AISM (2008) 60:257–271
DOI 10.1007/s10463-006-0106-0

Consistent estimation of the dimensionality in sliced inverse regression

Guy Martial Nkiet

Received: 27 July 2005 / Revised: 5 June 2006 / Published online: 13 January 2007

Abstract Several methods have been proposed in the literature in order to estimate the dimensionality in sliced inverse regression. Most of these methods are based on sequential tests for the nullity of the last eigenvalues of suitable operators. We first establish non consistency for estimators resulting from these methods. Then, we propose an estimator obtained by minimizing a suitable penalization of a statistic based on eigenvalues. A consistency property is established for this estimator and a simulation study is undertaken to evaluate its finite sample performance.

Keywords Consistency · Dimensionality · Estimation · Sliced inverse regression

1 Introduction

Let X be a random variable defined on a probability space $(\Omega, \mathcal{A}, P)$ and valued into a p-dimensional euclidean space E with inner product $\langle\cdot,\cdot\rangle_E$ and associated norm $\|\cdot\|_E$. When $E = \mathbb{R}^p$ we use the usual inner product $\langle\cdot,\cdot\rangle_{\mathbb{R}^p}$ in $\mathbb{R}^p$ defined by $\langle \mathbf{u}, \mathbf{v}\rangle_E = \mathbf{u}^{'}\mathbf{v} = \sum_{i=1}^{p} u_i v_i$, where $\mathbf{u} = \left(u_1, \ldots, u_p\right)^{'}$ and $\mathbf{v} = \left(v_1, \ldots, v_p\right)^{'}$. Given an univariate response variable Y, we consider the model

$$Y = f\left(\langle b_1, X\rangle_E, \ldots, \langle b_K, X\rangle_E, \varepsilon\right) \tag{1}$$

G. M. Nkiet (✉)
Département de Mathématiques et Informatique, Faculté des Sciences, Université des Sciences et Techniques de Masuku, BP 943, Franceville, Gabon
e-mail: gnkiet@hotmail.com

where K is an integer such that $K < p$, the vectors $b_1, \dots, b_K$ of E are unknown, ε is a random variable that is independent of X, and f is an arbitrary unknown function on $\mathbb{R}^{K+1}$. This model was first introduced by Li (1991); it expresses the fact that the information in X about Y depends only on the projection of X onto the subspace of E spanned by $\{b_1, \dots, b_K\}$, called effective dimension-reduction (EDR) space. In SIR, the problem of estimating the EDR space comes down, under a fairly general condition, to the spectral analysis of the operator $T = V^{-1/2}\Gamma V^{-1/2}$ where V is the covariance operator of X and Γ is an approximation of the covariance operator of the conditional expectation $\mathbb{E}\,(X|\,Y)$, obtained by slicing the range of Y. The main characteristic of formulation (1) is that it permits to reduce the number of variables to use for the estimation of f, since $K < p$. So, estimating K, say the dimensionality, is a very important problem. Several approaches have been proposed in the literature for this problem. First, Li (1991) suggested estimating K by testing successively the nullity of the $p - \ell$ smallest eigenvalues of T, starting at $\ell = 0$; for doing that he proposed a chi squared test under normal assumption for X. Since the resulting method may be nonrobust as it is sensitive to departures from normality, several extensions were studied. Schott (1994) built a method that requires elliptically symmetric distribution and, more recently, Velilla (1998) and Bura and Cook (2001) proposed methods for testing for dimension that do not depend on specific assumptions on the distribution of the regressor. All the preceding methods use sequences of tests. Only Ferré (1998) proposed an approach that is not based on tests. His method consists of estimating the dimensionality by minimizing a convenient measure of the closeness of subspaces of the EDR space and their estimates. However, for this method it is assumed that X has a symmetric elliptic distribution, and that the eigenvalues of T are distinct.

We believe that sequential tests are really not appropriate for the estimation of the dimensionality in the model (1) because the resulting estimate necessarily depends on the nominal significance levels that are used for the related tests and, as it is shown in page 7 of Sect. 2, the corresponding estimator is not consistent. So, there is an interest in introducing a direct estimation method for the dimensionality which does not require testing procedures to be used. Such an approach has recently been adopted for canonical analysis in Nkiet (2005). In this paper, we consider the case of sliced inverse regression, arguably the most popular sufficient dimension reduction method since its introduction in 1991 by Li. We propose an estimation procedure which is based on the minimization of a penalized statistic constructed by using the eigenvalues of a consistent estimator of T. This method only requires that the regressor X have finite second order moment and that, for all $u \in E$, $\mathbb{E}\,(\langle u, X\rangle_E \mid \langle b_1, X\rangle_E, \dots, \langle b_K, X\rangle_E) = c_0 + c_1\,\langle b_1, X\rangle_E + \cdots + c_K\,\langle b_K, X\rangle_E$ for some real constants $c_0, c_1, \dots, c_K$. This latter condition just is the the condition 3.1 of Li (1991). The proposed method is introduced in Sect. 3 and consistency of the related estimator is established. Finally, Sect. 4 is devoted to the presentation of some simulations which was made in order to evaluate finite sample performance of this estimator, and to compare it with estimators resulting from methods based on sequential tests. All the proofs of lemmas and theorems are carried out in Sect. 5.

2 Background and motivation

Denoting by $\mathbb{E}$ the expectation with respect to P, we assume that $\mathbb{E}\left(\|X\|_E^2\right) < +\infty$ and that the covariance operator of X equals the identity of E. It is well known that this latter assumption does not restrict the generality for studying the dimensionality in the model (1) (see Li, 1991; Bura and Cook, 2001). In addition, we suppose that the condition 3.1 of Li (1991) holds. It ensures the EDR space to be generated by eigenvectors of an operator Γ that is given below. Let $\{I_h\}_{1\leq h\leq H}$ be a partition of the range of Y such that, for any $h \in \{1,\ldots,H\}$, one has $p_h := P\left(Y \in I_h\right) > 0$. We consider

$$\mu = \mathbb{E}\left(X\right), \quad \mu_h = \mathbb{E}\left(X|\, Y \in I_h\right), \quad \tau_h = \mu_h - \mu,$$

then the aforementioned operator Γ is an approximation to the covariance operator of $\mathbb{E}\left(X|\, Y\right)$; it is given by

$$\Gamma = \sum_{h=1}^{H} p_h\, \tau_h^{\otimes^2},$$

where $\otimes$ is the usual tensor product between vectors: $u \otimes v$ is the linear map defined by $(u \otimes v)\, x = \langle u, x\rangle\; v$, where $\langle\cdot,\cdot\rangle$ is the inner product of the Euclidean space which u belongs to, and we write $u^{\otimes^2}$ instead of $u \otimes u$. This tensor product is related to some well known matrix operations (see Dauxois et al., 1994). When $E = \mathbb{R}^p$, the operator $\mathbf{u} \otimes \mathbf{v}$ equals the matrix $\mathbf{v}\mathbf{u}'$; so, we can simply write: $\Gamma = \sum_{h=1}^{H} p_h\, \tau_h \tau_h'$. In this paper, we use tensor products and operators instead of matrix notation; this approach is interesting since it allows an easy extension to the case where functional variates are considered, which has been of great recent interest (see Dauxois et al., 2001; Ferré and Yao, 2003).

Estimation of the dimensionality is usually made by estimating the rank d of the operator Γ (e.g., Li, 1991; Velilla, 1998; Bura and Cook, 2001), which is a lower bound to the dimension of the EDR space. Since $d \leq K < p$, this rank belongs to $\{1,\ldots,p-1\}$ and satisfies

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d > \lambda_{d+1} = \cdots = \lambda_p = 0,$$

where λ_i denotes the ith largest eigenvalue of Γ. The problem of estimating d has often been tackled in the literature through the use of sequences of tests for the nullity of the last eigenvalues of a suitable operator (e.g., Li, 1991; Schott, 1994; Velilla, 1998; Bura and Cook, 2001). Our purpose in this section is to show that such an approach yields an estimator of the dimensionality that is not consistent as it does not converge in probability to d, and also that the related performance depends on the significance levels of the individual tests.

Using sequential tests for the estimation of the dimensionality comes down to testing hypotheses of the form $H_\ell : \xi_\ell = 0$, where $\ell \in \{1, \ldots, p-1\}$ and ξ_ℓ is a given nonnegative index. Beginning with $\ell = 1$, one has to do a test for H_ℓ based on a statistic $\widehat{\xi}_\ell^{(n)}$. If H_ℓ is rejected, then one increments ℓ by 1 and repeats the procedure until H_ℓ is not rejected, or until $\ell = p-1$. Then, the estimate of d equals the first integer ℓ for which H_ℓ is not rejected, or equals p if all the hypotheses are rejected. Let us assume that $\widehat{\xi}_\ell^{(n)}$ converges in probability, as $n \to +\infty$, to ξ_ℓ, and that there exists a real $a > 0$ such that, under H_ℓ, $n^a\widehat{\xi}_\ell^{(n)}$ converges in distribution, as $n \to +\infty$, to a distribution with a continuous cumulative distribution function F_ℓ. This situation is common in previous methods; for example, the usual test for dimension considered by Li (1991) and Bura and Cook (2001) is based on the test statistic $\xi_\ell^{(n)} = \sum_{i=\ell+1}^{p} \lambda_i^{(n)}$ which converges almost surely to $\xi_\ell = \sum_{i=\ell+1}^{p} \lambda_i$, and for which we have, under H_ℓ, the convergence of $n\xi_\ell^{(n)}$ to a (weighted) chi-square distribution. Letting $\alpha_\ell \in \left]0,1\right[$ be the nominal significance level of the test for H_ℓ and putting $s_\ell = F_\ell^{-1}(1-\alpha_\ell)$, the hypothesis H_ℓ is rejected if $n^a\widehat{\xi}_\ell^{(n)} > s_\ell$. This test is consistent since $\widehat{\xi}_\ell^{(n)}$ converges in probability to ξ_ℓ. Indeed, under the alternative hypothesis $\overline{H}_\ell$ defined as $\xi_\ell > 0$, we can consider a real $\delta \in \left]0, \xi_\ell\right[$ and an integer n_0 such that $n \geq n_0$ implies $n^{-a}s_\ell < \xi_\ell - \delta$. Therefore, for $n \geq n_0$,

$$P_{\overline{H}_\ell}\left(\left|\widehat{\xi}_\ell^{(n)} - \xi_\ell\right| \leq \delta\right) \leq P_{\overline{H}_\ell}\left(\widehat{\xi}_\ell^{(n)} \geq \xi_\ell - \delta\right) \leq P_{\overline{H}_\ell}\left(n^a\widehat{\xi}_\ell^{(n)} > s_\ell\right)$$

and, consequently, $\lim_{n\to+\infty} P_{\overline{H}_\ell}(n^a\widehat{\xi}_\ell^{(n)} > s_\ell) = 1$. Putting $I = \{1, \ldots, p-1\}$, it is easily seen that the estimator of d resulting from the preceding sequence of tests is given by:

$$\widehat{d}_{ST}^{(n)} = \min\left\{\ell \in I \;/\; n^a\widehat{\xi}_\ell^{(n)} \leq s_\ell\right\}, \tag{2}$$

with the convention $\min(\emptyset) = p$. We will now prove that this estimator does not converge in probability to d. For any $\ell \in \{1, \ldots, d-1\}$, since $\overline{H}_\ell$ is true, we have

$$\lim_{n\to+\infty} P\big(n^a\widehat{\xi}_\ell^{(n)} > s_\ell\big) = 1.$$

Hence,

$$\lim_{n\to+\infty} P\left(\bigcap_{\ell=1}^{d-1}\left\{n^a\widehat{\xi}_\ell^{(n)} > s_\ell\right\}\right) = 1,$$

and since

$$\bigcap_{\ell=1}^{d-1}\left\{n^{a}\widehat{\xi}_{\ell}^{(n)} > s_\ell\right\} \subset \left\{n^{a}\widehat{\xi}_{d}^{(n)} \leq s_d\right\} \cup \bigcap_{\ell=1}^{d-1}\left\{n^{a}\widehat{\xi}_{\ell}^{(n)} > s_\ell\right\}$$

we also have

$$\lim_{n\to+\infty} P\left(\left\{n^{a}\widehat{\xi}_{d}^{(n)} \leq s_d\right\} \cup \bigcap_{\ell=1}^{d-1}\left\{n^{a}\widehat{\xi}_{\ell}^{(n)} > s_\ell\right\}\right) = 1.$$

Moreover, the equality

$$\left\{\widehat{d}_{ST}^{(n)} = d\right\} = \left\{n^{a}\widehat{\xi}_{d}^{(n)} \leq s_d\right\} \cap \bigcap_{\ell=1}^{d-1}\left\{n^{a}\widehat{\xi}_{\ell}^{(n)} > s_\ell\right\}$$

implies

$$\begin{aligned} P\left(\widehat{d}_{ST}^{(n)} = d\right) &= P\left(n^{a}\widehat{\xi}_{d}^{(n)} \leq s_d\right) + P\left(\bigcap_{\ell=1}^{d-1}\left\{n^{a}\widehat{\xi}_{\ell}^{(n)} > s_\ell\right\}\right) \\ &\quad - P\left(\left\{n^{a}\widehat{\xi}_{d}^{(n)} \leq s_d\right\} \cup \bigcap_{\ell=1}^{d-1}\left\{n^{a}\widehat{\xi}_{\ell}^{(n)} > s_\ell\right\}\right); \end{aligned}$$

therefore,

$$\lim_{n\to+\infty} P\left(\widehat{d}_{ST}^{(n)} = d\right) = \lim_{n\to+\infty} P\left(n^{a}\widehat{\xi}_{d}^{(n)} \leq s_d\right) = F_d\,(s_d) = 1 - \alpha_d < 1. \tag{3}$$

Since $\widehat{d}_{ST}^{(n)}$ and d are valued into a subset of $\mathbb{N}$, the formula in (3) is equivalent to the fact that $\widehat{d}_{ST}^{(n)}$ does not converge in probability to d as $n \to +\infty$. So, $\widehat{d}_{ST}^{(n)}$ is a non consistent estimator of d which is an undesirable attribute. However, in practice, the effect of the aforementioned non consistency will not be relevant for values that are usually taken for nominal significance level. Indeed, usually in applications, $\alpha_d = 0.1,\ 0.05,\ 0.01$ and, therefore, the limit in (3) is greater than or equal to 0.9. But a consistent estimator for d would be, of course, preferable to $\widehat{d}_{ST}^{(n)}$, and there is an interest in finding such an estimator.

3 A consistent estimator

Let $\{(X_i, Y_i)\}_{1\leq i\leq n}$ be independent, identically distributed random variables each with the same distribution as (X, Y). We consider the sample mean $\overline{X}_n =$

$n^{-1}\sum_{i=1}^{n} X_i$, and, for $h \in \{1,\ldots,H\}$, the random variables

$$\widehat{n}_h = \sum_{i=1}^{n} \mathbb{I}_{\{Y_i \in I_h\}}, \quad \overline{X}_n^{(h)} = \frac{1}{\widehat{n}_h}\sum_{i=1}^{n} \mathbb{I}_{\{Y_i \in I_h\}} X_i, \quad \widehat{\tau}_h^{(n)} = \overline{X}_n^{(h)} - \overline{X}_n,$$

where $\mathbb{I}_A$ denotes the indicator function of the set A. Then, we consider the moment estimator of Γ:

$$\widehat{\Gamma}_n = \sum_{h=1}^{H} \frac{\widehat{n}_h}{n}\, \widehat{\tau}_h^{(n)\otimes^2}.$$

It is a strongly consistent estimator since, from an obvious application of the strong law of large numbers, we have its almost sure uniform convergence to Γ, as $n \to +\infty$. Denoting by $\widehat{\lambda}_i^{(n)}$ the i-th largest eigenvalue of $\widehat{\Gamma}_n$, we take as estimator of ξ_ℓ $(\ell = 0,\ldots,p-1)$ the random variable

$$\widehat{\xi}_\ell^{(n)} = \sum_{i=\ell+1}^{p} \widehat{\lambda}_i^{(n)}.$$

This statistic is a strongly consistent estimator of ξ_ℓ. Indeed, the almost sure uniform convergence of $\widehat{\Gamma}_n$ to Γ (as $n \to +\infty$) ensures that each $\widehat{\lambda}_i^{(n)}$ converges almost surely, as $n \to +\infty$, to λ_i (see Gohberg and Krejn 1971 or Lemma 2.1 in Tyler 1981) and, therefore, that $\widehat{\xi}_\ell^{(n)}$ converges almost surely to ξ_ℓ as $n \to +\infty$. Note that this statistic is classically used in the literature for estimating the dimensionality via sequential tests for the nullity of the last eigenvalues of Γ under normality assumption (see Li, 1991) or for more general distributions (Velilla, 1998; Bura and Cook, 2001).

Our goal is to make an appropriate penalization of $\widehat{\xi}_\ell^{(n)}$ in order to obtain a consistent estimator of the dimensionality. Let $(k_n)_{n\in\mathbb{N}^*}$ be a sequence of maps from $J := \{0,\ldots,p-1\}$ to $\mathbb{R}_+$ such that there exist a real $\beta \in \left]0, 1/2\right[$ and a striclty increasing function $k : J \to \mathbb{R}_+$ satisfying:

$$\lim_{n\to+\infty} \left(n^{\beta}\, k_n\right) = k. \tag{4}$$

Then, we consider the real random variable

$$\widehat{\phi}_\ell^{(n)} = \widehat{\xi}_\ell^{(n)} + k_n(\ell) \quad (\ell \in J); \tag{5}$$

this statistic is a strongly consistent estimator for ξ_ℓ. Indeed, (4) implies the equality: $\lim_{n\to+\infty} k_n(\ell) = 0$, and since $\widehat{\xi}_\ell^{(n)}$ converges almost surely to ξ_ℓ, as $n \to +\infty$, so does $\widehat{\phi}_\ell^{(n)}$.

We define our estimator of d as the random variable

$$\widehat{d}^{(n)} := \min\left\{\ell \in J \;/\; \widehat{\phi}_\ell^{(n)} = \min_{j\in J}\left(\widehat{\phi}_j^{(n)}\right)\right\}. \tag{6}$$

From (6), it is easily seen that the value $\widehat{d}^{(n)} = p$ can not be taken; that does not cause any problem for the estimation of d since, from the hypotheses, one has $d \leq K < p$. Now, we will state a consistency property for this estimator. This is obtained from the following lemma that gives the limiting distribution of the random operator $\widehat{U}_n = \sqrt{n}\left(\widehat{\Gamma}_n - \Gamma\right)$. Note that, since this lemma just requires that $\mathbb{E}\left(\|X\|_E^2\right) < +\infty$, it is an extension of Theorem 1 in Sarraco (1997) that gives the limiting distribution of $\widehat{U}_n$ under the assumption that X has an elliptic distribution. For two Euclidean spaces F and G, we denote by $\mathcal{L}(F, G)$ the space of operators (linear maps) from F to G; when $F = G$, it is simply denoted by $\mathcal{L}(F)$. Any operator T of $\mathcal{L}\left(\mathbb{R}^H, E \times \mathbb{R}^H\right)$ can be writen as

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

where $T_1 \in \mathcal{L}\left(\mathbb{R}^H, E\right)$, $T_2 \in \mathcal{L}\left(\mathbb{R}^H\right)$; then, we denote by π_1 and π_2 the projectors defined by

$$\pi_1 \;:\; T \in \mathcal{L}\left(\mathbb{R}^H, E \times \mathbb{R}^H\right) \mapsto T_1 \in \mathcal{L}\left(\mathbb{R}^H, E\right),$$
$$\pi_2 \;:\; T \in \mathcal{L}\left(\mathbb{R}^H, E \times \mathbb{R}^H\right) \mapsto T_2 \in \mathcal{L}\left(\mathbb{R}^H\right).$$

Besides, we consider the random vector defined as $W := \sum_{h=1}^H \mathbb{1}_{\{Y\in I_h\}}\, e_h$, where $\{e_1, \ldots, e_H\}$ is the canonical basis of $\mathbb{R}^H$, and the $E\times\mathbb{R}^H$- valued random vector:

$$Z = \begin{pmatrix} X \\ W \end{pmatrix}.$$

Then, we have:

Lemma 1 *Put*

$$U = \sum_{h=1}^{H}\left\{\langle \pi_2(\mathcal{N})\, e_h, e_h\rangle_{\mathbb{R}^H}\left(\mu^{\otimes^2} - \mu_h^{\otimes^2}\right) + (\pi_1(\mathcal{N})\, e_h) \otimes \tau_h + \tau_h \otimes (\pi_1(\mathcal{N})\, e_h)\right\},$$

where $\mathcal{N}$ is a random variable valued into $\mathcal{L}\left(\mathbb{R}^H, E \times \mathbb{R}^H\right)$ and having a centered normal distribution with covariance operator equal to that of $W \otimes Z$. Then $\widehat{U}_n \to U$, in distribution, as $n \to +\infty$.

This lemma permits to obtain the following theorem which gives a consistency property for $\widehat{d}^{(n)}$ as an estimator of d.

Theorem 1 *For the procedure in* (6), $\widehat{d}^{(n)} \to d$, *in probability, as* $n \to +\infty$.

The penalty term $k_n(\ell)$ is introduced in (5) in order to ensure the previous consistency property. As it can be seen in the simultion study in Sect. 4, the method for estimating d that is proposed above is sensitive to the choice of this term. A study for making an optimal choice for k_n may perhaps be done, but we do not pursue that subject in this paper.

4 Simulation results

This section studies empirical aspects of the previous method for estimating the dimensionality. As this method gives an estimate which does not use sequential tests, we call it a direct estimation method (DEM). In order to evaluate its performance on finite samples, we compute percentages of correct estimation (PCE) of the dimensionality over 1,000 replications. We first compute PCE from DEM, with penalty $k_n(\ell) = n^{-0.45}\ell$, and also from two classical methods based on sequential tests that give estimator defined as in (2): the first method (CST) uses the Chi-squared tests of Li (1991) and the second method (ACS) uses the adjusted version of weighted chi-squared tests, as proposed in Bura and Cook (2003). These two latter methods are used with nominal significance level $\alpha = 0.05,\ 0.10$. For sample sizes $n = 50, 100, 500, 700, 900$, and number of slices $H = 7, 10$ and 15, we generate 1000 independent replicates of a structure, with $K = 2$, of the form

$$Y = X_1\,(X_1 + X_5 + 1) + 0.5\,\varepsilon. \tag{7}$$

The dimension p of the regressor X is taken to be $p = 5$; therefore, $X = (X_1, X_2, X_3, X_4, X_5)'$. The error ε in (7) is $N(0,1)$ and is independent from the predictors. For generating X, two different models are used:

* Model A: X is $N_5(0, I_5)$, where I_5 denotes the identity matrix of $\mathbb{R}^5$.

* Model B: X is defined by

$$X_1 = V_3 + V_4 + \frac{W_1}{6}, \quad X_2 = -V_3 + V_4 + \frac{W_1}{6}, \tag{8}$$

$$X_3 = -V_4 + \frac{W_1}{3}, \quad X_4 = V_1 + V_2, X_5 = -V_1 + V_2, \tag{9}$$

where V_1, V_2, V_3, V_4 and W_1 are independent random variables such that V_1, V_2, V_3, V_4 have the exponential distribution with parameter 1, and W_1 has the uniform distibution on $[0, 3]$. The relations (8) and (9) ensure that the regressors in Model B satisfy the condition 3.1 of Li (1991) (see Velilla, 1998, p. 1092). The percentages of correct estimation of the dimensionality obtained from DEM, CST and ACS are reported in Table 1 for Model A, and in Table 2 for Model B. In the normal regressors case, DEM is better than CST and ACS for large sample size (that is, $n = 500, 700, 900$). For moderate sample size ($n = 50, 100$), the conclusion is the same, excepted for $H = 15$ where DEM is outperformed by ACS. In Table 2, the situation is slightly different: DEM gives the best performance

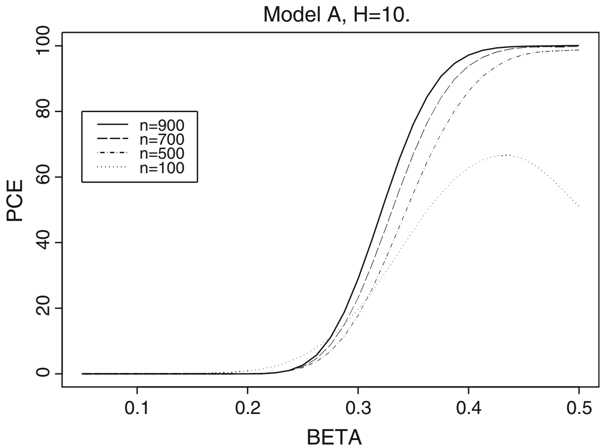

Fig. 1 Model A, $H = 10$

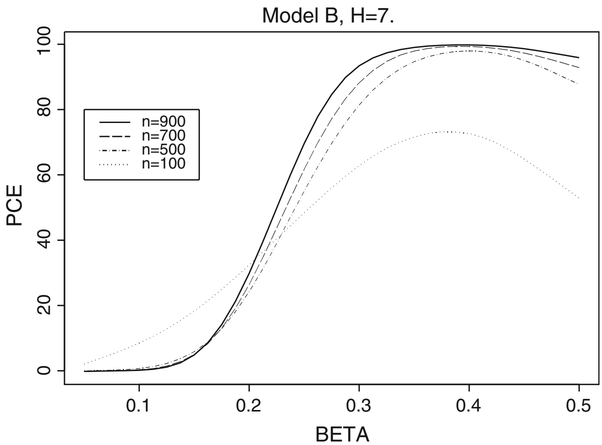

Fig. 2 Model B, $H = 7$

only for $H = 7$. For $H = 10$, it is better than CST for large sample size, but it is outperformed by the two classical methods for moderate sample size. When $H = 15$ it has the worst performance. These results also show that all methods are sensitive to number of slices.

In order to assess of the importance of the choice of β, we compute PCE, over 1000 replications, from DEM with $k_n(\ell) = n^{-\beta}\ell$, and for values of β from 0.05 up to 0.5. The results are given in Fig. 1 for Model A with $H = 10$, and in Fig. 2 for Model B with $H = 7$. It is seen that, when the regressors are normal and the sample size is large, DEM has bad performance for small values of β, and performs better for large values: PCE increases as β becomes more large, until reaching 100% for $\beta > 0.4$. The situation is similar for Model B, excepted for large values of β for which a slight decrease of PCE is observed.

Table 1 Percentages of correct estimation for Model A

n	Method	$H=7$	$H=10$	$H=15$
50	DEM	63.90	47.50	4.10
	CST ($\alpha=0.05$)	4.90	3.70	2.30
	ACS ($\alpha=0.05$)	7.90	11.70	21.00
	CST ($\alpha=0.10$)	8.40	7.80	4.70
	ACS ($\alpha=0.10$)	16.20	20.20	28.40
100	DEM	62.20	69.00	24.40
	CST ($\alpha=0.05$)	21.70	15.80	11.30
	ACS ($\alpha=0.05$)	25.00	23.10	25.60
	CST ($\alpha=0.10$)	31.70	24.70	19.50
	ACS ($\alpha=0.10$)	34.50	34.20	35.80
500	DEM	95.70	98.60	97.80
	CST ($\alpha=0.05$)	95.40	96.00	94.90
	ACS ($\alpha=0.05$)	95.50	95.10	94.20
	CST ($\alpha=0.10$)	91.70	92.00	92.90
	ACS ($\alpha=0.10$)	91.20	91.30	91.20
700	DEM	98.60	99.90	99.30
	CST ($\alpha=0.05$)	96.60	96.20	95.30
	ACS ($\alpha=0.05$)	96.60	95.80	94.80
	CST ($\alpha=0.10$)	92.50	92.00	91.60
	ACS ($\alpha=0.10$)	92.30	91.80	91.70
900	DEM	99.30	100	100
	CST ($\alpha=0.05$)	94.90	95.60	95.90
	ACS ($\alpha=0.05$)	95.00	95.40	95.00
	CST ($\alpha=0.10$)	88.30	90.90	91.60
	ACS ($\alpha=0.10$)	88.50	90.50	90.90

5 Proofs

5.1 Proof of Lemma 1

For $i\in\{1,\ldots,n\}$, we consider the random variables $W_i:=\sum_{h=1}^{H}\mathbb{1}_{\{Y_i\in I_h\}}\,e_h$ and

$$Z_i=\begin{pmatrix}X_i\\W_i\end{pmatrix}.$$

Clearly,

$$W\otimes Z=\begin{pmatrix}\sum_{h=1}^{H}\mathbb{1}_{\{Y\in I_h\}}e_h\otimes X\\ \sum_{h=1}^{H}\mathbb{1}_{\{Y\in I_h\}}e_h^{\otimes 2}\end{pmatrix},\quad W_i\otimes Z_i=\begin{pmatrix}\sum_{h=1}^{H}\mathbb{1}_{\{Y\in I_h\}}e_h\otimes X_i\\ \sum_{h=1}^{H}\mathbb{1}_{\{Y_i\in I_h\}}e_h^{\otimes 2}\end{pmatrix};$$

and since $\mathbb{E}\left(\mathbb{1}_{\{Y\in I_h\}}X\right)=p_h\mu_h$ and $n^{-1}\sum_{i=1}^{n}\mathbb{1}_{\{Y_i\in I_h\}}X_i=\frac{\widehat{n}_h}{n}\overline{X}_n^{(h)}$, we deduce that, putting $\widehat{H}_n:=\sqrt{n}\left[n^{-1}\sum_{i=1}^{n}W_i\otimes Z_i-\mathbb{E}\left(W\otimes Z\right)\right]$, $\widehat{H}_{12}^{(n)}:=\pi_1\left(\widehat{H}_n\right)$

Table 2 Percentages of correct estimation for Model B

n	Method	$H = 7$	$H = 10$	$H = 15$
50	DEM	53.80	15.70	0.70
	CST ($\alpha = 0.05$)	48.90	55.6	56.60
	ACS ($\alpha = 0.05$)	12.60	24.30	29.90
	CST ($\alpha = 0.10$)	52.70	55.70	56.70
	ACS ($\alpha = 0.10$)	20.80	30.10	37.30
100	DEM	65.8	29.80	1.90
	CST ($\alpha = 0.05$)	64.80	64.80	60.20
	ACS ($\alpha = 0.05$)	28.10	36.10	41.80
	CST ($\alpha = 0.10$)	62.70	62.50	56.70
	ACS ($\alpha = 0.10$)	37.90	45.80	51.30
500	DEM	95.10	82.30	37.10
	CST ($\alpha = 0.05$)	73.90	64.90	50.90
	ACS ($\alpha = 0.05$)	95.40	96.40	96.00
	CST ($\alpha = 0.10$)	63.10	53.80	41.90
	ACS ($\alpha = 0.10$)	94.40	93.20	92.60
700	DEM	97.80	89.50	55.30
	CST ($\alpha = 0.05$)	70.30	56.40	49.10
	ACS ($\alpha = 0.05$)	97.00	96.00	95.70
	CST ($\alpha = 0.10$)	59.90	48.60	40.50
	ACS ($\alpha = 0.10$)	94.50	91.80	91.90
900	DEM	98.90	93.20	66.00
	CST ($\alpha = 0.05$)	66.10	56.00	44.10
	ACS ($\alpha = 0.05$)	95.10	95.70	94.60
	CST ($\alpha = 0.10$)	58.40	45.00	36.80
	ACS ($\alpha = 0.10$)	91.30	91.80	88.50

and $\widehat{H}_{22}^{(n)} := \pi_2\left(\widehat{H}_n\right)$, we have:

$$\widehat{H}_{12}^{(n)} = \sum_{h=1}^{H} e_h \otimes \left[\sqrt{n}\left(\frac{\widehat{n}_h}{n}\overline{X}_n^{(h)} - p_h\mu_h\right)\right] \tag{10}$$

$$\widehat{H}_{22}^{(n)} = \sum_{h=1}^{H} \left[\sqrt{n}\left(\frac{\widehat{n}_h}{n} - p_h\right)\right] e_h^{\otimes 2} \tag{11}$$

Moreover, we have for $h \in \{1, \ldots, H\}$:

$$\sqrt{n}\left(\widehat{\Gamma}_n - \Gamma\right) = \sum_{h=1}^{H} \left\{\sqrt{n}\left(\frac{\widehat{n}_h}{n} - p_h\right)\widehat{\tau}_h^{(n)\otimes 2} + p_h\left[\sqrt{n}\left(\widehat{\tau}_h^{(n)} - \tau_h\right)\right] \otimes \widehat{\tau}_h^{(n)} \right.$$
$$\left. + p_h\,\tau_h \otimes \left[\sqrt{n}\left(\widehat{\tau}_h^{(n)} - \tau_h\right)\right]\right\}$$

and

$$\sqrt{n}\left(\widehat{\tau}_h^{(n)}-\tau_h\right)=\frac{n\sqrt{n}}{\widehat{n}_h}\left[\left(\frac{\widehat{n}_h}{n}\overline{X}_n^{(h)}-p_h\,\mu_h\right)-\left(\frac{\widehat{n}_h}{n}-p_h\right)\mu_h\right]-\sqrt{n}\left(\overline{X}_n-\mu\right).$$

Then using the equalities $\mu=\sum_{h=1}^{H}p_h\,\mu_h$ and $\overline{X}_n=\sum_{h=1}^{H}\frac{\widehat{n}_h}{n}\,\overline{X}_n^{(h)}$, together with (10) and (11), we finally obtain $\sqrt{n}(\widehat{\tau}_h^{(n)}-\tau_h)=\widehat{a}_n^{(h)}(\widehat{H}_n)$, where $\widehat{a}_n^{(h)}$ is the random operator

$$A\in\mathcal{L}\left(\mathbb{R}^H,E\times\mathbb{R}^H\right)\mapsto\frac{n}{\widehat{n}_h}\left(\pi_1(A)\,e_h-\langle\pi_2(A)\,e_h,e_h\rangle_{\mathbb{R}^H}\,\mu_h\right)$$
$$-\sum_{\ell=1}^{H}\pi_1(A)\,e_\ell\in E.$$

Consequently, we have $\sqrt{n}\left(\widehat{\Gamma}_n-\Gamma\right)=\widehat{\Psi}_n\left(\widehat{H}_n\right)$ where $\widehat{\Psi}_n$ is the random operator:

$$A\mapsto\sum_{h=1}^{H}\left\{\langle\pi_2(A)\,e_h,e_h\rangle_{\mathbb{R}^H}\;\widehat{\tau}_h^{(n)^{\otimes 2}}+p_h\,\widehat{a}_n^{(h)}(A)\otimes\widehat{\tau}_h^{(n)}+p_h\;\tau_h\otimes\widehat{a}_n^{(h)}(A)\right\}.$$

Since, from the law of large numbers, $\frac{\widehat{n}_h}{n}$ converges almost surely to p_h, as $n\to+\infty$, it is clear that $\left(\widehat{a}_n^{(h)}\right)_{n\in\mathbb{N}^*}$ converges almost surely uniformly to the operator $a^{(h)}$ defined by $a^{(h)}(A)=p_h^{-1}\left(\pi_1(A)\,e_h-\langle\pi_2(A)\,e_h,e_h\rangle_{\mathbb{R}^H}\,\mu_h\right)-\sum_{\ell=1}^{H}\pi_1(A)\,e_\ell$. This property together with the almost sure convergence $\widehat{\tau}_h^{(n)}\to\tau_h$, as $n\to+\infty$, implies that $\left(\widehat{\Psi}_n\right)_{n\in\mathbb{N}^*}$ converges almost surely uniformly to the operator Ψ defined by

$$\Psi(A)=\sum_{h=1}^{H}\left\{\langle\pi_2(A)\,e_h,e_h\rangle_{\mathbb{R}^H}\;\tau_h^{\otimes 2}+p_h\,a^{(h)}(A)\otimes\tau_h+p_h\;\tau_h\otimes a^{(h)}(A)\right\}$$

A more explicit expression can be given for Ψ. Indeed, since

$$\sum_{h=1}^{H}p_h\tau_h=\sum_{h=1}^{H}p_h\,\mu_h-\mu=0,$$

we deduce

$$\sum_{h=1}^{H}\sum_{\ell=1}^{H}p_h\left(\pi_1(A)\,e_\ell\right)\otimes\tau_h=\left(\sum_{\ell=1}^{H}\pi_1(A)\,e_\ell\right)\otimes\left(\sum_{h=1}^{H}p_h\tau_h\right)=0,$$

thus

$$\sum_{h=1}^{H} p_h\, a^{(h)}(A) \otimes \tau_h = \sum_{h=1}^{H} \left\{ (\pi_1(A)\, e_h) \otimes \tau_h - \langle \pi_1(A)\, e_h, e_h \rangle_{\mathbb{R}^H}\ \mu_h \otimes \tau_h \right\}.$$

Therefore, using the equality $\tau_h^{\otimes^2} - \mu_h \otimes \tau_h - \tau_h \otimes \mu_h = \mu^{\otimes^2} - \mu_h^{\otimes^2}$, we finally obtain:

$$\Psi(A) = \sum_{h=1}^{H} \left\{ \langle \pi_2(A)\, e_h, e_h \rangle_{\mathbb{R}^H} \left(\mu^{\otimes^2} - \mu_h^{\otimes^2} \right) + (\pi_1(A)\, e_h) \otimes \tau_h + \tau_h \otimes (\pi_1(A)\, e_h) \right\}.$$

The central limit theorem gives the convergence in distribution, as $n \to +\infty$, of $\widehat{H}_n$ to a random variable $\mathcal{N}$ having a centered normal distribution in $\mathcal{L}\left(\mathbb{R}^H, E \times \mathbb{R}^H\right)$ with covariance operator equal to that of $W \otimes Z$. Further, denoting by $\|\cdot\|$ the usual norm for operators defined by $\|A\| = \sqrt{\operatorname{tr}(AA^*)}$, and by $\|\cdot\|_\infty$ the uniform convergence norm of operators given by $\|A\|_\infty = \sup_{x \neq 0} \|Ax\| / \|x\|$, we have the inequality

$$\left\| \widehat{\Psi}_n\left(\widehat{H}_n\right) - \Psi\left(\widehat{H}_n\right) \right\| \leq \left\| \widehat{\Psi}_n - \Psi \right\|_\infty \left\| \widehat{H}_n \right\|$$

which implies that the sequence $\varepsilon_n := \widehat{\Psi}_n\left(\widehat{H}_n\right) - \Psi\left(\widehat{H}_n\right)$ converges in probability, as $n \to +\infty$, to 0; therefore, $\widehat{\Psi}_n\left(\widehat{H}_n\right)$ and $\Psi\left(\widehat{H}_n\right)$ have the same limiting distribution (because $\widehat{\Psi}_n\left(\widehat{H}_n\right) = \Psi\left(\widehat{H}_n\right) + \varepsilon_n$). Since Ψ is continuous (because it is a linear map from a finite-dimensional space to another) and since $\widehat{H}_n$ converges in distribution to $\mathcal{N}$, we deduce that this limiting distribution is the distribution of $U = \Psi(\mathcal{N})$.

5.2 Proof of Theorem 1

Since

$$\left\{ \widehat{d}^{(n)} = d \right\} = \left(\bigcap_{0 \leq \ell < d} \left\{ \widehat{\phi}_d^{(n)} < \widehat{\phi}_\ell^{(n)} \right\} \right) \cap \left(\bigcap_{d < \ell \leq p-1} \left\{ \widehat{\phi}_d^{(n)} \leq \widehat{\phi}_\ell^{(n)} \right\} \right)$$

with the convention $\bigcap_{\ell \in \emptyset} A_\ell = \Omega$ for any family (A_ℓ) of subsets of Ω, it suffices to prove

$$\lim_{n \to +\infty} P\left(\bigcap_{0 \leq \ell < d} \left\{ \widehat{\phi}_d^{(n)} < \widehat{\phi}_\ell^{(n)} \right\} \right) = 1 \tag{12}$$

and

$$\lim_{n \to +\infty} P\left(\bigcap_{d < \ell \leq p-1} \left\{ \widehat{\phi}_d^{(n)} \leq \widehat{\phi}_\ell^{(n)} \right\} \right) = 1. \tag{13}$$

First, $\widehat{\phi}_d^{(n)}$ converges almost surely to $\xi_d = 0$, as $n \to +\infty$. Further, for any integer ℓ such that $\ell < d$, we have the almost sure convergence of $\widehat{\phi}_\ell^{(n)}$ to $\xi_\ell > 0$. Using these convergence results and the inclusion

$$\left\{\widehat{\phi}_\ell^{(n)} > \xi_\ell - \delta\right\} \cap \left\{\widehat{\phi}_d^{(n)} < \delta\right\} \subset \left\{\widehat{\phi}_d^{(n)} < \widehat{\phi}_\ell^{(n)}\right\}$$

which holds for any $\ell < d$ and any $\delta \in]0, \xi_\ell/2[$, we obtain the equality

$$\lim_{n\to+\infty} P\left(\left\{\widehat{\phi}_d^{(n)} < \widehat{\phi}_\ell^{(n)}\right\}\right) = 1$$

that gives (12) since it holds for all $\ell \in \{1, \ldots, d-1\}$. If $d \geq p-1$, then (13) is obviously obtained from the equality $\bigcap_{\ell\in\emptyset}\left\{\widehat{\phi}_d^{(n)} \leq \widehat{\phi}_\ell^{(n)}\right\} = \Omega$. Now suppose that $d < p-1$ and consider an integer $\ell \in \{d+1, \ldots, p-1\}$, then:

$$\begin{aligned} n^\beta\left(\widehat{\phi}_\ell^{(n)} - \widehat{\phi}_d^{(n)}\right) &= -n^{\beta-1/2}\sum_{i=d+1}^{\ell}\left(\sqrt{n}\widehat{\lambda}_i^{(n)}\right) + n^\beta\left(h_n(\ell) - h_n(d)\right) \\ &= -n^{\beta-1/2}a_\ell'\left[\sqrt{n}\left(\widehat{\Lambda}^{(n)} - \Lambda\right)\right] + n^\beta\left(h_n(\ell) - h_n(d)\right), \end{aligned}$$

where $\widehat{\Lambda}^{(n)} = \left(\widehat{\lambda}_1^{(n)}, \cdots, \widehat{\lambda}_p^{(n)}\right)'$, $\Lambda = (\lambda_1, \cdots, \lambda_p)'$ and $a_\ell = \left(0_d', \mathbb{1}_{\ell-d}', 0_{p-\ell}'\right)'$. Using Lemma 1 and known results (see Eaton and Tyler, 1991), it is easily seen that $\sqrt{n}\left(\widehat{\Lambda}^{(n)} - \Lambda\right)$ converges in distribution to a random vector W. Thus $a_\ell'\left[\sqrt{n}\left(\widehat{\Lambda}^{(n)} - \Lambda\right)\right]$ converges in distribution to $a_\ell'W$ as $n \to +\infty$ and since $\beta < 1/2$, we deduce that $n^{\beta-1/2}a_\ell'\left[\sqrt{n}\left(\widehat{\Lambda}^{(n)} - \Lambda\right)\right]$ converges in probability to 0 as $n \to +\infty$. Consequently, $n^\beta\left(\widehat{\phi}_\ell^{(n)} - \widehat{\phi}_d^{(n)}\right)$ converges in probability to $k(\ell) - k(d)$ as $n \to +\infty$. Since k is strictly increasing, we have $k(\ell) - k(d) > 0$. Let δ be a real satisfying $0 < \delta < k(\ell) - k(d)$, the latter convergence in probability gives:

$$\lim_{n\to+\infty} P\left(n^\beta\left(\widehat{\phi}_\ell^{(n)} - \widehat{\phi}_d^{(n)}\right) > k(\ell) - k(d) - \delta\right) = 1.$$

Finally, since $\left\{n^\beta\left(\widehat{\phi}_\ell^{(n)} - \widehat{\phi}_d^{(n)}\right) > k(\ell) - k(d) - \delta\right\} \subset \left\{\widehat{\phi}_d^{(n)} \leq \widehat{\phi}_\ell^{(n)}\right\}$, it follows that

$$\lim_{n\to+\infty} P\left(\widehat{\phi}_d^{(n)} \leq \widehat{\phi}_\ell^{(n)}\right) = 1, \tag{14}$$

and (13) comes from the fact that (14) holds for any $\ell \in \{d+1, \ldots, p-1\}$.

Acknowledgments The author thanks the two referees and the associate editor for their remarks which permited to improve the original version of the paper.

References

Bura, E., Cook, D. (2001). Extending SIR: the weighted chi-square test. *Journal of the American Statistical Association*, *96*, 996–1003.

Bura, E., Cook, D. (2003). Assessing corrections to the weighted chi-squared test for dimension. *Communications in Statistics - Simulation and Computation*, *32*, 127–146.

Dauxois, J., Ferré, L., Yao, A.F. (2001). Un modèle semi-paramétrique pour variables aléatoires hilbertiennes. *Comptes Rendus de l'Académie des Sciences de Paris, Série I*, *333*, 947–952.

Dauxois, J., Romain, Y., Viguier, S. (1994). Tensor products and statistics. *Linear Algebra and its Applications*, *210*, 59–88.

Eaton, M.L., Tyler, D.E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *The Annals of Statistics*, *19*, 260–271.

Ferré, L. (1998). Determining the dimensionality in sliced inverse regression and related methods. *Journal of the American Statistical Association*, *93*, 132–140.

Ferré, L., Yao, A.F. (2003). Functional sliced inverse regression analysis. *Statistics*, *37*, 475–488.

Gohberg, I.C., Krejn, M.G. (1971). *Introduction à la théorie des opérateurs linéaires autoadjoints dans un espace hilbertien*. Paris: Dunod.

Li, K.C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, *86*, 319–342.

Nkiet, G.M. (2005). On estimation of the dimensionality in linear canonical analysis. *Statistics and Probability Letters*, *75*, 103–112.

Sarraco, J. (1997). An asymptotic theory for sliced inverse regression. *Communications in Statistics - Theory and Methods*, *26*, 2141–2171.

Schott, J.R. (1994). Determining the dimensionality in sliced inverse regression. *Journal of the American Statistical Association*, *89*, 141–148.

Tyler, D.E. (1981). Asymptotic inference for eigenvectors. *The Annals of Statistics*, *9*, 725–736.

Velilla, S. (1998). Assessing the number of linear components in a general regression problem. *Journal of the American Statistical Association*, *93*, 1088–1098.