Supplementary material to: Estimation of a groupwise additive multiple-index model and its applications

Tao Wang, Jun Zhang, Hua Liang and Lixing Zhu

Yale University, Shenzhen University, George Washington University

and Hong Kong Baptist University

Supplementary Material

This supplementary file covers the regularity conditions and proofs.

S1 Regularity conditions

Technical Conditions for Theorem 1. Let $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ be the smallest and largest eigenvalues of any square matrix $\mathbf{A}$, respectively. Let $\boldsymbol{\Lambda} = \mathrm{Cov}\{(Y - \boldsymbol{\beta}_{LS}^{\top}\boldsymbol{V})\boldsymbol{V}\}$ and $\mathcal{I}(\boldsymbol{\eta}, C) = \{i : 1 \le i \le n, |\boldsymbol{v}_i^{\top}\boldsymbol{\eta}| \le C\}$. To study the asymptotical behavior of $\hat{\boldsymbol{\beta}}_{LS}$, the following regularity conditions are needed:

(A1) there are some constants $a > 0$ and $C_1 > 0$ such that, for all $\boldsymbol{\eta}$ with $\|\boldsymbol{\eta}\|_2 = 1$,

$$\lim_{n\to\infty} P\left(\sum_{i\in\mathcal{I}(\boldsymbol{\eta},C_1)} (\boldsymbol{v}_i^{\top}\boldsymbol{\eta})^2 > an\right) = 1.$$

(A2) for any $\varepsilon > 0$, there exists a constant $C_2 > 0$ such that, for all $\boldsymbol{\eta}$ with $\|\boldsymbol{\eta}\|_2 = 1$,

$$\lim_{n\to\infty} P\left(\sum_{i\notin\mathcal{I}(\boldsymbol{\eta},C_2)} (\boldsymbol{v}_i^{\top}\boldsymbol{\eta})^2 \le \varepsilon n\right) = 1.$$

(A3) there is a constant $C_3 > 0$ such that $\lim_{n\to\infty} P(\max_{i=1,\dots,n} \|\boldsymbol{v}_i\|_2^2 \le C_3 n^2) = 1$.

(A4) all fourth moments of the predictors $\boldsymbol{V}$ are bounded above by some constant $C_4 > 0$.

(A5) $0 < L_1 \le \lambda_{\min}(\boldsymbol{\Sigma}_{\boldsymbol{V}}) \le \lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{V}}) \le L_2 < \infty$ for some constants L_1 and L_2.

(A6) $\mathrm{var}(Y - \boldsymbol{\beta}_{LS}^{\top}\boldsymbol{V}) < \infty$.

(A7) $0 < L_3 \leq \lambda_{\min}(\mathbf{\Lambda}) \leq \lambda_{\max}(\mathbf{\Lambda}) \leq L_4 < \infty$ for some constant L_3 and L_4.

Technical Conditions for Theorem 3. Let $\mathbf{\Sigma}_1 = \text{Cov}\{(\boldsymbol{V}_{11}^\top, \ldots, \boldsymbol{V}_{K1}^\top)^\top\}$, where $\boldsymbol{V}_{k1}$ consists of the first p_{k0} components of $\boldsymbol{V}_k$. Let $d_0 = p_{10} + \cdots + p_{K0}$. The following technical conditions are needed. There are $0 \leq c_1 < 1/2, c_1 < c_2 \leq 1, B_1, B_2, B_3, B_4, B_5 > 0$ such that, for some $k > 0$ and $\delta > 0$,

(D1) $0 < B_1 \leq \lambda_{\min}(\mathbf{\Sigma}_1) \leq \lambda_{\max}(\mathbf{\Sigma}_1) \leq B_2 < \infty$,

(D2) $d_0 = O(n^{c_1})$ and $n^{(1-c_2)/2} \min_{j:\beta_j \neq 0} |\beta_j| \geq B_3$,

(D3) $E(Y - \boldsymbol{\beta}_{11}^\top \boldsymbol{V}_{11} - \cdots - \boldsymbol{\beta}_{K1}^\top \boldsymbol{V}_{K1})^{2k} \leq B_4$,

(D4) all m-th moments of the predictors $\boldsymbol{V}$, with $m = 4(c_2 - c_1)k + 4 + \delta$, are bounded above by B_5,

(D5) $n^{1/2}\lambda/\log n \to \infty, \lambda = o(n^{-(1-c_2+c_1)/2}), d = o(n^{(c_2-c_1)k})$ and $d = o(\lambda^{2k} n^k)$.

S2 Proofs

Proof of Proposition 1. Without loss of generality, we assume that $E(\boldsymbol{V}) = \boldsymbol{0}_{d\times 1}$. We can use the basic properties of conditional expectation to obtain

$$\begin{aligned} \mathbf{\Sigma}_{\boldsymbol{V}}^{-1} E(\boldsymbol{V}Y) &= \mathbf{\Sigma}_{\boldsymbol{V}}^{-1} E[E\{E(\boldsymbol{V}|\mathbf{S}^\top \boldsymbol{V}, Y)|Y\}Y] \\ &= \mathbf{\Sigma}_{\boldsymbol{V}}^{-1} E[E\{E(\boldsymbol{V}|\mathbf{S}^\top \boldsymbol{V})|Y\}Y] \\ &= \mathbf{\Sigma}_{\boldsymbol{V}}^{-1} E[E\{\mathbf{\Sigma}_{\boldsymbol{V}}\mathbf{S}(\mathbf{S}^\top \mathbf{\Sigma}_{\boldsymbol{V}}\mathbf{S})^{-1}\mathbf{S}^\top \boldsymbol{V}|Y\}Y] \\ &= \mathbf{S}(\mathbf{S}^\top \mathbf{\Sigma}_{\boldsymbol{V}}\mathbf{S})^{-1}\mathbf{S}^\top E(\boldsymbol{V}Y), \end{aligned}$$

where the second equality is valid since Y is independent of $\boldsymbol{V}$ conditioned on $\mathbf{S}^\top \boldsymbol{V}$, and the third quality follows from the linearity condition. Set

$$\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)^\top = (\mathbf{S}^\top \mathbf{\Sigma}_{\boldsymbol{V}}\mathbf{S})^{-1}\mathbf{S}^\top E(\boldsymbol{V}Y).$$

Then, $\boldsymbol{\beta}_{LS} = \mathbf{\Sigma}_{\boldsymbol{V}}^{-1}\text{Cov}(\boldsymbol{V}, Y) = (\phi_1\boldsymbol{\beta}_1^\top, \phi_2\boldsymbol{\beta}_2^\top, \ldots, \phi_K\boldsymbol{\beta}_K^\top)^\top$. The proof is complete.

Proof of Theorem 1. The proof of this theorem can be found in Wu and Li (2011); see also Portnoy (1984).

Lemma 1. *Under the conditions of Proposition 2,* $\boldsymbol{\phi} \neq \boldsymbol{0}_{2\times 1}$.

Proof of Lemma 1. We prove it by contradiction. Suppose that $\boldsymbol{\phi} = \boldsymbol{0}_{2\times 1}$. It then follows that $E(\boldsymbol{X}Y) = \boldsymbol{0}_{p\times 1}$ and $E(\boldsymbol{Z}Y) = \boldsymbol{0}_{q\times 1}$, that is,

$$\mathbf{\Sigma}_{\boldsymbol{X}}\boldsymbol{\alpha} + E\{\boldsymbol{X}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \boldsymbol{0}_{p\times 1}, \quad (S2.1)$$

$$\mathbf{\Sigma}_{\boldsymbol{Z}\boldsymbol{X}}\boldsymbol{\alpha} + E\{\boldsymbol{Z}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \boldsymbol{0}_{q\times 1}. \quad (S2.2)$$

Under (B1), $E\{\boldsymbol{X}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \boldsymbol{0}_{p\times 1}$, and so by (S2.1) $\boldsymbol{\alpha} = \boldsymbol{0}_{p\times 1}$. Under (B2),

$$E\{\boldsymbol{X}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = E\{E(\boldsymbol{X}|\boldsymbol{Z})g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \boldsymbol{\Sigma}_{\boldsymbol{XZ}}\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}E\{\boldsymbol{Z}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\}. \tag{S2.3}$$

From (S2.1)-(S2.3), we have $(\boldsymbol{\Sigma}_{\boldsymbol{X}} - \boldsymbol{\Sigma}_{\boldsymbol{XZ}}\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{ZX}})\boldsymbol{\alpha} = 0$. Because $\boldsymbol{\Sigma}_{\boldsymbol{X}} - \boldsymbol{\Sigma}_{\boldsymbol{XZ}}\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{ZX}}$ is positive definite, $\boldsymbol{\alpha} = \boldsymbol{0}_{p\times 1}$.

Therefore, under (B0), if either (B1) or (B2) holds, there is a contradiction. The proof is complete.

Proof of Proposition 2. We prove the statement by contradiction. By Lemma 1, there are two possibilities: $(\phi_1 \neq 0, \phi_2 = 0)$ and $(\phi_1 = 0, \phi_2 \neq 0)$. Since $\boldsymbol{\beta}_{LS} = \mathbf{S}\boldsymbol{\phi}$, we have

$$\begin{aligned} E(\boldsymbol{X}Y) &= \phi_1\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\alpha} + \phi_2\boldsymbol{\Sigma}_{\boldsymbol{XZ}}\boldsymbol{\gamma}, \\ E(\boldsymbol{Z}Y) &= \phi_1\boldsymbol{\Sigma}_{\boldsymbol{ZX}}\boldsymbol{\alpha} + \phi_2\boldsymbol{\Sigma}_{\boldsymbol{Z}}\boldsymbol{\gamma}. \end{aligned}$$

If $\phi_1 \neq 0$ and $\phi_2 = 0$, then

$$\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\alpha} + E\{\boldsymbol{X}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \phi_1\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\alpha}, \tag{S2.4}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{ZX}}\boldsymbol{\alpha} + E\{\boldsymbol{Z}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \phi_1\boldsymbol{\Sigma}_{\boldsymbol{ZX}}\boldsymbol{\alpha}. \tag{S2.5}$$

By (S2.4) and (S2.5),

$$E\{\boldsymbol{Z}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \boldsymbol{\Sigma}_{\boldsymbol{ZX}}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}E\{\boldsymbol{X}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\}. \tag{S2.6}$$

Under (B1), $\boldsymbol{\Sigma}_{\boldsymbol{ZX}} = \boldsymbol{0}_{q\times 1}$, so by (S2.5) $E\{\boldsymbol{Z}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \boldsymbol{0}_{q\times 1}$. Under (B2), by (S2.3) and (S2.6),

$$(\boldsymbol{\Sigma}_{\boldsymbol{Z}} - \boldsymbol{\Sigma}_{\boldsymbol{ZX}}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{XZ}})\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}E\{\boldsymbol{Z}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \boldsymbol{0}_{q\times 1}.$$

Because $\boldsymbol{\Sigma}_{\boldsymbol{Z}} - \boldsymbol{\Sigma}_{\boldsymbol{ZX}}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{XZ}}$ is positive definite, $E\{\boldsymbol{Z}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \boldsymbol{0}_{q\times 1}$.

So given (B0), a contradiction occurs if either (B1) or (B2) holds. Thus, we can conclude that if $\phi_1 \neq 0$, then $\phi_2 \neq 0$.

On the other hand, if we assume that $\phi_1 = 0$ and $\phi_2 \neq 0$, then

$$\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\alpha} + E\{\boldsymbol{X}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \phi_2\boldsymbol{\Sigma}_{\boldsymbol{XZ}}\boldsymbol{\gamma}, \tag{S2.7}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{ZX}}\boldsymbol{\alpha} + E\{\boldsymbol{Z}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \phi_2\boldsymbol{\Sigma}_{\boldsymbol{Z}}\boldsymbol{\gamma}. \tag{S2.8}$$

By (S2.7) and (S2.8),

$$\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\alpha} + E\{\boldsymbol{X}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \boldsymbol{\Sigma}_{\boldsymbol{XZ}}\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{ZX}}\boldsymbol{\alpha} + \boldsymbol{\Sigma}_{\boldsymbol{XZ}}\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}E\{\boldsymbol{Z}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\}. \tag{S2.9}$$

Under (B1), $E\{\boldsymbol{X}g(\boldsymbol{\gamma}^\top \boldsymbol{Z})\} = \boldsymbol{0}_{p\times 1}$, so by (S2.7) $\boldsymbol{\alpha} = \boldsymbol{0}_{p\times 1}$. Under (B2), by (S2.3) and (S2.9),

$$(\boldsymbol{\Sigma}_{\boldsymbol{X}} - \boldsymbol{\Sigma}_{\boldsymbol{XZ}}\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{ZX}})\boldsymbol{\alpha} = \boldsymbol{0}_{p\times 1}.$$

Because $\boldsymbol{\Sigma}_{\boldsymbol{X}} - \boldsymbol{\Sigma}_{\boldsymbol{XZ}}\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{ZX}}$ is positive definite, $\boldsymbol{\alpha} = \boldsymbol{0}_{p\times 1}$.

Similarly, given (B0), there is a contradiction in each case. Thus, if $\phi_2 \neq 0$, then we must have $\phi_1 \neq 0$. The proof is complete.

Proof of Theorem 2. For the sake of simplicity, we assume for the moment that d is a large but fixed constant. Let $\hat{\varphi}_1 = T_{n1}/T_{n2}$. Then, we can write

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = (\hat{\varphi}_1 - \varphi_1)\boldsymbol{\beta}_1 + \varphi_1(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + (\hat{\varphi}_1 - \varphi_1)(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1). \tag{S2.10}$$

We need to consider the term $\hat{\varphi}_1 - \varphi$. The whole proof is divided into two main steps.

Step 1. We consider the numerator T_{n1} of $\hat{\varphi}_1$. Note that

$$\begin{aligned}
T_{n1} &= \frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{\mu}(\boldsymbol{z}_i^\top\hat{\boldsymbol{\beta}}_2;\hat{\boldsymbol{\beta}}_2)\}\{\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top\hat{\boldsymbol{\beta}}_2;\hat{\boldsymbol{\beta}}_2)\}^\top\hat{\boldsymbol{\beta}}_1 \\
&= \frac{1}{n}\sum_{i=1}^{n}\{y_i - \mu(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2)\}\{\boldsymbol{x}_i - \boldsymbol{\mu}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2)\}^\top\boldsymbol{\beta}_1 \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\{y_i - \mu(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2)\}\{\boldsymbol{x}_i - \boldsymbol{\mu}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2)\}^\top(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + R_{n1} \\
&\equiv I_{n1} + R_{n1},
\end{aligned}$$

where

$$\begin{aligned}
R_{n1} &= \frac{1}{n}\sum_{i=1}^{n}\{y_i - \mu(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2)\}\{\boldsymbol{\mu}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) - \hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top\hat{\boldsymbol{\beta}}_2;\hat{\boldsymbol{\beta}}_2)\}^\top\hat{\boldsymbol{\beta}}_1 \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\{\mu(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) - \hat{\mu}(\boldsymbol{z}_i^\top\hat{\boldsymbol{\beta}}_2;\hat{\boldsymbol{\beta}}_2)\}\{\boldsymbol{x}_i - \boldsymbol{\mu}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2)\}^\top\hat{\boldsymbol{\beta}}_1 \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\{\mu(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) - \hat{\mu}(\boldsymbol{z}_i^\top\hat{\boldsymbol{\beta}}_2;\hat{\boldsymbol{\beta}}_2)\}\{\boldsymbol{\mu}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) - \hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top\hat{\boldsymbol{\beta}}_2;\hat{\boldsymbol{\beta}}_2)\}^\top\hat{\boldsymbol{\beta}}_1 \\
&\equiv R_{n11} + R_{n12} + R_{n13}.
\end{aligned}$$

First, consider R_{n11}. It has the following decomposition:

$$\begin{aligned}
R_{n11} &= \frac{1}{n}\sum_{i=1}^{n}\{y_i - \mu(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2)\}\{\boldsymbol{\mu}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) - \hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2)\}^\top\hat{\boldsymbol{\beta}}_1 \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\{y_i - \mu(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2)\}\{\hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) - \hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top\hat{\boldsymbol{\beta}}_2;\hat{\boldsymbol{\beta}}_2)\}^\top\hat{\boldsymbol{\beta}}_1 \\
&= \frac{1}{n}\sum_{i=1}^{n}\{\epsilon_i + \varphi_1\boldsymbol{\beta}_1^\top\tilde{\boldsymbol{x}}_i\}\{\boldsymbol{\mu}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) - \hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2)\}^\top\hat{\boldsymbol{\beta}}_1 \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\{\epsilon_i + \varphi_1\boldsymbol{\beta}_1^\top\tilde{\boldsymbol{x}}_i\}\{\hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) - \hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top\hat{\boldsymbol{\beta}}_2;\hat{\boldsymbol{\beta}}_2)\}^\top\hat{\boldsymbol{\beta}}_1 \\
&\equiv R_{n11}^{(1)} + R_{n11}^{(2)},
\end{aligned}$$

where $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - E(\boldsymbol{x}_i|\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)$. An application of Proposition 1 (iii) in Cui, Härdle, and Zhu (2011) yields that

$$
\begin{aligned}
& \hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top\hat{\boldsymbol{\beta}}_2;\hat{\boldsymbol{\beta}}_2) - \hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) \\
= & \left\{\boldsymbol{\mu}_1'(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) + O_P\left(\sqrt{h^4 + \frac{1}{nh^3}}\right)\right\}^\top (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2).
\end{aligned}
$$

It follows that $R_{n11}^{(2)} = o_P(n^{-1/2})$. Using equivalent kernel (Fan and Gijbels (1996), page 64), we can show that

$$
\begin{aligned}
R_{n11}^{(1)} &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n K_h(\boldsymbol{z}_j^\top\boldsymbol{\beta}_2 - \boldsymbol{z}_i^\top\boldsymbol{\beta}_2)(\epsilon_i + \varphi_1\boldsymbol{\beta}_1^\top\tilde{\boldsymbol{x}}_i) \\
&\quad \frac{\{\boldsymbol{\mu}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) - \boldsymbol{x}_j\}^\top\hat{\boldsymbol{\beta}}_1}{f_{\boldsymbol{\beta}_2^\top\boldsymbol{Z}}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)}\{1 + o_P(1)\} \\
&\equiv -\frac{K(0)}{n^2h}\sum_{i=1}^n(\epsilon_i + \varphi_1\boldsymbol{\beta}_1^\top\tilde{\boldsymbol{x}}_i)\frac{\tilde{\boldsymbol{x}}_i^\top\hat{\boldsymbol{\beta}}_1}{f_{\boldsymbol{\beta}_2^\top\boldsymbol{Z}}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)}\{1 + o_P(1)\} + \frac{1}{2}U_n^\top\hat{\boldsymbol{\beta}}_1\{1 + o_P(1)\} \\
&= \frac{1}{2}U_n^\top\hat{\boldsymbol{\beta}}_1\{1 + o_P(1)\} + o_P\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

where $f_{\boldsymbol{\beta}_2^\top\boldsymbol{Z}}(\cdot)$ is the density function of $\boldsymbol{\beta}_2^\top\boldsymbol{Z}$ and U_n is a U-statistic with kernel function

$$
\begin{aligned}
& H(\boldsymbol{x}_i, \boldsymbol{z}_i^\top\boldsymbol{\beta}_2, \epsilon_i; \boldsymbol{x}_j, \boldsymbol{z}_j^\top\boldsymbol{\beta}_2, \epsilon_j) \\
= & \frac{1}{2}K_h(\boldsymbol{z}_j^\top\boldsymbol{\beta}_2 - \boldsymbol{z}_i^\top\boldsymbol{\beta}_2)(\epsilon_i + \varphi_1\boldsymbol{\beta}_1^\top\tilde{\boldsymbol{x}}_i)\frac{\boldsymbol{\mu}_1(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) - \boldsymbol{x}_j}{f_{\boldsymbol{\beta}_2^\top\boldsymbol{Z}}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)} \\
& + \frac{1}{2}K_h(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2 - \boldsymbol{z}_j^\top\boldsymbol{\beta}_2)(\epsilon_j + \varphi_1\boldsymbol{\beta}_1^\top\tilde{\boldsymbol{x}}_j)\frac{\boldsymbol{\mu}_1(\boldsymbol{z}_j^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) - \boldsymbol{x}_i}{f_{\boldsymbol{\beta}_2^\top\boldsymbol{Z}}(\boldsymbol{z}_j^\top\boldsymbol{\beta}_2)}.
\end{aligned}
$$

A simple calculation yields

$$
E\{H(\boldsymbol{x}_1, \boldsymbol{z}_1^\top\boldsymbol{\beta}_2, \epsilon_1; \boldsymbol{x}_2, \boldsymbol{z}_2^\top\boldsymbol{\beta}_2, \epsilon_2)\} = 0
$$

and

$$
E\{\|H(\boldsymbol{x}_1, \boldsymbol{z}_1^\top\boldsymbol{\beta}_2, \epsilon_1; \boldsymbol{x}_2, \boldsymbol{z}_2^\top\boldsymbol{\beta}_2, \epsilon_2)\|_2^2\} = o(n).
$$

By invoking Lemma 3.1 of Powell, Stock, and Stoker (1989), we obtain

$$
R_{n11}^{(1)} = \frac{2}{n}\sum_{i=1}^n \tilde{H}(\boldsymbol{x}_i, \boldsymbol{z}_i^\top\boldsymbol{\beta}_2, \epsilon_i) + o_P\left(\frac{1}{\sqrt{n}}\right),
$$

where

$$
\begin{aligned}
& \tilde{H}(\boldsymbol{x}_1, \boldsymbol{z}_1^\top\boldsymbol{\beta}_2, \epsilon_1) \\
= & E\{H(\boldsymbol{x}_1, \boldsymbol{z}_1^\top\boldsymbol{\beta}_2, \epsilon_1; \boldsymbol{x}_2, \boldsymbol{z}_2^\top\boldsymbol{\beta}_2, \epsilon_2)|\boldsymbol{x}_1, \boldsymbol{z}_1^\top\boldsymbol{\beta}_2, \epsilon_1\} \\
= & \frac{1}{2}\frac{\epsilon_1 + \varphi_1\boldsymbol{\beta}_1^\top\tilde{\boldsymbol{x}}_1}{f_{\boldsymbol{\beta}_2^\top\boldsymbol{Z}}(\boldsymbol{z}_1^\top\boldsymbol{\beta}_2)}E[\{\boldsymbol{\mu}_1(\boldsymbol{z}_1^\top\boldsymbol{\beta}_2;\boldsymbol{\beta}_2) - \boldsymbol{x}_2\}K_h(\boldsymbol{z}_2^\top\boldsymbol{\beta}_2 - \boldsymbol{z}_1^\top\boldsymbol{\beta}_2)|\boldsymbol{x}_1, \boldsymbol{z}_1^\top\boldsymbol{\beta}_2, \epsilon_1].
\end{aligned}
$$

It is easy to obtain that $E\{\tilde{H}(\boldsymbol{x}_1, \boldsymbol{z}_1^\top \boldsymbol{\beta}_2, \epsilon_1)\} = 0$ and $E\{\|\tilde{H}(\boldsymbol{x}_1, \boldsymbol{z}_1^\top \boldsymbol{\beta}_2, \epsilon_1)\|_2^2\} = O(h^2)$. It then follows that $R_{n11}^{(1)} = o_P(n^{-1/2})$. Thus, $R_{n11} = o_P(n^{-1/2})$.

Similarly, we obtain $R_{n12} = o_P(n^{-1/2})$.

Next, we consider R_{n13}. By Lemma A.4 in Wang et al. (2010) and the Cauchy-Schwarz inequality, for $\log^2 n = o(nh^2)$,

$$
\begin{aligned}
R_{n13} &\leq \left[\frac{1}{n}\sum_{i=1}^n \{\mu(\boldsymbol{z}_i^\top \boldsymbol{\beta}_2; \boldsymbol{\beta}_2) - \hat{\mu}(\boldsymbol{z}_i^\top \hat{\boldsymbol{\beta}}_2; \hat{\boldsymbol{\beta}}_2)\}^2\right]^{1/2} \\
&\quad \times \left(\frac{1}{n}\sum_{i=1}^n [\{\boldsymbol{\mu}_1(\boldsymbol{z}_i^\top \boldsymbol{\beta}_2; \boldsymbol{\beta}_2) - \hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top \hat{\boldsymbol{\beta}}_2; \hat{\boldsymbol{\beta}}_2)\}^\top \hat{\boldsymbol{\beta}}_1]^2\right)^{1/2} \\
&= O_p\left(\frac{\log n}{nh}\right) = o_P\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

Thus, we can arrive at $R_{n1} = o_P(n^{-1/2})$.

Step 2. We then deal with the denominator T_{n2} of $\hat{\varphi}_1$. Write

$$
\begin{aligned}
T_{n2} &= \frac{1}{n}\sum_{i=1}^n [\{\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_1(\boldsymbol{z}_i^\top \hat{\boldsymbol{\beta}}_2; \hat{\boldsymbol{\beta}}_2)\}^\top \hat{\boldsymbol{\beta}}_1]^2 \\
&= \frac{1}{n}\sum_{i=1}^n \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top \boldsymbol{\beta}_1 + \frac{2}{n}\sum_{i=1}^n \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + R_{n2} \\
&\equiv I_{n2} + R_{n2}.
\end{aligned}
$$

Similarly to the proof of R_{n1}, we can obtain that $R_{n2} = o_P(n^{-1/2})$.

Note that

$$
I_{n1} = \frac{1}{n}\sum_{i=1}^n (\epsilon_i + \varphi_1 \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i) \tilde{\boldsymbol{x}}_i^\top \boldsymbol{\beta}_1 + \frac{1}{n}\sum_{i=1}^n (\epsilon_i + \varphi_1 \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i) \tilde{\boldsymbol{x}}_i^\top (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)
$$

and

$$
I_{n2} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top \boldsymbol{\beta}_1 + \frac{2}{n}\sum_{i=1}^n \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1).
$$

By the Slutsky's theorem, we have

$$
\begin{aligned}
\hat{\varphi}_1 - \varphi_1 &= \frac{I_{n1}}{I_{n2}} - \varphi_1 + o_P\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i \tilde{\boldsymbol{x}}_i^\top \boldsymbol{\beta}_1 + \frac{1}{n}\sum_{i=1}^n (\epsilon_i - \varphi_1 \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i) \tilde{\boldsymbol{x}}_i^\top (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)}{\frac{1}{n}\sum_{i=1}^n \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top \boldsymbol{\beta}_1 + \frac{2}{n}\sum_{i=1}^n \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)} + o_P\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i \tilde{\boldsymbol{x}}_i^\top \boldsymbol{\beta}_1 - \frac{1}{n}\sum_{i=1}^n \varphi_1 \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)}{\frac{1}{n}\sum_{i=1}^n \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top \boldsymbol{\beta}_1} + o_P\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

Therefore, by (S2.10),

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} &= \boldsymbol{\beta}_1\boldsymbol{\beta}_1^\top \frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i \tilde{\boldsymbol{x}}_i}{\frac{1}{n}\sum_{i=1}^n \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top \boldsymbol{\beta}_1} - \boldsymbol{\beta}_1\boldsymbol{\beta}_1^\top \frac{\frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top}{\frac{1}{n}\sum_{i=1}^n \boldsymbol{\beta}_1^\top \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top \boldsymbol{\beta}_1} \varphi_1(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \\
&\quad + \varphi_1(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + o_P\left(\frac{1}{\sqrt{n}}\right) \\
&= \boldsymbol{\alpha}\boldsymbol{\alpha}^\top \frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i \tilde{\boldsymbol{x}}_i}{\frac{1}{n}\sum_{i=1}^n \boldsymbol{\alpha}^\top \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top \boldsymbol{\alpha}} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\top \frac{\frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top}{\frac{1}{n}\sum_{i=1}^n \boldsymbol{\alpha}^\top \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top \boldsymbol{\alpha}} \varphi_1(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \\
&\quad + \varphi_1(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + o_P\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

The proof is complete.

Proof of Theorem 3. See Wang, Xu, and Zhu (2012).

References

Cui, X., Härdle, W., and Zhu, L. X. (2011). The EFM approach for single-index models. *Ann. Statist.* **39**, 1658–1688.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman & Hall, New York.

Portnoy, S. (1984). Asymptotic behavior of M-estimates of p regression parameters when p^2/n is large. I. Consistency. *Ann. Statist.* **12**, 1298–1309.

Powell, J. L., Stock, J. H, and Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica* **57**, 1403–1430.

Wang, J. L., Xue, L. G., Zhu, L. X., and Chong, Y. S. (2010). Estimation for a partial-linear single-index model. *Ann. Statist.* **38**, 246–274.

Wang, T., Xu, P. R., and Zhu, L. X. (2012). Non-convex penalized estimation in high-dimensional models with single-index structure. *J. Multivariate Anal.* **109**, 221–235.

Wu, Y. and Li, L. (2011). Asymptotic properties of sufficient dimension reduction with a diverging number of predictors. *Statist. Sinica* **31**, 707–730.