Statistica Sinica **5**(1995), 727-736

ASYMPTOTICS OF SLICED INVERSE REGRESSION

Li-Xing Zhu and Kai W. Ng

Chinese Academy of Sciences and University of Hong Kong

Abstract. Sliced Inverse Regression is a method for reducing the dimension of the explanatory variables x in non-parametric regression problems. Li (1991) discussed a version of this method which begins with a partition of the range of y into slices so that the conditional covariance matrix of x given y can be estimated by the sample covariance matrix within each slice. After that the mean of the conditional covariance matrix is estimated by averaging the sample covariance matrices over all slices. Hsing and Carroll (1992) have derived the asymptotic properties of this procedure for the special case where each slice contains only two observations. In this paper we consider the case that each slice contains an arbitrary but fixed number of y_i and more generally the case when the number of y_i per slice goes to infinity. The asymptotic properties of the associated eigenvalues and eigenvectors are also obtained.

Key words and phrases: Asymptotics, sliced inverse regression, dimension reduction, eigenvalues and eigenvectors.

1. Introduction

Sliced Inverse Regression (SIR) is a useful technique for dimension reduction in non-parametric regression problems where the response variable y depends on K unknown linear combinations of the explanatory variables $(x_1, \ldots, x_p) = \boldsymbol{x}^\top$, but the exact form of dependence is unknown. These regression problems can be represented by the model

$$y = f(\boldsymbol{\beta}_1^\top \boldsymbol{x}, \boldsymbol{\beta}_2^\top \boldsymbol{x}, \ldots, \boldsymbol{\beta}_K^\top \boldsymbol{x}, \varepsilon), \tag{1.1}$$

where f is an unknown function defined on R^{K+1}, $K < p$, $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ are unknown $p \times 1$ vectors, and ε and $\boldsymbol{x}$ are independent random variables. The essential feature of the model is that instead of the p-dimensional $\boldsymbol{x}$ we need only the K-dimensional variables $(\boldsymbol{\beta}_1^\top \boldsymbol{x}, \ldots, \boldsymbol{\beta}_K^\top \boldsymbol{x})$ for predicting y. As Li (1991) pointed out, since the function f is unknown, the parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ are not identifiable; however, the linear space, B, spanned by the $\boldsymbol{\beta}$'s is identifiable. The space B is called the effective dimension-reduction (e.d.r.) space and any basis of B is called a set of e.d.r. directions. Naturally, the first step of solving the regression problem (1.1) is to estimate the e.d.r. space B. Then the form of the function f

can be explored based on the estimated e.d.r. space. Of course, since f and the $\boldsymbol{\beta}$'s are irretrievably confounded, the form of f usually depends on the particular choice of the representation of the e.d.r. space.

Under model (1.1) and the condition that the conditional expectation of any linear function of $x_1, \ldots, x_p$ given $\boldsymbol{\beta}_1^\top \boldsymbol{x}, \ldots, \boldsymbol{\beta}_K^\top \boldsymbol{x}$ is also a linear function of $\boldsymbol{\beta}_1^\top \boldsymbol{x}, \ldots, \boldsymbol{\beta}_K^\top \boldsymbol{x}$, Li (1991, Theorem 3.1) shows that the centered "inverse regression" curve $E(\boldsymbol{x}|y) - E(\boldsymbol{x})$ is confined to a K-dimensional linear subspace spanned by $\boldsymbol{\Sigma}_x \boldsymbol{\beta}_1, \boldsymbol{\Sigma}_x \boldsymbol{\beta}_2, \ldots, \boldsymbol{\Sigma}_x \boldsymbol{\beta}_K$, where $\boldsymbol{\Sigma}_x$ denotes the covariance matrix of $\boldsymbol{x}$. This relates the inverse regression $E(\boldsymbol{x}|y)$ to the e.d.r. space B. Let $\boldsymbol{z} = \boldsymbol{\Sigma}_x^{-1/2}(\boldsymbol{x} - E(\boldsymbol{x}))$ be the standardization of $\boldsymbol{x}$; then the e.d.r. space B is spanned by $\boldsymbol{\Sigma}_x^{-1/2}\boldsymbol{\eta}_1, \ldots, \boldsymbol{\Sigma}_x^{-1/2}\boldsymbol{\eta}_K$, where $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_K$ are the column eigenvectors associated with the K largest eigenvalues of the covariance matrix $\mathrm{Cov}(E(\boldsymbol{z}|\boldsymbol{y}))$. Alternatively, the $\boldsymbol{\eta}$'s are obtained as the eigenvectors associated with the K smallest eigenvalues of the average covariance matrix $E(\mathrm{Cov}(\boldsymbol{z}|\boldsymbol{y}))$. Therefore, there are two versions of sliced inverse regression corresponding to these two eigen analyses. The asymptotic results for the first version can be found, for example, in Li (1991) and Duan and Li (1991). In this paper we consider the SIR corresponding to the second eigen analysis suggested in the Remark 5.3 of Li (1991) but we deal directly with the original covariance matrix

$$\boldsymbol{\Lambda} = E(\mathrm{Cov}(\boldsymbol{x}|y)) \tag{1.2}$$

and its eigenvalues and eigenvectors. The corresponding SIR method first estimates the conditional covariance matrix $\mathrm{Cov}(\boldsymbol{x}|y)$ by the sample covariance matrix of the explanatory variables for different ranges of y and then estimates $\boldsymbol{\Lambda}$ by the average $\boldsymbol{\Lambda}_n$ of these sample covariance matrices. Hsing and Carroll (1992) have shown the root n consistency of this estimator of $\boldsymbol{\Lambda}$ for the special case that each slice contains two of the ordered $y_{(i)},\ i = 1, 2, \ldots, n$, under some smoothness conditions for the inverse regression curve $\boldsymbol{m}(y) = E(\boldsymbol{x}|y)$. In this paper, we establish the root n consistency for the case where the number c of $y_{(i)}$ in each of the H slices is an arbitrarily fixed number, and for the case in which $c = c_n \to \infty$ as $n \to \infty$. We also investigate the limiting behaviour of the eigenvalues and eigenvectors of $\boldsymbol{\Lambda}_n$.

2. Assumptions and Main Results

Let the data $(y_i, \boldsymbol{x}_i),\ i = 1, \ldots, n$, be ordered according to the values of y_i, and denote the ordered data set by $(y_{(i)}, \boldsymbol{x}_{(i)}),\ i = 1, \ldots, n$, where $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$. The $\boldsymbol{x}_{(i)}$ are called the concomitants of order statistics by Yang (1977). We introduce a double subscripts (h, j) where the first refers to the slice number and the second refers to the order number of an observation in the given slice.

That is,

$$y_{(h,j)} = y_{(c(h-1)+j)}, \qquad \boldsymbol{x}_{(h,j)} = \boldsymbol{x}_{(c(h-1)+j)}. \tag{2.1}$$

Using this notation the estimator $\boldsymbol{\Lambda}_n$ based on H slices of c data points can be written as

$$\begin{aligned}\boldsymbol{\Lambda}_n &= \frac{1}{H}\sum_{h=1}^{H}\left\{\frac{1}{c-1}\sum_{j=1}^{c}\left(\boldsymbol{x}_{(h,j)} - \frac{1}{c}\sum_{\ell=1}^{c}\boldsymbol{x}_{(h,\ell)}\right)\left(\boldsymbol{x}_{(h,j)} - \frac{1}{c}\sum_{\ell=1}^{c}\boldsymbol{x}_{(h,\ell)}\right)^{\top}\right\}\\ &= \frac{1}{H}\sum_{h=1}^{H}\left\{\frac{1}{c(c-1)}\sum\sum_{1\le j<\ell\le c}(\boldsymbol{x}_{(h,\ell)} - \boldsymbol{x}_{(h,j)})(\boldsymbol{x}_{(h,\ell)} - \boldsymbol{x}_{(h,j)})^{\top}\right\}\\ &= \frac{1}{n(c-1)}\sum_{h=1}^{H}\left\{\sum\sum_{1\le j<\ell\le c}(\boldsymbol{x}_{(h,\ell)} - \boldsymbol{x}_{(h,j)})(\boldsymbol{x}_{(h,\ell)} - \boldsymbol{x}_{(h,j)})^{\top}\right\}. \end{aligned}\tag{2.2}$$

In practice, the number of observations in the last slice may be less than c. But this does not change the results in this paper because the number of slices, H, is very large.

Denote the inverse regression curve and its residual respectively by

$$\boldsymbol{m}(y) = E(\boldsymbol{x}|y) \quad \text{and} \quad \boldsymbol{\varepsilon} = \boldsymbol{x} - \boldsymbol{m}(y). \tag{2.3}$$

Note that the concomitants $\boldsymbol{\varepsilon}_{(i)} = \boldsymbol{x}_{(i)} - \boldsymbol{m}(y_{(i)})$ are conditionally independent with mean zero given the order statistics $y_{(i)}$ (c.f. Yang (1977)).

Following Hsing and Carroll (1992) we need some smoothness conditions on the inverse regression curve $\boldsymbol{m}(y)$. Let $\Pi_n(B)$ be the collection of all the n-point partitions $-B \le y^*_{(1)} \le \cdots \le y^*_{(n)} \le B$ of the closed interval $[-B, B]$, where $B > 0$ and $n \ge 1$. Any vector-valued or real-valued function $\boldsymbol{m}(y)$ is said to have a *total variation of order* r if for any fixed $B > 0$

$$\lim_{n\to\infty}\frac{1}{n^r}\sup_{\Pi_n(B)}\sum_{i=1}^{n}\|\boldsymbol{m}(y^*_{(i+1)}) - \boldsymbol{m}(y^*_{(i)})\| = 0. \tag{2.4}$$

Furthermore, if there exist a non-decreasing real-valued function M and a real number B_0 such that for any two points, say y_1 and y_2, both in $(-\infty, -B_0]$ or both in $[B_0, \infty)$,

$$\|\boldsymbol{m}(y_1) - \boldsymbol{m}(y_2)\| \le |M(y_1) - M(y_2)|, \tag{2.5}$$

then we say that the function $\boldsymbol{m}(y)$ is *non-expansive in the metric of* M *on both sides of* B_0.

The following theorem concerns the asymptotic distributions of $\boldsymbol{\Lambda}_n$ as defined in (2.1) when c, the number of observations per slice, is fixed. For convenience, we denote by $\text{vech}(\boldsymbol{S})$ the vectorization of a symmetric $p \times p$ matrix $\boldsymbol{S}$; that is, vech($\boldsymbol{S}$) is a vector consisting of the $p(p+1)/2$ distinct elements of $\boldsymbol{S}$, in the order of row by row (or equivalently, column by column).

Theorem 1. *Assume the following four conditions:*
(i) $E(\|\boldsymbol{x}\|^4) < \infty$.
(ii) *The inverse regression function* $\boldsymbol{m}(y)$ *has a total variation of order* $r = 1/4$.
(iii) $\boldsymbol{m}(y)$ *is non-expansive in the metric of* $M(y)$ *on both sides of a positive number* B_0 *such that*

$$M^4(t)P(y > t) \to 0 \quad \textit{as } t \to \infty \quad \text{as } n \to \infty. \tag{2.6}$$

(iv) *The elements of* $\boldsymbol{V}(y) = \mathrm{Cov}(\boldsymbol{\varepsilon}|y) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top|y)$, *as functions of* y, *all have total variation of order* $r = 1$.
Then the vectorization of the matrix $\sqrt{n}(\boldsymbol{\Lambda}_n - \boldsymbol{\Lambda})$, *namely* $\sqrt{n}$ vech $(\boldsymbol{\Lambda}_n - \boldsymbol{\Lambda})$, *is asymptotically multinormal as* $n \to \infty$ *with zero means and covariance matrix*

$$\mathrm{Cov}(\mathrm{vech}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)) + \frac{2}{c-1}E\{(\mathrm{vech}(\boldsymbol{V}(y)))((\mathrm{vech}(\boldsymbol{V}(y)))^\top)\}. \tag{2.7}$$

This theorem shows that the asymptotic covariance matrix of the estimates is smaller for larger number of observations per slice. In other words, too many slices with too few observations is not very good for the variances of the estimates. If we take the limit as $c \to \infty$ (after $n \to \infty$), the second term disappears and the first term of (2.7) becomes the asymptotic covariance matrix. One may interpret this as the covariance matrix of the estimates when the number of observations per slice is large, but still small when compared with the total sample size. Note that with the repeated limits, there is no restriction on c. If we restrict the manner in which c and n go to infinity simultaneously, as the next theorem shows, we can relax the smoothness conditions on the inverse regression curve at the cost of stronger conditions on the moments of the covariates $\boldsymbol{x}$. Since the conditions on the moments of $\boldsymbol{x}$ are easier to check than those on the inverse regression curve, this trade-off is often worthwhile.

Theorem 2. *Assume the following four conditions:*
(i) *There exists a positive number* b *such that* $E(\|\boldsymbol{x}\|^{4+b}) < \infty$.
(ii) $\boldsymbol{m}(y)$ *has a total variation of order* $r > 0$.
(iii) $\boldsymbol{m}(y)$ *is non-expansive in the metric of* $M(y)$ *on both sides of a positive number* B_0 *such that*

$$M^{4+b}(t)P(y > t) \to 0 \quad \textit{as } t \to \infty. \tag{2.8}$$

(iv) $c = O(n^\alpha)$, *where* $\alpha = \frac{1}{2} - \max\{2r,\, r + \frac{1}{4+b},\, \frac{2}{4+b}\} > 0$.

Then $\sqrt{n}\,\mathrm{vech}(\boldsymbol{\Lambda}_n - \boldsymbol{\Lambda})$ *is asymptotically multinormal as* $n \to \infty$ *with zero means and covariance matrix*

$$\mathrm{Cov}(\mathrm{vech}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)). \tag{2.9}$$

With suitable choices of b and r, we can obtain a wide range of c so that the asymptotic normality holds. If $\boldsymbol{x}$ is normally distributed and $\boldsymbol{m}(y)$ has bounded variation on every finite interval, the conclusion of Theorem 2 holds for $c = O(\sqrt{n}/\log n)$. Furthermore, if $\boldsymbol{x}$ has a bounded support and $\boldsymbol{m}(y)$ has bounded first derivatives on every finite interval, the conclusion holds for $c = o(\sqrt{n})$.

Based on the above theorems we can obtain the root n consistency and asymptotic normality of the eigenvalues and eigenvectors of $\boldsymbol{\Lambda}_n$. Let $\lambda_1(\boldsymbol{\Lambda}) \geq \cdots \geq \lambda_p(\boldsymbol{\Lambda}) \geq 0$ and $\boldsymbol{b}_1(\boldsymbol{\Lambda}), \ldots, \boldsymbol{b}_p(\boldsymbol{\Lambda})$ be respectively the eigenvalues and the associated eigenvectors of $\boldsymbol{\Lambda}$. The spectral decomposition of $\boldsymbol{\Lambda}$ is

$$\boldsymbol{\Lambda} = \sum_{i=1}^{p} \lambda_i(\boldsymbol{\Lambda})\boldsymbol{b}_i(\boldsymbol{\Lambda})\boldsymbol{b}_i(\boldsymbol{\Lambda})^\top. \tag{2.10}$$

Similarly the spectral decomposition of $\boldsymbol{\Lambda}_n$ is

$$\boldsymbol{\Lambda}_n = \sum_{i=1}^{p} \lambda_i(\boldsymbol{\Lambda}_n)\boldsymbol{b}_i(\boldsymbol{\Lambda}_n)\boldsymbol{b}_i(\boldsymbol{\Lambda}_n)^\top. \tag{2.11}$$

Theorem 3. *Under the conditions either in Theorem* 1 *or in Theorem* 2*,*

$$\sqrt{n}(\lambda_i(\boldsymbol{\Lambda}_n) - \lambda_i(\boldsymbol{\Lambda})) = O_p(1) \tag{2.12}$$

as $n \to \infty$*.*

Theorem 4. *Under the conditions in Theorem* 1 *and the assumption that all non-zero* $\lambda_i(\boldsymbol{\Lambda})$ *are distinct, we have*

$$\sqrt{n}\{\lambda_i(\boldsymbol{\Lambda}_n) - \lambda_i(\boldsymbol{\Lambda})\} \xrightarrow{L} \boldsymbol{b}_i(\boldsymbol{\Lambda})^\top \boldsymbol{W}\boldsymbol{b}_i(\boldsymbol{\Lambda}) \tag{2.13}$$

and

$$\sqrt{n}\{\boldsymbol{b}_i(\boldsymbol{\Lambda}_n) - \boldsymbol{b}_i(\boldsymbol{\Lambda})\} \xrightarrow{L} \boldsymbol{A}_i\boldsymbol{W}\boldsymbol{b}_i, \tag{2.14}$$

as $n \to \infty$*, where*

$$\boldsymbol{A}_i = \sum_{\ell=1,\neq i}^{p} \boldsymbol{b}_\ell(\boldsymbol{\Lambda})\boldsymbol{b}_\ell(\boldsymbol{\Lambda})^\top \Big/ \{\lambda_i(\boldsymbol{\Lambda}) - \lambda_\ell(\boldsymbol{\Lambda})\} \tag{2.15}$$

and $\boldsymbol{W}$ *is a random* $p \times p$ *symmetric matrix such that* $\mathrm{vech}(\boldsymbol{W})$ *has a multivariate normal distribution with zero means and covariance matrix given by* (2.7)*.*

Furthermore, if the conditions of Theorem 1 *are replaced by the conditions of Theorem* 2*, the asymptotic distributions of* (2.13) *and* (2.14) *still hold except that in this case the covariance matrix of* $\mathrm{vech}(\boldsymbol{W})$ *is given by* (2.9)*.*

3. Outline of Proofs

Since the proof of Theorem 1 and part of the proof of Theorem 2 follow the line of thoughts along Hsing and Carroll (1992), we only present an outline of the proofs. Readers may refer to Zhu and Ng (1993) for more details. First, make use of the fact that $\boldsymbol{x} = \boldsymbol{m}(y) + \varepsilon$ by (2.3) and hence that

$$\begin{aligned}
&(\boldsymbol{x}_{(h,\ell)} - \boldsymbol{x}_{(h,j)})(\boldsymbol{x}_{(h,\ell)} - \boldsymbol{x}_{(h,j)})^\top \\
&= (\boldsymbol{m}(y_{(h,\ell)}) - \boldsymbol{m}(y_{(h,j)}))(\boldsymbol{m}(y_{(h,\ell)}) - \boldsymbol{m}(y_{(h,j)}))^\top \\
&\quad + (\varepsilon_{(h,\ell)} - \varepsilon_{(h,j)})(\boldsymbol{m}(y_{(h,\ell)}) - \boldsymbol{m}(y_{(h,j)}))^\top \\
&\quad + (\boldsymbol{m}(y_{(h,\ell)}) - \boldsymbol{m}(y_{(h,j)}))(\varepsilon_{(h,\ell)} - \varepsilon_{(h,j)})^\top + (\varepsilon_{(h,\ell)} - \varepsilon_{(h,j)})(\varepsilon_{(h,\ell)} - \varepsilon_{(h,j)})^\top. \qquad (3.1)
\end{aligned}$$

Substituting (3.1) into (2.2) and expanding the triple summation we obtain a 4-term expression for $\boldsymbol{\Lambda}_n$. Therefore we can write

$$\sqrt{n}(\boldsymbol{\Lambda}_n - \boldsymbol{\Lambda}) = \boldsymbol{T}_1 + \boldsymbol{T}_2 + \boldsymbol{T}_3 + \boldsymbol{T}_4, \qquad (3.2)$$

where the first three terms are respectively the first 3 terms of (3.1) after being operated on by the triple summation $n^{-1/2}(c-1)^{-1}\sum_{h=1}^{H}\sum\sum_{1\le j<\ell\le c}$ and the last term is

$$\boldsymbol{T}_4 = \frac{1}{\sqrt{n}(c-1)} \sum_{h=1}^{H} \sum\sum_{1\le j<\ell\le c} \left\{ (\varepsilon_{(h,\ell)} - \varepsilon_{(h,j)})(\varepsilon_{(h,\ell)} - \varepsilon_{(h,j)})^\top - 2\boldsymbol{\Lambda} \right\}. \qquad (3.3)$$

We can prove that the first three terms of (3.2) all converge to zero in probability as $n \to \infty$ and the asymptotic distribution of $\mathrm{vech}(\boldsymbol{T}_4)$ is multivariate normal with zero mean vector and covariance matrix given by (2.7). The following is a sketch.

Rearranging the inner double summation and then moving the summation over h to the inside, we obtain

$$\begin{aligned}
\boldsymbol{T}_1 &= \frac{1}{\sqrt{n}(c-1)} \sum_{h=1}^{H} \sum_{m=1}^{c-1} \sum_{j=1}^{c-m} \{\boldsymbol{m}(y_{(h,j+m)}) - \boldsymbol{m}(y_{(h,j)})\}\{\boldsymbol{m}(y_{(h,j+m)}) - \boldsymbol{m}(y_{(h,j)})\}^\top \\
&= \frac{1}{\sqrt{n}(c-1)} \sum_{m=1}^{c-1} \sum_{\ell=1}^{m} \boldsymbol{C}_{\ell m}, \qquad (3.4)
\end{aligned}$$

where

$$\boldsymbol{C}_{\ell m} = \sum_{h=1}^{H} \sum_{*} \{\boldsymbol{m}(y_{(h,\ell+jm)}) - \boldsymbol{m}(y_{(h,\ell+(j-1)m)})\}\{\boldsymbol{m}(y_{(h,\ell+jm)}) - \boldsymbol{m}(y_{(h,\ell+(j-1)m)})\}^\top \qquad (3.5)$$

and $*$ is a summation over j subject to the restriction $\ell + jm \leq c - m$. For a fixed p such that $0 < p < 1/2$, let us divide the outer summation over h into three summations: from 1 to Hp, then $Hp+1$ to $H(1-p)$ and then $H(1-p)+1$ to H, so that we can write

$$\boldsymbol{C}_{\ell m} = \boldsymbol{C}_{\ell m}^{(1)} + \boldsymbol{C}_{\ell m}^{(2)} + \boldsymbol{C}_{\ell m}^{(3)}. \tag{3.6}$$

Parallel to the proof of Lemma 1 of Hsing and Carroll we can show under the conditions of Theorem 1 that for each element in the symmetric matrix $\boldsymbol{C}_{\ell m}^{(i)}$, $i = 1, 2, 3$, the maximum over all ℓ and m has order $o_p(\sqrt{n})$; then $\boldsymbol{T}_1$ converges to zero in probability.

Next consider $\boldsymbol{T}_2$ and $\boldsymbol{T}_3$. Since they can be treated in a similar way, we only show the case of $\boldsymbol{T}_2$. Again we only demonstrate the proof for the first diagonal element of $\boldsymbol{T}_2$. Denote this element by T_2 and use $m(y)$ for the corresponding element of $\boldsymbol{m}(y)$ which gives rise to T_2. Let $\bar{\varepsilon}_{(n)}$ and $\bar{\varepsilon}_{(1)}$ be respectively the largest and smallest of the corresponding ε_i's. It is clear that

$$\begin{aligned} T_2 &\leq \frac{\bar{\varepsilon}_{(n)} - \bar{\varepsilon}_{(1)}}{\sqrt{n}(c-1)} \sum_{h=1}^{H} \sum\sum_{1 \leq j < \ell \leq c} |m(y_{(h,\ell)}) - m(y_{(h,j)})| \\ &\leq \frac{\bar{\varepsilon}_{(n)} - \bar{\varepsilon}_{(1)}}{\sqrt{n}(c-1)} \sum_{m=1}^{c-1} \sum_{j=1}^{n-m} |m(y_{(j+m)}) - m(y_{(j)})| \\ &\leq \frac{\bar{\varepsilon}_{(n)} - \bar{\varepsilon}_{(1)}}{\sqrt{n}(c-1)} \sum_{m=1}^{c-1} \sum_{i=1}^{m} \sum_{j=1}^{n-1} |m(y_{(j+1)}) - m(y_{(j)})| \\ &\leq \frac{2c}{\sqrt{n}} |\bar{\varepsilon}_{(n)} - \bar{\varepsilon}_{(1)}| \sum_{j=1}^{n-1} |m(y_{(j+1)}) - m(y_{(j)})|. \end{aligned}$$

Applying Lemma A.1 of Hsing and Carroll (1992), we obtain

$$n^{-1/4} |\bar{\varepsilon}_{(n)} - \bar{\varepsilon}_{(1)}| \xrightarrow{p} 0.$$

Also, condition (ii) implies

$$\lim_{n \to \infty} n^{-1/4} \sum_{h=1}^{n-1} |m(y_{(h+1)}) - m(y_{(h)})| = 0.$$

This completes the proof for T_2.

For dealing with $\boldsymbol{T}_4$, we write $\boldsymbol{T}_4$ as the sum of the following two terms

$$\boldsymbol{T}_4^{(1)} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j^\top - \boldsymbol{\Lambda}) \tag{3.7}$$

and

$$\boldsymbol{T}_4^{(2)} = \frac{1}{\sqrt{n}(c-1)} \sum_{h=1}^{H} \sum_{j \neq \ell} \boldsymbol{\varepsilon}_{(h,\ell)} \boldsymbol{\varepsilon}_{(h,j)}^{\top} = \frac{1}{\sqrt{n}(c-1)} \sum_{h=1}^{H} G_h. \tag{3.8}$$

Due to the conditional uncorrelation between $\boldsymbol{T}_4^{(1)}$ and $\boldsymbol{T}_4^{(2)}$, we establish the fact that the asymptotic distribution of $\boldsymbol{T}_4$ is a convolution of two normal distributions with zero mean; hence, it is a normal distribution with zero mean and the variance defined in (2.7).

By the Central Limit Theorem, the asymptotic distribution of $\boldsymbol{a}^{\top}\mathrm{vech}(\boldsymbol{T}_4^{(1)})$ is normal with zero mean and variance $\boldsymbol{a}^{\top} E\,\mathrm{Cov}(\mathrm{vech}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\top}))\boldsymbol{a}$. On the other hand, observe that conditionally each term G_h has mean zero and independent of each other. Thus parallel to the argument used in Theorem 2.3 of Hsing and Carroll (1992), $\boldsymbol{T}_4^{(2)}$ is asymptotically normal with zero mean. To show that the asymptotic variance of $\boldsymbol{T}_4^{(2)}$ is equal to the second term defined in (2.7), consider only the first diagonal element of $\boldsymbol{T}_4^{(2)}$ for simplicity. The plain face ε is used to denote the first element of $\boldsymbol{\varepsilon}$. The variance of the first diagonal element of $\boldsymbol{T}_4^{(2)}$ is equal to

$$\begin{aligned}
&E\left\{\frac{1}{\sqrt{n}(c-1)} \sum_{h=1}^{H} \sum_{j \neq \ell} \varepsilon_{(h,\ell)} \varepsilon_{(h,j)}\right\}^2 = \frac{1}{n(c-1)^2} E\left\{E\left(\left\{\sum_{h=1}^{H} \sum_{j \neq \ell} \varepsilon_{(h,\ell)} \varepsilon_{(h,j)}\right\}^2 \middle| \mathcal{F}_n\right)\right\} \\
&= \frac{1}{n(c-1)^2} \sum_{h=1}^{H} \sum_{\substack{j \neq \ell \\ k \neq i}} E\left\{E\left(\left\{\varepsilon_{(h,i)}\varepsilon_{(h,j)}\varepsilon_{(h,k)}\varepsilon_{(h,\ell)}\right\} \middle| \mathcal{F}_n\right)\right\} \\
&= \frac{2}{n(c-1)^2} \sum_{h=1}^{H} \sum_{j \neq \ell} E\left\{\varepsilon^2_{(h,\ell)}\varepsilon^2_{(h,j)}\right\} = \frac{2}{n(c-1)^2} \sum_{h=1}^{H} \sum_{j \neq \ell} E\left\{V(y_{(h,\ell)})V(y_{(h,j)})\right\} \\
&= \frac{2}{n(c-1)} \sum_{h=1}^{H} \sum_{j=1} E\{V^2(y_{(h,j)})\} \\
&\quad + \frac{2}{n(c-1)^2} \sum_{h=1}^{H} \sum_{j \neq \ell} E\left\{V(y_{(h,\ell)})[V(y_{(h,j)}) - V(y_{(h,\ell)})]\right\} \\
&= E\Big\{\frac{2}{n(c-1)} \sum_{j=1}^{n} V^2(y_j)\Big\} + o(1) \to \frac{2}{(c-1)} E\{V^2(y)\},
\end{aligned} \tag{3.9}$$

where $V(y) = \mathrm{Cov}(\varepsilon|y)$. The last equation is obtained using an argument which needs condition (iv) and is analogous to the proof of Hsing and Carroll's Theorem 2.3. This completes the proof of Theorem 1.

The convergence in probability for $\boldsymbol{T}_1, \boldsymbol{T}_2$ and $\boldsymbol{T}_3$ of (3.2) under the conditions of Theorem 2 can be shown similarly to Theorem 1. Indeed, under the conditions of Theorem 2, the first three terms of (3.2) can be expressed as follows:

$$\boldsymbol{T}_1 = o_p(c\,n^{-1/2+\max\{2r,2/(4+b)\}}), \tag{3.10}$$

$$\boldsymbol{T}_i = o_p(c\,n^{-1/2+1/(4+b)+\max\{r,1/(4+b)\}}), \qquad i = 2, 3. \tag{3.11}$$

The last term of (3.2), $\boldsymbol{T}_4$, can be decomposed as the sum of (3.7) and (3.8). Analogous to the proof of Theorem 1, it suffices to show that $\boldsymbol{T}_4^{(2)}$ converges to zero in probability. We again demonstrate this for the first diagonal element of $\boldsymbol{T}_4^{(2)}$. Expanding as in (3.9), up to the third equality, we have

$$\begin{aligned} E\left\{\frac{1}{\sqrt{n}(c-1)}\sum_{h=1}^{H}\sum_{j\neq\ell}\varepsilon_{(h,\ell)}\varepsilon_{(h,j)}\right\}^2 &\leq \frac{2}{n(c-1)^2}\sum_{h=1}^{H}\left\{\sum_{i=1}^{c}(E(\varepsilon_{(h,i)}^4))^{1/2}\right\}^2 \\ \leq \frac{2c}{n(c-1)^2}\sum_{h=1}^{H}\sum_{i=1}^{c}E(\varepsilon_{(h,i)}^4) &\leq \frac{4}{n(c-1)}\sum_{i=1}^{n}E(\boldsymbol{\varepsilon}_i^4) \\ = \frac{4}{c-1}E(\varepsilon_1^4) \to 0 \quad \text{as} \quad c \to \infty. \end{aligned}$$

Thus $\boldsymbol{T}_4^{(2)}$ converges to zero in probability and the proof of Theorem 2 is completed.

For the proof of Theorem 3, let $\lambda_1(\boldsymbol{\Lambda}_n) \geq \lambda_2(\boldsymbol{\Lambda}_n) \geq \cdots \geq \lambda_p(\boldsymbol{\Lambda}_n) \geq 0$ and

$$P_j(\boldsymbol{\Lambda}_n) = \sum_{i=w(\lambda_j-1)+1}^{w(\lambda_j)} \boldsymbol{b}_i(\boldsymbol{\Lambda}_n)\boldsymbol{b}_i(\boldsymbol{\Lambda}_n)^\top, \tag{3.12}$$

where $w(\lambda_j)$ is the multiplicity of λ_i, $\sum_{i=1}^{q} w(\lambda_i) = p$. In view of Theorem 1 and Theorem 2, Lemma 4.1 of Tyler (1981) implies that

$$\sqrt{n}(P_j(\boldsymbol{\Lambda}_n) - P_j(\boldsymbol{\Lambda})) = O_p(1), \quad j = 1, \ldots, q. \tag{3.13}$$

On the other hand, $\boldsymbol{\Lambda}_n - \boldsymbol{\Lambda}$ can be expressed as

$$\begin{aligned} \boldsymbol{\Lambda}_n - \boldsymbol{\Lambda} &= \sum_{i=1}^{p}\lambda_i(\boldsymbol{\Lambda}_n)\boldsymbol{b}_i(\boldsymbol{\Lambda}_n)\boldsymbol{b}_i^\top(\boldsymbol{\Lambda}_n) - \sum_{i=1}^{p}\lambda_i(\boldsymbol{\Lambda})\boldsymbol{b}_i(\boldsymbol{\Lambda})\boldsymbol{b}_i^\top(\boldsymbol{\Lambda}) \\ &= \sum_{i=1}^{p}\{\lambda_i(\boldsymbol{\Lambda}_n)-\lambda_i(\boldsymbol{\Lambda})\}\boldsymbol{b}_i^\top(\boldsymbol{\Lambda}_n)\boldsymbol{b}_i^\top(\boldsymbol{\Lambda}_n)+\sum_{i=1}^{q}\lambda_i(\boldsymbol{\Lambda})\{P_i(\boldsymbol{\Lambda}_n)-P_i(\boldsymbol{\Lambda})\} \end{aligned} \tag{3.14}$$

Therefore we obtain the conclusion of Theorem 3:

$$\begin{aligned} &\sqrt{n}(\lambda_j(\boldsymbol{\Lambda}_n) - \lambda_j(\boldsymbol{\Lambda})) \\ &= \boldsymbol{b}_j^\top(\boldsymbol{\Lambda}_n)\sqrt{n}(\boldsymbol{\Lambda}_n - \boldsymbol{\Lambda})\boldsymbol{b}_j(\boldsymbol{\Lambda}_n) - \sum_{i=1}^{p}\lambda_i(\boldsymbol{\Lambda})\boldsymbol{b}_j^\top(\boldsymbol{\Lambda}_n)\{P_i(\boldsymbol{\Lambda}_n) - P_i(\boldsymbol{\Lambda})\}\boldsymbol{b}_j(\boldsymbol{\Lambda}_n) \\ &= O_p(1). \end{aligned}$$

Using the root n consistency in Theorem 3, we can prove Theorem 4 in the same way as for Theorem 2.2 of Zhu and Fang (1993). We omit the proof here.

Acknowledgment

L. X. Zhu was supported by the NSF of China and a Hong Kong UPGC grant and K. W. Ng was supported by a Hong Kong UPGC grant. The authors are grateful to the Editors and a referee for their valuable comments and suggestions which improved greatly the presentation of the paper. The authors especially would like to thank Professor K. C. Li for his encouragement and suggestions which simplified the proof of Theorem 1.

References

Derfringer, G. (1968). Nene iterationsverfahren in der faktorenanalyse. Biometrische Zeitschrift **10**, 57-75.

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.

Hall, P. (1989). On projection pursuit regression. Ann. Statist. **17**, 573-588.

Hardle, 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. J. (1985). Projection pursuit (with discussion). Ann. Statist. **13**, 435-475.

Li, K. C. (1989). Data visualization with SIR: A transformation based projection pursuit method. UCLA Statistical Technique Report #24.

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

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

Yang, S. S. (1977). General distribution theory of the concomitants of order statistics. Ann. Statist. **5**, 996-1002.

Zhu, L. X. and Fang, K. T. (1993). Asymptotics for kernel estimate of sliced inverse regression. Submitted to the Ann. Statist.

Zhu, L. X. and Fang, K. T. (1992). Projection pursuit approximation for nonparametric regression. Proceeding of Order Statistic and Nonparametric: Theory and Methods (Edited by P. K. Sen and I. A. Salama), 455-469, North-Holland.

Zhu, L. X. and Ng, K. W. (1993). Asymptotics of sliced inverse regression. Research Report Serial No. 49, Department of Statistics, University of Hong Kong.

Institute of Applied Mathematics, Science Academy of China, Beijing, 100080.

Department of Statistics, University of Hong Kong, Pokfulam Road, Hong Kong.

(Received September 1993; accepted March 1995)