Partial linear varying multi-index coefficient model for integrative gene-environment interactions

Xu Liu[1], Yuehua Cui[1] and Runze Li[2]
[1]*Michigan State University,* [2]*Pennsylvania State University*

Supplementary Material

Summary

This supplementary file contains 4 sections. The first section describes the algorithm for the estimation method. The second section illustrates the detailed bandwidth selection and the empirical bias bandwidth selection (EBBS) methods. Some additional simulations about the estimation and further details about case study are given in Section 3. The detailed proofs of Lemmas and Theorems are given in Section 4.

S1 Algorithm

The following describes the detailed algorithm of the estimation method in Section 2.

- **Step 0**. Choose initial values for $\boldsymbol{\beta}_l$, denoted by $\boldsymbol{\beta}_l^{(old)}$, $l = 0, 1$.
- **Step 1**. Calculate $u_l = \mathbf{X}\boldsymbol{\beta}_l^{(old)}$, and $\mathbf{B}_r(u_l)$, $l = 0, 1$.
- **Step 2**. Use ordinary least squares estimation to obtain $\widehat{\boldsymbol{\alpha}}$, and $\widehat{\lambda}(\boldsymbol{\beta})$ by (3) in the paper.
- **Step 3**. Estimate $\boldsymbol{\beta}^{(new)}$ by

$$\widehat{\boldsymbol{\beta}}^{(new)} = \arg\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}_\beta} \tilde{R}((\widehat{\boldsymbol{\alpha}}^T, \boldsymbol{\beta}^T)^T, \widehat{\lambda}(\boldsymbol{\beta})),$$

  then update $\boldsymbol{\beta}_l^{(old)}$ by $\boldsymbol{\beta}_l^{(old)} = \boldsymbol{\beta}_l^{(new)}/\|\boldsymbol{\beta}_l^{(new)}\|_2$, $l = 0, 1$. The Newton-Raphson algorithm is used to calculate $\boldsymbol{\beta}$ here.
- **Step 4**. Repeat Steps 1-3 until convergence. This gives the whole parametric estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\alpha}}^T, \widehat{\boldsymbol{\beta}}^T)^T$ and $\widehat{\lambda}_l(\widehat{\boldsymbol{\theta}})$, which results in $\tilde{m}_l(\widehat{\boldsymbol{\beta}}_l^T\mathbf{X}, \widehat{\boldsymbol{\beta}})$, $l = 0, 1$.
- **Step 5**. Given the parametric estimator $\widehat{\boldsymbol{\theta}}$ in Step 4, calculate the pseudo responses,

$$\tilde{Y}_{i0} = Y_i - \widehat{\boldsymbol{\alpha}}_0^T\mathbf{Z}_i - \tilde{m}_1(\widehat{\boldsymbol{\beta}}_1^T\mathbf{X}_i, \widehat{\boldsymbol{\beta}})G_i - \widehat{\boldsymbol{\alpha}}_1^T\mathbf{Z}_iG_i,$$
$$\tilde{Y}_{i1} = Y_i - \widehat{\boldsymbol{\alpha}}_0^T\mathbf{Z}_i - \tilde{m}_0(\widehat{\boldsymbol{\beta}}_0^T\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) - \widehat{\boldsymbol{\alpha}}_1^T\mathbf{Z}_iG_i.$$

- **Step 6**. Compute $\widehat{m}_1(u_1, \widehat{\boldsymbol{\beta}})$ by (5) in the paper based on new data $\{\tilde{Y}_{i1}, \mathbf{X}_i, \mathbf{Z}_i, G_i\}_{i=1}^n$, and $\widehat{m}_0(u_0, \widehat{\boldsymbol{\beta}})$ based on new data $\{\tilde{Y}_{i0}, \mathbf{X}_i, \mathbf{Z}_i, G_i\}_{i=1}^n$.

Remark S1: Unlike the algorithms proposed by Carroll et al. (1997) and Cui et al. (2011) which estimate the parametric parameters and nonparametric functions iteratively, we do not need to iterate the two steps for the kernel estimator. Thus, it is faster than their algorithm.

Remark S2: In Step 3, we need to calculate the first derivative of $\mathbf{B}_r(u_l)$. Denote by $\mathbf{B}_r'(u_l)^T$ the first derivative of $\mathbf{B}_r(u_l)$. According to de Boor (2001), $\mathbf{B}_r'(u_l) = \mathbf{B}_{r-1}(u_l)^T \mathbf{W}_1$, where $\mathbf{B}_{r-1}(u_l) = (B_{s,r-1}(u_l) : 2 \le s \le J_n)^T$, and

$$\mathbf{W}_1 = (r-1)\begin{pmatrix} \frac{-1}{\xi_{r+1}-\xi_2} & \frac{1}{\xi_{r+1}-\xi_2} & 0 & \cdots & 0 \\ 0 & \frac{-1}{\xi_{r+2}-\xi_3} & \frac{1}{\xi_{r+2}-\xi_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{-1}{\xi_{N+2r-1}-\xi_{N+r}} & \frac{1}{\xi_{N+2r-1}-\xi_{N+r}} \end{pmatrix}_{(J_n-1)\times J_n}. \tag{A.1}$$

In $\mathbf{W}_1$, ξ_j, $j = 1, \cdots, N_n + r$ are defined in Section 2.2 in our paper.

S2 Bandwidth selection

S2.1 The EBBS algorithm

The BSBK estimator $\widehat{m}_l(u_l, \widehat{\boldsymbol{\beta}})$ is sensitive to the choice of bandwidth h_l, $l = 0, 1$. Fortunately, the bandwidth selection is not related to the parametric estimators as well as the B-spline function estimators $\tilde{m}_0(\widehat{\boldsymbol{\beta}}_0)$ and $\tilde{m}_1(\widehat{\boldsymbol{\beta}}_1)$. Thus, the bandwidths are chosen based on $\widehat{\boldsymbol{\theta}}$ and new data $\tilde{Y}_{i0}$ and $\tilde{Y}_{i1}$, $i = 1, \cdots, n$, defined in Section 2.3. Bandwidth selection has been intensively studied in nonparametric literature, see Sepanski et al. (1994) and Ruppert et al. (1995) for good discussions. In fact, the bandwidth h can be chosen by any bandwidth selector that minimizes the mean squared error (MSE) of the estimator, for example, the cross-validation. Here we employ a bandwidth selection method called empirical bias bandwidth selection (EBBS) (Ruppert et al. 1995; Carroll and Ruppert 1998; Liu et al. 2014).

The basic idea behind EBBS is as follows. To facilitate, we only consider the choice of h_0, and denote it by h by omitting the subscript in current section. Give parametric estimator $\widehat{\boldsymbol{\theta}}$ and new data $(\tilde{Y}_{i0}, X_i, Z_i, G_i)$ and $\widehat{u}_i = X_i^T \widehat{\boldsymbol{\beta}}_0$, $i = 1, \cdots, n$, which is used for the local linear estimator $\widehat{m}_0(u_0, \widehat{\boldsymbol{\beta}})$. Fix h^* and u_0, we can assume that $\zeta(h^*, \boldsymbol{\tau})$ is a known form of the bias function of $m_0(u_0)$ by Theorem 4, such as linear form $\zeta(h, \gamma) = \tau_1 h^{s+1} + \cdots + \tau_t h^{s+t}$, where $t \ge 1$, s is the degree of polynomial kernel regression used, and $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_t)^T$ is unknown. In this work, we take $s = 1$ because the linear kernel regression is used. Let $\widehat{m}_0(u_0, \widehat{\boldsymbol{\beta}}, h)$ be the BSBK estimator of $m_0(u_0)$, where $\widehat{m}_0(u_0, \widehat{\boldsymbol{\beta}}, h)$ is $\widehat{m}_0(u_0, \widehat{\boldsymbol{\beta}})$ in Theorem 4 emphasizing that it relates to bandwidth h. According to Theorem 4, the BSBK estimator $\widehat{m}_0(u_0, \widehat{\boldsymbol{\beta}}, h)$ can be well approximated by a function of h, $\tau_0 + \zeta(h, \boldsymbol{\tau}) + o_p(h^{s+t})$, where $\tau_0 = m_0(u_0)$ is the limit.

Denote by $\{h_1, \cdots, h_K\}$ a grid of bandwidths in a neighborhood $\mathcal{H}_0$ of h^*, where $K \geq t+1$. Let $(\widehat{\tau}_0, \widehat{\boldsymbol{\tau}})$ minimize $\sum_{k=1}^{K}\{\widehat{m}_0(u_0, \widehat{\boldsymbol{\beta}}, h_k) - \tau_0 - \zeta(h_k, \boldsymbol{\tau})\}^2$. Then by Theorem 4, the bias of $m_0(u_0)$ could be estimated at h^* by $\zeta(h^*, \widehat{\boldsymbol{\tau}})$ if $\mathcal{H}_0$ is small enough.

Noting that the MSE of $\widehat{m}_0(u_0, \widehat{\boldsymbol{\beta}}, h)$ is a function of bandwidth h, the data-driven optimal bandwidth is selected to minimize the following MSE function,

$$\mathrm{MSE}(h, u_0) = \mathrm{bias}^2\left(\widehat{m}_0(u_0, \widehat{\boldsymbol{\beta}}, h)\right) + \mathrm{var}\left(\widehat{m}_0(u_0, \widehat{\boldsymbol{\beta}}, h)\right).$$

By Section 2.4 and the previous discussion, it is easy to estimate $\mathrm{MSE}(h_k, u_0)$ at each h_k, $k = 1, \cdots, K$. The two quantities t and J need to be determined in practice. As discussed by Carroll and Ruppert (1998), we do a grid search for bandwidth $\mathcal{H}_1 = \{h_1, \cdots, h_M\}$. Given (K_1, K_2) such that $K_1 + K_2 \geq t$, we use above method to calculate $\mathrm{MSE}(h_k, u_0)$ for each h_k, $k \in \{K_1 + 1, \cdots, M - K_2\}$ with $\mathcal{H}_0 = \{h_j, j = k - K_1, \cdots, k + K_2\}$. In our simulation and real data analysis, we select (t, M, K_1, K_2) following Carroll and Ruppert (1998).

S2.2 Bandwidth selection in simulation studies

Here we demonstrate how to select the bandwidth for the simulation study in Section 4.1. Initializing a bandwidth $h_l^* = \widehat{\sigma}_{u_l} \cdot n^{-1/5}$, where $\widehat{\sigma}_{u_l}$ is the sample standard deviation of the estimator $u_l = \widehat{\boldsymbol{\beta}}_l^T \mathbf{X}$, $l = 0, 1$, we give a grid of bandwidths $\mathcal{H}_1 = h^* \times R$ in which R is chosen from the interval $[0.1, 2.1]$ by 0.1 increment. Here we choose $(t, K_1, K_2) = (1, 2, 2)$. One can choose other combinations and more details are referred to Ruppert et al. (1995), Carroll and Ruppert (1998) and Liu et al. (2014).

S3 Simulation and Case Studies

S3.1 Performance of estimation

Consider the PLVMICM model

$$Y = m_0(\boldsymbol{\beta}_0^T \mathbf{X}) + \boldsymbol{\alpha}_0^T \mathbf{Z} + m_1(\boldsymbol{\beta}_1^T \mathbf{X})G + \boldsymbol{\alpha}_1^T \mathbf{Z} G + \varepsilon, \quad \text{(A.2)}$$

where the setup is the same as that in Section 4.1 in the paper. Table S1 reports the Bias, SD, SE and CP, which are defined in Section 4.1 of the paper. In this Table, the results for $n = 1000$ are added. For easy comparison, we repeat the results listed in the main context for $n = 200$ and $n = 500$. It is clear that the performance improves as the sample size n increases.

The estimation for function $m_0(\cdot)$ under different MAFs and sample sizes is shown in Figure S1. Overall, the function can be reasonably estimated with high accuracy indicated with narrow confidence bands under different simulation combinations. In Figure S2 we added the estimation for function $m_1(\cdot)$ under $n = 1000$. The result is consistent with what we observed in the main context.

We also considered the PLVMICM model given in model (9) in the main context with two genetic components and tested if both $m_1(\cdot)$ and $m_2(\cdot)$ are simultaneously linear following

Table S1: Simulation results under $p_A = 0.1, 0.3, 0.5$ and sample size $n = 200, 500, 1000$.

			$p_A = 0.1$				$p_A = 0.3$				$p_A = 0.5$			
n	Param	True	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
200	α_{01}	0.500	4.4E-04	0.016	0.016	95.2	3.1E-04	0.020	0.020	95.2	9.9E-04	0.026	0.026	95.1
	α_{02}	0.500	-1.6E-04	0.016	0.016	95.3	4.1E-04	0.020	0.020	95.3	5.6E-04	0.026	0.026	95.8
	α_{11}	0.300	9.4E-05	0.040	0.039	94.1	6.0E-04	0.024	0.024	94.1	6.7E-05	0.022	0.022	95.2
	α_{12}	0.300	-1.1E-03	0.040	0.039	95.0	-1.1E-03	0.023	0.024	95.9	-4.4E-04	0.021	0.022	96.3
	β_{01}	0.620	-3.7E-04	0.011	0.011	94.7	-1.7E-03	0.012	0.013	94.8	-2.1E-03	0.014	0.014	94.5
	β_{02}	0.555	3.3E-04	0.012	0.012	95.3	1.0E-03	0.013	0.013	96.4	1.5E-03	0.014	0.015	96.6
	β_{03}	0.555	-2.7E-04	0.012	0.012	94.0	4.2E-04	0.013	0.013	95.3	3.1E-04	0.015	0.015	95.4
	β_{11}	0.577	1.4E-03	0.028	0.027	92.9	-4.0E-04	0.015	0.015	95.5	-7.5E-05	0.012	0.012	95.1
	β_{12}	0.577	-3.4E-04	0.029	0.028	93.5	9.5E-05	0.015	0.015	95.3	2.9E-04	0.011	0.012	96.2
	β_{13}	0.577	-3.2E-03	0.028	0.027	94.3	-2.6E-04	0.015	0.015	96.1	-5.7E-04	0.012	0.012	96.0
500	α_{01}	0.500	-3.2E-04	0.010	0.010	95.8	-5.5E-04	0.012	0.012	95.2	-4.0E-04	0.016	0.016	96.1
	α_{02}	0.500	1.9E-04	0.010	0.010	94.1	2.0E-04	0.013	0.012	94.2	3.8E-04	0.016	0.016	94.6
	α_{11}	0.300	5.6E-04	0.023	0.022	93.7	9.9E-04	0.015	0.014	93.8	6.5E-04	0.013	0.013	94.5
	α_{12}	0.300	1.2E-05	0.023	0.022	94.0	2.6E-04	0.015	0.014	93.8	2.0E-04	0.013	0.013	94.1
	β_{01}	0.620	-4.6E-04	0.007	0.007	95.2	-1.0E-03	0.008	0.008	95.7	-1.2E-03	0.009	0.009	94.9
	β_{02}	0.555	1.2E-04	0.007	0.007	95.5	4.3E-04	0.008	0.008	95.1	5.5E-04	0.009	0.009	95.1
	β_{03}	0.555	2.6E-04	0.007	0.007	94.2	5.2E-04	0.008	0.008	94.1	5.2E-04	0.009	0.009	94.4
	β_{11}	0.577	5.2E-04	0.015	0.016	95.0	3.0E-05	0.009	0.009	96.6	-8.5E-06	0.007	0.007	95.9
	β_{12}	0.577	-3.4E-04	0.016	0.016	94.0	-8.0E-06	0.009	0.009	95.6	1.0E-04	0.007	0.007	96.3
	β_{13}	0.577	-8.3E-04	0.016	0.016	94.5	-2.3E-04	0.009	0.009	95.2	-2.3E-04	0.007	0.007	94.8
1000	α_{01}	0.500	-5.1E-05	0.007	0.007	95.1	8.4E-05	0.009	0.009	95.4	2.3E-04	0.011	0.011	96.6
	α_{02}	0.500	-2.3E-04	0.007	0.007	95.2	-2.2E-04	0.009	0.009	95.2	6.1E-05	0.011	0.011	96.3
	α_{11}	0.300	1.2E-04	0.015	0.015	95.6	1.5E-04	0.010	0.010	95.3	1.3E-04	0.009	0.009	94.8
	α_{12}	0.300	1.1E-03	0.015	0.015	96.0	6.1E-04	0.010	0.010	96.5	2.8E-04	0.009	0.009	94.6
	β_{01}	0.620	-3.6E-04	0.005	0.005	94.4	-7.1E-04	0.005	0.005	95.1	-8.3E-04	0.006	0.006	95.2
	β_{02}	0.555	1.7E-04	0.005	0.005	94.4	3.5E-04	0.006	0.006	93.2	4.3E-04	0.007	0.006	95.1
	β_{03}	0.555	1.6E-04	0.005	0.005	95.0	3.6E-04	0.006	0.006	95.4	3.8E-04	0.007	0.006	94.4
	β_{11}	0.577	3.9E-04	0.011	0.011	94.5	1.4E-05	0.006	0.006	94.9	5.2E-05	0.005	0.005	94.7
	β_{12}	0.577	-7.8E-04	0.011	0.011	95.2	-1.9E-04	0.006	0.006	96.1	-1.2E-04	0.005	0.005	96.1
	β_{13}	0.577	8.6E-05	0.011	0.011	95.6	7.3E-05	0.006	0.006	95.6	2.5E-06	0.005	0.005	94.4

Theorem 6. The results are depicted in Figure S3 which are quite similar to the results of the one component test.

S3.2 Case Study

Table S2 lists the results for testing the index loading parameters, where $p_{\beta_j}, j = 1, 2, 3$ refers to the p-value for testing $H_0 : \beta_{j1} = 0$ and $p_{\beta_{jk}}$ refers to the p-value for testing $H_0 : \beta_{j1} = \beta_{k1} = 0$.

The plots of the other 4 SNPs showing statistical significance are given in Figure S4 along with their 95% confidence bands.

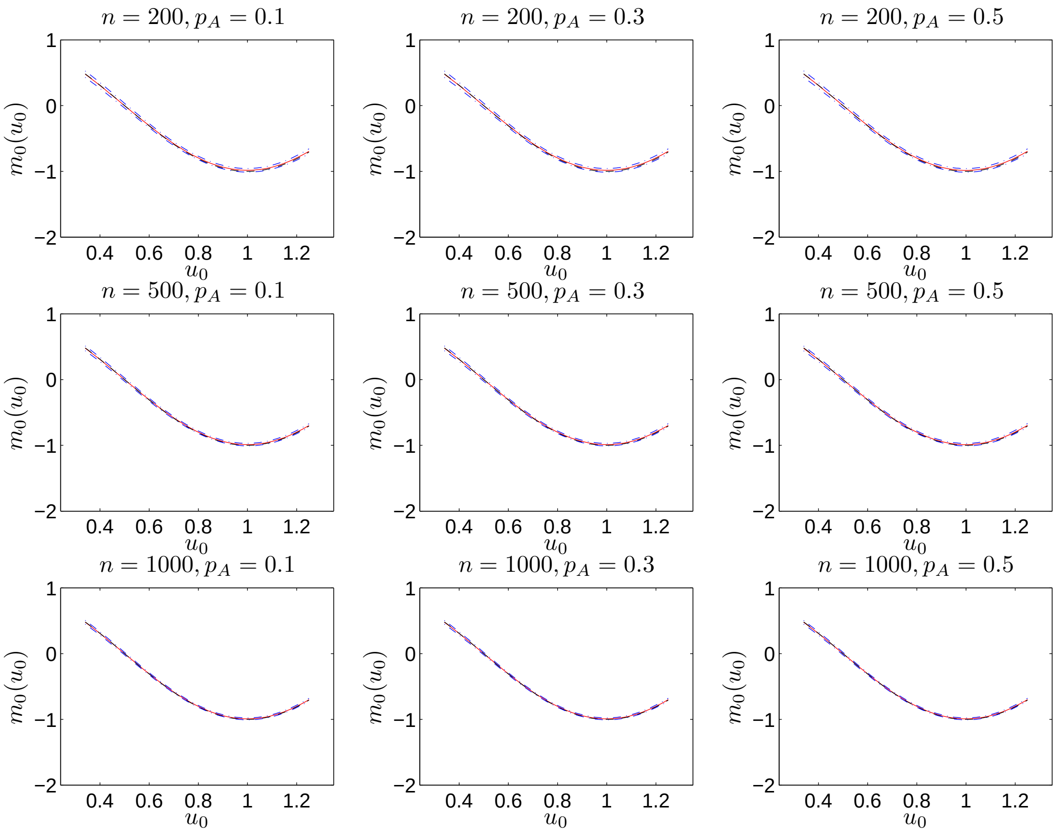

Figure S1: *The estimation of function $m_0(\cdot)$ under different MAFs and sample sizes. The estimated and true functions are denoted by the solid and dashed lines, respectively. The 95% confidence band is denoted by the dotted-dash line.*

Table S2: List of p-values for testing the index loading parameters.

SNP ID	p_{β_1}	p_{β_2}	p_{β_3}	$p_{\beta_{12}}$	$p_{\beta_{13}}$	$p_{\beta_{23}}$
rs16884481	2.93E-07	6.04E-08	2.39E-02	7.27E-07	1.18E-07	6.64E-02
rs10946428	6.41E-09	3.91E-09	2.32E-04	4.47E-08	3.17E-09	5.12E-03
rs6904348	1.86E-08	7.28E-09	5.80E-04	6.30E-08	7.75E-09	7.04E-03
rs10806925	2.79E-10	1.58E-11	6.17E-03	3.89E-08	7.63E-10	3.85E-05
rs9465873	4.73E-13	2.73E-12	1.20E-04	2.34E-12	3.77E-11	6.76E-06
rs12662218	2.22E-08	2.53E-09	6.37E-02	8.32E-08	1.50E-08	3.24E-05

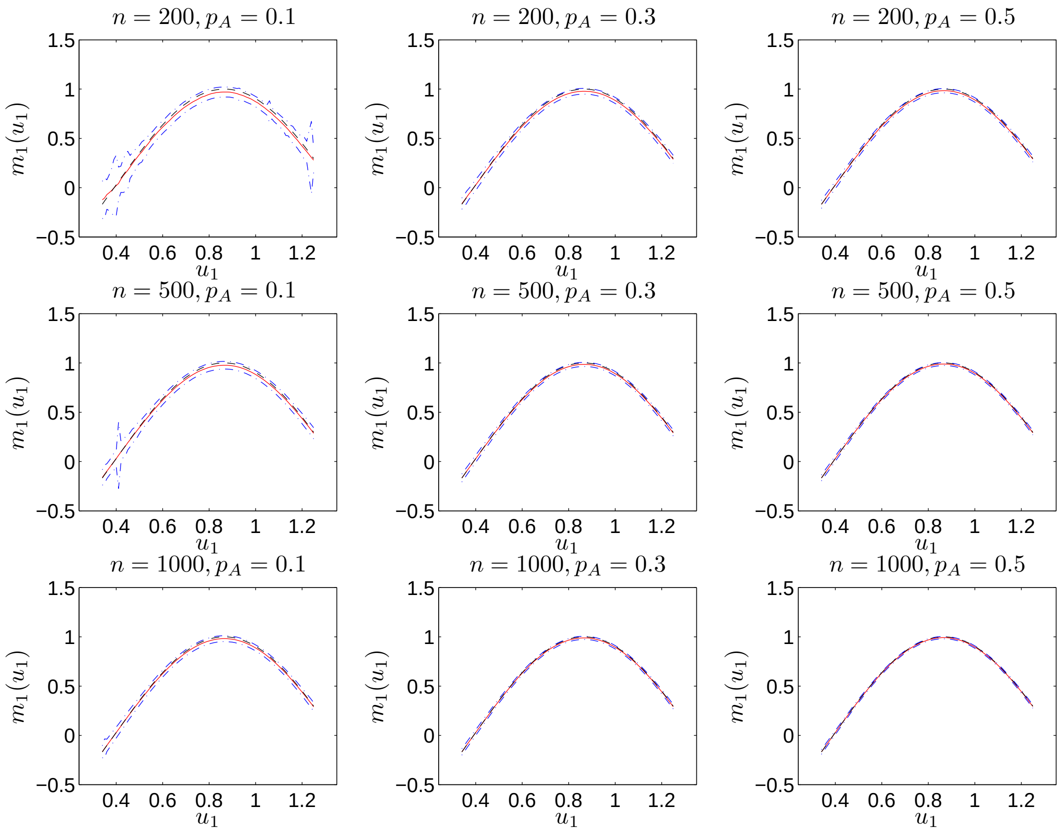

Figure S2: *The estimation of function $m_1(\cdot)$ under different MAFs and sample sizes. The estimated and true functions are denoted by the solid and dashed lines, respectively. The 95% confidence band is denoted by the dotted-dash line.*

S4 Proofs

In this section, we provide the technical details for the proofs of lemmas and theorems.

Let $m(\mathbf{V}_i, \boldsymbol{\beta}) = m_0(\mathbf{V}_i, \boldsymbol{\beta}_0) + m_1(\mathbf{V}_i, \boldsymbol{\beta}_1)G_i$ and $\mathbf{m} = (m(\mathbf{V}_1, \boldsymbol{\beta}), \cdots, m(\mathbf{V}_n, \boldsymbol{\beta}))^T$. Denote $Y_{z,i} = Y_i - \mathbf{Z}_i^T \boldsymbol{\alpha}_0^0 - \mathbf{Z}_i^T \boldsymbol{\alpha}_1^0 G_i$, $Y_z = (Y_{z,1}, \cdots, Y_{z,n})^T$, $\mathbf{e} = (\varepsilon_1, \cdots, \varepsilon_n)^T$, $\mathbb{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)^T$, $\mathbb{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_n)^T$, $\mathbf{G} = (G_1, \cdots, G_n)^T$, $\mathbb{G} = (\mathbf{1}_n, \mathbf{G})$ and $\widehat{\Lambda}(\boldsymbol{\theta}) = (\widehat{\boldsymbol{\alpha}}^T, \widehat{\lambda}(\boldsymbol{\beta})^T)^T$. Then $\widehat{\Lambda}(\boldsymbol{\theta})$ can be decomposed into $\widehat{\Lambda}(\boldsymbol{\theta}) = \widehat{\Lambda}_m(\boldsymbol{\theta}) + \widehat{\Lambda}_e(\boldsymbol{\theta})$ by (3) in the paper with

$$\begin{aligned} \widehat{\Lambda}_m(\boldsymbol{\theta}) &= (\mathbf{D}(\tilde{\mathbf{Z}}, \boldsymbol{\beta})^T \mathbf{D}(\tilde{\mathbf{Z}}, \boldsymbol{\beta}))^{-1} \mathbf{D}(\tilde{\mathbf{Z}}, \boldsymbol{\beta})^T (\mathbf{m} + \tilde{\mathbf{Z}}\boldsymbol{\alpha}), \\ \widehat{\Lambda}_e(\boldsymbol{\theta}) &= (\mathbf{D}(\tilde{\mathbf{Z}}, \boldsymbol{\beta})^T \mathbf{D}(\tilde{\mathbf{Z}}, \boldsymbol{\beta}))^{-1} \mathbf{D}(\tilde{\mathbf{Z}}, \boldsymbol{\beta})^T \mathbf{e}, \end{aligned} \tag{A.3}$$

where $\tilde{\mathbf{Z}} = (\tilde{\mathbf{Z}}_1, \cdots, \tilde{\mathbf{Z}}_n)^T$ with $\tilde{\mathbf{Z}}_i = (\mathbf{Z}_i^T, \mathbf{Z}_i^T G_i)^T$, defined in Section 2.2. Let $\boldsymbol{\Theta}$ be the

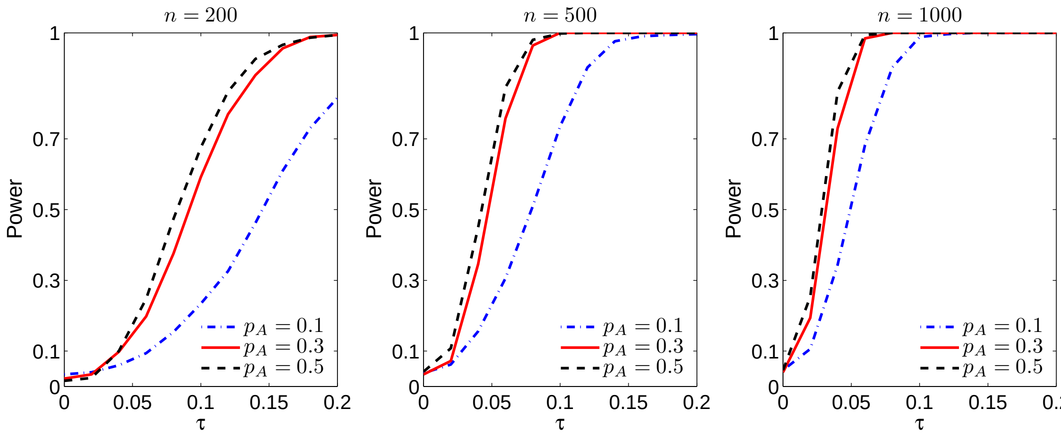

Figure S3: *The empirical size and power function of testing nonparametric functions $m_1(\cdot)$ and $m_2(\cdot)$ simultaneously under different sample sizes and MAFs.*

parametric space of $\boldsymbol{\theta}$. Define

$$\begin{aligned} \mathbf{U}(\boldsymbol{\beta}) &= E[D_i(\boldsymbol{\beta})D_i(\boldsymbol{\beta})^T], \quad \widehat{\mathbf{U}}(\boldsymbol{\theta}) = \frac{1}{n}\mathbf{D}(\boldsymbol{\beta})^T\mathbf{D}(\boldsymbol{\beta}), \\ \mathbf{U}(\tilde{\mathbf{Z}},\boldsymbol{\beta}) &= E[D_i(\tilde{\mathbf{Z}},\boldsymbol{\beta})D_i(\tilde{\mathbf{Z}},\boldsymbol{\beta})^T], \quad \widehat{\mathbf{U}}(\tilde{\mathbf{Z}},\boldsymbol{\beta}) = \frac{1}{n}\mathbf{D}(\tilde{\mathbf{Z}},\boldsymbol{\beta})^T\mathbf{D}(\tilde{\mathbf{Z}},\boldsymbol{\beta}) \end{aligned} \tag{A.4}$$

where $D_i(\boldsymbol{\beta}) = (D_{i,sl}(\boldsymbol{\beta}_l)\mathbb{G}_{il}, 1 \le s \le J_n, l = 0, 1)^T$ and $\mathbf{D}(\boldsymbol{\beta}) = (D_1(\boldsymbol{\beta}), \cdots, D_n(\boldsymbol{\beta}))^T$, which is a $n \times 2J_n$ matrix.

Lemma S.1. *Let assumptions* (A1) and (A4) *be satisfied. For any vector* $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_0^T, \boldsymbol{\zeta}_1^T)^T$ *with* $\boldsymbol{\zeta}_l = (\zeta_{s,l} : 1 \le s \le J_n)^T$ *and* $||\boldsymbol{\zeta}_l|| = 1$, $l = 0, 1$, *there exists constants* $0 < c_U < C_U < \infty$, *such that for any* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and for large enough* n,

$$c_U J_n^{-1} \le \boldsymbol{\zeta}^T\mathbf{U}(\boldsymbol{\beta})\boldsymbol{\zeta} \le C_U J_n^{-1}, \quad \text{and} \quad C_U^{-1}J_n \le \boldsymbol{\zeta}^T\mathbf{U}(\boldsymbol{\beta})^{-1}\boldsymbol{\zeta} \le c_U^{-1}J_n, \tag{A.5}$$

$$\sup_{1\le s,s'\le J_n, 0\le l\le 1}\left|n^{-1}\sum_{i=1}^n D_{i,sl}(\boldsymbol{\beta}_l)D_{i,s'l}(\boldsymbol{\beta}_l) - E[D_{i,sl}(\boldsymbol{\beta}_l)D_{i,s'l}(\boldsymbol{\beta}_l)]\right| = O\left(\sqrt{J_n^{-1}n^{-1}\log n}\right), \quad a.s., \tag{A.6}$$

$$\sup_{1\le s,s'\le J_n, l\ne l'}\left|n^{-1}\sum_{i=1}^n D_{i,sl}(\boldsymbol{\beta}_l)D_{i,s'l'}(\boldsymbol{\beta}_l) - E[D_{i,sl}(\boldsymbol{\beta}_l)D_{i,s'l'}(\boldsymbol{\beta}_l)]\right| = O\left(J_n^{-1}\sqrt{n^{-1}\log n}\right), \quad a.s., \tag{A.7}$$

and with probability approaching 1,

$$c_U J_n^{-1} \le \boldsymbol{\zeta}^T\widehat{\mathbf{U}}(\boldsymbol{\beta})\boldsymbol{\zeta} \le C_U J_n^{-1}, \quad \text{and} \quad C_U^{-1}J_n \le \boldsymbol{\zeta}^T\widehat{\mathbf{U}}(\boldsymbol{\beta})\boldsymbol{\zeta} \le c_U^{-1}J_n. \tag{A.8}$$

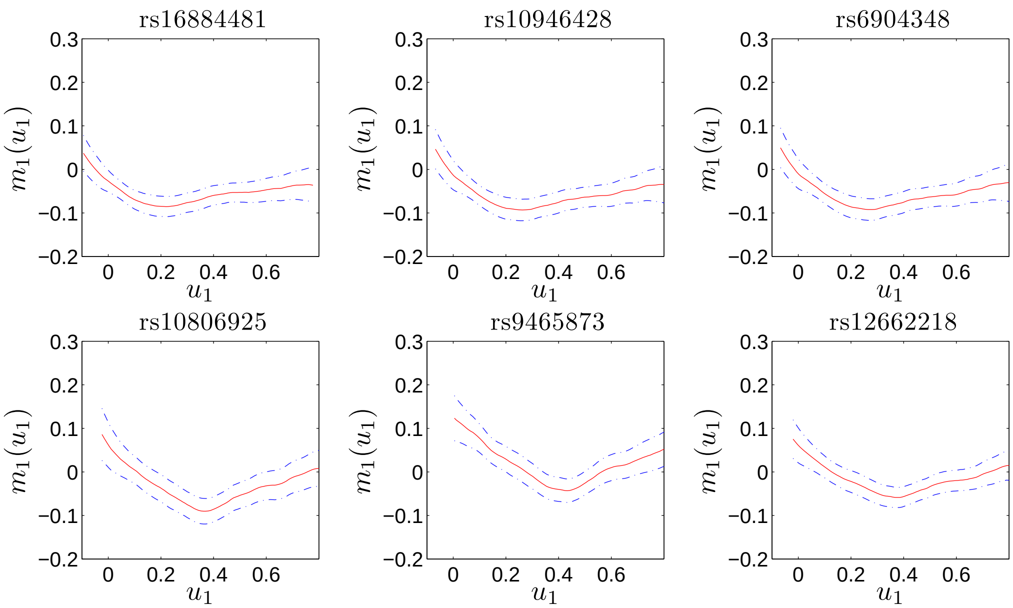

Figure S4: *Plot of the estimate (solid curve) of nonparametric function $m_1(u_1)$ for the 4 SNPs along with their 95% confidence bands (dash-dotted lines).*

Proof of Lemma S.1: By Theorem 5.4.2 of DeVore and Lorentz (1993) and assumption (A1), for any vector $\boldsymbol{\zeta}_l$ with $||\boldsymbol{\zeta}_l|| = 1$ and for large enough n, there exist constants $0 < c_l \leq C_l < \infty$, $l = 0, 1$, for any $\boldsymbol{\beta} \in \boldsymbol{\Theta}_\beta$, such that

$$c_l J_n^{-1} \leq \boldsymbol{\zeta}_l^T E\left[\mathbf{B}_r(\boldsymbol{\beta}_l^T \mathbf{X}_i)\mathbf{B}_r(\boldsymbol{\beta}_l^T \mathbf{X}_i)^T\right]\boldsymbol{\zeta}_l \leq C_l J_n^{-1}.$$

Let $\pi_{il} = \sum_{s=1}^{J_n} \zeta_{s,l} B_{s,q}(\boldsymbol{\beta}_l^T \mathbf{X}_i)$ and $\pi_i = (\pi_{i0}, \pi_{i1})^T$. By assumptions (A1) and (A4), for large enough n, we have

$$\begin{aligned}
\boldsymbol{\zeta}^T E\left[\mathbf{U}(\boldsymbol{\beta})\right]\boldsymbol{\zeta} =& E\left[\pi_{i0} + \pi_{i1} G_i\right]^2 \\
\leq& C_G E[\pi_i^T \pi_i] \\
=& C_G \sum_{l=0,1} \boldsymbol{\zeta}_l^T E\left[\mathbf{B}_r(\boldsymbol{\beta}_l^T \mathbf{X}_i)\mathbf{B}_r(\boldsymbol{\beta}_l^T \mathbf{X}_i)^T\right]\boldsymbol{\zeta}_l \\
\leq& 2C_G \min(C_1, C_2) J_n^{-1}.
\end{aligned}$$

As the same way, we have $\boldsymbol{\zeta}^T E\left[\mathbf{U}(\boldsymbol{\beta})\right]\boldsymbol{\zeta} \geq 2c_G \max(c_1, c_2) J_n^{-1}$. The second result in (A.5) can be shown directly from the first inequalities. Similarly it is easy to prove that (A.8) holds. (A.6) and (A.7) can be shown by Bernstein's inequality as Bosq (1998).

□

Lemma S.2. *Let assumptions* (A1), (A3) and (A4) *be satisfied. For any* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $||n^{-1}\mathbf{D}(\boldsymbol{\beta})^T\mathbf{e}||_2 = O_p(n^{-1/2})$.

Proof of Lemma S.2: By the law of large numbers, with probability approaching 1, we have

$$\begin{aligned}||n^{-1}\mathbf{D}(\boldsymbol{\beta})^T\mathbf{e}||_2^2 =& \sum_{1\le s\le J_n, l=0,1}\left[n^{-1}\sum_{i=1}^{n} B_{s,q}(\boldsymbol{\beta}_l^T\mathbf{X}_i)\mathbb{G}_{il}\varepsilon_i\right]^2\\ =&n^{-1}\sum_{1\le s\le J_n, l=0,1} E\left[B_{s,q}(\boldsymbol{\beta}_l^T\mathbf{X}_i)\mathbb{G}_{il}\varepsilon_i\right]^2 + o_p(n^{-1})\\ =&O_p(n^{-1}).\end{aligned}$$

□

Lemma S.3. *Let assumptions* (A1) and (A4) *be satisfied. There exists a constant* $0 \le c_D \le \infty$, *such that for any* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and for large enough* n,

$$\|n^{-1}\widehat{\mathbf{U}}(\tilde{\mathbf{Z}},\boldsymbol{\beta})^{-1}\mathbf{D}(\tilde{\mathbf{Z}},\boldsymbol{\beta})\mathbf{1}_n\|_\infty \le C_D.$$

Proof of Lemma S.3: Let $S_n = \widehat{\mathbf{U}}(\tilde{\mathbf{Z}},\boldsymbol{\beta})$ with

$$S_n = \begin{pmatrix} S_{11} & S_{12}\\ S_{21} & S_{22}\end{pmatrix},$$

where $S_{11} = n^{-1}\tilde{\mathbf{Z}}^T\tilde{\mathbf{Z}}$, $S_{12} = S_{21}^T = n^{-1}\tilde{\mathbf{Z}}^T\mathbf{D}(\boldsymbol{\beta})$ and $S_{22} = \widehat{\mathbf{U}}(\boldsymbol{\beta})$. Denote $S_{22.1} = S_{22} - S_{21}S_{11}^{-1}S_{12}$. For any $\boldsymbol{\zeta}$ as given in Lemma S.1, we have

$$\begin{aligned}\boldsymbol{\zeta}^T S_{22.1}\boldsymbol{\zeta} =&\boldsymbol{\zeta}^T\widehat{\mathbf{U}}(\boldsymbol{\beta})\boldsymbol{\zeta} - n^{-2}\boldsymbol{\zeta}^T\mathbf{D}(\boldsymbol{\beta})^T\tilde{\mathbf{Z}}S_{11}^{-1}\tilde{\mathbf{Z}}^T\mathbf{D}(\boldsymbol{\beta})\boldsymbol{\zeta}\\ =&\boldsymbol{\zeta}^T\widehat{\mathbf{U}}(\boldsymbol{\beta})\boldsymbol{\zeta} - c_z\boldsymbol{\zeta}^T\widehat{\mathbf{U}}(\boldsymbol{\beta})\boldsymbol{\zeta} = (1-c_z)\boldsymbol{\zeta}^T\widehat{\mathbf{U}}(\boldsymbol{\beta})\boldsymbol{\zeta},\end{aligned}$$

which is also followed by $\boldsymbol{\zeta}^{-1}S_{22.1}^{-1}\boldsymbol{\zeta} = c_s\boldsymbol{\zeta}^T\widehat{\mathbf{U}}(\boldsymbol{\beta})^{-1}\boldsymbol{\zeta}$, where c_z and c_s are constants. Thus, we have

$$\begin{aligned}&\|n^{-1}(\mathbf{0}_{2J_n\times 2q}, I_{2J_n})\widehat{\mathbf{U}}(\tilde{\mathbf{Z}},\boldsymbol{\beta})^{-1}\mathbf{D}(\tilde{\mathbf{Z}},\boldsymbol{\beta})^T\mathbf{1}_n\|_\infty\\ =&\|n^{-1}\begin{pmatrix}-S_{22.1}^{-1}S_{21}S_{11}^{-1} & S_{22.1}^{-1}\end{pmatrix}\mathbf{D}(\tilde{\mathbf{Z}},\boldsymbol{\beta})^T\mathbf{1}_n\|_\infty\\ =&\|n^{-1}(-S_{22.1}^{-1}S_{21}S_{11}^{-1}\tilde{\mathbf{Z}}^T + S_{22.1}^{-1}\mathbf{D}(\boldsymbol{\beta})^T)\mathbf{1}_n\|_\infty\\ \le&\|n^{-1}S_{22.1}^{-1}\mathbf{D}(\boldsymbol{\beta})^T\tilde{\mathbf{Z}}S_{11}^{-1}\tilde{\mathbf{Z}}^T\mathbf{1}_n\|_\infty + \|n^{-1}S_{22.1}^{-1}\mathbf{D}(\boldsymbol{\beta})^T\mathbf{1}_n\|_\infty\\ \le&c_1\|n^{-1}\widehat{\mathbf{U}}(\boldsymbol{\beta})^{-1}\mathbf{D}(\boldsymbol{\beta})\mathbf{1}_n\|_\infty\\ \le&c_1\|\widehat{\mathbf{U}}(\boldsymbol{\beta})^{-1}\|_\infty\|n^{-1}\mathbf{D}(\boldsymbol{\beta})\mathbf{1}_n\|_\infty,\end{aligned}$$

where c_1 is a constant. By Bernstein's inequality in Bosq (1998), it can be shown that $||n^{-1}\sum_{i=1}^n \mathbf{D}_i(\boldsymbol{\beta}^0)||_\infty = O_p(J_n^{-1})$. We have $\|n^{-1}(\mathbf{0}_{2J_n\times 2q}, I_{2J_n})\widehat{\mathbf{U}}(\tilde{\mathbf{Z}},\boldsymbol{\beta})^{-1}\mathbf{D}(\tilde{\mathbf{Z}},\boldsymbol{\beta})^T\mathbf{1}_n\|_\infty \le$

c_2, where c_2 is a constant. As the proof above, we have

$$
\begin{aligned}
&\|n^{-1}(I_{2q}, \mathbf{0}_{2q\times 2J_n})\widehat{\mathbf{U}}(\tilde{\mathbf{Z}}, \boldsymbol{\beta})^{-1}\mathbf{D}(\tilde{\mathbf{Z}}, \boldsymbol{\beta})^T\mathbf{1}_n\|_\infty \\
=&\|n^{-1}\left(S_{11}^{-1} + S_{11}^{-1}S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1} \quad -S_{11}^{-1}S_{12}S_{22.1}^{-1}\right)\mathbf{D}(\tilde{\mathbf{Z}}, \boldsymbol{\beta})^T\mathbf{1}_n\|_\infty \\
=&\|n^{-1}\left(S_{11}^{-1}\tilde{\mathbf{Z}}^T + S_{11}^{-1}S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1}\tilde{\mathbf{Z}}^T - S_{11}^{-1}S_{12}S_{22.1}^{-1}\mathbf{D}(\boldsymbol{\beta})^T\right)\mathbf{1}_n\|_\infty \\
\le&\|n^{-1}S_{11}^{-1}\tilde{\mathbf{Z}}^T\mathbf{1}_n\|_\infty + \|n^{-3}S_{11}^{-1}\tilde{\mathbf{Z}}^T\mathbf{D}(\boldsymbol{\beta})S_{22.1}^{-1}\mathbf{D}(\boldsymbol{\beta})^T\tilde{\mathbf{Z}}S_{11}^{-1}\tilde{\mathbf{Z}}^T\mathbf{1}_n\|_\infty \\
&+ \|n^{-2}S_{11}^{-1}\tilde{\mathbf{Z}}^T\mathbf{D}(\boldsymbol{\beta})S_{22.1}^{-1}\mathbf{D}(\boldsymbol{\beta})^T\mathbf{1}_n\|_\infty \\
\le& c_3\|n^{-1}S_{11}^{-1}\tilde{\mathbf{Z}}^T\mathbf{1}_n\|_\infty \\
\le& c_4,
\end{aligned}
$$

combining with above proof, which arrives at the second part of Lemma S.3, where c_3 and c_4 are constants.

□

The following Lemma states the convergence rate of the nonparametric estimators $\tilde{m}_l(u_l, \boldsymbol{\beta}^0)$, $l = 0, 1$ and their first derivatives $\tilde{m}_l'(u_l, \boldsymbol{\beta}^0)$.

Lemma S.4. *Let assumptions* (A1)-(A4) *be satisfied. We have*

(a) under $N \to \infty$ and $nN^{-1} \to \infty$, as $n \to \infty$,

$$|\tilde{m}_l(u_l, \boldsymbol{\beta}^0) - m_l(u_l)| = O_p(\sqrt{N/n} + N^{-r})$$

uniformly for any $u_l \in [a_l, b_l]$, and

$$\left\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\right\|_2 = O_p(\sqrt{N/n} + N^{-r});$$

(b) under $N \to \infty$ and $nN^{-3} \to \infty$, as $n \to \infty$,

$$|\tilde{m}_l'(u_l, \boldsymbol{\beta}^0) - m_l'(u_l)| = O_p(\sqrt{N^3/n} + N^{1-r})$$

uniformly for any $u_l \in [a_l, b_l]$.

Proof of Lemma S.4: According to the result of de Boor (2001), for $m_l(\cdot)$ satisfying assumption (A2), there is a function $m_l^0(u_l) = \mathbf{B}_r(u_l)^T\lambda_l$, such that

$$\sup_{u_l\in[a_l,b_l]} |m_l^0(u_l) - m_l(u_l)| = O(J_n^{-r}). \tag{A.9}$$

Let $\mathbb{B}_r(\mathbf{u}) = (\mathbf{B}_r(u_0)^T, \mathbf{B}_r(u_1)^T)^T$, $\tilde{\mathbb{B}}_r(\mathbf{u}) = (\mathbf{1}_{2q}^T, \mathbb{B}_r(\mathbf{u})^T)^T$. and $\lambda = (\lambda_0, \lambda_1)^T$, where $\mathbf{1}_{2q}$ is $2q$-vector with all elements 1 and $\mathbf{u} = (u_0, u_1)^T$. Let $\Lambda(\boldsymbol{\theta}) = (\boldsymbol{\alpha}^T, \lambda^T)^T$. Thus by Lemma S.3

and (A.9), for each $\mathbf{u} \in [a_0, b_0] \times [a_1, b_1]$,

$$\begin{aligned}\left|\tilde{\mathbb{B}}_r(\mathbf{u})^T(\widehat{\Lambda}_m(\boldsymbol{\theta}^0) - \Lambda(\boldsymbol{\theta}^0))\right| &= \left|n^{-1}\tilde{\mathbb{B}}_r(\mathbf{u})^T\widehat{\mathbf{U}}(\tilde{\mathbf{Z}}, \boldsymbol{\beta}^0)^{-1}\mathbf{D}(\tilde{\mathbf{Z}}, \boldsymbol{\beta}^0)^T(\mathbf{m} + \tilde{\mathbf{Z}}\boldsymbol{\alpha}^0 - \mathbf{D}(\tilde{\mathbf{Z}}, \boldsymbol{\beta}^0)\lambda)\right| \\ &= \left|n^{-1}\tilde{\mathbb{B}}_r(\mathbf{u})^T\widehat{\mathbf{U}}(\tilde{\mathbf{Z}}, \boldsymbol{\beta}^0)^{-1}\mathbf{D}(\tilde{\mathbf{Z}}, \boldsymbol{\beta}^0)^T(\mathbf{m} - \mathbf{D}(\boldsymbol{\beta}^0)\lambda)\right| \\ &\leq \left|2q + \mathbb{B}_r(\mathbf{u})^T\mathbf{1}_{2J_n}\right| \|n^{-1}\widehat{\mathbf{U}}(\tilde{\mathbf{Z}}, \boldsymbol{\beta}^0)^{-1}\mathbf{D}(\tilde{\mathbf{Z}}, \boldsymbol{\beta}^0)^T\mathbf{1}_n\|_\infty O(J_n^{-r}) \\ &= O_p(J_n^{-r}).\end{aligned} \tag{A.10}$$

Furthermore, for each $\mathbf{u} \in [a_0, b_0] \times [a_1, b_1]$, by Lemma S.3 and (A.3) and (A.8) and assumption (A3), with probability approaching 1, we have

$$\begin{aligned}&E\left[\tilde{\mathbb{B}}_r(\mathbf{u})^T\widehat{\Lambda}_e(\boldsymbol{\theta}^0)\middle|\mathbf{X}, \mathbf{Z}, G\right]^2 \\ \leq& E\left[C'_D\mathbf{1}_{2q}^T(\tilde{\mathbf{Z}}^T\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}^T\mathbf{e}\middle|\mathbf{X}, \mathbf{Z}, G\right]^2 \\ &+ E\left[\left|C_Dn^{-1}\mathbb{B}_r(\mathbf{u})^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\mathbf{e}\right|\middle|\mathbf{X}, \mathbf{Z}, G\right]^2 \\ =&O_p(1/n) + C_D^2E\left[n^{-1}\mathbb{B}_r(\mathbf{u})^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\mathbf{e}\middle|\mathbf{X}, \mathbf{Z}, G\right]^2 \\ =&n^{-2}C_D^2\mathbb{B}_r(\mathbf{u})^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^TE[\mathbf{e}^{\otimes 2}|\mathbf{X}, \mathbf{Z}, G]\mathbf{D}(\boldsymbol{\beta}^0)\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbb{B}_r(\mathbf{u}) + O_p(1/n) \\ =&n^{-1}C_D^2C_\sigma\mathbb{B}_r(\mathbf{u})^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbb{B}_r(\mathbf{u}) + O_p(1/n) \\ =&O_p(J_n/n),\end{aligned} \tag{A.11}$$

which implies by the law of large numbers that for each $\mathbf{u} \in [a_0, b_0] \times [a_1, b_1]$, $\left\|\tilde{\mathbb{B}}_r(\mathbf{u})\widehat{\Lambda}_e\right\|_2 = O_p(\sqrt{J_n/n})$. Thus, combining (A.10) and (A.11), we have, uniformly for each $\mathbf{u} \in [a_0, b_0] \times [a_1, b_1]$,

$$\left|\tilde{\mathbb{B}}_r(\mathbf{u})^T\widehat{\Lambda}(\boldsymbol{\theta}^0) - \tilde{\mathbb{B}}_r(\mathbf{u})^T\Lambda(\boldsymbol{\theta}^0)\right| = O_p(\sqrt{J_n/n} + J_n^{-r}),$$

which, by (A.9), leads to Lemma S.5 (a). Noting that $\|\mathbf{W}_1\|_\infty = O(J_n)$ where $\mathbf{W}_1$ is defined in (A.1), one can show similarly that the second part of Lemma S.4 holds.
□

Let $\bar{\mathbf{X}}_k = (X_{1k}, \cdots, X_{nk})^T$ and

$$\widehat{\boldsymbol{\zeta}}_k^X = \underset{\boldsymbol{\zeta}_k^X \in \mathcal{R}^{2J_n}}{\arg\min} \sum_{i=1}^n \left\|X_{ik} - D_i(\boldsymbol{\beta}^0)\boldsymbol{\zeta}_k^X\right\|_2^2.$$

It is obvious that $\widehat{\boldsymbol{\zeta}}_k = n^{-1}\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\bar{\mathbf{X}}_k$ by ordinary least squares. Let $\widehat{\boldsymbol{\zeta}}^X = (\widehat{\boldsymbol{\zeta}}_1^X, \cdots, \widehat{\boldsymbol{\zeta}}_p^X)$ be a $2J_n \times p$ matrix and $\mathbf{P}_n(\mathbf{X}_i) = D_i(\boldsymbol{\beta}^0)^T\widehat{\boldsymbol{\zeta}}^X$. As the same way, we can define $\mathbf{P}_n(\mathbf{Z}_i) = D_i(\boldsymbol{\beta}^0)^T\widehat{\boldsymbol{\zeta}}^Z$, where $\widehat{\boldsymbol{\zeta}}^Z = (\widehat{\boldsymbol{\zeta}}_1^Z, \cdots, \widehat{\boldsymbol{\zeta}}_q^Z)$ is a $2J_n \times q$ matrix and

$$\widehat{\boldsymbol{\zeta}}_k^Z = \underset{\boldsymbol{\zeta}_k^Z \in \mathcal{R}^{2J_n}}{\arg\min} \sum_{i=1}^n \left\|Z_{ik} - D_i(\boldsymbol{\beta}^0)\boldsymbol{\zeta}_k^Z\right\|_2^2.$$

Introduce the shorthand notations

$$\begin{aligned}\mathbf{X}_{m,i} =&(m_0'(\mathbf{X}_i^T\boldsymbol{\beta}_0^0)\mathbf{X}_i^T, m_1'(\mathbf{X}_i^T\boldsymbol{\beta}_1^0)G_i\mathbf{X}_i^T)^T,\\ \mathbf{X}_{\widehat{m},i} =&(\mathbf{B}_r'(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0)^T\widehat{\lambda}_0(\widehat{\boldsymbol{\theta}})\mathbf{X}_i^T, \mathbf{B}_r'(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1)^T\widehat{\lambda}_1(\widehat{\boldsymbol{\theta}})G_i\mathbf{X}_i^T)^T,\\ \mathbb{X}_m =&(\mathbf{X}_{m,1}^T, \cdots, \mathbf{X}_{m,n}^T)^T,\\ \mathbb{X}_{\widehat{m}} =&(\mathbf{X}_{\widehat{m},1}^T, \cdots, \mathbf{X}_{\widehat{m},n}^T)^T.\end{aligned}$$

Let $\tilde{\mathbf{P}}_n(\mathbf{X}_i) = (m_i'(\mathbf{X}_i^T\boldsymbol{\beta}_0^0)\mathbf{P}_n(\mathbf{X}_i)^T, m_i'(\mathbf{X}_i^T\boldsymbol{\beta}_0^0)G_i\mathbf{P}_n(\mathbf{X}_i)^T)^T$, $\tilde{\mathbf{P}}_n(\mathbf{Z}_i) = (\mathbf{P}_n(\mathbf{Z}_i)^T, G_i\mathbf{P}_n(\mathbf{Z}_i)^T)^T$, $\tilde{\mathbf{P}}(\mathbf{X}_i) = (m_i'(\mathbf{X}_i^T\boldsymbol{\beta}_0^0)\mathbf{P}(\mathbf{X}_i)^T, m_i'(\mathbf{X}_i^T\boldsymbol{\beta}_0^0)G_i\mathbf{P}(\mathbf{X}_i)^T)^T$ and $\tilde{\mathbf{P}}(\mathbf{Z}_i) = (\mathbf{P}(\mathbf{Z}_i)^T, G_i\mathbf{P}(\mathbf{Z}_i)^T)^T$

Lemma S.5. *Let assumptions* (A1)-(A4) *be satisfied, and* $nN^{-4}\to\infty$ *and* $nN^{-2r-2}\to 0$*, as* $n\to\infty$*. We have*

$$\begin{aligned}D_i(\boldsymbol{\beta}^0)^T\{\widehat{\lambda}(\widehat{\boldsymbol{\beta}})-\lambda\} =&n^{-1}D_i(\boldsymbol{\beta}^0)^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\mathbf{e}\\ &-\tilde{\mathbf{P}}(\mathbf{X}_i)^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0)-\tilde{\mathbf{P}}(\mathbf{Z}_i)^T(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)\\ &+o_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|_2)+o_p(\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\|_2)+o_p(n^{-1/2}),\end{aligned}\tag{A.12}$$

where λ *is defined in Lemma S.4.*

Proof of Lemma S.5: The estimate of $\lambda(\boldsymbol{\theta})$ solves equation

$$0 = n^{-1}\sum_{i=1}^n(Y_i-\tilde{\mathbf{Z}}_i^T\widehat{\boldsymbol{\alpha}}-D_i(\widehat{\boldsymbol{\beta}})^T\widehat{\lambda}(\widehat{\boldsymbol{\beta}}))D_i(\widehat{\boldsymbol{\beta}}).$$

Recalling $\widehat{\mathbf{U}}(\boldsymbol{\beta}^0) = O_p(J_n^{-1})$ and $nN^{-2r-2}\to 0$, via Taylor series, we have

$$\begin{aligned}0 =&n^{-1}\sum_{i=1}^n D_i(\boldsymbol{\beta}^0)\{Y_i-\tilde{\mathbf{Z}}_i^T\boldsymbol{\alpha}-D_i(\boldsymbol{\beta}^0)^T\lambda\}\\ &-n^{-1}\sum_{i=1}^n D_i(\boldsymbol{\beta}^0)D_i(\boldsymbol{\beta}^0)^T\{(\widehat{\lambda}(\widehat{\boldsymbol{\beta}})-\lambda)+o_p(\sqrt{N/n}+N^{-r})\}\\ &-n^{-1}\sum_{i=1}^n D_i(\boldsymbol{\beta}^0)X_{\widehat{m},i}^T\{(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0)+o_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|_2)\}\\ &-n^{-1}\sum_{i=1}^n D_i(\boldsymbol{\beta}^0)\tilde{\mathbf{Z}}_i^T\{(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)+o_p(\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\|_2)\}\\ =&n^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\mathbf{e}+\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)(\widehat{\lambda}(\widehat{\boldsymbol{\beta}})-\lambda)+o_p(\sqrt{N/n}+N^{-r})\\ &+n^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\mathbb{X}_{\widehat{m}}\{(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0)+o_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|_2\}\\ &+n^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\tilde{\mathbf{Z}}\{(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)+o_p(\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\|_2)\}+o_p(n^{-1/2}).\end{aligned}$$

Thus, we have

$$\begin{aligned}&D_i(\boldsymbol{\beta}^0)^T\{\widehat{\lambda}(\widehat{\boldsymbol{\beta}})-\lambda\}\\ =&n^{-1}D_i(\boldsymbol{\beta}^0)^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\mathbf{e}\\ &+n^{-1}D_i(\boldsymbol{\beta}^0)^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\mathbb{X}_{\widehat{m}}\{(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0)+o_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|_2\}\\ &+n^{-1}D_i(\boldsymbol{\beta}^0)^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\tilde{\mathbf{Z}}\{(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)+o_p(\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\|_2\}+o_p(n^{-1/2}).\end{aligned}\tag{A.13}$$

Along the same arguments of Ma and Song (2015), we have

$$\begin{aligned} n^{-1}D_i(\boldsymbol{\beta}^0)^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T m_0'(\mathbf{X}_i^T\boldsymbol{\beta}_0^0)\mathbb{X} &= m_0'(\mathbf{X}_i^T\boldsymbol{\beta}_0^0)\mathbf{P}_n(\mathbf{X}_i) + O_p(J_n^{1-r}), \\ n^{-1}D_i(\boldsymbol{\beta}^0)^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T m_1'(\mathbf{X}_i^T\boldsymbol{\beta}_1^0)\mathbb{X}\mathbf{G} &= m_1'(\mathbf{X}_i^T\boldsymbol{\beta}_1^0)G_i\mathbf{P}_n(\mathbf{X}_i) + O_p(J_n^{1-r}), \end{aligned}$$

which result in

$$D_i(\boldsymbol{\beta}^0)^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\mathbb{X}_{\widehat{m}} = \tilde{\mathbf{P}}_n(\mathbf{X}_i) + O_p(J_n^{1-r}).$$

Similarly, we have

$$D_i(\boldsymbol{\beta}^0)^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\tilde{\mathbf{Z}} = \tilde{\mathbf{P}}_n(\mathbf{Z}_i) + O_p(J_n^{1-r}).$$

Similar as Lemma S.4, we can show that $\|\mathbf{P}_n(\mathbf{X}_i) - \mathbf{P}(\mathbf{X}_i)\|_\infty = O_p(\sqrt{N/n} + N^{-r})$ and $\|\mathbf{P}_n(\mathbf{Z}_i) - \mathbf{P}(\mathbf{Z}_i)\|_\infty = O_p(\sqrt{N/n} + N^{-r})$ which implies that $\tilde{\mathbf{P}}_n(\mathbf{X}_i) = \tilde{\mathbf{P}}(\mathbf{X}_i) + O_p(\sqrt{N/n} + N^{-r})$ and $\tilde{\mathbf{P}}_n(\mathbf{Z}_i) = \tilde{\mathbf{P}}(\mathbf{Z}_i) + O_p(\sqrt{N/n} + N^{-r})$. Together with (A.13), this leads to the result of Lemma S.5.
□

Lemma S.6. *Let assumptions* (A1)-(A5) *be satisfied, and* $nN^{-4} \to \infty$ *and* $nN^{-2r-2} \to 0$, *as* $n \to \infty$. *We have*

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = \{E[\phi(\mathbf{V}, \boldsymbol{\beta}^0)^{\otimes 2}]\}^{-1} n^{-1/2} \sum_{i=1}^n \boldsymbol{\Psi}_i(\mathbf{X}, \mathbf{Z})\varepsilon_i(1 + O_p(J_n^{1-r})) + o_p(n^{-1/2}) \qquad \text{(A.14)}$$

where $\phi(\mathbf{V}, \boldsymbol{\beta}^0)$ *is defined in Section 2.4 and*

$$\boldsymbol{\Psi}_i(\mathbf{X}, \mathbf{Z}) = \begin{pmatrix} \tilde{Z}_i - \tilde{\mathbf{P}}(\mathbf{Z}_i) \\ X_{\widehat{m},i} - \tilde{\mathbf{P}}(\mathbf{X}_i) \end{pmatrix}.$$

Proof of Lemma S.6: Let τ_0 and τ_1 be the Lagrange multipliers, the estimates of $\boldsymbol{\beta}$ solve the following equation

$$0 = \begin{pmatrix} \tau_0\widehat{\boldsymbol{\beta}}_0 \\ \tau_1\widehat{\boldsymbol{\beta}}_1 \end{pmatrix} + n^{-1}\sum_{i=1}^n \left\{Y_i - \tilde{\mathbf{Z}}_i^T\widehat{\boldsymbol{\alpha}}_0 - D_i(\widehat{\boldsymbol{\beta}})^T\widehat{\lambda}(\widehat{\boldsymbol{\beta}})\right\} \begin{pmatrix} D_i'(\widehat{\boldsymbol{\beta}})^T\widehat{\lambda}_0(\widehat{\boldsymbol{\beta}})\mathbf{X}_i \\ D_i'(\widehat{\boldsymbol{\beta}})^T\widehat{\lambda}_1(\widehat{\boldsymbol{\beta}})G_i\mathbf{X}_i \end{pmatrix}.$$

By Taylor expansion, we obtain

$$
\begin{aligned}
0 &= \begin{pmatrix} \tau_0 \widehat{\boldsymbol\beta}_0 \\ \tau_1 \widehat{\boldsymbol\beta}_1 \end{pmatrix} + n^{-1} \sum_{i=1}^n \left\{ Y_i - \tilde{\mathbf{Z}}_i^T \boldsymbol\alpha^0 - D_i(\boldsymbol\beta^0)^T \lambda(\boldsymbol\theta^0) \right\} \mathbf{X}_{\widehat{m},i} \\
&\quad - n^{-1} \mathbb{X}_{\widehat{m}}^T \mathbb{X}_m \{ (\widehat{\boldsymbol\beta} - \boldsymbol\beta^0) + o_p(\|\widehat{\boldsymbol\beta} - \boldsymbol\beta^0\|_2) \} \\
&\quad - n^{-1} \mathbb{X}_{\widehat{m}}^T \tilde{\mathbf{Z}} \{ (\widehat{\boldsymbol\alpha} - \boldsymbol\alpha^0) + o_p(\|\widehat{\boldsymbol\alpha} - \boldsymbol\alpha^0\|_2) \} \\
&\quad - n^{-1} \sum_{i=1}^n X_{\widehat{m},i} D_i(\boldsymbol\beta^0)^T \{ \widehat\lambda(\widehat{\boldsymbol\beta}) - \lambda(\boldsymbol\theta^0) + o_p(\sqrt{N/n} + N^{-r}) \} \\
&= \begin{pmatrix} \tau_0 \widehat{\boldsymbol\beta}_0 \\ \tau_1 \widehat{\boldsymbol\beta}_1 \end{pmatrix} + n^{-1} \mathbb{X}_{\widehat{m}}^T \mathbf{e} - n^{-1} \mathbb{X}_{\widehat{m}}^T \mathbb{X}_m \{ (\widehat{\boldsymbol\beta} - \boldsymbol\beta^0) + o_p(\|\widehat{\boldsymbol\beta} - \boldsymbol\beta^0\|_2) \} \\
&\quad - n^{-1} \mathbb{X}_{\widehat{m}}^T \tilde{\mathbf{Z}} \{ (\widehat{\boldsymbol\alpha} - \boldsymbol\alpha^0) + o_p(\|\widehat{\boldsymbol\alpha} - \boldsymbol\alpha^0\|_2) \} \\
&\quad - n^{-2} \sum_{i=1}^n X_{\widehat{m},i} D_i(\boldsymbol\beta^0)^T \widehat{\mathbf{U}}(\boldsymbol\beta^0)^{-1} \mathbf{D}(\boldsymbol\beta^0)^T \mathbf{e} \\
&\quad + n^{-1} \sum_{i=1}^n X_{\widehat{m},i} \tilde{\mathbf{P}}(\mathbf{X}_i)^T \{ (\widehat{\boldsymbol\beta} - \boldsymbol\beta^0) + o_p(\|\widehat{\boldsymbol\beta} - \boldsymbol\beta^0\|_2) \} \\
&\quad + n^{-1} \sum_{i=1}^n X_{\widehat{m},i} \tilde{\mathbf{P}}(\mathbf{Z}_i)^T \{ (\widehat{\boldsymbol\alpha} - \boldsymbol\alpha^0) + o_p(\|\widehat{\boldsymbol\alpha} - \boldsymbol\alpha^0\|_2) \} \\
&= \begin{pmatrix} \tau_0 \widehat{\boldsymbol\beta}_0 \\ \tau_1 \widehat{\boldsymbol\beta}_1 \end{pmatrix} - n^{-1} \sum_{i=1}^n X_{\widehat{m},i} \left[X_{m,i} - \tilde{\mathbf{P}}(\mathbf{X}_i) \right]^T \{ (\widehat{\boldsymbol\beta} - \boldsymbol\beta^0) + o_p(\|\widehat{\boldsymbol\beta} - \boldsymbol\beta^0\|_2) \} \\
&\quad - n^{-1} \sum_{i=1}^n X_{\widehat{m},i} \left[\tilde{Z}_i - \tilde{\mathbf{P}}(\mathbf{Z}_i) \right]^T \{ (\widehat{\boldsymbol\alpha} - \boldsymbol\alpha^0) + o_p(\|\widehat{\boldsymbol\alpha} - \boldsymbol\alpha^0\|_2) \} \\
&\quad + n^{-1} \mathbb{X}_{\widehat{m}}^T \mathbf{e} - n^{-2} \sum_{j=1}^n \varepsilon_j D_j(\boldsymbol\beta^0) \widehat{\mathbf{U}}(\boldsymbol\beta^0)^{-1} \sum_{i=1}^n D_i(\boldsymbol\beta^0)^T X_{\widehat{m},i} + o_p(n^{-1/2}),
\end{aligned}
$$

$$
\begin{aligned}
&= \begin{pmatrix} \tau_0 \widehat{\boldsymbol\beta}_0 \\ \tau_1 \widehat{\boldsymbol\beta}_1 \end{pmatrix} - n^{-1} \sum_{i=1}^n X_{\widehat{m},i} \left[X_{m,i} - \tilde{\mathbf{P}}(\mathbf{X}_i) \right]^T \{ (\widehat{\boldsymbol\beta} - \boldsymbol\beta^0) + o_p(\|\widehat{\boldsymbol\beta} - \boldsymbol\beta^0\|_2) \} \\
&\quad - n^{-1} \sum_{i=1}^n X_{\widehat{m},i} \left[\tilde{Z}_i - \tilde{\mathbf{P}}(\mathbf{Z}_i) \right]^T \{ (\widehat{\boldsymbol\alpha} - \boldsymbol\alpha^0) + o_p(\|\widehat{\boldsymbol\alpha} - \boldsymbol\alpha^0\|_2) \} \\
&\quad + n^{-1} \sum_{i=1}^n \left[X_{\widehat{m},i} - \tilde{\mathbf{P}}(\mathbf{X}_i) \right] \varepsilon_i \{ 1 + O_p(J_n^{1-r}) \} + o_p(n^{-1/2}).
\end{aligned}
$$

Similarly, the estimates of $\boldsymbol\alpha$ solve the following equation

$$
0 = n^{-1} \sum_{i=1}^n \left\{ Y_i - \tilde{\mathbf{Z}}_i^T \widehat{\boldsymbol\alpha} - D_i(\widehat{\boldsymbol\beta})^T \widehat\lambda(\widehat{\boldsymbol\beta}) \right\} \tilde{\mathbf{Z}}_i.
$$

Then by Taylor expansion, we obtain

$$
\begin{aligned}
0 =& n^{-1}\sum_{i=1}^{n}\left\{Y_i - \tilde{\mathbf{Z}}_i^T\boldsymbol{\alpha}^0 - D_i(\boldsymbol{\beta}^0)^T\lambda(\boldsymbol{\theta}^0)\right\}\tilde{\mathbf{Z}}_i \\
& - n^{-1}\tilde{\mathbf{Z}}^T\mathbb{X}_{\widehat{m}}\{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + o_p(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_2)\} \\
& - n^{-1}\tilde{\mathbf{Z}}^T\tilde{\mathbf{Z}}\{(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) + o_p(\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\|_2)\} \\
& - n^{-1}\sum_{i=1}^{n}\tilde{Z}_i D_i(\boldsymbol{\beta}^0)^T\{\widehat{\lambda}(\widehat{\boldsymbol{\beta}}) - \lambda + o_p(\sqrt{N/n} + N^{-r})\} \\
=& n^{-1}\tilde{\mathbf{Z}}^T\mathbf{e} - n^{-1}\tilde{\mathbf{Z}}^T\mathbb{X}_{\widehat{m}}\{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + o_p(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_2)\} \\
& - n^{-1}\tilde{\mathbf{Z}}^T\tilde{\mathbf{Z}}\{(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) + o_p(\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\|_2)\} \\
& - n^{-2}\sum_{i=1}^{n}\tilde{Z}_i D_i(\boldsymbol{\beta}^0)^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\mathbf{e} \\
& + n^{-1}\sum_{i=1}^{n}\tilde{Z}_i\tilde{\mathbf{P}}(\mathbf{X}_i)^T\{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + o_p(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_2)\} \\
& + n^{-1}\sum_{i=1}^{n}\tilde{Z}_i\tilde{\mathbf{P}}(\mathbf{Z}_i)^T\{(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) + o_p(\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\|_2)\} \\
=& - n^{-1}\sum_{i=1}^{n}\tilde{Z}_i\left[X_{\widehat{m},i} - \tilde{\mathbf{P}}(\mathbf{X}_i)\right]^T\{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + o_p(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_2)\} \\
& - n^{-1}\sum_{i=1}^{n}\tilde{Z}_i\left[\tilde{Z}_i - \tilde{\mathbf{P}}(\mathbf{Z}_i)\right]^T\{(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) + o_p(\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\|_2)\} \\
& + n^{-1}\tilde{\mathbf{Z}}^T\mathbf{e} - n^{-2}\sum_{j=1}^{n}\varepsilon_j D_j(\boldsymbol{\beta}^0)\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\sum_{i=1}^{n}D_i(\boldsymbol{\beta}^0)^T\tilde{Z}_i + o_p(n^{-1/2}), \\
=& - n^{-1}\sum_{i=1}^{n}\tilde{Z}_i\left[X_{\widehat{m},i} - \tilde{\mathbf{P}}(\mathbf{X}_i)\right]^T\{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + o_p(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_2)\} \\
& - n^{-1}\sum_{i=1}^{n}\tilde{Z}_i\left[\tilde{Z}_i - \tilde{\mathbf{P}}(\mathbf{Z}_i)\right]^T\{(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) + o_p(\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\|_2)\} \\
& + n^{-1}\sum_{i=1}^{n}\left[\tilde{Z}_i - \tilde{\mathbf{P}}(\mathbf{Z}_i)\right]\varepsilon_i\{1 + O_p(J_n^{1-r})\} + o_p(n^{-1/2}),
\end{aligned}
$$

which leads to

$$
\begin{aligned}
&\begin{pmatrix} \mathbf{0} \\ \tau_0\widehat{\boldsymbol{\beta}}_0 \\ \tau_1\widehat{\boldsymbol{\beta}}_1 \end{pmatrix} + n^{-1}\sum_{i=1}^{n}\begin{pmatrix} \tilde{Z}_i \\ X_{\widehat{m},i} \end{pmatrix}\begin{pmatrix} \tilde{Z}_i - \tilde{\mathbf{P}}(\mathbf{Z}_i) \\ X_{\widehat{m},i} - \tilde{\mathbf{P}}(\mathbf{X}_i) \end{pmatrix}^T\{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) + o_p(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|_2)\} \\
=& n^{-1}\sum_{i=1}^{n}\begin{pmatrix} \tilde{Z}_i - \tilde{\mathbf{P}}(\mathbf{Z}_i) \\ X_{\widehat{m},i} - \tilde{\mathbf{P}}(\mathbf{X}_i) \end{pmatrix}\varepsilon_i\{1 + O_p(J_n^{1-r})\} + o_p(n^{-1/2}).
\end{aligned}
$$

Let

$$P_{\boldsymbol\beta} \equiv \begin{pmatrix} I & 0 & 0 \\ 0 & I-\widehat{\boldsymbol\beta}_0\widehat{\boldsymbol\beta}_0^T & 0 \\ 0 & 0 & I-\widehat{\boldsymbol\beta}_1\widehat{\boldsymbol\beta}_1^T \end{pmatrix} = \begin{pmatrix} 0 & 0 & I \\ 0 & I-\boldsymbol\beta_0^0(\boldsymbol\beta_0^0)^T & 0 \\ 0 & 0 & I-\boldsymbol\beta_1^0(\boldsymbol\beta_1^0)^T \end{pmatrix} + o_p(1).$$

According to the constraints $\|\boldsymbol\beta_0\|_2 = 1$ and $\|\boldsymbol\beta_1\|_2 = 1$, we have

$$\begin{aligned} n^{-1}P_{\boldsymbol\beta}\sum_{i=1}^n \begin{pmatrix} \tilde Z_i \\ X_{\widehat m,i}\end{pmatrix}\begin{pmatrix} \tilde Z_i - \tilde{\mathbf{P}}(\mathbf{Z}_i) \\ X_{\widehat m,i} - \tilde{\mathbf{P}}(\mathbf{X}_i)\end{pmatrix}^T \{(\widehat{\boldsymbol\theta}-\boldsymbol\theta^0) + o_p(\|\widehat{\boldsymbol\theta}-\boldsymbol\theta^0\|_2)\} \\ = n^{-1}P_{\boldsymbol\beta}\sum_{i=1}^n \begin{pmatrix} \tilde Z_i - \tilde{\mathbf{P}}(\mathbf{Z}_i) \\ X_{m,i} - \tilde{\mathbf{P}}(\mathbf{X}_i)\end{pmatrix}\varepsilon_i\{1+O_p(J_n^{1-r})\} + o_p(n^{-1/2}). \end{aligned}$$

By Lemma S.4, it can be shown that $\|X_{\widehat m,i} - X_{m,i}\| = O_p(\sqrt{N^3/n} + N^{1-r})$. Thus, by the law of large numbers, we have

$$n^{-1}\sum_{i=1}^n \begin{pmatrix} \tilde Z_i \\ X_{\widehat m,i}\end{pmatrix}\begin{pmatrix} \tilde Z_i - \tilde{\mathbf{P}}(\mathbf{Z}_i) \\ X_{m,i} - \tilde{\mathbf{P}}(\mathbf{X}_i)\end{pmatrix}^T = E\left\{\boldsymbol\phi(\mathbf{V},\boldsymbol\beta^0)^{\otimes 2}\right\} + O_p(\sqrt{N^3/n} + N^{1-r}) + O_p(n^{-2}).$$

This completes the proof of Lemma S.6.

□

Proof of Theorem 1: Because the observations $\mathbf{V}_1, \cdots, \mathbf{V}_n$ are independent, by Lindeberg-Feller central limit theorem, it is easy to prove that

$$n^{-1/2}\sum_{i=1}^n \begin{pmatrix} \tilde Z_i \\ X_{m,i} - \tilde{\mathbf{P}}(\mathbf{X}_i) - \tilde{\mathbf{P}}(\mathbf{Z}_i)\end{pmatrix}\varepsilon_i \xrightarrow{\mathcal{L}} N(\mathbf{0}, \tilde\Sigma),$$

where $\tilde\Sigma = E\left[\sigma(\mathbf{V})^2\boldsymbol\phi(\mathbf{V},\boldsymbol\beta^0)^{\otimes 2}\right]$. Then combining Lemma S.6, the proof of Theorem 1 can be completed by Slusky's theorem.

□

Proof of Theorem 2: Because $B_{s,l}(u_l)$ $(s = 1, \cdots, J_n, l = 0, 1)$ have the banded first derivatives for any $u_l \in [a_l, b_l]$, by (A.13), (A.11) and Theorem 1, we have, for any $u_l \in [a_l, b_l]$,

$$\begin{aligned} |\tilde m_l(u_l, \widehat{\boldsymbol\beta}) - \tilde m_l(u_l, \boldsymbol\beta^0)| =& |D(\widehat{\boldsymbol\beta})^T\widehat\lambda(\widehat{\boldsymbol\beta}) - D(\boldsymbol\beta^0)^T\lambda(\boldsymbol\theta^0)| \\ \le& |D(\boldsymbol\beta^0)^T\{\widehat\lambda(\widehat{\boldsymbol\beta}) - \lambda(\boldsymbol\theta^0)\}| + |\{D(\widehat{\boldsymbol\beta}) - D(\boldsymbol\beta^0)\}^T\widehat\lambda(\widehat{\boldsymbol\beta})| \\ \le& |n^{-1}D(\boldsymbol\beta^0)^T\widehat{\mathbf{U}}(\boldsymbol\beta^0)^{-1}\mathbf{D}(\boldsymbol\beta^0)^T\mathbf{e}| + O_p(n^{-1/2}) \\ =& O_p(\sqrt{N/n}). \end{aligned}$$

Then, combining Lemma S.4, we have

$$\begin{aligned} \sup_{u_l\in[a_l,b_l]}|\tilde m_l(u_l, \widehat{\boldsymbol\beta}) - m_l(u_l)| \le& \sup_{u_l\in[a_l,b_l]}|\tilde m_l(u_l, \widehat{\boldsymbol\beta}) - \tilde m_l(u_l, \boldsymbol\beta^0)| + \sup_{u_l\in[a_l,b_l]}|\tilde m_l(u_l, \boldsymbol\beta^0) - m_l(u_l)| \\ =& O_p(\sqrt{N/n} + N^{-r}). \end{aligned}$$

This completes the proof of Theorem 2.

□

Lemma S.7. *Let assumptions* (A1)-(A6) *be satisfied, and* $nN^{-4} \to \infty$ *and* $nN^{-2r-2} \to 0$*, as* $n \to \infty$*. We have*

$$\sup_{u_l \in [a_l, b_l]} \left| \widehat{m}_l^O(u_l, \widehat{\boldsymbol{\beta}}) - m_l(u_l) \right| = O_p(n^{-2/5}\sqrt{\log n}),$$

and for any $u_l \in [a_l, b_l]$,

$$\sqrt{nh_l}\left\{ \widehat{m}_l^O(u_l, \widehat{\boldsymbol{\beta}}) - m_l(u_l) - b_l(u_l)h_l^2 \right\} \xrightarrow{\mathcal{L}} N(0, v_l(u_l)),$$

where

$$\begin{aligned} b_l(u_l) =& \mu_1 m_l''(u_l)/2, \\ v_l(u_l) =& \{E[\tilde{G}_l^2|u_l]\}^{-2}\|K\|_2^2 E[\tilde{G}_l^2\sigma^2(\mathbf{v})|u_l] f_l(u_l)^{-1}. \end{aligned}$$

and $\tilde{G}_0 = 1$ *and* $\tilde{G}_1 = G$.

Proof of Lemma S.7: Noting that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\| = O_p(n^{-1/2})$ by Theorem 1, it is not hard to see that, for any $u_l \in [a_l, b_l]$, $l = 0, 1$,

$$\begin{aligned} \mathbf{Y}_l^O(\widehat{\boldsymbol{\theta}}) =& \mathbf{Y}_l^O(\boldsymbol{\theta}^0) + O_p(n^{-1/2}), \\ \mathbf{W}(u_l, \widehat{\boldsymbol{\beta}}) =& \mathbf{W}(u_l, \boldsymbol{\beta}^0) + O_p(n^{-1/2}) \\ \tilde{\mathbf{X}}(u_l, \widehat{\boldsymbol{\beta}}) =& \tilde{\mathbf{X}}(u_l, \boldsymbol{\beta}^0) + O_p(n^{-1/2}), \end{aligned}$$

which implies that

$$\sup_{u_l \in [a_l, b_l]} \left| \widehat{m}_l^O(u_l, \widehat{\boldsymbol{\beta}}) - \widehat{m}_l^O(u_l, \boldsymbol{\beta}^0) \right| = O_p(n^{-1/2}). \tag{A.15}$$

For $l = 1$, given the true parameter $\boldsymbol{\theta}^0$ and true function $m_0(\cdot)$, the estimator

$$\tilde{m}_1^O(u_1, \boldsymbol{\beta}^0) = (1, 0)\{\tilde{\mathbf{X}}(u_1, \boldsymbol{\beta}_1^0)^T \mathbf{W}(u_1, \boldsymbol{\beta}_1^0)\tilde{\mathbf{X}}(u_1, \boldsymbol{\beta}_1^0)\}^{-1}\tilde{\mathbf{X}}(u_1, \boldsymbol{\beta}_1^0)^T \mathbf{W}(u_1, \boldsymbol{\beta}_1^0)\mathbf{Y}_1^O(\boldsymbol{\theta}^0)$$

which is in fact the Nadaraya-Watson estimator based on model $E\{Y_1^O(\boldsymbol{\theta}^0)|\mathbf{V}, \mathbf{X}^T\boldsymbol{\beta}_1^0 = u_1\} = m_1(u_1)$, the same as $\tilde{m}_1^O(u_l, \boldsymbol{\beta}^0)$. By Theorem 2.5 and 2.6 in Li and Racine (2007), we have

$$\begin{aligned} \sup_{u_l \in [a_l, b_l]} \left| \widehat{m}_l^O(u_l, \widehat{\boldsymbol{\beta}}) - m_l(u_l) \right| \le& \sup_{u_l \in [a_l, b_l]} \left| \widehat{m}_l^O(u_l, \widehat{\boldsymbol{\beta}}) - \tilde{m}_l^O(u_l, \boldsymbol{\beta}^0) \right| + \sup_{u_l \in [a_l, b_l]} \left| \widehat{m}_l^O(u_l, \boldsymbol{\beta}^0) - m_l(u_l) \right| \\ =& O_p(n^{-1/2}) + O_p(n^{-2/5}\sqrt{\log n}) \\ =& O_p(n^{-2/5}\sqrt{\log n}), \end{aligned}$$

which leads to the first part of Lemma S.7, and

$$\sqrt{nh_l}\left\{ \tilde{m}_l^O(u_l, \boldsymbol{\beta}^0) - m_l(u_l) - b_l(u_l)h_l^2 \right\} \xrightarrow{\mathcal{L}} N(0, v_l(u_l)).$$

The second part can be shown by combining (A.15) and the asymptotic normality of $\tilde{m}_l^O(u_1, \boldsymbol{\beta}^0)$.
□

To prove Theorem 3, we need additional nations and two more Lemmas. Define

$$\begin{aligned}
\boldsymbol{\Phi}_{l,m}(u_l,\boldsymbol{\theta}) =& n^{-1}\tilde{\mathbf{X}}(u_l,\boldsymbol{\beta}_l)^T\mathbf{W}(u_l,\boldsymbol{\beta}_l)\{\tilde{\mathbf{m}}_{l,m}(\mathbf{X}^T\boldsymbol{\beta}_l,\boldsymbol{\beta})-\mathbf{m}_l(\mathbf{X}^T\boldsymbol{\beta}_l)\},\\
\boldsymbol{\Phi}_{l,e}(u_l,\boldsymbol{\theta}) =& n^{-1}\tilde{\mathbf{X}}(u_l,\boldsymbol{\beta}_l)^T\mathbf{W}(u_l,\boldsymbol{\beta}_l)\tilde{\mathbf{m}}_{l,e}(\mathbf{X}^T\boldsymbol{\beta}_l,\boldsymbol{\beta}),\\
\mathbf{m}_l(\mathbf{X}^T\boldsymbol{\beta}_l) =& (m_l(\mathbf{X}_1^T\boldsymbol{\beta}_l)\cdots,m_l(\mathbf{X}_n^T\boldsymbol{\beta}_l))^T\\
\tilde{\mathbf{m}}_{l,m}(\mathbf{X}^T\boldsymbol{\beta}_l,\boldsymbol{\beta}) =& (\tilde{m}_{l,m}(\mathbf{X}_1^T\boldsymbol{\beta}_l,\boldsymbol{\beta}),\cdots,\tilde{m}_{l,m}(\mathbf{X}_n^T\boldsymbol{\beta}_l,\boldsymbol{\beta}))^T,\\
\tilde{\mathbf{m}}_{l,e}(\mathbf{X}^T\boldsymbol{\beta}_l,\boldsymbol{\beta}) =& (\tilde{m}_{l,e}(\mathbf{X}_1^T\boldsymbol{\beta}_l,\boldsymbol{\beta}),\cdots,\tilde{m}_{l,e}(\mathbf{X}_n^T\boldsymbol{\beta}_l,\boldsymbol{\beta}))^T,
\end{aligned}$$

where $\tilde{m}_{l,m}(\mathbf{X}_i^T\boldsymbol{\beta}_l,\boldsymbol{\beta})$ and $\tilde{m}_{l,e}(\mathbf{X}_i^T\boldsymbol{\beta}_l,\boldsymbol{\beta})$ are defined by

$$\tilde{m}_{l,m}(u_l,\boldsymbol{\beta})=\mathbf{B}_q(u_l)^T\widehat{\lambda}_{l,m}(\boldsymbol{\beta})\quad\text{and}\quad\tilde{m}_{l,e}(u_l,\boldsymbol{\beta})=\mathbf{B}_q(u_l)^T\widehat{\lambda}_{l,e}(\boldsymbol{\beta}).$$

Note that $\tilde{m}_l(u_l,\boldsymbol{\beta})=\tilde{m}_{l,m}(u_l,\boldsymbol{\beta})+\tilde{m}_{l,e}(u_l,\boldsymbol{\beta}),\quad l=0,1.$

Lemma S.8. *Let assumptions* (A1)-(A6) *be satisfied, and* $N\to\infty$*,* $h\to0$*,* $nN^{-1}\to\infty$ *and* $nh\to\infty$*, as* $n\to\infty$*. We have, for* $l=0,1$*,*

$$\sup_{u_l\in[a_l,b_l]}\left|\boldsymbol{\Phi}_{l,m}^{(1)}(u_l,\boldsymbol{\beta}^0)\right|+\sup_{u_l\in[a_l,b_l]}\left|\boldsymbol{\Phi}_{l,m}^{(2)}(u_l,\boldsymbol{\beta}^0)\right|=O_p(J_n^{-r}),$$

where $\boldsymbol{\Phi}_{l,m}^{(1)}(u_l,\boldsymbol{\beta}^0)=(1,0)\boldsymbol{\Phi}_{l,m}(u_l,\boldsymbol{\beta}^0)$ *and* $\boldsymbol{\Phi}_{l,m}^{(2)}(u_l,\boldsymbol{\beta}^0)=(0,1)\boldsymbol{\Phi}_{l,m}(u_l,\boldsymbol{\beta}^0)$*.*

Proof of Lemma S.8: We first prove the case for the varying-index function (i.e., $l=1$). The case for $l=0$ can be proved similarly. By (A.9) and (A.10), we have $\max_{i=1,\cdots,n}|\tilde{m}_{l,m}(\mathbf{X}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0)-m_l(\mathbf{X}_i^T\boldsymbol{\beta}_l^0)|=O_p(J_n^{-r})$, $l=0,1$. It is easy to show by Lemma A.2 in Xia and Li (1999) that

$$\sup_{u_1\in[a_1,b_1]}\left|n^{-1}\sum_{i=1}^n G_iK_{h_1}(\mathbf{X}_i^T\boldsymbol{\beta}_1^0-u_1)\right|=O_p(1).$$

By the definition of $\boldsymbol{\Phi}_{l,m}(u_l,\boldsymbol{\beta}^0)$, we have

$$\begin{aligned}
\sup_{u_1\in[a_1,b_1]}\left|\boldsymbol{\Phi}_{1,m}^{(1)}(u_1,\boldsymbol{\beta}^0)\right|\le&\sup_{u_1\in[a_1,b_1]}\left|n^{-1}\sum_{i=1}^n G_iK_{h_1}(\mathbf{X}_i^T\boldsymbol{\beta}_1^0-u_1)\right|\max_{i=1,\cdots,n}|\tilde{m}_{l,m}(\mathbf{X}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0)-m_l(\mathbf{X}_i^T\boldsymbol{\beta}_l^0)|\\
=&O_p(J_n^{-r}).
\end{aligned}$$

and,

$$\sup_{u_1\in[a_1,b_1]}\left|\boldsymbol{\Phi}_{1,m}^{(2)}(u_1,\boldsymbol{\beta}^0)\right|=O_p(J_n^{-r}).$$

This completes the proof of Lemma S.8.
□

Lemma S.9. *Let assumptions* (A1)-(A6) *be satisfied, and* $N\to\infty$*,* $h\to0$ *and* $\sqrt{N^2\log n/(nh)}\to0$*, as* $n\to\infty$*. We have, for* $l=0,1$*,*

$$\sup_{u_l\in[a_l,b_l]}\left|\boldsymbol{\Phi}_{l,e}^{(1)}(u_l,\boldsymbol{\beta}^0)\right|+\sup_{u_l\in[a_l,b_l]}\left|\boldsymbol{\Phi}_{l,e}^{(2)}(u_l,\boldsymbol{\beta}^0)\right|=O_p(n^{-1/2}),$$

where $\boldsymbol{\Phi}_{l,e}^{(1)}(u_l,\boldsymbol{\beta}^0)=(1,0)\boldsymbol{\Phi}_{l,e}(u_l,\boldsymbol{\beta}^0)$ *and* $\boldsymbol{\Phi}_{l,e}^{(2)}(u_l,\boldsymbol{\beta}^0)=(0,1)\boldsymbol{\Phi}_{l,e}(u_l,\boldsymbol{\beta}^0)$*.*

Proof of Lemma S.9: We first prove the case for the varying-index function (i.e., $l = 1$). The case for $l = 0$ can be proved similarly. Let $\eta(u_1) = (\eta_0^T(u_1), \eta_1^T(u_1))^T$ with

$$\eta_l(u_1) = n^{-1}\sum_{i=1}^{n} G_i K_{h_1}(\mathbf{X}_i^T\boldsymbol{\beta}_1^0 - u_1)B_q(\mathbf{X}_i^T\boldsymbol{\beta}_l^0), l = 0, 1.$$

By Lemma A.2 in Xia and Li (1999), we have

$$\sup_{u_1\in[a_1,b_1]} \|\eta_l(u_1) - E\{\eta_l(u_1)\}\|_\infty = O(\sqrt{\log n/(nh)}).$$

By de Boor (2001), we have, for any $u_1 \in [a_1, b_1]$,

$$\begin{aligned}
\|E\{\eta_1(u_1)\}\|_\infty =&\|E\{K_{h_1}(\mathbf{X}_i^T\boldsymbol{\beta}_1^0 - u_1)E[G_iB_q(\mathbf{X}_i^T\boldsymbol{\beta}_0^0)|\mathbf{X}_i^T\boldsymbol{\beta}_1^0]\}\|_\infty\\
=&\mu_0 f_1(u_1)|E\{G_i|\mathbf{X}_i^T\boldsymbol{\beta}_1^0 = u_1\}|\|E\{B_q(\mathbf{X}_i^T\boldsymbol{\beta}_0^0)\}\|_\infty + O_p(h_1^2)O_p(\|E\{B_q(\mathbf{X}_i^T\boldsymbol{\beta}_0^0)\}\|_\infty)\\
=&O_p(J_n^{-1}).
\end{aligned}$$

similarly, $\|E\{\eta_1(u_1)\}\|_\infty = O_p(J_n^{-1})$ holds. Thus, it can be show that

$$\sup_{u_1\in[a_1,b_1]} \|\eta_l(u_1)\|_\infty = O(J_n^{-1}).$$

By the definition of $\boldsymbol{\Phi}_{l,e}(u_l, \boldsymbol{\beta}^0)$, we have

$$\begin{aligned}
\boldsymbol{\Phi}_{1,e}^{(1)}(u_1, \boldsymbol{\beta}^0) =&n^{-1}\sum_{i=1}^{n} G_iK_{h_1}(\mathbf{X}_i^T\boldsymbol{\beta}_1^0 - u_1)\tilde{m}_{1,e}(\mathbf{X}_i^T\boldsymbol{\beta}_0^0, \boldsymbol{\beta}^0)\\
=&n^{-1}(\mathbf{0}^T, \eta_1(u_1)^T)^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^T\mathbf{e}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\sup_{u_1\in[a_1,b_1]} E\left(\boldsymbol{\Phi}_{1,e}^{(1)}(u_1, \boldsymbol{\beta}^0)|\mathbf{X}, \mathbf{Z}, G\right)^2\\
&\quad= \sup_{u_1\in[a_1,b_1]} n^{-2}(\mathbf{0}^T, \eta_1(u_1)^T)^T\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^TE\{\mathbf{e}\mathbf{e}^T|\mathbf{X}, \mathbf{Z}, G\}\mathbf{D}(\boldsymbol{\beta}^0)\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}(\mathbf{0}^T, \eta_1(u_1)^T)\\
&\quad\le \sup_{u_1\in[a_1,b_1]} n^{-1}C_\sigma\|\eta_1(u_1)\|_2^2\|\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)^{-1}\|_2\\
&\quad=O(n^{-1}),
\end{aligned}$$

which implies that

$$\sup_{u_1\in[a_1,b_1]} \left|\boldsymbol{\Phi}_{1,e}^{(1)}(u_1, \boldsymbol{\beta}^0)\right| = O_p(n^{-1/2}).$$

By the same way, we also have

$$\sup_{u_1\in[a_1,b_1]} \left|\boldsymbol{\Phi}_{1,e}^{(2)}(u_1, \boldsymbol{\beta}^0)\right| = O_p(n^{-1/2}).$$

This completes the proof of Lemma S.9.

□

Proof of Theorem 3: By the estimators in (2.5) and (2.6), we have, for $l=2$,

$$\begin{aligned}&\widehat{m}_1(u_1,\boldsymbol{\beta}^0)-\widehat{m}_1^O(u_1,\boldsymbol{\beta}^0)\\ &=(1,0)\{\tilde{\mathbf{X}}(u_1,\boldsymbol{\beta}_1^0)^T\mathbf{W}(u_1,\boldsymbol{\beta}_1^0)\tilde{\mathbf{X}}(u_1,\boldsymbol{\beta}_1^0)\}^{-1}\tilde{\mathbf{X}}(u_1,\boldsymbol{\beta}_1^0)^T\mathbf{W}(u_1,\boldsymbol{\beta}_1^0)\{\mathbf{Y}_1(\boldsymbol{\theta}^0)-\mathbf{Y}_1^O(\boldsymbol{\theta}^0)\}\\ &=-(1,0)\{n^{-1}\tilde{\mathbf{X}}(u_1,\boldsymbol{\beta}_1^0)^T\mathbf{W}(u_1,\boldsymbol{\beta}_1^0)\tilde{\mathbf{X}}(u_1,\boldsymbol{\beta}_1^0)\}^{-1}\{\boldsymbol{\Phi}_{l,m}(u_1,\boldsymbol{\beta}^0)+\boldsymbol{\Phi}_{l,e}(u_1,\boldsymbol{\beta}^0)\}.\end{aligned}$$

It is obvious by Cai et al (2000) that

$$\sup_{u_1\in[a_1,b_1]}\|\{n^{-1}\tilde{\mathbf{X}}(u_1,\boldsymbol{\beta}_1^0)^T\mathbf{W}(u_1,\boldsymbol{\beta}_1^0)\tilde{\mathbf{X}}(u_1,\boldsymbol{\beta}_1^0)\}^{-1}\|_2=O_p(1).$$

Combining Lemma S.8 and Lemma S.9, this gives

$$\sup_{u_1\in[a_1,b_1]}|\widehat{m}_1(u_1,\boldsymbol{\beta}^0)-\widehat{m}_1^O(u_1,\boldsymbol{\beta}^0)|=O_p(n^{-1/2}+J_n^{-r}).\tag{A.16}$$

Because $\mathbf{B}_r(u_l)$, $l=0,1$, have bounded first derivatives, we have $D(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_l)=D(\mathbf{X}_i^T\boldsymbol{\beta}_l^0)+O_p(n^{-1/2})$ and $\widehat{\mathbf{U}}(\widehat{\boldsymbol{\beta}})=\widehat{\mathbf{U}}(\boldsymbol{\beta}^0)+O_p(n^{-1/2})$ by $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0\|_2=O_p(n^{-1/2})$, which followed by

$$\tilde{m}_l(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_l,\widehat{\boldsymbol{\beta}})=\tilde{m}_l(\mathbf{X}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\theta}^0)+O_p(n^{-1/2}).$$

Similar to (A.15), we have, for $l=0,1$,

$$\sup_{u_l\in[a_l,b_l]}\left|\widehat{m}_l(u_l,\widehat{\boldsymbol{\beta}})-\widehat{m}_l(u_l,\boldsymbol{\beta}^0)\right|=O_p(n^{-1/2}),$$

which combining (A.15) and (A.16) implies that

$$\begin{aligned}\sup_{u_1\in[a_1,b_1]}|\widehat{m}_1(u_1,\widehat{\boldsymbol{\beta}})-\widehat{m}_1^O(u_1,\widehat{\boldsymbol{\beta}})|\le&\sup_{u_1\in[a_1,b_1]}\left|\widehat{m}_1(u_1,\widehat{\boldsymbol{\beta}})-\widehat{m}_1(u_1,\boldsymbol{\beta}^0)\right|+\sup_{u_l\in[a_l,b_l]}\left|\widehat{m}_1^O(u_1,\widehat{\boldsymbol{\beta}})-\widehat{m}_1^O(u_1,\boldsymbol{\beta}^0)\right|\\ &+\sup_{u_1\in[a_1,b_1]}|\widehat{m}_1(u_1,\boldsymbol{\beta}^0)-\widehat{m}_1^O(u_1,\boldsymbol{\beta}^0)|\\ =&O_p(n^{-1/2}+J_n^{-r}).\end{aligned}$$

This completes Theorem 3. □

Proof of Theorem 4: Due to $nh^5=O(1)$, we have $\sqrt{nh_l}n^{-2/5}=o_p(1)$. By Theorem 3, we have

$$\sqrt{nh_l}\left\{\widehat{m}_l(u_l,\widehat{\boldsymbol{\beta}})-m_l(u_l)-b_l(u_l)h_l^2\right\}=\sqrt{nh_l}\left\{\widehat{m}_l^O(u_l,\widehat{\boldsymbol{\beta}})-m_l(u_l)-b_l(u_l)h_l^2\right\}+o_p(1).$$

Thus Theorem 4 can be shown straightforwardly by Lemma S.7. □

Lemma S.10. *Let assumptions* (A1)-(A6) *be satisfied, and $h\to 0$ and $nh^3\to\infty$, as $n\to\infty$. We have, for any $\boldsymbol{\theta}\in\boldsymbol{\Theta}$,*

$$E\left\{\left\|\frac{\partial\widehat{m}_1^O(\mathbf{X}^T\boldsymbol{\beta}_l,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_l}-m_l'(\mathbf{X}^T\boldsymbol{\beta}_l)(\mathbf{X}-\mathbf{P}(\mathbf{X}))\right\|_2^2\right\}=O_p(h^4+n^{-1}h^{-3}),$$

where $\boldsymbol{\Theta}$ is the parameter space for $\boldsymbol{\theta}$.

Proof of Lemma S.10: This Lemma can be completed along the lines of Proposition 1 (iii) in Cui et al. (2011). □

To prove Theorem 5, we need the following Lemma. Similar to Ma and Song (2015), assuming that the nonparametric function $m_0(u_0)$ is known, we can construct a GLR statistic based on the new data $(Y_1^O, \mathbf{X}, \mathbf{Z}, G)$ given in Section 2.3 of the main paper and show its asymptotic property. Then our interested GLR statistic based on $(\tilde{Y}_1, \mathbf{X}, \mathbf{Z}, G)$ can be constructed by plugging in the BSBK estimators of the nonparametric functions. Its asymptotic distribution can be shown in Theorem 3.

Consider hypothesis test (7) in the main context. Let $\widehat{\boldsymbol{\theta}}$ be the BSBK estimate of $\boldsymbol{\theta}$ proposed in Section 2.2 of the paper. Assuming $m_0(u_0)$ be known, similar to Ma and Song (2015), let $\widehat{m}^O_{1,H_0}(u_1)$ and $\widehat{m}^O_{1,H_1}(u_1)$ be the estimates based on new data $(Y_1^O, \mathbf{X}, \mathbf{Z}, G)$ under H_0 and H_1, respectively. The resulting residual sums of squares under H_0 and H_1 in hypothesis test (7) are

$$\mathrm{RSS}_1^O(H_0) = \sum_{i=1}^n \left\{ Y_i - \widehat{\boldsymbol{\alpha}}^T \tilde{\mathbf{Z}}_i - m_0(\widehat{\boldsymbol{\beta}}_0^T \mathbf{X}_i) - m_1(\widehat{\boldsymbol{\beta}}_1^T \mathbf{X}_i) G_i \right\}^2,$$
$$\mathrm{RSS}_1^O(H_1) = \sum_{i=1}^n \left\{ Y_i - \widehat{\boldsymbol{\alpha}}^T \tilde{\mathbf{Z}}_i - m_0(\widehat{\boldsymbol{\beta}}_0^T \mathbf{X}_i) - \widehat{m}^O_{1,H_1}(\widehat{\boldsymbol{\beta}}_1^T \mathbf{X}_i) G_i \right\}^2.$$

The following GLR statistic can be used to test the hypothesis in (7) in the main context when $m_0(u_0)$ is known,

$$T_1^O = \frac{n}{2} \frac{\mathrm{RSS}_1^O(H_0) - \mathrm{RSS}_1^O(H_1)}{\mathrm{RSS}_1^O(H_1)}. \tag{A.17}$$

Let $a_K = \{K(0) - 1/2 \int K^2(u)du\}[\int \{K(u) - 1/2K * K(u)\}du]^{-1}$, where $K * K(u)$ denotes the convolution of K. Denote by Ω_l the support of $\boldsymbol{\beta}_l^T \mathbf{x}$, and by $|\Omega_l|$ the length of Ω_l, $l = 0, 1$.

The following Lemma states the asymptotic distribution of T_1^O.

Lemma S.11. *Suppose that assumptions* (A.1)-(A.6) *in the Appendix hold, and $nN^{-4} \to \infty$ and $nN^{-2r-2} \to 0$, then under H_0 in* (7) *in the main context, when $m_0(u_0)$ is known, and $m_1^0(u_1)$ is a linear function of u_1,*

$$\sigma_{1n}^{-1}(T_1^O - \mu_{1n}) \xrightarrow{\mathcal{L}} N(0,1),$$

where $\sigma_{1n}^2 = 2|\Omega_1| \int \left\{K(u) - 1/2K * K^2(u)\right\}^2 du/h_1$ and $\mu_{1n} = |\Omega_1| \left\{K(0) - \frac{1}{2}\int K^2(u)du\right\}/h_1$. Furthermore, the scaled T_1^O follows an asymptotic χ^2-distribution with d_1 degrees of freedom, that is,

$$a_K T_1^O \overset{a}{\sim} \chi^2_{d_1},$$

where $d_1 = a_K \mu_{1n}$.

Proof of Lemma S.11: Define

$$\mathrm{RSS}_1^*(H_0) = \sum_{i=1}^n \left\{ Y_i - \tilde{\mathbf{Z}}_i^T \boldsymbol{\alpha}^0 - m_0(\mathbf{X}_i^T \boldsymbol{\beta}_0^0) - m_1(\mathbf{X}_i^T \boldsymbol{\beta}_1^0) G_i \right\}^2,$$

$$\mathrm{RSS}_1^*(H_1) = \sum_{i=1}^n \left\{ Y_i - \tilde{\mathbf{Z}}_i^T \boldsymbol{\alpha}^0 - m_0(\mathbf{X}_i^T \boldsymbol{\beta}_0^0) - \widehat{m}_{1,H_1}^O(\mathbf{X}_i^T \boldsymbol{\beta}_1^0) G_i \right\}^2,$$

$$T_1^* = \frac{n}{2} \frac{\mathrm{RSS}_1^*(H_0) - \mathrm{RSS}_1^*(H_1)}{\mathrm{RSS}_1^*(H_1)}.$$

We can rewrite $n^{-1}\{\mathrm{RSS}_1^*(H_0) - \mathrm{RSS}_1^O(H_0)\}$ as

$$\begin{aligned}
&\{\mathrm{RSS}_1^*(H_0) - \mathrm{RSS}_1^O(H_0)\} \\
=&\sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i^T(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) + m_0(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_0) - m_0(\mathbf{X}_i^T \boldsymbol{\beta}_0^0) + (m_1(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_1) - m_1(\mathbf{X}_i^T \boldsymbol{\beta}_1^0)) G_i \right\}^2 \\
&+ 2\sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i^T(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) + m_0(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_0) - m_0(\mathbf{X}_i^T \boldsymbol{\beta}_0^0) + (m_1(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_1) - m_1(\mathbf{X}_i^T \boldsymbol{\beta}_1^0)) G_i \right\} \\
&\times \left\{ Y_i - \widehat{\boldsymbol{\alpha}}^T \tilde{\mathbf{Z}}_i - m_0(\widehat{\boldsymbol{\beta}}_0^T \mathbf{X}_i) - \widehat{m}_{1,H_0}^O(\widehat{\boldsymbol{\beta}}_1^T \mathbf{X}_i) G_i \right\} \\
&= O_p(1),
\end{aligned}$$

and similarly

$$\begin{aligned}
&\{\mathrm{RSS}_1^*(H_1) - \mathrm{RSS}_1^O(H_1)\} \\
=&\sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i^T(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) + m_0(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_0) - m_0(\mathbf{X}_i^T \boldsymbol{\beta}_0^0) + (\widehat{m}_{1,H_1}^O(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_1) - \widehat{m}_{1,H_1}^O(\mathbf{X}_i^T \boldsymbol{\beta}_1^0)) G_i \right\}^2 \\
&+ 2\sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i^T(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) + m_0(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_0) - m_0(\mathbf{X}_i^T \boldsymbol{\beta}_0^0) + (\widehat{m}_{1,H_1}^O(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_1) - \widehat{m}_{1,H_1}^O(\mathbf{X}_i^T \boldsymbol{\beta}_1^0)) G_i \right\} \\
&\times \left\{ Y_i - \widehat{\boldsymbol{\alpha}}^T \tilde{\mathbf{Z}}_i - m_0(\widehat{\boldsymbol{\beta}}_0^T \mathbf{X}_i) - \widehat{m}_{1,H_1}^O(\widehat{\boldsymbol{\beta}}_1^T \mathbf{X}_i) G_i \right\} \\
&= O_p(1),
\end{aligned}$$

which implies that

$$\mathrm{RSS}_1^O(H_0) - \mathrm{RSS}_1^O(H_1) = \mathrm{RSS}_1^*(H_0) - \mathrm{RSS}_1^*(H_1) + O_p(1).$$

Along the lines of the proof of Theorem 7, we have

$$n^{-1}\mathrm{RSS}_1^O(H_1) = \sigma^2 + o_p(1), \quad \text{and} \quad n^{-1}\mathrm{RSS}_1^*(H_1) = \sigma^2 + o_p(1).$$

Therefore, it can be show that $T_1^O = T_1^* + O_p(1)$. It remains to show that

$$\sigma_{1n}^{-1}(T_1^* - \mu_{1n}) \xrightarrow{\mathcal{L}} N(0,1).$$

This can be shown along the lines of Fan et al. (2001).

□

Proof of Theorem 5: It is easy to see that

$$
\begin{aligned}
&\{\mathrm{RSS}_1^O(H_0) - \mathrm{RSS}_1(H_0)\} \\
=&\sum_{i=1}^n \Big\{\widehat{m}_{0,H_0}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) - m_0(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) + (\widehat{m}_{1,H_0}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1) - m_1(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1))G_i\Big\}^2 \\
&+ 2\sum_{i=1}^n \Big\{\widehat{m}_{0,H_0}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) - m_0(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) + (\widehat{m}_{1,H_0}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1) - m_1(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1))G_i\Big\} \\
&\times \Big\{Y_i - \widehat{\boldsymbol{\alpha}}^T\tilde{\mathbf{Z}}_i - \widehat{m}_{0,H_0}(\widehat{\boldsymbol{\beta}}_0^T\mathbf{X}_i) - \widehat{m}_{1,H_0}(\widehat{\boldsymbol{\beta}}_1^T\mathbf{X}_i)G_i\Big\} \\
\equiv&\mathbb{J}_{31} + \mathbb{J}_{32},
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&\{\mathrm{RSS}_1^O(H_1) - \mathrm{RSS}_1(H_1)\} \\
=&\sum_{i=1}^n \Big\{\widehat{m}_{0,H_1}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) - m_0(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) + (\widehat{m}_{1,H_1}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1) - \widehat{m}_{1,H_1}^O(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1))G_i\Big\}^2 \\
&+ 2\sum_{i=1}^n \Big\{\widehat{m}_{0,H_1}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) - m_0(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) + (\widehat{m}_{1,H_1}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1) - \widehat{m}_{1,H_1}^O(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1))G_i\Big\} \\
&\times \Big\{Y_i - \widehat{\boldsymbol{\alpha}}^T\tilde{\mathbf{Z}}_i - \widehat{m}_{0,H_1}(\widehat{\boldsymbol{\beta}}_0^T\mathbf{X}_i) - \widehat{m}_{1,H_1}(\widehat{\boldsymbol{\beta}}_1^T\mathbf{X}_i)G_i\Big\} \\
\equiv&\mathbb{J}_{41} + \mathbb{J}_{42}.
\end{aligned}
$$

Furthermore, we can show that

$$
\begin{aligned}
\mathbb{J}_{31} - \mathbb{J}_{41} =& \sum_{i=1}^n \Big\{\widehat{m}_{0,H_0}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) - \widehat{m}_{0,H_1}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) + (\widehat{m}_{1,H_0}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1) - m_1(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1))G_i \\
&- (\widehat{m}_{1,H_1}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1) - \widehat{m}_{1,H_1}^O(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1))G_i\Big\}^2 \\
&+ 2\sum_{i=1}^n \Big\{\widehat{m}_{0,H_0}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) - \widehat{m}_{0,H_1}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) + (\widehat{m}_{1,H_0}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1) - m_1(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1))G_i \\
&- (\widehat{m}_{1,H_1}^O(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1) - \widehat{m}_{1,H_1}^O(\mathbf{X}_i^T\boldsymbol{\beta}_1^0))G_i\Big\} \\
&\times \Big\{\widehat{m}_{0,H_1}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) - m_0(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_0) + (\widehat{m}_{1,H_1}(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1) - \widehat{m}_{1,H_1}^O(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_1))G_i\Big\} \\
=&O_p(1),
\end{aligned}
$$

and similarly $\mathbb{J}_{12} - \mathbb{J}_{22} = O_p(1)$. As the same proof of Lemma S.11, $T_1 = T_1^O + o(1)$ holds by showing the following claims that

$$
\begin{aligned}
\mathrm{RSS}_1(H_0) - \mathrm{RSS}_1(H_1) &= \mathrm{RSS}_1^O(H_0) - \mathrm{RSS}_1^O(H_1) + O_p(1), \\
n^{-1}\mathrm{RSS}_1(H_1) = \sigma^2 + o_p(1),\ &\text{and } n^{-1}\mathrm{RSS}_1^O(H_1) = \sigma^2 + o_p(1).
\end{aligned}
$$

Thus, T_1 and T_1^O have the same asymptotical distribution, which completes the proof of Theorem 5 by Lemma S.11.

□

Proof of Theorem 6: Define

$$
\begin{aligned}
\mathrm{RSS}_2^O(H_0) &= \sum_{i=1}^n \left\{ Y_i - \tilde{\mathbf{Z}}_i^T \widehat{\boldsymbol{\alpha}} - m_0(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_0) - m_1(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_1) G_{i1} - m_2(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_2) G_{i2} \right\}^2, \\
\mathrm{RSS}_2^O(H_1) &= \sum_{i=1}^n \left\{ Y_i - \tilde{\mathbf{Z}}_i^T \widehat{\boldsymbol{\alpha}} - \widehat{m}_0^O(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_0) - \widehat{m}_1^O(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_1) G_{i1} - \widehat{m}_2^O(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_2) G_{i2} \right\}^2, \\
\mathrm{RSS}_2^*(H_0) &= \sum_{i=1}^n \left\{ Y_i - \tilde{\mathbf{Z}}_i^T \boldsymbol{\alpha}^0 - m_0(\mathbf{X}_i^T \boldsymbol{\beta}_0^0) - m_1(\mathbf{X}_i^T \boldsymbol{\beta}_1^0) G_{i1} - m_2(\mathbf{X}_i^T \boldsymbol{\beta}_2^0) G_{i2} \right\}^2, \\
\mathrm{RSS}_2^*(H_1) &= \sum_{i=1}^n \left\{ Y_i - \tilde{\mathbf{Z}}_i^T \boldsymbol{\alpha}^0 - m_0(\mathbf{X}_i^T \boldsymbol{\beta}_0^0) - \widehat{m}_1^O(\mathbf{X}_i^T \boldsymbol{\beta}_1^0) G_{i1} - \widehat{m}_2^O(\mathbf{X}_i^T \boldsymbol{\beta}_2^0) G_{i2} \right\}^2, \\
T_2^O &= \frac{n}{2} \frac{\mathrm{RSS}_2^O(H_0) - \mathrm{RSS}_2^O(H_1)}{\mathrm{RSS}_2^O(H_1)}, \\
T_2^* &= \frac{n}{2} \frac{\mathrm{RSS}_2^*(H_0) - \mathrm{RSS}_2^*(H_1)}{\mathrm{RSS}_2^*(H_1)},
\end{aligned}
$$

where $\tilde{\mathbf{Z}}_i = (\mathbf{Z}_i^T, \mathbf{Z}_i^T G_{i1}, \mathbf{Z}_i^T G_{i2})^T$. Along the lines of the proof of Lemma S.11, we have

$$
\begin{aligned}
\mathrm{RSS}_2^O(H_0) - \mathrm{RSS}_2^O(H_1) &= \mathrm{RSS}_2^*(H_0) - \mathrm{RSS}_2^*(H_1) + O_p(1), \\
n^{-1}\mathrm{RSS}_2^O(H_1) = \sigma^2 + o_p(1), &\text{ and } n^{-1}\mathrm{RSS}_2^*(H_1) = \sigma^2 + o_p(1).
\end{aligned}
$$

Therefore, it can be show that $T_2^O = T_2^* + o_p(1)$. Following the same way as the proof of Theorem 5, we can show that $T_2 = T_2^O + o(1)$ following the claims below,

$$
\begin{aligned}
\mathrm{RSS}_2(H_0) - \mathrm{RSS}_2(H_1) &= \mathrm{RSS}_2^O(H_0) - \mathrm{RSS}_2^O(H_1) + O_p(1), \\
n^{-1}\mathrm{RSS}_2(H_1) &= \sigma^2 + o_p(1).
\end{aligned}
$$

It remains to show that

$$
\sigma_{2n}^{-1}(T_2^* - \mu_{2n}) \xrightarrow{\mathcal{L}} N(0,1).
$$

Let $U_{l,i} = X_i^T \boldsymbol{\beta}_l^0$, $U_l = \mathbf{X}^T \boldsymbol{\beta}_l^0$ and $\mathbb{G}_i = (1, G_{i1}, G_{i2})^T$. By Lemma S.7, we have

$$
\widehat{m}_l^O(u_l, \boldsymbol{\theta}^0) - m_l(u_l) = \{a_{l,n}(u_l) + R_n(u_l)\}(1 + o_p(1)),
$$

where

$$
\begin{aligned}
a_{ln}(u_l) &= \frac{1}{nh} d_l(u_l) \sum_{i=1}^n \varepsilon_i \mathbb{G}_{il} K((U_{l,i} - u_l)/h_l), \\
R_{ln}(u_l) &= \frac{1}{nh} d_l(u_l) \sum_{i=1}^n \left\{ m_{l0}(U_{l,i}) \mathbb{G}_{l,i} - \tau_l(u_l)^T \check{X}_{il} \right\} \mathbb{G}_{l,i} K((U_{l,i} - u_l)/h_l),
\end{aligned}
$$

and $d_l(u_l) = \left\{ E[\mathbb{G}_l^2 | U_l = u_l] f_l(u_l) \right\}^{-1}$, $\tau_l(u_l) = (m_{l0}(u_l), h m_{l0}'(u_l))^T$ and $\check{X}_{il} = \check{X}_i(u_l) =$

$(\mathbb{G}_{il}, \mathbb{G}_{il}(U_{l,i}-u_l)/h_l)^T$. Define

$$R_n^{(1)} = \sum_{l=1}^{2}\sum_{i=1}^{n} \varepsilon_i R_{l,n}(U_{l,i})\mathbb{G}_{l,i},$$
$$R_n^{(2)} = \sum_{l=1}^{2}\sum_{i=1}^{n} a_{l,n}(U_{l,i})\mathbb{G}_{l,i}^2 R_{l,n}(U_{l,i}),$$
$$R_n^{(3)} = \frac{1}{2}\sum_{i=1}^{n}\left\{\sum_{l=1}^{2}\mathbb{G}_{l,i}R_{l,n}(U_{l,i})\right\}^2.$$

Thus, as the proof of Theorem 5 in Fan et al. (2001), we have

$$\begin{aligned}-2\sigma^2 T_2^* =& -\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left\{\sum_{k=1}^{n}\varepsilon_k\sum_{l=1}^{2}d_l(U_{l,i})\mathbb{G}_{l,k}\mathbb{G}_{l,i}K_{h_l}(U_{l,k}-U_{l,i})\right\}\\ &+\frac{1}{2n^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\varepsilon_i\varepsilon_j\left\{\sum_{l=1}^{2}d_l(U_{l,k})\mathbb{G}_{l,k}\mathbb{G}_{l,i}K_{h_l}(U_{l,i}-U_{l,k})\right\}\\ &\times\left\{\sum_{l=1}^{2}d_l(U_{l,k})\mathbb{G}_{l,k}\mathbb{G}_{l,j}K_{h_l}(U_{l,j}-U_{l,k})\right\}-R_n^{(1)}+R_n^{(2)}+R_n^{(3)}+O_p(n^{-1}h^{-2})\\ \equiv&\tilde{T}_n+\tilde{S}_n-R_n^{(1)}+R_n^{(2)}+R_n^{(3)}+O_p(n^{-1}h^{-2}).\end{aligned}$$

By some direct but tedious calculations, we have

$$\tilde{R}_n \equiv -R_n^{(1)}+R_n^{(2)}+R_n^{(3)} = O_p(nh^4+n^{1/2}h^2). \tag{A.18}$$

It is obvious that

$$\begin{aligned}-\tilde{T}_n =& 2\sigma^2K(0)\sum_{l=1}^{2}h_l^{-1}E\{f_l(U_l)^{-1}\}\\ &+\frac{1}{n}\sum_{i\neq k}^{n}\varepsilon_i\varepsilon_k\sum_{l=1}^{2}d_l(U_{l,i})\mathbb{G}_{l,k}\mathbb{G}_{l,i}K_{h_l}(U_{l,k}-U_{l,i})+o_p(h^{-1/2}).\end{aligned} \tag{A.19}$$

Let $2\tilde{S}_n = S_{n1}+S_{n2}$, where

$$\begin{aligned}S_{n1} =& \frac{1}{n^2}\sum_{i=1}^{n}\varepsilon_i^2\sum_{k=1}^{n}\left\{\sum_{l=1}^{2}d_l(U_{l,k})\mathbb{G}_{l,k}\mathbb{G}_{l,i}K_{h_l}(U_{l,i}-U_{l,k})\right\}^2,\\ S_{n2} =& \frac{1}{n^2}\sum_{i\neq j}^{n}\varepsilon_i\varepsilon_j\sum_{k=1}^{n}\left\{\sum_{l=1}^{2}d_l(U_{l,k})\mathbb{G}_{l,k}\mathbb{G}_{l,i}K_{h_l}(U_{l,i}-U_{l,k})\right\}\\ &\times\left\{\sum_{l=1}^{2}d_l(U_{l,k})\mathbb{G}_{l,k}\mathbb{G}_{l,j}K_{h_l}(U_{l,j}-U_{l,k})\right\}.\end{aligned}$$

It is easy to see that

$$S_{n1} = \varrho_n(1+o(1))+O_p(n^{-3/2}h^{-2})+O_p((nh^2)^{-1})+o_p(h^{-1/2}),$$

where

$$\varrho_n = \frac{\sigma^2}{n(n-1)} \sum_{i \neq k}^{n} \left\{ \sum_{l=1}^{2} d_l(U_{l,k}) \mathbb{G}_{l,k} \mathbb{G}_{l,i} K_{h_l}(U_{l,i} - U_{l,k}) \right\}^2 .$$

Using Hoeffiding's decomposition (Serfling 1980) for the variance of U-statistics, by tedious calculations we can show that the variance of ϱ_n is

$$\varrho_n = O_p(n^{-1}h^{-2}).$$

It is straightforward to calculate the expectation of ϱ_n

$$E\varrho_n = \sigma^2 \left\{ \sum_{l=1}^{2} h_l^{-1} E f_l(U_l)^{-1} \right\} \int K^2(t)dt + o_p(h^{-1}),$$

which implies that

$$S_{n1} = \sigma^2 \sum_{l=1}^{2} h_l^{-1} E \left\{ f_l(U_l)^{-1} \right\} \int K^2(t)dt + o_p(h^{-1/2}). \tag{A.20}$$

S_{n2} can be further decomposed as $S_{n2} = S_{n21} + S_{n22}$, where

$$S_{n21} = \frac{1}{n^2} \sum_{i \neq j}^{n} \varepsilon_i \varepsilon_j \sum_{k \neq i,j}^{n} \left\{ \sum_{l=1}^{2} d_l(U_{l,k}) \mathbb{G}_{l,k} \mathbb{G}_{l,i} K_{h_l}(U_{l,i} - U_{l,k}) \right\} \times \left\{ \sum_{l=1}^{2} d_l(U_{l,k}) \mathbb{G}_{l,k} \mathbb{G}_{l,j} K_{h_l}(U_{l,j} - U_{l,k}) \right\},$$

$$\begin{aligned} S_{n22} =& \frac{1}{n^2} \sum_{i \neq j}^{n} \varepsilon_i \varepsilon_j \left\{ \sum_{l=1}^{2} d_l(U_{l,j}) \mathbb{G}_{l,j} \mathbb{G}_{l,i} K_{h_l}(U_{l,i} - U_{l,j}) \right\} \left\{ \sum_{l=1}^{2} d_l(U_{l,j}) \mathbb{G}_{l,j}^2 K_{h_l}(0) \right\} \\ &+ \frac{1}{n^2} \sum_{i \neq j}^{n} \varepsilon_i \varepsilon_j \left\{ \sum_{l=1}^{2} d_l(U_{l,i}) \mathbb{G}_{l,j} \mathbb{G}_{l,i} K_{h_l}(U_{l,i} - U_{l,j}) \right\} \left\{ \sum_{l=1}^{2} d_l(U_{l,i}) \mathbb{G}_{l,i}^2 K_{h_l}(0) \right\}. \end{aligned}$$

It can be show by tedious calculation that

$$\text{var}(S_{n22}) = O(n^{-2}h^{-3}),$$

which implies that

$$S_{n22} = o(h^{-1/2}). \tag{A.21}$$

Let

$$Q_{ijk} = \left\{ \sum_{l=1}^{2} d_l(U_{l,k}) \mathbb{G}_{l,k} \mathbb{G}_{l,i} K_{h_l}(U_{l,i} - U_{l,k}) \right\} \left\{ \sum_{l=1}^{2} d_l(U_{l,k}) \mathbb{G}_{l,k} \mathbb{G}_{l,j} K_{h_l}(U_{l,j} - U_{l,k}) \right\}.$$

It can be shown that

$$E\left\{n^{-1}\sum_{k\neq i,j}^{n}Q_{ijk}-E\left[Q_{ijk}|(U_{l,i},U_{l,j})_{l=2,3}\right]\right\}^2\leq n^{-2}\sum_{k\neq i,j}^{n}EQ_{ijk}^2=O((nh^2)^{-1}),$$

which results in that

$$\begin{aligned}S_{n21}=&\frac{n-2}{n^2}\sum_{i\neq j}^{n}\varepsilon_i\varepsilon_j E\left[Q_{ijk}|(U_{l,i},U_{l,j})_{l=0,1}\right]+o_p(h^{-1/2})\\=&\frac{1}{n}\sum_{i\neq j}^{n}\varepsilon_i\varepsilon_j\sum_{l=1}^{2}h_l^{-1}d_l(U_{l,i})\mathbb{G}_{l,j}\mathbb{G}_{l,i}\tilde{K}((U_{l,i}-U_{l,j})/h_l)+o_p(h^{-1/2}).\end{aligned}\tag{A.22}$$

By (A.18)-(A.22), we have

$$T_1^*=\mu_{2n}+\frac{1}{2}\Upsilon(n)\sqrt{b(h)}-\tilde{R}_n+o_p(h^{-1/2}),$$

where

$$\begin{aligned}b(h)=&\sum_{l=1}^{2}h_l^{-1}Ef_l(U_l)^{-1},\\\mu_{2n}=&\left\{K(0)-\frac{1}{2}\int K^2(t)dt\right\}\sum_{l=1}^{2}h_l^{-1}Ef_l(U_l)^{-1},\\\Upsilon(n)=&\frac{1}{n\sqrt{b(h)}\sigma^2}\sum_{i\neq j}^{n}\varepsilon_i\varepsilon_j\sum_{l=1}^{2}h_l^{-1}d_l(U_{l,i})\mathbb{G}_{l,j}\mathbb{G}_{l,i}\{K((U_{l,i}-U_{l,j})/h_l)-1/2\tilde{K}((U_{l,i}-U_{l,j})/h_l)\}.\end{aligned}$$

It remains to prove that

$$\Upsilon(n)\xrightarrow{\mathcal{L}}N(0,v^2)$$

with $v^2=2\int\{K(t)-\frac{1}{2}\tilde{K}(t)\}^2dt$.

Let

$$\Phi_{ij}=\varepsilon_i\varepsilon_j\sum_{l=1}^{2}d_l(U_{l,i})\mathbb{G}_{l,i}\mathbb{G}_{l,j}\left\{K_{h_l}((U_{l,i}-U_{l,j}))-\frac{1}{2}\tilde{K}_{h_l}((U_{l,i}-U_{l,j}))\right\}$$

and

$$\Upsilon(n)=\sum_{i<j}\Upsilon_{ij}.$$

where $\Upsilon_{ij}=\frac{1}{n\sqrt{b(h)}\sigma^2}(\Phi_{ij}+\Phi_{ji})$. Define $\sigma_{2n}^2=\mathrm{Var}\,(\Upsilon(n))$. According to Proposition 3.2 in de Jong (1987), it suffices to check the following conditions:

(a) $\Upsilon(n)$ is clean [see de Jong (1987) for the definition],

(b) $v_n^2\to v^2$,

(c) ζ_1 is of lower order than v_n^4,

(d) ζ_2 is of lower order than v_n^4,

(e) ζ_3 is of lower order than v_n^4,

where

$$\zeta_1 = E \sum_{1 \le i < j \le n} \Upsilon_{ij}^4,$$
$$\zeta_2 = E \sum_{1 \le i < j < k \le n} \{\Upsilon_{ij}^2 \Upsilon_{ik}^2 + \Upsilon_{ji}^2 \Upsilon_{jk}^2 + \Upsilon_{ki}^2 \Upsilon_{kj}^2\},$$
$$\zeta_3 = E \sum_{1 \le i < j < k < l \le n} \{\Upsilon_{ij} \Upsilon_{ik} \Upsilon_{lj} \Upsilon_{lk} + \Upsilon_{ij} \Upsilon_{il} \Upsilon_{kj} \Upsilon_{kl} + \Upsilon_{ik} \Upsilon_{il} \Upsilon_{jk} \Upsilon_{jl}\}.$$

We check each of the conditions as follows. Condition (a) holds obviously. Then we calculate the variance σ_n^2 as follows. (a) implies that $\sigma_n^2 = E\Upsilon(n)^2$. By de Jong (1987), we have

$$\sigma_{2n}^2 = E \sum_{1 \le i < j \le n} \Upsilon_{ij}^2 = \frac{4}{n^2 b(h) \sigma^4} E \sum_{i \ne j} \Phi_{ij}^2.$$

It is easy to see that

$$E \sum_{i \ne j} \Phi_{ij}^2 = n^4 \sigma^4 b(h) \int \{K(t) - \frac{1}{2} \tilde{K}(t)\}^2 dt,$$

which implies that (b) holds. Noting that

$$E \sum_{1 \le i < j \le n} \{\Phi_{ij}^4\} = O(n^2 h^{-3}), E \sum_{1 \le i < j \le n} \{\Phi_{ij}^3 \Phi_{ji}\} = O(n^2 h^{-2}), E \sum_{1 \le i < j \le n} \{\Phi_{ij}^2 \Phi_{ji}^2\} = O(n^2 h^{-2}),$$

then we have $\zeta_1 = O(h^2 n^{-4}) O(n^2 h^{-3}) = O(n^{-2} h^{-1})$. Similarly conditions (d) can be shown by noting that

$$E \sum_{1 \le i < j < k \le n} \{\Upsilon_{ij}^2 \Upsilon_{ik}^2\} = h^2 n^{-4} O(n^3 h^{-2}) = O(n^{-1}),$$

which implies that $\zeta_2 = O(n^{-1})$. It is obvious by straightforward calculations that,

$$E \sum_{1 \le i < j < k < l \le n} \{\Phi_{ij} \Phi_{ik} \Phi_{lj} \Phi_{lk}\} = O(n^4 h^{-1}),$$
$$E \sum_{1 \le i < j < k < l \le n} \{\Phi_{ji} \Phi_{ik} \Phi_{lj} \Phi_{lk}\} = O(n^4 h^{-1}),$$
$$E \sum_{1 \le i < j < k < l \le n} \{\Phi_{ji} \Phi_{ki} \Phi_{lj} \Phi_{lk}\} = O(n^4 h^{-1}),$$

which results in

$$E \sum_{1 \le i < j < k < l \le n} \{\Upsilon_{ij} \Upsilon_{ik} \Upsilon_{lj} \Upsilon_{lk}\} = O(h^2 n^{-4}) O(n^4 h^{-1}) = O(h).$$

Therefore, we have $\zeta_3 = O(h)$, which implies that condition (e) holds. This completes the proof of Theorem 6.

□

Proof of Theorem 7: This proof is similar to Liang et al. (2010). Thus we only provide a sketch of the proof here. We first prove $n^{-1}R(H_1) = E\{\sigma(\mathbf{V})\} + o_p(1)$. Let $\widehat{m}(\mathbf{X}, \boldsymbol{\beta}) = \widehat{m}_0(\mathbf{X}^T\boldsymbol{\beta}_0, \boldsymbol{\beta}) + \widehat{m}_1(\mathbf{X}^T\boldsymbol{\beta}_1, \boldsymbol{\beta})G$ and correspondingly $\widehat{m}^O(\mathbf{X}, \boldsymbol{\beta}) = \widehat{m}_0^O(\mathbf{X}^T\boldsymbol{\beta}_0, \boldsymbol{\beta}) + \widehat{m}_1^O(\mathbf{X}^T\boldsymbol{\beta}_1, \boldsymbol{\beta})G$. By Theorem 3 and (A.15), $n^{-1}R(H_1)$ can be decomposed as following

$$
\begin{aligned}
n^{-1}R(H_1) =& \frac{1}{n}\sum_{i=1}^n \left\{y_i - \tilde{\mathbf{Z}}^T\widehat{\boldsymbol{\alpha}} - \widehat{m}(\mathbf{X}_i, \widehat{\boldsymbol{\beta}})\right\}^2 \\
=& \frac{1}{n}\sum_{i=1}^n \left\{y_i - \tilde{\mathbf{Z}}^T\boldsymbol{\alpha}^0 - \widehat{m}^O(\mathbf{X}_i, \boldsymbol{\beta}^0)\right\}^2 + o_p(n^{-2/5}) + O_p(n^{-1/2}) \\
=& \frac{1}{n}\sum_{i=1}^n \left\{\varepsilon_i - (\widehat{m}^O(\mathbf{X}_i, \boldsymbol{\beta}^0) - m(\mathbf{V}_i, \boldsymbol{\beta}^0)\right\}^2 + o_p(n^{-2/5}) \\
\equiv& \mathbb{I}_1 + \mathbb{I}_2 + \mathbb{I}_3 + + o_p(n^{-2/5}),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{I}_1 =& \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2, \\
\mathbb{I}_2 =& -2\frac{1}{n}\sum_{i=1}^n \{\widehat{m}^O(\mathbf{X}_i, \boldsymbol{\beta}^0) - m(\mathbf{V}_i, \boldsymbol{\beta}^0)\}\varepsilon_i, \\
\mathbb{I}_3 =& \frac{1}{n}\sum_{i=1}^n \left\{\widehat{m}^O(\mathbf{X}_i, \boldsymbol{\beta}^0) - m(\mathbf{V}_i, \boldsymbol{\beta}^0\right\}^2.
\end{aligned}
$$

It is easy to see by the law of large numbers that $\mathbb{I}_1 = E\{\sigma(\mathbf{V})\} + O_p(n^{-1/2})$. By Theorem 2.6 in Li and Racine (2007), we have $\max_i |\widehat{m}^O(\mathbf{X}_i, \boldsymbol{\beta}^0) - m(\mathbf{X}_i, \boldsymbol{\beta}^0)| = O_p(\sqrt{\log n/(nh)})$, which results in $\mathbb{I}_2 = O_p(\sqrt{\log n/(n^2h)})$ and $\mathbb{I}_3 = O_p(\log n/(nh))$. This leads to $n^{-1}R(H_1) = E\{\sigma(\mathbf{V})\} + o_p(1)$.

Let $B_n = \mathbf{A}\Gamma_n^{-1}\mathbf{A}^T$, where $\mathbf{A}$ is defined in Section 3.2 and

$$
\Gamma_n = \sum_{i=1}^n \widehat{\boldsymbol{\Psi}}_i(\mathbf{X}, \mathbf{Z})\widehat{\boldsymbol{\Psi}}_i(\mathbf{X}, \mathbf{Z})^T, \text{ with } \widehat{\boldsymbol{\Psi}}_i(\mathbf{X}, \mathbf{Z}) = \begin{pmatrix} X_{\widehat{m},i} - \tilde{\mathbf{P}}_n(\mathbf{X}_i) \\ \tilde{Z}_i - \tilde{\mathbf{P}}_n(\mathbf{Z}_i) \end{pmatrix}.
$$

The difference of $R(H_0) - R(H_1)$ can be decomposed as

$$
\begin{aligned}
R(H_0) - R(H_1) =& \sum_{i=1}^n \left\{\tilde{\mathbf{Z}}^T(\widehat{\boldsymbol{\alpha}}_{H_0} - \widehat{\boldsymbol{\alpha}}_{H_1}) + (\widehat{m}(\mathbf{X}_i, \widehat{\boldsymbol{\theta}}_{H_0}) - \widehat{m}(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_{H_1}))\right\}^2 \\
& + 2\sum_{i=1}^n \left\{\tilde{\mathbf{Z}}^T(\widehat{\boldsymbol{\alpha}}_{H_0} - \widehat{\boldsymbol{\alpha}}_{H_1}) + (\widehat{m}(\mathbf{X}_i, \widehat{\boldsymbol{\theta}}_{H_0}) - \widehat{m}(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_{H_1}))\right\} \\
& \times \left\{y_i - \tilde{\mathbf{Z}}^T\widehat{\boldsymbol{\alpha}}_{H_1} - \widehat{m}(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_{H_1})\right\} \\
\equiv& \mathbb{I}_4 + \mathbb{I}_5.
\end{aligned}
$$

By Lemma S.5 and Lemma S.10, $\mathbb{I}_4$ can be rewritten as

$$\begin{aligned}
\mathbb{I}_4 =& \sum_{i=1}^{n} \left\{ \tilde{\mathbf{Z}}^T (\widehat{\boldsymbol{\alpha}}_{H_0} - \widehat{\boldsymbol{\alpha}}_{H_1}) + (\widehat{m}(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_{H_0}) - \widehat{m}(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_{H_1})) \right\}^2 \\
=& \sum_{i=1}^{n} \left\{ \tilde{\mathbf{Z}}^T (\widehat{\boldsymbol{\alpha}}_{H_0} - \widehat{\boldsymbol{\alpha}}_{H_1}) + (\widehat{m}^O(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_{H_0}) - \widehat{m}^O(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_{H_1})) \right\}^2 + o(1) \\
=& \sum_{i=1}^{n} \Big\{ \tilde{\mathbf{Z}}^T (\widehat{\boldsymbol{\alpha}}_{H_0} - \widehat{\boldsymbol{\alpha}}_{H_1}) + D_i(\widehat{\boldsymbol{\beta}}_{H_0}^T \mathbf{X}_i)^T (\widehat{\lambda}_{H_0}(\widehat{\boldsymbol{\beta}}_{H_0}) - \widehat{\lambda}_{H_1}(\widehat{\boldsymbol{\beta}}_{H_1})) \\
&+ (D_i(\widehat{\boldsymbol{\beta}}_{H_0}^T \mathbf{X}_i) - D_i(\widehat{\boldsymbol{\beta}}_{H_1}^T \mathbf{X}_i)^T)^T \widehat{\lambda}_{H_1}(\widehat{\boldsymbol{\beta}}_{H_1}) + o_p(\|\widehat{\boldsymbol{\theta}}_{H_0} - \widehat{\boldsymbol{\theta}}_{H_1}\|_2)) \Big\}^2 + o(1) \\
=& (\widehat{\boldsymbol{\theta}}_{H_0} - \widehat{\boldsymbol{\theta}}_{H_1})^T \sum_{i=1}^{n} \widehat{\boldsymbol{\Psi}}_i(\mathbf{X}, \mathbf{Z})^{\otimes 2} (\widehat{\boldsymbol{\theta}}_{H_0} - \widehat{\boldsymbol{\theta}}_{H_1}) + o(1).
\end{aligned}$$

As the estimators of $\boldsymbol{\theta}$ under the null and alternative hypotheses have the following relationship

$$\widehat{\boldsymbol{\theta}}_{H_0} = \widehat{\boldsymbol{\theta}}_{H_1} + \Gamma_n^{-1} \mathbf{A}^T B_n^{-1} (\gamma - \mathbf{A}\widehat{\boldsymbol{\theta}}_{H_1}),$$

we have

$$\begin{aligned}
\mathbb{I}_4 =& \left(\gamma - \mathbf{A}\widehat{\boldsymbol{\theta}}_{H_1}\right)^T B_n^{-1} \mathbf{A} \Gamma_n^{-1} \sum_{i=1}^{n} \widehat{\boldsymbol{\Psi}}_i(\mathbf{X}, \mathbf{Z})^{\otimes 2} \Gamma_n^{-1} \mathbf{A}^T B_n^{-1} \left(\gamma - \mathbf{A}\widehat{\boldsymbol{\theta}}_{H_1}\right) + o(1) \\
=& \left(\gamma - \mathbf{A}\widehat{\boldsymbol{\theta}}_{H_1}\right)^T B_n^{-1} \left(\gamma - \mathbf{A}\widehat{\boldsymbol{\theta}}_{H_1}\right) + o(1)
\end{aligned}$$

Therefore, under the null hypothesis, $\sigma^{-2}\mathbb{I}_4 \xrightarrow{\mathcal{L}} \chi_k^2$ and under the alternative hypothesis $\sigma^{-2}\mathbb{I}_4$ asymptotically follows a noncentral Chi-square distribution with k degrees of freedom and non-centrality parameter ϕ. It remains to show that $\mathbb{I}_5 = o_p(1)$. As above arguments, we have

$$\begin{aligned}
\mathbb{I}_5 =& 2\sum_{i=1}^{n} \left\{ \tilde{\mathbf{Z}}^T (\widehat{\boldsymbol{\alpha}}_{H_0} - \widehat{\boldsymbol{\alpha}}_{H_1}) + (\widehat{m}(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_{H_0}) - \widehat{m}(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_{H_1})) \right\} \\
&\times \left\{ y_i - \tilde{\mathbf{Z}}^T \widehat{\boldsymbol{\alpha}}_{H_1} - \widehat{m}(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_{H_1}) \right\} \\
=& 2\sum_{i=1}^{n} \left\{ \widehat{\boldsymbol{\Psi}}_i(\mathbf{X}, \mathbf{Z})(\widehat{\boldsymbol{\theta}}_{H_0} - \widehat{\boldsymbol{\theta}}_{H_1}) + O_p(n^{-1/2}) \right\} \left\{ \varepsilon_i + O_p(n^{-1/2}) \right\} \\
=& o_p(1).
\end{aligned}$$

This completes the proof of Theorem 7. □

Bibliography

Bosq, D. (1998). *Nonparametric statistics for stochastic processes.* Springer-Verlag, New York.

Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J. Am. Stat. Assoc.* **95**, 888-902.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Am. Stat. Assoc.* **92**, 477-489.

Carroll, R. J., Ruppert, D. and Welsh, A. H. (1998). Local estimating equations. *J. Am. Stat. Assoc.* **93**, 214-227.

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

de Boor, C. *A practical guide to splines*, Springer, New York.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Prob. Theo. Relat. Fiel.* **75** 262-277.

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation*. Springer, New York.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. stat* **29**, 153-193.

Li, Q. and Racine, R. S. (2007). *Nonparametric Econometrics: Theory and practice*. Princeton University Press, Princeton, N. J.

Liang, H., Liu, X., Li, R. and Tsai, C. L. (2010). Estimation and testing for partially linear singleindex models. *Ann. Stat.* **38**, 3811-3836.

Liu, X., Jiang, H. and Zhou, Y. (2014). Local empirical likelihood inference for varying-coefficient density-ratio models based on case-control data. *J. Am. Stat. Assoc.* **109**, 635-646.

Ma, S. and Song, P. X. (2015). Varying index coefficient models. *J. Am. Stat. Assoc.* in press.

Ruppert, D., Sheathers, S. J. and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Am. Stat. Assoc.* **90**, 1257-1270.

Sepanski, J. H., Knickerbocker, R. and Carroll, R. J. (1994). A semiparametric correction for attenuation. *J. Am. Stat. Assoc.* **89**, 1366-1373.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons Inc., New York.

Xia, Y. C. and Li, W. K. (1999). On single-index coefficient regression models," *J. Am. Stat. Assoc.* **94**, 1275-1285.