Functional Linear Regression Model for Nonignorable Missing Scalar Responses

Tengfei Li[1], *Fengchang Xie*[2], *Xiangnan Feng*[3], *Joseph G. Ibrahim*[4],

Hongtu Zhu[1,4], *and for the Alzheimers Disease Neuroimaging Initiative*

[1] *University of Texas MD Anderson Cancer Center,*

[2] *Nanjing Normal University,* [3] *The Chinese University of Hong Kong, and*

[4] *University of North Carolina at Chapel Hill*

Supplementary Material

S1 Simulations

In this section, we conducted two sets of Monte Carlo simulations to evaluate the finite-sample performance of $\hat{\boldsymbol{\theta}}$ by using four different evaluation measures.

In the first set, we simulated datasets from ETFLR given by

$$Y = 0.5\int_0^1 |\sin(4\pi t)|Z(t)dt + \beta_1 W + \alpha_1 + \epsilon,$$

$$\text{logit}[\Pr(\delta = 1)] = \phi Y - \int_0^1 \sin(4\pi t)Z(t)dt - W + \alpha_2,$$

where $W \sim N(0,1)$, $\epsilon \sim N(0,\sigma^2)$, $Z = Z(t)$ is the standard Brownian motion, and Z, W, and ϵ are mutually independent. Moreover, we set $\alpha_1 = 0$ and $\beta_1 = 0.5$, whereas σ, α_2

and ϕ were varied for comparison purposes. We used α_2 to determine the missingness rate and ϕ to measure how the model is different from MAR. Moreover, we set $\mathbf{t} = \{0, 0.01, \ldots, 0.99, 1\}$ and $K = 100$, and approximated $\int f(t)dt \approx \sum_{k=0}^{100} f(0.01k) \times 0.01$. The values of α and ϕ are key factors for controlling the missingness rate in each scenario. Specifically, the missingness rates are 66%, 60%, 56%, and 52%, when (ϕ, α) takes the values $(1, -1)$, $(2, -1)$, $(1, -0.5)$, and $(2, -0.5)$, respectively. For each simulated dataset, we used GCV to select $k_n \in \{1, 2, \cdots, 20\}$ and empirically fixed $h = n^{-1/3}\sigma$ and $w_0 = 1/2$. Moreover, the Gaussian Kernel $K(t) = \exp(-0.5t^2)/\sqrt{2\pi}$ is utilized in (2.9).

We considered four competing estimates as follows:

- (i) MCAR. Use complete observations to estimate the parameters.
- (ii) MNAR($\hat{k}_n$). Set ϕ as the true value ϕ_0.
- (iii) MNAR($\hat{\phi}, \hat{k}_n$). Calculate $\hat{\phi}$ according to the first approach in subsection 2.2.3.
- (iv) MAR. Set $\phi = 0$.

We simulated $S = 5,000$ data sets for each combination of $(\alpha, \phi, n, N/n)$, in which n and N denote the total sample size and that of validation dataset, respectively. Each test set has the same size as the training set. Let $\hat{y}_i^{(s)}$, $\hat{\boldsymbol{\theta}}^{(s)}$, $\hat{\beta}^{(s)}$, and $\hat{\alpha}_1$ represent the i-th predicted response, the estimated $\boldsymbol{\theta}$, β, and α_1 in the s-th simulation, respectively. We consider four evaluation criteria as follows:

- The prediction bias:

  Bias $= S^{-1} \sum_{s=1}^{S} | \sum_{i=1}^{n} (\hat{y}_i^{(s)} - y_i^{(s)})/n |$;

- The nonfunctional bias:

  $\text{Bias}_n = S^{-1}\sum_{s=1}^{S} |n^{-1}\sum_{i=1}^{n}[(\hat{\beta}_1^{(s)} - 0.5)W + \hat{\alpha}_1^{(s)}]|$;

- Mean integrated squared error (MISE_1):

$$\text{MISE}_1 = \text{Median}[\{\sum_{j=1}^{J}(\hat{\boldsymbol{\theta}}^{(s)}(t_j) - \boldsymbol{\theta}(t_j))^2(t_j - t_{j-1})\}_{s\leq S}] \approx \text{Median}[\int(\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t))^2 dt].$$

- Median integrated squared error (MISE_2):

$$\text{MISE}_2 = \text{Mean}[\{\sum_{j=1}^{J}(\hat{\boldsymbol{\theta}}^{(s)}(t_j) - \boldsymbol{\theta}(t_j))^2(t_j - t_{j-1})\}_{s\leq S}] \approx \text{Mean}[\int(\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t))^2 dt].$$

- Mean squared error for nonfunctional:

$$\text{MSE} = \text{Mean}\{(\hat{\beta}^{(s)} - 0.5)^2 + (\hat{\alpha}_1^{(s)} - 0)^2\}_{s\leq S}.$$

Table A summarizes the simulation results in all scenarios. In terms of Bias, Bias_n, and MSE, MNAR($\hat{k}_n$) outperforms all other methods and MNAR($\hat{\phi}, \hat{k}_n$) is the second best, indicating the advantage of incorporating the nonfunctional part estimation and response prediction for the MNAR methods. When either σ or ϕ becomes larger, MNARs are much better than their competing methods. As both n and N/n increase, the performance of MNAR($\hat{\phi}, \hat{k}_n$) is more similar to that of MNAR($\hat{k}_n$). In terms of MISE_1 and MISE_2, MNAR and MAR have similar performance, whereas MCAR is not very stable and has large MISE_1 particularly when n is small.

In the second set, we generated simulation data sets from ETFLR given by

$$Y = 0.5\langle\boldsymbol{\theta}, Z\rangle + \beta_1 W + \alpha_1 + \epsilon,$$

$$\text{logit}[\Pr(\delta = 1)] = \phi Y - \langle g, Z\rangle + \beta_2 W + \alpha_2,$$

where $\epsilon \sim N(0, \sigma^2)$. We fixed $\beta_1 = \beta_2 = 0$, $\alpha_1 = 0$, and $\alpha_2 = -1$. We consider an image pool consisting of 1457 two-dimensional images and randomly selected $Z_i(\cdot)'s$ from the pool. Figure A (left) presents several randomly selected images.

We consider two scenarios for $(\boldsymbol{\theta}, g)$. In both scenarios, we randomly selected an image from the image pool for each simulated data set. In the first scenario, we chose 7 images for $\boldsymbol{\theta}$ from the image pool. For each $\boldsymbol{\theta}$ image, we generated 1000 simulated data sets. In the second scenario, we generated 5,000 simulated data sets and randomly selected $\boldsymbol{\theta}$ from the image pool in each simulated data set. We evaluated MNAR($\hat{k}_n$), MCAR, MAR and MNAR($\hat{\phi}, \hat{k}_n$) by using the prediction Bias and MISE_1, since they are close to Bias_n and MISE_2, respectively, in the second set.

Table B presents the simulation results for both scenarios. MNARs outperform MCAR and MAR, indicating that selecting the correct missing data mechanism can improve both prediction and estimation accuracy. Figure A (right) presents some randomly selected estimated $\hat{\boldsymbol{\theta}}$'s. These $\hat{\boldsymbol{\theta}}$'s can be quite different from $\boldsymbol{\theta}$ since the basis functions constructed from FPCA may not accurately capture the variation of $\boldsymbol{\theta}$. However, for missing data problem, it remains largely unclear how to choose a set of efficient basis functions to accurately recover both the missing data and the functional signal. We will address this issue in our future research.

Table A. Simulation results based on 5000 replications for the first simulation setting.

test	(ϕ, α, σ^2),	$(n, N/n)$	MNAR($\hat{k}_n$)	MCAR	MAR	MNAR($\hat{\phi}, \hat{k}_n$)
Bias	(1,-1,.25)	(300,1/30)	0.090	0.159	0.158	0.144
Bias_n	(1,-1,.25)	(300,1/30)	0.129	0.188	0.192	0.179
MISE_1	(1,-1,.25)	(300,1/30)	0.166	1.336	0.186	0.195
MISE_2	(1,-1,.25)	(300,1/30)	0.065	0.070	0.065	0.066
MSE	(1,-1,.25)	(300,1/30)	0.011	0.017	0.018	0.017
Bias	(1,-1,.64)	(300,1/30)	0.137	0.376	0.378	0.242
Bias_n	(1,-1,.64)	(300,1/30)	0.189	0.411	0.393	0.278
MISE_1	(1,-1,.64)	(300,1/30)	0.223	3.340	0.199	0.205
MISE_2	(1,-1,.64)	(300,1/30)	0.077	0.099	0.089	0.082
MSE	(1,-1,.64)	(300,1/30)	0.024	0.070	0.067	0.042
Bias	(2,-1,.25)	(300,1/30)	0.200	0.534	0.536	0.267
Bias_n	(2,-1,.25)	(300,1/30)	0.262	0.608	0.563	0.324
MISE_1	(2,-1,.25)	(300,1/30)	0.179	3.260	0.165	0.158
MISE_2	(2,-1,.25)	(300,1/30)	0.074	0.101	0.077	0.074
MSE	(2,-1,.25)	(300,1/30)	0.033	0.132	0.117	0.047
Bias	(2,-1,.64)	(300,1/30)	0.138	0.260	0.264	0.187
Bias_n	(2,-1,.64)	(300,1/30)	0.174	0.301	0.272	0.213
MISE_1	(2,-1,.64)	(300,1/30)	0.097	1.349	0.104	0.099
MISE_2	(2,-1,.64)	(300,1/30)	0.062	0.069	0.066	0.062
MSE	(2,-1,.64)	(300,1/30)	0.014	0.034	0.030	0.020
Bias	(1,-1,.64)	(600,1/10)	0.116	0.375	0.377	0.166
Bias_n	(1,-1,.64)	(600,1/10)	0.160	0.406	0.397	0.204
MISE_1	(1,-1,.64)	(600,1/10)	0.143	0.798	0.130	0.140
MISE_2	(1,-1,.64)	(600,1/10)	0.072	0.089	0.077	0.075
MSE	(1,-1,.64)	(600,1/10)	0.065	0.070	0.069	0.065
Bias	(1,-1,.64)	(300,1/10)	0.137	0.376	0.378	0.200
Bias_n	(1,-1,.64)	(300,1/10)	0.189	0.406	0.393	0.240
MISE_1	(1,-1,.64)	(300,1/10)	0.223	1.443	0.199	0.202
MISE_2	(1,-1,.64)	(300,1/10)	0.077	0.088	0.089	0.079
MSE	(1,-1,.64)	(300,1/10)	0.024	0.068	0.067	0.032
Bias	(1,-0.5,.64)	(300,1/30)	0.105	0.315	0.318	0.203
Bias_n	(1,-0.5,.64)	(300,1/30)	0.153	0.352	0.340	0.237
MISE_1	(1,-0.5,.64)	(300,1/30)	0.207	3.912	0.190	0.211
MISE_2	(1,-0.5,.64)	(300,1/30)	0.072	0.089	0.077	0.075
MSE	(1,-0.5,.64)	(300,1/30)	0.016	0.052	0.050	0.031
Bias	(1,-0.5,.25)	(600,1/3)	0.059	0.134	0.134	0.078
Bias_n	(1,-0.5,.25)	(600,1/3)	0.094	0.159	0.162	0.112
MISE_1	(1,-0.5,.25)	(600,1/3)	0.095	0.131	0.103	0.096
MISE_2	(1,-0.5,.25)	(600,1/3)	0.057	0.057	0.058	0.057
MSE	(1,-0.5,.25)	(600,1/3)	0.005	0.011	0.011	0.006
Bias	(2,-0.5,.25)	(300,1/30)	0.103	0.216	0.220	0.150
Bias_n	(2,-0.5,.25)	(300,1/30)	0.140	0.257	0.239	0.180
MISE_1	(2,-0.5,.25)	(300,1/30)	0.094	1.100	0.098	0.099
MISE_2	(2,-0.5,.25)	(300,1/30)	0.059	0.067	0.061	0.060
MSE	(2,-0.5,.25)	(300,1/30)	0.010	0.025	0.023	0.015
Bias	(2,-0.5,.25)	(300,1/5)	0.103	0.217	0.219	0.131
Bias_n	(2,-0.5,.25)	(300,1/5)	0.140	0.251	0.239	0.166
MISE_1	(2,-0.5,.25)	(300,1/5)	0.094	0.335	0.098	0.094
MISE_2	(2,-0.5,.25)	(300,1/5)	0.059	0.062	0.061	0.059
MSE	(2,-0.5,.25)	(300,1/5)	0.010	0.024	0.023	0.012
Bias	(2,-0.5,.25)	(600,1/2)	0.092	0.217	0.219	0.110
Bias_n	(2,-0.5,.25)	(600,1/2)	0.131	0.252	0.248	0.149
MISE_1	(2,-0.5,.25)	(600,1/2)	0.078	0.136	0.079	0.076
MISE_2	(2,-0.5,.25)	(600,1/2)	0.055	0.056	0.056	0.055
MSE	(2,-0.5,.25)	(600,1/2)	0.007	0.023	0.022	0.009

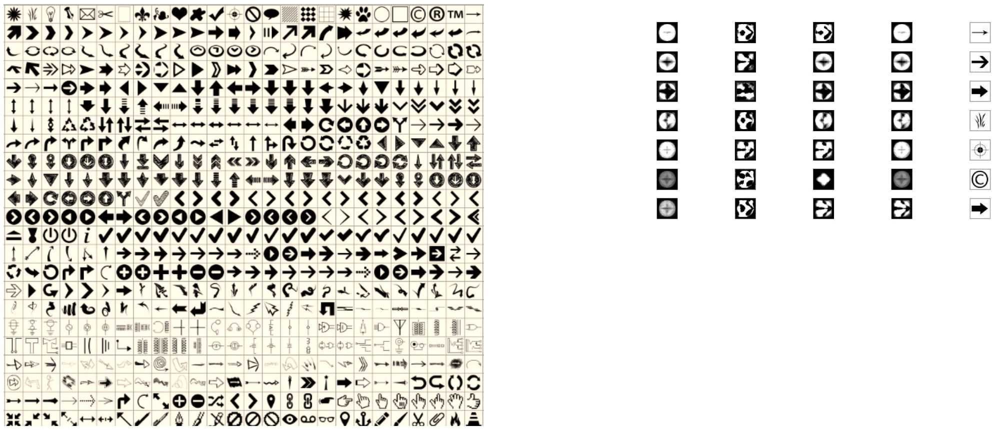

Figure A. The image pool used in the second simulation setting (left); and the selected 7 images in the second simulation setting—true value and their estimates (right), where the 5 columns represent the corresponding image estimates using the MNAR($\hat{k}_n$), MCAR, MAR, and MNAR($\hat{\phi}, \hat{k}_n$), and the true image, respectively.

Table B. Simulation results based on 2-dimensional image covariates for the second simulation setting. Seven identification (ID) numbers were assigned to 7 randomly chosen images (see Figure A) each producing 1000 replications of data sets. ID='-' indicates data set were generated with randomly chosen image coefficient in each replication.

test	ID	(ϕ, σ^2),	$(n, N/n)$	MNAR($\hat{k}_n$)	MCAR	MAR	MNAR($\hat{\phi}, \hat{k}_n$)
Bias	1	(1,.25)	(600,1/10)	0.060	0.077	0.071	0.059
MISE_1	1	(1,.25)	(600,1/10)	0.636	0.901	0.704	0.634
Bias	2	(1,.25)	(1000,1/10)	0.063	0.081	0.076	0.062
MISE_1	2	(1,.25)	(1000,1/10)	0.689	0.902	0.751	0.696
Bias	3	(1,.25)	(600,1/10)	0.064	0.082	0.076	0.063
MISE_1	3	(1,.25)	(600,1/10)	0.731	0.958	0.770	0.737
Bias	4	(1,.25)	(600,1/10)	0.060	0.078	0.072	0.059
MISE_1	4	(1,.25)	(600,1/10)	0.660	0.910	0.726	0.669
Bias	5	(1,.25)	(600,1/10)	0.060	0.078	0.072	0.059
MISE_1	5	(1,.25)	(600,1/10)	0.642	0.899	0.701	0.649
Bias	6	(1,.25)	(600,1/10)	0.064	0.083	0.077	0.064
MISE_1	6	(1,.25)	(600,1/10)	0.673	0.893	0.758	0.674
Bias	7	(1,.25)	(600,1/10)	0.064	0.082	0.076	0.064
MISE_1	7	(1,.25)	(600,1/10)	0.733	0.934	0.767	0.741
Bias	-	(1,.25)	(600,1/5)	0.065	0.082	0.077	0.064
MISE_1	-	(1,.25)	(600,1/5)	0.671	0.814	0.746	0.670
Bias	-	(1,.64)	(600,1/5)	0.114	0.200	0.192	0.113
MISE_1	-	(1,.64)	(600,1/5)	0.707	1.141	0.960	0.708
Bias	-	(2,.25)	(600,1/5)	0.074	0.097	0.093	0.074
MISE_1	-	(2,.25)	(600,1/5)	0.629	0.778	0.686	0.630
Bias	-	(1,.64)	(1000,1/3)	0.114	0.199	0.194	0.113
MISE_1	-	(1,.64)	(1000,1/3)	0.676	0.885	0.867	0.676

S2 Assumptions

Let C be a generic constant.

(**A.1**) $\boldsymbol{\theta}(\cdot) \in L^2([0,1]) = \{f : [0,1] \to \mathbb{R} | \int_0^1 f^2(t)dt < \infty\}$.

(**A.2**) The function $Z \in \mathbb{H}$, is centered: $\mathrm{E}(Z) = 0$ and has the decomposition

$$Z(\cdot) = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\xi_j v_j(\cdot),$$

where the ξ_j's are independent real random variables with zero mean and unit variance. For all $j, l \in \mathbb{N}$, there exists a constant b such that $\mathrm{E}|\xi_j|^l \leq l! b^{l-2} \mathrm{E}(|\xi_j|^2)/2$.

(**A.3**) $\lambda_j - \lambda_{j+1} \geq Cj^{-a-1}$ for $j \geq 1$ and $a > 1$.

(**A.4**) $\langle \boldsymbol{\theta}_0, v_j \rangle \leq Cj^{-b}$ for $j \geq 1$ and $b > 1 + a/2$.

(**A.5**) For any $C > 0$, there exists a $\tau_0 > 0$ such that

$$\sup_{s \in [0,1]} \{\mathrm{E}(Z(t))^C\} < \infty \text{ and } \sup_{t_1, t_2 \in [0,1]} \mathrm{E}|t_1 - t_2|^{-\tau_0} |Z(t_1) - Z(t_2)|^C < \infty.$$

(**A.6**) $\max\{E(\|W\|^2), \sigma^2\} \leq C < \infty$, and ϵ is independent of Z and W.

(**A.7**) $k_n \to \infty$ and $k_n^{5a+3} n^{-1} \to 0$ as $n \to \infty$.

(**A.8**) $\mathfrak{J} \triangleq \mathrm{E}(W^{\otimes 2}) - \sum_{j=1}^{\infty} \mathrm{E}(W\xi_j)\mathrm{E}(W^T \xi_j) > 0$.

(**A.9**) The true value $\phi = \phi_0$ is known.

(**B.1**) There exists a constant C_1 such that $\max(\|Z\|, \|W\|) \leq C_1$.

(**B.2**) There exists a monotone and continuous function $G_1 : \mathbb{R} \to \mathbb{R}$ such that $G_1(G(Z, W)) \leq \|Z\| + \|W\|$ and the Lipschitz's condition holds such that $|G(Z_1, W_1) - G(Z_2, W_2)| < C(\|Z_1 - Z_2\| + \|W_1 - W_2\|)$.

(B.3) K is a kernel of type I (Martinez, 2013) if the function $K : \mathbb{R} \to [0, \infty)$ satisfies $\int_0^\infty K(u)du = 1$ and there exist constants c_1 and $c_2 \in \mathbb{R}$ such that $c_1 1_{u\in[0,1]} \leq K(u) \leq c_2 1_{u\in[0,1]}$ holds for $0 < c_1 < c_2 < \infty$.

(B.4) There exist a function ψ and a constant $\delta > 0$ such that $\forall \tau \in (0, \delta)$ and $(z, x) \in H_0 \triangleq \{(z, x)|z \in H, x \in \mathbb{R}^p, \max\{\|z\|, \|x\|\} \leq C_1\}$, we have $\psi_{z,x}(\tau \geq \psi(\tau) \geq 0$, where $\psi_{z,x}(\tau) \triangleq \Pr[(Z, W) \in \{(\tilde{z}, \tilde{x})| \ \|\tilde{z} - z\| \leq \tau, \|\tilde{x} - x\| \leq \tau\}]$.

(B.5) $k_n^{a+1}[h + 1/\sqrt{n\psi(h)}] \to 0$ as $n \to \infty$.

(B.6) The weight parameter $w_0 \in (0, 1)$.

S3 Proofs

Lemma 1–Lemma 9 and Lemma 10–Lemma 13 are listed in the following for proofs of Theorem 1 and 2, respectively. Proofs of Lemma 1, 2, and Lemma 3- (i) can be found in Lemma 3.3 of Hall and Hosseini-Nasab (2009), Theorem 3 of Hall and Hosseini-Nasab (2006), and Proposition 18 of Crambes and Andr (2013), respectively. Before continuing, we define the following operators and notations.

First we define $x_n \preceq y_n$ or $x_n = O_p(y_n)$ for random sequences (x_n) and (y_n), if for any $\tau > 0$, there exist $M_\tau > 0$,and $N > 0$ such that for any $n > N$, $\Pr(|x_n/y_n| > M_\tau) < \tau$; $x_n \succeq y_n$, or $y_n = O_p(x_n)$, if for any $\tau > 0$, there exists $M_\tau > 0$ such that $\Pr(|y_n/x_n| > M_\tau) < \tau$; $x_n \ll y_n$ or $x_n = o_p(y_n)$, if for any $\tau > 0$, $\Pr(|x_n/y_n| > \tau) \to 0$; $x_n \gg y_n$ or $y_n = o_p(x_n)$, if for any $\tau > 0$, $\Pr(|y_n/x_n| > \tau) \to 0$; $x_n \sim y_n$, if $x_n \preceq y_n$ as well as $x_n \succeq y_n$. Second, for an arbitrary bivariate function $f : \mathfrak{D}_x \times \mathfrak{D}_y \mapsto \mathfrak{R}$, random variables $\xi : \Omega \mapsto \mathfrak{D}_x$, and $\eta : \Omega \mapsto \mathfrak{D}_y$, if $\mathrm{E}[f(x,\eta)] < \infty$ for any $x \in \mathfrak{D}_x$, define the notation $\mathrm{E}_{-\xi} f(\xi,\eta)$ by $\mathrm{E}_{-\xi} f(\xi,\eta) = g(\xi)$ where $g : \mathfrak{D}_x \mapsto \mathfrak{R}, g(x) = \mathrm{E}[f(x,\eta)]$ for any $x \in \mathfrak{D}_x$. Note that if ξ and η are independent, $\mathrm{E}_{-\xi} f(\xi,\eta) = \mathrm{E}[f(\xi,\eta)|\xi]$. Third, we denote

$$r_j^* = \frac{\sum_{i=1}^n [\delta_i M_j(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_{1,0}) + (1-\delta_i) m^0_{M_j,i,\gamma}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_{1,0})]}{n\lambda_j},$$

$$\hat{r}_j^* = \frac{\sum_{i=1}^n [\delta_i M_j(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_{1,0}) + (1-\delta_i) \hat{m}_{M_j,i,\gamma}](Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_{1,0})}{n\lambda_j},$$

$\gamma_0 = -\phi_0$, $\zeta_j = \min_{k \le j} |\lambda_k - \lambda_{k+1}|$, and define $F(\langle Z_l, \boldsymbol{\theta}_0 \rangle, G(\langle Z_l, W_l \rangle), W_l)$ as the following conditional expectation

$$\mathrm{E}(\mathrm{E}[\langle Z_l, \boldsymbol{\theta}_0\rangle \delta_l \exp(\gamma_0 Y_l)|Z_l, W_l]|\langle Z_l, \boldsymbol{\theta}_0\rangle, G(\langle Z_l, W_l\rangle), W_l).$$

Furthermore, since $Z_i, i = 1, 2, \cdots, n$ are n independent and identically distributed realizations of Z, from Assumption (A.2), there exist random variables $\xi_j^{(i)}, i \le n, j \in \mathbb{Z}_+$ such that $Z_i = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \xi_j^{(i)} v_j, i, j \in \mathbb{Z}_+$, where $\xi_j^{(i)}, i = 1, 2, \cdots, n$ are n independent and identically distributed realizations of ξ_j, and $\xi_j^{(i)}$ are mutually independent for $i \le n, j \in \mathbb{Z}_+$. Finally, for a kernel function $K : R \mapsto [0, +\infty)$, and $K_h : R \mapsto [0, +\infty), K_h(\cdot) = K(\cdot/h)$ we define random functions $K_h^{(l)}(\cdot)$ as

$$K_h^{(l)}(\cdot) = K_h(\cdot)\delta_l \exp(\gamma_0 Y_l)\langle Z_l, \boldsymbol{\theta}_0\rangle,$$

and $\tilde{K}_h^{(l)}(\cdot)$ as

$$\tilde{K}_h^{(l)}(\cdot) = K_h(\cdot)\delta_l \exp(\gamma_0 Y_l),$$

for $l = 1, 2, \cdots, n$ and $j \in \mathbb{Z}_+$.

Lemma 1. *Assume that with probability 1, X is left-continuous at each point (or right-continuous at each point), and that Conditions (B3) and (B4) hold. Then, for each $C > 0$, $\mathrm{E}(\|\hat{\mathcal{K}} - \mathcal{K}\|^C) <$ constant $* n^{-C/2}$, where $\mathcal{K}$ and $\hat{\mathcal{K}}$ are the covariance and the sample covariance function of the process $Z(\cdot)$, and*

$$\|\hat{\mathcal{K}} - \mathcal{K}\| \triangleq \sqrt{\int_{[0,1]^2} [\hat{\mathcal{K}}(s_1, s_2) - \mathcal{K}(s_1, s_2)]^2 ds_1 ds_2}$$

Lemma 2 *Under Assumptions (A.2) and (A.3), we have*

$$\|\hat{v}_j - v_j\| \le 8^{1/2}\zeta_j^{-1}\|\hat{\mathcal{K}} - \mathcal{K}\| \text{ for any } j.$$

Lemma 3

(i) Under Assumptions (A.2) and (A.3), when j is large enough,

$$0 < constant \times j^{-a} \leq \lambda_j < constant \times j^{-1};$$

(ii) under Assumptions (A.2) and (A.5), we have

$$\Pr(\overline{\lim_{k\to\infty}} \bigcup_{j\leq k} |\hat{\lambda}_j - \lambda_j| > (\lambda_j - \lambda_{j+1})/2) = 0,$$

which implies

$$\Pr(\overline{\lim_{k\to\infty}} \bigcup_{j\leq k} \hat{\lambda}_j < \lambda_{j+1}) = 0.$$

Proof of Lemma 3, (i).

From $\mathrm{E}\langle Z, Z\rangle = \sum_{j=1}^{\infty} \langle Z, v_j\rangle^2 = \sum_{j=1}^{\infty} \lambda_j < \infty$, we have $\lambda_j \ll j^{-1}$; from $\lambda_j = \sum_{k=j}^{\infty}(\lambda_k - \lambda_{k+1})$ and Assumption (A.3) we have $\lambda_j \gg j^{-a}$.

□

Lemma 4. *Under Assumptions (A.1), (A.2), (A.6) and (A.9), we have*

(i)

$$m^0_{\tilde{M},i,\gamma}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) = \mathrm{E}[\widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) | \delta_i = 0, Z_i, W_i];$$

(ii)

$$L_4 \triangleq \frac{1}{n}\sum_{i=1}^{n}[\delta_i \widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) + (1-\delta_i) m^0_{\tilde{M},i,\gamma}(Y_i, W_i; \boldsymbol{\beta}_{1,0})] - \mathrm{E}[\langle \boldsymbol{\theta}_0, Z\rangle W] = O_p(1/\sqrt{n}).$$

Proof. Part (i) is shown in the following.

$$
\begin{aligned}
m^0_{\tilde{M},i,\gamma}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) &= \frac{\mathrm{E}\{\delta_i \widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) \exp(\gamma Y_i) | Z_i, W_i\}}{\mathrm{E}\{\delta_i \exp(\gamma Y_i) | Z_i, W_i\}} \\
&= \frac{\mathrm{E}\{\mathrm{E}[\delta_i \widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) \exp(\gamma Y_i) | Z_i, Y_i, W_i] | Z_i, W_i\}}{\mathrm{E}\{\mathrm{E}[\delta_i \exp(\gamma Y_i) | Z_i, Y_i, W_i] | Z_i, W_i\}} \\
&= \frac{\mathrm{E}\{\mathrm{E}[\delta_i | Z_i, W_i, Y_i] \widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) \exp(\gamma Y_i) | Z_i, W_i\}}{\mathrm{E}\{\mathrm{E}[\delta_i | Z_i, W_i, Y_i] \exp(\gamma Y_i) | Z_i, W_i\}} \\
&= \frac{\mathrm{E}\{\Pr(\delta_i = 1 | Y_i, Z_i, W_i) \widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) \exp(\gamma Y_i) | Z_i, W_i\}}{\mathrm{E}\{\Pr(\delta_i = 1 | Y_i, Z_i, W_i) \exp(\gamma Y_i) | Z_i, W_i\}} \\
&= \frac{\mathrm{E}\{\Pr(\delta_i = 0 | Y_i, Z_i, W_i) \widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) | Z_i, W_i\}}{\mathrm{E}\{\Pr(\delta_i = 0 | Y_i, Z_i, W_i) | Z_i, W_i\}} \\
&= \frac{\mathrm{E}\{(1-\delta_i) \widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) | Z_i, W_i\}}{\mathrm{E}\{(1-\delta_i) | Z_i, W_i\}} = \mathrm{E}[\widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) | \delta_i = 0, Z_i, W_i].
\end{aligned}
$$

To prove (ii), first we calculate the expectation as below.

$$
\begin{aligned}
& \mathrm{E}[\delta_i \widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) + (1-\delta_i) m^0_{\tilde{M},i,\gamma}(Y_i, W_i; \boldsymbol{\beta}_{1,0})] \\
= \; & \mathrm{E}\left\{\delta_i \widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) + (1-\delta_i) \mathrm{E}[\widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) | \delta_i = 0, Z_i, W_i]\right\} \\
= \; & \Pr(\delta_i = 1) \mathrm{E}[\widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) | \delta_i = 1] + \Pr(\delta_i = 0) \mathrm{E}\{\mathrm{E}[\widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) | \delta_i = 0, Z_i, W_i] | \delta_i = 0\} \\
= \; & \Pr(\delta_i = 1) \mathrm{E}[\widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) | \delta_i = 1] + \Pr(\delta_i = 0) \mathrm{E}\{\widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) | \delta_i = 0\} \\
= \; & \mathrm{E}[\widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0})] = \mathrm{E}[\langle \boldsymbol{\theta}_0, Z \rangle W].
\end{aligned}
$$

Second we calculate the variance, using the independence across different subjects.

$$
\begin{aligned}
& \mathrm{E}^2\{\frac{1}{n} \sum_{i=1}^n [\delta_i \widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) + (1-\delta_i) m^0_{\tilde{M},i,\gamma}(Y_i, W_i; \boldsymbol{\beta}_{1,0})] - \mathrm{E}[\langle \boldsymbol{\theta}_0, Z \rangle W]\} \\
= \; & \frac{1}{n^2} \mathrm{E}^2 \sum_{i=1}^n \{[\delta_i \widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) + (1-\delta_i) m^0_{\tilde{M},i,\gamma}(Y_i, W_i; \boldsymbol{\beta}_{1,0})] - \mathrm{E}[\langle \boldsymbol{\theta}_0, Z \rangle W]\} \\
= \; & \frac{1}{n} \mathrm{E}^2 \{[\delta_i \widetilde{M}(Y_i, W_i; \boldsymbol{\beta}_{1,0}) + (1-\delta_i) m^0_{\tilde{M},i,\gamma}(Y_i, W_i; \boldsymbol{\beta}_{1,0})] - \mathrm{E}[\langle \boldsymbol{\theta}_0, Z \rangle W]
\end{aligned}
$$

$$\begin{aligned}
&\leq \frac{1}{n}\mathrm{E}^2[\delta_i\widetilde{M}(Y_i,W_i;\boldsymbol{\beta}_{1,0})+(1-\delta_i)m^0_{\tilde{M},i,\gamma}(Y_i,W_i;\boldsymbol{\beta}_{1,0})] \\
&= \frac{1}{n}\mathrm{E}^2\{\delta_i\widetilde{M}(Y_i,W_i;\boldsymbol{\beta}_{1,0})+(1-\delta_i)\mathrm{E}[\widetilde{M}(Y_i,W_i;\boldsymbol{\beta}_{1,0})|\delta_i=0,Z_i,W_i]\} \\
&\leq \frac{2}{n}\mathrm{E}^2[\widetilde{M}(Y_i,W_i;\boldsymbol{\beta}_{1,0})]\}=O_p(1/n).
\end{aligned}$$

Finally, we have

$$\frac{1}{n}\sum_{i=1}^{n}[\delta_i\widetilde{M}(Y_i,W_i;\boldsymbol{\beta}_{1,0})+(1-\delta_i)m^0_{\tilde{M},i,\gamma}(Y_i,W_i;\boldsymbol{\beta}_{1,0})]=\mathrm{E}[\langle\boldsymbol{\theta}_0,Z\rangle W]+O_p(1/\sqrt{n}).$$

□

Lemma 5. *Suppose (Z,W,Y,δ) is independently and identically distributed with $(Z_i,W_i,Y_i,\delta_i), i=1,2,\cdots,n$. Then under Assumptions (A.1), (A.2), (A.4), (A.6), (A.7), and (A.9), we have*

(i)

$$\mathrm{E}r_j^*=\langle\boldsymbol{\theta}_0,v_j\rangle.$$

(ii) $L_5 \triangleq$

$$\mathrm{E}\sum_{j=1}^{k_n}r_j^*[\delta W\langle Z,v_j\rangle+(1-\delta)\frac{\mathrm{E}\{\delta\langle Z,v_j\rangle W\exp(\gamma Y)|X,V\}}{\mathrm{E}\{\delta\exp(\gamma Y)|X,V\}}]-\mathrm{E}[\langle\boldsymbol{\theta}_0,Z\rangle W]=O(k_n^{1/2-b}).$$

Proof. Similar to the proof of Lemma 1, we have

$$\begin{aligned}
\mathrm{E}r_j^* &= \mathrm{E}M_j(Y_i,Z_i,W_i,v_j;\boldsymbol{\beta}_1)/\lambda_j=\langle\boldsymbol{\theta}_0,v_j\rangle; \qquad \text{(S3.1)}\\
U_1 &\triangleq \mathrm{E}\sum_{j=1}^{k_n}\langle\boldsymbol{\theta}_0,v_j\rangle[\delta W\langle Z,v_j\rangle+(1-\delta)\frac{\mathrm{E}\{\delta\langle Z,v_j\rangle W\exp(\gamma Y)|X,V\}}{\mathrm{E}\{\delta\exp(\gamma Y)|X,V\}}] \\
&= \sum_{j=1}^{k_n}\langle\boldsymbol{\theta}_0,v_j\rangle\mathrm{E}[\langle Z,v_j\rangle W]=\mathrm{E}[\langle\boldsymbol{\theta}_0,Z\rangle W]-\sum_{j=k_n+1}^{\infty}\langle\boldsymbol{\theta}_0,v_j\rangle\mathrm{E}[\langle Z,v_j\rangle W] \qquad \text{(S3.2)}
\end{aligned}$$

Hence (i) has been proved. To prove (ii), using the independence of (Z, W, Y, δ) with $(Z_i, W_i, Y_i, \delta_i), i = 1, 2, \cdots, n$, and following (S3.2), we have

$$\begin{aligned} L_5 &= \mathrm{E}\sum_{j=1}^{k_n}\mathrm{E}(r_j^*)[\delta W\langle Z, v_j\rangle + (1-\delta)\frac{\mathrm{E}\{\delta\langle Z, v_j\rangle W\exp(\gamma Y)|X, V\}}{\mathrm{E}\{\delta\exp(\gamma Y)|X, V\}}] - \mathrm{E}[\langle\boldsymbol{\theta}_0, Z\rangle W] \\ &= U_1 - \mathrm{E}[\langle\boldsymbol{\theta}_0, Z\rangle W] = -\sum_{j=k_n+1}^{\infty}\langle\boldsymbol{\theta}_0, v_j\rangle\mathrm{E}[\langle Z, v_j\rangle W]. \end{aligned}$$

It follows that

$$\begin{aligned} |L_5| = \sum_{j=k_n+1}^{\infty}\langle\boldsymbol{\theta}_0, v_j\rangle\mathrm{E}[\langle Z, v_j\rangle W] &= \sum_{j=k_n+1}^{\infty}\sqrt{\lambda_j}\langle\boldsymbol{\theta}_0, v_j\rangle\mathrm{E}(\xi_j W) \\ \leq \sum_{j=k_n+1}^{\infty}\sqrt{\lambda_j}\langle\boldsymbol{\theta}_0, v_j\rangle\sqrt{\mathrm{E}(W^2)\mathrm{E}\xi_j^2} &= \sqrt{\mathrm{E}W^2}\sum_{j=k_n+1}^{\infty}\sqrt{\lambda_j}\langle\boldsymbol{\theta}_0, v_j\rangle. \end{aligned}$$

Together with $\lambda_j \preceq j^{-1}$ in Lemma 3 and $\langle\boldsymbol{\theta}_0, v_j\rangle \preceq j^{-b}$ in Assumption (A.4), we have $L_5 = O(k_n^{1/2-b})$. □

Lemma 6. *Under Assumptions (A.1)–(A.7), and (A.9), we have*

(i)

$$\sum_{j=1}^{k_n} j^x\lambda_j(r_j^* - \langle\boldsymbol{\theta}_0, v_j\rangle)^2 = O_p(k_n^{1+x}/n), \text{ for any } x > -1.$$

(ii)

$$\sup_{j\leq k_n}\zeta_j\lambda_j|\Delta_2(r_j)| = O_p(k_n^{a-1}/\sqrt{n}),$$

where

$$\Delta_2(r_j) = r_j - \frac{\sum_{i=1}^n[\delta_i M_j(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_1) + (1-\delta_i)m^0_{M_j,i,\gamma}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_1)]}{n\hat{\lambda}_j}.$$

(iii)

$$\sum_{j=1}^{k_n}\lambda_j|r_j - r_j^*|^2 = O_p(k_n^{4a+2}n^{-1} + k_n^{-2b});$$

(iv)

$$\sum_{j=1}^{k_n} \mathrm{E}\frac{[\frac{1}{n}\sum_{i=1}^{n}\big(\delta_i W_i\langle Z_i, v_j\rangle + (1-\delta_i)\mathrm{E}(W_i\langle Z_i, v_j\rangle|Z_i, W_i, \delta_i=0)\big)]^2}{\lambda_j} = O(1);$$

(v)

$$\sup_j \zeta_j|\frac{1}{n}\sum_{i=1}^{n}\delta_i W_i\langle Z_i, \hat{v}_j - v_j\rangle| + \sup_j \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{-\hat{v}_j}[\zeta_j|W_i\langle Z_i, \hat{v}_j - v_j\rangle|\big|Z_i, W_i, \delta_i=0] = O_p(1/\sqrt{n}).$$

(vi) $L_6 \triangleq$

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\Delta_1\left\{\sum_{j=1}^{k_n} r_j\big[\delta_i W_i\langle Z_i, \hat{v}_j\rangle + (1-\delta_i)\mathrm{E}_{-\hat{v}_j}\{\langle Z_i, \hat{v}_j\rangle W|Z_i, W_i, \delta_i=0\}\big]\right\}\\ = \;& O_p(k_n^{2a+1}n^{-1/2} + k_n^{-b}),\end{aligned}$$

where

$$\begin{aligned}&\Delta_1\left\{\sum_{j=1}^{k_n} r_j\big[\delta_i W_i\langle Z_i, \hat{v}_j\rangle + (1-\delta_i)\mathrm{E}_{-\hat{v}_j}\{\langle Z_i, \hat{v}_j\rangle W|Z_i, W_i, \delta_i=0\}\big]\right\}\\ = &\left\{\sum_{j=1}^{k_n} r_j\big[\delta_i W_i\langle Z_i, \hat{v}_j\rangle + (1-\delta_i)\mathrm{E}_{-\hat{v}_j}\{\langle Z_i, \hat{v}_j\rangle W|Z_i, W_i, \delta_i=0\}\big]\right\}\\ - &\left\{\sum_{j=1}^{k_n} r_j^*\big[\delta_i W_i\langle Z_i, v_j\rangle + (1-\delta_i)\mathrm{E}\{\langle Z_i, v_j\rangle W|Z_i, W_i, \delta_i=0\}\big]\right\}.\end{aligned}$$

Proof of Lemma 6 (i). Using conclusion 1 of Lemma 5, we have

$$\begin{aligned}&\mathrm{E}\sum_{j=1}^{k_n} j^x\lambda_j(r_j^* - \langle\boldsymbol{\theta}_0, v_j\rangle)^2\\ = \;&\mathrm{E}\sum_{j=1}^{k_n} j^x(\sum_{i=1}^{n}\frac{\delta_i M_j(Y_i, Z_i, W_i, v_j;\boldsymbol{\beta}_1) + (1-\delta_i)m^0_{M_j,i,\gamma}(Y_i, Z_i, W_i, v_j;\boldsymbol{\beta}_1)}{n\sqrt{\lambda_j}} - \sqrt{\lambda_j}\langle\boldsymbol{\theta}_0, v_j\rangle)^2\\ = \;&\sum_{j=1}^{k_n} j^x\frac{1}{n^2}\mathrm{E}\sum_{i=1}^{n}(\frac{\delta_i(M_j(Y_i, Z_i, W_i, v_j;\boldsymbol{\beta}_1) - \lambda_j\langle\boldsymbol{\theta}_0, v_j\rangle)}{\sqrt{\lambda_j}}\end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{\sum_{i=1}^{n}(1-\delta_i)(m^0_{M_j,i,\gamma}(Y_i,Z_i,W_i,v_j;\boldsymbol{\beta}_1)-\lambda_j\langle\boldsymbol{\theta}_0,v_j\rangle)}{\sqrt{\lambda_j}})^2 \\
&\le \quad \frac{2}{n}\sum_{j=1}^{k_n} j^x \mathrm{E}(\frac{\delta_i(M_j(Y_i,Z_i,W_i,v_j;\boldsymbol{\beta}_1)-\lambda_j\langle\boldsymbol{\theta}_0,v_j\rangle)}{\sqrt{\lambda_j}})^2 \\
&+ \quad \frac{2}{n}\sum_{j=1}^{k_n} j^x \mathrm{E}(\frac{(1-\delta_i)(m^0_{M_j,i,\gamma}(Y_i,Z_i,W_i,v_j;\boldsymbol{\beta}_1)-\lambda_j\langle\boldsymbol{\theta}_0,v_j\rangle)}{\sqrt{\lambda_j}})^2 \triangleq \frac{2}{n}(A^{(6,1)}+B^{(6,1)}).
\end{aligned}
$$

In the following the order of $A^{(6,1)}$ and $B^{(6,1)}$ are calculated separately.

$$
\begin{aligned}
\mathrm{E}A^{(6,1)} &= \sum_{j=1}^{k_n} j^x \mathrm{E}(\frac{\delta_i(M_j(Y_i,Z_i,W_i,v_j;\boldsymbol{\beta}_1)-\lambda_j\langle\boldsymbol{\theta}_0,v_j\rangle)}{\sqrt{\lambda_j}})^2 \\
&\le \sum_{j=1}^{k_n} j^x 2\mathrm{E}[(\frac{M_j(Y_i,Z_i,W_i,v_j;\boldsymbol{\beta}_1)}{\sqrt{\lambda_j}})^2+(\frac{(\lambda_j\langle\boldsymbol{\theta}_0,v_j\rangle)}{\sqrt{\lambda_j}})^2] \\
&\le \sum_{j=1}^{k_n} j^x 2[\mathrm{E}(\frac{\langle Z,v_j\rangle(\langle\boldsymbol{\theta},Z\rangle+\epsilon)}{\sqrt{\lambda_j}})^2+(j^{-1/2-b})^2] \\
&= \sum_{j=1}^{k_n} j^x 2[\mathrm{E}(\xi_j(\langle\boldsymbol{\theta},Z\rangle+\epsilon))^2+(j^{-1/2-b})^2] \\
&\le (\sqrt{\mathrm{E}\xi_1^4}\sqrt{\mathrm{E}(\langle\boldsymbol{\theta},Z\rangle+\epsilon)^4}+constant)\sum_{j=1}^{k_n} j^x = O(k_n^{1+x});
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}B^{(6,1)} &= \sum_{j=1}^{k_n} j^x \mathrm{E}(\frac{(1-\delta_i)(\mathrm{E}(M_j(Y_i,Z_i,W_i,v_j;\boldsymbol{\beta}_1)|Z_i,W_i,\delta_i=0)-\lambda_j\langle\boldsymbol{\theta}_0,v_j\rangle)}{\sqrt{\lambda_j}})^2 \\
&\le \sum_{j=1}^{k_n} j^x 2\mathrm{E}[(\frac{\mathrm{E}(M_j(Y_i,Z_i,W_i,v_j;\boldsymbol{\beta}_1)|Z_i,W_i,\delta_i=0)}{\sqrt{\lambda_j}})^2+(\frac{(\lambda_j\langle\boldsymbol{\theta}_0,v_j\rangle)}{\sqrt{\lambda_j}})^2] \\
&\le \sum_{j=1}^{k_n} j^x 2\mathrm{E}[\mathrm{E}(\frac{M_j^2(Y_i,Z_i,W_i,v_j;\boldsymbol{\beta}_1)}{\lambda_j}|Z_i,W_i,\delta_i=0)+(\frac{(\lambda_j\langle\boldsymbol{\theta}_0,v_j\rangle)}{\sqrt{\lambda_j}})^2] \\
&\le \sum_{j=1}^{k_n} j^x 2\mathrm{E}[(\frac{M_j(Y_i,Z_i,W_i,v_j;\boldsymbol{\beta}_1)}{\sqrt{\lambda_j}})^2+(\frac{(\lambda_j\langle\boldsymbol{\theta}_0,v_j\rangle)}{\sqrt{\lambda_j}})^2] = O(k_n^{1+x}),
\end{aligned}
$$

Combine them together, and we have $\mathrm{E}\sum_{j=1}^{k_n} j^x\lambda_j(r_j^*-\langle\boldsymbol{\theta}_0,v_j\rangle)^2 = O((A^{(6,1)}+B^{(6,1)})/n) = O(k_n^{1+x}/n)$. □

Proof of Lemma 6 (ii). Denote

$$\Delta_2(M_j(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1)) = M_j(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1) - M_j(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_1),$$

and

$$\Delta_2(m^0_{M_j,i,\gamma}(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1)) = m^0_{M_j,i,\gamma}(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1) - m^0_{M_j,i,\gamma}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_1).$$

From Lemma 3 (i) we have $\lambda_j/\lambda_{j+1} \leq k_n^{a-1}$ when j is sufficiently large. Then $\sup_{j\leq k_n} \lambda_j/\lambda_{j+1} \leq k_n^{a-1}$ when k_n is sufficiently large. Together with Lemma 3 (ii) we have $\mathrm{E}\sup_{j\leq k_n} \zeta_j\lambda_j|\Delta_2(r_j)| \leq$

$$\begin{aligned}
& \mathrm{E}\sup_{j\leq k_n} \frac{1}{n}\sum_{i=1}^n \zeta_j \frac{\lambda_j}{\hat{\lambda}_j}[|\delta_i\Delta_2(M_j(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1))| + |(1-\delta_i)\Delta_2(m^0_{M_j,i,\gamma}(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1))|] \\
\leq\ & \frac{1}{n}\sum_{i=1}^n \mathrm{E}\sup_{j\leq k_n} \zeta_j \frac{\lambda_j}{\lambda_{j+1}}[|\delta_i\Delta_2(M_j(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1))| + |(1-\delta_i)\Delta_2(m^0_{M_j,i,\gamma}(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1))|] \\
=\ & \mathrm{E}\sup_{j\leq k_n} \zeta_j \frac{\lambda_j}{\lambda_{j+1}}[|\delta_i\Delta_2(M_j(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1))| + |(1-\delta_i)\Delta_2(m^0_{M_j,i,\gamma}(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1))|] \\
\leq\ & \mathrm{E}\sup_{j\leq k_n} \zeta_j \frac{\lambda_j}{\lambda_{j+1}}|\Delta_2\big(M_j(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1))| + \mathrm{E}\sup_{j\leq k_n} \zeta_j \frac{\lambda_j}{\lambda_{j+1}}|\Delta_2\big(m^0_{M_j,i,\gamma}(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1))| \\
\leq\ & \sup_{j\leq k_n} \frac{\lambda_j}{\lambda_{j+1}}[\mathrm{E}\sup_{j\leq k_n} \zeta_j|\Delta_2(M_j(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1))| + \mathrm{E}\sup_{j\leq k_n} \zeta_j|\Delta_2(m^0_{M_j,i,\gamma}(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1))|] \\
\leq\ & constant * k_n^{a-1}[\mathrm{E}\sup_{j\leq k_n} \zeta_j|\Delta_2(M_j(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1))| + \mathrm{E}\sup_{j\leq k_n} \zeta_j|\Delta_2(m^0_{M_j,i,\gamma}(Y_i, Z_i, W_i, \hat{v}_j; \boldsymbol{\beta}_1))|] \\
\triangleq\ & constant * k_n^{a-1}(A^{(6,2)} + B^{(6,2)}).
\end{aligned}$$

Next the two terms $A^{(6,2)}$ and $B^{(6,2)}$ are calculated separately. It follows from the Cauchy's inequality that

$$\begin{aligned}
A^{(6,2)} &= \mathrm{E}\sup_{j\leq k_n} \zeta_j|\langle Z_i, \hat{v}_j - v_j\rangle(\langle Z_i, \boldsymbol{\theta}\rangle + \epsilon_i)| \\
&= \mathrm{E}\sup_{j\leq k_n} |\langle(\langle Z_i, \boldsymbol{\theta}\rangle + \epsilon_i)Z_i, \zeta_j(\hat{v}_j - v_j)\rangle| \\
&\leq \mathrm{E}\sqrt{\sup_{j\leq k_n} |\langle(\langle Z_i, \boldsymbol{\theta}\rangle + \epsilon_i)Z_i, (\langle Z_i, \boldsymbol{\theta}\rangle + \epsilon_i)Z_i\rangle\langle\zeta_j(\hat{v}_j - v_j), \zeta_j(\hat{v}_j - v_j)\rangle|}
\end{aligned}$$

$$= \sqrt{\mathrm{E}\langle Z, Z\rangle(\langle Z, \boldsymbol{\theta}\rangle + \epsilon)^2}\{\mathrm{E}\sup_j \zeta_j^2\|\hat{v}_j - v_j\|^2\}^{1/2}\big(1 + o(1)\big).$$

From Lemma 1 and Lemma 2, $\{\mathrm{E}\sup_{j\le k_n}\zeta_j^2\|\hat{v}_j - v_j\|^2\}^{1/2} \le constant * n^{-1/2}$. Then $A^{(6,2)} = O_p(1/\sqrt{n})$. Using Lemma 4, by Jensen's inequality, we have

$$\begin{aligned}
B^{(6,2)} &= \mathrm{E}\left\{\sup_{j\le k_n}\zeta_j\big|\mathrm{E}_{-\hat{v}_j}[\langle Z_i, \hat{v}_j - v_j\rangle(\langle Z_i, \boldsymbol{\theta}\rangle + \epsilon_i)|Z_i, W_i, \delta_i = 0]\big|\right\} \\
&\le \mathrm{E}\left\{\mathrm{E}_{-(\hat{v}_1,\cdots)}[\sup_{j\le k_n}\zeta_j|\langle Z_i, \hat{v}_j - v_j\rangle(\langle Z_i, \boldsymbol{\theta}\rangle + \epsilon_i)||Z_i, W_i, \delta_i = 0]\right\} \\
&= \mathrm{E}\left\{\mathrm{E}_{-(\hat{v}_1,\cdots)}[\sup_{j\le k_n}|\langle(\langle Z_i, \boldsymbol{\theta}\rangle + \epsilon_i)Z_i, \zeta_j(v_j - v_j)\rangle||Z_i, W_i, \delta_i = 0]\right\} \\
&\le \mathrm{E}\sqrt{\langle Z, Z\rangle(\langle Z, \boldsymbol{\theta}\rangle + \epsilon)^2}\{\mathrm{E}\mathrm{E}_{-(\hat{v}_1,\cdots)}\sup_j \zeta_j^2\|\hat{v}_j - v_j\|^2\}^{1/2}(1 + o(1)) \\
&= \mathrm{E}\sqrt{\langle Z, Z\rangle(\langle Z, \boldsymbol{\theta}\rangle + \epsilon)^2}\{\mathrm{E}\sup_j \zeta_j^2\|\hat{v}_j - v_j\|^2\}^{1/2}(1 + o(1)).
\end{aligned}$$

Then $B^{(6,2)} = O_p(1/\sqrt{n})$. □

Proof of Lemma 6 (iii). From definitions of r_j, r_j^*, and $\Delta_2(r_j)$ in Lemma 6, (ii), we have

$$|r_j - r_j^*| \le |\Delta_2(r_j)| + |\frac{1}{\lambda_j} - \frac{1}{\hat{\lambda}_j}|\hat{\lambda}_j|r_j^*|.$$

It follows that

$$\begin{aligned}
(\sum_{j=1}^{k_n}\lambda_j|r_j - r_j^*|^2) &\le [\sum_{j=1}^{k_n}\lambda_j 2(\Delta_2(r_j)^2 + (\frac{1}{\lambda_j} - \frac{1}{\hat{\lambda}_j})^2\hat{\lambda}_j^2 r_j^{*2})] \\
&= \sum_{j=1}^{k_n} 2\lambda_j\Delta_2(r_j)^2 + \sum_{j=1}^{k_n} 2(\frac{1}{\lambda_j} - \frac{1}{\hat{\lambda}_j})^2\hat{\lambda}_j^2 r_j^{*2}\lambda_j \triangleq A^{(6,3)} + B^{(6,3)}.
\end{aligned}$$

Next the two terms $A^{(6,3)}$ and $B^{(6,3)}$ are calculated. From Lemma 3, we have $\mathrm{E}A^{(6,3)} \le$

$$2\mathrm{E}(\sum_{j=1}^{k_n}\lambda_j|\frac{1}{n}\sum_{i=1}^{n}\langle Z_i, \hat{v}_j - v_j\rangle(\langle Z_i, \boldsymbol{\theta}_0\rangle + \epsilon_i)|^2/\hat{\lambda}_j)$$

$$
\begin{aligned}
&+ \quad 2\mathrm{E}(\sum_{j=1}^{k_n}\lambda_j\Big|\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{-\hat{v}_j}[\langle Z_i,\hat{v}_j-v_j\rangle(\langle Z_i,\boldsymbol{\theta}_0\rangle+\epsilon_i)|Z_i,W_i,\delta_i=0]\Big|^2/\hat{\lambda}_j)\\
&\leq \quad 2\sup_{j\leq k_n}\frac{\lambda_j}{\lambda_{j+1}}*(\sum_{j=1}^{k_n}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}|\langle Z_i,\hat{v}_j-v_j\rangle(\langle Z_i,\boldsymbol{\theta}_0\rangle+\epsilon_i)|^2)\\
&+ \quad 2\sup_{j\leq k_n}\frac{\lambda_j}{\lambda_{j+1}}*(\sum_{j=1}^{k_n}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\mathrm{E}^2_{-\hat{v}_j}[\langle Z_i,\hat{v}_j-v_j\rangle(\langle Z_i,\boldsymbol{\theta}_0\rangle+\epsilon_i)|Z_i,W_i,\delta_i=0])\\
&\leq \quad 2\sup_{j\leq k_n}\frac{\lambda_j}{\lambda_{j+1}}*(\sum_{j=1}^{k_n}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}|\langle Z_i,\hat{v}_j-v_j\rangle(\langle Z_i,\boldsymbol{\theta}_0\rangle+\epsilon_i)|^2)\\
&+ \quad 2\sup_{j\leq k_n}\frac{\lambda_j}{\lambda_{j+1}}*(\sum_{j=1}^{k_n}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\mathrm{E}_{-\hat{v}_j}[\langle Z_i,\hat{v}_j-v_j\rangle^2(\langle Z_i,\boldsymbol{\theta}_0\rangle+\epsilon_i)^2|Z_i,W_i,\delta_i=0])\\
&\leq \quad constant*k_n^{a-1}\sum_{j=1}^{k_n}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\langle Z_i,Z_i\rangle\|\hat{v}_j-v_j\|^2(\langle Z_i,\boldsymbol{\theta}_0\rangle+\epsilon_i)^2]\\
&+ \quad constant*k_n^{a-1}\sum_{j=1}^{k_n}\frac{1}{n}\sum_{i=1}^{n}\sqrt{\mathrm{E}\langle Z_i,Z_i\rangle^2(\langle Z_i,\boldsymbol{\theta}_0\rangle+\epsilon_i)^4}\sqrt{\mathrm{E}[\|\hat{v}_j-v_j\|^4]}\\
&\leq \quad constant*k_n^{a-1}\sum_{j=1}^{k_n}\zeta_j^{-2}n^{-1}=O(k_n^{a-1}\sum_{j=1}^{k_n}j^{3a+2}/n)=O(k_n^{4a+2}n^{-1}).
\end{aligned}
$$

The last inequality holds from Lemma 1 and Lemma 2. Next, from Lemma 3, we have

$B^{(6,3)} =$

$$
\begin{aligned}
& \sum_{j=1}^{k_n}2(\frac{1}{\lambda_j}-\frac{1}{\hat{\lambda}_j})^2\hat{\lambda}_j^2 r_j^{*2}\lambda_j\\
= & \sum_{j=1}^{k_n}2\frac{(\lambda_j-\hat{\lambda}_j)^2}{\lambda_j^2}(r_j^*-\langle\boldsymbol{\theta}_0,v_j\rangle+\langle\boldsymbol{\theta}_0,v_j\rangle)^2\lambda_j\\
\leq & \sum_{j=1}^{k_n}4\frac{(\lambda_j-\lambda_{j+1})^2}{\lambda_j^2}(r_j^*-\langle\boldsymbol{\theta}_0,v_j\rangle)^2\lambda_j+\sum_{j=1}^{k_n}4\frac{(\lambda_j-\lambda_{j+1})^2}{\lambda_j^2}\langle\boldsymbol{\theta}_0,v_j\rangle^2\lambda_j\\
\leq & \sum_{j=1}^{k_n}4(r_j^*-\langle\boldsymbol{\theta}_0,v_j\rangle)^2\lambda_j+\sum_{j=1}^{k_n}4\langle\boldsymbol{\theta}_0,v_j\rangle^2\lambda_j
\end{aligned}
$$

$$\begin{aligned}
&\leq \quad constant \times [\sum_{j=1}^{k_n} \lambda_j (r_j^* - \langle \boldsymbol{\theta}_0, v_j \rangle)^2 + \sum_{j=1}^{k_n} j^{-1} j^{-2b}] \\
&= \quad O_p(k_n/n) + O_p(k_n^{-2b}) = O_p(k_n/n + k_n^{-2b}).
\end{aligned}$$

Finally we have $\mathrm{E}(\sum_{j=1}^{k_n} \lambda_j |r_j - r_j^*|^2) = A^{(6,3)} + B^{(6,3)} = O(k_n^{4a+2}/n + k_n^{-2b})$.

Proof of Lemma 6 (iv). The left side of the equation is equal to or less than

$$\begin{aligned}
& \sum_{j=1}^{k_n} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E} \frac{(\delta_i W_i \langle Z_i, v_j \rangle + (1-\delta_i)\mathrm{E}(W_i \langle Z_i, v_j \rangle | Z_i, W_i, \delta_i = 0))^2}{\lambda_j} \\
\leq \quad & \frac{1}{n} \sum_{j=1}^{k_n} \sum_{i=1}^{n} \mathrm{E}[\delta_i W_i \langle Z_i, v_j \rangle]^2 / \lambda_j + \frac{1}{n} \sum_{j=1}^{k_n} \sum_{i=1}^{n} \mathrm{E}[(1-\delta_i)\mathrm{E}(W_i \langle Z_i, v_j \rangle | Z_i, W_i, \delta_i = 0)]^2 / \lambda_j \\
\triangleq \quad & A^{(6,4)} + B^{(6,4)},
\end{aligned}$$

We only need to calculate the order of $A^{(6,4)}$ and $B^{(6,4)}$ separately. We have $A^{(6,4)} =$

$$\begin{aligned}
& \sum_{j=1}^{k_n} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[\delta_i W_i < Z_i, v_j >]^2 / \lambda_j \\
= \quad & \sum_{j=1}^{k_n} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[\delta_i W_i \xi_j^{(i)}]^2 \leq \sum_{j=1}^{k_n} \mathrm{E}[W_i \xi_j^{(i)}]^2 \leq \mathrm{E}[W_i^2],
\end{aligned}$$

and $B^{(6,4)} =$

$$\begin{aligned}
& \mathrm{E} \sum_{j=1}^{k_n} [\frac{1}{n} \sum_{i=1}^{n} (1-\delta_i)\mathrm{E}(W_i \langle Z_i, v_j \rangle | Z_i, W_i, \delta_i = 0)]^2 / \lambda_j \\
\leq \quad & \sum_{j=1}^{k_n} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}(W_i \langle Z_i, v_j \rangle | Z_i, W_i, \delta_i = 0)]^2 / \lambda_j \\
\leq \quad & \sum_{j=1}^{k_n} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}(W_i^2 \langle Z_i, v_j \rangle^2 | Z_i, W_i, \delta_i = 0)] / \lambda_j \\
= \quad & \sum_{j=1}^{k_n} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[W_i \xi_j^{(i)}]^2 = \sum_{j=1}^{k_n} \mathrm{E}[W_i \xi_j^{(i)}]^2 \leq \mathrm{E}[W_i^2].
\end{aligned}$$

□

Proof of Lemma 6 (v). Using Lemma 1 and Lemma 2, we have

$$\begin{aligned}
& \mathrm{E}\sup_j \zeta_j \Big|\frac{1}{n}\sum_{i=1}^n \delta_i W_i \langle Z_i, \hat{v}_j - v_j\rangle\Big| \\
\leq & \sqrt{\mathrm{E}\delta_i^2 W_i^2 \langle Z_i, Z_i\rangle}\sqrt{\mathrm{E}\sup_j \zeta_j^2 \|\hat{v}_j - v_j\|^2} = O_p(1/\sqrt{n}),
\end{aligned}$$

and

$$\begin{aligned}
& \mathrm{E}\sup_j \zeta_j \Big|\frac{1}{n}\sum_{i=1}^n \mathrm{E}_{-\hat{v}_j}[W_i\langle Z_i, \hat{v}_j - v_j\rangle | Z_i, W_i, \delta_i = 0]\Big| \\
\leq & \mathrm{E}\sup_j \frac{1}{n}\sum_{i=1}^n \mathrm{E}_{-\hat{v}_j}[\zeta_j |W_i\langle Z_i, \hat{v}_j - v_j\rangle|\big| Z_i, W_i, \delta_i = 0] \\
\leq & \sqrt{\mathrm{E}[\mathrm{E}_{-\hat{v}_j}(W_i^2\langle Z_i, Z_i\rangle | W_i, Z_i, \delta_i = 0)]}\sqrt{\mathrm{E}\mathrm{E}_{-\hat{v}_j}\sup_j \zeta_j \|\hat{v}_j - v_j\|^2} \\
= & O_p(1/\sqrt{n}).
\end{aligned}$$

This uses the similar technique to $A^{(6,2)}$ in the proof Lemma 6 (ii).

Proof of Lemma 6 (vi). Denote

$$\Delta_1[\sum_{j=1}^{k_n} r_j(\delta_i W_i \langle Z_i, \hat{v}_j\rangle)] = \sum_{j=1}^{k_n} r_j(\delta_i W_i\langle Z_i, \hat{v}_j\rangle) - \sum_{j=1}^{k_n} r_j^*(\delta_i W_i \langle Z_i, v_j\rangle).$$

Then we have the following decomposition.

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^n \Delta_1[\sum_{j=1}^{k_n} r_j(\delta_i W_i\langle Z_i, \hat{v}_j\rangle)] \\
= & \sum_{j=1}^{k_n}(r_j - r_j^*)\frac{1}{n}\sum_{i=1}^n \delta_i W_i \langle Z_i, \hat{v}_j - v_j\rangle + \sum_{j=1}^{k_n}(r_j - r_j^*)\frac{1}{n}\sum_{i=1}^n \delta_i W_i\langle Z_i, v_j\rangle \\
+ & \sum_{j=1}^{k_n}(r_j^* - \langle \boldsymbol{\theta}_0, v_j\rangle)\frac{1}{n}\sum_{i=1}^n \delta_i W_i\langle Z_i, \hat{v}_j - v_j\rangle + \sum_{j=1}^{k_n}\langle \boldsymbol{\theta}_0, v_j\rangle \frac{1}{n}\sum_{i=1}^n \delta_i W_i \langle Z_i, \hat{v}_j - v_j\rangle \\
\triangleq & A^{(6,6)} + B^{(6,6)} + C^{(6,6)} + D^{(6,6)}.
\end{aligned}$$

Next we bound the $A^{(6,6)}, B^{(6,6)}, C^{(6,6)}$, and $D^{(6,6)}$ separately.

From the Cauchy's inequality, by conclusions 3 and 4 of this lemma, we have $|B^{(6,6)}| \leq$

$$\begin{aligned} & \sqrt{\sum_{j=1}^{k_n} \lambda_j (r_j - r_j^*)^2} \sqrt{\sum_{j=1}^{k_n} [\frac{1}{n} \sum_{i=1}^{n} \delta_i W_i \langle Z_i, v_j \rangle]^2 / \lambda_j} \\ = & \sqrt{O_p(k_n^{4a+2} n^{-1} + k_n^{-2b}) O_p(1)} = O_p(k_n^{2a+1} n^{-1/2} + k_n^{-b}), \end{aligned}$$

and $|A^{(6,6)}| \leq$

$$\begin{aligned} & \sum_{j=1}^{k_n} |\Delta_2(r_j) \frac{1}{n} \sum_{i=1}^{n} \delta_i W_i \langle Z_i, \hat{v}_j - v_j \rangle| \\ + & \sum_{j=1}^{k_n} |\frac{1}{\lambda_j} - \frac{1}{\hat{\lambda}_j}| \hat{\lambda}_j | r_j^* - \langle \boldsymbol{\theta}_0, v_j \rangle || \frac{1}{n} \sum_{i=1}^{n} \delta_i W_i \langle Z_i, \hat{v}_j - v_j \rangle | \\ + & \sum_{j=1}^{k_n} |\frac{1}{\lambda_j} - \frac{1}{\hat{\lambda}_j}| \hat{\lambda}_j | \langle \boldsymbol{\theta}_0, v_j \rangle || \frac{1}{n} \sum_{i=1}^{n} \delta_i W_i \langle Z_i, \hat{v}_j - v_j \rangle | \\ \triangleq & A_1^{(6,6)} + A_2^{(6,6)} + A_3^{(6,6)}, \end{aligned}$$

where $\Delta_2(r_j)$ was defined in Lemma 6, (ii). To bound $A^{(6,6)}$, we only need to bound $A_1^{(6,6)}, A_2^{(6,6)}$ and $A_3^{(6,6)}$ separately. From Lemma 6, (ii) and Lemma 6, (v), we have

$$A_1^{(6,6)} \leq \sup_{j \leq k_n} \zeta_j \lambda_j |\Delta_2(r_j)| \sup_{j \leq k_n} \zeta_j | \frac{1}{n} \sum_{i=1}^{n} \delta_i W_i \langle Z_i, \hat{v}_j - v_j \rangle | \sum_{j=1}^{k_n} 1/(\lambda_j \zeta_j^2) = O_p(k_n^{4a+2} n^{-1}).$$

By the Cauchy's inequality, we have

$$\begin{aligned} & A_2^{(6,6)} + |C^{(6,6)}| \leq \{ \sum_{j=1}^{k_n} \frac{(|\frac{1}{\lambda_j} - \frac{1}{\hat{\lambda}_j}| \hat{\lambda}_j / \sqrt{\lambda_j} + 1/\sqrt{\lambda_j})^2}{\zeta_j^2} \lambda_j (r_j^* - \langle \boldsymbol{\theta}_0, v_j \rangle)^2 \\ \times & \sum_{j=1}^{k_n} [\frac{\sum_{i=1}^{n} \zeta_j \delta_i W_i \langle Z_i, \hat{v}_j - v_j \rangle}{n}]^2 \}^{0.5}. \end{aligned}$$

Using the fact that

$$(|\frac{1}{\lambda_j} - \frac{1}{\hat{\lambda}_j}|\hat{\lambda}_j/\sqrt{\lambda_j} + 1/\sqrt{\lambda_j})^2/\zeta_j^2 \leq constant \times (j^{\frac{a}{2}})^2 j^{2a+2} = constant \times j^{3a+2}$$

implied by Lemma 3, together with Lemma 6, (i), and Lemma 6, (v), we have

$$A_2^{(6,6)} + |C^{(6,6)}| = \sqrt{O_p(k_n^{3a+3}/n)O_p(1/n)} = O_p(k_n^{(3a+3)/2}/n).$$

Similarly, we have

$$A_3^{(6,6)} + |D^{(6,6)}| = \sum_{j=1}^{k_n} j^{-b}/\zeta_j O_p(1/\sqrt{n}) = O_p(\max(k_n^{a+2-b}, \log n)n^{-1/2}).$$

Therefore, $\frac{1}{n}\sum_{i=1}^n \Delta_1[\sum_{j=1}^{k_n} r_j(\delta_i W_i\langle Z_i, \hat{v}_j\rangle)] =$

$$A_1^{(6,6)} + A_2^{(6,6)} + A_3^{(6,6)} + |B^{(6,6)}| + |C^{(6,6)}| + |D^{(6,6)}| = O_p(k_n^{2a+1}n^{-1/2} + k_n^{-b}).$$

Similarly,

$$\frac{1}{n}\sum_{i=1}^n \Delta_1[\sum_{j=1}^{k_n} r_j(1-\delta_i)\mathrm{E}_{-\hat{v}_j}(W_i\langle Z_i, \hat{v}_j\rangle|Z_i, W_i, \delta_i = 0)] = O_p(k_n^{2a+1}n^{-1/2} + k_n^{-b}).$$

Finally $L_6 =$

$$\frac{1}{n}\sum_{i=1}^n \Delta_1[\sum_{j=1}^{k_n} r_j(\delta_i W_i\langle Z_i, \hat{v}_j\rangle)] + \frac{1}{n}\sum_{i=1}^n \Delta_1[\sum_{j=1}^{k_n} r_j(1-\delta_i)\mathrm{E}_{-\hat{v}_j}(W_i\langle Z_i, \hat{v}_j\rangle|Z_i, W_i, \delta_i = 0)]$$

which is $O_p(k_n^{2a+1}n^{-1/2} + k_n^{-b})$.

□

Lemma 7. *Under Assumptions (A.1)–(A.4), (A.6), (A.7) and (A.9), we have* $L_7 \triangleq$

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n\{\sum_{j=1}^{k_n} r_j^*[\delta_i W_i\langle Z_i, v_j\rangle + (1-\delta_i)\mathrm{E}\{\langle Z_i, v_j\rangle W|Z_i, W_i, \delta_i = 0\}]\} \\ - \; &\mathrm{E}\{\sum_{j=1}^{k_n} r_j^*[\delta W\langle Z, v_j\rangle + (1-\delta)\mathrm{E}\{\langle Z, v_j\rangle W|X, V, \delta_i = 0\}]\} = O_p(\sqrt{k_n/n}). \end{aligned}$$

Proof. We use the following decomposition.

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\{\sum_{j=1}^{k_n} r_j^*[\delta_i W_i\langle Z_i, v_j\rangle + (1-\delta_i)\mathrm{E}\{\langle Z_i, v_j\rangle W | Z_i, W_i, \delta_i = 0\}]\} \\
- \quad & \mathrm{E}\{\sum_{j=1}^{k_n} r_j^*[\delta W\langle Z, v_j\rangle + (1-\delta)\mathrm{E}\{\langle Z, v_j\rangle W | X, V, \delta_i = 0\}]\} \\
\leq \quad & \frac{1}{n}\sum_{i=1}^{n}\{\sum_{j=1}^{k_n}(r_j^* - \langle v_j, \boldsymbol{\theta}_0\rangle)\big[\delta_i W_i\langle Z_i, v_j\rangle + (1-\delta_i)\mathrm{E}\{\langle Z_i, v_j\rangle W | Z_i, W_i, \delta_i = 0\}\big]\} \\
+ \quad & \sum_{j=1}^{k_n}\langle \boldsymbol{\theta}_0, v_j\rangle[\frac{1}{n}\sum_{i=1}^{n}\delta_i W_i\langle Z_i, v_j\rangle + (1-\delta_i)\mathrm{E}(\langle Z_i, v_j\rangle W | Z_i, W_i, \delta_i = 0) \\
- \quad & \mathrm{E}\{W\langle Z, v_j\rangle\}] \triangleq A^{(7)} + B^{(7)}.
\end{aligned}$$

Denote

$$B_1^{(7)} \triangleq \mathrm{E}\sum_{j=1}^{k_n}\left(\frac{1}{n}\sum_{i=1}^{n}[\delta_i W_i\langle Z_i, v_j\rangle + (1-\delta_i)\mathrm{E}(\langle Z_i, v_j\rangle W | Z_i, W_i, \delta_i = 0) - \mathrm{E}\{W\langle Z, v_j\rangle\}]\right)^2.$$

Using the Jensen's inequality and the Cauchy's inequality, after algebraic calculations, we have $B_1^{(7)} =$

$$\begin{aligned}
& \frac{1}{n}\sum_{j=1}^{k_n}\mathrm{E}[\delta_i W_i\langle Z_i, v_j\rangle + (1-\delta_i)\mathrm{E}(\langle Z_i, v_j\rangle W | Z_i, W_i, \delta_i = 0) - \mathrm{E}\{W\langle Z, v_j\rangle\}]^2 \\
\leq \quad & \frac{3}{n}\sum_{j=1}^{k_n}[\mathrm{E}(W\langle Z, v_j\rangle)^2 + \mathrm{E}(\mathrm{E}(W\langle Z, v_j\rangle | Z_i, W_i, \delta_i = 0))^2 + \mathrm{E}^2(W\langle Z, v_j\rangle)] \\
\leq \quad & \frac{3}{n}\sum_{j=1}^{k_n}\lambda_j(\mathrm{E}W^2\xi_j^2 + \mathrm{E}W^2\xi_j^2 + \mathrm{E}^2(W\xi_j)) \leq \frac{9}{n}\sum_{j=1}^{\infty}\lambda_j\sqrt{\mathrm{E}W^4\mathrm{E}\xi_1^4} = O(1/n).
\end{aligned}$$

It follows that

$$B^{(7)} \leq \sqrt{(\sum_{j=1}^{k_n}\langle \boldsymbol{\theta}_0, v_j\rangle^2)B_1^{(7)}} \leq \|\boldsymbol{\theta}_0\|\sqrt{B_1} = O_p(1/\sqrt{n}).$$

Before we continue with $A^{(7)}$, first we denote

$$\begin{aligned} A_1^{(7)} &\triangleq \sum_{j=1}^{k_n} \frac{\Big[\frac{1}{n}\sum_{i=1}^n \big(\delta_i W_i\langle Z_i, v_j\rangle + (1-\delta_i)\mathrm{E}\{\langle Z_i, v_j\rangle W_i | Z_i, W_i, \delta_i = 0\}\big)\Big]^2}{\lambda_j}. \\ \mathrm{E}A_1^{(7)} &\leq \sum_{j=1}^{k_n} \frac{\sum_{i=1}^n \Big[\mathrm{E}\big(\delta_i W_i\langle Z_i, v_j\rangle + (1-\delta_i)\mathrm{E}\{\langle Z_i, v_j\rangle W_i | Z_i, W_i, \delta_i = 0\}\big)\Big]^2}{n\lambda_j} \\ &= \sum_{j=1}^{k_n} \frac{\mathrm{E}\big(W_i^2\langle Z_i, v_j\rangle^2 + \mathrm{E}(\mathrm{E}^2\{\langle Z_i, v_j\rangle W_i | Z_i, W_i, \delta_i = 0\})\big)}{\lambda_j} \\ &\leq \sum_{j=1}^{k_n} \frac{\mathrm{E}\big(W_i^2\langle Z_i, v_j\rangle^2 + \mathrm{E}(\mathrm{E}\{\langle Z_i, v_j\rangle^2 W_i^2 | Z_i, W_i, \delta_i = 0\})\big)}{\lambda_j} \\ &= \sum_{j=1}^{k_n} \frac{2\mathrm{E}(W_i^2\xi_j^2\lambda_j)}{\lambda_j} = \sum_{j=1}^{k_n} 2\mathrm{E}(\xi_j^2 W^2) \leq 2\mathrm{E}W^2. \end{aligned}$$

Then the following equation

$$A^{(7)} \leq \sqrt{\sum_{j=1}^{k_n} \lambda_j (r_j^* - \langle v_j, \boldsymbol{\theta}_0\rangle)^2 A_1^{(7)}} = \sqrt{O_p(k_n/n)O_p(1)} = O_p(\sqrt{\frac{k_n}{n}}),$$

holds by the conclusion 1 of Lemma 6.

Finally we have $L_7 \leq A^{(7)} + B^{(7)} = O_p(\sqrt{k_n/n})$.

□

Lemma 8. *Under Assumptions (A.1)–(A.7) and (A.9), we have*

$$\|\sum_{j=1}^{k_n} r_j \hat{v}_j - \sum_{j=1}^{k_n} \langle \boldsymbol{\theta}_0, v_j\rangle v_j\| = O_p(k_n^{5a/2+3/2}/\sqrt{n} + k_n^{a/2+1/2-b}).$$

Proof. First we make the decomposition of the formula.

$$\|\sum_{j=1}^{k_n} r_j \hat{v}_j - \sum_{j=1}^{k_n} \langle \boldsymbol{\theta}_0, v_j\rangle v_j\|$$

$$\leq \ \|\sum_{j=1}^{k_n}(r_j - \langle \boldsymbol{\theta}_0, v_j\rangle)(\hat{v}_j - v_j)\| + \|\sum_{j=1}^{k_n}(r_j - \langle \boldsymbol{\theta}_0, v_j\rangle)v_j\|$$

$$+ \ \|\sum_{j=1}^{k_n}\langle \boldsymbol{\theta}_0, v_j\rangle(\hat{v}_j - v_j)\| \triangleq A^{(8)} + B^{(8)} + C^{(8)}.$$

We will discuss $A^{(8)}, B^{(8)}$, and $C^{(8)}$ separately. To calculate $A^{(8)}$ we define $A_1^{(8)} \triangleq \|\sum_{j=1}^{k_n}(r_j^* - \langle \boldsymbol{\theta}_0, v_j\rangle)(\hat{v}_j - v_j)\|$, and $A_2 \triangleq \|\sum_{j=1}^{k_n}(r_j - r_j^*)(\hat{v}_j - v_j)\|$. Using the conclusion of Lemma1, Lemma 2, and Lemma 6, (i), we have $A_1^{(8)} \leq$

$$\sqrt{\sum_{j=1}^{k_n}\lambda_j(r_j^* - \langle \boldsymbol{\theta}_0, v_j\rangle)^2/(\zeta_j^2\lambda_j)}\sqrt{\sum_{j=1}^{k_n}\zeta_j^2\|\hat{v}_j - v_j\|^2} = \sqrt{O_p(k_n^{3a+2+1}/n)O_p(k_n/n)}$$

which equals $O_p(k_n^{3a/2+2}/n)$; using the conclusion of Lemma 1, Lemma 2, and Lemma 6, (iii), we have $A_2^{(8)} \leq$

$$\sqrt{\sum_{j=1}^{k_n}\lambda_j(r_j - r_j^*)^2}\sqrt{\sum_{j=1}^{k_n}\|\hat{v}_j - v_j\|^2/\lambda_j} = \sqrt{O_p(k_n^{4a+2}/n + k_n^{-2b})O_p(k_n^{3a+3}/n)}.$$

which equals $O_p(k_n^{7a/2+5/2}/n + k_n^{3a/2+3/2-b}/\sqrt{n})$. Put them together and we get

$$A^{(8)} \leq A_1^{(8)} + A_2^{(8)} = O_p(k_n^{5a/2+3/2}/n + k_n^{a/2+1/2-b}/\sqrt{n}).$$

To calculate $B^{(8)}$ we define $B_1^{(8)} \triangleq \|\sum_{j=1}^{k_n}(r_j^* - \langle \boldsymbol{\theta}_0, v_j\rangle)v_j\|$, and $B_2^{(8)} \triangleq \|\sum_{j=1}^{k_n}(r_j - r_j^*)v_j\|$. Using the conclusion of Lemma 6, (i), we have $B_1^{(8)} \leq$

$$\sqrt{\sum_{j=1}^{k_n}\lambda_j(r_j^* - \langle \boldsymbol{\theta}_0, v_j\rangle)^2/\lambda_j} = \sqrt{O_p(k_n^{a+1}/n)} = O_p(k_n^{(a+1)/2}/\sqrt{n});$$

using Lemma 1, 2, and 6, (iii), we have $B_2^{(8)} \leq$

$$\sqrt{\sum_{j=1}^{k_n}\lambda_j(r_j - r_j^*)^2}\sqrt{\sum_{j=1}^{k_n}\|v_j\|^2/\lambda_j} = \sqrt{O_p(k_n^{4a+2}/n + k_n^{-2b})O_p(k_n^{a+1})},$$

which equals $O_p(k_n^{5a/2+3/2}/\sqrt{n}+k_n^{a/2+1/2-b})$. Put them together, and we have

$$B^{(8)} \leq B_1^{(8)} + B_2^{(8)} = O_p(k_n^{5a/2+3/2}/\sqrt{n}+k_n^{a/2+1/2-b}).$$

Using Lemma 1 and Lemma 2, the following inequality holds for $C^{(8)}$.

$$\|C^{(8)}\| \leq \sum_{j=1}^{k_n} \zeta_j O_p(n^{-1/2})|\langle \boldsymbol{\theta}_0, v_j\rangle| \leq \sum_{j=1}^{k_n} j^{a+1-b} O(n^{-1/2}) \leq O(k_n^{(a+2-b)_+}/\sqrt{n}).$$

Finally, we have

$$\|\sum_{j=1}^{k_n} r_j \hat{v}_j - \sum_{j=1}^{k_n} \langle \boldsymbol{\theta}_0, v_j\rangle v_j\| \leq A^{(8)} + B^{(8)} + C^{(8)} = O_p(k_n^{5a/2+3/2}/\sqrt{n}+k_n^{a/2+1/2-b}).$$

□

Lemma 9. *Under Assumptions (A.1)–(A.9), we have*

$$\|\partial U(\boldsymbol{\beta}_0)/\partial(\boldsymbol{\beta}^T) - \mathfrak{J}\| = o_p(1).$$

Proof of Lemma 9. After straightforward algebraic calculations, $\partial U(\boldsymbol{\beta}_0)/\partial(\boldsymbol{\beta}^T) =$

$$\frac{1}{n}\sum_{i=1}^{n}[\delta_i W_i W_i^T + (1-\delta_i)\mathrm{E}(W_i W_i^T|Z_i, W_i, \delta_i = 0)] - \sum_{j=1}^{k_n} R(\hat{v}_j, W)R(\hat{v}_j, W^T)/\hat{\lambda}_j,$$

where

$$R(\hat{v}_j, W) = \frac{1}{n}\sum_{i=1}^{n}(\delta_i\langle Z_i, \hat{v}_j\rangle W_i + (1-\delta_i)\mathrm{E}(\langle Z_i, \hat{v}_j\rangle W_i|Z_i, W_i, \delta_i = 0)).$$

It follows that $\partial U(\boldsymbol{\beta}_0)/\partial(\boldsymbol{\beta}^T) - \mathfrak{J} =$

$$
\begin{aligned}
& \{\frac{1}{n}\sum_{i=1}^{n}[\delta_i W_i W_i^T + (1-\delta_i)\mathrm{E}(W_i W_i^T | Z_i, W_i, \delta_i = 0)] - \mathrm{E}WW^T\} \\
- & \{\sum_{j=1}^{k_n}\tilde{r}_j R(\hat{v}_j, W_i^T) - \sum_{j=1}^{k_n}\tilde{r}_j^* R(v_j, W_i^T)\} \\
- & \{\sum_{j=1}^{k_n}\tilde{r}_j^* R(v_j, W_i^T) - \sum_{j=1}^{k_n}[\mathrm{E}\langle Z, v_j\rangle W][\mathrm{E}\langle Z, v_j\rangle W^T]/\lambda_j\} \\
- & \sum_{j=k_n+1}^{\infty}[\mathrm{E}\langle Z, v_j\rangle W][\mathrm{E}\langle Z, v_j\rangle W^T]/\lambda_j \triangleq A^{(9)} + B^{(9)} + C^{(9)} + D^{(9)},
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{r}_j &= \frac{\sum_{i=1}^{n}[\delta_i \tilde{M}_j(Z_i, W_i, \hat{v}_j) + (1-\delta_i) m^0_{\tilde{M}_j,i,\gamma}(Z_i, W_i, \hat{v}_j)]}{n\hat{\lambda}_j}, \\
\tilde{r}_j^* &= \frac{\sum_{i=1}^{n}[\delta_i \tilde{M}_j(Z_i, W_i, v_j) + (1-\delta_i) m^0_{\tilde{M}_j,i,\gamma}(Z_i, W_i, v_j)]}{n\lambda_j},
\end{aligned}
$$

and $\tilde{M}_j(Z_i, W_i, v_j) = \langle Z_i, v_j\rangle W_i$. Note that $\|A^{(9)}\| = o_p(1)$ holds by law of large numbers;

$$
\begin{aligned}
B^{(9)} &= \sum_{j=1}^{k_n}(\tilde{r}_j - \tilde{r}_j^*)[\frac{1}{n}\sum_{i=1}^{n}\langle Z_i, (\hat{v}_j - v_j)\rangle W_i^T] \\
&+ \sum_{j=1}^{k_n}(\tilde{r}_j - \tilde{r}_j^*)[\frac{1}{n}\sum_{i=1}^{n}\langle Z_i, v_j\rangle W_i^T] + \sum_{j=1}^{k_n}\tilde{r}_j^*[\frac{1}{n}\sum_{i=1}^{n}\langle Z_i, \hat{v}_j - v_j\rangle W_i^T],
\end{aligned}
$$

which equals $O_p(k_n^{2a+1}n^{-1/2} + k_n^{a-b-1})$ using the technique similar to the proof of Lemma 6, (vi); similar to the proof of Lemma 7, $\|C\| = O_p(\sqrt{k_n/n})$; $\|D^{(9)}\| = o(1)$ since

$$
\sum_{j=1}^{\infty}[\mathrm{E}\langle Z, v_j\rangle W][\mathrm{E}\langle Z, v_j\rangle W^T]/\lambda_j = \sum_{j=1}^{\infty}\mathrm{E}\xi_j W \mathrm{E}\xi_j W^T \le \mathrm{E}WW^T
$$

holds by Assumption (A.8). In conclusion,

$$
\|\partial U(\boldsymbol{\beta}_0)/\partial(\boldsymbol{\beta}^T) - \mathfrak{J}\| \le \|A^{(9)}\| + \|B^{(9)}\| + \|C^{(9)}\| + \|D^{(9)}\| = o_p(1).
$$

$\square$

Proof of Theorem 1. First we denote $e \triangleq \partial U(\boldsymbol{\beta}_1)/\partial \boldsymbol{\beta}_1^T - \mathfrak{J}$. From Lemma 4–Lemma 7, we have

$$U(\boldsymbol{\beta}_{1,0}) = L_5 - L_4 + L_6 + L_7 = O_p(k_n^{2a+1} n^{-1/2} + k_n^{1/2-b}),$$

where the 'L_i' is defined in Lemma $i, i = 4, 5, 6, 7$. Then from $\|e\| \to 0$ of Lemma 9, together with Equation (1.5) of Stewart (1969), we have $\|\hat{\boldsymbol{\beta}}_{1,0} - \boldsymbol{\beta}_{1,0}\| =$

$$\begin{aligned}
& \|(\frac{\partial U(\boldsymbol{\beta}_1)}{\partial \boldsymbol{\beta}_1^T})^{-1} U(\boldsymbol{\beta}_{1,0})\| \le \|(\frac{\partial U(\boldsymbol{\beta}_1)}{\partial \boldsymbol{\beta}_1^T})^{-1}\| \|U(\boldsymbol{\beta}_{1,0})\| = \|(\mathfrak{J} + e)^{-1}\| \|U(\boldsymbol{\beta}_{1,0})\| \\
\le\ & [\|(\mathfrak{J} + e)^{-1} - \mathfrak{J}^{-1}\| + \|\mathfrak{J}^{-1}\|] \|U(\boldsymbol{\beta}_{1,0})\| \le [\|(\mathfrak{J} + e)^{-1} - \mathfrak{J}^{-1}\| / \|\mathfrak{J}^{-1}\| + 1] \|U(\boldsymbol{\beta}_{1,0})\| \|\mathfrak{J}^{-1}\| \\
\le\ & [\frac{\|\mathfrak{J}^{-1}\| \|e\|}{1 - \|\mathfrak{J}^{-1}\| \|e\|} + 1] \|U(\boldsymbol{\beta}_{1,0})\| \|\mathfrak{J}^{-1}\| \le \frac{\|U(\boldsymbol{\beta}_{1,0})\| \|\mathfrak{J}^{-1}\|}{1 - \|\mathfrak{J}^{-1}\| \|e\|} = O_p(\|U(\boldsymbol{\beta}_{1,0})\|)
\end{aligned}$$

which also equals $O_p(k_n^{2a+1} n^{-1/2} + k_n^{1/2-b})$. Then the first part of the theorem is proved.

For the second part of the theorem, we make the following decompositions. $\boldsymbol{\theta}(\hat{\boldsymbol{\beta}}_1) - \boldsymbol{\theta}_0 =$

$$\begin{aligned}
& \{\sum_{j=1}^{k_n} (r_j(\hat{\boldsymbol{\beta}}_1) - r_j(\boldsymbol{\beta}_{1,0})) \hat{v}_j\} + \{\sum_{j=1}^{k_n} r_j(\boldsymbol{\beta}_{1,0}) \hat{v}_j - \sum_{j=1}^{k_n} \langle \boldsymbol{\theta}_0, v_j \rangle v_j\} + \{-\sum_{j=k_n+1}^{\infty} \langle \boldsymbol{\theta}_0, v_j \rangle v_j\} \\
\triangleq\ & A^{\mathbf{T1}} + B^{\mathbf{T1}} + C^{\mathbf{T1}}.
\end{aligned}$$

From Lemma 8, we have $\|B^{\mathbf{T1}}\| = O_p(k_n^{5a/2+3/2}/\sqrt{n} + k_n^{a/2+1/2-b})$. Therefore we only need to calculate the three terms $A^{\mathbf{T1}}, B^{\mathbf{T1}}$ and $C^{\mathbf{T1}}$. Note that

$$\|C^{\mathbf{T1}}\| = \sqrt{\langle C, C \rangle} = \sqrt{\sum_{j=k_n+1}^{\infty} \langle \boldsymbol{\theta}_0, v_j \rangle^2} = O(\sqrt{\sum_{j>k_n} j^{-2b}}) = O_p(k_n^{1/2-b}),$$

and $A^{\mathbf{T1}}$ has the following decompositions, $A^{\mathbf{T1}} = (A_1^{\mathbf{T1}} + A_2^{\mathbf{T1}} + A_3^{\mathbf{T1}} + A_4^{\mathbf{T1}})(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,0})$, where

$$A_1^{\mathbf{T1}} = \sum_{j=1}^{k_n} (\tilde{r}_j - \mathrm{E}\langle Z, v_j\rangle W/\lambda_j)(\hat{v}_j - v_j); A_2^{\mathbf{T1}} = \sum_{j=1}^{k_n} (\tilde{r}_j - \mathrm{E}\langle Z, v_j\rangle W/\lambda_j) v_j;$$

$$A_3^{\mathbf{T1}} = \sum_{j=1}^{k_n} \mathrm{E}\langle Z, v_j\rangle W/\lambda_j(\hat{v}_j - v_j); A_4^{\mathbf{T1}} = \sum_{j=1}^{k_n} \mathrm{E}\langle Z, v_j\rangle W/\lambda_j v_j.$$

In the following we will calculate the four terms $A_1^{\mathbf{T1}}, A_2^{\mathbf{T1}}, A_3^{\mathbf{T1}}$ and $A_4^{\mathbf{T1}}$. We have $\|A_3^{\mathbf{T1}}\| \leq$

$$\sqrt{\mathrm{E}\sum_{j=1}^{k_n} \langle Z, v_j\rangle^2 W^2/\lambda_j^2 \sum_{j=1}^{k_n} \|\hat{v}_j - v_j\|^2} = O_p(k_n^{(a+1)/2}) \sqrt{O_p(\sum_{j=1}^{k_n} 1/\zeta_j^2)\frac{1}{n}} = O_p(k_n^{(3a+4)/2}/\sqrt{n})$$

using Lemma 1, 2 and the Cauchy's inequality; we have $\|A_4^{\mathbf{T1}}\| \leq$

$$\sqrt{\mathrm{E}\sum_{j=1}^{k_n} \langle Z, v_j\rangle^2 W^2/\lambda_j^2} = \sqrt{\sum_{j=1}^{k_n} \mathrm{E}(\xi_j^2 W^2)/\lambda_j} = O_p(k_n^{(a+1)/2})$$

by Assumption 2; similar to Lemma 6, (i), we have

$$\sum_{j=1}^{k_n} j^x \lambda_j (\tilde{r}_j^* - \mathrm{E}\langle Z, v_j\rangle W/\lambda_j)^2 = O_p(k_n^{1+x}/n), \text{ for any } x \neq -1,$$

so that the following equation holds,

$$\|A_{21}^{\mathbf{T1}}\| \triangleq \sqrt{\sum_{j=1}^{k_n} [\tilde{r}_j^* - \mathrm{E}\langle Z, v_j\rangle W/\lambda_j]^2} = O_p(k_n^{(a+1)/2});$$

similar to Lemma 6, (iii), we have

$$\sum_{j=1}^{k_n} \lambda_j |\tilde{r}_j - \tilde{r}_j^*|^2 = O_p(k_n^{4a+2} n^{-1} + k_n^{-2b}),$$

so that the following equation holds,

$$\|A_{22}^{\mathbf{T1}}\| \triangleq \sqrt{\sum_{j=1}^{k_n} [\tilde{r}_j - \tilde{r}_j^*]^2} \leq \sqrt{\sum_{j=1}^{k_n} \lambda_j [\tilde{r}_j - \tilde{r}_j^*]^2/\lambda_{k_n}}$$

$$= \sqrt{O_p(k_n^{4a+2} n^{-1} + k_n^{-2b}) O_p(k_n^a)} = O_p(k_n^{5/2a+1} n^{-1/2} + k_n^{a/2-b});$$

consequently, we have $\|A_2^{\mathbf{T1}}\| \le \|A_{21}^{\mathbf{T1}}\| + \|A_{22}^{\mathbf{T1}}\| = O_p(k_n^{(a+1)/2})$. Similar to the derivation of $\|A_3^{\mathbf{T1}}\|$, we can prove that $\|A_1^{\mathbf{T1}}\| = O_p(\|A_2^{\mathbf{T1}}\|)$.

In all,

$$
\begin{aligned}
& \|A^{\mathbf{T1}}\| \le \|A_1^{\mathbf{T1}} + A_2^{\mathbf{T1}} + A_3^{\mathbf{T1}} + A_4^{\mathbf{T1}}\| \|\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,0}\| \\
= \; & O_p(k_n^{(a+1)/2}) O_p(k_n^{2a+1} n^{-1/2} + k_n^{1/2-b}) = O_p(k_n^{5/2a+3/2} n^{-1/2} + k_n^{1+a/2-b}).
\end{aligned}
$$

and finally, we get

$$
\begin{aligned}
& \|\boldsymbol{\theta}(\hat{\boldsymbol{\beta}}_1) - \boldsymbol{\theta}_0\| \le \|A^{\mathbf{T1}}\| + \|B^{\mathbf{T1}}\| + \|C^{\mathbf{T1}}\| \\
= \; & O_p(k_n^{5/2a+3/2} n^{-1/2} + k_n^{1+a/2-b}).
\end{aligned}
$$

□

Lemma 10. *Under Assumption A.1–A.9 and B.1–B.5, we have*

(i)

$$
F(\langle Z_l, \boldsymbol{\theta}_0\rangle, G(\langle Z_l, W_l\rangle), W_l) = \mathrm{E}[\langle Z_l, \boldsymbol{\theta}_0\rangle \delta_l \exp(\gamma_0 Y_l) | Z_l, W_l];
$$

(ii) for any $C_1 > 0$, there exists a constant $C_2 > 0$, such that

$$
\sup_{\sum_{i=1}^3 \|x_i\| \le C_1} \max_{i=1,2,3} \left|\frac{\partial F(x_1, x_2, x_3)}{\partial x_i}\right| \le C_2, \quad \sup_{\sum_{i=1}^3 \|x_i\| \le C_1} F(x_1, x_2, x_3) \le C_2,
$$

holds.

Proof of Lemma 10 (i) By algebraic calculations, we have

$$\begin{aligned}
& \mathrm{E}[\langle Z_l, \boldsymbol{\theta}_0\rangle \delta_l \exp(\gamma Y_l)|Z_l, W_l] \\
= & \langle Z_l, \boldsymbol{\theta}_0\rangle \mathrm{E}[\Pr(\delta_l = 1|Z_l, W_l, Y_l)\exp(\gamma_0 Y_l)|Z_l, W_l] \\
= & \langle Z_l, \boldsymbol{\theta}_0\rangle \mathrm{E}\{\frac{\exp(G(Z_l, W_l))}{1+\exp(\langle g, Z_l\rangle + \langle \boldsymbol{\beta}_{2,0}, W_l\rangle + \phi_0 Y_l)}|Z_l, W_l\} \\
= & \int \frac{\langle Z_l, \boldsymbol{\theta}_0\rangle \exp(G(Z_l, W_l))}{1+\exp(G(Z_l, W_l)+\phi_0 y)} \frac{1}{\sqrt{2\pi}} \exp(-\frac{(y - \langle Z_l, \boldsymbol{\theta}_0\rangle - \boldsymbol{\beta}_{1,0}^T W_l)^2}{2\sigma^2}) dy,
\end{aligned}$$

which is a measurable function of $\langle Z_l, \boldsymbol{\theta}_0\rangle, G(Z_l, W_l)$ and W.

□

Proof of Lemma 10 (ii) Continue with Lemma 10, (i), and we obtain the explicit form of the function F in the following.

$$F(x_1, x_2, x_3) = \frac{1}{\sqrt{2\pi}} \int \frac{x_1 \exp(x_2)}{1+\exp(x_2 + \boldsymbol{\beta}_{2,0}^T x_3 + \phi_0 y)} \exp(-\frac{(y - x_1 - \boldsymbol{\beta}_{1,0}^T x_3)^2}{2\sigma^2}) dy,$$

and

$$\begin{aligned}
\frac{\partial F(x_1, x_2, x_3)}{\partial x_1} = & \frac{1}{\sqrt{2\pi}} \int \frac{\exp(x_2)}{1+\exp(x_2+\phi_0 y)} \exp(-\frac{(y - x_1 - \boldsymbol{\beta}_{1,0}^T x_3)^2}{2\sigma^2}) dy \\
+ & \frac{1}{\sqrt{2\pi}} \int \frac{x_1 \exp(x_2)}{1+\exp(x_2+\phi_0 y)} \exp(-\frac{(y - x_1 - \boldsymbol{\beta}_{1,0}^T x_3)^2}{2\sigma^2}) \frac{y - x_1 - \boldsymbol{\beta}_{1,0}^T x_3}{\sigma^2} dy.
\end{aligned}$$

Note that both F and $\partial F/\partial x_1$ are continuous functions of (x_1, x_2, x_3) within the compact set

$$\{(x_1, x_2, x_3)| \sum_{i=1}^{3} \|x_i\| \le C_1\}.$$

Therefore, there exists $C_{2,1} > 0, C_{2,0} > 0$ such that

$$\sup_{\sum_{i=1}^3 \|x_i\| \le C_1} |\frac{\partial F(x_1, x_2, x_3)}{\partial x_1}| \le C_{2,1}, \sup_{\sum_{i=1}^3 \|x_i\| \le C_1} F(x_1, x_2, x_3) \le C_{2,0}$$

Similarly, there exist constants $C_{2,i} > 0, i = 2, 3$ such that

$$\sup_{\sum_{i=1}^{3}\|x_i\| \le C_1} \left|\frac{\partial F(x_1, x_2, x_3)}{\partial x_i}\right| \le C_{2,i},$$

and then we have

$$\sup_{\sum_{i=1}^{3}\|x_i\| \le C_1} \max_i \left|\frac{\partial F(x_1, x_2, x_3)}{\partial x_i}\right| \le \max_{i \le 3}(C_{2,i}),$$

which completes the proof. □

Lemma 11. *Under Assumption A.1–A.9 and B.1–B.5, we have*

(i) Given a constant function $z \in \mathbb{H}$, a constant vector $x \in \mathbb{R}^p$, and a constant $w_0 \in (0, 1)$, there exist constants $0 \le c_1 \le c_2 < \infty$ such that

$$c_1\psi_{z,x}(h) \le \mathrm{E}[K_h(w_0\sqrt{\sum_{j_1=1}^{k_n}\langle Z - z, v_{j_1}\rangle^2} + (1 - w_0)\|W - x\|)] \le c_2\phi_x(h/(1 - w_0)),$$

where $\psi_{z,x}(h)$ is defined in Assumption (B.4) and $\phi_x(h) \triangleq \Pr(W \in \{\tilde{x}|\ \|\tilde{x} - x\| \le h\})$.

(ii) The following two inequalities

$$\begin{aligned}
&\sup_{z,x}\mathrm{E}\Big\{\Big[\frac{1/n\sum_{l=1}^{n}K_h^{(l)}(w_0\sqrt{\sum_{j_1=1}^{k_n}\langle Z_l - Z_i, v_{j_1}\rangle^2} + (1 - w_0)\|W_l - W_i\|)}{\mathrm{E}[K_h(w_0\sqrt{\sum_{j_1=1}^{k_n}\langle Z - Z_i, v_{j_1}\rangle^2} + (1 - w_0)\|W - W_i\|)|Z_i, W_i]} \\
- &\ \mathrm{E}(\langle Z_i, \boldsymbol{\theta}_0\rangle\delta_i\exp(\gamma Y_i)|Z_i, W_i)\Big]^2\Big|Z_i = z, W_i = x\Big\} \\
\triangleq &\ L_{11}^{(1)} \le constant \times (h^2 + \frac{1}{n\psi(h)}),
\end{aligned}$$

and

$$\sup_{z,x} \mathrm{E}\Big\{ \Big[\frac{1/n \sum_{l=1}^{n} \tilde{K}_h^{(l)}(w_0 \sqrt{\sum_{j_1=1}^{k_n} \langle Z_l - Z_i, v_{j_1}\rangle^2} + (1-w_0)\|W_l - W_i\|)}{\mathrm{E}[K_h(w_0\sqrt{\sum_{j_1=1}^{k_n}\langle Z - Z_i, v_{j_1}\rangle^2} + (1-w_0)\|W - W_i\|)|Z_i, W_i]}$$

$$- \quad \mathrm{E}(\delta_i \exp(\gamma Y_i)|Z_i, W_i)\Big]^2 | Z_i = z, W_i = x \Big\}$$

$$\triangleq \quad L_{11}^{(2)} \le constant \times (h^2 + \frac{1}{n\psi(h)})$$

hold.

Proof of Lemma 11 (i). From Assumption (B.3) and the definition of type I kernel in Martinez C. A. (2013), we have

$$\begin{aligned}
& \mathrm{E}[K_h(w_0\sqrt{\sum_{j_1=1}^{k_n}\langle Z - z, v_{j_1}\rangle^2} + (1-w_0)\|W - x\|)] \\
\ge \quad & c_1 \Pr((Z,W) \in \{(\tilde{z},\tilde{x})|\ w_0\sqrt{\sum_{j=1}^{k_n}\langle \tilde{z} - z, v_j\rangle^2} + (1-w_0)\|\tilde{x} - x\| \le h\}) \\
\ge \quad & c_1 \Pr((Z,W) \in \{(\tilde{z},\tilde{x})|\ \sqrt{\sum_{j=1}^{k_n}\langle \tilde{z} - z, v_j\rangle^2} \le h, \|\tilde{x} - x\| \le h\}) \\
\ge \quad & c_1 \Pr((Z,W) \in \{(\tilde{z},\tilde{x})|\ \|\tilde{z} - z\| \le h, \|\tilde{x} - x\| \le h\}) = c_1\psi_{z,x}(h),
\end{aligned}$$

and

$$\begin{aligned}
& \mathrm{E}[K_h(w_0\sqrt{\sum_{j_1=1}^{k_n}\langle Z - z, v_{j_1}\rangle^2} + (1-w_0)\|W - x\|)] \\
\le \quad & c_2 \Pr((Z,W) \in \{(\tilde{z},\tilde{x})|\ w_0\sqrt{\sum_{j=1}^{k_n}\langle \tilde{z} - z, v_j\rangle^2} + (1-w_0)\|\tilde{x} - x\| \le h\}) \\
\le \quad & c_2 \Pr((Z,W) \in \{(\tilde{z},\tilde{x})|\ (1-w_0)\|\tilde{x} - x\| \le h\}) = c_2\phi_x(\frac{h}{1-w_0}),
\end{aligned}$$

which complete the proof. □

Proof of Lemma 11 (ii). Before the proof, note that from Assumption (B.3) the kernel function $K(\cdot)$ satisfies,

$$\int K(t)dt = 1; \int tK(t)dt = 0; \int K^2(t)dt \leq constant.$$

We divide the proof into two parts. In Part 1, we calculate the bias while in Part 2 the variance is calculated.

Part 1. For simplicity, denote

$$D_l^{(i)} = w_0 \sqrt{\sum_{j_1=1}^{k_n} \langle Z_l - Z_i, v_{j_1} \rangle^2} + (1 - w_0)\|W_l - W_i\|,$$

$$R_0^{(i)} = \mathrm{E}[K_h(w_0 \sqrt{\sum_{j_1=1}^{k_n} \langle Z - Z_i, v_{j_1} \rangle^2} + (1 - w_0)\|W - W_i\|)|Z_i, W_i],$$

and

$$E_0^{(i)} = \mathrm{E}(\langle Z_i, \boldsymbol{\theta}_0 \rangle \delta_i \exp(\gamma_0 Y_i)|Z_i, W_i).$$

In this step we calculate the bias $A^{(11)}(Z_i, W_i)$ first. Using Lemma 10, (i), Assumption B.3, and Lemma 11, (i), we have

$$\begin{aligned}
A^{(11)}(Z_i, W_i) &\triangleq \mathrm{Bias}(\frac{1/n \sum_{l=1}^n K_h^{(l)}(D_l^{(i)})}{R_0^{(i)}}|Z_i, W_i) = \mathrm{E}[\frac{1/n \sum_{l=1}^n K_h^{(l)}(D_l^{(i)})}{R_0^{(i)}}|Z_i, W_i] - E_0^{(i)} \\
&= \mathrm{E}[\frac{\frac{1}{n}\sum_{l=1}^n K_h(D_l^{(i)})\delta_l \exp(\gamma_0 Y_l)\langle Z_l, \boldsymbol{\theta}_0\rangle}{R_0^{(i)}}|Z_i, W_i] - E_0^{(i)} \\
&= \mathrm{E}[\{1/n \sum_{l=1}^n K_h(D_l^{(i)})\mathrm{E}[\delta_l \exp(\gamma_0 Y_l)\langle Z_l, \boldsymbol{\theta}_0\rangle|Z_i, W_i, Z_l, W_l]|Z_i, W_i\}/R_0^{(i)}] - E_0^{(i)} \\
&= \mathrm{E}[1/n \sum_{l=1}^n K_h(D_l^{(i)})F(\langle Z_l, \boldsymbol{\theta}_0\rangle, G(Z_l, W_l), W_l)|Z_i, W_i]/R_0^{(i)} - E_0^{(i)} \\
&= \mathrm{E}[1/n \sum_{l=1}^n K_h(D_l^{(i)})1_{D_l^{(i)} \leq h} F(\langle Z_l, \boldsymbol{\theta}_0\rangle, G(Z_l, W_l), W_l)|Z_i, W_i]/R_0^{(i)} - E_0^{(i)}
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathrm{E}[1/n\sum_{l=1}^{n}K_h(D_l^{(i)})1_{\|Z_l-Z_i\|\leq\frac{h}{w_0}}1_{\|W_l-W_i\|\leq\frac{h}{1-w_0}}F(\langle Z_l,\boldsymbol{\theta}_0\rangle,G(Z_l,W_l),W_l)|Z_i,W_i]/R_0^{(i)}-E_0^{(i)}\\
&\leq \mathrm{E}[1/n\sum_{l=1}^{n}K_h(D_l^{(i)})|Z_i,W_i][F(\langle Z_i,\boldsymbol{\theta}_0\rangle,G(Z_i,W_i),W_i)+\mathit{constant}\times h]/R_0^{(i)}-E_0^{(i)}\\
&= R_0^{(i)}(E_0^{(i)}+\mathit{constant}\times h)]/R_0^{(i)}-E_0^{(i)} \qquad \text{(S3.3)}
\end{aligned}
$$

which equals $constant * h$. Next we prove that this constant exists uniformly for $Z_i = z$ and $W_i = w$; in other words, there exists constant C such that

$$
\Pr(\sup_{Z_i,W_i} A^{(11)}(Z_i,W_i)\leq Ch)=1.
$$

From Assumption (B.1), we have $|\langle Z_i,\boldsymbol{\theta}_0\rangle|\leq\|Z_i\|\|\boldsymbol{\theta}_0\|\leq constant$ and $\|W_l\|\leq constant$; from Assumption (B.2), we have $|G(Z,W)|\leq G_1^{-1}(\|Z\|+\|W\|)\leq constant$. It follows from Lemma 10, (ii) that there exist constant C_2, such that

$$
\begin{aligned}
& |F(\langle Z_l,\boldsymbol{\theta}_0\rangle,G(Z_l,W_l),W_l)-F(\langle Z_i,\boldsymbol{\theta}_0\rangle,G(Z_i,W_i),W_i)|\\
\leq\ & |F(\langle Z_l,\boldsymbol{\theta}_0\rangle,G(Z_l,W_l),W_l)-F(\langle Z_i,\boldsymbol{\theta}_0\rangle,G(Z_l,W_l),W_l)|\\
+\ & |F(\langle Z_i,\boldsymbol{\theta}_0\rangle,G(Z_l,W_l),W_l)-F(\langle Z_i,\boldsymbol{\theta}_0\rangle,G(Z_i,W_i),W_l)|\\
+\ & |F(\langle Z_i,\boldsymbol{\theta}_0\rangle,G(Z_i,W_i),W_l)-F(\langle Z_i,\boldsymbol{\theta}_0\rangle,G(Z_i,W_i),W_i)|\\
\leq\ & C_2[\langle|Z_l-Z_i|,\boldsymbol{\theta}_0\rangle+|G(Z_l,W_l)-G(Z_i,W_i))|+\|W_l-W_i\|]\\
\leq\ & \mathit{constant}*[\langle|Z_l-Z_i|,\boldsymbol{\theta}_0\rangle+\|Z_l-Z_i\|+\|W_l-W_i\|+\|W_l-W_i\|],
\end{aligned}
$$

The last inequality holds from the Lipschitz's condition in Assumption B.2, and the constant here is irrelevant to Z_l, Z_i, W_l and W_i. So that

$$
\begin{aligned}
& 1_{\|Z_l-Z_i\|\leq\frac{h}{w_0}}1_{\|W_l-W_i\|\leq\frac{h}{1-w_0}}F(\langle Z_l,\boldsymbol{\theta}_0\rangle,G(Z_l,W_l),W_l)\\
\leq\ & 1_{\|Z_l-Z_i\|\leq\frac{h}{w_0}}1_{\|W_l-W_i\|\leq\frac{h}{1-w_0}}[F(\langle Z_i,\boldsymbol{\theta}_0\rangle,G(Z_i,W_i),W_i)+\mathit{constant}\times h],
\end{aligned}
$$

and the constant here is irrelevant to Z_l, Z_i, W_l and W_i. This illustrate the constant in (S3.3) is irrelevant to Z_l, Z_i, W_l and W_i. It follows that

$$\Pr(\sup_{Z_i,W_i} A^{(11)}(Z_i, W_i) \le Ch) = 1.$$

Part 2. First we calculate

$$B^{(11)}(Z_i, W_i) \triangleq \mathrm{E}\big[[K_h^{(l)}(D_l^{(i)})]^2|Z_i, W_i\big]/[R_0^{(i)}]^2,$$

for $l \neq i$. Using Lemma 10, (ii), we have $B^{(11)}(Z_i, W_i) =$

$$\begin{aligned}
&\frac{\mathrm{E}\big[K_h^2(D_l^{(i)})\delta_l[\exp(\gamma_0 Y_l)\langle Z_l, \boldsymbol{\theta}_0\rangle]^2|Z_i, W_i\big]}{[R_0^{(i)}]^2} \\
= \;& \mathrm{E}\big[K_h^2(D_l^{(i)})\langle Z_l, \boldsymbol{\theta}_0\rangle \times \mathrm{E}[\delta_l \exp(2\gamma_0 Y_l)\langle Z_l, \boldsymbol{\theta}_0\rangle|Z_i, W_i, Z_l, W_l]|Z_i, W_i\big]/[R_0^{(i)}]^2 \\
= \;& \mathrm{E}[K_h^2(D_l^{(i)})\exp(\gamma_0 Y_l)\langle Z_l, \boldsymbol{\theta}_0\rangle F(\langle Z_l, \boldsymbol{\theta}_0\rangle, \langle Z_l, g\rangle, W_l; 2\gamma_0)|Z_i, W_i]/[R_0^{(i)}]^2 \\
= \;& \textit{constant} \times \mathrm{E}[K_h^2(D_l)|Z_i, W_i]/[R_0^{(i)}]^2 \\
\le \;& \textit{constant} \times \mathrm{E}[K_h(D_l^{(i)})|Z_i, W_i]/[R_0^{(i)}]^2 = \textit{constant} \times \frac{1}{R_0^{(i)}}.
\end{aligned}$$

The last inequality is from Assumption (B.3).

Then for $l_0 \neq i$, we have

$$\begin{aligned}
&\mathrm{Var}[1/n\sum_{l=1}^{n}[K_h^{(l)}(D_l^{(i)})]/R_0^{(i)}|Z_i, W_i] \\
\le \;& \frac{1}{n^2[R_0^{(i)}]^2} n\mathrm{Var}\big[[K_h^{(l_0)}(D_{l_0}^{(i)})]|Z_i, W_i\big] = \frac{1}{n}\mathrm{Var}\big[[K_h^{(l_0)}(D_{l_0}^{(i)})]/R_0^{(i)}|Z_i, W_i\big] \\
\le \;& \frac{1}{n}\mathrm{E}\big[K_h^2(D_{l_0}^{(i)})/[R_0^{(i)}]^2|Z_i, W_i\big] = O_p(\frac{1}{nR_0^{(i)}}),
\end{aligned}$$

where the O_p, o_p term and the relevant constant hold uniformly with respect to Z_i, W_i by Assumption (B.1).

Finally, we have

$$
\begin{aligned}
L_{11}^{(1)} &\leq \sup_{Z_i,W_i}[A^{(11)}(Z_i,W_i)]^2 + \sup_{Z_i,W_i}\mathrm{Var}(\frac{1/n\sum_{l=1}^n K_h^{(l)}(D_0)}{R_0}|Z_i,W_i) \\
&\leq \textit{constant} \times (h^2 + \sup_{Z_i,W_i}\frac{1}{n\psi_{Z_i,X_i}(h)}) \\
&\leq \textit{constant} \times (h^2 + \frac{1}{n\psi(h)}).
\end{aligned}
$$

It can be proved in the same way that

$$
L_{11}^{(2)} \leq \textit{constant} \times (h^2 + \frac{1}{n\psi(h)}).
$$

□

Lemma 12. *Under Assumption A.1–A.9 and B.1–B.5, we have*

(i)

$$
\sup_{j\leq k_n}\zeta_j\lambda_j|\Delta_2(\hat{r}_j)| = O_p(k_n^{a-1}/\sqrt{n}),
$$

where

$$
\Delta_2(\hat{r}_j) = \hat{r}_j - \frac{\sum_{i=1}^n[\delta_i M_j(Y_i,Z_i,W_i,v_j;\boldsymbol{\beta}_{1,0}) + (1-\delta_i)\hat{m}_{M_j,i,\gamma}(Y_i,Z_i,W_i,v_j;\boldsymbol{\beta}_{1,0})]}{n\hat{\lambda}_j};
$$

(ii)

$$
\sum_{j=1}^{k_n}\lambda_j|\hat{r}_j - \hat{r}_j^*|^2 = O_p(k_n^{4a+2}n^{-1} + k_n^{-2b});
$$

(iii)

$$
\sum_{j=1}^{\infty}\lambda_j^2|\hat{r}_j^* - r_j^*|^2 = O_p[h^2 + \frac{1}{n\psi(h)}].
$$

(iv) $L_{12} \triangleq$

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\Delta_1\left\{\sum_{j=1}^{k_n}\hat{r}_j\big[\delta_i W_i\langle Z_i,\hat{v}_j\rangle + \sum_{l=0}^{n} w_{l,0}(Z_i,W_i;\gamma)\langle Z_l,\hat{v}_j\rangle W_l\big]\right\} \\
= & \; O_p(k_n^{2a+1}n^{-1/2} + k_n^{-b} + k_n^{(4a+3)/2}n^{-1/2}[h + \frac{1}{\sqrt{n\psi(h)}}]),
\end{aligned}
$$

where

$$
\begin{aligned}
& \Delta_1\left\{\sum_{j=1}^{k_n}\hat{r}_j\big[\delta_i W_i\langle Z_i,\hat{v}_j\rangle + \sum_{l=0}^{n} w_{l,0}(Z_i,W_i;\gamma)\langle Z_l,\hat{v}_j\rangle W_l\big]\right\} \\
= & \sum_{j=1}^{k_n}\hat{r}_j\big[\delta_i W_i < Z_i,\hat{v}_j > + \sum_{l=0}^{n} w_{l,0}(Z_i,W_i;\gamma)\langle Z_l,\hat{v}_j\rangle W_l\big] \\
- & \sum_{j=1}^{k_n}\hat{r}_j^*\big[\delta_i W_i\langle Z_i,v_j\rangle + \sum_{l=0}^{n} w_{l,0}(Z_i,W_i;\gamma)\langle Z_l,v_j\rangle W_l\big].
\end{aligned}
$$

Proof of Lemma 12 (i). The proof is exactly the same as Lemma 6, (ii) except that it uses $A_2^{(12,1)}$ to replace $A_2^{(6,2)}$, where $A_2^{(12,1)}$ is expressed in the following (here $w_{l,0}^{(i)} \triangleq w_{l,0}(Z_i,W_i;\gamma)$ which is defined by (2.7)). $A_2^{(12,1)} \triangleq$

$$
\begin{aligned}
& \mathrm{E}[\sup_{j\le k_n}\zeta_j\sum_{l=1}^{n} w_{l,0}^{(i)}\langle Z_l,\hat{v}_j - v_j\rangle(\langle Z_l,\boldsymbol{\theta}_0\rangle + \epsilon_l)] \\
= & \sum_{l_0=1}^{n}\Pr(\arg\max_{l\le n} w_{l,0}^{(i)} = l_0)\mathrm{E}[\sup_{j\le k_n}\zeta_j\langle Z_{l_0},\hat{v}_j - v_j\rangle(\langle Z_{l_0},\boldsymbol{\theta}_0\rangle + \epsilon_{l_0})|\arg\max_{l\le n} w_{l,0}^{(i)} = l_0] \\
\le & \sup_{l_0\le n}\mathrm{E}[\sup_{j\le k_n}\zeta_j\langle Z_{l_0},\hat{v}_j - v_j\rangle(\langle Z_{l_0},\boldsymbol{\theta}_0\rangle + \epsilon_{l_0})|\arg\max_{l\le n} w_{l,0}^{(i)} = l_0] \\
= & \mathrm{E}[\sup_{j\le k_n}\zeta_j\langle Z_{l_0},\hat{v}_j - v_j\rangle(\langle Z_{l_0},\boldsymbol{\theta}_0\rangle + \epsilon_{l_0})|\arg\max_{l\le n} w_{l,0}^{(i)} = l_0] \\
\le & \sqrt{E[\langle Z,Z\rangle(\langle Z,\boldsymbol{\theta}_0\rangle + \epsilon)^2|\arg\max_{l\le n} w_{l,0}^{(i)} = l_0]} \\
& \sqrt{\mathrm{E}[\sup_{j}\zeta_j^2\|\hat{v}_j - v_j\|^2\big|\arg\max_{l\le n} w_{l,0}^{(i)} = l_0]} \\
\le & \; \textit{constant} \times O_p(n^{-1/2}).
\end{aligned}
$$

□

Proof of Lemma 12 (ii). The proof is exactly the same as Lemma 6, (iii), except using

$$A^{(12,2)} = \sum_{j=1}^{k_n} 2\lambda_j \Delta_2(\hat{r}_j)^2$$

to replace $A^{(6,3)}$, where its expectation is calculated in the following (here $w_{l,0}^{(i)} \triangleq w_{l,0}(Z_i, W_i; \gamma)$ which is defined by (2.7)). Since $w_{l,0}^{(i)}$ is positive, we have $\sum_{l=1}^{n}(w_{l,0}^{(i)})^2 \le (\sum_{l=1}^{n} w_{l,0}^{(i)})^2 = 1$, and it follows that $\mathrm{E}A^{(12,2)} =$

$$\begin{aligned}
& 2\mathrm{E}(\sum_{j=1}^{k_n} \lambda_j \big|\frac{1}{n}\sum_{i=1}^{n} \langle Z_i, \hat{v}_j - v_j\rangle(\langle Z_i, \boldsymbol{\theta}_0\rangle + \epsilon_i)\big|^2/\hat{\lambda}_j) \\
+ \quad & 2\mathrm{E}(\sum_{j=1}^{k_n} \lambda_j \big|\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{n} w_{l,0}^{(i)}[\langle Z_l, \hat{v}_j - v_j\rangle(\langle Z_l, \boldsymbol{\theta}_0\rangle + \epsilon_l)]\big|^2/\hat{\lambda}_j) \\
\le \quad & 2(\sup_{j\le k_n} \frac{\lambda_j}{\lambda_{j+1}}) \times (\sum_{j=1}^{k_n} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}|\langle Z_i, \hat{v}_j - v_j\rangle(\langle Z_i, \boldsymbol{\theta}_0\rangle + \epsilon_i)|^2) \\
+ \quad & 2(\sup_{j\le k_n} \frac{\lambda_j}{\lambda_{j+1}}) \times (\sum_{j=1}^{k_n} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\big|\sum_{l=1}^{n} w_{l,0}^{(i)}[\langle Z_l, \hat{v}_j - v_j\rangle(\langle Z_l, \boldsymbol{\theta}_0\rangle + \epsilon_l)]\big|^2) \\
\le \quad & 2(\sup_{j\le k_n} \frac{\lambda_j}{\lambda_{j+1}}) \times (\sum_{j=1}^{k_n} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}|\langle Z_i, \hat{v}_j - v_j\rangle(\langle Z_i, \boldsymbol{\theta}_0\rangle + \epsilon_i)|^2) \\
+ \quad & 2(\sup_{j\le k_n} \frac{\lambda_j}{\lambda_{j+1}}) \times (\sum_{j=1}^{k_n} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\sum_{l=1}^{n}(w_{l,0}^{(i)})^2 \sum_{l=1}^{n}[\langle Z_l, \hat{v}_j - v_j\rangle(\langle Z_l, \boldsymbol{\theta}_0\rangle + \epsilon_l)]^2) \\
\le \quad & 4(\sup_{j\le k_n} \frac{\lambda_j}{\lambda_{j+1}}) \times (\sum_{j=1}^{k_n} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}|\langle Z_i, \hat{v}_j - v_j\rangle(\langle Z_i, \boldsymbol{\theta}_0\rangle + \epsilon_i)|^2) \\
\le \quad & \textit{constant} \times k_n^{a-1} \sum_{j=1}^{k_n} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[\langle Z_i, Z_i\rangle \|\hat{v}_j - v_j\|^2(\langle Z_i, \boldsymbol{\theta}_0\rangle + \epsilon_i)^2] \\
\le \quad & \textit{constant} \times k_n^{a-1} \sum_{j=1}^{k_n} \zeta_j^{-2} n^{-1} = O(k_n^{a-1}\sum_{j=1}^{k_n} j^{3a+2}/n) = O(k_n^{4a+2} n^{-1}).
\end{aligned}$$

□

Proof of Lemma 12 (iii). After straightforward algebraic calculation, $\sum_{j=1}^{k_n} \lambda_j^2(\hat{r}_j^* - r_j^*)^2 =$

$$
\begin{aligned}
& \sum_{j=1}^{k_n} [\frac{1}{n} \sum_{i=1}^{n} \left(\hat{m}_{M_j,i,\gamma_0}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_{1,0}) - m^0_{M_j,i,\gamma_0}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_{1,0}) \right)]^2 \\
\leq \; & \sum_{j=1}^{k_n} \frac{1}{n} \sum_{i=1}^{n} \left(\hat{m}_{M_j,i,\gamma_0}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_{1}) - m^0_{M_j,i,\gamma_0}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_{1,0}) \right)^2.
\end{aligned}
$$

Next we calculate $|\hat{m}_{M_j,i,\gamma_0}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_1) - m^0_{M_j,i,\gamma_0}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_1)|$. Denote

$$
A_i^{(12,3)} = 1/n \sum_{l=1}^{n} K_h^{(l)}(w_0 \sqrt{\sum_{j_1=1}^{k_n} \langle Z_l - Z_i, v_{j_1} \rangle^2} + (1-w_0)\|W_l - W_i\|),
$$

$$
B_i^{(12,3)} = 1/n \sum_{l=1}^{n} \tilde{K}_h^{(l)}(w_0 \sqrt{\sum_{j_1=1}^{k_n} \langle Z_l - Z_i, v_{j_1} \rangle^2} + (1-w_0)\|W_l - W_i\|),
$$

$$
C_i^{(12,3)} = \mathrm{E}(\delta_i \exp(\gamma_0 Y_i) \langle \boldsymbol{\theta}_0, Z_i \rangle | Z_i, W_i), D_i^{(12,3)} = \mathrm{E}(\delta_i \exp(\gamma_0 Y_i) | Z_i, W_i),
$$

and

$$
E_i^{(12,3)} = \mathrm{E}[K_h(w_0 \sqrt{\sum_{j_1=1}^{k_n} \langle Z - Z_i, v_{j_1} \rangle^2} + (1-w_0)\|W - W_i\|) | Z_i, W_i].
$$

By Condition (B.1), we have $D_i^{(12,3)} \geq$

$$
\begin{aligned}
& \inf_{\max\{\|z\|,\|x\|\} \leq C_1} \mathrm{E}(\delta_i | Z_i = z, W_i = x) \\
= \; & \inf_{\max\{\|z\|,\|x\|\} \leq C_1} \int \frac{\phi y + G(z,x)}{1 + \exp(\phi_0 Y_i + G(z,x))} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y - \boldsymbol{\beta}_{1,0}^T x - \langle \boldsymbol{\theta}_0, z \rangle)^2} dy \\
\geq \; & constant > 0.
\end{aligned}
$$

The last inequality is because the integral is a continuous function of $(x, G(z,x), \langle z, \boldsymbol{\theta}_0 \rangle)$ and each item of $x, G(z,x), \langle z, \boldsymbol{\theta}_0 \rangle$ is positive in a compact set from Condition (B.1) and

(B.2). From

$$|\hat{m}_{M_j,i,\gamma_0}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_1) - m^0_{M_j,i,\gamma_0}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_1)| = \langle Z_i, v_j\rangle |\frac{A_i^{(12,3)}}{B_i^{(12,3)}} - \frac{C_i^{(12,3)}}{D_i^{(12,3)}}|$$

$$+ \quad \frac{1/n\sum_{l=1}^n K_h^{(l)}(w_0\sqrt{\sum_{j_1=1}^{k_n}\langle Z_l - Z_i, v_{j_1}\rangle^2} + (1-w_0)\|W_l - W_i\|)\langle Z_l - Z_i, v_j\rangle}{B_i^{(12,3)}},$$

we have $\sum_{j=1}^{\infty}\sum_{i=1}^{n}|\hat{m}_{M_j,i,\gamma_0}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_1) - m^0_{M_j,i,\gamma_0}(Y_i, Z_i, W_i, v_j; \boldsymbol{\beta}_1)|^2/n \le F_1^{(12,3)} + F_2^{(12,3)}$, where

$$F_1^{(12,3)} = \sum_{j=1}^{\infty}\frac{2}{n}\sum_{i=1}^{n}\langle Z_i, v_j\rangle|\frac{A_i^{(12,3)}}{B_i^{(12,3)}} - \frac{C_i^{(12,3)}}{D_i^{(12,3)}}|^2;$$

$$F_2^{(12,3)} = \sum_{j=1}^{\infty}\frac{2}{n}\sum_{i=1}^{n}[\frac{1/n\sum_{l=1}^n K_h^{(l)}(w_0\sqrt{\sum_{j_1=1}^{k_n}\langle Z_l - Z_i, v_{j_1}\rangle^2} + (1-w_0)\|W_l - W_i\|)\langle Z_l - Z_i, v_j\rangle}{B_i^{(12,3)}}]^2,$$

Next we calculate the two terms $F_1^{(12,3)}$ and $F_2^{(12,3)}$ separately. By Lemma 11, (ii), the following equations hold uniformly with respect to (z, x).

$$\frac{A_i^{(12,3)}}{E_i^{(12,3)}} - C_i^{(12,3)} \le constant \times \sqrt{(h^2 + \frac{1}{n\psi(h)})}; \frac{B_i^{(12,3)}}{E_i^{(12,3)}} - D_i \le constant \times \sqrt{h^2 + \frac{1}{n\psi(h)}}.$$

It follows that $\sup_{z,x}|A_i^{(12,3)}/B_i^{(12,3)} - C_i^{(12,3)}/D_i^{(12,3)}| =$

$$|\frac{A_i^{(12,3)}/E_i^{(12,3)} - C_i^{(12,3)}}{D_i^{(12,3)} + (B_i^{(12,3)}/E_i^{(12,3)} - D_i^{(12,3)})} - \frac{C_i^{(12,3)}(D_i^{(12,3)} - B_i^{(12,3)}/E_i^{(12,3)})}{[D_i^{(12,3)} + (B_i^{(12,3)}/E_i^{(12,3)} - D_i^{(12,3)})]D_i^{(12,3)}}|$$

$$\le \quad constant \times \sqrt{(h^2 + \frac{1}{n\psi(h)})}.$$

Then we have

$$F_1^{(12,3)} \le constant \times (h^2 + \frac{1}{n\psi(h)})\sum_{j=1}^{\infty}\frac{1}{n}\sum_{i=1}^{n}\langle Z_i, v_j\rangle^2 = O_p(h^2 + \frac{1}{n\psi(h)}).$$

Note that $F_2^{(12,3)}$ can be simplified as

$$\frac{2}{n}\sum_{i=1}^{n}[\frac{1/n\sum_{l=1}^n K_h^{(l)}(w_0\sqrt{\sum_{j_1=1}^{k_n}\langle Z_l - Z_i, v_{j_1}\rangle^2} + (1-w_0)\|W_l - W_i\|)\|Z_l - Z_i\|^2}{B_i^{(12,3)}}]^2$$

which equals (similar to Lemma 11 (ii))

$$
\begin{aligned}
& \frac{2}{n}\sum_{i=1}^{n}[\frac{1/n\sum_{l=1}^{n}K_h^{(l)}(w_0\sqrt{\sum_{j_1=1}^{k_n}\langle Z_l-Z_i,v_{j_1}\rangle^2}+(1-w_0)\|W_l-W_i\|)1_{\|Z_l-Z_i\|\le\frac{h}{w_0}}\|Z_l-Z_i\|^2}{B_i^{(12,3)}}]^2 \\
= \ & O_p(h^2).
\end{aligned}
$$

Finally, we have

$$
\sum_{j=1}^{k_n}\lambda_j^2(\hat{r}_j^*-r_j^*)^2=F_1^{(12,3)}+F_2^{(12,3)}=O_p(h^2+\frac{1}{n\psi(h)}).
$$

□

Proof of Lemma 12 (iv). Denote

$$
\Delta_1[\sum_{j=1}^{k_n}\hat{r}_j(\delta_iW_i\langle Z_i,\hat{v}_j\rangle)]=\sum_{j=1}^{k_n}\hat{r}_j(\delta_iW_i\langle Z_i,\hat{v}_j\rangle)-\sum_{j=1}^{k_n}\hat{r}_j^*(\delta_iW_i\langle Z_i,v_j\rangle),
$$

and

$$
\Delta_1[\sum_{j=1}^{k_n}\hat{r}_j(\sum_{l=1}^{n}w_{l,0}\delta_lW_l\langle Z_l,\hat{v}_j\rangle)]=\sum_{j=1}^{k_n}\hat{r}_j(\sum_{l=1}^{n}w_{l,0}\delta_lW_l\langle Z_l,\hat{v}_j\rangle)-\sum_{j=1}^{k_n}\hat{r}_j^*(\sum_{l=1}^{n}w_{l,0}\delta_lW_l\langle Z_l,v_j\rangle).
$$

We have the decomposition similar to the proof of Lemma 6, (vi).

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\Delta_1[\sum_{j=1}^{k_n}\hat{r}_j(\delta_iW_i\langle Z_i,\hat{v}_j\rangle)] \\
= \ & \sum_{j=1}^{k_n}(\hat{r}_j-\hat{r}_j^*)\frac{1}{n}\sum_{i=1}^{n}\delta_iW_i\langle Z_i,\hat{v}_j-v_j\rangle+\sum_{j=1}^{k_n}(\hat{r}_j-\hat{r}_j^*)\frac{1}{n}\sum_{i=1}^{n}\delta_iW_i\langle Z_i,v_j\rangle \\
+ \ & \sum_{j=1}^{k_n}(\hat{r}_j^*-r_j^*)\frac{1}{n}\sum_{i=1}^{n}\delta_iW_i\langle Z_i,\hat{v}_j-v_j\rangle+\sum_{j=1}^{k_n}(r_j^*-\langle\boldsymbol{\theta}_0,v_j\rangle)\frac{1}{n}\sum_{i=1}^{n}\delta_iW_i\langle Z_i,\hat{v}_j-v_j\rangle \\
+ \ & \sum_{j=1}^{k_n}\langle\boldsymbol{\theta}_0,v_j\rangle\frac{1}{n}\sum_{i=1}^{n}\delta_iW_i\langle Z_i,\hat{v}_j-v_j\rangle\triangleq A^{(12,4)}+B^{(12,4)}+C^{(12,4)}+D^{(12,4)}+E^{(12,4)}.
\end{aligned}
$$

Similar to the proof of Lemma 6, (vi), and using Lemma 12, (i), and Lemma 12, (ii), we have $A^{(12,4)}+B^{(12,4)}+D^{(12,4)}+E^{(12,4)}=O_p(k_n^{2a+1}n^{-1/2}+k_n^{-b})$. We only need to calculate the order of the term $C^{(12,4)}$. Using the Cauchy's inequality and Lemma 12, (iii), we have

$$\begin{aligned} C^{(12,4)} &\leq \sqrt{\sum_{j=1}^{k_n} \lambda_j^2 (\hat{r}_j^* - r_j^*)^2 \sum_{j=1}^{k_n} [(\frac{1}{n}\sum_{i=1}^{n} \delta_i W_i \langle Z_i, \hat{v}_j - v_j\rangle)/\lambda_j]^2} \\ &= \sqrt{O_p[h^2 + \frac{1}{n\psi(h)}]\sum_{j=1}^{k_n} 1/\zeta_j^2/\lambda_j^2 \times 1/n} = O_p(\frac{k_n^{(4a+3)/2}[h + \frac{1}{\sqrt{n\psi(h)}}]}{\sqrt{n}}). \end{aligned}$$

Then we have $\frac{1}{n}\sum_{i=1}^{n} \Delta_1[\sum_{j=1}^{k_n} \hat{r}_j(\delta_i W_i \langle Z_i, \hat{v}_j\rangle)] =$

$$A^{(12,4)} + B^{(12,4)} + C^{(12,4)} + D^{(12,4)} + E^{(12,4)} = O_p(k_n^{2a+1} n^{-1/2} + k_n^{-b} + \frac{k_n^{(4a+3)/2}}{\sqrt{n}}[h + \frac{1}{\sqrt{n\psi(h)}}]).$$

Similarly, we can also prove

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \Delta_1[\sum_{j=1}^{k_n} \hat{r}_j(\sum_{l=1}^{n} w_{l,0}\delta_l W_l \langle Z_l, \hat{v}_j\rangle)] \\ = \;& O_p(k_n^{2a+1} n^{-1/2} + k_n^{-b} + k_n^{(4a+3)/2} n^{-1/2}[h + \frac{1}{\sqrt{n\psi(h)}}]). \end{aligned}$$

The conclusion holds based on the above results. □

Lemma 13. *Under Assumption A.1–A.9 and B.1–B.5, we have*

$$\begin{aligned} L_{13} \;&\triangleq\; \frac{1}{n}\sum_{i=1}^{n}\{\sum_{j=1}^{k_n} \hat{r}_j^* \big[\delta_i W_i \langle Z_i, v_j\rangle + (1-\delta_i)\sum_{l=0}^{n} w_{l,0}(Z_i, W_i; \gamma)\langle Z_l, v_j\rangle W_l\big]\} \\ &- \frac{1}{n}\sum_{i=1}^{n}\{\sum_{j=1}^{k_n} r_j^* \big|\delta_i W_i \langle Z_i, v_j\rangle + (1-\delta_i)\mathrm{E}(\langle Z_i, v_j\rangle W | Z_i, W_i, \delta_i = 0)\big|\} \\ &= O_p(k_n^{(a+1)/2}[h + 1/\sqrt{n\psi(h)}]). \end{aligned}$$

Proof of Lemma 13. First we decompose L_{13} in the following.

$$
\begin{aligned}
L_{13} \quad \leq & \quad \frac{1}{n}\sum_{i=1}^{n}\{\sum_{j=1}^{k_n}|\hat{r}_j^* - r_j^*| \times |\mathrm{E}(\langle Z_i, v_j\rangle W|Z_i, W_i, \delta_i = 0) - \sum_{l=0}^{n} w_{l,0}(Z_i, W_i;\gamma)\langle Z_l, v_j\rangle W_l|\} \\
+ & \quad \frac{1}{n}\sum_{i=1}^{n}\{\sum_{j=1}^{k_n}|\hat{r}_j^* - r_j^*| \times \left[\delta_i W_i\langle Z_i, v_j\rangle + (1-\delta_i)\mathrm{E}(\langle Z_i, v_j\rangle W|Z_i, W_i, \delta_i = 0)\right] \\
+ & \quad \frac{1}{n}\sum_{i=1}^{n}\{\sum_{j=1}^{k_n}|r_j^* - \langle \boldsymbol{\theta}_0, v_j\rangle| \times |\mathrm{E}(\langle Z_i, v_j\rangle W|Z_i, W_i, \delta_i = 0) - \sum_{l=0}^{n} w_{l,0}(Z_i, W_i;\gamma)\langle Z_l, v_j\rangle W_l|\} \\
+ & \quad \frac{1}{n}\sum_{i=1}^{n}\{\sum_{j=1}^{k_n}|\langle \boldsymbol{\theta}_0, v_j\rangle| \times |\mathrm{E}(\langle Z_i, v_j\rangle W|Z_i, W_i, \delta_i = 0) - \sum_{l=0}^{n} w_{l,0}(Z_i, W_i;\gamma)\langle W_l, v_j\rangle W_l|\} \\
\triangleq & \quad A^{(13)} + B^{(13)} + C^{(13)} + D^{(13)}.
\end{aligned}
$$

Define $\widehat{K}_h^{(l)}(\cdot) = K_h(\cdot)\delta_l \exp(\gamma_0 Y_l)W_l$. Similar to the proof of Lemma 11, (ii), we have

$$
\begin{aligned}
& \sup_{z,x} \mathrm{E}\{[\frac{1/n\sum_{l=1}^{n}\widehat{K}_h^{(l)}(w_0\sqrt{\sum_{j_1=1}^{k_n}\langle Z_l - Z_i, v_{j_1}\rangle^2} + (1-w_0)\|W_l - W_i\|)}{\mathrm{E}[K_h(w_0\sqrt{\sum_{j_1=1}^{k_n}\langle Z - Z_i, v_{j_1}\rangle^2} + (1-w_0)\|W - W_i\|)|Z_i, W_i]} \\
- & \quad \mathrm{E}(W_i\delta_i\exp(\gamma Y_i)|Z_i, W_i)]^2|Z_i = z, W_i = x\} \\
\leq & \quad \mathit{constant} \times (h^2 + \frac{1}{n\psi(h)}).
\end{aligned}
$$

Consequently, similar to the proof of Lemma 12, (iii), we get

$$
\sum_{j=1}^{k_n}|\mathrm{E}(\langle Z_i, v_j\rangle W|Z_i, W_i, \delta_i = 0) - \sum_{l=0}^{n} w_{l,0}(Z_i, W_i;\gamma)\langle Z_l, v_j\rangle W_l|^2 = O_p(h^2 + \frac{1}{n\psi(h)}).
$$

Then from Lemma 6, we have

$$
\begin{aligned}
C^{(13)} \leq & \sqrt{\sum_{j=1}^{k_n}(r_j^* - \langle \boldsymbol{\theta}_0, v_j\rangle)^2} \\
\times & \quad \sqrt{\sum_{j=1}^{k_n}|\mathrm{E}(\langle Z_i, v_j\rangle W|Z_i, W_i, \delta_i = 0) - \sum_{l=0}^{n} w_{l,0}(Z_i, W_i;\gamma)\langle Z_l, v_j\rangle W_l|^2} \\
= & \quad \sqrt{O_p(k_n^{a+1}/n)O_p(h^2 + \frac{1}{n\psi(h)})} = O_p(k_n^{(a+1)/2}n^{-1/2}[h + \frac{1}{\sqrt{n\psi(h)}}]),
\end{aligned}
$$

and

$$D^{(13)} \leq \sqrt{\sum_{j=1}^{k_n} \langle \boldsymbol{\theta}_0, v_j \rangle^2}$$
$$\times \sqrt{\sum_{j=1}^{k_n} |\mathrm{E}(\langle Z_i, v_j \rangle W | Z_i, W_i, \delta_i = 0) - \sum_{l=0}^{n} w_{l,0}(Z_i, W_i; \gamma) \langle Z_l, v_j \rangle W_l|^2}$$
$$= \sqrt{O_p(1) O_p(h^2 + \frac{1}{n\psi(h)})} = O_p(h + \frac{1}{\sqrt{n\psi(h)}}).$$

Similarly, using the Cauchy's inequality and Lemma 12, (iii), we have

$$A^{(13)} \leq \sqrt{\sum_{j=1}^{k_n} (\hat{r}_j^* - r_j^*)^2}$$
$$\times \sqrt{\sum_{j=1}^{k_n} |\mathrm{E}(\langle Z_i, v_j \rangle W | Z_i, W_i, \delta_i = 0) - \sum_{l=0}^{n} w_{l,0}(Z_i, W_i; \gamma) \langle Z_l, v_j \rangle W_l|^2}$$
$$= \sqrt{O_p(k_n^{2a}[h^2 + \frac{1}{n\psi(h)}]) O_p(h^2 + \frac{1}{n\psi(h)})} = O_p(k_n^a [h^2 + \frac{1}{n\psi(h)}]).$$

Similarly, using the Cauchy's inequality and Lemma 6, (iv), we have

$$B^{(13)} \leq \sqrt{\sum_{j=1}^{k_n} \lambda_j^2 (\hat{r}_j^* - r_j^*)^2}$$
$$\times \sqrt{\sum_{j=1}^{k_n} \{\frac{1}{n} \sum_{i=1}^{n} [\delta_i W_i \langle Z_i, v_j \rangle + (1 - \delta_i) \mathrm{E}(\langle Z_i, v_j \rangle W | Z_i, W_i, \delta_i = 0)]\}^2 / \lambda_j^2}$$
$$= \sqrt{O_p(h^2 + \frac{1}{n\psi(h)}) O_p(k_n^{a+1})} = O_p(k_n^{(a+1)/2} [h^2 + \frac{1}{n\psi(h)}]).$$

Finally, we get $L_{13} = O_p(A^{(13)} + B^{(13)} + C^{(13)} + D^{(13)}) = O_p(k_n^{(a+1)/2}[h + 1/\sqrt{n\psi(h)}])$.

□

Proof of Theorem 2.

From Lemma 4, 5, 7, 12 and 13, we get

$$\begin{aligned} &\tilde{U}(\boldsymbol{\beta}_{1,0}) = L_5 - L_4 + L_7 + L_{12} + L_{13} \\ = \;& O_p(k_n^{1/2-b} + k_n n^{-1/2} + k_n^{2a+1} n^{-1/2} + k_n^{-b} + k_n^{(4a+3)/2} n^{-1/2}[h + \frac{1}{\sqrt{n\psi(h)}}] + k_n^{(a+1)/2}[h + 1/\sqrt{n\psi(h)}]) \\ = \;& O_p(k_n^{1/2-b} + k_n^{2a+1} n^{-1/2} + k_n^{(a+1)/2}[h + 1/\sqrt{n\psi(h)}]), \end{aligned}$$

where the 'L_i' is defined in Lemma $i, i = 4, 5, 7, 12, 13$. Similar to Lemma 9, we have $\|\partial^2 \tilde{U}(\boldsymbol{\beta}_{1,0})/(\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T) - \mathfrak{I}\| = o(1)$. Then we have

$$\begin{aligned} &\|\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,0}\| = \|\mathfrak{I}^{-1} \tilde{U}(\boldsymbol{\beta}_{1,0})\|(1 + o_p(1)) \\ = \;& O_p(k_n^{1/2-b} + k_n^{2a+1} n^{-1/2} + k_n^{(a+1)/2}[h + 1/\sqrt{n\psi(h)}]), \end{aligned}$$

and the first part of Theorem 2 is finished.

We continue with the second part of the theorem. For $\widetilde{\boldsymbol{\theta}} = \sum_{j=1}^{k_n} r_j(\widetilde{\boldsymbol{\beta}}_1)\hat{v}_j$, we have

$$\begin{aligned} &\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ = \;& \{\sum_{j=1}^{k_n}(r_j(\widetilde{\boldsymbol{\beta}}_1) - r_j(\boldsymbol{\beta}_{1,0}))\hat{v}_j\} + \{\sum_{j=1}^{k_n} r_j(\boldsymbol{\beta}_{1,0})\hat{v}_j - \sum_{j=1}^{k_n}\langle \boldsymbol{\theta}_0, v_j\rangle v_j\} \\ + \;& \{-\sum_{j=k_n+1}^{\infty}\langle \boldsymbol{\theta}_0, v_j\rangle v_j\} + \sum_{j=1}^{k_n}(\hat{r}_j(\widetilde{\boldsymbol{\beta}}_1) - r_j(\widetilde{\boldsymbol{\beta}}_1))\hat{v}_j\} \\ \triangleq \;& A^{\mathbf{T2}} + B^{\mathbf{T2}} + C^{\mathbf{T2}} + D^{\mathbf{T2}}. \end{aligned}$$

By the proof of Theorem 1, $\|C^{\mathbf{T2}}\| = \|C^{\mathbf{T1}}\| = O_p(k_n^{1/2-b})$; $\|B^{\mathbf{T2}}\| = \|B^{\mathbf{T1}}\| = O_p(k_n^{5a/2+3/2}/\sqrt{n} + k_n^{a/2+1/2-b})$; and

$$\begin{aligned} &\|A^{\mathbf{T2}}\| \le \|A_1^{\mathbf{T2}} + A_2^{\mathbf{T2}} + A_3^{\mathbf{T2}} + A_4^{\mathbf{T2}}\|\|\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,0}\| \\ = \;& O_p(k_n^{(a+1)/2})O_p(k_n^{2a+1} n^{-1/2} + k_n^{1/2-b} + k_n^{(a+1)/2}[h + 1/\sqrt{n\psi(h)}]) \\ = \;& O_p(k_n^{5/2a+3/2} n^{-1/2} + k_n^{1+a/2-b} + k_n^{a+1}[h + 1/\sqrt{n\psi(h)}]), \end{aligned}$$

where $A_1^{\mathbf{T2}} = A_1^{\mathbf{T1}}, A_2^{\mathbf{T2}} = A_1^{\mathbf{T1}}, A_3^{\mathbf{T2}} = A_1^{\mathbf{T1}}$ and $A_4^{\mathbf{T2}} = A_1^{\mathbf{T1}}$, and $A_i^{\mathbf{T1}}$ are defined in Theorem 1, for $i = 1, 2, 3, 4$. So that we only need to calculate $\|D^{\mathbf{T2}}\|$. First we define

$$\tilde{\hat{r}}_j = \frac{\sum_{i=1}^n [\delta_i \tilde{M}_j(Z_i, W_i, \hat{v}_j) + (1-\delta_i)\hat{m}_{\tilde{M}_j,i,\gamma}(Z_i, W_i, \hat{v}_j)]}{n\hat{\lambda}_j},$$

and

$$\tilde{\hat{r}}_j^* = \frac{\sum_{i=1}^n [\delta_i \tilde{M}_j(Z_i, W_i, v_j) + (1-\delta_i)\hat{m}_{\tilde{M}_j,i,\gamma}(Z_i, W_i, v_j)]}{n\lambda_j},$$

where the definition of $\tilde{M}_j(Z_i, W_i, \hat{v}_j)$ can be found in the proof of Lemma 9. Then similar to the proof of Theorem 1, we have

$$\|D^{\mathbf{T2}}\| = \sqrt{\sum_{j=1}^{k_n}[\hat{r}_j(\widetilde{\boldsymbol{\beta}}_1) - r_j(\widetilde{\boldsymbol{\beta}}_1)]^2} = \sqrt{\sum_{j=1}^{k_n}[\hat{r}_j(\boldsymbol{\beta}_{1,0}) - r_j(\boldsymbol{\beta}_{1,0}) + (\tilde{\hat{r}}_j - \tilde{r}_j)(\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,0})]^2}$$

$$\leq \sqrt{\sum_{j=1}^{k_n}[\hat{r}_j(\boldsymbol{\beta}_{1,0}) - r_j(\boldsymbol{\beta}_{1,0})]^2} + \sqrt{\sum_{j=1}^{k_n}[(\tilde{\hat{r}}_j - \tilde{r}_j)(\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,0})]^2} \triangleq D_1^{\mathbf{T2}} + D_2^{\mathbf{T2}},$$

By the conclusions of Lemma 12, (ii), Lemma 12, (iii), Lemma 6, (iii), and the Cauchy's inequality, we have

$$\begin{aligned}
D_1^{\mathbf{T2}} &= \sqrt{\sum_{j=1}^{k_n}[(\hat{r}_j - \hat{r}_j^*) + (\hat{r}_j^* - r_j^*) + (r_j^* - r_j)]^2} \\
&\leq \sqrt{\sum_{j=1}^{k_n}(\hat{r}_j - \hat{r}_j^*)^2} + \sqrt{\sum_{j=1}^{k_n}(\hat{r}_j^* - r_j^*)^2} + \sqrt{\sum_{j=1}^{k_n}(r_j^* - r_j)^2} \\
&\leq \sqrt{\sum_{j=1}^{k_n}\lambda_j(\hat{r}_j - \hat{r}_j^*)^2 \sum_{j=1}^{k_n}\frac{1}{\lambda_j}} + \sqrt{\sum_{j=1}^{k_n}\lambda_j^2(\hat{r}_j^* - r_j^*)^2 \sum_{j=1}^{k_n}\frac{1}{\lambda_j^2}} + \sqrt{\sum_{j=1}^{k_n}\lambda_j(r_j^* - r_j)^2 \sum_{j=1}^{k_n}\frac{1}{\lambda_j}} \\
&= \sqrt{O_p(k_n^{4a+2}n^{-1} + k_n^{-2b}) \times k_n^{a+1}} + \sqrt{O_p(h^2 + \frac{1}{n\psi(h)}) \times k_n^{2a+1}} \\
&+ \sqrt{O_p(k_n^{4a+2}n^{-1} + k_n^{-2b}) \times k_n^{a+1}} \\
&= O_p(k_n^{5a/2+3/2}n^{-1/2} + k_n^{1+a/2-b} + k_n^{a+1/2}[h + \frac{1}{\sqrt{n\psi(h)}}]).
\end{aligned}$$

Similar to the calculation of $D_1^{\mathbf{T2}}$, we have

$$\sqrt{\sum_{j=1}^{k_n}[(\tilde{\hat{r}}_j - \tilde{r}_j)]^2} = O_p(k_n^{5a/2+3/2}n^{-1/2} + k_n^{1+a/2-b} + k_n^{a+1/2}[h + \frac{1}{\sqrt{n\psi(h)}}]).$$

Note that $D_2^{\mathbf{T2}} = \sqrt{\sum_{j=1}^{k_n}[(\tilde{\hat{r}}_j - \tilde{r}_j)]^2}\|\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,0}\|$, which implies $D_2^{\mathbf{T2}} = o_p(D_1^{\mathbf{T2}})$.

Finally, we get the conclusion.

$$\begin{aligned} & \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \leq \|A^{\mathbf{T2}}\| + \|B^{\mathbf{T2}}\| + \|C^{\mathbf{T2}}\| + \|D^{\mathbf{T2}}\| \\ = \ & O_p(k_n^{5a/2+3/2}n^{-1/2} + k_n^{1+a/2-b} + k_n^{a+1}[h + \frac{1}{\sqrt{n\psi(h)}}]). \end{aligned}$$

□

Proof of Theorem 3

First of all, minimizing the following expression

$$(\mathbf{Y} - Z^*\boldsymbol{r} - \mathbf{W}\boldsymbol{\beta})^T\Sigma(\mathbf{Y} - Z^*\boldsymbol{r} - \mathbf{W}\boldsymbol{\beta}) + (Z^*\boldsymbol{r})^T(I_n - \Sigma)Z^*\boldsymbol{r},$$

where $\Sigma = \left\{D + \mathrm{diag}\left[\Xi(I_n - D)\mathbf{1}_n\right]\right\}$, is equivalent to solving the following eqnarray

$$\begin{cases} \frac{1}{n}Z^{*T}\left\{D + \mathrm{diag}\left[\Xi(I_n - D)\mathbf{1}_n\right]\right\}(\mathbf{Y} - \mathbf{W}\boldsymbol{\beta}) - \frac{1}{n}Z^{*T}Z^*\boldsymbol{r} = 0, \\ \frac{1}{n}\mathbf{W}^T\left\{D + \mathrm{diag}\left[\Xi(I_n - D)\mathbf{1}_n\right]\right\}(\mathbf{Y} - Z^*\boldsymbol{r} - \mathbf{W}\boldsymbol{\beta}) = 0. \end{cases} \tag{S3.4}$$

From the definitions of Subsection 2.2.2 of the main text, $Z^* = \begin{pmatrix} \bar{Z}_1\bar{V}_{k_n} \\ \vdots \\ \bar{Z}_n\bar{V}_{k_n} \end{pmatrix}$, where

$\bar{Z}_i \triangleq (\bar{Z}_{i,1}, \bar{Z}_{i,2}, \cdots, \bar{Z}_{i,n}), i = 1, 2, \cdots, n,$ and

$$\Xi(I_n - D)\mathbf{1}_n = \begin{pmatrix} \sum_{i=1}^n (1-\delta_i) w_{1,i} \\ \sum_{i=1}^n (1-\delta_i) w_{2,i} \\ \vdots \\ \sum_{i=1}^n (1-\delta_i) w_{n,i} \end{pmatrix}.$$

Then we have the following equivalence.

$$\frac{1}{n} Z^{*T} \left\{ D + \mathrm{diag}\big[\Xi(I_n - D)\mathbf{1}_n\big] \right\} (\mathbf{Y} - \mathbf{W}\boldsymbol{\beta}) - \frac{1}{n} Z^{*T} Z^* \boldsymbol{r} = 0$$

$$\Longleftrightarrow$$

$$\frac{1}{n} \begin{pmatrix} \bar{Z}_1 \bar{V}_{k_n} \\ \vdots \\ \bar{Z}_n \bar{V}_{k_n} \end{pmatrix}^T \mathrm{diag}\left\{ \delta_1 + \sum_{i=1}^n (1-\delta_i) w_{1,i}, \cdots, \delta_n + \sum_{i=1}^n (1-\delta_i) w_{n,i} \right\}$$

$$\times \begin{pmatrix} Y_1 - W_1 \boldsymbol{\beta} \\ \vdots \\ Y_n - W_n \boldsymbol{\beta} \end{pmatrix} - \hat{\Lambda} \boldsymbol{r} = 0$$

$$\Longleftrightarrow$$

$$\frac{1}{n} \sum_{i=1}^n (\bar{Z}_i \bar{V}_{k_n})^T [\delta_i + \sum_{k=1}^n (1-\delta_k) w_{i,k}] (Y_i - W_i \boldsymbol{\beta}) - \sum_{i=1}^n [\delta_i + \sum_{k=1}^n (1-\delta_k) w_{i,k}] \frac{\hat{\Lambda}}{n} \boldsymbol{r} = 0$$

$$\Longleftrightarrow$$

$$\frac{1}{n} \sum_{i=1}^n (\bar{Z}_i \bar{V}_{k_n})^T \delta_i (Y_i - W_i \boldsymbol{\beta}) + \sum_{i,k \le n} (\bar{Z}_i \bar{V}_{k_n})^T (1-\delta_k) w_{i,k} (Y_i - W_i \boldsymbol{\beta})$$

$$- \sum_{i=1}^n [\delta_i + \sum_{k=1}^n (1-\delta_k) w_{i,k}] \frac{\hat{\Lambda}}{n} \boldsymbol{r} = 0$$

$$\Longleftrightarrow$$

$$\frac{1}{n}\sum_{i=1}^{n}(\bar{Z}_i\bar{V}_{k_n})^T\delta_i(Y_i - W_i\boldsymbol{\beta}) + \sum_{i,l\le n}(\bar{Z}_l\bar{V}_{k_n})^T(1-\delta_i)w_{l,i}(Y_l - W_l\boldsymbol{\beta})$$
$$-\sum_{i=1}^{n}[\delta_i + \sum_{k=1}^{n}(1-\delta_k)w_{i,k}]\frac{\hat{\Lambda}}{n}\boldsymbol{r} = 0$$
$$\Longleftrightarrow$$
$$\frac{1}{n}\sum_{i=1}^{n}[\delta_i(\bar{Z}_i\bar{V}_{k_n})^T(Y_i - W_i\boldsymbol{\beta}) + (1-\delta_i)\sum_{l=1}^{n}w_{l,i}(\bar{Z}_l\bar{V}_{k_n})^T(Y_l - W_l\boldsymbol{\beta})]$$
$$-\sum_{i=1}^{n}[\delta_i + \sum_{l=1}^{n}(1-\delta_i)w_{l,i}]\frac{\hat{\Lambda}}{n}\boldsymbol{r} = 0$$
$$\Longleftrightarrow$$
$$\frac{1}{n}\sum_{i=1}^{n}\{\delta_i[(\bar{Z}_i\bar{V}_{k_n})^T(Y_i - W_i\boldsymbol{\beta}) - \begin{pmatrix}\hat{\lambda}_1\boldsymbol{r}_1\\ \vdots \\ \hat{\lambda}_{k_n}\boldsymbol{r}_{k_n}\end{pmatrix}] + (1-\delta_i)\times$$
$$\sum_{l=1}^{n}w_{l,i}[(\bar{Z}_l\bar{V}_{k_n})^T(Y_l - W_l\boldsymbol{\beta}) - \begin{pmatrix}\hat{\lambda}_1\boldsymbol{r}_1\\ \vdots \\ \hat{\lambda}_{k_n}\boldsymbol{r}_{k_n}\end{pmatrix}]\} = 0$$
$$\Longleftrightarrow$$
$$\sum_{i=1}^{n}[\delta_i\psi_1(Y_i, Z_i, W_i; \boldsymbol{r}, \boldsymbol{\beta}_1) + (1-\delta_i)\hat{m}_{\psi_1,i,\gamma}(Y_i, Z_i, W_i; \boldsymbol{r}, \boldsymbol{\beta}_1)] = 0,$$

where

$$\psi_1(Y_i, Z_i, W_i; \boldsymbol{r}, \boldsymbol{\beta}_1) = (\bar{Z}_i\bar{V}_{k_n})^T(Y_i - W_i\boldsymbol{\beta}) - \begin{pmatrix}\hat{\lambda}_1\boldsymbol{r}_1\\ \vdots \\ \hat{\lambda}_{k_n}\boldsymbol{r}_{k_n}\end{pmatrix},$$

is a discretized form of

$$\begin{pmatrix} \langle Z_i, \hat{v}_1\rangle(Y_i - W_i\boldsymbol{\beta}) - \hat{\lambda}_1 r_1 \\ \vdots \\ \langle Z_i, \hat{v}_{k_n}\rangle(Y_i - W_i\boldsymbol{\beta}) - \hat{\lambda}_{k_n} r_{k_n} \end{pmatrix}. \tag{S3.5}$$

Similarly, the second equation of (S3.4) is equivalent to

$$\sum_{i=1}^{n}[\delta_i\psi_2(Y_i, Z_i, W_i; \boldsymbol{r}, \boldsymbol{\beta}_1) + (1-\delta_i)\hat{m}_{\psi_2,i,\gamma}(Y_i, Z_i, W_i; \boldsymbol{r}, \boldsymbol{\beta}_1)] = 0,$$

where $\psi_2(Y_i, Z_i, W_i; \boldsymbol{r}, \boldsymbol{\beta}_1) = [Y_i - \boldsymbol{\beta}_1^T W_i - \bar{Z}_i\bar{V}_{k_n}\boldsymbol{r}]W_i$ is a discretized form of

$$W_i^T[Y_i - \boldsymbol{\beta}_1^T W_i - \sum_{j=1}^{k_n} r_j\langle Z_i, v_j\rangle]. \tag{S3.6}$$

Compare (S3.5) and (S3.6) with (2.5), and we get the conclusion in Theorem 3.

□

Proof of Corollary 1

The proof is the same as that of Theorem 2 if we use

$$\psi(h) = \inf_{(z,x)\in\mathbb{H}_0}\psi_{z,x}(h); \psi_{z,x}(h) = \Pr[(Z, W) \in \{(\tilde{z}, \tilde{x})|\ w_0\|\tilde{z}-z\| + (1-w_0)\|\tilde{x}-x\| \leq h\}],$$

to replace of the corresponding notation in Theorem 2.

□

References

Crambes, C., and Andr, M. (2013). Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli* **19.5B**, 2627–2651.

Ferraty, F., and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* Springer Series in Statistics. Springer, New York.

Hall, P., and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society, Series B* **68**, 109–126.

Hall, P., and Hosseini-Nasab, M. (2009). Theory for high-order bounds in functional principal components analysis. *Mathematical Proceedings of the Cambridge Philosophical Society* **146**(1).

Stewart, G. W. (1969). On the continuity of the generalized inverse. *SIAM Journal on Applied Mathematics* **17**(1), 33–45.

Tengfei Li and Hongtu Zhu

Department of Biostatistics,

University of Texas MD Anderson Cancer Center

Houston TX 77030, USA

E-mail: tengfeili2006@gmail.com, hzhu5@mdanderson.org

Ibrahim, J. G. and Hongtu Zhu

Department of Biostatistics,

University of North Carolina at Chapel Hill

Chapel Hill, NC 27599, USA

E-mail: jibrahim@email.unc.edu and hzhu@bios.unc.edu

Fengchang Xie

School of Mathematical Sciences,

Nanjing Normal University,

Nanjing 210023, CHINA

E-mail: fcxie@njnu.edu.cn

Xiangnan Feng

Department of Statistics,

The Chinese University of Hong Kong

Hong Kong, China

E-mail: fengxiangnan123@gmail.com