Statistica Sinica **13**(2003), 297-310

AN ASYMPTOTIC THEORY FOR SIR$_\alpha$ METHOD

Ali Gannoun and Jérôme Saracco

Université Montpellier II

Abstract: Sliced Inverse Regression (SIR) is a nonparametric method for achieving dimension reduction in regression problems. It is widely applicable, very easy to implement on a computer and requires no nonparametric smoothing devices such as kernel regression or smoothing splines regression. The first moment-based SIR has been extensively studied. However, one major restriction is its vulnerability to symmetric dependencies. Methods based on second moments have been suggested as a remedy, one is called SIR$_\alpha$. In this paper, we establish the asymptotic normality of the SIR$_\alpha$ estimates.

Key words and phrases: Asymptotics, eigen-elements, semiparametric regression model, sliced inverse regression (SIR).

1. Introduction

Sliced Inverse Regression (SIR) methods, introduced by Li (1991), examine the relationship between a univariate response variable y and a p-multidimensional regressor variable x with expectation μ and covariance matrix Σ. In contrast to regression analysis, the aim is not the estimation of an unknown regression function, but that of the influence of the regressor space on the response variable. The corresponding model assumes that the dependency between the regressors and the response variable is described by linear combinations of the regressors. If only a few linear combinations are needed in comparison to the dimensionality of the explanatory variable, the aim of reducing dimension is reached. Note that the regression function need not be estimated. The underlying semiparametric regression model is rather general:

$$y = g(\beta_1' x, \ldots, \beta_K' x, \varepsilon). \tag{1}$$

Here g is an unknown function and ε is an unknown random error independent of x. The goal is to estimate the space spanned by the K linearly independent β's, called the effective dimension reduction (e.d.r.) space. When K is small ($K << p$), the data can be effectively reduced by projecting x along the e.d.r. directions for futher study of their relationship with y.

While there are several possible variations, the basic principle of SIR methods is to reverse the role of y and x. Instead of regressing the univariate y on the multivariate x, the multivariate x is regressed on the univariate y. Estimates based on the first moment $E(x|y)$ have been studied extensively (see for instance Duan and Li (1991), Li (1991), Carrol and Li (1992), Hsing and Carrol (1992), Zhu and Ng (1995), Kötter (1996), Saracco (1997), Aragon and Saracco (1997)). Estimates based on the second moment have also been suggested (see for instance Li (1991), Cook and Weisberg (1991) or Kötter (2000)).

One crucial condition for the success of SIR methods is:

$$E(b'x|\beta_1'x, \ldots, \beta_K'x) \quad \text{is linear for any } b. \tag{2}$$

Note that (2) is satisfied when x has an elliptically symmetric distribution. It does not seem possible to verify (2), 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 (2) holds approximatively for many high dimensional data sets. Thus, a blind application of SIR methods without checking (2) can still be helpful in finding the e.d.r. directions. A diagnostic check is then recommended after using SIR methods.

Standard work with SIR concentrates on the use of the inverse regression curve $E(x|y)$ to find the e.d.r. space, constituting the SIR-I approach. A natural extension is to consider the conditional covariance, $\text{Cov}\,(x|y)$, as y varies. Like the inverse regression curve, the orientation of this second-moment curve in the space of $p \times p$ symmetric matrices can be used to determine the e.d.r. directions. This leads to the SIR-II approach. Conjugate information can be used for increasing the chance of discovering the e.d.r. directions. 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. This is the idea of the SIR$_\alpha$ method. The following is a short summary of the underlying theoretical results for these approaches. Let T denote a monotonic transformation of y. Under (2), Li (1991) established the following geometric properties of the model (1).

• **SIR-I approach.** The centered inverse regression curve, $E(x|T(y)) - E(x)$ as y varies, is contained in the linear subspace of $\mathbb{R}^p$ spanned by the vectors $\Sigma\beta_1, \ldots,\ \Sigma\beta_K$. A straightforward consequence is that the covariance matrix, $M_I = \text{Cov}\,(E(x|T(y)))$, is degenerate in any direction Σ-orthogonal to the β_k's. Therefore, the eigenvectors associated with the nonnull K eigenvalues of $\Sigma^{-1}M_I$ are e.d.r. directions.

• **SIR-II approach.** Using the conditional covariance Cov $(x|T(y))$, the eigenvectors with the largest eigenvalues of the matrix $\Sigma^{-1}M_{II}$ are the e.d.r. directions, where

$$M_{II}=E\{(\mathrm{Cov}\,(x|T(y))-E(\mathrm{Cov}\,(x|T(y))))\Sigma^{-1}(\mathrm{Cov}\,(x|T(y))-E(\mathrm{Cov}\,(x|T(y))))'\}.$$

• **SIR$_{\boldsymbol\alpha}$ approach.** One convenient choice is to conjugate information from SIR-I and SIR-II. Then we consider, for $\alpha \in [0,1]$, the matrix $\Sigma^{-1}M_\alpha$ where $M_\alpha = (1-\alpha)M_I\Sigma^{-1}M_I + \alpha M_{II}$. Straightforwardly, the eigenvectors associated with the largest K eigenvalues of $\Sigma^{-1}M_\alpha$ are the e.d.r. directions.

In the following, we assume that these K e.d.r. directions, denoted by $b_1,\ldots,$ b_K, span the e.d.r. space. Let us remark that, when $\alpha = 0$ (resp. $\alpha = 1$), SIR$_\alpha$ is equivalent to SIR-I (resp. SIR-II).

The above properties involve only first and second order moments of x and y. Li (1991) proposed a transformation T, called a slicing, which categorizes the response y into a new response with $H > K$ levels. We assume the support of y is partitioned into H slices $s_1,\ldots,s_h,\ldots,s_H$. With such transformation T, the matrices of interest are now written as

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

where $p_h = P(y \in s_h)$, $m_h = E(x|y \in s_h)$, $V_h = \mathrm{Cov}\,(x|y \in s_h)$, $\overline{V} = \sum_{h=1}^{H} p_hV_h$ and $K_h = \sqrt{p_h}(V_h - \overline{V})$.

Let $\mathbb{I}[.]$ be the indicator function and let $\mathbb{I}_h = \mathbb{I}[y \in s_h]$. Then $p_h = E(\mathbb{I}_h)$, $m_h = E(x\mathbb{I}_h)/p_h$ and $V_h = E(xx'\mathbb{I}_h)/p_h - (E(x\mathbb{I}_h)/p_h)(E(x\mathbb{I}_h)/p_h)'$. So, it is straightforward to estimate these matrices and therefore the e.d.r. directions.

Let $\{(y_i,x_i'),\ i = 1,\ldots,n\}$ be a sample of observations from model (1). The empirical mean and covariance matrix of the x_i's are given by $\overline{x} = n^{-1}\sum_{i=1}^{n} x_i$ and $\widehat{\Sigma} = \overline{xx'} - \overline{x}\ \overline{x}'$ where $\overline{xx'} = n^{-1}\sum_{i=1}^{n} x_ix_i'$. Moreover, let us write $\mathbb{I}_{hi} = \mathbb{I}[y_i \in s_h]$, $\overline{\mathbb{I}_h} = n^{-1}\sum_{i=1}^{n}\mathbb{I}_{hi}$, $\overline{x\mathbb{I}_h} = n^{-1}\sum_{i=1}^{n} x_i\mathbb{I}_{hi}$, $\overline{xx'\mathbb{I}_h} = n^{-1}\sum_{i=1}^{n} x_ix_i'\mathbb{I}_{hi}$, $\widehat{V}_h = (\overline{xx'\mathbb{I}_h}/\overline{\mathbb{I}_h}) - (\overline{x\mathbb{I}_h}/\overline{\mathbb{I}_h})(\overline{x\mathbb{I}_h}/\overline{\mathbb{I}_h})'$ and $\widehat{\overline{V}} = \sum_{h=1}^{H}\overline{\mathbb{I}_h}\widehat{V}_h$.

By substituting empirical versions of these moments for their theoretical counterparts, M_I and M_{II} are then estimated by

$$\widehat{M_I} = \sum_{h=1}^{H}\overline{\mathbb{I}_h}((\overline{x\mathbb{I}_h}/\overline{\mathbb{I}_h})-\overline{x})((\overline{x\mathbb{I}_h}/\overline{\mathbb{I}_h})-\overline{x})' \quad\text{and}\quad \widehat{M}_{II} = \sum_{h=1}^{H}\widehat{K}_h\widehat{\Sigma}^{-1}\widehat{K}_h, \tag{3}$$

where $\widehat{K_h} = (\overline{\mathbb{I}_h})^{-1/2}(\widehat{V}_h - \widehat{\overline{V}})$.

Finally, M_α is estimated by $\widehat{M_\alpha} = (1-\alpha)\widehat{M_I}\widehat{\Sigma}^{-1}\widehat{M_I} + \alpha\widehat{M}_{II}$. Then the eigenvectors $\hat{b}_1,\ldots,\hat{b}_K$, associated with the K largest eigenvalues of $\widehat{\Sigma}^{-1}\widehat{M_\alpha}$, are the K estimated e.d.r. directions. The e.d.r. space E is estimated by $\hat{E}$, the linear subspace generated by the $\hat{b}_k$'s.

Li (1991) and Saracco (2001) have shown that each estimated e.d.r. direction converges to an e.d.r. direction at rate root n for SIR-I, SIR-II or the SIR_α method. Determining the number K (of indices) in model (1) is considered by Li (1991), Schott (1994) and Ferré (1998). We do not examine this topic here, and assume that K is known.

This paper focuses on the asymptotic normality of the SIR_α estimator of the e.d.r. space when the support of y is partitioned into H fixed slices. In Section 2, we state the main results. First, the asymptotic distribution of $\sqrt{n}\mathrm{vec}(\widehat{\Sigma}^{-1}\widehat{M}_\alpha - \Sigma^{-1}M_\alpha)$ is obtained in Theorem 1. Then the asymptotic distribution of the eigenprojector on the estimated e.d.r. space is derived in Theorem 2, as well as the asymptotic distributions of each estimated e.d.r. direction and its corresponding eigenvalue, respectively, in Theorems 3 and 4. The proofs are in the Appendix.

2. Main Results

From now on, for each $s\times s$ matrix $D = \left(d^{(jk)}\right)$, let $\mathrm{vec}(D) = (d^{(11)}, \ldots, d^{(s1)}, d^{(21)}, d^{(22)}, \ldots, d^{(ss)})'$ be the s^2-dimensional column vector of all elements of D. Let $D_1 \otimes D_2$ denote the Kronecker product of the matrices D_1 and D_2 (see Tyler (1981) for some useful properties of the Kronecker product). In the sequel, the notation $X_n \longrightarrow_d X$ means that X_n converges in distribution to X as $n \to \infty$.

The assumptions which are necessary to state our results are gathered together below for easy reference.

(A1) $\{(y_i, x_i'), i = 1, \ldots, n\}$ is a sample of independent observations from model (1).

(A2) The support of y is partitioned into H fixed slices $s_1, \ldots, s_h, \ldots, s_H$ such that $p_h \neq 0$.

(A3) The covariance matrix Σ is positive definite.

(A4) The $K+1$ largest eigenvalues of $\Sigma^{-1}M_\alpha$ are non-null and satisfy: $\lambda_1 > \cdots > \lambda_K > \lambda_{K+1}$, where $K+1 \leq p$.

2.1. Asymptotic distribution of $\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{M}}_{\boldsymbol{\alpha}}$

Theorem 1. *Under assumptions* (A1), (A2) *and* (A3),

$$\sqrt{n}(\widehat{\Sigma}^{-1}\widehat{M}_\alpha - \Sigma^{-1}M_\alpha) \longrightarrow_d \Phi, \tag{4}$$

where Φ *is such that* $\mathrm{vec}(\Phi)$ *is normally distributed with mean zero and covariance matrix* $\mathcal{C}$ *defined at* (6).

2.2. Asymptotic normality of the eigenelements of the estimated e.d.r. space

Let us denote by $w = \{\lambda_1, \ldots, \lambda_K\}$ the set of the K eigenvalues associated with the e.d.r. space. Let $P = \sum_{\lambda_k \in w} P_{\lambda_k}$ be the Σ-orthogonal eigenprojector on the e.d.r. space, where $P_{\lambda_k} = b_k b_k' \Sigma$.

Recall that the sample version of SIR$_\alpha$ uses the estimated matrix $\widehat{\Sigma}^{-1}\widehat{M_\alpha}$. Then take $\widehat{w} = \{\widehat{\lambda}_1, \ldots, \widehat{\lambda}_K\}$ as the set of the K largest eigenvalues of this matrix. The $\widehat{\Sigma}$-orthogonal eigenprojector onto the estimated e.d.r. space $\hat{E}$ is $\widehat{P} = \sum_{\widehat{\lambda}_k \in \widehat{w}} \widehat{P}_{\widehat{\lambda}_k}$ where $\widehat{P}_{\widehat{\lambda}_k} = \widehat{b}_k \widehat{b}_k' \widehat{\Sigma}$.

Starting from the limit distribution obtained in (4), we can derive the asymptotic distribution of the eigenelements describing the estimated e.d.r. space, namely the eigenprojector onto the estimated e.d.r. space (in Theorem 2), the estimated e.d.r. directions (in Theorem 3) and their corresponding eigenvalues (in Theorem 4).

Theorem 2. *Under the assumptions* (A1), (A2), (A3) *and* (A4), *we have* $\sqrt{n}(\widehat{P} - P) \longrightarrow_d \Phi_P$, *where* Φ_P *is such that* $\mathrm{vec}(\Phi_P)$ *is normally distributed with mean zero and covariance matrix* $\mathcal{C}_P$ *defined at* (9).

Theorem 3. *Under the assumptions* (A1), (A2), (A3) *and* (A4), *we get* $\sqrt{n}(\hat{b}_k - b_k) \longrightarrow_d \Phi_{b_k}$, *where* Φ_{b_k} *has the normal distribution with mean zero and covariance matrix* $\mathcal{C}_{b_k}$ *defined at* (13).

Theorem 4. *Under the assumptions* (A1), (A2), (A3) *and* (A4), *we get* $\sqrt{n}(\hat{\lambda}_k - \lambda_k) \longrightarrow_d \Phi_{\lambda_k}$, *where* Φ_{λ_k} *has a normal distribution with mean zero and variance* $\mathcal{C}_{\lambda_k} = [b_k' \otimes b_k'\Sigma]\mathcal{C}[b_k \otimes \Sigma b_k]$.

2.3. Concluding remark

The asymptotic covariance matrices $\mathcal{C}$, $\mathcal{C}^*$, $\mathcal{C}_P$, $\mathcal{C}_{b_k}$ and $\mathcal{C}_{\lambda_k}$ depend (directly or via eigenelements) on the theoretical moments, which can be easily estimated by their empirical counterparts. Therefore, it is straighforward to derive consistent estimates of the asymptotic covariance matrices. Note that the theoretical results of this paper are given for a fixed α in [0,1]. Clearly, if $\alpha = 0$, asymptotic results concern the SIR-I method. Similarly, if $\alpha = 1$, the theory works for the SIR-II approach. In practice, we should choose a value for α using the available data, and adaptively. Some optimal choices have been studied by Saracco (2001) and Gannoun and Saracco (2001).

Acknowledgements

The authors are grateful to the Editor, an associate editor and two anonymous referees for contributing to the improvement of this paper through many

useful remarks, suggestions, and detailed comments.

Appendix

A.1. Proof of theorem 1

Throughout the proof, I_s denotes the $s \times s$ identity matrix, and $0_{s_1,s_2}$ stands for the $s_1 \times s_2$ null matrix. The proof is divided into five steps and relies on the Central Limit Theorem and the Delta method (see for example Serfling (1980), Theorem A, p.122).

Step 1*: Application of the Central Limit Theorem.*
For $i = 1, \ldots, n$, define the $(H + pH + p + p^2 + p^2H)$-dimensional column vector

$$U_i = (\mathbb{I}_{1i}, \ldots, \mathbb{I}_{Hi}, x_i'\mathbb{I}_{1i}, \ldots, x_i'\mathbb{I}_{Hi}, x_i', \mathrm{vec}(x_i x_i')', \mathrm{vec}(x_i x_i'\mathbb{I}_{1i})', \ldots, \mathrm{vec}(x_i x_i'\mathbb{I}_{Hi})')'.$$

Under (A1), the vectors U_i, $i = 1, \ldots, n$ are independent and identically distributed. For $h = 1, \ldots, H$, write $\widetilde{m}_h = E(x\mathbb{I}_h)$ and $\widetilde{V}_h = E(xx'\mathbb{I}_h)$, then the mean μ_U of U is

$$\mu_U = (p_1, \ldots, p_H, \widetilde{m}_1', \ldots, \widetilde{m}_H', \mu', \mathrm{vec}(\Sigma + \mu\mu')', \mathrm{vec}(\widetilde{V}_1)', \ldots, \mathrm{vec}(\widetilde{V}_H)')'.$$

To give the expression to the covariance matrix of U, we need additional notation. For $h = 1, \ldots, H$, $\widetilde{M}_h = E(x(x' \otimes x')\mathbb{I}[y \in s_h])$, and $\widetilde{N}_h = E((xx') \otimes (xx')\mathbb{I}_h)$. Moreover, set $M = E(x(x' \otimes x'))$ and $N = E((xx') \otimes (xx'))$. The covariance matrix Σ_U is then

$$\Sigma_U = \begin{bmatrix} B_{11} & B_{12} & B_{13} & B_{14} & B_{15} \\ B_{12}' & B_{22} & B_{23} & B_{24} & B_{25} \\ B_{13}' & B_{23}' & B_{33} & B_{34} & B_{35} \\ B_{14}' & B_{24}' & B_{34}' & B_{44} & B_{45} \\ B_{15}' & B_{25}' & B_{35}' & B_{45}' & B_{55} \end{bmatrix},$$

where for i, $j = 1, \ldots, 5$, the blocks B_{ij} are the following:

$$B_{11} = \begin{bmatrix} p_1(1-p_1) & -p_2p_1 & \cdots & -p_Hp_1 \\ -p_1p_2 & p_2(1-p_2) & \cdots & -p_Hp_2 \\ \vdots & \vdots & \ddots & \vdots \\ -p_1p_H & \cdots & \cdots & p_H(1-p_H) \end{bmatrix},$$

$$B_{12} = \begin{bmatrix} (1-p_1)\widetilde{m}_1' & -p_1\widetilde{m}_2' & \cdots & -p_1\widetilde{m}_H' \\ -p_2\widetilde{m}_1' & (1-p_2)\widetilde{m}_2' & \cdots & -p_2\widetilde{m}_H' \\ \vdots & \vdots & \ddots & \vdots \\ -p_H\widetilde{m}_1' & \cdots & \cdots & (1-p_H)\widetilde{m}_H' \end{bmatrix},$$

$$B_{13}=\begin{bmatrix}\widetilde{m}_1'-p_1\mu'\\ \vdots\\ \widetilde{m}_H'-p_H\mu'\end{bmatrix},\qquad B_{14}=\begin{bmatrix}\mathrm{vec}(\widetilde{V}_1)'-p_1\mathrm{vec}(\Sigma+\mu\mu')'\\ \vdots\\ \mathrm{vec}(\widetilde{V}_H)'-p_H\mathrm{vec}(\Sigma+\mu\mu')'\end{bmatrix},$$

$$B_{15}=\begin{bmatrix}(1-p_1)\mathrm{vec}(\widetilde{V}_1)' & -p_1\mathrm{vec}(\widetilde{V}_1)' & \dots & -p_1\mathrm{vec}(\widetilde{V}_H)'\\ -p_2\mathrm{vec}(\widetilde{V}_1)' & (1-p_2)\mathrm{vec}(\widetilde{V}_2)' & \dots & -p_2\mathrm{vec}(\widetilde{V}_H)'\\ \vdots & \vdots & \ddots & \vdots\\ -p_H\mathrm{vec}(\widetilde{V}_1)' & \dots & \dots & (1-p_H)\mathrm{vec}(\widetilde{V}_H)'\end{bmatrix},$$

$$B_{22}=\begin{bmatrix}\widetilde{V}_1-\widetilde{m}_1\widetilde{m}_1' & -\widetilde{m}_1\widetilde{m}_2' & \dots & -\widetilde{m}_1\widetilde{m}_H'\\ -\widetilde{m}_2\widetilde{m}_1' & \widetilde{V}_2-\widetilde{m}_2\widetilde{m}_2' & \dots & -\widetilde{m}_2\widetilde{m}_H'\\ \vdots & \vdots & \ddots & \vdots\\ -\widetilde{m}_H\widetilde{m}_1' & \dots & \dots & \widetilde{V}_H-\widetilde{m}_H\widetilde{m}_H'\end{bmatrix},$$

$$B_{23}=\begin{bmatrix}\widetilde{V}_1-\widetilde{m}_1\mu'\\ \vdots\\ \widetilde{V}_H-\widetilde{m}_H\mu'\end{bmatrix},\qquad B_{24}=\begin{bmatrix}\widetilde{M}_1-\widetilde{m}_1\mathrm{vec}(\Sigma+\mu\mu')'\\ \vdots\\ \widetilde{M}_H-\widetilde{m}_H\mathrm{vec}(\Sigma+\mu\mu')'\end{bmatrix},$$

$$B_{25}=\begin{bmatrix}\widetilde{M}_1-\widetilde{m}_1\mathrm{vec}(\widetilde{V}_1)' & -\widetilde{m}_1\mathrm{vec}(\widetilde{V}_2)' & \dots & -\widetilde{m}_1\mathrm{vec}(\widetilde{V}_H)'\\ -\widetilde{m}_2\mathrm{vec}(\widetilde{V}_1)' & \widetilde{M}_2-\widetilde{m}_2\mathrm{vec}(\widetilde{V}_1)' & \dots & -\widetilde{m}_2\mathrm{vec}(\widetilde{V}_H)'\\ \vdots & \vdots & \ddots & \vdots\\ -\widetilde{m}_H\mathrm{vec}(\widetilde{V}_1)' & \dots & \dots & \widetilde{M}_H-\widetilde{m}_1\mathrm{vec}(\widetilde{V}_H)'\end{bmatrix},$$

$$B_{33}=\Sigma,\qquad B_{34}=\left[M-\mu\mathrm{vec}(\Sigma+\mu\mu')'\right],$$

$$B_{35}=\left[\widetilde{M}_1-\mu\mathrm{vec}(\widetilde{V}_1)'\ \cdots\ \widetilde{M}_H-\mu\mathrm{vec}(\widetilde{V}_H)'\right],$$

$$B_{44}=\left[N-\mathrm{vec}(\Sigma+\mu\mu')\mathrm{vec}(\Sigma+\mu\mu')'\right],$$

$$B_{45}=\left[\widetilde{N}_1-\mathrm{vec}(\Sigma+\mu\mu')\mathrm{vec}(\widetilde{V}_1)'\ \cdots\ \widetilde{N}_H-\mathrm{vec}(\Sigma+\mu\mu')\mathrm{vec}(\widetilde{V}_H)'\right],$$

$$B_{55}=\begin{bmatrix}\widetilde{N}_1-\mathrm{vec}(\widetilde{V}_1)\mathrm{vec}(\widetilde{V}_1)' & -\mathrm{vec}(\widetilde{V}_1)\mathrm{vec}(\widetilde{V}_2)' & \cdots & -\mathrm{vec}(\widetilde{V}_1)\mathrm{vec}(\widetilde{V}_H)\\ -\mathrm{vec}(\widetilde{V}_2)\mathrm{vec}(\widetilde{V}_1)' & \widetilde{N}_2-\mathrm{vec}(\widetilde{V}_2)\mathrm{vec}(\widetilde{V}_2)' & \cdots & -\mathrm{vec}(\widetilde{V}_2)\mathrm{vec}(\widetilde{V}_H)'\\ \vdots & \vdots & \ddots & \vdots\\ -\mathrm{vec}(\widetilde{V}_H)\mathrm{vec}(\widetilde{V}_1)' & \dots & \dots & \widetilde{N}_H-\mathrm{vec}(\widetilde{V}_H)\mathrm{vec}(\widetilde{V}_H)'\end{bmatrix}.$$

From the Central Limit Theorem, $\sqrt{n}(\overline{U}-\mu_U)\longrightarrow_d \mathcal{N}(0,\Sigma_U)$.

Step 2: Asymptotic distribution of the random variables comprising $\widehat{M}_I$, $\widehat{M}_{II}$ and $\widehat{\Sigma}$.

In order to use the Delta method, we stack the variables comprising $\widehat{M}_I$, $\widehat{M}_{II}$ and $\widehat{\Sigma}$ into the vector

$$\begin{aligned}\overline{U}_1 = (&\overline{\mathbb{I}_1},\ldots,\overline{\mathbb{I}_H},(\overline{x\mathbb{I}_1}/\overline{\mathbb{I}_1})',\ldots,(\overline{x\mathbb{I}_H}/\overline{\mathbb{I}_H})',\overline{x}',\mathrm{vec}(\overline{xx'})',\mathrm{vec}(\overline{xx'\mathbb{I}_1})'/\overline{\mathbb{I}_1},\ldots,\\ &\mathrm{vec}(\overline{xx'\mathbb{I}_1})'/\overline{\mathbb{I}_H})'.\end{aligned}$$

We define the function f_1 from $\mathbb{R}^{H+pH+p+p^2+p^2H}$ to $\mathbb{R}^{H+pH+p+p^2+p^2H}$ by

$$\begin{aligned}&f_1\left((a',b_1',\ldots,b_H',c',d',e_1',\ldots,e_H')'\right)\\ &=(a',b_1'/a_1,\ldots,b_H'/a_H,c',d',e_1'/a_1,\ldots,e_H'/a_H)',\end{aligned}$$

where $a=(a_1,\ldots,a_H)'\in\mathbb{R}^H$ (assumed nonnull), $b_h\in\mathbb{R}^p$, $c\in\mathbb{R}^p$, $d\in\mathbb{R}^{p^2}$ and $e_h\in\mathbb{R}^{p^2H}$ are column vectors. Under (A2), it is clear that $\overline{U}_1=f_1(\overline{U})$.

Let us also define $\mu_1=f_1(\mu_U)=(p_1,\ldots,p_H,m_1',\ldots,m_H',\mu',\mathrm{vec}(\Sigma+\mu\mu'),\mathrm{vec}(\widetilde{V}_1)/p_1,\ldots,\mathrm{vec}(\widetilde{V}_H)/p_H)'$ and $\Sigma_1=F_1'\Sigma_UF_1$, where $F_1=\frac{\partial f_1'}{\partial u}\Big|_E$. Here and subsequently the notation $g|_E$ is the evaluation of g at the expectation of its argument. After straightforward calculations, we get

$$F_1=\left[\begin{array}{c|c|c|c|c} I_H & \begin{matrix}-\frac{\widetilde{m}_1'}{(p_1)^2} & & \\ & \ddots & \\ & & -\frac{\widetilde{m}_H'}{(p_H)^2}\end{matrix} & 0_{H,p} & 0_{H,p^2} & \begin{matrix}-\frac{\mathrm{vec}(\widetilde{V}_1)'}{(p_1)^2} & & \\ & \ddots & \\ & & -\frac{\mathrm{vec}(\widetilde{V}_H)'}{(p_H)^2}\end{matrix}\\ \hline 0_{Hp,H} & \begin{matrix}I_p/p_1 & & \\ & \ddots & \\ & & I_p/p_H\end{matrix} & 0_{Hp,p} & 0_{Hp,p^2} & 0_{Hp,p^2H}\\ \hline 0_{p,H} & 0_{p,pH} & I_p & 0_{p,p^2} & 0_{p,p^2H}\\ \hline 0_{p^2,H} & 0_{p^2,pH} & 0_{p^2,p} & I_{p^2} & 0_{p^2,p^2H}\\ \hline 0_{p^2H,H} & 0_{p^2H,pH} & 0_{p^2H,p} & 0_{p^2H,p} & \begin{matrix}I_{p^2}/p_1 & & \\ & \ddots & \\ & & I_{p^2}/p_H\end{matrix}\end{array}\right].$$

Since f_1 satisfies the required conditions of the Delta method theorem and from *Step* 1, we get $\sqrt{n}(\overline{U}_1-\mu_1)\longrightarrow_d\mathcal{N}(0,\Sigma_1)$.

Step 3*: Asymptotic distribution of* $\sqrt{n}\left(\begin{bmatrix}\mathrm{vec}(\widehat{M}_I)\\ \mathrm{vec}(\widehat{\Sigma})\\ \mathrm{vec}(\widehat{K}_1)\\ \vdots\\ \mathrm{vec}(\widehat{K}_H)\end{bmatrix}-\begin{bmatrix}\mathrm{vec}(M_I)\\ \mathrm{vec}(\Sigma)\\ \mathrm{vec}(K_1)\\ \vdots\\ \mathrm{vec}(K_H)\end{bmatrix}\right)$.

Let us define f_2 on $I\!R^{H+pH+p+p^2+p^2H}$ to $I\!R^{p^2+p^2+p^2H}$ as follows:

$$f_2((a', b'_1, \ldots, b'_H, c', d', e'_1, \ldots, e'_H)')$$
$$= \begin{pmatrix} \sum_{h=1}^H a_h \mathrm{vec}[(b_h - c)(b_h - c)'] \\ d - \mathrm{vec}(cc') \\ \sqrt{a_1}\{e_1 - \mathrm{vec}(b_1 b'_1) - \sum_{j=1}^H a_j [e_j - \mathrm{vec}(b_j b'_j)]\} \\ \vdots \\ \sqrt{a_H}\{e_H - \mathrm{vec}(b_H b'_H) - \sum_{j=1}^H a_j [e_j - \mathrm{vec}(b_j b'_j)]\} \end{pmatrix}.$$

It is clear that $f_2(\overline{U}_1) = (\mathrm{vec}(\widehat{M}_I)', \mathrm{vec}(\widehat{\Sigma})', \mathrm{vec}(\widehat{K}_1)', \ldots, \mathrm{vec}(\widehat{K}_H)')'$ and $f_2(\mu_1) = (\mathrm{vec}(M_I)', \mathrm{vec}(\Sigma)', \mathrm{vec}(K_1)', \ldots, \mathrm{vec}(K_H)')'$.

Let us also define $\Sigma_2 = F'_2 \Sigma_1 F_2$

where $F_2 = \frac{\partial f'_2}{\partial u}\Big|_E = \begin{bmatrix} E_1 & 0_{H,p^2} & E_2 \\ E_3 & 0_{pH,p^2} & E_4 \\ E_5 & E_6 & 0_{p,p^2H} \\ 0_{p^2,p^2} & I_{p^2} & 0_{p^2,p^2H} \\ 0_{p^2H,p^2} & 0_{p^2H,p^2} & E_7 \end{bmatrix}$, with $E_1 = \begin{bmatrix} (m_1-\mu)' \otimes (m_1-\mu)' \\ \vdots \\ (m_H-\mu)' \otimes (m_H-\mu)' \end{bmatrix}$,

$$E_2 = \begin{bmatrix} \frac{1}{2\sqrt{p_1}}[(1-3p_1)\mathrm{vec}(V_1)' - \zeta_1] & -\sqrt{p_2}\mathrm{vec}(V_1)' & \cdots & -\sqrt{p_H}\mathrm{vec}(V_1)' \\ -\sqrt{p_1}\mathrm{vec}(V_2)' & \ddots & \ldots & \vdots \\ \vdots & \ldots & \ddots & \vdots \\ -\sqrt{p_1}\mathrm{vec}(V_H)' & \ldots & \ldots & \frac{1}{2\sqrt{p_H}}[(1-3p_H)\mathrm{vec}(V_H)' - \zeta_H] \end{bmatrix},$$

$$E_3 = \begin{bmatrix} p_1[I_p \otimes (m_1 - \mu)' + (m_1 - \mu)' \otimes I_p] \\ \vdots \\ p_H[I_p \otimes (m_H - \mu)' + (m_H - \mu)' \otimes I_p] \end{bmatrix},$$

$$E_4 = \begin{bmatrix} \sqrt{p_1}(p_1 - 1)\xi_1 & \sqrt{p_2}p_1\xi_1 & \cdots & \sqrt{p_H}p_1\xi_1 \\ \sqrt{p_1}p_2\xi_2 & \ddots & \ldots & \vdots \\ \vdots & \ldots & \ddots & \vdots \\ \sqrt{p_1}p_H\xi_H & \ldots & \ldots & \sqrt{p_H}(p_H - 1)\xi_H \end{bmatrix}$$

$$E_5 = \sum_{h=1}^H p_h[I_p \otimes (\mu - m_h)' + (\mu - m_h)' \otimes I_p], \qquad E_6 = -(I_p \otimes \mu' + \mu' \otimes I_p),$$

$$E_7 = \begin{bmatrix} \sqrt{p_1}(p_1 - 1)I_{p^2} & \sqrt{p_2}p_1 I_{p^2} & \ldots & \sqrt{p_H}p_1 I_{p^2} \\ \sqrt{p_1}p_2 I_{p^2} & \ddots & \ldots & \vdots \\ \vdots & \ldots & \ddots & \vdots \\ \sqrt{p_1}p_H I_{p^2} & \ldots & \ldots & \sqrt{p_H}(p_H - 1)I_{p^2} \end{bmatrix},$$

where $\zeta_h = \sum_{j \neq h} p_j \mathrm{vec}(V_j)'$ and $\xi_h = (I_p \otimes m_h' + m_h' \otimes I_p)$, for $h = 1, \ldots, H$.

Since the Delta method applies to f_2, from *Step* 2 we get $\sqrt{n}(f_2(\overline{U}_1) - f_2(\mu_1)) \longrightarrow_d \mathcal{N}(0, \Sigma_2)$.

Step 4*: Asymptotic distribution of* $\sqrt{n}\left(\begin{bmatrix} \mathrm{vec}(\widehat{M}_I) \\ \mathrm{vec}(\widehat{\Sigma}^{-1}) \\ \mathrm{vec}(\widehat{K}_1) \\ \vdots \\ \mathrm{vec}(\widehat{K}_H) \end{bmatrix} - \begin{bmatrix} \mathrm{vec}(M_\alpha) \\ \mathrm{vec}(\Sigma^{-1}) \\ \mathrm{vec}(K_1) \\ \vdots \\ \mathrm{vec}(K_H) \end{bmatrix}\right)$.

Let R be the matrix

$$R = \begin{bmatrix} I_{p^2} & 0_{p^2,p^2} & 0_{p^2,p^2H} \\ 0_{p^2,p^2} & -(\Sigma^{-1} \otimes \Sigma^{-1}) & 0_{p^2,p^2H} \\ 0_{p^2H,p^2} & 0_{p^2H,p^2} & I_{p^2H} \end{bmatrix}. \tag{5}$$

Under (A3), and using the first order approximation $\widehat{\Sigma}^{-1} \doteq \Sigma^{-1} - \Sigma^{-1}(\widehat{\Sigma} - \Sigma)\Sigma^{-1}$ and *Step* 3, we get

$$\sqrt{n}\left(\begin{bmatrix} \mathrm{vec}(\widehat{M}_I) \\ \mathrm{vec}(\widehat{\Sigma}^{-1}) \\ \mathrm{vec}(\widehat{K}_1) \\ \vdots \\ \mathrm{vec}(\widehat{K}_H) \end{bmatrix} - \begin{bmatrix} \mathrm{vec}(M_I) \\ \mathrm{vec}(\Sigma^{-1}) \\ \mathrm{vec}(K_1) \\ \vdots \\ \mathrm{vec}(K_H) \end{bmatrix}\right) \doteq R\sqrt{n}\left(\begin{bmatrix} \mathrm{vec}(\widehat{M}_I) \\ \mathrm{vec}(\widehat{\Sigma}) \\ \mathrm{vec}(\widehat{K}_1) \\ \vdots \\ \mathrm{vec}(\widehat{K}_H) \end{bmatrix} - \begin{bmatrix} \mathrm{vec}(M_I) \\ \mathrm{vec}(\Sigma) \\ \mathrm{vec}(K_1) \\ \vdots \\ \mathrm{vec}(K_H) \end{bmatrix}\right),$$

which converges in distribution to $\mathcal{N}(0, C_R)$ where $C_R = R\Sigma_2 R'$.

Step 5*: Asymptotic distribution of* $\sqrt{n}\mathrm{vec}\left(\widehat{\Sigma}^{-1}\widehat{M}_\alpha - \Sigma^{-1}M_\alpha\right)$.

For the $p \times p$ matrices A, B, C_1, $\ldots$, C_H, let f_3 be defined from $\mathbb{R}^{p^2+p^2+p^2H}$ to $\mathbb{R}^{p^2}$ by

$$f_3(\mathrm{vec}(A)', \mathrm{vec}(B)', \mathrm{vec}(C_1)', \ldots, \mathrm{vec}(C_H)') = \mathrm{vec}\left(B\left[(1-\alpha)ABA + \alpha\sum_{h=1}^{H} C_h B C_h\right]\right).$$

It is clear that $f_3\left(\mathrm{vec}(\widehat{M}_I)', \mathrm{vec}(\widehat{\Sigma}^{-1})', \mathrm{vec}(\widehat{K}_1)', \ldots, \mathrm{vec}(\widehat{K}_H)'\right) = \mathrm{vec}\left(\widehat{\Sigma}^{-1}\widehat{M}_\alpha\right)$ and $f_3\left(\mathrm{vec}(M_I)', \mathrm{vec}(\Sigma^{-1})', \mathrm{vec}(K_1)', \ldots, \mathrm{vec}(K_H)'\right) = \mathrm{vec}\left(\Sigma^{-1}M_\alpha\right)$.

Let us also define

$$\mathcal{C} = F_3' C_R F_3, \tag{6}$$

where

$$F_3 = \frac{\partial f_3'}{\partial u}\Big|_E = \begin{bmatrix} (1-\alpha)(I_p \otimes \Sigma^{-1}M_\alpha + M_\alpha\Sigma^{-1} \otimes I_p)(I_p \otimes \Sigma^{-1}) \\ (1-\alpha)(I_p \otimes \Sigma^{-1}M_\alpha + M_\alpha\Sigma^{-1} \otimes I_p)(M_\alpha \otimes I_p) \\ \quad +\alpha\sum_{h=1}^{H}(I_p \otimes \Sigma^{-1}K_h + K_h\Sigma^{-1} \otimes I_p)(K_h \otimes I_p) \\ \alpha(I_p \otimes \Sigma^{-1}K_1 + K_1\Sigma^{-1} \otimes I_p)(I_p \otimes \Sigma^{-1}) \\ \vdots \\ \alpha(I_p \otimes \Sigma^{-1}K_H + K_H\Sigma^{-1} \otimes I_p)(I_p \otimes \Sigma^{-1}) \end{bmatrix}. \tag{7}$$

Since the Delta method applies to f_3 and, from *Step* 4, a final application of the Delta method leads to $\sqrt{n}\mathrm{vec}(\widehat{\Sigma}^{-1}\widehat{M}_\alpha - \Sigma^{-1}M_\alpha) \longrightarrow_d \mathcal{N}(0, \mathcal{C})$.

A.2. Proof of Theorem 2

Let M be a square matrix of order p. As in Tyler (1981), we set $||M|| = [\max$ eigenvalue of $(\Sigma^{-1}M'\Sigma M)]^{1/2}$. The Moore-Penrose generalized inverse of M is denoted by M^+.

Under the assumptions of the Theorem 2, $\widehat{\Sigma}^{-1}\widehat{M}_\alpha$ converges in probability to $\Sigma^{-1}M_\alpha$ (see for instance Li (1991) or Saracco (2001)). Thus with probability 1, for n sufficiently large, we have

$$\left|\left|\widehat{\Sigma}^{-1}\widehat{M}_\alpha - \Sigma^{-1}M_\alpha\right|\right| \leq \lambda_K/2. \tag{8}$$

From Theorem 1 and (8), we are now in position to apply Lemma 4.1 of Tyler (1981), so

$$\begin{aligned}\hat{P} = P - \sum_{\lambda_k \in w} &[P_{\lambda_k}(\widehat{\Sigma}^{-1}\widehat{M}_\alpha - \Sigma^{-1}M_\alpha)(\Sigma^{-1}M_\alpha - \lambda_k I_p)^+ \\ &+(\Sigma^{-1}M_\alpha - \lambda_k I_p)^+(\widehat{\Sigma}^{-1}\widehat{M}_\alpha - \Sigma^{-1}M_\alpha)P_{\lambda_k}] + \hat{E}_o\end{aligned}$$

where $||\hat{E}_o|| \leq (1+\frac{\lambda_1-\lambda_K}{\lambda_K})(\frac{2}{\lambda_K}||\widehat{\Sigma}^{-1}\widehat{M}_\alpha - \Sigma^{-1}M_\alpha||)^2(1-\frac{2}{\lambda_K}||\widehat{\Sigma}^{-1}\widehat{M}_\alpha - \Sigma^{-1}M_\alpha||)^{-1}$.

Let $\Phi_P = -\sum_{\lambda_k \in w}[P_{\lambda_k}\Phi(\Sigma^{-1}M_\alpha - \lambda_k I_p)^+ + (\Sigma^{-1}M_\alpha - \lambda_k I_p)^+\Phi P_{\lambda_k}]$. From the above, it follows that $\sqrt{n}(\hat{P} - P) \longrightarrow_d \Phi_P$. We remark that $\mathrm{vec}(\Phi_P) = C_w\mathrm{vec}(\Phi)$ where $C_w = -\sum_{\lambda_k \in w}[(M_\alpha\Sigma^{-1} - \lambda_k I_p)^+ \otimes P_{\lambda_k} + P'_{\lambda_k} \otimes (\Sigma^{-1}M_\alpha - \lambda_k I_p)^+]$. Then $\mathrm{vec}(\Phi_P)$ follows the normal distribution $\mathcal{N}(0, \mathcal{C}_P)$ where

$$\mathcal{C}_P = C_w\mathcal{C}C_w'. \tag{9}$$

Note that $(\Sigma^{-1}M_\alpha - \lambda_k I_p)^+$ can be replaced by $S_{\lambda_k} = \sum_{\lambda_l \neq \lambda_k}\frac{1}{\lambda_l - \lambda_k}P_{\lambda_l}$.

A.3. Proof of Theorem 3

This result is a straightforward application of the Lemma 2 of Saracco (1997).

We first show that

$$\sqrt{n}\left(\begin{bmatrix}\mathrm{vec}\left(\widehat{\Sigma}^{-1}\widehat{M}_\alpha\right)\\ \mathrm{vec}\left(\widehat{\Sigma}\right)\end{bmatrix}-\begin{bmatrix}\mathrm{vec}\left(\Sigma^{-1}M_\alpha\right)\\ \mathrm{vec}(\Sigma)\end{bmatrix}\right)\longrightarrow_d\begin{pmatrix}\mathrm{vec}(\Phi)\\ \mathrm{vec}(\Phi_\Sigma)\end{pmatrix}, \tag{10}$$

where $\begin{pmatrix}\mathrm{vec}(\Phi)\\ \mathrm{vec}(\Phi_\Sigma)\end{pmatrix}$ follows the normal distribution $N(0,\mathcal{C}^*)$ with $\mathcal{C}^*$ given by (12). The proof is based on slight modifications of *Steps* 4 and 5 in the proof of Theorem 1. Take $R^*=\begin{bmatrix}R\\ 0_{p^2,p^2}\ \ I_{p^2}\ \ 0_{p^2,p^2H}\end{bmatrix}$ where R is defined in (5). Then, by the use of *Step* 3 in the proof of Theorem 1,

$$\begin{aligned}
&\sqrt{n}([\mathrm{vec}(\widehat{M}_I)',\mathrm{vec}(\widehat{\Sigma}^{-1})',\mathrm{vec}(\widehat{K}_1)',\ldots,\mathrm{vec}(\widehat{K}_H)',\mathrm{vec}(\widehat{\Sigma})']'\\
&\qquad-[\mathrm{vec}(M_I)',\mathrm{vec}(\Sigma^{-1})',\mathrm{vec}(K_1)',\ldots,\mathrm{vec}(K_H)',\mathrm{vec}(\Sigma)']')\\
\doteq\ & R^*\sqrt{n}([\mathrm{vec}(\widehat{M}_I)',\mathrm{vec}(\widehat{\Sigma}^{-1})',\mathrm{vec}(\widehat{K}_1)',\ldots,\mathrm{vec}(\widehat{K}_H)')']' \qquad (11)\\
&\qquad-[\mathrm{vec}(M_I)',\mathrm{vec}(\Sigma^{-1})',\mathrm{vec}(K_1)',\ldots,\mathrm{vec}(K_H)']')\\
\longrightarrow_d\ & \mathcal{N}(0,C_{R^*}),
\end{aligned}$$

where $C_{R^*}=R^*\Sigma_2R^{*\prime}$.

For the $p\times p$ matrices A, B, C_1, $\ldots$, C_H, D, let f_3^* be defined from $\mathbb{R}^{p^2+p^2+p^2H+p^2}$ to $\mathbb{R}^{p^2+p^2}$ by

$$\begin{aligned}
&f_3^*([\mathrm{vec}(A)',\mathrm{vec}(B)',\mathrm{vec}(C_1)',\ldots,\mathrm{vec}(C_H)',\mathrm{vec}(D)']')\\
&=\begin{pmatrix}\mathrm{vec}(B[(1-\alpha)ABA+\alpha\sum_{h=1}^{H}C_hBC_h])\\ \mathrm{vec}(D)\end{pmatrix}.
\end{aligned}$$

It is clear that

$$f_3^*([\mathrm{vec}(\widehat{M}_I)',\mathrm{vec}(\widehat{\Sigma}^{-1})',\mathrm{vec}(\widehat{K}_1)',\ldots,\mathrm{vec}(\widehat{K}_H)',\mathrm{vec}(\widehat{\Sigma})']')=\begin{pmatrix}\mathrm{vec}\left(\widehat{\Sigma}^{-1}\widehat{M}_\alpha\right)\\ \mathrm{vec}\left(\widehat{\Sigma}\right)\end{pmatrix},$$

$$f_3^*([\mathrm{vec}(M_I)',\mathrm{vec}(\Sigma^{-1})',\mathrm{vec}(K_1)',\ldots,\mathrm{vec}(K_H)',\mathrm{vec}(\Sigma)']')=\begin{pmatrix}\mathrm{vec}(\Sigma^{-1}M_\alpha)\\ \mathrm{vec}(\Sigma)\end{pmatrix}.$$

Let us also define

$$\mathcal{C}^*=F_3^{*\prime}C_{R^*}F_3^*, \tag{12}$$

where $F_3^*=\frac{\partial f_3^{*\prime}}{\partial u}\Big|_E=\begin{bmatrix}F_3 & 0_{p^2+p^2+Hp^2,p^2}\\ 0_{p^2,p^2} & I_{p^2}\end{bmatrix}$ with F_3 defined in (7). Since the Delta method applies to f_3^* and, from (11), a final application of the Delta method leads to (10).

Therefore, we get $\sqrt{n}(\hat{b}_k - b_k) \longrightarrow_d \Phi_{b_k}$ where $\Phi_{b_k} = (\Sigma^{-1}M_\alpha - \lambda_k I_p)^+ \Phi b_k - \frac{1}{2}(b_k' \Phi_\Sigma b_k) b_k = M_k^* \begin{pmatrix} \mathrm{vec}(\Phi) \\ \mathrm{vec}(\Phi_\Sigma) \end{pmatrix}$ with $M_k^* = [b_k' \otimes (\Sigma^{-1}M_\alpha - \lambda_k I_p)^+ \ -\frac{1}{2} b_k (b_k' \otimes b_k')]$. Then Φ_{b_k} follows the normal distribution $\mathcal{N}(0, \mathcal{C}_{b_k})$ where

$$\mathcal{C}_{b_k} = M_k^* \mathcal{C}^* M_k^{*\prime}. \tag{13}$$

A.4. Proof of Theorem 4

Using Theorem 1, Corollary 4 of Saracco (1997) gives $\Phi_{\lambda_k} = b_k' \Sigma \Phi b_k$. Then we use only the fact that $\Phi_{\lambda_k} = (b_k' \otimes b_k' \Sigma)\mathrm{vec}(\Phi)$ to complete the proof.

References

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

Carroll, R. J. and Li, K. C.(1992). Measurement error regression with unknown link: dimension reduction and data visualization. *J. Amer. Statist. Assoc.* **87**, 1040-1050.

Cook, R. D. and Weisberg, S. (1991). Discussion of "Sliced inverse regression". *J. Amer. Statist. Assoc.* **86**, 328-332.

Duan N. and K. C. Li (1991). Slicing regression: a link-free regression method. *Ann. Statist.* **19**, 505-530.

Ferré, L. (1998). Determining the dimension in Sliced Inverse Regression and related methods. *J. Amer. Statist. Assoc.* **93**, 132-140.

Gannoun, A. and Saracco, J. (2001). Two Cross Validation Criteria for SIR$_\alpha$ and PSIR$_\alpha$ methods in view of prediction. Submitted for publication.

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

Hsing, T. and Carroll, R. J. (1992). An asympotic theory for Sliced Inverse regression. *Ann. Statist.* **20**, 1040-1061.

Kötter, T. (1996). An asymptotic result for Sliced Inverse Regression. *Comput. Statist.* **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. *J. Amer. Statist. Assoc.* **86**, 316-342.

Saracco, J. (1997). An asymptotic theory for Sliced Inverse Regression. *Comm. Statist. Theory Methods* **26**, 2141-2171.

Saracco, J. (2001). Pooled Slicing methods versus Slicing methods. *Comm. Statist. Simulation Comput.* **30**, 489-511.

Serfling, R. L. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley, New York.

Schott, J. R. (1994). Determining the dimensionality in Sliced Inverse Regression. *J. Amer. Statist. Assoc.* **89**, 141-148.

Tyler, D. E. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9**, 725-736.

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

Laboratoire de Probabilités et de Statistique, Département des Sciences Mathématiques, CC 051, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France.

E-mail: gannoun@stat.math.univ-montp2.fr

Laboratoire de Probabilités et de Statistique, Département des Sciences Mathématiques, CC 051, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France.
E-mail: saracco@stat.math.univ-montp2.fr

(Received May 2001; accepted August 2002)