ASYMPTOTICS FOR KERNEL ESTIMATE OF SLICED INVERSE REGRESSION[1]

By Li-Xing Zhu and Kai-Tai Fang

Chinese Academy of Sciences and Hong Kong Baptist University

To explore nonlinear structures hidden in high-dimensional data and to estimate the effective dimension reduction directions in multivariate nonparametric regression, Li and Duan proposed the sliced inverse regression (SIR) method which is simple to use. In this paper, the asymptotic properties of the kernel estimate of sliced inverse regression are investigated. It turns out that regardless of the kernel function, the asymptotic distribution remains the same for a wide range of smoothing parameters.

1. Introduction. In many statistical applications involving high-dimensional data two tasks will be faced:

1. exploring nonlinear structures such as clusters hidden in data;
2. overcoming the difficulty caused by the sparseness of the data.

Constrained by the three-dimensional physical world in which we live, the ability to look at more than three variables with human eyes is a real challenge. Yet, in statistical applications, there is a tendency to trust visual perception more than mathematical speculation. These issues have led to the development of dimension reduction techniques, for example, the projection pursuit (PP) method [cf. Huber (1985)] including PP regression [Friedman and Stuetzle (1981), Hall (1989) and Zhu and Fang (1992)] and PP clustering [Friedman and Tukey (1974)], the generalized additive model [Hastie and Tibshirani (1987)] and so on.

Alternatively, Li (1991) considers the regression model

$$(1.1) \qquad Y = F\big(\boldsymbol{\beta}_1^\mathsf{T}\mathbf{x}, \ldots, \boldsymbol{\beta}_k^\mathsf{T}\mathbf{x}, \varepsilon\big),$$

where $\mathbf{x}$ is a d-dimensional random vector, the $\boldsymbol{\beta}$'s are unknown vectors, the random variable ε is independent of $\mathbf{x}$ and F is an arbitrary unknown function on R^{k+1}. This model means that the projection of the d-dimensional explanatory $\mathbf{x}$ onto the k-dimensional subspace, $(\boldsymbol{\beta}_1^\mathsf{T}\mathbf{x}, \ldots, \beta_k^\mathsf{T}\mathbf{x})$, captures all we need to know about the response variable Y. When **k** is small, we may reach the goal of data reduction by estimating the $\boldsymbol{\beta}$'s efficiently.

A method, called sliced inverse regression (SIR), is used to search for the nonlinear structure in data and to estimate the projection directions $\boldsymbol{\beta}$'s. See

Duan and Li (1991), Hsing and Carroll (1992) and Li (1991) for results on SIR.

SIR can be used as a descriptive tool for finding interesting structures without relying on the regression model (1.1). Suppose that $\mathbf{X}_i^{\mathsf{T}} = (X_{i,1}, \ldots, X_{i,d+1})$, $i = 1, \ldots, n$, are i.i.d. $(d+1)$-dimensional observations from $\mathbf{X}^{\mathsf{T}} = (X_1, \ldots, X_{d+1})$ with density $f(\mathbf{X})$. To explore the structure in the observations, Li (1989) suggested the following method. First, we choose a component of $\mathbf{X}$, say X_m, as the dependent variable and consider the rest of $\mathbf{X}$, $\mathbf{X}^{(m)} = (X_1, \ldots, X_{m-1}, X_{m+1}, \ldots, X_{d+1})^{\mathsf{T}}$, as the regressor. Then we look for the interesting projection of $\mathbf{X}^{(m)}$, $\boldsymbol{\beta}_m^{\mathsf{T}} \mathbf{X}^{(m)}$, so that when it is plotted against X_m, we might find nonlinear structures in the data. For this purpose, Li suggested a projection index which is defined as the largest possible R-squared value for fitting a transformation of X_m linearly on $\boldsymbol{\beta}^{\mathsf{T}} \mathbf{X}^{(m)}$, that is, for any $m = 1, \ldots, d+1$,

$$B_m^2(\beta) = \max_{\mathbf{A} \in \mathscr{L}^2(\mathbf{X})} \left[\mathrm{Corr}\big(A(X_m), \boldsymbol{\beta}^{\mathsf{T}} \mathbf{X}^{(m)} \big) \right]^2,$$

where the notation "Corr" means the correlation coefficient of two random variables and $\mathscr{L}^2(\mathbf{X})$ is a class of square integrable functions with respect to the distribution of $\mathbf{X}$. It is easy to see [also see Li (1989), Theorem 2.1] that for a given $\boldsymbol{\beta}$, $A(X_m) = \mathbb{E}(\boldsymbol{\beta}^{\mathsf{T}} \mathbf{X}^{(m)} | X_m)$, and hence

$$(1.2) \qquad B_m^2(\beta) = \frac{\boldsymbol{\beta}^{\mathsf{T}} \mathrm{Cov}\big(\mathbb{E}(\mathbf{X}^{(m)} | X_m) \big) \boldsymbol{\beta}}{\boldsymbol{\beta}^{\mathsf{T}} \mathrm{Cov}(\mathbf{X}^{(m)}) \boldsymbol{\beta}},$$

where the notation "$\mathbb{E}$" means the expectation.

Hence a projection direction β_m can be found by maximizing $B_m^2(\beta)$ for a given m; that is, β_m is such that

$$(1.3) \qquad \begin{aligned} B_m^2(\beta_m) &= \sup_{\boldsymbol{\beta} \in R^d} \frac{\boldsymbol{\beta}^{\mathsf{T}} \mathrm{Cov}\big(\mathbb{E}(\mathbf{X}^{(m)} | X_m) \big) \beta}{\boldsymbol{\beta}^{\mathsf{T}} \mathrm{Cov}(\mathbf{X}^{(m)}) \beta} \\ &= \text{the largest eigenvalue of } \boldsymbol{\Lambda}, \end{aligned}$$

where

$$(1.4) \qquad \boldsymbol{\Lambda} = \big\{ \mathrm{Cov}(\mathbf{X}^{(m)}) \big\}^{-1/2} \mathrm{Cov}\big(\mathbb{E}(\mathbf{X}^{(m)} | X_m) \big) \big\{ \mathrm{Cov}(\mathbf{X}^{(m)}) \big\}^{-1/2},$$

assuming that $\mathrm{Cov}(\mathbf{X}^{(m)})$ is a positive-definite matrix. It is well known that $\mathrm{Cov}(\mathbf{X}^{(m)})$ can be estimated well by a sample covariance matrix. Hence how to estimate $\mathrm{Cov}(\mathbb{E}(\mathbf{X}^{(m)} | X_m))$ is the main issue.

The matrix $\mathrm{Cov}(\mathbb{E}(\mathbf{X}^{(m)} | X_m))$ can be estimated directly by some nonparametric method such as the **k**-slice estimate proposed by Li (1991). This estimate is based on the following procedure. Divide the range of X_m into n/k (rounded to the nearest integer) slices, $S_1, \ldots, S_{n/k}$, where each slice contains k sample points of X_m. Within each slice, compute the sample covariance matrix of concomitants $\mathbf{X}^{(m)}$ associated with X_m's, denoted by $\mathbf{C}_i$, $i = 1, \ldots, n/k$. Finally, a weighted sum of $\mathbf{C}_i$'s is used as an estimate of $\boldsymbol{\Lambda}$. One can refer to Section 4 of Li (1991) for more details. Hsing and Carroll

(1992) presented the asymptotic normality of Li's two-slice estimation. The issue on appropriate choice of **k** remains debatable; see the discussions in Li (1991).

In this paper we consider the kernel method for estimating $\mathrm{Cov}(\mathbb{E}(\mathbf{X}^{(m)}|X_m))$ and its eigenvalues and eigenvectors, then $\mathbf{\Lambda}$ and the associated eigenvalues and eigenvectors. In our results root n consistency and asymptotic normality are established. It turns out that the asymptotic variance of our kernel estimate does not depend on the kernel function used. It remains the same over a wide range of bandwidth. This theoretical result supports the empirical finding that SIR is not sensitive to the number of slices used; see Li (1991).

In the next section we shall state the main results of the paper and leave the proofs to Section 4. Some lemmas which are needed in the proofs shall be given in Section 3.

2. Main results. For simplicity of notation, we write Y and $\mathbf{Z}$ for X_m and $\mathbf{X}^{(m)}$, respectively, for given m and let $f(Y)$ denote the density function of Y. Put $\mathbf{Z}$ and its independent copies $\mathbf{Z}_j$ as

$$\mathbf{Z} = (Z_1, \ldots, Z_d)^{\mathsf{T}}, \qquad \mathbf{Z}_j = (Z_{1j}, \ldots, Z_{dj})^{\mathsf{T}}, \qquad j = 1, \ldots, n.$$

Our objective is to estimate, based on $(\mathbf{Z}_j, Y_j)$'s, $\mathrm{Cov}(\mathbb{E}(\mathbf{Z}|Y))$, its eigenvalues and the corresponding eigenvectors. Define

$$\mathbf{R}(Y) \triangleq (R_1(Y), \ldots, R_d(Y))^{\mathsf{T}} = (\mathbb{E}(Z_1|Y), \ldots, \mathbb{E}(Z_d|Y))^{\mathsf{T}} \triangleq \mathbb{E}(\mathbf{Z}|Y),$$

$$\mathbf{g}(Y) = (g_1(Y), \ldots, g_d(Y))^{\mathsf{T}} = (R_1(Y)f(Y), \ldots, R_d(Y)f(Y))^{\mathsf{T}}$$

and

$$\mathbf{\Lambda}_1 \triangleq \mathrm{Cov}(\mathbf{R}(Y)) = \mathrm{Cov}(\mathbb{E}(\mathbf{Z}|Y)).$$

We may estimate $\mathbf{R}(Y)$ by a kernel method

$$\hat{\mathbf{g}}_i(Y) = \frac{1}{nh} \sum_{j=1}^{n} Z_{ij} K\big((Y - Y_j)h^{-1}\big),$$

$$\hat{\mathbf{g}}(Y) = (\hat{g}_1(Y), \ldots, \hat{g}_d(Y))^{\mathsf{T}},$$

$$\hat{f}(Y) = \frac{1}{nh} \sum_{j=1}^{n} K\big((Y - Y_j)h^{-1}\big)$$

and

$$\hat{\mathbf{R}}(Y) = \hat{\mathbf{g}}(Y)/\hat{f}(Y),$$

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

However, to avoid technical difficulties due to small values in the denominator, we have to use a modified version of the kernel estimate. For each fixed $b > 0$, let

$$\hat{f}_b(Y) = \max\{\hat{f}(Y), b\}, \qquad f_b(Y) = \max\{f(Y), b\}$$

and

$$\mathbf{R}_b(Y) = \frac{\mathbf{R}(Y)f(Y)}{f_b(Y)} = (R_{b1}(Y), \ldots, R_{bd}(Y))^{\mathsf{T}}.$$

We shall use $\hat{\mathbf{R}}_b(Y) = \hat{\mathbf{g}}(Y)/\hat{f}_b(Y)$ instead of $\hat{\mathbf{R}}(Y)$. Without loss of generality, let $E\mathbf{Z} = \mathbf{0}$. Then an estimate of $\mathbf{\Lambda}_1$ can be constructed as follows:

$$\mathbf{\Lambda}_{1n} = \frac{1}{n} \sum_{j=1}^{n} \left(\hat{\mathbf{R}}_b(Y_j)\right)\left(\hat{\mathbf{R}}_b(Y_j)\right)^{\mathsf{T}}.$$

To present our main theorem, we need some notation. For a symmetric $(d \times d)$ matrix $\mathbf{c} = (c_{il})_{d \times d}$, let $\mathrm{Vech}(\mathbf{c}) = (c_{11}, \ldots, c_{d1}, c_{22}, \ldots, c_{d2}, c_{33}, \ldots, c_{dd})^{\mathsf{T}}$ be a $([d(d+1)]/2)$-dimensional vector. Define

$$V(\mathbf{Z}, Y) = \left(\tfrac{1}{2}Z_i R_l(Y) + \tfrac{1}{2}Z_l R_i(Y)\right)_{d \times d},$$

$$\sigma_\lambda^2 = \mathbf{\lambda}^{\mathsf{T}} \mathrm{Cov}\left(\mathrm{Vech}(V(\mathbf{Z}, Y))\right)\lambda \quad \text{for } \mathbf{\lambda} \in R^{d(d+1)/2}.$$

We now present the main results and put the conditions next to them.

THEOREM 2.1. *Under conditions* 1 *to* 6 *given in Remark* 2.2, *as* $n \to \infty$, *we have*

$$(2.1) \qquad \sqrt{n}\,(\mathbf{\Lambda}_{1n} - \mathbf{\Lambda}_1) \Rightarrow \mathbf{H},$$

where $\mathbf{\lambda}^{\mathsf{T}} \mathrm{Vech}(\mathbf{H})$ *has a normal distribution* $N(0, \sigma_\lambda^2)$ *for any* $\mathbf{\lambda} \neq \mathbf{0}$ *and the notation* "$\Rightarrow$" *stands for the convergence in distribution.*

From Theorem 2.1 we can derive the $\sqrt{n}$-consistency of eigenvalues and of corresponding eigenvectors by using perturbation theory. The result is stated in the following theorem. First introduce some notation.

Let $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_d(\mathbf{A}) \geq 0$ and $\mathbf{b}_i(\mathbf{A}) = (b_{1i}(\mathbf{A}), \ldots, b_{di}(\mathbf{A}))^{\mathsf{T}}$, $i = 1, \ldots, d$, denote, respectively, the eigenvalues and the corresponding eigenvectors of a $d \times d$ matrix A.

THEOREM 2.2. *Suppose that nonzero* $\lambda_i(\mathbf{\Lambda}_1)$'s *are distinct and that conditions* 1 *to* 6 *given in Remark* 2.2 *hold. Then for each nonzero eigenvalue* $\lambda_i(\mathbf{\Lambda}_1)$ *and the corresponding eigenvector* $\mathbf{b}_i(\mathbf{\Lambda}_1)$, *we have*

$$\sqrt{n}\,(\lambda_i(\mathbf{\Lambda}_{1n}) - \lambda_i(\mathbf{\Lambda}_1))$$

$$(2.2) \qquad = \sqrt{n}\,\mathbf{b}_i(\mathbf{\Lambda}_1)^{\mathsf{T}}(\mathbf{\Lambda}_{1n} - \mathbf{\Lambda}_1)\mathbf{b}_i(\mathbf{\Lambda}_1) + o_p(\sqrt{n}\,\|\mathbf{\Lambda}_{1n} - \mathbf{\Lambda}_1\|)$$

$$\Rightarrow \mathbf{b}_i(\mathbf{\Lambda}_1)^{\mathsf{T}}\mathbf{H}\mathbf{b}_i(\mathbf{\Lambda}_1),$$

where $\mathbf{H}$ *is given in Theorem* 2.1, *and*

$$
\sqrt{n}\,\big(\mathbf{b}_i(\boldsymbol{\Lambda}_{1n}) - \mathbf{b}_i(\boldsymbol{\Lambda}_1)\big)
$$

(2.3)
$$
= \sqrt{n}\,\sum_{\substack{l=1\\ l\neq i}}^{d} \frac{\mathbf{b}_l(\boldsymbol{\Lambda}_1)\mathbf{b}_i(\boldsymbol{\Lambda}_1)^{\mathsf{T}}(\boldsymbol{\Lambda}_{1n} - \boldsymbol{\Lambda}_1)\mathbf{b}_i(\boldsymbol{\Lambda}_1)}{\big(\lambda_j(\boldsymbol{\Lambda}_1) - \lambda_l(\boldsymbol{\Lambda}_1)\big)} + o_p\big(\sqrt{n}\,\|\boldsymbol{\Lambda}_{1n} - \boldsymbol{\Lambda}_1\|\big)
$$

$$
\Rightarrow \sum_{\substack{l=1\\ l\neq i}}^{d} \frac{\mathbf{b}_l(\boldsymbol{\Lambda}_1)\mathbf{b}_i(\boldsymbol{\Lambda}_1)^{\mathsf{T}}\mathbf{H}\mathbf{b}_i(\boldsymbol{\Lambda}_1)}{\big(\lambda_j(\boldsymbol{\Lambda}_{1n}) - \lambda_l(\boldsymbol{\Lambda}_1)\big)}
$$

as $n \to \infty$, *where* $\|\boldsymbol{\Lambda}_{1n}\| = \sum_{1\leq i,\,j\leq d}|a_{ij}|$ *is the* l_1-*norm of* $\boldsymbol{\Lambda}_1$.

REMARK 2.1. Based on Theorem 2.1, it is not hard to obtain the asymptotic distribution for the estimate of $\boldsymbol{\Lambda}$ given in (1.4) and the associated eigenvalues and eigenvectors of $\boldsymbol{\Lambda}$, where the estimate of $\boldsymbol{\Lambda}$ is $\widehat{\mathrm{Cov}(\mathbf{Z})}^{-1/2}\boldsymbol{\Lambda}_{1n} \times \widehat{\mathrm{Cov}(\mathbf{Z})}^{-1/2}$ and $\widehat{\mathrm{Cov}(\mathbf{Z})}$ is the sample covariance matrix of $\mathbf{Z}$. The perturbation formula as given by (3.29) or (3.36) in Kato (1983), pages 92 and 93, together with the standard delta argument should suffice; see also Tyler (1981) and Li (1991). The derivation can be simplified by noting the affine invariance property of our problem. One needs only to work out the special case that $\mathrm{Cov}(\mathbf{Z}) = \mathbf{I}$. In this case $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_1 = \mathrm{Cov}(R(Y))$. We omit the details.

REMARK 2.2. The following six conditions are required for Theorems 2.1 and 2.2.

1. All $g_i(y) = R_i(y)f(y)$, $i = 1,\ldots,d$, and $f(y)$ are 3-times differentiable and their third derivatives satisfy the following condition: there exists a neighborhood of the origin, say U, and a constant $c > 0$ such that, for any $u \in U$,

$$
\big|f^{(3)}(y + u) - f^{(3)}(y)\big| \leq c|u|,
$$

$$
\big|g_i^{(3)}(y + u) - g_i^{(3)}(y)\big| \leq c|u|, \qquad i = 1,\ldots,d.
$$

2. For each pair $1 \leq i, l \leq d$ and for any $u \in U$,

$$
\big|R_i(y + u)R_l(y + u) - R_i(y)R_l(y)\big| \leq c|u|.
$$

3. $\mathbb{E}|Z_i|^4 < \infty$, $i = 1,\ldots,d$.
4. The continuous kernel function $K(\cdot)$ has the following properties:
   (a) the support of $K(\cdot)$ is the interval $[-1, 1]$;
   (b) $K(\cdot)$ is symmetric about 0;
   (c) $\int_{-1}^{1}K(u)\,du = 1$, and $\int_{-1}^{1}u^i K(u)\,du = 0$, $i = 1, 2, 3$.
5. As $n \to \infty$, $h \sim n^{-c_1}$, $b \sim n^{-c_2}$ with the positive numbers c_1 and c_2 satisfying that $1/8 + c_2/4 < c_1 < 1/4 - c_2$, and the notation "$\sim$" means that two quantities have the same convergence order.
6. $\sqrt{n}\,\mathbb{E}[R_i(Y)R_l(Y)I(f(Y) \leq b)] = o(1)$ as $n \to \infty$, for $1 \leq i, l \leq d$, where $I(\cdot)$ is the indicator function and b satisfies condition 5.

Conditions 1 and 2 are concerned with the smoothness of the density function of Y and regression curve $\mathbf{R}(Y)$. Condition 3 is necessary for the asymptotic normality of Λ_{1n}.

3. Some lemmas. To prove succinctly the theorems, we put some lemmas in this section. Without any confusion, we shall denote a constant by c which may take different values for each appearance (independent of n) throughout this section.

LEMMA 3.1 [cf. Rao (1983), Theorem 2.1.8]. *Suppose that conditions* 1, 2, 4 *and* 5 *hold. Then, as* $n \to \infty$,

$$(3.1) \qquad \sup_{y} |\hat{f}(y) - f(y)| = O(h^4 + n^{-1/2}h^{-1}\log n) \quad a.s.$$

The following lemma is a slightly modified version of Theorem 2.37 of Pollard (1984) [or see Zhu (1993), Theorem 3.1].

LEMMA 3.2. *Suppose that conditions* 1 *to* 6 *are fulfilled. Then, for any fixed* $\varepsilon > 0$,

$$P\left\{ \sup_{y} |\hat{g}_i(y) - \mathbb{E}\hat{g}_i(y)| > 8n^{-1/2}h^{-1}\varepsilon \right\}$$

$$(3.2) \qquad \leq 2c\left(\frac{\varepsilon}{\sqrt{n}\,d} \right)^{-4} \exp\left\{ -\frac{1}{2}\varepsilon^2 \big/ \big(32\,d(\log n)^{1/2} \big) \right\}$$

$$+ 8cd^{-8}\exp(-nd^2)$$

$$+ E(Z_i)^4 I\big(|Z_i| > cd^{-1/2}(\log n)^{1/4} \big),$$

where

$$d \geq \sup_{y} \left\{ \mathrm{Var}\left(K\left(\frac{y-Y}{h} \right) \right) \right\}^{1/2}.$$

By this lemma the uniform convergence rate of $\hat{g}_i(y)$ can be achieved.

LEMMA 3.3. *Suppose that the conditions in Lemma* 3.1 *are satisfied. Then*

$$(3.3) \qquad \sup_{y} |\hat{g}_i(y) - g_i(y)| = O_p(h^4 + n^{-1/2}h^{-1}\log n).$$

PROOF. Since the kernel function K is uniformly continuous on $[-1, 1]$, writing $c_1 = \sup_{-1 \leq u \leq 1}|K(u)|$, we have

$$\sup_{y} \left(\mathrm{Var}\left(K\left(\frac{y-Y}{h} \right) \right) \right)^{1/2} \leq \sup_{y} \left(\int K^2\left(\frac{y-Y}{h} \right) f(Y)\,dY \right)^{1/2} \leq c_1.$$

Choose $\varepsilon = \log n$, then as $n \to \infty$, we have

$$(3.4) \qquad \sup_y |\hat{g}_i(y) - \mathbb{E}\hat{g}_i(y)| = O_p(n^{-1/2}h^{-1}\log n).$$

On the other hand, expanding $g_i(y)$ in a Taylor series with the Lagrange form of the remainder term [or see Rao (1983), page 47] and using condition 1, we can see that

$$\sup_y |\mathbb{E}\hat{g}_i(y) - g_i(y)|$$

$$= \sup_y \left| \int \frac{1}{h} K\left(\frac{y-Y}{h}\right) \{R_i(Y)f(Y) - R_i(y)f(y)\}\, dY \right|$$

$$= \sup_y \left| \int K(u)\{R_i(y-uh)f(y-uh) - R_i(y)f(y)\}\, du \right|$$

$$= \sup_y \left| \int K(u)(\tilde{g}_i(z-u) - \tilde{g}_i(y))(uh)^3\, du \right| \Big/ 6$$

$$\le c \int |K(u)u^4|\, du \cdot h^4,$$

completing the proof. □

4. Proof of Theorem 2.1. Clearly, we need to deal with only the ilth element

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\hat{g}_i(Y_j)\hat{g}_l(Y_j)\right)/\hat{f}_b^2(Y_j) \text{ of the matrix } \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\hat{\mathbf{R}}_b(Y)\right)\left(\hat{\mathbf{R}}_b(Y)\right)^{\mathsf{T}}.$$

The proof will be done in several steps. Define

$$R_{b,i}(y) = \frac{g_i(y)}{f_b(y)},$$

$$I_{i,l}^{(1)}(y) = \frac{g_i(y)g_l(y)}{f_b^2(y)} = R_{bi}(y)R_{bl}(y),$$

$$I_{i,l}^{(2)}(y) = \frac{g_i(y)\hat{g}_l(y) + g_l(y)\hat{g}_i(y)}{f_b^2(y)} = \frac{R_{bi}(y)\hat{g}_l(y)}{f_b(y)} + \frac{R_{bl}(y)\hat{g}_i(y)}{f_b(y)},$$

$$I_{i,l}^{(3)}(y) = 2R_{bi}(y)R_{bl}(y)\frac{\hat{f}_b(y)}{f_b(y)}.$$

Step 1. Show that

$$(4.1) \qquad \begin{aligned} &\frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\frac{\hat{g}_i(Y_j)\hat{g}_l(Y_j)}{\hat{f}_b^2(Y_j)}\right) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{I_{il}^{(1)}(Y_j) + I_{il}^{(2)}(Y_j) - I_{il}^{(3)}(Y_j)\right\} + o_p(1). \end{aligned}$$

It is easy to see that

$$
\text{(4.2)} \quad \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\frac{\hat{g}_i(Y_j)\hat{g}_l(Y_j)}{\hat{f}_b^2(Y_j)} \right) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ I_{il}^{(1)}(Y_j) + I_{il}^{(2)}(Y_j) - I_{il}^{(3)}(Y_j) \right\}
$$

$$
- J_{n1} + J_{n2} + J_{n3} - J_{n4},
$$

where

$$
J_{n1} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ g_i(Y_j)\big(\hat{g}_l(Y_j) - g_l(Y_j)\big) + g_l(Y_j)\big(\hat{g}_i(Y_j) - g_i(Y_j)\big) \right\}
$$

$$
\times \frac{\hat{f}_b^2(Y_j) - f_b^2(Y_j)}{\hat{f}_b^2(Y_j)f_b^2(Y_j)},
$$

$$
J_{n2} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\big(\hat{g}_i(Y_j) - g_i(Y_j)\big)\big(\hat{g}_l(Y_j) - g_l(Y_j)\big)}{\hat{f}_b^2(Y_j)},
$$

$$
J_{n3} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} R_{bi}(Y_j) R_{bl}(Y_j) \frac{\big(\hat{f}_b^2(Y_j) - f_b^2(Y_j)\big)^2}{f_b^2(Y_j)\hat{f}_b^2(Y_j)}
$$

and

$$
J_{n4} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \big(\hat{f}_b(Y_j) - f_b(Y_j)\big)^2 \frac{R_{bi}(Y_j) R_{bl}(Y_j)}{f_b^2(Y_j)}.
$$

Hence it suffices to show that, for obtaining (4.1),

$$
\text{(4.3)} \qquad\qquad |J_{ni}| = o_p(1), \qquad i = 1, \dots, 4.
$$

Since

$$
\text{(4.4)} \quad \begin{aligned} &\big(\hat{f}_b^2(Y) - f_b^2(Y)\big) = \big(\hat{f}_b(Y) - f_b(Y)\big)^2 + 2f_b(Y)\big(\hat{f}_b(Y) - f_b(Y)\big), \\ &\big|\hat{f}_b(y) - f_b(y)\big| \le \big|\hat{f}(y) - f(y)\big| \quad \text{and} \quad |R_{bi}(y)| \le |R_i(y)|, \end{aligned}
$$

then, by Lemmas 3.1 and 3.3 and condition 5, we have

$$
|J_{n1}| = O_p\Bigg(\Big\{ \big(h^4 + n^{-1/2}h^{-1}\log n\big)^3 b^{-3}n^{1/2}
$$

$$
+ \big(h^4 + n^{-1/2}h^{-1}\log n\big)^2 b^{-2}n^{1/2} \Big\}
$$

$$
\text{(4.5)} \qquad\qquad \times \left\{ \frac{1}{n} \sum_{j=1}^{n} \big(|R_i(Y_j)| + |R_l(Y_j)|\big) \right\} \Bigg)
$$

$$
= O_p\big(n^{-8c_1 + 2c_2 + 1/2} + n^{2c_1 + 2c_2 - 1/2}(\log n)^2 \big) = o_p(1),
$$

where the second equation is due to the law of large numbers for $n^{-1}\sum_{j=1}^{n}|R_i(Y_j)|$.

Similarly, we can prove (4.3) for J_{ni}, $i = 2, 3, 4$.

Step 2. Show that

$$
\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ I_{il}^{(1)}(Y_j) - \mathbb{E}\left[I_{il}^{(1)}(Y) \right] \right\}
$$

(4.6)

$$
= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ R_i(Y_j) R_l(Y_j) - \mathbb{E}\left[R_i(Y) R_l(Y) \right] \right\} + o_p(1).
$$

All we need to do is to show that

$$
\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \left(R_i(Y_j) R_l(Y_j) - R_{bi}(Y_j) R_{bl}(Y_j) \right) \right.
$$

(4.7)

$$
\left. - \mathbb{E}\left(R_i(Y) R_l(Y) - R_{bi}(Y) R_{bl}(Y) \right) \right\} = o_p(1).
$$

By the independence of Y_j's and $b \to 0$, we have

$$
\mathbb{E}\left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \left(R_j(Y_j) R_l(Y_j) - R_{bi}(Y_j) R_{bl}(Y_j) \right) \right.\right.
$$

$$
\left.\left. - \mathbb{E}\left(R_i(Y) R_l(Y) - R_{bi}(Y) R_{bl}(Y) \right) \right\}\right\}^2
$$

$$
= \operatorname{Var}\left(R_i(Y) R_l(Y) - R_{bi}(Y) R_{bl}(Y) \right)
$$

$$
\leq \mathbb{E}\left(R_i(Y) R_l(Y) - R_{bi}(Y) R_{bl}(Y) \right)^2
$$

$$
= \int_{\{f(Y)<b\}} \left(R_i(Y) R_l(Y) - R_{bi}(Y) R_{bl}(Y) \right)^2 f(Y) \, dY
$$

$$
\leq 2 \int_{\{f(Y)<b\}} R_i^2(Y) R_l^2(Y) f(Y) \, dY \to 0 \quad \text{as } b \to 0,
$$

which implies (4.7), and (4.6) is proved.

Step 3. Show that

$$
\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ I_{il}^{(2)}(Y_j) - \mathbb{E}\left[I_{il}^{(2)}(Y) \right] \right\}
$$

$$
= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ R_{bi}(Y_j) R_{bl}(Y_j) + \frac{1}{2} Z_{lj} R_{bi}(Y_j) \frac{f(Y_j)}{f_b(Y_j)} \right.
$$

(4.8)

$$
\left. + \frac{1}{2} Z_{ij} R_{bl}(Y_j) \frac{f(Y_j)}{f_b(Y_j)} \right\}
$$

$$
- \mathbb{E}\left\{ R_{bi}(Y) R_{bl}(Y) + \frac{1}{2} Z_l R_{bi}(Y) \frac{f(Y)}{f_b(Y)} + \frac{1}{2} Z_i R_{bl}(Y) \frac{f(Y)}{f_b(Y)} \right\}
$$

$$
+ o_p(1).
$$

To analyze $n^{-1/2}\sum_{j=1}^n I_{il}^{(2)}(Y_j)$, we need only to work on

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n R_{bi}(Y)\frac{\hat{g}_l(Y_j)}{f_b(Y_j)}.$$

Approximate this sum by a U-statistic with a varying kernel. Applying the similar argument used by Stone (1984), Nolan and Pollard (1987), Härdle and Stoker (1989), and Powell, Stock and Stoker (1989), we have

(4.9)
$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{j=1}^n & \left\{R_{bi}(Y_j)\frac{\hat{g}_l(Y_j)}{f_b(Y_j)} - \mathbb{E}\left[R_{bi}(Y_j)\frac{\hat{g}_l(Y_j)}{f_b(Y_j)}\right]\right\}\\
& = \frac{1}{\sqrt{n}}\sum_{j=1}^n u_{hbi}(w_{lj}) - \mathbb{E}[u_{hbi}(w_l)] + o_p(1),
\end{aligned}$$

where

$$u_{hbi}(w_{lj}) = \frac{1}{2h}\int K\big((Y_j - y)h^{-1}\big)\left\{\frac{y_l R_{bi}(Y_j)}{f_b(Y_j)} + \frac{Z_{lj}R_{bi}(y)}{f_b(y)}\right\}f_1(z_l, y)\,dz_l\,dy$$

and $f_1(z_l, y)$ is the density function of $w_l = (Z_l, Y)$. Let $f(y)/f_b(y) \doteq \varepsilon_b(y)$; we have

$$\begin{aligned}
u_{hbi}(w_{lj}) &= \frac{1}{2h}\int K\big((Y_j - y)h^{-1}\big)\left\{\frac{y_l R_{bi}(Y_j)}{f_b(Y_j)} + \frac{Z_{lj}R_{bi}(y)}{f_b(y)}\right\}f_1(z_l, y)\,dz_l\,dy\\[4pt]
&= \frac{R_{bi}(Y_j)}{2f_b(Y_j)}\int \frac{1}{h}K\big((Y_j - y)h^{-1}\big)R_l(y)f(y)\,dy\\[4pt]
&\quad + \frac{1}{2}Z_{lj}\int \frac{1}{h}K\big((Y_j - y)h^{-1}\big)\frac{R_{bi}(y)f(y)}{f_b(y)}\,dy\\[4pt]
\text{(4.10)}\quad &= \frac{R_{bi}(Y_j)R_l(Y_j)f(Y_j)}{2f_b(Y_j)} + \frac{Z_{lj}R_{bi}(Y_j)f(Y_j)}{2f_b(Y_j)}\\[4pt]
&\quad + \frac{R_{bi}(Y_j)}{2f_b(Y_j)}\int \frac{1}{h}K\big((Y_j - y)h^{-1}\big)\big\{R_l(y)f(y) - R_l(Y_j)f(Y_j)\big\}\,dy\\[4pt]
&\quad + \frac{1}{2}Z_{lj}\int \frac{1}{h}K\big((Y_j - y)h^{-1}\big)\left\{\frac{R_{bi}(y)f(y)}{f_b(y)} - \frac{R_{bi}(Y_j)f(Y_j)}{f_b(Y_j)}\right\}dy\\[4pt]
&\doteq \frac{1}{2}R_i(Y_j)R_l(Y_j)\varepsilon_b^2(Y_j) + \frac{1}{2}Z_{lj}R_i(Y_j)\varepsilon_b^2(Y_j)\\[4pt]
&\quad + J_{nil}^{(5)}(Y_j) + J_{nil}^{(6)}(Y_j).
\end{aligned}$$

In order to show (4.8), it is sufficient to show that

$$\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} J_{nil}^{(5)}(Y_j)\right) = \mathrm{Var}(J_{nil}^{(5)}(Y)) = o(1)$$

and

$$\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} J_{nil}^{(6)}(Y_j)\right) = \mathrm{Var}(J_{nil}^{(6)}(Y)) = o(1).$$

By conditions 1 and 5, there exists a constant $c > 0$ such that

$$\mathrm{Var}(J_{nil}^{(5)}(Y)) \le \mathbb{E}(J_{nil}^{(5)}(Y))^2$$

(4.11)
$$\le c \int \left(\frac{R_{bi}(y)}{f_b(y)}\right)^2 \left(\int |K(u)u|\,du\right)^2 f(y)\,dy \cdot h^4$$

$$\le ch^4 b^{-2}\,\mathbb{E}R_i^2(Y) = O(n^{-4c_1+2c_2}) = o(1).$$

Moreover,

$$\frac{R_{bi}(Y_1 - uh)f(Y_1 - uh)}{f_b(Y_1 - uh)} - \frac{R_{bi}(Y_1)f(Y_1)}{f_b(Y_1)} \to 0$$

as $h \to 0$ and $b \to 0$. Applying the dominated convergence theorem, we have

$$\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} J_{nil}^{(6)}(Y_j)\right) = \mathrm{Var}(J_{nil}^{(6)}(Y_1)) \le \mathbb{E}(J_{nil}^{(6)}(Y_1))^2$$

(4.12)
$$= \frac{1}{4}\mathbb{E}\left\{(Z_i^2|Y_1)\left\{\int K(u)\left(\frac{R_{bi}(Y_1 - uh)f(Y_1 - uh)}{f_b(Y_1 - uh)}\right.\right.\right.$$
$$\left.\left.\left. - \frac{R_{bi}(Y_1)f(Y_1)}{f_b(Y_1)}\right)du\right\}^2\right\}$$

$$= o(1),$$

and completing the proof of Step 3 from (4.9) to (4.12).

Step 4. Prove the following three equations:

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{I_{il}^{(3)}(Y_j) - \mathbb{E}[I_{il}^{(3)}(Y_j)]\right\}$$

(4.13)
$$= \frac{2}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{R_{bi}(Y_j)R_{bl}(Y_j)f(Y_j)}{f_b(Y_j)}\right.$$
$$\left. - \mathbb{E}\left[\frac{R_{bi}(Y)R_{bl}(Y)f(Y)}{f_b(Y)}\right]\right\} + o_p(1),$$

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ R_{bi}(Y_j) R_{bl}(Y_j) + \frac{Z_{lj} R_{bi}(Y_j) f(Y_j)}{2 f_b(Y_j)} + \frac{Z_{ij} R_{bl}(Y_j) f(Y_j)}{2 f_b(Y_j)} \right.$$

$$- \mathbb{E}\left[R_{bi}(Y) R_{bl}(Y) + \frac{R_l(Y) R_{bi}(Y) f(Y)}{2 f_b(Y)} \right.$$

(4.14)

$$\left. \left. + \frac{R_i(Y) R_{bl}(Y) f(Y)}{2 f_b(Y)} \right] \right\}$$

$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ R_i(Y_j) R_l(Y_j) + \frac{1}{2} Z_{lj} R_i(Y_j) + \frac{1}{2} Z_{ij} R_l(Y_j) \right.$$

$$\left. - 2 E[R_i(Y) R_l(Y)] \right\} + o_p(1)$$

and

(4.15)

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \frac{R_{bi}(Y_j) R_{bl}(Y_j) f(Y_j)}{f_b(Y_j)} - \mathbb{E}\left[\frac{R_{bi}(Y) R_{bl}(Y) f(Y)}{f_b(Y)} \right] \right\}$$

$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ R_i(Y_j) R_l(Y_j) - \mathbb{E}[R_i(Y) R_l(Y)] \right\} + o_p(1).$$

Since the arguments for proving (4.13) to (4.15) are essentially the same as those used in the proofs of Steps 2 and 3, we omit the details here.

Up to now, we have showed that, combining Steps 1 to 4,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\hat{g}_i(Y_j) \hat{g}_l(Y_j) / \hat{f}_b^2(Y_j) - \mathbb{E}[\hat{v}_{bil}(Y)] \right)$$

$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \frac{1}{2} (Z_l R_i(Y) + Z_i R_l(Y)) - \mathbb{E}(R_i(Y) R_l(Y)^{\top}) \right\} + o_p(1),$$

where

$$\hat{v}_{bil}(Y) = R_{bi}(Y) R_{bl}(Y) + \frac{R_{bi}(Y) \hat{g}_l(Y)}{f_b(Y)} + \frac{R_{bl}(Y) \hat{g}_i(Y)}{f_b(Y)}$$

$$- 2 \frac{R_{bi}(Y) R_{bl}(Y) \hat{f}_b(Y)}{f_b(Y)}, \qquad 1 \le i, l \le d.$$

Consequently, the last step is to prove that

(4.16) $$\sqrt{n} \, \mathbb{E}[\hat{v}_{bil}(Y)] = \sqrt{n} \, \mathbb{E}[R_i(Y) R_l(Y)] + o(1).$$

It is easy to see that

$$0 \le 1 - \varepsilon_b^2(Y) = \left(1 - \frac{f^2(Y)}{b^2} \right) I(f(Y) < b) \le I(f(Y) < b).$$

Furthermore, by condition 6, we have

$$\sqrt{n}\left|\mathbb{E}[R_{bi}(Y)R_{bl}(Y)] - \mathbb{E}[R_i(Y)R_l(Y)]\right|$$

(4.17)
$$= \sqrt{n}\left|\mathbb{E}\left[R_i(Y)R_l(Y)\left(1 - \frac{f^2(Y)}{f_b(Y)}\right)\right]\right|$$

$$\leq \sqrt{n}\,\mathbb{E}\left|R_i(Y)R_l(Y)I(f(Y) < b)\right| = o(1).$$

Invoking condition 6 again, we obtain

$$\sqrt{n}\,\mathbb{E}\left[\frac{R_{bi}(Y)\hat{g}_l(Y)}{f_b(Y)}\right]$$

$$= \sqrt{n}\int \frac{1}{h}K((Y - Y_1)h^{-1})\frac{Z_l R_{bi}(Y)f_1(Y_1, Z_l)f(Y)}{f_b(Y)}\,dZ_l\,dY_1\,dY$$

$$= \sqrt{n}\int \frac{1}{h}K((Y - Y_1)h^{-1})\frac{R_l(Y_1)R_{bi}(Y)f(Y_1)f(Y)}{f_b(Y)}\,dY_1\,dY$$

$$= \sqrt{n}\int K(u)\frac{R_l(Y - uh)f(Y - uh)R_{bi}(Y)f(Y)}{f_b(Y)}\,du\,dY$$

(4.18)
$$= \sqrt{n}\,\mathbb{E}[R_l(Y)R_{bi}(Y)] + \sqrt{n}\int K(u)$$

$$\times\{R_l(Y - uh)f(Y - uh) - R_l(Y)f(Y)\}\frac{R_{bi}(Y)f(Y)}{f_b(Y)}\,du\,dY$$

$$- \sqrt{n}\int K(u)\{f_b(Y) - f(Y)\}\frac{R_l(Y)R_{bi}(Y)f(Y)}{f_b(Y)}\,du\,dY$$

$$\triangleq \sqrt{n}\,\mathbb{E}[R_l(Y)R_i(Y)] + J_{nil}^{(7)} + o(1).$$

Moreover, by conditions 1, 4 and 5, we have

(4.19)
$$|J_{nil}^{(7)}| \leq c\sqrt{n}\,h^4 b^{-1}\int |K(u)|u^4 R_i(Y)f(Y)\,du\,dY$$

$$= O(n^{-4c_1 + c_2 + 1/2}) = o(1).$$

Together with (4.18), (4.19) and the expression of (4.15), we need only to show that

(4.20) $$\sqrt{n}\,\mathbb{E}\left[\frac{R_{bi}(Y)R_{bl}(Y)\hat{f}_b(Y)}{f_b(Y)}\right] = \sqrt{n}\,\mathbb{E}[R_i(Y)R_l(Y)] + o(1).$$

For this purpose we express $\hat{f}_b$ as a sum of five terms, where c is a positive constant to be specified later:

$$
\begin{aligned}
\hat{f}_b(Y) &= \hat{f}(Y)I\big(\hat{f}(Y) \geq b\big) + bI\big(\hat{f}(Y) < b\big) \\
&= \hat{f}(Y)I\big(f(Y) > b + 2c(h^4 + n^{-1/2}h^{-1}\log n)\big) \\
&\quad + f(Y)\big\{I\big(\hat{f}(Y) \geq b\big) - I\big(f(Y) \geq b + 2c(h^4 + n^{-1/2}h^{-1}\log n)\big)\big\} \\
&\quad + \big(\hat{f}(Y) - f(Y)\big)\big\{I\big(\hat{f}(Y) \geq b\big) \\
&\qquad\qquad\qquad\qquad -I\big(f(Y) \geq b + 2c(h^4 + n^{-1/2}h^{-1}\log n)\big)\big\} \\
&\quad + bI\big(f(Y) < b - 2c(h^4 + n^{-1/2}h^{-1}\log n)\big) \\
&\quad + b\big\{I\big(\hat{f}(Y) < b\big) - I\big(f(Y) < b\big) - 2c(h^4 + n^{-1/2}h^{-1}\log n)\big\} \\
&\hat{=} J_{n8}(Y) + J_{n9}(Y) + J_{n10}(Y) + J_{n11}(Y) + J_{n12}(Y).
\end{aligned}
$$

(4.21)

It is now clear that (4.20) is equivalent to

$$
(4.22) \quad \sqrt{n}\,\mathbb{E}\left[\frac{J_{n8}(Y)R_{bi}(Y)R_{bl}(Y)}{f_b(Y)}\right] = \sqrt{n}\,\mathbb{E}\big[R_i(Y)R_l(Y)\big] + o(1)
$$

and

$$
(4.23) \quad \sqrt{n}\,\mathbb{E}\left[\frac{J_{nk}(Y)R_{bi}(Y)R_{bl}(Y)}{f_b(Y)}\right] = o(1), \qquad k = 9,\ldots,12.
$$

One can show (4.22) using the argument for proving (4.18) and (4.19). For (4.23) we break it into four separate cases to proceed. By Lemma 3.1 there exists a positive constant c such that

$$
\sup_y \big|\hat{f}(Y) - f(Y)\big| \leq 2c(h^4 + n^{-1/2}h^{-1}\log n) \quad \text{a.s.}
$$

Since the set $\{f(Y) > b + 2c(h^4 + n^{-1/2}h^{-1}\log n)\}$ is contained in the set $\{\hat{f}(Y) > b\}$ with probability 1, we can see, by condition 6, that

$$
\begin{aligned}
&\sqrt{n}\left|\mathbb{E}\left[\frac{J_{n9}(Y)R_{bi}(Y)R_{bl}(Y)}{f_b(Y)}\right]\right| \\
&\quad \leq \sqrt{n}\,\mathbb{E}\big\{\big|R_i(Y)R_l(Y)\big| \\
&\qquad\qquad \times \big|I\big(\hat{f}(Y) \geq b\big) - I\big(f(Y) \geq b + 2c(h^4 + n^{-1/2}h^{-1}\log n)\big)\big|\big\} \\
&\quad = \sqrt{n}\,\mathbb{E}\big\{\big|R_i(Y)R_l(Y)\big| \\
&\qquad\qquad \times I\big(f(Y) \leq b + 2c(h^4 + n^{-1/2}h^{-1}\log n), \hat{f}(Y) \geq b\big)\big\} \\
&\quad \leq \sqrt{n}\,\mathbb{E}\big|R_i(Y)R_l(Y)\big|I\big(f(Y) \leq b + 2c(h^4 + n^{-1/2}h^{-1}\log n)\big) \\
&\quad = o(1).
\end{aligned}
$$

(4.24)

Using a similar argument for (4.24), we obtain

$$\sqrt{n}\left|\mathbb{E}\left[\frac{J_{n10}(Y)R_{bi}(Y)R_{bl}(Y)}{f_b(Y)}\right]\right| = o(1),$$

$$\sqrt{n}\left|\mathbb{E}\left[\frac{J_{n11}(Y)R_{bi}(Y)R_{bl}(Y)}{f_b(Y)}\right]\right|$$
$$\leq \sqrt{n}\left|\mathbb{E}\left[R_{bi}(Y)R_{bl}(Y)I\big(f(Y) < b - 2c(h^4 + n^{-1/2}h^{-1}\log n)\big)\right]\right|$$
$$= o(1)$$

and

$$\sqrt{n}\left|\mathbb{E}\left[\frac{J_{n12}(Y)R_{bi}(Y)R_{bl}(Y)}{f_b(Y)}\right]\right|$$
$$\leq \sqrt{n}\left|\mathbb{E}\left[R_{bi}(Y)R_{bl}(Y)I\big(f(Y) < b + 2c(h^4 + n^{-1/2}h^{-1}\log n)\big)\right]\right|$$
$$= o(1).$$

So we have shown (4.20), and hence (4.16), completing the proof of Theorem 2.1. $\square$

5. Proof of Theorem 2.2. The standard perturbation theory gives the following chain-rule formulas [see also Sun (1988)]

$$\frac{\partial \lambda_i}{\partial a_{kl}} = \mathbf{b}_i\frac{\partial \mathbf{\Lambda}_1}{\partial a_{kl}}\mathbf{b}_i,$$

$$\frac{\partial \mathbf{b}_i}{\partial a_{kl}} = \sum_{\substack{m=1 \\ m \neq i}}^{d} \frac{\mathbf{b}_m\mathbf{b}_m^{\mathsf{T}}(\partial \mathbf{\Lambda}_1/\partial a_{kl})\mathbf{b}_i}{(\lambda_i - \lambda_m)}.$$

By Theorem 2.1, the perturbation $\mathbf{\Lambda}_{1n} - \mathbf{\Lambda}_1$ is small when n is large enough. Then, by a Taylor expansion, we have

(5.1)
$$\sqrt{n}\left(\lambda_i(\mathbf{\Lambda}_{1n}) - \lambda_i(\mathbf{\Lambda}_1)\right)$$
$$= \sqrt{n}\,\mathbf{b}_i(\mathbf{\Lambda}_1)^{\mathsf{T}}(\mathbf{\Lambda}_{1n} - \mathbf{\Lambda}_1)\mathbf{b}_i(\mathbf{\Lambda}_1) + \mathbf{R}_i(\mathbf{\Lambda}_1^*),$$

where $\mathbf{R}_i(\mathbf{\Lambda}_1^*) = \sqrt{n}\,\mathbf{b}_i(\mathbf{\Lambda}_1^*)^{\mathsf{T}}(\mathbf{\Lambda}_{1n} - \mathbf{\Lambda})\mathbf{b}_i(\mathbf{\Lambda}_1^*) - \sqrt{n}\,\mathbf{b}_i(\mathbf{\Lambda}_1)^{\mathsf{T}}(\mathbf{\Lambda}_{1n} - \mathbf{\Lambda})\mathbf{b}_i(\mathbf{\Lambda}_1)$ and $(\mathbf{\Lambda}_1^* - \mathbf{\Lambda}_1) \to \mathbf{0}$ in probability. The proof of (2.2) can be completed by proving $\mathbf{R}_i(\mathbf{\Lambda}_1^*) \to \mathbf{0}$ in probability. In view of the expansion in (5.1), this is easily done, so we omit the details. The assertion (2.3) can be proved similarly. The proof is complete. $\square$

REFERENCES

DUAN, N. and LI, K. C. (1991). Slicing regression: a link free regression method. *Ann. Statist.* **19** 505–530.

FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823.

FRIEDMAN, J. H. and TUKEY, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Comput.* **C-23** 881–889.

HALL, P. (1989). On projection pursuit regression. *Ann. Statist.* **17** 583–588.

HÄRDLE, W. and STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** 986–995.

HASTIE, T. and TIBSHIRANI, R. (1986). Generalized additive models. *Statist. Sci.* **1** 297–318.

HSING, T. and CARROLL, R. J. (1992). An asymptotic theory for sliced inverse regression. *Ann. Statist.* **20** 1040–1061.

HUBER, P. (1985). Projection pursuit (with discussion). *Ann. Statist.* **13** 435–475.

KATO, T. (1983). *Perturbation Theory for Linear Operation*, 2nd ed. Springer, New York.

LI, K. C. (1989). Data visualization with SIR: a transformation based projection pursuit method. *UCLA Statist. Ser.* **24**.

LI, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.* **86** 316–342.

NOLAN, D. and POLLARD, D. (1987). *U*-processes: rates of convergence. *Ann. Statist.* **15** 780–799.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

POWELL, J. L., STOCK, J. H. and STOKER, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica* **57** 1403–1430.

RAO, B. L. S. P. (1983). *Nonparametric Functional Estimation*. Academic Press, Orlando, FL.

STONE, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimate. *Ann. Statist.* **12** 1285–1297.

SUN, S. G. (1988). Analytic expressions for the derivatives of the eigenvalues and eigenvectors of a matrix. *Adv. in Math.* (*Beijing*) **17** 391–397 (in Chinese).

TYLER, D. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9** 725–736.

ZHU, L. X. (1993). Convergence rates of the empirical processes indexed by the classes of functions with applications. *J. Systems Sci. Math. Sci.* **13** 33–41 (in Chinese).

ZHU, L. X. and FANG, K. T. (1992). Projection pursuit approximation for nonparametric regression. *Proceedings of the Order Statistic and Nonparametric: Theory and Methods* (P. K. Sen and I. A. Salama, eds.) 455–469. North-Holland, Amsterdam.

PROBABILITY LABORATORY
INSTITUTE OF APPLIED MATHEMATICS
CHINESE ACADEMY OF SCIENCES
BEIJING, 100080
CHINA

DEPARTMENT OF MATHEMATICS
HONG KONG BAPTIST UNIVERSITY
224 WATERLOO ROAD
KOWLOON
HONG KONG