Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 98 (2007) 970–991
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

On kernel method for sliced average variance estimation☆

Li-Ping Zhu[a,b], Li-Xing Zhu[a,b,∗]

[a] *East China Normal University, Shanghai, China*
[b] *Hong Kong Baptist University, Hong Kong, China*

Received 4 August 2005
Available online 19 January 2007

Abstract

In this paper, we use the kernel method to estimate sliced average variance estimation (SAVE) and prove that this estimator is both asymptotically normal and root n consistent. We use this kernel estimator to provide more insight about the differences between slicing estimation and other sophisticated local smoothing methods. Finally, we suggest a Bayes information criterion (BIC) to estimate the dimensionality of SAVE. Examples and real data are presented for illustrating our method.

AMS 2000 subject classification: 62H12; 62J02

Keywords: Asymptotic normality; Bandwidth selection; Dimension reduction; Kernel estimation; Sliced average variance estimation; Sliced inverse regression; Slicing estimation

1. Introduction

The goal of regression analysis is to understand how the conditional distribution of the response Y given a p-dimensional predictor vector $X = (X_1, \ldots, X_p)^T$ depends on the value assumed by X. Since in many statistical applications the dimension p is large, the statistical analysis becomes difficult. Therefore, it is very important to reduce the dimension p without much loss of information on regression. This has been achieved through the development of sufficient dimension reduction methods. A case in point is the dimension reduction subspace [3,4] which is defined as

☆ The research described here was supported by a Grant (HKBU 7058/05P) from the Research Grants Council of Hong Kong, and a Grant (FRG/05-06/II-02) from Hong Kong Baptist University, Hong Kong, China.

∗ Corresponding author. Hong Kong Baptist University, Hong Kong, China.
E-mail address: lzhu@hkbu.edu.hk (L.-X. Zhu).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2006.11.005

the column space of any $p \times K$ $(K \leqslant p)$ matrix B such that

$$Y \perp\!\!\!\perp X | B^T X, \tag{1.1}$$

where "$\perp\!\!\!\perp$" stands for independence. This means that $B^T X$ is a sufficient statistic for the regression of Y on X. The central dimension reduction (CDR) subspace [4], indicated with $S_{Y|X}$, is defined as the intersection of all dimension reduction subspaces satisfying (1.1). Throughout this article, we assume that CDR space exists unless stated otherwise. Sliced Inverse Regression (SIR) [12] and Sliced Average Variance Estimation (SAVE) [7,5] are two promising tools for identifying and estimating the CDR subspace. SIR uses the mean regression of X given Y, and SAVE is based on the conditional variance of X given Y. To estimate the SIR matrix and then CDR space, Li [12] proposed a simple and useful estimation scheme which has become one of the standard methods in this area. The idea is to divide the whole space of Y into several slices and then to estimate the SIR matrix through the average of sample covariance of X in each slice. This slicing estimation method can also be applied to estimate the SAVE matrix (see, e.g. [7]).

The consistency of this slicing estimation is clearly of importance. Hsing and Carroll [11] and Zhu and Ng [24] proved the asymptotic normality and the root n consistency of the SIR matrix estimator when the number of data points in each slice, say c, ranges from 2 to $\sqrt{n}$, where n is the sample size. Clearly when c is fixed, the estimator is *very undersmoothing*. Their results actually provide theoretical support for Li's [12] empirical study showing that the estimator is not very sensitive to the number of slices. A relevant work is Zhu et al. [27].

In contrast, as Cook and Critchley [6] and Ye and Weiss [19] pointed out, although SAVE is more comprehensive than SIR in the sense that the space spanned by the eigenvectors associated with the non-zero eigenvalues of the SAVE matrix contains the corresponding space based on SIR, the large and finite sample behavior of the slicing estimator of the SAVE matrix greatly depend on the choice of the number of slices as revealed by Cook [5], Cook and Critchley [6] and Zhu et al. [27] empirical studies. This feature was then confirmed by Li and Zhu's [14] theoretical results. Specifically, for continuous Y, when c is fixed, the slicing estimator of the SAVE matrix does not converge. When $c \to \infty$ and $c/\sqrt{n} \to 0$ as $n \to \infty$, the convergence rate becomes $1/c$, and the asymptotic normality does not hold. This inconsistency/slow convergence rate of the slicing estimator deteriorates the performance of SAVE. Only when Y is taking finite values, the root n consistency holds. Li and Zhu [14] proposed a bias corrected method to achieve root n consistency, but the number of slices needs to be selected carefully. How to determine this number via a data-driven selection algorithm still remains an open problem.

Clearly, any local smoothing method can be applied to estimate the SIR matrix. Zhu and Fang [23] used kernel methods to obtain the asymptotic normality of the kernel estimator of SIR when the number of data points in each window ranges from the rate $n^{1/2}$ to $n^{(2d-1)/(2d)}$ where $d \geqslant 2$ presents the degree of smoothness for all related functions to be specified in Section A.1. Note that the kernel estimator can be viewed as a smoothed moving slicing estimator. Li et al. [13] constructed a moving slicing estimator which is used in the contour regression. However, a significant difference between the slicing and kernel estimation is: regarding the number of data points in each slice as a tuning parameter, there is no overlap of ranges for the consistency of the estimators: 2 to $\sqrt{n}$ for slicing estimator and $n^{1/2}$ to $n^{(2d-1)/2d}$ for the kernel estimator.

The above existing results motivate us to pose the following questions which should be of interest and importance for SAVE: can a kernel estimator of the SAVE matrix be asymptotically normal? Compared with the slicing estimation, would it also have a completely different range for the selection of tuning parameters? In this paper, we prove the kernel estimator for the SAVE matrix to be root n consistent for a range of bandwidths, which is similar to the result with the kernel estimator for the SIR matrix.

Furthermore, an important issue in the area of dimension reduction is the estimation of the dimension of CDR space. Therefore, another thrust of our article is that we recommend Bayes information criterion (BIC) to consistently estimate the dimension. The method is a modification of Zhu et al. [25] method.

The paper is organized as follows. In the next section we present the asymptotic results. In Section 3, we discuss the dimensionality determination of CDR space. Simulation results and the illustration of a real application are reported in Section 4, where we also propose a new criterion to choose bandwidth. The appendix contains the proofs of the theoretical results.

2. Asymptotic behavior of the kernel estimator

Write $\Sigma_X > 0$ as the covariance matrix of X. We note that when the standardized variable Z of X, $Z = \Sigma_X^{-\frac{1}{2}}\Big(X - E(X)\Big)$, is used, CDR space $S_{Y|Z} = \Sigma_X^{\frac{1}{2}} S_{Y|X}$ (see [4, Chapters 10 and 11]). Throughout this paper, we use Z and Y to estimate $S_{Y|Z}$ for simplicity. Denote by $f(y)$, the density functions of Y and let Z and its independent copies z_j be

$$Z = (Z_1, \ldots, Z_p)^T, \quad z_j = (z_{1j}, \ldots, z_{pj})^T, \quad j = 1, \ldots, n.$$

We define $A^2 = AA$ for squared symmetric matrix A. Then the SAVE matrix is defined as

$$\Lambda = E\Big(I_p - Cov(Z|Y)\Big)^2 = \bigg(I_p - 2E\Big(Cov(Z|Y)\Big) + E\Big(Cov(Z|Y)\Big)^2\bigg).$$

When SAVE is used to identify the subspace $S_{Y|Z}$, we need to assume the following two conditions:

$$E(Z|P_{S_{Y|Z}}Z) = P_{S_{Y|Z}}Z, \tag{2.1}$$

$$Cov(Z|P_{S_{Y|Z}}Z) = I_p - P_{S_{Y|Z}}. \tag{2.2}$$

where $P_{(.)}$ stands for the projection operator in the standard inner product (see, [4]). Our objective is then to estimate, based on (z_j, y_j)'s, the SAVE matrix Λ, its eigenvalues and the corresponding eigenvectors.

For notational simplicity, write

$$R_{kl}(y) = E(Z_k Z_l|Y = y), \quad G_{kl}(y) = R_{kl}(y)f(y), \quad 1 \leqslant k, l \leqslant p,$$

$$r(y) = E(Z|Y = y) = \Big(E(Z_1|Y = y), \ldots, E(Z_p|Y = y)\Big)^T =: \Big(r_1(y), \ldots, r_p(y)\Big)^T,$$

$$g(y) = \Big(r_1(y)f(y), \ldots, r_p(y)f(y)\Big)^T =: \Big(g_1(y), \ldots, g_p(y)\Big)^T. \tag{2.3}$$

The kernel estimators of $r_k(y)$ and $R_{kl}(y)$ are defined by

$$\hat{g}_i(y) = \frac{1}{nh}\sum_{j=1}^{n} z_{ij} K\Big(\frac{y - y_j}{h}\Big),$$

$$\hat{g}(y) = (\hat{g}_1(y), \ldots, \hat{g}_p(y))^T, \quad \hat{f}(y) = \frac{1}{nh}\sum_{j=1}^{n} K\Big(\frac{y - y_j}{h}\Big),$$

$$\hat{r}(y) = (\hat{r}_1(y), \ldots, \hat{r}_p(y))^T = \hat{g}(y)/\hat{f}(y),$$

$$\hat{G}_{kl}(y) = \frac{1}{nh}\sum_{j=1}^{n} z_{kj} z_{lj} K\Big(\frac{y - y_j}{h}\Big), \quad \hat{R}_{kl}(y) = \hat{G}_{kl}(y)/\hat{f}(y), \tag{2.4}$$

where h is a bandwidth and $K(.)$ is a kernel function.

To define a kernel estimator of Λ, we further introduce some notations. Let $\delta_{kl} = 1$ if $k = l$ and $\delta_{kl} = 0$ otherwise. The klth element λ_{kl} of Λ can be written as, by using the notations of (2.3)

$$\lambda_{kl} = \delta_{kl} - 2E\Big(R_{kl}(Y) - r_k(Y)r_l(Y)\Big) + E\Bigg(\sum_{i=1}^{p}\Big(R_{ki}(Y)R_{il}(Y) - R_{ki}(Y)r_i(Y)r_l(Y) - r_k(Y)r_i(Y)R_{il}(Y) + r_k(Y)r_l(Y)r_i^2(Y)\Big)\Bigg).$$

Through replacing the unknowns by their kernel estimators, the corresponding estimator $\lambda_{n,kl}$ in Λ_n can be

$$\lambda_{n,kl} = \delta_{kl} - \frac{2}{n}\sum_{j=1}^{n}\Big(\hat{R}_{kl}(y_j) - \hat{r}_k(y_j)\hat{r}_l(y_j)\Big) + \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{p}\Bigg(\hat{R}_{ki}(y_j)\hat{R}_{il}(y_j) - \hat{R}_{ki}(y_j)\hat{r}_i(y_j)\hat{r}_l(y_j) - \hat{r}_k(y_j)\hat{r}_i(y_j)\hat{R}_{il}(y_j) + \hat{r}_k(y_j)\hat{r}_l(y_j)\hat{r}_i^2(y_j)\Bigg). \tag{2.5}$$

To present our main theorems, we adopt vectorization of a matrix as follows. For a symmetric $(p \times p)$ matrix $C = (c_{kl})_{p\times p}$, let $V_{ech}(C) = (c_{11}, \ldots, c_{p1}, c_{22}, \ldots, c_{p2}, c_{33}, \ldots, c_{pp})$ be a $p(p+1)/2$ dimensional vector.

We are now in the position to introduce the theoretical results. Define the klth element of matrix $H(Z, Y)$ as

$$\begin{aligned} H_{kl}(Z, Y) = & -\lambda_{kl} + \delta_{kl} - 2\Big(Z_k Z_l - Z_k r_l(Y) - Z_l r_k(Y) + r_l(Y)r_k(Y)\Big) \\ & + \sum_{i=1}^{p}\Bigg(Z_k Z_i R_{il}(Y) + Z_l Z_i R_{ik}(Y) - R_{ki}(Y)R_{il}(Y) - Z_l Z_i r_i(Y)r_k(Y) \\ & - Z_i R_{li}(Y)r_k(Y) - Z_k R_{li}(Y)r_i(Y) + 2r_k(Y)R_{li}(Y)r_i(Y) - Z_k Z_i r_i(Y)r_l(Y) \\ & - Z_i R_{ki}(Y)r_l(Y) - Z_l R_{ki}(Y)r_i(Y) + 2r_l(Y)R_{ki}(Y)r_i(Y) + Z_k r_i^2(Y)r_l(Y) \\ & + Z_l r_i^2(Y)r_k(Y) + 2Z_i r_i(Y)r_k(Y)r_l(Y) - 3r_i^2(Y)r_k(Y)r_l(Y)\Bigg), \end{aligned} \tag{2.6}$$

and for any $\lambda \in R^{p(p+1)/2}$,

$$\sigma_\lambda^2 = \lambda^T Cov(V_{ech}(H(Z, Y))^T)\lambda.$$

The asymptotic normality is stated in the following theorem.

Theorem 1. *In addition to* (2.1) *and* (2.2), *assume that conditions* (1)–(6) *in Section* A.1 *hold. Then as* $n \to \infty$, *we have*

$$\sqrt{n}(\Lambda_n - \Lambda) \to H \quad in\ distribution, \tag{2.7}$$

where $\lambda^T V_{ech}(H)$ *is distributed as* $N(0, \sigma_\lambda^2)$ *for any* $\lambda \neq 0$.

From Theorem 1, we can derive the asymptotic normality of the eigenvalues and of the corresponding eigenvectors by using perturbation theory. The following result is similar to that of SIR obtained by Zhu and Fang [23]. The proof is also quite similar to that for the SIR matrix estimator, hence we omit the details.

Let $\lambda_1(A)\geqslant\lambda_2(A)\geqslant\cdots\geqslant\lambda_p(A)\geqslant 0$ and $b_i(A)=(b_{1i}(A),\ldots,b_{pi}(A))^T$, $i=1,\ldots,p$, denote, respectively, the eigenvalues and their corresponding eigenvectors of a $p\times p$ matrix A.

Theorem 2. *In addition to the conditions of Theorem* 1, *assume that the nonzero $\lambda_i(\Lambda)$'s are distinct. Then for each nonzero eigenvalue $\lambda_i(\Lambda)$ and the corresponding eigenvector $b_i(\Lambda)$, we have*

$$\begin{aligned}&\sqrt{n}(\lambda_i(\Lambda_n)-\lambda_i(\Lambda))\\&=\sqrt{n}b_i(\Lambda)^T(\Lambda_n-\Lambda)b_i(\Lambda)+o_p(\sqrt{n}\|\Lambda_n-\Lambda\|)\\&\to b_i(\Lambda)^T H b_i(\Lambda)\quad \textit{in distribution},\end{aligned}\tag{2.8}$$

where H is given in Theorem 1, *and as $n\to\infty$*,

$$\begin{aligned}&\sqrt{n}(b_i(\Lambda_n)-b_i(\Lambda))\\&=\sqrt{n}\sum_{l=1,l\neq i}^{p}\frac{b_i(\Lambda)b_i(\Lambda)^T(\Lambda_n-\Lambda)b_i(\Lambda)}{\lambda_j(\Lambda)-\lambda_l(\Lambda)}+o_p(\sqrt{n}\|\Lambda_n-\Lambda\|)\\&\to\sum_{l=1,l\neq i}^{p}\frac{b_i(\Lambda)b_i(\Lambda)^T H b_i(\Lambda)}{\lambda_j(\Lambda)-\lambda_l(\Lambda)}\quad \textit{in distribution},\end{aligned}\tag{2.9}$$

where $\|\Lambda_n-\Lambda\|=\sum_{1\leqslant i,j\leqslant p}|a_{ij}|$.

3. Determination of the dimension of $S_{Y|Z}$

The determination of the dimension of $S_{Y|Z}$ is another important issue in the area of dimension reduction. A popular method is the sequential chi-square test method proposed by Li [12]. This method and some other later developed methods, such as Schott [15], Velilla [18], Bura and Cook [1], Ferre [9], are particularly useful in SIR. However, most of these methods depend on the asymptotic normality of the estimators. When SAVE is involved, we either do not have asymptotic normality if the slicing estimator is employed or cannot obtain an asymptotic distribution of the estimator with an easily implemented limiting variance if kernel estimator is used. Hence, we suggest a modified BIC [16] for estimating the dimension of CDR space.

Zhu et al. [25] used a similar algorithm when the dimension of X diverges. The major merit of this methodology is that only the convergence of the estimator of relevant matrix is enough to guarantee the consistency of the estimator of the dimension. We note that this is a general method which can be applied to SAVE. To avoid the inconvenience of selecting the constant in the penalty term, we suggest here a modified version of Zhu et al. [25] algorithm.

Recall the definition of $\lambda_i(A)$ in Section 2. Let $\Omega=\Lambda+I_p$ and $\Omega_n=\Lambda_n+I_p$. Clearly, $\lambda_i(\Omega)=\lambda_i(\Lambda)+1$. Determination of the dimension of $S_{Y|Z}$ now becomes the estimation of K, the number of the eigenvalues of Ω being greater than 1. We define

$$\log L(\lambda(\Omega))=\frac{n}{2}\log|\Omega|-\frac{n}{2}\operatorname{tr}(\Omega^{-1}\Omega_n),\tag{3.1}$$

where $\lambda(\Omega)=(\lambda_1(\Omega),\ldots,\lambda_p(\Omega))^T$. Let Θ_k be the set consisting of all values such that $\lambda_1(\Omega)\geqslant\lambda_2(\Omega)\geqslant\cdots\geqslant\lambda_k(\Omega)>1$ and $\lambda_{k+1}(\Omega)=\cdots=\lambda_p(\Omega)=1$. In addition, let τ denote the number of $\lambda_i(\Omega_n)>1$. Clearly, $\tau=p\geqslant K$ holds almost surely as n tends to infinity. According to

Zhu et al. [25], which is based on Zhao et al. [20,21], we can have an explicit equivalent form of $\sup_{\lambda(\Omega)\in\Theta_k} \log L(\lambda(\Omega))$, that is

$$\sup_{\lambda(\Omega)\in\Theta_k} \log L(\lambda(\Omega)) := \log L_k = \frac{n}{2} \sum_{i=1+\min(\tau,k)}^{p} (\log \lambda_i(\Omega_n) + 1 - \lambda_i(\Omega_n)).$$

Differing from Zhu et al. [25], we consider directly $\lambda_i(\Omega)$ as the parameters. From the above presentation, the supremum of $\log L(\lambda(\Omega))$ over Θ_k only involves $p-k$ parameters. Using exactly the idea of Schwarz's [16] BIC, we define the criterion by

$$G(k) = \log L_k + (p-k)\log n.$$

The second term of $G(k)$ is a penalty and $(p-k)$ equals to the number of $\lambda_i(\Omega)$ needed to be estimated. Similar to Schwarz [16], we include the factor $\log n$ in the penalty. Then the estimator of K is defined as the maximizer $\hat{K}$ of $G(k)$ over $k \in \{0, \ldots, p-1\}$, that is,

$$G(\hat{K}) = \max_{0\leqslant k\leqslant p-1} G(k). \tag{3.2}$$

Theorem 3. *Under the conditions of Theorem* 1, $\hat{K}$ *converges to* K *in probability*.

Remark 3.1. Zhu et al. [25] proposed that the penalty can be of the form c/W_n where c is the number of data points in each slice and W_n is a sequence converging to infinity as n tends to $+\infty$. However, how to select W_n is of concern. In contrast, we simply use $\log n$ as was used by Schwarz [16] and we will see from the simulation results that our BIC works well.

4. Simulations and an application

4.1. Simulations

In this section, we conduct simulation studies to evaluate the performance of kernel estimation and to compare it with the existing methods. Also the efficiency of BIC criterion for the determination of dimension is assessed here. We adopt the criterion proposed by Li [12] to measure the distance between the estimated CDR space and the true CDR space $S_{Y|Z}$. That is, when the estimated CDR space is spanned by $b_i(\Lambda_n)$'s that are associated with the k largest eigenvalues, we use the squared trace correlation, the average of the squared canonical correlation coefficients between $b_i^T(\Lambda_n)z$'s and $\beta_i^T z$'s which span the true CDR space. See Li [12] for more details. For ease of presentation, we denote Li's [12] criterion by R^2. In our simulation results, we will report the median of R^2 from a total of 200 Monte Carlo samples.

In this simulation, we considered the following models with structural dimension $k = 1, 2$.

Model 1: $y = (\beta^T x)^2 \times \epsilon$;

Model 2: $y = (\beta^T x)^2 + \epsilon$;

Model 3: $y = (\beta_1^T x)^3 + (\beta_2^T x)^2 + \epsilon$;

Model 4: $y = (\beta_1^T x)^2 + (\beta_2^T x)^2 \times \epsilon$.

In these models, the covariable x and the error ϵ are independent, and follow respectively normal $N(0, I_{10})$ and $N(0, 1)$, where I_{10} is the 10×10 identity matrix. In the simulations,

Table 1
The empirical R^2 with $n = 400$

		Model 1	Model 2	Model 3	Model 4
SAVE	Kernel	0.9106	0.9724	0.9246	0.9496
SAVE	Slicing ($H = 2$)	0.0567	0.9404	0.6498	0.5122
SAVE	Slicing ($H = 5$)	0.9445	0.9481	0.8678	0.8968
SAVE	Slicing ($H = 10$)	0.9282	0.9359	0.0915	0.8797
SAVE	Slicing ($H = 20$)	0.8996	0.9180	0.4322	0.7927
SAVE	Slicing ($H = 50$)	0.7515	0.8203	0.3191	0.5562

$\beta = (1, 1, 0, 0, \ldots, 0)$ for models 1 and 2, and $\beta_1 = (1, 0, 0, 0, \ldots, 0)$, $\beta_2 = (0, 1, 0, 0, \ldots, 0)$ for models 3 and 4. The basic experiment was replicated to obtain 200 data sets, each of size $n = 400$.

Throughout this section, we used the kernel $K(u) = 15/16(1 - u^2)^2 I_{(|u| \leqslant 1)}$ to estimate the SAVE matrix because this commonly used kernel function possesses some optimality properties [10]. Another important issue in kernel smoothing is the choice of bandwidth h. Note that undersmoothing is needed. This also occurs in model checking, see Zhu [22] and Zhu and Ng [26]. Therefore, we have to select a smaller bandwidth than the one that is optimal in the sense of nonparametric regression estimation. According to Assumption (5), we can select a bandwidth at the convergence rate $n^{-1/3}$. Following the idea of Carroll et al. [2] and Stute and Zhu [17], we propose an algorithm which can be easily implemented. Specifically, we first choose the optimal bandwidth h_{opt} in terms of the generalized cross-validation (GCV) criterion. It is of the rate $n^{-1/5}$. Then, we use $h_{\text{final}} = n^{-\frac{2}{15}} h_{\text{opt}}$ as the resulting bandwidth.

For the sake of comparison, we report in Table 1 the values of R^2 obtained through both kernel and slicing estimation of the SAVE matrix. From the simulation results we can see that the kernel estimator has some advantages. First, the bandwidth in kernel smoothing can be easily selected using the existing data-driven algorithm whereas there is no good a priori estimator of H in practice or theory. What makes the slicing estimation worse is the performance of slicing estimator is sensitive to the number of slices H. Moreover, R^2 for slicing estimation has a large variation for different choices of H. When H is selected properly in model 1, the slicing estimator can perform well. For example, when $H = 5$ and 10, the value of R^2 is slightly larger than the corresponding value obtained by the kernel estimator. However, the improvement in such cases is marginal. The results indicate that for models 2–4, the kernel estimator clearly outperforms the slicing estimator.

Therefore, when SAVE is used, kernel estimation is worthy of recommendation.

Now let us investigate the efficiency of the BIC criterion for determining the structural dimension. The sample size was 400, and the basic experiments were repeated 200 times. The proportions of decisions for dimension using the kernel and the slicing estimator of the SAVE matrix are reported in Tables 2 and 3, respectively. From these two tables, we can see that the slicing estimation tends to overestimate the dimension. Moreover, the number of slices has a significant impact for estimating dimension. Therefore, kernel estimation based determination clearly outperforms the slicing estimation based method.

4.2. An application: wheat protein data

Fearn [8] described a data set from an experiment performed to calibrate a near infrared reflectance (NIR) instrument for the measurement of the protein content in ground wheat samples.

Table 2
The frequency of decisions of dimension with $n = 400$ when the kernel estimator of SAVE is used

	$D = 0$	$D = 1$	$D = 2$	$D = 3$	$D = 4$	$D = 5$	$D = 6$	$D = 7$	$D = 8$	$D = 9$
Model 1	0.005	0.995	0	0	0	0	0	0	0	0
Model 2	0	1	0	0	0	0	0	0	0	0
Model 3	0	0.01	0.885	0.105	0	0	0	0	0	0
Model 4	0	0.06	0.94	0	0	0	0	0	0	0

D stands for dimension.

Table 3
The frequency of decisions of dimension with $n = 400$ when the slicing estimator of SAVE is used

	$D = 0$	$D = 1$	$D = 2$	$D = 3$	$D = 4$	$D = 5$	$D = 6$	$D = 7$	$D = 8$	$D = 9$
Model 1										
$H = 5$	0	0.64	0.335	0.025	0	0	0	0	0	0
$H = 6$	0	0.23	0.55	0.195	0.025	0	0	0	0	0
$H = 10$	0	0.01	0.075	0.48	0.36	0.07	0.005	0	0	0
$H = 20$	0	0	0	0.015	0.085	0.485	0.37	0.045	0	0
Model 2										
$H = 5$	0	0.675	0.305	0.02	0	0	0	0	0	0
$H = 6$	0	0.255	0.56	0.18	0.005	0	0	0	0	0
$H = 10$	0	0.005	0.08	0.46	0.41	0.045	0	0	0	0
$H = 20$	0	0	0	0.02	0.225	0.45	0.26	0.045	0	0
Model 3										
$H = 5$	0	0.045	0.815	0.14	0	0	0	0	0	0
$H = 6$	0	0	0.51	0.425	0.06	0.005	0	0	0	0
$H = 10$	0	0	0.04	0.295	0.575	0.09	0	0	0	0
$H = 20$	0	0	0	0.02	0.12	0.48	0.34	0.04	0	0
Model 4										
$H = 5$	0	0	0.865	0.135	0	0	0	0	0	0
$H = 6$	0	0.01	0.465	0.475	0.05	0	0	0	0	0
$H = 10$	0	0	0.065	0.385	0.42	0.125	0.005	0	0	0
$H = 20$	0	0	0	0.01	0.135	0.5	0.315	0.04	0	0

The protein content measurements of each sample (y in percent) were made using the standard Kjeldahl method, and the six predictors, $L_1, \ldots, L_6$ were measurements on log(1/reflectance) of NIR radiation by the wheat samples of six wavelengths in the range 1680–2310 nm. The calibration is used to find a linear combination of the log values of reflectance which predicts protein content; the coefficients may then be programmed into the instrument so that the protein content of future unknown samples can be read directly. See also the description of Cook [4]. In the analysis, 50 samples of ground wheat were used. The problem here is to determine the structural dimension of y given the six predictors. Cook [4] used 2D added-variable plots and discovered that case 47 and case 24 stands apart from the empirical distribution of the remaining predictor values, and hence these cases are deleted. We now use the other 48 samples in the following analysis.

Cook [4] stated that there is strong collinearity between predictors. Therefore, the linear combination of L_3 and L_4, namely, $L_{34} = 0.856L_3 - 0.517L_4$ was used as a single predictor. We follow Cook's [4] suggestion and use L_{34}, instead of L_3 and L_4, in our analysis.

Table 4
The structural dimension K determined by BIC criterion

Kernel bandwidth	Estimation dim $= K$	Slicing $H(c)$	Estimation dim $= K$
$h_{\text{final}} = 1.187$	$K = 1$	$H = 24(c = 2)$	$K = 4$
		$H = 16(c = 3)$	$K = 4$
		$H = 12(c = 4)$	$K = 1$
		$H = 8(c = 6)$	$K = 1$
		$H = 4(c = 12)$	$K = 1$

Our proposed BIC is used to determine the structural dimension. Using GCV, we can select an optimal bandwidth h_{opt} and the final bandwidth $h_{\text{final}} = n^{-\frac{2}{15}} h_{\text{opt}} = 1.187$. By kernel estimation for SAVE, the dimension of CDR space is 1. For slicing estimation, we considered several values H of the number of slices. We report the dimension determined by BIC in Table 4 where c is the number of points in each slice.

These results indicate that kernel estimation can be conveniently used and slicing estimation performs equally well when H is small. When H becomes comparatively large, slicing method cannot obtain a good estimator of K. This also confirms the theoretical conclusion obtained by Li and Zhu [14].

A. Appendix

A.1. Assumptions

The following conditions are required for Theorems 1 and 2.

(1) All $g_k(y) = r_k(y)f(y)$, $G_{kl}(y) = R_{kl}(y)f(y)$ and $f(y)$ are d-times differentiable and their derivatives satisfy the following condition: letting $H_1(y)$ stand for $f(y)$, $g_k(y)$, $G_{kl}(y)$, respectively, there exists a neighborhood of the origin, say U, and a constant $c > 0$ such that, for any $u \in U$,

$$|H_1^{(d-1)}(y+u) - H_1^{(d-1)}(y)| \leqslant c|u|, \quad \sup_y |H_1^{(1)}(y)| \leqslant c.$$

The constant c can take different value at different places (independent of n) throughout this section.

(2) For each pair $1 \leqslant k, l \leqslant d$ and for any $u \in U$, $|H_2^{(d-1)}(y+u) - H_2^{(d-1)}(y)| \leqslant c|u|$ where $H_2(y)$ stands for, respectively, $R_{kl}(y), r_l(y), r_k(y)r_l(y), R_{ki}(y)R_{il}(y), R_{kl}(y)r_l(y), R_{kl}(y)r_l(y)r_i(y)$ and $r_k(y)r_l(y)r_i(y)$ for each pair $1 \leqslant k, l, i \leqslant d$ and for any $u \in U$.

(3) $E|Z_k Z_l|^4 < \infty$, $k, l = 1 \ldots, d$.

(4) The kernel function $K(.)$ has the following properties:
  (a) the support of $K(.)$ is the interval $[-1, 1]$;
  (b) $K(.)$ is symmetric about 0;
  (c) $\int_{-1}^{1} K(u)\,du = 1$, $\int_{-1}^{1} u^i K(u)\,du = 0$, $i = 1, \ldots, d-1$ and $\int_{-1}^{1} u^d K(u)\,du < \infty$;
  (d) $R_2(K) = \int_{-1}^{1} K^2(u)\,du < \infty$.

(5) As $n \to \infty$, $h \sim n^{-c_1}$ with a positive number c_1 satisfying $\frac{1}{2d} < c_1 < \frac{1}{2}$, the notation "$\sim$" means that h and n^{-c_1} have the same convergence order.
(6) $\inf_y f(y) \geqslant c > 0$ for some positive constant c.

Remark A.1. Conditions (1) and (2) are concerned with the smoothness of the density function of Y and regression curve $R(y)$. These conditions are commonly used. Condition (3) is necessary for the asymptotic normality of Λ_n. Condition (4) is for the use of dth order kernel. Condition (5) shows the range of bandwidths for asymptotic normality. Clearly, it is fairly wide, but an undersmoothing is needed because the optimal bandwidth $O(n^{-\frac{1}{2d+1}})$ is not in this range. On the other hand, since $hn^{1/2} \to \infty$ as $n \to \infty$, thus the number of data points within each slice cannot be too small. Therefore, compared with the results of Li and Zhu [14], this confirms the significant difference between the slicing and kernel estimation in this circumstance.

A.2. Lemmas

Since the proof of Theorem 1 is rather long, we split the proof into several lemmas. The following lemmas present the results that the elements of Λ_n can be written as U-statistics and then can be approximated by sums of $i.i.d.$ random variables.

Lemma A.1. *Suppose conditions* (1), (4)–(6) *are satisfied. Then*

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\hat{h}_1(y_j) - h_1(y_j)\right)\left(\hat{h}_2(y_j) - h_2(y_j)\right) \Big/ \hat{f}^2(y_j) = o_p(1),$$

where both $h_1(.)$ and $h_2(.)$ can be $f(.)$, $g_k(.)$ and $G_{kl}(.)$ for each pair $1 \leqslant k, l \leqslant q$.

Lemma A.2. *Suppose conditions* (1)–(6) *are satisfied. Then*

$$\frac{1}{\sqrt{n}}\left(\sum_{j=1}^{n} H(y_j)\hat{f}(y_j) - \sum_{j=1}^{n} H(y_j)f(y_j)\right) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big(H(y_j)f(y_j) - E(H(Y)f(Y))\Big) + o_p(1),$$

where $H(.)$ can be $\frac{R_{kl}(.)}{f(.)}$, $\frac{r_k(.)r_l(.)}{f(.)}$, $\frac{R_{ki}(.)R_{il}(.)}{f(.)}$, $\frac{r_k(.)r_l(.)r_i(.)}{f(.)}$ and $\frac{R_{kl}(.)r_k(.)r_l(.)}{f(.)}$.

Lemma A.3. *Suppose conditions* (1)–(6) *are satisfied. Then*

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big(H(y_j)\hat{g}_k(y_j) - H(y_j)g_k(y_j)\Big) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big(z_{kj} H(y_j) f(y_j) - EH(Y)g_k(Y)\Big) + o_p(1),$$

where $H(.)$ can be $\frac{r_l(.)}{f(.)}$ $\frac{r_l(.)r_k(.)}{f(.)}$, $\frac{r_l(.)r_k(.)r_i(.)}{f(.)}$ and $\frac{R_{ki}(.)r_l(.)}{f(.)}$.

Lemma A.4. *Suppose conditions* (1)–(6) *are satisfied. Then*

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(H(y_j)\hat{G}_{kl}(y_j)-H(y_j)G_{kl}(y_j)\Big)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(z_{kj}z_{lj}H(y_j)f(y_j)-EH(Y)G_{kl}(Y)\Big)+o_p(1),$$

where $H(.)$ *can be* $\frac{1}{f(.)}$, $\frac{R_{kl}(.)}{f(.)}$ *and* $\frac{r_k(.)r_l(.)}{f(.)}$.

The proofs are left to Section A.4.

A.3. Proofs of the theorems

Proof of Theorem 1. We need only to deal with the klth element $\lambda_{n,kl}$ of Λ_n. The proof has been divided into five steps. In each step, our major target is to approximate each term by a sum of $i.i.d.$ random variables. First recall the definitions of $R_{kl}(\cdot)=G_{kl}(\cdot)/f(\cdot)$ and $\hat{R}_{kl}(\cdot)=\hat{G}_{kl}(\cdot)/\hat{f}(\cdot)$ defined in (2.3). Looking at the formula of $\lambda_{n,kl}$ of (2.5), in the following we deal with the involved estimators to derive asymptotic linear representations in (2.6). Recall the definition of the conditional expectation in (2.3). Without confusion, we write $E(\cdot|Y=y)=E(\cdot|y)$ and $\hat{E}(\cdot|Y=y)=\hat{E}(\cdot|y)$ throughout this proof. The proof can be done through the asymptotic linear representations of U-statistics in the lemmas.

Step 1: By Lemma A.1, we have

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\hat{R}_{kl}(y_j)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(\hat{G}_{kl}(y_j)-G_{kl}(y_j)+G_{kl}(y_j)\Big)\left(\frac{f(y_j)-\hat{f}(y_j)}{f(y_j)\hat{f}(y_j)}+\frac{1}{f(y_j)}\right)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{\hat{G}_{kl}(y_j)-G_{kl}(y_j)}{f(y_j)}+\frac{G_{kl}(y_j)}{f(y_j)}+\frac{G_{kl}(y_j)\Big(f(y_j)-\hat{f}(y_j)\Big)}{f(y_j)^2}\right)+o_p(1).$$

Therefore, by Lemmas A.2 and A.4

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(\hat{R}_{kl}(y_j)-ER_{kl}(y_j)\Big)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{\hat{G}_{kl}(y_j)-G_{kl}(y_j)}{f(y_j)}+\Big(R_{kl}(y_j)-ER_{kl}(y_j)\Big)\right.$$
$$\left.+\frac{G_{kl}(y_j)\Big(f(y_j)-\hat{f}(y_j)\Big)}{f(y_j)^2}\right)+o_p(1)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(z_{kj}z_{lj}-ER_{kl}(Y)\Big)+o_p(1). \tag{A.1}$$

Step 2: By Lemma A.1, we have

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\hat{r}_k(y_j)\hat{r}_l(y_j)$$

$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(\hat{g}_k(y_j)-g_k(y_j)+g_k(y_j)\Big)\Big(\hat{g}_l(y_j)-g_l(y_j)+g_l(y_j)\Big)$$

$$\times\left(\frac{\Big(f(y_j)-\hat{f}(y_j)\Big)^2+2f(y_j)\Big(f(y_j)-\hat{f}(y_j)\Big)}{f^2(y_j)\hat{f}^2(y_j)}+\frac{1}{f^2(y_j)}\right)$$

$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(\hat{g}_k(y_j)-g_k(y_j)+g_k(y_j)\Big)\Big(\hat{g}_l(y_j)-g_l(y_j)+g_l(y_j)\Big)$$

$$\times\left(\frac{2f(y_j)\Big(f(y_j)-\hat{f}(y_j)\Big)}{f^4(y_j)}+\frac{1}{f^2(y_j)}\right)+o_p(1)$$

$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{-2g_k(y_j)g_l(y_j)\Big(\hat{f}(y_j)-f(y_j)\Big)}{f^3(y_j)}+\frac{g_l(y_j)\Big(\hat{g}_k(y_j)-g_k(y_j)\Big)}{f^2(y_j)}\right.$$

$$\left.+\frac{g_k(y_j)\Big(\hat{g}_l(y_j)-g_l(y_j)\Big)}{f^2(y_j)}+\frac{g_k(y_j)g_l(y_j)}{f^2(y_j)}\right)+o_p(1).$$

Lemmas A.2 and A.3 imply

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(\hat{r}_k(y_j)\hat{r}_l(y_j)-Er_k(Y)r_l(Y)\Big)$$

$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(z_{kj}r_l(y_j)+z_{lj}r_k(y_j)-r_k(y_j)r_l(y_j)-Er_k(Y)r_l(Y)\Big)+o_p(1). \qquad \text{(A.2)}$$

Step 3: Using Lemma A.1, we have

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\hat{R}_{ki}(y_j)\hat{R}_{il}(y_j)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\Big(\hat{R}_{ki}(y_j)-R_{ki}(y_j)\Big)R_{il}(y_j)$$

$$+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\Big(\hat{R}_{il}(y_j)-R_{il}(y_j)\Big)R_{ki}(y_j)$$

$$+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}R_{ki}(y_j)R_{il}(y_j)+o_p(1).$$

Therefore, Lemma A.4 yields

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\left(\hat{R}_{ki}(y_j)\hat{R}_{il}(y_j)-E\Big(R_{ki}(Y)R_{il}(Y)\Big)\right)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\Bigg(z_{kj}z_{ij}R_{il}(y_j)+z_{lj}z_{ij}R_{ik}(y_j)-R_{ki}(y_j)R_{il}(y_j)$$
$$-E\Big(R_{ki}(Y)R_{il}(Y)\Big)\Bigg)+o_p(1). \tag{A.3}$$

Step 4: Use Lemma A.1 again to obtain

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\left(\hat{r}_i^2(y_j)\hat{r}_k(y_j)\hat{r}_l(y_j)\right)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\Bigg(r_i^2(y_j)r_l(y_j)\Big(\hat{r}_k(y_j)-r_k(y_j)\Big)+2r_k(y_j)r_l(y_j)r_i(y_j)\Big(\hat{r}_i(y_j)-r_i(y_j)\Big)$$
$$+r_i^2(y_j)r_k(y_j)\Big(\hat{r}_l(y_j)-r_l(y_j)\Big)+r_i^2(y_j)r_k(y_j)r_l(y_j)\Bigg)+o_p(1).$$

By Lemmas A.2 and A.3, we have

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\left(\hat{r}_i^2(y_j)\hat{r}_k(y_j)\hat{r}_l(y_j)-E\Big(r_i^2(Y)r_k(Y)r_l(Y)\Big)\right)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\Bigg(z_{kj}r_i^2(y_j)r_l(y_j)+z_{lj}r_i^2(y_j)r_k(y_j)+2z_{ij}r_i(y_j)r_k(y_j)r_l(y_j)$$
$$-3r_i^2(y_j)r_k(y_j)r_l(y_j)-E\Big(r_i^2(Y)r_k(Y)r_l(Y)\Big)\Bigg)+o_p(1). \tag{A.4}$$

Step 5: Invoking Lemma A.1 again, we derive

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\hat{R}_{ki}(y_j)\hat{r}_i(y_j)\hat{r}_l(y_j)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\Bigg(\Big(\hat{R}_{ki}(y_j)-R_{ki}(y_j)\Big)r_i(y_j)r_l(y_j)+\Big(\hat{r}_i(y_j)-r_i(y_j)\Big)R_{ki}(y_j)r_l(y_j)$$
$$+\Big(\hat{r}_l(y_j)-r_l(y_j)\Big)r_i(y_j)R_{ki}(y_j)+r_l(y_j)r_i(y_j)R_{ki}(y_j)\Bigg).$$

Therefore, by Lemmas A.2–A.4, we achieve

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\left(\hat{R}_{ki}(y_j)\hat{r}_i(y_j)\hat{r}_l(y_j)-E\Big(r_l(Y)R_{ki}(Y)r_i(Y)\Big)\right)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\Bigg(z_{kj}z_{ij}r_i(y_j)r_l(y_j)+z_{ij}R_{ki}(y_j)r_l(y_j)+z_{lj}R_{ki}(y_j)r_i(y_j)$$
$$-2r_l(y_j)R_{ki}(y_j)r_i(y_j)-E\Big(r_l(Y)R_{ki}(Y)r_i(Y)\Big)\Bigg)+o_p(1). \tag{A.5}$$

Using the similar arguments, we have

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\left(\hat{R}_{li}(y_j)\hat{r}_i(y_j)\hat{r}_k(y_j)-E\Big(r_k(Y)R_{li}(Y)r_i(Y)\Big)\right)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{p}\Bigg(z_{lj}z_{ij}r_i(y_j)r_k(y_j)+z_{ij}R_{li}(y_j)r_k(y_j)+z_{kj}R_{li}(y_j)r_i(y_j)$$
$$-2r_k(y_j)R_{li}(y_j)r_i(y_j)-E\Big(r_k(Y)R_{li}(Y)r_i(Y)\Big)\Bigg)+o_p(1). \tag{A.6}$$

Finally, by combining the results of (A.1)–(A.6), we have proved that $\lambda_{n,kl}$ can be written asymptotically as a sum of $i.i.d.$ random variables. Hence, Central Limit Theorem yields the desired result with the variance of (2.6). □

Proof of Theorem 3. Let K be the true value of the dimension of Ψ. Note that

$$G(K)-G(k)=\log L_K-\log L_k-(K-k)\log n.$$

With probability one, we have that from Theorem 2, in probability for large n, $\lambda_i(\Omega_n)>1, i=1,\ldots,K$ and $\min(\tau,K)=K$.

If $k<K$, then $\min(\tau,k)=k$. Thus for large n,

$$\log L_K-\log L_k=-\frac{n}{2}\sum_{i=k+1}^{K}(\log\lambda_i(\Omega_n)+1-\lambda_i(\Omega_n))=\frac{n}{2}W_n(K,k),$$

where

$$W_n(K,k)=-\sum_{i=k+1}^{K}(\log\lambda_i(\Omega_n)+1-\lambda_i(\Omega_n)).$$

We have, for large n,

$$\lim_{n\to\infty}W_n(K,k)=W(K,k)\equiv-\sum_{i=k+1}^{K}(\log\lambda_i(\Omega_n)+1-\lambda_i(\Omega_n))>0.$$

Hence, in probability, we have that for large n

$$G(K)-G(k)>\tfrac{1}{4}nW(K,k)-(K-k)\log n>0. \tag{A.7}$$

If $k > K$, we note that for $i = K+1, \ldots, k$, $\lambda_i(\Omega_n) - 1 = O_p(1/\sqrt{n})$, and $\log \lambda_i(\Omega_n) + (1-\lambda_i(\Omega_n)) = -(1-\lambda_i(\Omega_n))^2/2 + o_p(1/n) = O_p(1/n)$. Furthermore, $K - k < 0$. Then in probability

$$G(K) - G(k) > 0. \tag{A.8}$$

It follows from (A.7) and (A.8) that $\hat{K} \to K$. □

A.4. Proofs of the lemmas

Proof of Lemma A.1. We only need to show this lemma when $h_1(.)$ and $h_2(.)$ are the same because this lemma can be proven easily by the Cauchy inequality when $h_1(.)$ and $h_2(.)$ are different. Also we only prove the case with $h_1(.) = h_1(.) = g_k(.)$ because the proof for other cases are essentially the same. First of all, we show that $\frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\hat{g}_k(y_j) - g_k(y_j)\right)^2$ is $o_p(1)$ by rewriting it as a U-statistic. The method is exactly identical to the one developed by Zhu and Fang [23] to prove their results. The following is an outline of the proof.

First, we can easily obtain that, invoking the conditions,

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{j=1}^n \hat{g}_k^2(y_j) &= \frac{\sqrt{n}}{n^3} \sum_{j=1}^n \sum_{i=1}^n \sum_{l=1}^n z_{ki} z_{kl} \frac{1}{h^2} K\Big(\frac{y_j - y_i}{h}\Big) K\Big(\frac{y_j - y_l}{h}\Big) \\
&= \frac{\sqrt{n}}{C_n^3} \sum_{i<j<l} z_{ki} z_{kl} \frac{1}{h^2} K\Big(\frac{y_j - y_i}{h}\Big) K\Big(\frac{y_j - y_l}{h}\Big) + o_p(1) \\
&=: \frac{\sqrt{n}}{C_n^3} \sum_{i<j<l} h(z_{ki}, z_{kl}, z_{kj}, y_i, y_j, y_k) + o_p(1) \\
&=: \sqrt{n} U_{n1} + o_p(1).
\end{aligned}$$

Similarly, we can derive that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n g_k(y_j) \hat{g}_k(y_j) =: \sqrt{n} U_{n2} + o_p(1).$$

Furthermore, by the properties of the conditional expectation, it is easy to derive that

$$\begin{aligned}
\sqrt{n} E(U_{n1}) &= \sqrt{n} E\left(Z_{k2} Z_{k3} \frac{1}{h^2} K\left(\frac{Y_1 - Y_2}{h}\right) K\left(\frac{Y_1 - Y_3}{h}\right)\right) \\
&= \sqrt{n} E\left(E\left(Z_{k2} Z_{k3} \frac{1}{h^2} K\left(\frac{y_1 - Y_2}{h}\right) K\left(\frac{y_1 - Y_3}{h}\right) \middle| y_1 \right)\right) \\
&= \sqrt{n} E\left(E\left(\frac{1}{h} K\left(\frac{y_1 - Y_2}{h}\right) r_k(Y_2) \middle| y_1 \right) E\left(\frac{1}{h} K\left(\frac{y_1 - Y_3}{h}\right) r_k(Y_3) \middle| y_1 \right)\right) \\
&= \sqrt{n} E\left(\int \frac{1}{h} K\left(\frac{y_1 - y_2}{h}\right) r_k(y_2) f(y_2)\, dy_2 \int \frac{1}{h} K\left(\frac{y_1 - y_3}{h}\right) r_k(y_3) f(y_3)\, dy_3 \right) \\
&= \sqrt{n} E\left(g_k^2(Y_1) + O(h^d) \right) = \sqrt{n} E\left(g_k^2(Y_1) \right) + o(1).
\end{aligned}$$

Also we have

$$\sqrt{n}E(U_{n2}) = \sqrt{n}E\left(Z_{k2}\frac{1}{h}K\left(\frac{Y_1-Y_2}{h}\right)g_k(Y_1)\right)$$
$$= \sqrt{n}\int \frac{1}{h}K\left(\frac{y_1-y_2}{h}\right)r_k(y_2)g_k(y_1)f(y_2)f(y_1)\,dy_2\,dy_1$$
$$= \sqrt{n}E\left(g_k^2(Y_1)\right) + O(\sqrt{n}h^d) = \sqrt{n}E\left(g_k^2(Y_1)\right) + o(1).$$

Therefore, we have

$$E\left(\frac{1}{\sqrt{n}}\sum_{j=1}^n\left(\hat{g}_k(Y_j) - g_k(Y_j)\right)^2\right)$$
$$= E\left(\frac{1}{\sqrt{n}}\sum_{j=1}^n\left(\hat{g}_k^2(Y_j) - 2\hat{g}_k(Y_j)g_k(Y_j) + g_k^2(Y_j)\right)\right)$$
$$= \sqrt{n}E(U_{n1}) - 2\sqrt{n}E(U_{n2}) + E\left(\frac{1}{\sqrt{n}}\sum_{j=1}^n g_k^2(Y_j)\right) + o(1)$$
$$= o(1).$$

Since $\hat{f}(y_j)$ is uniformly consistent to $f(y_j)$ over j, by condition 6 and Markov inequality, we obtain

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n\left(\frac{\hat{g}_k(y_j)-g_k(y_j)}{\hat{f}(y_j)}\right)^2 = \frac{1}{\sqrt{n}}\sum_{j=1}^n\left(\frac{\hat{g}_k(y_j)-g_k(y_j)}{f(y_j)}\right)^2 + o_p(1)$$
$$\leqslant \frac{1}{\sqrt{n}}\sum_{j=1}^n b^{-2}\left(\hat{g}_k(y_j)-g_k(y_j)\right)^2 + o_p(1) = o_p(1).$$

The proof is concluded. □

Since the proof of Lemmas A.2–A.4 are essentially the same, we only prove Lemma A.3 here.

Proof of Lemma A.3. We first write $\frac{1}{\sqrt{n}}\sum_{j=1}^n H(y_j)\hat{g}_k(y_j)$ as a U-statistic:

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n H(y_j)\hat{g}_k(y_j)$$
$$= \frac{1}{n^{3/2}}\sum_{i=1}^n\sum_{j=1}^n z_{ki}H(y_j)\frac{1}{h}K\left(\frac{y_j-y_i}{h}\right)$$
$$= \sqrt{n}\frac{1}{C_n^2}\sum_{i<j}\frac{z_{ki}H(y_j)\frac{1}{h}K\left(\frac{y_j-y_i}{h}\right) + z_{kj}H(y_i)\frac{1}{h}K\left(\frac{y_j-y_i}{h}\right)}{2} + o_p(1)$$
$$=: \sqrt{n}\frac{1}{C_n^2}\sum_{i<j}u_h(z_{ki},y_i,z_{kj},y_j) + o_p(1) = \sqrt{n}U_n + o_p(1), \quad (A.9)$$

where $o_p(1)$ is the sum of all terms with $i = j$. To prove this lemma, we then show that U_n can be approximated by its projection

$$\hat{U}_n = \sum_{j=1}^{n} E(U_n|z_{kj}, y_j) - (n-1)Eu_h(Z_{k1}, Y_1, Z_{k2}, Y_2), \tag{A.10}$$

where $u_h(.)$ is the kernel of the U-statistic U_n. Note that $\hat{U}_n$ is not a sum of $i.i.d.$ random variables. In the following we prove that $\hat{U}_n$ can be asymptotically equivalent to a sum of $i.i.d.$ random variables. To compute EU_n first, we can obtain that

$$\begin{aligned} EU_n &= Eu_h(Z_{k1}, Y_1; Z_{k2}, Y_2) = E\left(Z_{k1} H(Y_2) \frac{1}{h} K\left(\frac{Y_2 - Y_1}{h} \right) \right) \\ &= E\left(H(Y_2) \frac{1}{h} K\left(\frac{Y_2 - Y_1}{h} \right) r_k(Y_1) \right) = E\Big(H(Y) g_k(Y) \Big) + O(h^d). \end{aligned}$$

Note that

$$\begin{aligned} u_1(z_{k1}, y_1) =: & E\Big(u_h(z_{k1}, y_1; Z_{k2}, Y_2) | z_{k1}, y_1 \Big) \\ &= \frac{z_{k1}}{2} \int H(y_2) \frac{1}{h} K\Big(\frac{y_2 - y_1}{h} \Big) f(y_2)\, dy_2 \\ &\quad + E\left(\frac{H(y_1)}{2} \frac{1}{h} K\left(\frac{y_2 - y_1}{h} \right) r_k(y_2) \middle| z_{k1}, y_1 \right) \\ &= \frac{z_{k1}}{2} \int H(y_1 + ht) K(t) f(y_1 + ht)\, dt + \frac{H(y_1)}{2} \int K(t) g_k(y_1 + ht)\, dt \\ &= \frac{z_{k1} H(y_1) f(y_1) + H(y_1) g_k(y_1)}{2} + O_p(h^d). \end{aligned}$$

Thus, the centered conditional expectation is as follows:

$$\begin{aligned} \tilde{u}_{h1}(z_{k1}, y_1) &= E\Big(u_h(z_{k1}, y_1; Z_{k2}, Y_2) | z_{k1}, y_1 \Big) - E\Big(u_h(Z_{k1}, Y_1; Z_{k2}, Y_2) \Big) \\ &= \frac{z_{k1} H(y_1) f(y_1) + H(y_1) g_k(y_1)}{2} - E\Big(H(Y) g_k(Y) \Big) + O_p(h^d). \end{aligned}$$

From the above, we have

$$\begin{aligned} E(U_n|z_{k1}, y_1) &= \frac{1}{n(n-1)} \sum_{i \neq j} E\Big(u_h(Z_{ki}, Z_{kj}, Y_i, Y_j) | z_{k1}, y_1 \Big) \\ &= \frac{1}{n(n-1)} \left(\sum_{i \neq j, i \text{ or } j \neq 1} E\Big(u_h(Z_{ki}, Z_{kj}, Y_i, Y_j) | z_{k1}, y_1 \Big) \right. \\ &\quad \left. + \sum_{i \neq j, i \text{ or } j = 1} E\Big(u_h(Z_{ki}, Z_{kj}, Y_i, Y_j) | z_{k1}, y_1 \Big) \right) \\ &= \frac{1}{n(n-1)} \left(\sum_{i \neq j, i \text{ or } j \neq 1} Eu_h(Z_{ki}, Z_{kj}, Y_i, Y_j) \right. \\ &\quad \left. + \sum_{i \neq j, i \text{ or } j = 1} E\Big(u_h(Z_{ki}, Z_{kj}, Y_i, Y_j) | z_{k1}, y_1 \Big) \right) \end{aligned}$$

$$= \frac{1}{n(n-1)}\Bigg((n-1)(n-2)Eu_h(Z_{k1}, Z_{k2}, Y_1, Y_2)$$
$$+ 2(n-1)E\Big(u_h(z_{k1}, Z_{k2}, y_1, Y_2)|z_{k1}, y_1\Big)\Bigg)$$
$$= \frac{n-2}{n}Eu_h(Z_{k1}, Z_{k2}, Y_1, Y_2) + \frac{2}{n}E\Big(u_h(z_{k1}, Z_{k2}, y_1, Y_2)|z_{k1}, y_1\Big).$$

We will see that the projection $\hat{U}_n$ of the U-statistic U_n can be approximated as the sum of a sequence of $i.i.d$ random variables as follows:

$$\hat{U}_n - Eu_h(Z_{k1}, Z_{k2}, Y_1, Y_2)$$
$$= \sum_{j=1}^{n} E(U_n|z_{kj}, y_j) - (n-1)Eu_h(Z_{k1}, Z_{k2}, Y_1, Y_2) - Eu_h(Z_{k1}, Z_{k2}, Y_1, Y_2)$$
$$= \frac{2}{n}\sum_{j=1}^{n} E\Big(u_h(Z_{ki}, z_{kj}, Y_i, y_j)|z_{kj}, y_j\Big) - Eu_h(Z_{k1}, Z_{k2}, Y_1, Y_2)$$
$$- Eu_h(Z_{k1}, Z_{k2}, Y_1, Y_2)$$
$$= \frac{2}{n}\sum_{j=1}^{n}\Bigg(E\Big(u_h(Z_{ki}, z_{kj}, Y_i, y_j)|z_{kj}, y_j\Big) - Eu_h(Z_{k1}, Z_{k2}, Y_1, Y_2)\Bigg)$$
$$= \frac{2}{n}\sum_{j=1}^{n}\Big(u_{h1}(z_{kj}, y_j) - Eu_h(Z_{k1}, Z_{k2}, Y_1, Y_2)\Big) = \frac{2}{n}\sum_{j=1}^{n}\tilde{u}_{h1}(z_{kj}, y_j)$$
$$= \frac{2}{n}\sum_{j=1}^{n}\Bigg(\frac{z_{kj}H(y_j)f(y_j) + H(y_j)g_k(y_j)}{2} - E\Big(H(Y)g_k(Y)\Big)\Bigg) + O_p(h^d).$$

We have obtained the simplified form of $\hat{U}_n$.

In the following three steps we will verify that U_n can be approximated by its projection $\hat{U}_n$ at a rate $1/\sqrt{nh}$, that is,

$$\sqrt{n}(\hat{U}_n - U_n) = O_p(1/\sqrt{nh}).$$

Step L1: $E\Big(u_h(Z_{k1}, Y_1, Z_{k2}, Y_2)\Big)^2 = O(1/h)$ where $u_h(\cdot)$ is defined in (A.10).

Clearly

$$E\Big(u_h(Z_{k1}, Y_1, Z_{k2}, Y_2)\Big)^2$$
$$\leqslant 2E\Bigg(Z_{k1}^2 H^2(Y_2)\frac{1}{h^2}K^2\Bigg(\frac{Y_2 - Y_1}{h}\Bigg)\Bigg) + 2E\Bigg(Z_{k2}^2 H^2(Y_1)\frac{1}{h^2}K^2\Bigg(\frac{Y_2 - Y_1}{h}\Bigg)\Bigg).$$

It is easy to see that

$$E\Bigg(Z_{k1}^2 H^2(Y_2)\frac{1}{h^2}K^2\Bigg(\frac{Y_2 - Y_1}{h}\Bigg)\Bigg) = E\Bigg(R_{kk}(Y_1)H^2(Y_2)\frac{1}{h^2}K^2\Bigg(\frac{Y_2 - Y_1}{h}\Bigg)\Bigg)$$
$$= \int R_{kk}(y_1)H^2(y_2)\frac{1}{h^2}K^2\Big(\frac{y_2 - y_1}{h}\Big)f(y_1)f(y_2)\,dy_1\,dy_2$$

$$
\begin{aligned}
&= \int G_{kk}(y_1) H^2(y_2) \frac{1}{h^2} K^2\Big(\frac{y_2 - y_1}{h}\Big) f(y_2)\, dy_1\, dy_2 \\
&= \int G_{kk}(y_2 - ht) H^2(y_2) \frac{1}{h} K^2(t) f(y_2)\, dt\, dy_2 \\
&= \int G_{kk}(y_2) H^2(y_2) \frac{1}{h} K^2(t) f(y_2)\, dt\, dy_2 + O(1) \\
&= \frac{R_2(K)}{h} E\Big(Z_k^2 H^2(Y) f(Y)\Big) + O(1) = O(1/h),
\end{aligned}
$$

and

$$
E\left(Z_{k2}^2 H^2(Y_1) \frac{1}{h^2} K^2\left(\frac{Y_2 - Y_1}{h}\right)\right) = \frac{R_2(K)}{h} E\Big(Z_k^2 H^2(Y) f(Y)\Big) + O(1) = O(1/h).
$$

This concludes the proof. □

Together with Step L1 and the computation of $E(U_n)$ right below (A.11), we have

$$
\xi_2 := var\Big(u_h(Z_{k1}, Y_1; Z_{k2}, Y_2)\Big) = O\Big(\frac{1}{h}\Big).
$$

Step L2: $U_n - \hat{U}_n$ is a U-statistic. It can be obtained as

$$
\begin{aligned}
U_n - \hat{U}_n &= U_n - \sum_{j_0=1}^{n} E(U_n | z_{kj_0}, y_{j_0}) + (n-1) E u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) \\
&= \frac{1}{n(n-1)} \sum_{i \neq j} u_h(z_{ki}, y_i, z_{kj}, y_j) \\
&\quad - \sum_{j_0=1}^{n} E\left(\frac{1}{n(n-1)} \sum_{i \neq j} u_h(z_{ki}, y_i, z_{kj}, y_j) | z_{kj_0}, y_{j_0}\right) \\
&\quad + (n-1) E u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) \\
&= \frac{1}{n(n-1)} \sum_{i \neq j} u_h(z_{ki}, y_i, z_{kj}, y_j) \\
&\quad - \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{j_0=1}^{n} E\Big(u_h(Z_{ki}, Y_i, Z_{kj}, Y_j) | z_{kj_0}, y_{j_0}\Big) \\
&\quad + (n-1) E u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) \\
&= \frac{1}{n(n-1)} \sum_{i \neq j} u_h(z_{ki}, y_i, z_{kj}, y_j) \\
&\quad - \frac{1}{n(n-1)} \sum_{i \neq j} \Bigg(E\Big(u_h(z_{ki}, y_i, Z_{kj}, Y_j) | z_{ki}, y_i\Big) \\
&\quad + E\Big(u_h(Z_{ki}, Y_i, z_{kj}, y_j) | z_{kj}, y_j\Big) + (n-2) E u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) \Bigg) \\
&\quad + (n-1) E u_h(Z_{ki}, Y_i, Z_{kj}, Y_j)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{n(n-1)} \sum_{i \neq j} u_h(z_{ki}, y_i, z_{kj}, y_j) - \frac{1}{n(n-1)} \sum_{i \neq j} \Big(u_{h1}(z_{ki}, y_i) \\
&\quad + u_{h1}(z_{kj}, y_j) \Big) + E u_h(Z_{ki}, Y_i, Z_{kj}, Y_j) \\
&= \frac{1}{n(n-1)} \sum_{i \neq j} \Big(u_h(z_{ki}, y_i, z_{kj}, y_j) - u_{h1}(z_{ki}, y_i) - u_{h1}(z_{kj}, y_j) \Big) \\
&\quad + E u_h(Z_{ki}, Y_i, Z_{kj}, Y_j) \\
&= \frac{1}{n(n-1)} \sum_{i \neq j} H(z_{ki}, y_i, z_{kj}, y_j) = \frac{1}{C_n^2} \sum_{i<j} H(z_{ki}, y_i, z_{kj}, y_j).
\end{aligned}
$$

Clearly, $H(\cdot)$ is a symmetric kernel of a U-Statistic. □

Step L3: $\sqrt{n}(U_n - \hat{U}_n) = O_p(\frac{1}{\sqrt{nh}})$. For this, we only need to compute the convergence rate of $E(U_n - \hat{U}_n)^2$.

We can easily obtain that $E\Big(H(Z_{k1}, Y_1, Z_{k2}, Y_2)\Big)=0$. Moreover, we can prove that $E(H|z_{kj}, y_j)=0$ for any j. Actually

$$
\begin{aligned}
H_1(z_{k1}, y_1) &=: E(H|z_{k1}, y_1) \\
&= E\Big(u_h(z_{k1}, y_1, Z_{k2}, Y_2)|z_{k1}, y_1]\Big) - E\Big(u_{h1}(z_{k1}, y_1)|z_{k1}, y_1\Big) \\
&\quad - E\Big(u_{h1}(Z_{k2}, Y_2)|z_{k1}, y_1\Big) + E\Big(u_h(Z_{k1}, Y_1, Z_{k2}, Y_2)\Big) \\
&= u_{h1}(z_{k1}, y_1) - u_{h1}(z_{k1}, y_1) - E\Big(u_{h1}(Z_{k2}, Y_2)|z_{k1}, y_1\Big) \\
&\quad + E u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) \\
&= -E u_{h1}(Z_{k2}, Y_2) + E u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) = 0.
\end{aligned}
$$

These imply that

$$
E(\sqrt{n}(U_n - \hat{U}_n))^2 = nVar(U_n - \hat{U}_n) = \frac{2E\Big(H(Z_{k1}, Y_1, Z_{k2}, Y_2)\Big)^2}{n-1}.
$$

The conclusion can be achieved if we can prove that $E\Big(H(Z_{k1}, Y_1, Z_{k2}, Y_2)\Big)^2 = O(\frac{1}{h})$. We can obtain this through the following calculation as:

$$
\begin{aligned}
&E\Big(H(Z_{k1}, Y_1, Z_{k2}, Y_2)\Big)^2 \\
&= E\Big(u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) - u_{h1}(Z_{k1}, Y_1) - u_{h1}(Z_{k2}, Y_2) + E u_h(Z_{k1}, Y_1, Z_{k2}, Y_2)\Big)^2 \\
&= E u_h^2(Z_{k1}, Y_1, Z_{k2}, Y_2) + E u_{h1}^2(Z_{k1}, Y_1) + E u_{h1}^2(Z_{k2}, Y_2) + E^2 u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) \\
&\quad - 2E\Big(u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) u_{h1}(Z_{k1}, Y_1)\Big) - 2E\Big(u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) u_{h1}(Z_{k2}, Y_2)\Big) \\
&\quad + 2E^2 u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) + 2E\Big(u_{h1}(Z_{k1}, Y_1) u_1(Z_{k2}, Y_2)\Big) \\
&\quad - 2E u_{h1}(Z_{k1}, Y_1) E u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) - 2E u_{h1}(Z_{k2}, Y_2) E u_h(Z_{k1}, Y_1, Z_{k2}, Y_2)
\end{aligned}
$$

$$= Eu_h^2(Z_{k1}, Y_1, Z_{k2}, Y_2)+Eu_{h1}^2(Z_{k1}, Y_1)+Eu_{h1}^2(Z_{k2}, Y_2)+E^2u_h(Z_{k1}, Y_1, Z_{k2}, Y_2)$$
$$-2Eu_{h1}^2(Z_{k1}, Y_1) - 2Eu_{h1}^2(Z_{k2}, Y_2) + 2E^2u_h(Z_{k1}, Y_1, Z_{k2}, Y_2)$$
$$+2E^2u_{h1}(Z_{k1}, Y_1) - 4Eu_{h1}(Z_{k1}, Y_1)Eu_h(Z_{k1}, Y_1, Z_{k2}, Y_2)$$
$$= Eu_h^2(Z_{k1}, Y_1, Z_{k2}, Y_2) + E^2u_h(Z_{k1}, Y_1, Z_{k2}, Y_2) - 2Eu_{h1}^2(Z_{k1}, Y_1) = O\Big(\frac{1}{h}\Big).$$

From the above results, we have that, together with (A.10),

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} H(y_j)\hat{g}_k(y_j)$$
$$= \sqrt{n}U_n + o_p(1) = \sqrt{n}\hat{U}_n + o_p(1)$$
$$= \frac{2}{\sqrt{n}}\sum_{j_0=1}^{n} E\Big(u_h(Z_{ki}, Z_{kj}, Y_i, Y_j)|z_{kj_0}, y_{j_0}\Big) - \sqrt{n}E\Big(u_h(Z_{ki}, Z_{kj}, Y_i, Y_j)\Big) + o_p(1)$$
$$= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(z_{kj}H(y_j)f(y_j) + H(y_j)g_k(y_j)\Big) - \sqrt{n}E(H(Y)g_k(Y)) + o_p(1).$$

Then

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(H(y_j)\hat{g}_k(y_j) - H(y_j)g_k(y_j)\Big)$$
$$= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(z_{kj}H(y_j)f(y_j) - EH(Y)g_k(Y)\Big) + o_p(1).$$

The proof is finished. □

Acknowledgment

The authors thank the associate editor and two referees for providing useful comments and suggestions which led to the great improvement of the presentation of the paper.

References

[1] E. Bura, R.D. Cook, Estimating the structural dimension of regressions via parametric inverse regression, J. Roy. Statist. Soc. B. 63 (2001) 393–410.
[2] R.J. Carroll, J. Fan, I. Gijbels, M.P. Wand, Generalized partially linear single-index models, J. Amer. Statist. Assoc. 92 (1997) 477–489.
[3] R.D. Cook, On the interpretation of regression plots, J. Amer. Statist. Assoc. 89 (1994) 177–189.
[4] R.D. Cook, Regression Graphics: Ideas for Studying Regressions through Graphics, Wiley, New York, 1998.
[5] R.D. Cook, SAVE: a method for dimension reduction and graphics in regression, Comm. Statist. Theory Methods 29 (2000) 2109–2121.
[6] R.D. Cook, F. Critchley, Identifying regression outliers and mixtures graphically, J. Amer. Statist. Assoc. 95 (2000) 781–794.
[7] R.D. Cook, S. Weisberg, Discussion to “Sliced inverse regression for dimension reduction”, J. Amer. Statist. Assoc. 86 (1991) 316–342.
[8] T. Fearn, A misuse of ridge regression in the calibration of near infrared reflectance instruments, Appl. Statist. 32 (1983) 73–79.
[9] L. Ferre, Determining the dimension in sliced inverse regression and related methods, J. Amer. Statist. Assoc. 93 (1998) 132–140.

[10] W. Härdle, E. Mammen, Comparing nonparametric versus parametric regression fits, Ann. Statist. 21 (1993) 1926–1947.
[11] T. Hsing, R.J. Carroll, An asymptotic theory for sliced inverse regression, Ann. Statist. 20 (1992) 1040–1061.
[12] K.C. Li, Sliced inverse regression for dimension reduction (with discussion), J. Amer. Statist. Assoc. 86 (1991) 316–342.
[13] B. Li, H. Zha, F. Chiaromonte, Contour regression: a general approach to dimension reduction, Ann. Statist. 33 (2005) 1580–1616.
[14] Y.X. Li, L.X. Zhu, When is sliced average variance estimation convergent? Technical Report, Department of Statistics and Actuarial Science, The University of Hong Kong. Hong Kong, 2004.
[15] J.R. Schott, Determining the dimensionality in sliced inverse regression, J. Amer. Statist. Assoc. 89 (1994) 141–148.
[16] G. Schwarz, Estimating the dimension of a model, Ann. Math. Statist. 30 (1978) 461–464.
[17] W. Stute, L.X. Zhu, Nonparametric checks for single-index models, Ann. Statist. 33 (2005) 1048–1083.
[18] S. Velilla, Assessing the number of linear components in a general regression problem, J. Amer. Statist. Assoc. 93 (1998) 1088–1098.
[19] Z. Ye, R. Weiss, Using the bootstrap to select one of a new class of dimension reduction methods, J. Amer. Statist. Assoc. 98 (2003) 968–979.
[20] L.C. Zhao, P.R. Krishnaiah, Z.D. Bai, On detection of the number of signals in presence of white noise, J. Multivariate Anal. 20 (1986) 1–25.
[21] L.C. Zhao, P.R. Krishnaiah, Z.D. Bai, On detection of the number of signals when the noise covariance matrix is arbitrary, J. Multivariate Anal. 20 (1986) 26–49.
[22] L.X. Zhu, Model checking of dimensional-reduction type for regression, Statist. Sin. 13 (2003) 283–296.
[23] L.X. Zhu, K.T. Fang, Asymptotics for kernel estimate of sliced inverse regression, Ann. Statist. 24 (1996) 1053–1068.
[24] L.X. Zhu, K.W. Ng, Asymptotics of sliced inverse regression, Statist. Sin. 5 (1995) 727–736.
[25] L.X. Zhu, B.Q. Miao, H. Peng, Sliced inverse regression with high-dimensional covariates, J. Amer. Statist. Assoc. 101 (2006) 630–643.
[26] L.X. Zhu, K.W. Ng, Checking the adequacy of a partial linear model, Statist. Sin. 13 (2003) 763–781.
[27] L.X. Zhu, M. Ohtaki, Y.X. Li, Hybrid methods of inverse regression based algorithms, Comp. Statist. Data. Anal. 51 (2007) 2621–2635.