APPENDIX B. ASSUMPTIONS AND THEOREM PROOFS

Appendix B.1. Conditions and Notations

Let $\mu_B(u) = E(X|B^\top X = u)$ and $w_B(u) = E\{XX^\top|B^\top X = u\}$. The following technical conditions are used in our proofs.

(C1) [Design of X] The covariate X is bounded; its density function $f(x)$ has bounded second order derivatives; functions $\mu_B(u)$ and $w_B(u)$ have bounded derivatives with respect to u and B for $B \in \{B : |BB^\top - B_0B_0^\top| \le c\}$ for some $c > 0$.

(C2) [Conditional distribution function] The conditional probability function $P(z_k = 1|\ B^\top X = u)$ has bounded fourth order derivatives in a small neighbor of B_0 for every $1 \le k \le H$.

(C3) [The CS] For any orthogonal matrix $B \in \mathbb{R}^{p\times d}$ and constant $c > 0$,

$$\inf_{\{B:|BB^\top - B_0B_0^\top|\ge c\}} \sum_{k=1}^{H} E[E\{z_k|B^\top X\} - E\{z_k|B_0^\top X\}]^2 > 0.$$

(C4) [Kernel function] Function $K_0(\cdot)$ is a symmetric univariate density function with bounded first order derivative. Furthermore, we need $n^2K_0(\log n^c) \to 0$ as $n \to \infty$ for some $c > 0$.

(C5) [Bandwidths] For working dimension d, the bandwidths $h_{(t)}$ satisfies $h_{(0)} \propto n^{-1/(p+4)}$, $h_{(t)} = \max\{\varsigma h_{(t-1)}, \hbar\}$ with $1/2 < \varsigma < 1$, and $\hbar \propto n^{-1/(d+4)}$.

If X is not bounded, the trimming scheme of Härdle et al. (1993) can be used. Thus the requirement for bounded X in (C1) can be removed but with more complicated proofs. Lower order of smoothness than (C2) is sufficient if consistency is the only

concern. Compared with Proposition 2, condition (C3) further indicates that the CS is unique; see Cook (1998b). The popular Gaussian kernel and many others satisfy (C4). Many bandwidths, such as the rule of thumb, satisfy (C5).

To simplify the notation, we consider only one slice in the following proofs. Hence, the subscript associated with the slice (i.e., k) can be omitted. For example, we denote z_{ik}, $G_k(u)$ and ϵ_{ik} by z_i, $G(u)$ and ϵ_i respectively. Similarly, we denote $a_{jk}^{(t)}$, $d_{jk}^{(t)}$, and $X_{ijk}^{(t)}$ by $a_j^{(t)}$, $d_j^{(t)}$, and $X_{ij}^{(t)}$, respectively. The main idea to prove Theorem 1 is to show that $\mathcal{S}_{y|x}$ is the attractor of our algorithm. Recall that the estimate of B_0 in the $t-$th iteration is $B_{(t)}$. It follows from Step 2 that

$$\begin{aligned}\Gamma^{(t+1)} &= \ell(B_0) + \Big\{ \sum_{j,i}^{n} \rho_j^{(t)} K_{h_{(t)}}(B_{(t)}^\top X_{ij}) X_{ij}^{(t)} (X_{ij}^{(t)})^\top \Big\}^{-1} \\ &\quad \times \sum_{j,i}^{n} \rho_j^{(t)} K_{h_{(t)}}(B_{(t)}^\top X_{ij}) X_{ij}^{(t)} \{ z_i - a_j^{(t)} - \ell(B_0)^\top X_{ij}^{(t)} \}, \qquad (0.1)\end{aligned}$$

where $X_{ij}^{(t)} = d_j^{(t)} \otimes X_{ij}$ as defined in the algorithm. By the decomposition in Step 3, we obtain the estimate $B_{(t+1)}$ in the next iteration. If the initial value $B_{(0)}$ is a consistent estimator of B_0, then by Lemmas 2 – 6 below, we can establish the following recurring relationship

$$\ell(B_{(t+1)}) - \ell(B_0) = \Theta\{\ell(B_{(t)}) - \ell(B_0)\} + e_{(t)}$$

with $|\Theta| < 1$ and $|e_{(t)}| = o(1)$ almost surely. Therefore, the true parameter $\ell(B_0)$ becomes an attractor of our algorithm. Such a recurring relationship can then be used to prove the convergence of the algorithm and the consistency of the final estimator. To ensure the convergence of the algorithm, we need to consider consistency with probability 1.

For any index set $\mathcal{Z}$ and random matrix $A_n(z)$, we say $A_n(z) = \mathcal{O}(a_n|z \in \mathcal{Z})$, or $A_n(z) = \mathcal{O}(a_n)$ for simplicity, if $\sup_{z\in\mathcal{Z}} |A_n(z)|/a_n \leq c$ almost surely for a universal constant c. Let $\delta_{kh} = \{\log n/(nh^k)\}^{1/2}$, $\delta_n = (\log n/n)^{1/2}$, $r_{kh} = h^2 + \delta_{kh}$, and $X_{ix} = X_i - x$. Let $\mathcal{B} = \{B : B^\top B = I_{d_0}\}$. Let $\mu_{kp} = \int K(v_1, \cdots, v_p) v_1^k dv_1 \cdots dv_p$. For ease of exposition, we further assume that $\mu_{0d_0} = 1$ and $\mu_{2d_0} = 1$. Let $f_B(u)$ be the density function of $B^\top X$, $\nu_B(x) = \mu_B(B^\top x) - x$, $\bar{w}_B(x) = w_B(B^\top x) - \mu_B(B^\top x)\mu_B^\top(B^\top x)$. For simplicity, we further denote $f_B(B^\top x)$ by $f_B(x)$, and $\mu_B(B^\top x)$ by $\mu_B(x)$. For any square matrix A, A^{-1} and A^+ denote, respectively, the inverse (if it exists) and the Moore-Penrose inverse. Let $\mathcal{D}_x$ be any impact set of $\mathbb{R}^p$.

Appendix B.2. Relevant Lemmas

In order to prove the theorems, we first introduce Lemma 1, which is about the uniform convergence of martingales. Its proofs can be found in Xia (2007). We then present another set of lemmas (Lemmas 2 – 6), which are also used for our theorem proof. Similar lemmas can be found in Xia (2007). To save space, the proof of those lemmas are omitted and can be obtained from the authors.

Lemma 1 (Uniform consistency for martingale). *Suppose $G_{n,i}(\chi)$ is a martingale difference sequence with respect to $\mathcal{F}_i = \sigma\{G_{n,\tau}(\chi), \tau \leq i\}$ with $\chi \in \mathcal{X}$ and $\mathcal{X}$ is a compact region in a multidimensional space such that (i) $|G_{n,i}(\chi)| < \xi_i$, where ξ_i are IID and $E\xi_1^{2r} < \infty$ for some $r > 2$; (ii) $EG_{n,k}^2(\chi) < a_n s(\chi)$ with $\inf s(\chi) > 0$, and (iii) $|G_{n,i}(\chi) - G_{n,i}(\tilde{\chi})| < n^{\alpha_1}|\chi - \tilde{\chi}|M_i$, where $M_i, i = 1, 2, ...$ are IID with $EM_1^2 < \infty$. If $a_n = cn^{-\delta}$ with $0 \leq \delta < 1 - 2/r$, then for any $\alpha_1' > 0$ we have*

$$\sup_{|\chi|\leq n^{\alpha_1'}} \left| n^{-1} s^{-1/2}(\chi) \sum_{i=1}^{n} G_{n,i}(\chi) \right| = O\{(n^{-1} a_n \log n)^{1/2}\}$$

almost surely. Suppose for any fixed n and k, $G_{n,i,k}(\theta)$ is a martingale difference

sequence with respect to $\mathcal{F}_{i,k} = \sigma\{G_{n,\tau,k}(\theta), \tau \leq i\}$ *such that (I)* $|G_{n,i,k}(\theta)| \leq \xi_i$*, (II)* $EG^2_{n,i,k}(\theta) < a_n$ *and (III)* $|G_{n,i,k}(\theta) - G_{n,i,k}(\tilde{\theta})| < n^{\alpha_2}|\theta - \tilde{\theta}|M_i$*, where* ξ_i, a_n *and* M_i *are defined above. If* $E|\varepsilon_k|^{2r} < \infty$ *and* $E\{\varepsilon_k | G_{n,i,j}(\theta), i < j, j = 1, ..., k-1\} = 0$*, then*

$$\sup_{\theta\in\Theta} \Big| n^{-2} \sum_{k=2}^{n} \Big\{ \sum_{i=1}^{k-1} G_{n,i,k}(\theta) \Big\} \varepsilon_k \Big| = O\{(a_n \log n)^{1/2}/n\} \text{ almost surely.}$$

Lemma 2 (General kernel smoother). *Let* $M(x) = G(B_0^\top x)$ *and*

$$\begin{pmatrix} a_x \\ d_x h \end{pmatrix} = \Big\{ \sum_{i=1}^{n} K_h(X_{ix}) \begin{pmatrix} 1 \\ X_{ix}/h \end{pmatrix} \begin{pmatrix} 1 \\ X_{ix}/h \end{pmatrix}^\top \Big\}^{-1} \sum_{i=1}^{n} K_h(X_{ix}) \begin{pmatrix} 1 \\ X_{ix}/h \end{pmatrix} z_i.$$

Under assumptions (C1), (C2) and (C4), if $h \propto n^{-\varrho}$ *with* $0 < \varrho < 1/p$*, then we have*

$$a_x = M(x) + \frac{\mu_{2p}}{2\mu_{0p}} \sum_{\kappa=1}^{p} \bigtriangledown^2_{\kappa,\kappa} M(x) h^2 + \mathcal{O}(h^3 + \delta_{ph} | x \in \mathcal{D}_x),$$

$$d_x = \bigtriangledown M(x) + \{\mu_{2p} n h f(x)\}^{-1} \sum_{i=1}^{n} K_h(X_{ix})(X_{ix}/h)\epsilon_i + \mathcal{O}(r_{ph} | x \in \mathcal{D}_x).$$

Lemma 3 (Kernel smoother in SR). *Let*

$$\begin{aligned} \Sigma_n^B(x) &= n^{-1} \sum_{i=1}^{n} K_h(B^\top X_{ix}) \begin{pmatrix} 1 \\ B^\top X_{ix}/h \end{pmatrix} \begin{pmatrix} 1 \\ B^\top X_{ix}/h \end{pmatrix}^\top, \\ \begin{pmatrix} a_x^B \\ d_x^B h \end{pmatrix} &= \{n\Sigma_n^B(x)\}^{-1} \sum_{i=1}^{n} K_h(B^\top X_{ix}) \begin{pmatrix} 1 \\ B^\top X_{ix}/h \end{pmatrix} z_i. \end{aligned}$$

Under assumptions (C1), (C2) and (C4), if $h \propto n^{-\varrho}$ *with* $0 < \varrho < 1/d_0$*, then*

$$\begin{aligned} a_x^B &= G(B_0^\top x) + \bigtriangledown^\top G(B_0^\top x)(B_0 - B)^\top \nu_B(x) + \frac{1}{2} \sum_{\kappa=1}^{d_0} \bigtriangledown^2_{\kappa,\kappa} G(B_0^\top x) h^2 \\ &\quad + \mathcal{V}_{1n}^B(x) + \mathcal{O}(\Delta_n^B | x \in \mathcal{D}_x, B \in \mathcal{B}), \\ d_x^B h &= \bigtriangledown G(B_0^\top x) h + Q_1^B(x) h^3 + \mathcal{V}_{2n}^B(x) + \mathcal{O}(\Delta_n^B | x \in \mathcal{D}_x, B \in \mathcal{B}), \end{aligned}$$

where $\Delta_n^B = h^4 + \delta_{d_0 h}^2 + h\delta_B + \delta_B^2$ *with* $\delta_B = |B - B_0|$,

$$Q_1^B(x) = \frac{1}{2} f_B^{-1}(x) \bigtriangledown^2 G(B_0^\top x) \bigtriangledown f_B(x) + \frac{1}{6} \mu_{4d_0} \{\bigtriangledown_{1,1,1}^3 G(B_0^\top x), ..., \bigtriangledown_{d_0,d_0,d_0}^3 G(B_0^\top x)\}^\top,$$

$$\mathcal{V}_{1n}^B(x) = \mathcal{E}_{n,1}^B(x) - h \bigtriangledown^\top f_B(x) \mathcal{E}_{n,2}^B(x), \quad \mathcal{V}_{2n}^B(x) = \mathcal{E}_{n,2}^B(x) - h \bigtriangledown f_B(x) \mathcal{E}_{n,1}^B(x),$$

$$\mathcal{E}_{n,1}^B(x) = \{n f_B(x)\}^{-1} \sum_{i=1}^n K_h(B^\top X_{ix}) \epsilon_i,$$

$$\mathcal{E}_{n,2}^B(x) = \{n f_B(x)\}^{-1} \sum_{i=1}^n K_h(B^\top X_{ix})(B^\top X_{ix}/h) \epsilon_i.$$

Lemma 4 (Denominator of SR). *Let* $\hat{f}_B(x) = n^{-1} \sum_{i=1}^n K_h(B^\top X_{ix})$, $\hat{\rho}_j^B = \rho(\hat{f}_B(X_j))/\hat{f}_B(X_j)$ *and* $X_{ij}^B = d_j^B \otimes X_{ij}$ *where* $d_j^B = d_{X_j}^B$. *Suppose (C1)–(C4) hold and* $h \propto n^{-\varrho}$ *with* $0 < \varrho < 1/d_0$ *and* $\delta_B/h \to 0$. *We have*

$$\begin{aligned} \Big\{ n^{-2} \sum_{j,i=1}^n \hat{\rho}_j^B K_h(B^\top X_{ij}) X_{ij}^B X_{ij}^{B^\top} \Big\}^{-1} &= (I_{d_0} \otimes B_0) M_0^{-1} (I_{d_0} \otimes B_0^\top) h^{-2} \\ &\quad + (I_{d_0} \otimes B_0) L_0 + L_0^\top (I_{d_0} \otimes B_0^\top) \\ &\quad + \frac{1}{2} \tilde{D}_0^+ + \mathcal{O}\{(\tilde{r}_{d_0 h} + \delta_B)/h | B \in \mathcal{B}\}, \end{aligned}$$

where $M_0 = E\{\rho(f_{B_0}(X_i)) \bigtriangledown G(B_0^\top X_i) \bigtriangledown^\top G(B_0^\top X_i)\}$, $\tilde{r}_{d_0 h} = h^2 + \delta_{d_0 h} + \delta_{d_0 h}^2/h^2$ *and* L_0 *is a constant matrix (details can be found in the proof) and*

$$\tilde{D}_0 = E\Big\{ \rho(f_{B_0}(X_i)) \bigtriangledown G(B_0^\top X_i) \bigtriangledown^\top G(B_0^\top X_i) \otimes \bar{w}_{B_0}(X_i) \Big\}.$$

Lemma 5 (Numerator of SR). *Suppose conditions (C1)–(C4) hold. If* $h_0 \propto n^{-\varrho}$

with $0 < \varrho < 1/d_0$ *and* $\delta_B/h \to 0$, *then*

$$n^{-2}\sum_{j,i=1}^{n}\hat{\rho}_j^B K_h(B^\top X_{ij}) d_j^B \otimes X_{ij}\{z_i - a_j^B - \ell(B_0)^\top X_{ij}^B\} \;=\; D_0\ell(B-B_0) + \Phi_n + \mathcal{O}\{\tilde{\Delta}_n^B | B \in \mathcal{B}\},$$

where $\tilde{\Delta}_n^B = h^4 + \delta_{d_0 h}^2 + \delta_B r_{d_0 h}/h + \delta_B^2 + \delta_n h$, $\Phi_n = n^{-1}\sum_{i=1}^n \rho(f_{B_0}(X_i)) \bigtriangledown G(B_0^\top X_i) \otimes \nu_{B_0}(X_i)\epsilon_i$ *and* $D_0 = E\Big\{\rho(f_{B_0}(X_i))\bigtriangledown G(B_0^\top X_i) \otimes \nu_{B_0}(X_i)[\bigtriangledown G(B_0^\top X_i) \otimes \nu_{B_0}(X_i)]^\top\Big\}$.

Lemma 6 (Initial estimator). *Suppose assumptions (C1)-(C4) hold. With bandwidth* $h \propto n^{-1/(p+4)}$, *the initial estimator* $B_{(0)}$ *is consistent estimator of a basis in* $\mathcal{S}(B_0)$, *i.e. there exists a rotation matrix* $Q : Q^\top Q = I_{d_0}$ *such that* $|B_{(0)} - B_0 Q| = O(h^2 + \delta_{ph} + \delta_{ph}^2/h^2)$ *almost surely.*

Appendix B.3. Proof of Theorem 1

Theorem 1. *Suppose conditions (C1)-(C5) in the appendix hold,* $d = d_0$, *and the final bandwidth is* $\hbar$, *then the SR estimator* $\hat{B}$ *is consistent with*

$$|\hat{B}\hat{B}^\top - B_0 B_0^\top| = O_p\{\hbar^4 + \log n/(n\hbar^{d_0}) + n^{-1/2}\}.$$

Proof: Without loss of generality, we assume that $Q = I_d$. Otherwise, we can simply define a new basis as $B_0 := B_0 Q$. Let $\delta_{(t)} = \delta_{B_{(t)}}$ denote the estimation error in the $t-$th iteration. By (0.1), Lemmas 4 and 5 and the facts that $(I_{d_0} \otimes B_0^\top)D_0 = 0$ and

$(I_{d_0} \otimes B_0^\top)\Phi_n = 0$, if $\delta_{(t)} \log n / h_{(t)} = o(1)$ and $\delta_n / h_{(t)}^2 = o(1)$ we have

$$\begin{aligned}
\Gamma^{(t+1)} &= \ell(B_0) + \tilde{D}_0^+ D_0 \ell(B_{(t)} - B_0) + \tilde{D}_0^+ \Phi_n + (I_{d_0} \otimes B_0) L_0 \mathcal{O}(c_n^{(t)}) \\
&\quad + \mathcal{O}\{\tilde{\Delta}_n^{B_{(t)}} + (\delta_{(t)} + \delta_n)(r_{d_0 h_{(t)}} + \delta_{(t)})/h_{(t)}\} \\
&= (I_{d_0} \otimes B_0)\{\ell(I_{d_0}) + \mathcal{O}(c_n^{(t)})\} + \tilde{D}_0^+ D_0 \ell(B_{(t)} - B_0) + \tilde{D}_0^+ \Phi_n \\
&\quad + \mathcal{O}\{\tilde{\Delta}_n^{B_{(t)}} + (\delta_{(t)} + \delta_n)(r_{d_0 h_{(t)}} + \delta_{(t)})/h_{(t)}\}, \qquad (0.2)
\end{aligned}$$

where $c_n^{(t)} = \tilde{\Delta}_n^{B_{(t)}}/h_{(t)}^2 + \delta_{(t)} + \delta_n$. Since $\delta_{d_0 h_{(t)}}/h_{(t)} = o(1)$, we have $\mathcal{M}(\Gamma^{(t+1)}) = B_0 \Lambda_n^{(t)} + \mathcal{O}\{\delta_n + \delta_{(t)} + \tilde{\Delta}_n^{B_{(t)}}\}$, where $\Lambda_n^{(t)} = I_{d_0} + \mathcal{O}(c_n^{(t)})$ and $\mathcal{M}(.)$ is defined in section 2.3. Note that $\tilde{\Lambda}_n^{(t+1)} = \{\mathcal{M}(\Gamma^{(t+1)})\}^\top \mathcal{M}(\Gamma^{(t+1)}) = (\Lambda_n^{(t)})^2 + \mathcal{O}\{\delta_n + \delta_{(t)} + \tilde{\Delta}_n^{B_{(t)}}\}$. If $c_n^{(t)} = o(1)$ almost surely, then by Step 3, $B_{(t+1)} = \mathcal{M}(\Gamma^{(t+1)})\{\tilde{\Lambda}_n^{(t+1)}\}^{-1} = B_0 + \mathcal{O}\{\delta_n + \delta_{(t)} + \tilde{\Delta}_n^{B_{(t)}}\}$. It follows that $\ell(B_{(t+1)}) = \ell(B_0) + \tilde{D}_0^+ D_0 \ell(B_{(t)} - B_0) + \tilde{D}_0^+ \Phi_n + \mathcal{O}\{\tilde{c}_n^{(t)}(\delta_{(t)} + \delta_n) + \tilde{\Delta}_n^{B_{(t)}}\}$, where $\tilde{c}_n^{(t)} = c_n^{(t)} + (r_{d_0 h_{(t)}} + \delta_{(t)})/h_{(t)}$. Thus

$$\ell(B_{(t+1)} - B_0) = \tilde{D}_0^+ D_0 \ell(B_{(t)} - B_0) + \tilde{D}_0^+ \Phi_n + \mathcal{O}\{\tilde{c}_n^{(t)}(\delta_{(t)} + \delta_n) + \tilde{\Delta}_n^{B_{(t)}}\}. \quad (0.3)$$

Recall that $\tilde{D}_0 = \{I_{d_0} \otimes \tilde{B}_0\}\tilde{V}_0^{-1}\{I_{d_0} \otimes \tilde{B}_0\}^\top$, $D_0 = \{I_{d_0} \otimes \tilde{B}_0\}V_0\{I_{d_0} \otimes \tilde{B}_0\}^\top$, where $\tilde{V}_0 = 2E\{\rho(f_{B_0}(X))\bigtriangledown G(B_0^\top X)\bigtriangledown^\top G(B_0^\top X) \otimes \tilde{B}_0^\top \tilde{w}_{B_0}(X)\tilde{B}_0\}$, and $V_0 = E\Big\{\rho(f_{B_0}(X))\bigtriangledown G(B_0^\top X)\ \bigtriangledown^\top G(B_0^\top X) \otimes \tilde{B}_0^\top \tilde{w}_{B_0}(X)\tilde{B}_0\Big\}$. Therefore, $\tilde{D}_0^+ D_0 = \tilde{B}_0 \tilde{B}_0^\top/2$. By (0.3),

$$\delta_{(t+1)} = \frac{1}{2}\delta_{(t)} + |\tilde{D}_0^+ \Phi_n| + \mathcal{O}\{\tilde{c}_n^{(t)}(\delta_{(t)} + \delta_n) + \tilde{\Delta}_n^{B_{(t)}}\} = \{\frac{1}{2} + \bar{c}_n^{(t)}\}\delta_{(t)} + |\tilde{D}_0^+ \Phi_n| + R_t, \quad (0.4)$$

where $|\bar{c}_n^{(t)}| \le c_0\{h_{(t)} + \delta_{d_0 h_{(t)}}/h_{(t)} + \delta_n/h_{(t)} + \delta_{(t)} r_{d_0 h_{(t)}}/h_{(t)}^3 + \delta_{(t)}/h_{(t)}\} < c_0(h_{(0)} + \delta_{d_0 \hbar}/\hbar + (\log n)^{1/2}\delta_{(t)}/h_{(t)})$ and $R_t = h_{(t)}^4 + \delta_{d_0 h_{(t)}}^2 + h_{(t)}\delta_n$. Note that (0.4) holds providing $\delta_{(t)} \log(n)\ /h_{(t)} \to 0$, which is true for $t = 0$ by Lemma 6. Furthermore, by (C5), if n

is sufficiently large, we have $c_n \stackrel{def}{=} \max_t\{\frac{1}{2}+\bar{c}_n^{(t)}\}\varsigma^{-1} < 1$. Thus

$$\begin{aligned}\delta_{(t+1)}\log n/h_{(t+1)} &= c_n\delta_{(t)}\log n/h_{(t)} + \{|\tilde{D}_0^+\Phi_n| + R_t\}\log n/h_{(t+1)} \\ &< c_n\delta_{(t)}\log n/h_{(t)} + \delta_n\log n/\hbar + h_{(0)}^3\log n + (n\hbar^{d_0+1})^{-1},\end{aligned}$$

implying that $\delta_{(t+1)}\log n/h_{(t+1)} \to 0$ for all t as n is large enough. Therefore, (0.4) applies to all t. As a consequence, we have from (0.4) that

$$\begin{aligned}\delta_{(t+1)} &< c_n^t\delta_{(0)} + \sum_{\tau=1}^{t} c_n^\tau|\tilde{D}_0^+\Phi_n| + \sum_{\tau=1}^{t} c_n^{t-\tau}R_\tau \\ &\to \frac{1}{1-c_n}|\tilde{D}_0^+\Phi_n| + \tilde{R}_n^*, \text{ as } t\to\infty,\end{aligned}$$

where $\tilde{R}_n^* = \mathcal{O}(\hbar^4+\delta_{d_0\hbar}^2+\hbar\delta_n)$. This means the algorithm converges. By the definition of the estimator and that $\Phi_n = O_P(n^{-1/2})$, it follows that $|\hat{B}-B_0| = O_P(\hbar^4+\delta_{d_0\hbar}^2+n^{-1/2})$. This completes the proof of Theorem 1.

Appendix B.4. Proof of Theorem 2

Theorem 2. *Suppose conditions (C1)-(C5) in the appendix hold. Moreover, the bandwidth $\hbar_d$ used for different dimension d satisfies $\hbar_d \propto n^{-1/(d+4)}$. Then, we have*

$$P(\hat{d} = d_0) \to 1 \text{ as } n\to\infty.$$

Proof: Let $B \in \mathbb{R}^{p\times d}$ with $B^\top B = I_d$. Define $\hat{f}_{B,j}(x) = n^{-1}\sum_{\substack{i=1\\i\neq j}}^n K_{\hbar_d}(B^\top X_{ij})$, $\hat{a}_{B,j}(x) = \{n\hat{f}_{B,j}(x)\}^{-1}\sum_{\substack{i=1\\i\neq j}}^n K_{\hbar_d}(B^\top X_{ij})z_i$, and $CV(B) = n^{-1}\sum_{j=1}^n w(X_j)\{z_j-\hat{a}_{B,j}(X_j)\}^2$, where $w(.)$ is a trimming function. We discuss two cases separately.

Case 1: $d < d_0$. It is easy to see that $|BB^\top - B_0B_0^\top| = 1$. By (C3), following the proof of Yao and Tong (1994) it is easy to see that there is a constant $\delta > 0$ such that

$\lim_{n\to\infty} P\{CV(B) > CV(B_0) + \delta\} = 1$ for all $B \in \mathbb{R}^{p\times d}$ with $B^\top B = I_d$. Note that by Theorem 1, with working dimension d_0 we have $\hat{B}_{d_0} \to B_0$. Thus, $CV(\hat{B}_{d_0}) \to CV(B_0)$ in probability. Hence for the estimator $\hat{B}_d \in \mathbb{R}^{p\times d}$ with working dimension d, we have

$$\lim_{n\to\infty} P\{CV(\hat{B}_d) > CV(\hat{B}_{d_0})\} = 1 \quad \text{if} \quad d < d_0. \tag{0.5}$$

Case 2: $d > d_0$. Write the model as $z_i = E\{G(B_0^\top X_i)|B^\top X_i\} + \xi_i$, where $\xi_i = \epsilon_i + G(B_0^\top X_i) - E\{G(B_0^\top X_i)|B^\top X_i\}$. By Theorem 2 of Cheng and Tong (1993), we have $CV(B) = n^{-1}\sum_{i=1}^n w(X_j)\xi_i^2 + c/(n\hbar_d^d)\{1 + o_P(1)\}$, where $c > 0$ is a constant.

$$\begin{aligned} n^{-1}\sum_{i=1}^n w(X_i)\xi_i^2 = & n^{-1}\sum_{i=1}^n w(X_i)\epsilon_i^2 + n^{-1}\sum_{i=1}^n w(X_i)[G(B_0^\top X_i) - E\{G(B_0^\top X_i)|B^\top X_i\}]^2 \\ & + n^{-1}\sum_{i=1}^n w(X_i)[G(B_0^\top X_i) - E\{G(B_0^\top X_i)|B^\top X_i\}]\epsilon_i. \end{aligned}$$

Let $c_B^2 = E[w(X_i)\{G(B_0^\top X_i) - E(G(B_0^\top X_i)|B^\top X_i)\}]^2$. By Lemma 1, we have the second term on the right hand side above is $c_B^2\{1+o_P(1)\}$ and the third term $O_P(c_B\delta_n)$. If $c_B = O(\delta_n)$, then $n^{-1}\sum_{i=1}^n w(X_i)\xi_i^2 = n^{-1}\sum_{i=1}^n w(X_i)\epsilon_i^2 + O_P(\delta_n^2)$. If $c_B \gg \delta_n$, then $n^{-1}\sum_{i=1}^n w(X_i)\xi_i^2 = n^{-1}\sum_{i=1}^n w(X_i)\epsilon_i^2 + c_B^2\{1 + o_P(1)\}$. Note that $\delta_n^2 = o\{(n\hbar_d^d)^{-1}\}$. Generally, we have $CV(B) = n^{-1}\sum_{i=1}^n w(X_i)\epsilon_i^2 + \max\{c_B^2,\ c/(n\hbar_d^d)\}\{1 + o_P(1)\}$. On the other hand, by Theorem 2 of Cheng and Tong (1993) again, we have

$$CV(B_0) = n^{-1}\sum_{i=1}^n w(X_i)\epsilon_i^2 + c_1/(n\hbar_{d_0}^{d_0})\{1 + o_P(1)\},$$

where $c_1 > 0$ is a constant. By (C5), we have $c_1/(n\hbar_{d_0}^{d_0}) = o\{c/(n\hbar_d^d)\}$. Thus

$$CV(B) - CV(B_0) = c_2/(n\hbar_d^d)\{1 + o_P(1)\}, \tag{0.6}$$

where $c_2 > 0$ is a constant independent of B and $o_p(1)$ is uniformly small over $\{B : B^\top B = I_d$. Suppose we can show that

$$CV(\hat{B}_{d_0}) - CV(B_0) = o_P\{(n\hbar_{d_0}^{d_0})^{-1}\}. \tag{0.7}$$

Then, from (0.6) and (0.7) we have for estimator $\hat{B}_d \in \mathbb{R}^{p\times d}$

$$\lim_{n\to\infty} P\{CV(\hat{B}_d) > CV(\hat{B}_{d_0})\} = 1 \quad \text{for all } d > d_0. \tag{0.8}$$

Therefore, Theorem 2 follows from (0.5) and (0.8).

Appendix B.5. Proof of (0.7)

To complete the proof, we need only to prove (0.7). Let $B \in \mathbb{R}_0^{p\times d_0}$ with $|BB^\top - B_0B_0^\top| \le \delta_B'$. By Theorem 1, we only need to consider $\delta_B' = o(\hbar_{d_0}^2)$. Write

$$\begin{aligned} CV(B) &= n^{-1}\sum_{j=1}^n w(X_j)\{z_j - \hat{a}_{B_0,j}(X_j) + \hat{a}_{B_0,j}(X_j) - \hat{a}_{B,j}(X_j)\}^2 \\ &= CV(B_0) + n^{-1}\sum_{j=1}^n w(X_j)\{\hat{a}_{B,j}(X_j) - \hat{a}_{B_0,j}(X_j)\}^2 \\ &\quad +2n^{-1}\sum_{j=1}^n w(X_j)\{z_j - \hat{a}_{B_0,j}(X_j)\}\{\hat{a}_{B,j}(X_j) - \hat{a}_{B_0,j}(X_j)\}. \end{aligned} \tag{0.9}$$

By (C1) and (C2), we have $E\{K_{\hbar_{d_0}}(B^\top X_{ix})z_i - K_{\hbar_{d_0}}(B_0^\top X_{ix})z_i\} = o(\hbar_{d_0}^2)$, and thus

$$E\{n^{-1}\sum_{\substack{i=1\\ i\neq j}}^n K_{\hbar_{d_0}}(B^\top X_{ix})z_i - n^{-1}\sum_{\substack{i=1\\ i\neq j}}^n K_{\hbar_{d_0}}(B_0^\top X_{ix})z_i\} = o(\hbar_{d_0}^2)$$

uniformly for B and x. Similarly, $E[K_{\hbar_{d_0}}(B^\top X_{ix})z_i - K_{\hbar_{d_0}}(B_0^\top X_{ix})z_i]^2 = o(\hbar_{d_0}^2)$. By

Lemma 1 and $\delta_n = o(\hbar_d^2)$, it follows that

$$n^{-1}\sum_{\substack{i=1\\ i\neq j}}^{n} K_{\hbar_{d_0}}(B^\top X_{ix})z_i - n^{-1}\sum_{\substack{i=1\\ i\neq j}}^{n} K_{\hbar_{d_0}}(B_0^\top X_{ix})z_i = o_P(\hbar_{d_0}^2 + \hbar_{d_0}\delta_n) = o_P(\hbar_{d_0}^2).$$

Similarly, $n^{-1}\sum_{\substack{i=1\\ i\neq j}}^{n} K_{\hbar_{d_0}}(B^\top X_{ix}) - n^{-1}\sum_{\substack{i=1\\ i\neq j}}^{n} K_{\hbar_{d_0}}(B_0^\top X_{ix}) = o_P(\hbar_{d_0}^2)$. It follows that $\hat{a}_{B,j}(X_j) - \hat{a}_{B_0,j}(X_j) = o_P(\hbar_{d_0}^2)$ and that

$$n^{-1}\sum_{j=1}^{n} w(X_j)\{\hat{a}_{B,j}(X_j) - \hat{a}_{B_0,j}(X_j)\}^2 = o_P(\hbar_{d_0}^4). \tag{0.10}$$

Note that $G(B_0^\top X_j) - \hat{a}_{B_0,j}(X_j) = O_P(r_{d_0}\hbar_{d_0})$. We have

$$\begin{aligned}
&n^{-1}\sum_{j=1}^{n} w(X_j)\{z_j - \hat{a}_{B_0,j}(X_j)\}\{\hat{a}_{B,j}(X_j) - \hat{a}_{B_0,j}(X_j)\} \\
&= n^{-1}\sum_{j=1}^{n} w(X_j)\{G(B_0^\top X_j) - \hat{a}_{B_0,j}(X_j)\}\{\hat{a}_{B,j}(X_j) - \hat{a}_{B_0,j}(X_j)\} \\
&\qquad\qquad + n^{-1}\sum_{j=1}^{n} w(X_j)\epsilon_j\{\hat{a}_{B,j}(X_j) - \hat{a}_{B_0,j}(X_j)\} \\
&= o_P(\hbar_{d_0}^4) + n^{-1}\sum_{j=1}^{n} w(X_j)\epsilon_j\{\hat{a}_{B,j}(X_j) - \hat{a}_{B_0,j}(X_j)\}.
\end{aligned} \tag{0.11}$$

Write

$$\begin{aligned}
\hat{a}_{B,j}(X_j) - \hat{a}_{B_0,j}(X_j) = &\{n\hat{f}_{B,j}(X_j)\hat{f}_{B_0,j}(X_j)\}^{-1}\sum_{i\neq j} K_{\hbar_{d_0}}(B^\top X_{ij})z_i\{\hat{f}_{B,j}(X_j) - \hat{f}_{B_0,j}(X_j)\} \\
&+\{n\hat{f}_{B,j}(X_j)\}^{-1}\sum_{i\neq j}\{K_{\hbar_{d_0}}(B_0^\top X_{ij}) - K_{\hbar_{d_0}}(B^\top X_{ij})\}z_i.
\end{aligned}$$

Next, we consider the second term on the right hand side above only. Similar idea and result can also be applied to the first term. By the fact that $\hat{f}_{B,j}(X_j) = f_{B,j}(X_j) +$

$O_P(r_{d_0}\hbar_{d_0})$ and that $\hat{a}_{B,j}(X_j) - \hat{a}_{B_0,j}(X_j) = o_P(\hbar_{d_0}^2)$, we have

$$
\begin{aligned}
& n^{-1}\sum_{j=1}^{n} w(X_j)\epsilon_j\{n\hat{f}_{B,j}(X_j)\}^{-1}\sum_{i\neq j}\{K_{\hbar_{d_0}}(B_0^\top X_{ij}) - K_{\hbar_{d_0}}(B^\top X_{ij})\}z_i \\
&= n^{-1}\sum_{j=1}^{n} w(X_j)\epsilon_j\{nf_B(X_j)\}^{-1}\sum_{i\neq j}\{K_{\hbar_{d_0}}(B_0^\top X_{ij}) - K_{\hbar_{d_0}}(B^\top X_{ij})\}z_i + o_P(\hbar_{d_0}^4) \\
&= n^{-1}\sum_{j=1}^{n} w(X_j)\epsilon_j\{nf_B(X_j)\}^{-1}\sum_{i\neq j}\{K_{\hbar_{d_0}}(B_0^\top X_{ij}) - K_{\hbar_{d_0}}(B^\top X_{ij})\}G(B_0^\top X_i) \\
&\quad + n^{-1}\sum_{j=1}^{n} w(X_j)\epsilon_j\{nf_B(X_j)\}^{-1}\sum_{i\neq j}\{K_{\hbar_{d_0}}(B_0^\top X_{ij}) - K_{\hbar_{d_0}}(B^\top X_{ij})\}\epsilon_i + o_P(\hbar_{d_0}^4).
\end{aligned}
$$

Applying Lemma 1, we have

$$
n^{-1}\sum_{j=1}^{n} w(X_j)\epsilon_j\{nf_B(X_j)\}^{-1}\sum_{i\neq j}\{K_{\hbar_{d_0}}(B_0^\top X_{ij}) - K_{\hbar_{d_0}}(B^\top X_{ij})\}G(B_0^\top X_i) = o_P(\hbar_{d_0}^4),
$$

$$
n^{-1}\sum_{j=1}^{n} w(X_j)\epsilon_j\{nf_B(X_j)\}^{-1}\sum_{i\neq j}\{K_{\hbar_{d_0}}(B_0^\top X_{ij}) - K_{\hbar_{d_0}}(B^\top X_{ij})\}\epsilon_i = o_P(\hbar_{d_0}^4).
$$

Therefore $\sup_B n^{-1}\sum_{j=1}^{n} w(X_j)\epsilon_j\{\hat{a}_{B,j}(X_j) - \hat{a}_{B_0,j}(X_j)\} = o_P(\hbar_{d_0}^4)$. Equation (0.7) follows from (0.9), (0.10) and (0.11) by noting that $\hbar_{d_0}^4 = (n\hbar_{d_0}^{d_0})^{-1}$ if $\hbar_{d_0} \propto n^{-1/(d_0+4)}$.

REFERENCES

Cook, R. D. (1998b), *Regression Graphics*, John Wiley, New York, NY.

Härdle, W., Hall, P. and Ichimura, H. (1993) "Optimal smoothing in single-index models", *The Annals of Statistics*, 21, 157–178.

Yao, Q. and Tong, H. (1994), "On subset selection in non-parametric stochastic regression," *Statistica Sinica*, 4, 51–70.