_{\alpha_j}^{(j)}, \tag{5}$$

for positive weigths w_j and parameters α_j. Each transformation T_j categorizes each response Y^j into a new response with $H_j > K$ levels; that is the support of each Y^j is partitioned into H_j fixed slices $s_1^{(j)},\ldots,s_{H_j}^{(j)}$. For $j = 1\ldots,q$, the matrices $M_{\alpha_j}^{(j)}$ are defined as follows:

$$M_{\alpha_j}^{(j)} = (1-\alpha_j)M_I^{(j)}\Sigma^{-1}M_I^{(j)} + \alpha_j M_{II}^{(j)},$$

where $M_I^{(j)}$ and $M_{II}^{(j)}$ are respectively the matrices used in the SIR-I and SIR-II approaches corresponding to the component Y^j based on a slicing T_j. For simplicity, assume that X has a multivariate normal distribution, then the eigenvectors, denoted by $b_1,\ldots,b_K$, associated with the largest K eigenvalues of $\Sigma^{-1}M_{\alpha,P}$ are e.d.r. directions.

It is straightforward to estimate the matrices $M_I^{(j)}$ and $M_{II}^{(j)}$ by substituting empirical versions of the moments for their theoretical counterparts, and therefore to obtain the estimation of the e.d.r. directions $\hat{b}_1,\ldots,\hat{b}_K$ which are the eigenvectors associated with the K largest eigenvalues of $\widehat{\Sigma}^{-1}\widehat{M}_{\alpha,P}$.

Remarks. From a practical point of view, as in Aragon (1997), two kinds of weights w_j can be used: equal weights or weights proportional to the major eigeinvalues found by a preliminary univariate SIR$_\alpha$ analysis of each component of Y. In (5), the parameters α_j are individually chosen for each

univariate SIR_α method. For the choice of the α_j's, the method based on the test approach of Saracco (2001) can be used. The asymptotic normality of $\widehat{\Sigma}^{-1}\widehat{M}_{\alpha,P}$ is obtained by Saracco (2005), as well as the asymptotic normality of the eigenprojector on the estimated e.d.r. space, and the asymptotic normality of each estimated e.d.r. direction and its corresponding eigenvalue.

3 Incorporating discrete covariables

In this section, we consider a multivariate extension of model (3) in which Y is assumed to be q-dimensional ($q > 1$). Let Y_i denote the ith observed q-dimensional vector Y. Let Y_i^j be the jth component of Y_i. Let us now introduce the multivariate semiparametric regression model: for $l = 1, \ldots, L$,

$$Y = \begin{cases} Y^1 = g_1^{(l)}(\beta_1' X, \ldots, \beta_K' X, \varepsilon_1^{(l)}) & \text{when } Z = l, \\ \vdots & \\ Y^q = g_q^{(l)}(\beta_1' X, \ldots, \beta_K' X, \varepsilon_q^{(l)}) & \text{when } Z = l. \end{cases} \tag{6}$$

The quantitative predictor $X \in \Re^p$ is the covariable with respect to which we will perform dimension reduction, while the discrete predictor Z is an additional categorical covariable that is not included in the reduction of the dimension. This covariable may represent one or more discrete covariables that identify L subpopulations.

For simplicity, from now on we assume that $K = 1$, thus we focus on the characterization and the estimation of an e.d.r. direction b colinear to β.

Conditions. As in the standard SIR approaches, a design condition is required for the consistency of the method. In our context, let us assume that X is elliptically symmetric for each subpopulation. Note that we then get the following linear conditions for each subpopulation:

$$\forall v \in \Re^p, \quad \mathbf{E}[v'X|\beta'X, Z = l] \text{ is linear in } \beta'X \text{ for each } l = 1, \ldots, L.$$

Moreover, an additional condition is necessary for the consistency of the method: we assume that the conditional variance $\mathbf{V}(X|\beta'X, Z = l)$ is non-random for each subpopulation. Alternatively, note that we can also make the assumption that, for each subpopulation l, X has a multivariate normal distribution.

Estimator of the e.d.r. space. We define a general estimator of the e.d.r. space. The idea of the estimator is based on the PMS_α approach. We propose to pool the "marginal" matrix obtained for each component Y^j of Y and for each level l of Z. Then, the population version of the pooled matrix

of interest is defined as follows:

$$\mathcal{M}_{q,L}^{P} = \sum_{j=1}^{q} \tilde{w}_q^{(j)} \left\{ \sum_{l=1}^{L} w_L^{(l)} \left(\Sigma^{(l)} \right)^{-1} M_{\alpha^{(j,l)}}^{(j,l)} \right\}, \tag{7}$$

where the matrices $M_{\alpha^{(j,l)}}^{(j,l)}$ are the M_α matrices corresponding to the subpopulation l for the component Y^j of Y, that is the SIR_α matrix defined for the pairs (X, Y^j) given $Z = l$. The weights $\{w_L^{(l)},\ l = 1, \ldots, L\}$ are the probability of the events $Z = l$, and the weights $\{\tilde{w}_q^{(j)},\ j = 1, \ldots, q\}$ are some positive weights such that $\sum_{j=1}^{q} \tilde{w}_q^{(j)} = 1$. Note that the parameter α of each M_α matrix can be individually adapted and is denoted by $\alpha^{(j,l)}$. In the sequel, we use the equal weights $\{\tilde{w}_q^{(j)} = \frac{1}{q},\ j = 1, \ldots, q\}$ for simplicity.

As in Chiaromonte et al. (2002), we will first consider the same simplifying assumption that all the covariance structure of X given $Z = l$ are the same across the L subpopulations:

$$\Sigma^{(l)} = \Sigma^\star, \quad l = 1, \ldots, L. \tag{8}$$

Hereafter, this common covariance assumption will be refered as the "homoscedastic case" in constrast to the more general "heteroscedastic case". In each case, we will describe the specific population and sample version of the matrix of interest (7).

3.1 Homoscedastic case

When $\Sigma^{(l)} = \Sigma^\star$ for each subpopulation $l = 1, \ldots, L$, the matrix $\mathcal{M}_{q,L}^{P}$ defined in (7) can be written this way:

$$(\Sigma^\star)^{-1} M_{q,L}^{P}, \tag{9}$$

where $M_{q,L}^{P} = \sum_{j=1}^{q} \tilde{w}_q^{(j)} \left\{ \sum_{l=1}^{L} w_L^{(l)} M_{\alpha^{(j,l)}}^{(j,l)} \right\}$. Clearly, if the design condition holds within each subpopulation, the eigenvector associated with the largest eigenvalue of $(\Sigma^\star)^{-1} M_{q,L}^{P}$ is an e.d.r. direction.

Sample version. We assume that an independent and identically distributed (i.i.d.) sample

$$\mathcal{S} = \{(X_i, Z_i, Y_i),\ i = 1, \ldots, n\}$$

is available from model (6). In order to get an estimator of the matrix defined in (9), the usual idea of the SIR approaches is to substitute empirical versions of all the moments for their theoretical counterparts.

Let $\mathcal{S}^{(l)} = \{(X_i, Y_i)$ such that $Z_i = l\}$ be the subsample corresponding to the subpopulation l. Let $\widehat{\Sigma}^{(l)}$ be the covariance matrix of X in each

subsample $\mathcal{S}^{(l)}$. An estimate of the common covariance $\Sigma^\star$ is given by: $\widehat{\Sigma}^\star = \sum_{l=1}^{L} \frac{n_l}{n} \widehat{\Sigma}^{(l)}$, where n_l is the size of the subsample $\mathcal{S}^{(l)}$.

Let us also introduce the subsample $\mathcal{S}^{(j,l)} = \{(X_i, Y_i^j) \text{ such that } Z_i = l\}$. Following the usual way of the PMS$_\alpha$ approach, we assume that, for each subpopulation l, the sample support of each component Y^j of Y is partitioned into $H^{(j,l)}$ fixed slices $s_1^{(j,l)}, \ldots, s_h^{(j,l)}, \ldots, s_{H^{(j,l)}}^{(j,l)}$. Let $n_h^{(j,l)}$ be the number of observations in slice h of the subsample $\mathcal{S}^{(j,l)}$ for the Y^j component. For each subsample $\mathcal{S}^{(j,l)}$, we calculate the corresponding intraslice mean vectors and intraslice covariance matrices as follows: for $h = 1, \ldots, H^{(j,l)}$, $j = 1, \ldots, q$ and $l = 1, \ldots, L$,

$$\bar{x}_h^{(j,l)} = \frac{1}{n_h^{(j,l)}} \sum_{i=1}^{n} X_i \mathbf{I}\left[Y_i^j \in s_h^{(j,l)}\right],$$

$$\text{and } \widehat{V}_h^{(j,l)} = \frac{1}{n_h^{(j,l)}} \sum_{i=1}^{n} (X_i - \bar{x}_h^{(j,l)})(X_i - \bar{x}_h^{(j,l)})' \mathbf{I}\left[Y_i^j \in s_h^{(j,l)}\right],$$

where $\mathbf{I}[.]$ is the indicator function. Let us define $\widehat{\overline{V}}^{(j,l)} = \sum_{h=1}^{H^{(j,l)}} \frac{n_h^{(j,l)}}{n_l} \widehat{V}_h^{(j,l)}$. The matrices $M_I^{(j,l)}$ and $M_{II}^{(j,l)}$ are then estimated as follows:

$$\widehat{M}_I^{(j,l)} = \sum_{h=1}^{H^{(j,l)}} \frac{n_h^{(j,l)}}{n_l} \left(\bar{x}_h^{(j,l)} - \bar{x}^{(j,l)}\right) \left(\bar{x}_h^{(j,l)} - \bar{x}^{(j,l)}\right)'$$

and

$$\widehat{M}_{II}^{(j,l)} = \sum_{h=1}^{H^{(j,l)}} \frac{n_h^{(j,l)}}{n_l} \left(\widehat{V}_h^{(j,l)} - \widehat{\overline{V}}^{(j,l)}\right) (\widehat{\Sigma}^\star)^{-1} \left(\widehat{V}_h^{(j,l)} - \widehat{\overline{V}}^{(j,l)}\right)'.$$

Then, we can define the estimators of the matrices $M_{\alpha^{(j,l)}}^{(j,l)}$ and $M_{q,L}^P$ by

$$\widehat{M}_{\alpha^{(j,l)}}^{(j,l)} = (1 - \alpha^{(j,l)}) \widehat{M}_I^{(j,l)} (\widehat{\Sigma}^\star)^{-1} \widehat{M}_I^{(j,l)} + \alpha^{(j,l)} \widehat{M}_{II}^{(j,l)}$$

and

$$\widehat{M}_{q,L}^P = \frac{1}{q} \sum_{j=1}^{q} \left\{ \sum_{l=1}^{L} \frac{n_l}{n} \widehat{M}_{\alpha^{(j,l)}}^{(j,l)} \right\}.$$

Finally, the estimated e.d.r. direction $\hat{b}$ is the eigenvector associated with the largest eigenvalue of $(\widehat{\Sigma}^\star)^{-1} \widehat{M}_{q,L}^P$.

In the sequel, this method is named *PMS$_\alpha$ homo*. When the value of α is fixed at zero by the user, the corresponding method is called *PMS-I homo*.

Asymptotics. Let us assume that $n_h^{(j,l)} \to +\infty$ as $n \to +\infty$. Convergence in probability at root n rate of the estimated e.d.r. direction $\hat{b}$ to the EDR direction b is easy to establish. For each subsample $\mathcal{S}^{(j,l)}$, we have $\widehat{M}_{\alpha^{(j,l)}}^{(j,l)} = M_{\alpha^{(j,l)}}^{(j,l)} + O_p(n^{-1/2})$ for a fixed parameter $\alpha^{(j,l)}$. Then, $\widehat{M}_{q,L}^P = M_{q,L}^P + O_p(n^{-1/2})$. Straightforwardly, since $(\widehat{\Sigma}^\star)^{-1}$ converges to $(\Sigma^\star)^{-1}$, the estimated e.d.r. direction $\hat{b}$ (principal eigenvector of $(\widehat{\Sigma}^\star)^{-1}\widehat{M}_{q,L}^P$) converges to the e.d.r. direction b (principal eigenvector of $(\Sigma^\star)^{-1} M_{q,L}^P$) at the same rate.

The asymptotic normality of $\sqrt{n}\left((\widehat{\Sigma}^\star)^{-1}\widehat{M}_{q,L}^P - (\Sigma^\star)^{-1}M_{q,L}^P\right)$ can be obtained as in Saracco (2005) using the delta method and the central limit theorem. From this result, we could derive the asymptotic normality of the eigenprojector on the e.d.r. space as well as the asymptotic distributions of the estimated e.d.r. direction and its corresponding eigenvalue.

Remarks. Note that the parameters $\alpha^{(j,l)}$ are individually chosen for each matrix $\widehat{M}_{\alpha^{(j,l)}}^{(j,l)}$ which corresponds to the univariate SIR$_\alpha$ method applied on the subsample $\mathcal{S}^{(j,l)}$. One can use the test approach proposed by Saracco (2001) in order to get appropriate values for the $\alpha^{(j,l)}$'s.

In this homoscedastic context and for $q = 1$, the partial sliced inverse regression approach of Chiaromonte et al. (2002) leads to the eigenvalue decomposition of the matrix $\widetilde{M}_L^P = \mathbf{E}[\mathbf{V}(E[\tilde{X}|Y,Z])|Z]$ where $\tilde{X} = (\Sigma^\star)^{-1/2}(X-\mu)$ is the standardized version of X. The eigenvector $\tilde{b}$ associated with the largest eigenvalue of $\widetilde{M}_L^P$ is a standardized e.d.r. direction. Returning to the X-scale, $b = (\Sigma^\star)^{-1/2}\tilde{b}$ is an e.d.r. direction. This characterization is equivalent to our approach when $\alpha = 0$.

3.2 Heteroscedastic case

In the heteroscedastic case, we consider the eigenvalue decomposition of the matrix $\mathcal{M}_{q,L}^P$ defined in (7) under the design condition. The important point to note here is that this matrix have no reason to be symmetric (with respect to a specific inner product) or positive definite. So it is possible to have some complex eigenvalues and eigenvectors. However the crucial fact is that, for each matrix $\left(\Sigma^{(l)}\right)^{-1} M_{\alpha^{(j,l)}}^{(j,l)}$ (with $j = 1,\ldots,q$ and $l = 1,\ldots,L$), the eigenvector associated with the largest eigenvalue is an e.d.r. direction. Then, since the matrix $\mathcal{M}_{q,L}^P$ is a convex combination of the matrices $\left(\Sigma^{(l)}\right)^{-1} M_{\alpha^{(j,l)}}^{(j,l)}$, it is straightforward to see that there exists an eigenvector b which is an e.d.r. direction associated with a real positive eigenvalue λ. From an algebraic point of view, there is no guarantee that this corresponding eigenvalue is the largest one (in descending order of modulus). Geometrically speaking, one can find pathological cases in which the largest eigenvalue is not associated with an e.d.r. direction. More precisely, let A_u and B_u (for

$u = 1, \ldots, U$) be symmetric positive definite matrices. Let v_u be the eigenvector associated with the largest eigenvalue of $A_u B_u$ and let us assume that v_u is colinear to a direction β. Let us consider the convex combination $N = \sum_{u=1}^{U} \gamma_u A_u B_u$ where the γ_u are positive and such that $\sum_{u=1}^{U} \gamma_u = 1$. The eigenvector associated with the largest eigenvalue of N is not necessary the one colinear to β. Unfortunately, to the best of our knowledge, there is no characterization of these pathological cases, nor necessary conditions allowing us to avoid these cases. It is an open problem which is under investigation. It is important to mention that the potential pathological case can only occur when one is underestimating the true dimension K, as it is illustrated in the pathological example given the Appendix. Moreover, we will see that, in our simulation studies (with $K = 1$), we never encounter this problem.

Sample version. In the heteroscedastic case, we have to estimate the matrix $\mathcal{M}_{q,L}^{P}$ by substituting empirical versions of the moments for their theoretical counterparts. Then the corresponding estimate is given by:

$$\widehat{\mathcal{M}}_{q,L}^{P} = \frac{1}{q} \sum_{j=1}^{q} \left\{ \sum_{l=1}^{L} \frac{n_l}{n} \left(\widehat{\Sigma}^{(l)} \right)^{-1} \widehat{M}_{\alpha^{(j,l)}}^{(j,l)} \right\}, \tag{10}$$

where the major modification is that the estimated common covariance matrix $\widehat{\Sigma}^{\star}$ is now replaced by the estimated marginal covariance matrix $\widehat{\Sigma}^{(l)}$ of the subsample $\mathcal{S}^{(l)}$ in the estimated matrices $\widehat{M}_{\alpha^{(j,l)}}^{(j,l)}$ and $\widehat{M}_{II}^{(j,l)}$. More precisely, these matrices are now written this way:

$$\widehat{M}_{\alpha^{(j,l)}}^{(j,l)} = (1 - \alpha^{(j,l)}) \widehat{M}_{I}^{(j,l)} \left(\widehat{\Sigma}^{(l)} \right)^{-1} \widehat{M}_{I}^{(j,l)} + \alpha^{(j,l)} \widehat{M}_{II}^{(j,l)}$$

where $\widehat{M}_{I}^{(j,l)}$ is as defined before and

$$\widehat{M}_{II}^{(j,l)} = \sum_{h=1}^{H^{(j,l)}} \frac{n_h^{(j,l)}}{n_l} \left(\widehat{V}_h^{(j,l)} - \widehat{\overline{V}}^{(j,l)} \right) \left(\widehat{\Sigma}^{(l)} \right)^{-1} \left(\widehat{V}_h^{(j,l)} - \widehat{\overline{V}}^{(j,l)} \right)'.$$

Let $\hat{b}$ be the eigenvector associated with the eigenvalue $\hat{\lambda}$ corresponding to λ. This vector $\hat{b}$ is the estimated e.d.r. direction.

In the sequel, this method is named *PMS$_\alpha$ hetero*. When the value of α is fixed at zero by the user, the corresponding method is called *PMS-I hetero*.

Asymptotics. As in the homoscedastic case, we have the root n consistency of the estimated e.d.r. direction to the corresponding e.d.r. direction: $\hat{b} = b + O_p(n^{-1/2})$.

Remark. For convenience, let us assume that $q = 1$. As in Carroll and Li (1995), an alternative optimization problem which combines the covariance matrix of each subpopulation is the following:

$$\max_b \sum_{l=1}^{L} w_L^{(l)} \frac{b' M_{\alpha^{(l)}}^{(l)} b}{b' \Sigma^{(l)} b}. \tag{11}$$

In the heteroscedastic case, there is no analytic solution for (11), but there exists numerical methods for performing this simultaneous eigendecomposition. In the homoscedastic case, the maximization problem (11) leads to a single eigenvalue decomposition of the matrix $(\Sigma^\star)^{-1} M_L^P$ where M_L^P is the following pooled covariance matrix: $M_L^P = \sum_{l=1}^{L} w_L^{(l)} M_{\alpha^{(l)}}^{(l)}$. The eigenvector associated with the largest eigenvalue is clearly an e.d.r. direction. Note that this characterization is equivalent to our method.
When $K = 1$, in order to get the common direction, there exists other ways such combining the eigenvectors v_t corresponding to the largest eigenvalues of the Lq matrices

$$(\hat{\Sigma}^{(l)})^{-1} \hat{M}_{\alpha^{(j,l)}}^{(j,l)}.$$

An immediate question is how to combine them. There are several ways to proceed. The simplest is to take the average of these vectors, but recall that only the direction is estimated and problems with orientation and norm of the v_t's could arise. A less simple alternative based on a combination via covariance weighting have been proposed by Carroll and Li (1995) for $L = 2$ and $q = 1$. Moreover it may be interesting to study the eigenvector corresponding to the largest eigenvalue of $\sum_t v_t v_t'$ or some other appropriately chosen matrix function of the v_t's. But the choice of a good or optimal function is not straightforward and is still an open problem.

4 Simulation study

The aim of this simulation study is twofold. First, we consider an homoscedastic case and we compare the performance of the PMS_α homo and PMS_α hetero methods versus the PMS-I homo and PMS-I hetero methods. Secondly, we focus on the heteroscedastic case and we evaluate the quality of the PMS_α homo method versus the PMS_α hetero method.

In all simulations, since only the direction of β is identifiable, in order to evaluate the quality of an estimate $\widehat{b}$ of the direction of β, we calculate the following efficiency measure:

$$\cos^2\left(\widehat{b}, \beta\right) = \frac{< \widehat{b}, \beta >^2}{||\widehat{b}||^2 ||\beta||^2}$$

where $||\cdot||$ is the usual euclidian norm associated to scalar product $< \cdot, \cdot >$. The closer this square cosine of the angle between $\widehat{b}$ and β is to one, the better is the estimation.

4.1 Simulation 1

Let us first recall that, for a variable X with mean 0, a symmetrically dependent model is a regression model such that: $Y = f(\beta' X) + \varepsilon$, where the distribution of $\beta' X$ is symmetric around 0 and the link function f is also symmetric around the vertical axis.

In this simulation, we consider an homoscedactic case and we generate simulated data from the following multivariate semiparametric regression model (with $q = 2,\ L = 2$):

$$\begin{cases} Y^1 = [(1-\rho)/8](\beta' X)^2 + \rho(\beta' X) + \varepsilon_1^{(1)} & \text{for } Z = 1, \\ Y^2 = 2(1-\rho)\sqrt{|2\beta' X|} + \rho(\beta' X) + \varepsilon_2^{(1)} & \text{for } Z = 1, \\ & \\ Y^1 = (1-\rho)|\beta' X| + \rho(\beta' X)^3 + \varepsilon_1^{(2)} & \text{for } Z = 2, \\ Y^2 = [(1-\rho)/8](\beta' X)^2 + \rho \exp(\beta' X/4) + \varepsilon_2^{(2)} & \text{for } Z = 2, \end{cases} \tag{12}$$

where $X|Z = l$ (for $l = 1, 2$) follows a 5-dimensional normal distribution with mean $\mu^{(l)} = \mathbf{0}_5$ and common covariance matrix

$$\Sigma^\star = \Sigma^{(1)} = \Sigma^{(2)} = \begin{bmatrix} 5 & 3.5 & 1.5 & 2.5 & 2.5 \\ 3.5 & 12.5 & 4 & 5 & 4 \\ 1.5 & 4 & 9 & 2.5 & 1.5 \\ 2.5 & 5 & 2.5 & 7 & 3 \\ 2.5 & 4 & 1.5 & 3 & 12 \end{bmatrix}.$$

Each $\varepsilon_j^{(l)}$ is standard normally distributed. We take $\beta = (1, 1, -1, -1, 0)'$.

In order to point out the efficiency of the methods based on SIR_α versus the SIR-I methods, we have introduced in the model (12) the parameter ρ which controls the symmetric dependency between the index $\beta' x$ and the response variables Y^j. When $\rho = 0$, all the "submodels" are symmetrically dependent. On the contrary, the symmetric dependency disappears as ρ increases. When $\rho = 1$, there is none symmetric dependency.

From model (12), $N = 100$ samples of size $n = 50$ and $n = 100$ were generated for each value of ρ in the set $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$. For each simulated sample, the e.d.r. direction was estimated with the four methods: PMS-I hetero, PMS-I homo, PMS_α hetero and PMS_α homo. Then, in order to compare the different estimates obtained with each of the four described methods, we have calculated, for each estimation, its corresponding square cosine.

Comments on the plots of the means of the squared cosines. We present on Figure 1, the mean of the $N = 100$ square cosines obtained for each of 6 values of ρ (from 0 to 1) for the four methods.

- For the two sample sizes, both PMS_α methods give reliable estimates of the direction of β for all values of ρ.

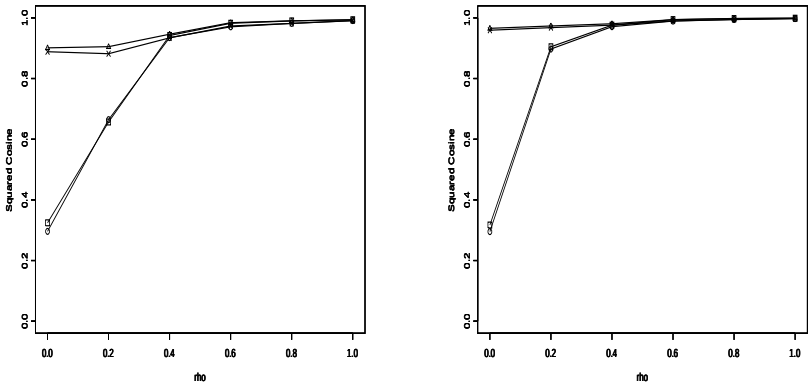

Figure 1: *Plots of the means of the squared cosines for different values of* ρ *and* $n = 50$*'s sample size on the left side and* $n = 100$*'s sample size on the right side. The two upper (resp. lower) curves correspond to the* PMS_α *(resp. PMS-I) hetero and* PMS_α *(resp. PMS-I) homo methods.*

- For both PMS-I methods, the quality of the estimation increases with increasing values of ρ. For $\rho \geq 0.4$, these hetero and homo methods give reliable estimates. The quality of the estimation increases when the sample size is larger. For $\rho = 0.2$ and $n = 100$, these methods give suitable results compared with $n = 50$.

Comments on the boxplots. We represent on Figure 2 the boxplots of the $N = 100$ square cosines obtained with the four methods for different values of ρ (0, 0.2, 0.4, 0.6) and $n = 100$'s sample size.

- For $\rho = 0$ (case of global symmetric dependency), the PMS-I homo and hetero methods can not recover the e.d.r. direction. The PMS_α hetero method gives the best quality of estimations even if the PMS_α homo method gives reliable estimations.
- For $\rho = 0.2$, the PMS_α methods are still better, and the PMS_α hetero also provides the best perfomance.
- For $\rho = 0.4$ or 0.6, the four methods give suitable estimates.
- For $\rho > 0.6$ (case of non-symmetrically dependent model), the results of the simulation have not been reported here because they do not present further of interest: indeed all methods give very good estimations in this context.

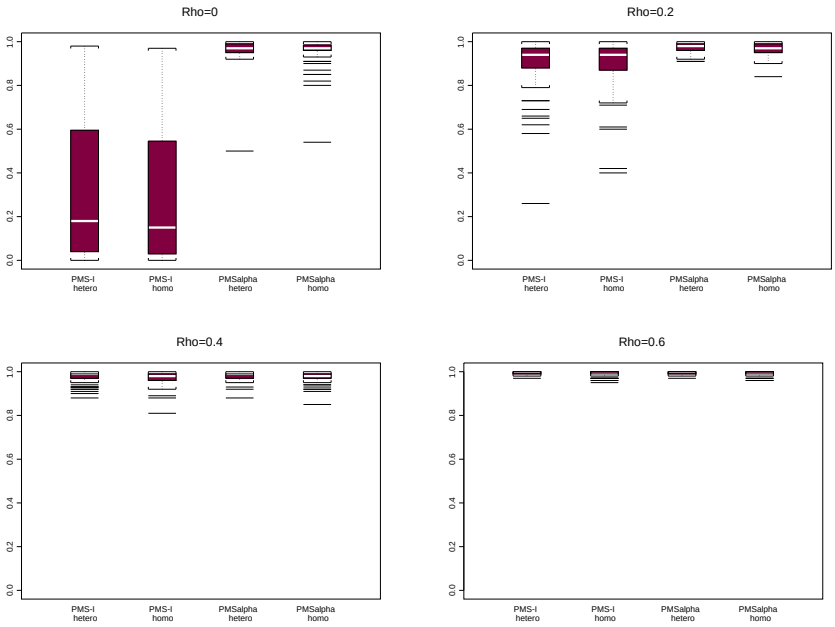

Figure 2: *Boxplots of the squared cosines for different values of* ρ *and* $n = 100$*'s sample size.*

The boxplots corresponding to the simulation for $n = 50$ (not presented here) give same conclusions about the performance of the four method. For PMS-I hetero and PMS-I homo methods, the performance of the estimation increases with increasing values of ρ. Methods based on SIR_α give the best results with a superority of PMS_α hetero method over the other ones.

An illustrated example. A sample of size $n = 100$ were generated from model (12) with $\rho = 0.5$. The e.d.r. direction was estimated with the PMS_α hetero method and the PMS_α homo method. The lists of the eigenvalues are the following: $\hat{\lambda}_{hetero} = \{0.640, 0.021, 0.011, 0.006, 0.004\}$ and $\hat{\lambda}_{homo} = \{0.660, 0.024, 0.013, 0.007, 0.006\}$. Clearly, there is a visible jump between the first and the second eigenvalues, then we retain only one e.d.r. direction. The two methods give excellent estimations with $\cos^2(\widehat{b}, \beta) \simeq 0.999$. The corresponding estimated e.d.r. directions are respectively $\widehat{b}_{hetero} = (0.68, 0.73, -0.70, -0.73, 0.01)$ and $\widehat{b}_{homo} = (-0.63, -0.72, 0.69, 0, 73, 0.01)$. The plots for each subpopulation of the response variables Y^1 and Y^2 versus the index $\beta' X$ are represented on Figure 3. In this figure, we exhibit the kernel estimates of the link functions $g_j^{(l)}$ between the variables of interest Y^j when $Z = l$ and the common estimated index. We used the gaussian kernel

and the bandwidths were chosen by cross validation.

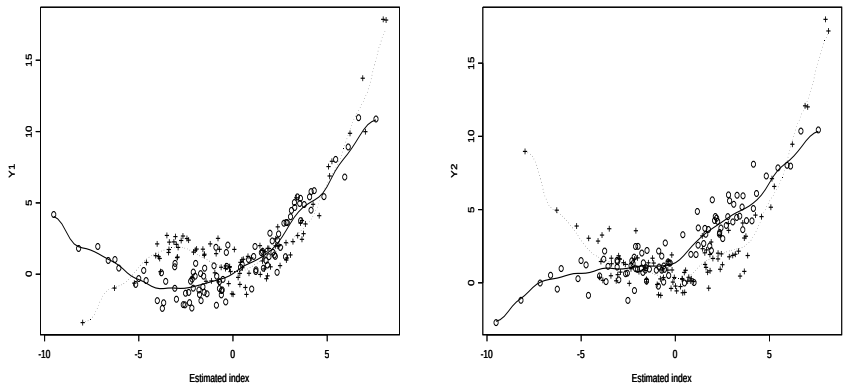

Figure 3: *Plots of the estimated index $X'\hat{b}$ versus Y^1 (first response variable) on the left side and Y^2 (second response variable) on the right side, with $\circ$ for $Z = 1$ and $+$ for $Z = 2$, with plots of the estimated link functions for $Z = 1$ (solid line) and for $Z = 2$ (dotted line)*

4.2 Simulation 2

To evaluate the performance between PMS_α hetero and PMS_α homo methods, we generate simulated data from the following semiparametric regression model ($q = 1,\ L = 2$):

$$\begin{cases} Y^1 = \beta' X + \varepsilon^{(1)} & \text{for } \ Z = 1 \\ Y^2 = \sqrt{5|\beta' X|} + \varepsilon^{(2)} & \text{for } \ Z = 2 \end{cases} \tag{13}$$

where $X|Z = l$ (for $l = 1, 2$) follows a 5-dimensional normal distribution with mean $\mu^{(l)} = \mathbf{0}_5$ and covariance matrix $\Sigma^{(l)}$ defined by:

$$\Sigma^{(1)} = \text{diag}(1, 5, 10, 15, 20) \ \text{ and } \ \Sigma^{(2)} = \Sigma^{(1)} + \theta V,$$

where $\theta \geq 0$ and

$$V = \begin{bmatrix} 26 & 4.6 & 5.4 & 7.6 & 5.6 \\ 4.6 & 21 & 4.6 & 6.4 & 3.6 \\ 5.4 & 4.6 & 16.8 & 7.6 & 5.2 \\ 7.6 & 6.4 & 7.6 & 16.2 & 6.4 \\ 5.6 & 3.6 & 5.2 & 6.4 & 6.8 \end{bmatrix}.$$

Each $\varepsilon^{(l)}$ is standard normally distributed. We take $\beta = (1, 1, -1, -1, 0)'$. The parameter θ switches homoscedastic case to heteroscedastic case. When

$\theta = 0$, we are in the homoscedastic case. Highest values of θ involve that the covariance matrices $\Sigma^{(1)}$ and $\Sigma^{(2)}$ become very different.

From model (13), $N = 100$ samples of size $n = 100$ were generated for each value of θ in the set $\{0, 2, 4, 6, 8, 10\}$. For each simulated sample, the e.d.r. direction was estimated with PMS_α homo and PMS_α hetero methods. In order to compare these two different estimates, we calculated, for each estimation, its corresponding square cosine.

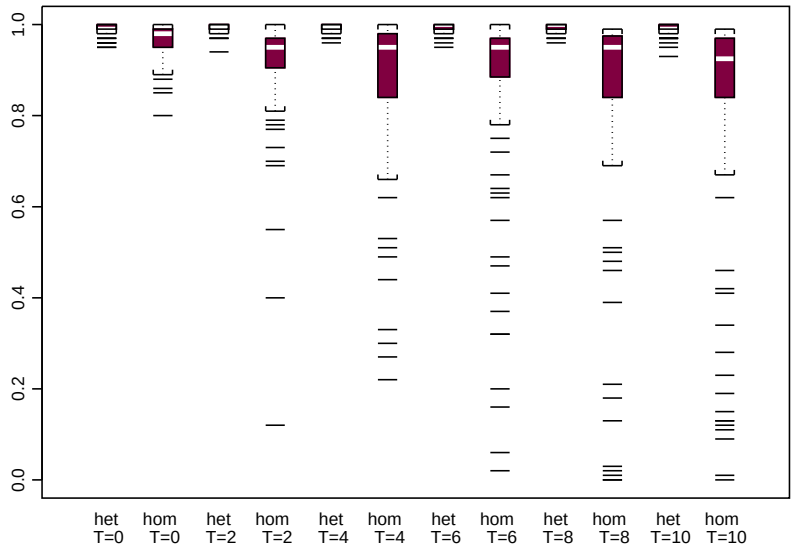

Figure 4: *Boxplots of the squared cosines for different values of θ.*

Comments on the boxplots. We represent on Figure 4 the boxplots of the $N = 100$ square cosines obtained with the PMS_α homo and PMS_α hetero methods for different values of θ (0, 2, 4, 6, 8, 10) and $n = 100$'s sample size.

- The PMS_α hetero method always gives the best quality of estimations. The corresponding squared cosines are always greater than 0.95.
- The PMS_α homo method only gives suitable results when $\theta = 0$ ($\Sigma^{(1)} = \Sigma^{(2)}$). The quality of the estimations decreases with increasing values of θ. This method seems to be very sensitive to deviation from the homoscedastic hypothesis.

5 Concluding remarks

Since a very large number of high-dimensional data sets do contain quantitative and categorical variables, the introduction of discrete predictors in reduction dimension model appears to be very useful. In this article, we presented an extension of the well-known dimension reduction methods, SIR_α and PMS, to regression analyses involving predictors of both types, with multiple responses. All assumptions used in our extension are similar to the ones employed in classical dimension reduction context. Contrary to Chiaromonte et al. (2002), we have abandoned the common covariance assumption (8) used in partial SIR. This is particulary helpful when the several subpopulations (identified by the discrete predictor) have very different covariance structures for the quantitative covariables. Note that Chiaromonte et al.'s framework does guarantee that the eigenvectors corresponding to the largest eigenvalues belong to the Partial Central Subspace as they defined it. Our method has been implemented in Splus and can be avalaible from the authors. An homoscadastic version and an heteroscedastic version of the method have been developed. Our approach perfoms well with small sample size. The link functions between the variables of interest and the common estimated index can be first nonparametrically estimated with a kernel method for instance, and subsequently parametrically modelled if necessary. Finally, we have not detailed the multiple indices case ($K > 1$); in this case, we focus on the estimated e.d.r. space which is spanned by the first K eigenvectors associated with the largest K eigenvalues of the matrices $(\widehat{\Sigma}^\star)^{-1}\widehat{M}_{q,L}^P$ (for the homoscedastic context) or $\widehat{\mathcal{M}}_{q,L}^P$ (for the heteroscedastic context). Moreover, other well known procedures for multidimensional dependent variable such as Alternating SIR (see Li et al., 2003, for instance) or nearest neighbor inverse regression of Hsing (1999) could be adapted to our approach in a way similar to the one detailed here for pooled marginal slicing. This is currently under investigations.

Appendix: an example of a pathological case

Let us introduce the following 2×2 symmetric positive definite matrices $\tilde{A}_1$ and $\tilde{B}_1$, and the multiplication of these two matrices:

$$\tilde{A}_1 = \begin{bmatrix} \epsilon & 0 \\ 0 & 1/\epsilon \end{bmatrix},\ \tilde{B}_1 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \text{ and } \tilde{C}_1 = \tilde{A}_1\tilde{B}_1 = \begin{bmatrix} 2\epsilon & -\epsilon \\ -1/\epsilon & 2/\epsilon \end{bmatrix}.$$

Let $\epsilon = 2 + \sqrt{3}$ (see below an explanation for the choice of this value for ϵ). The eigenvalues of $\tilde{C}_1$ are $\tilde{\lambda}_1 = 4 + \sqrt{13}$ and $\tilde{\lambda}_2 = 4 - \sqrt{13}$. Let $\tilde{A}_2 = \tilde{B}_1'$, $\tilde{B}_2 = \tilde{A}_1'$ and $\tilde{C}_2 = \tilde{A}_2\tilde{B}_2$. Since $\tilde{C}_2 = \tilde{C}_1'$, these two matrices have the same eigenvalues.

Let $\delta > 0$. We now consider the following 3×3 symmetric positive definite

matrices A_1 and B_1, and the multiplication of these two matrices:

$$A_1 = \begin{bmatrix} \sqrt{4+\sqrt{13}+\delta} & 0 & 0 \\ 0 & & \\ 0 & & \tilde{A}_1 \end{bmatrix}, \; B_1 = \begin{bmatrix} \sqrt{4+\sqrt{13}+\delta} & 0 & 0 \\ 0 & & \\ 0 & & \tilde{B}_1 \end{bmatrix},$$

$$\text{and } C_1 = A_1 B_1 = \begin{bmatrix} 4+\sqrt{13}+\delta & 0 & 0 \\ 0 & & \\ 0 & & \tilde{C}_1 \end{bmatrix}.$$

Clearly, the eigenvalues of C_1 are $\lambda_1 = 4+\sqrt{13}+\delta \; > \; \lambda_2 = \tilde{\lambda}_1 \; > \; \lambda_3 = \tilde{\lambda}_2$. Moreover, the eigenvector associated with the largest eigenvalue λ_1 is $b_1 = (1,0,0)'$.

Similarly, we also consider the following 3×3 symmetric positive definite matrices A_2 and B_2, and the multiplication of these two matrices:

$$A_2 = \begin{bmatrix} \sqrt{4+\sqrt{13}+\delta} & 0 & 0 \\ 0 & & \\ 0 & & \tilde{A}_2 \end{bmatrix}, \; B_2 = \begin{bmatrix} \sqrt{4+\sqrt{13}+\delta} & 0 & 0 \\ 0 & & \\ 0 & & \tilde{B}_2 \end{bmatrix},$$

$$\text{and } C_2 = A_2 B_2 = \begin{bmatrix} 4+\sqrt{13}+\delta & 0 & 0 \\ 0 & & \\ 0 & & \tilde{C}_2 \end{bmatrix}.$$

Clearly, C_2 have the same eigenvalues as C_1: $\lambda_1 > \lambda_2 > \lambda_3$, and the eigenvector associated with the largest eigenvalue λ_1 is also $b_1 = (1,0,0)'$.

Finally, let us define the matrix

$$N = \frac{1}{2}C_1 + \frac{1}{2}C_2 = \begin{bmatrix} 4+\sqrt{13}+\delta & 0 & 0 \\ 0 & 2\epsilon & -(\epsilon+1/\epsilon)/2 \\ 0 & -(\epsilon+1/\epsilon)/2 & 2/\epsilon \end{bmatrix}.$$

Note that when $\epsilon = 2+\sqrt{3}$, the matrix N has the following simplest expression:

$$N = \begin{bmatrix} 4+\sqrt{13}+\delta & 0 & 0 \\ 0 & 2\epsilon & -2 \\ 0 & -2 & 2/\epsilon \end{bmatrix}.$$

Straightforwardly, since $0 < \delta < 8-(4+\sqrt{13})$, the eigenvalues of N are: $\xi_1 = 8 \; > \; \xi_2 = \lambda_1 \; > \; \xi_3 = 0$. Hence, the eigenvector $b_1 = (1,0,0)'$ is now associated with the eigenvalue ξ_2 which is not the largest eigenvalue of N. This illustrates the mentioned pathological case.

Note that this numerical example involves the sums of two matrices, C_1 and C_2, each with two large nearly equal eigenvalues and a third relatively small eigenvalue. In our dimension reduction setting, one would most likely be using $K = 2$ here (instead of $K = 1$).

Acknowledgements

We would like to thank the two anonymous referees and the Associated Editor for providing helpful comments that greatly improved the final version of this article.

References

Aragon, Y. (1997). A Gauss implementation of multivariate sliced inverse regression. *Computational Statistics*, **12**, 355-372.

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

Barreda, L., Gannoun, A. & Saracco, J. (2003). Some extensions of multivariate SIR. *Submitted paper.*

Bura, E. & 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.

Bura, E. & Cook, R.D. (2001). Extending sliced inverse regression: the weighted chi-squared test. *Journal of the American Statistical Association*, **96**, 996-1003.

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

Carroll, R.J. & Li, K.C. (1995). Binary regressors in dimension reduction models: A new look at treatment comparisons. *Statistica Sinica*, **5**, 667-688.

Chiaromonte, F., Cook, R.D. & Li, B. (2002). Sufficient dimension reduction in regressions with categorical predictors. *The Annals of Statistics*, **30**, 475-497.

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

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

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

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

Gannoun, A. & Saracco, J. (2003b). Two Cross Validation Criteria for SIR_α and $PSIR_\alpha$ methods in view of prediction. To appear in *Computational Statistics*, **4**.

Gather, U., Hilker, T. & Becker, C. (2002). A note on outlier sensitivity of sliced inverse regression. *Statistics*, **36**, 271-281.

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

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

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

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., Aragon Y., Shedden, K. & 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. (1999). Sliced Inverse Regression under linear constraints. *Communications in Statistics - Theory and methods*, **28**(10), 2367-2393.

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. To appear in *Journal of Multivariate Analysis.*

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

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

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