PARTIAL FUNCTIONAL PARTIALLY LINEAR SINGLE-INDEX MODELS

Qingguo Tang[1], Linglong Kong[2], David Ruppert[3], and Rohana J. Karunamuni[2]

[1] *Nanjing University of Science and Technology*

[2] *University of Alberta and* [3]*Cornell University*

Supplementary Material

This part contains the proofs of Theorems 3.1 to 3.5, which depend on a number of preliminary lemmas.

Proofs of the theorems

We give the proofs of the main results here. Let $C > 0$ denote a generic constant of which the value may change from line to line. For a matrix $A = (a_{ij})$, set $\|A\|_\infty = \max_i \sum_j |a_{ij}|$ and $|A|_\infty = \max_{i,j} |a_{ij}|$. For a vector $v = (v_1, \ldots, v_k)^T$, set $\|v\|_\infty = \sum_{j=1}^k |v_j|$ and $|v|_\infty = \max_{1\le j\le k} |v_j|$. We write $Y_i = Y_i^* + \varepsilon_i$ with $Y_i^* = \int_{\mathcal{T}} a(t)X_i(t)dt + W_i^T\boldsymbol{\alpha}_0 + g(Z_i^T\boldsymbol{\beta}_0)$. Denote $\check{Y}_i = Y_i^* - \frac{1}{n}\sum_{l=1}^n Y_l^*\tilde{\xi}_{il}$, $\tilde{\varepsilon}_i = \varepsilon_i - \frac{1}{n}\sum_{l=1}^n \varepsilon_l\tilde{\xi}_{il}$ and $\check{\boldsymbol{Y}} = (\check{Y}_1, \ldots, \check{Y}_n)^T$, $\tilde{\boldsymbol{\varepsilon}} = (\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n)^T$. Then $\tilde{Y}_i = \check{Y}_i + \tilde{\varepsilon}_i$ and $\tilde{\boldsymbol{Y}} = \check{\boldsymbol{Y}} + \tilde{\boldsymbol{\varepsilon}}$. Define $\boldsymbol{P}(\boldsymbol{\beta}) = I_n - \tilde{\boldsymbol{B}}(\boldsymbol{\beta})(\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta})\tilde{\boldsymbol{B}}(\boldsymbol{\beta}))^{-1}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta})$, where I_n is the $n \times n$ identity matrix. By (3.5), (2.9) and (2.10), we have

$$\tilde{G}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{n}[(\check{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha})^T\boldsymbol{P}(\boldsymbol{\beta})(\check{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}) + 2(\check{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha})^T\boldsymbol{P}(\boldsymbol{\beta})\tilde{\varepsilon} + \tilde{\varepsilon}^T\boldsymbol{P}(\boldsymbol{\beta})\tilde{\varepsilon}]. \tag{A.1}$$

Lemma A.1. Suppose that Assumptions 1 to 4 and 5' hold. Then

$$\tfrac{1}{n}(\check{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha})^T(\check{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}) = \rho(\boldsymbol{\alpha}) + o_p(1),$$

where $\rho(\boldsymbol{\alpha}) = (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T E(\check{V}\check{V}^T)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) - 2\boldsymbol{b}_0^T E[\boldsymbol{B}_{\boldsymbol{\beta}_0}(Z^T\boldsymbol{\beta}_0)\check{V}^T](\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + \boldsymbol{b}_0^T\Gamma(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0)\boldsymbol{b}_0$, and $o_p(1)$ holds uniformly for $\boldsymbol{\alpha}$ in any bounded neighborhood of $\boldsymbol{\alpha}_0$.

Proof. Define $\check{\xi}_{il} = \sum_{j=1}^{m} \frac{\xi_{lj}\xi_{ij}}{\lambda_j}$, $\check{Y}_{i1} = Y_i^* - \frac{1}{n}\sum_{l=1}^{n} Y_l^* \check{\xi}_{il}$ and $\check{Y}_{i2} = \frac{1}{n}\sum_{l=1}^{n} Y_l^*(\tilde{\xi}_{il} - \check{\xi}_{il})$. Then $\check{Y}_i = \check{Y}_{i1} - \check{Y}_{i2}$ and

$$\frac{1}{n}\check{\boldsymbol{Y}}^T\check{\boldsymbol{Y}} = \frac{1}{n}\sum_{i=1}^{n}(\check{Y}_{i1}^2 - 2\check{Y}_{i1}\check{Y}_{i2} + \check{Y}_{i2}^2). \tag{A.2}$$

Denote $\check{Y}_{i21} = \sum_{j=1}^{m} \frac{1}{\lambda_j}[\frac{1}{n}\sum_{l=1}^{n} Y_l^*(\hat{\xi}_{lj} - \xi_{lj})]\xi_{ij}$, $\check{Y}_{i22} = \sum_{j=1}^{m}(\frac{1}{\hat{\lambda}_j} - \frac{1}{\lambda_j})(\frac{1}{n}\sum_{l=1}^{n} Y_l^*\hat{\xi}_{lj})\xi_{ij}$ and $\check{Y}_{i23} = \sum_{j=1}^{m} \frac{1}{\hat{\lambda}_j}(\frac{1}{n}\sum_{l=1}^{n} Y_l^*\hat{\xi}_{lj})(\hat{\xi}_{ij} - \xi_{ij})$. Then we have

$$\check{Y}_{i2}^2 \le 3(\check{Y}_{i21}^2 + \check{Y}_{i22}^2 + \check{Y}_{i23}^2). \tag{A.3}$$

From Lemma 5.1 of Hall and Horowitz (2007) it follows that

$$\hat{\xi}_{lj} - \xi_{lj} = \sum_{k\neq j} \frac{\xi_{lk}}{\hat{\lambda}_j - \lambda_k}\int \Delta\hat{\phi}_j\phi_k + \xi_{lj}\int(\hat{\phi}_j - \phi_j)\phi_j, \tag{A.4}$$

where $\Delta = \hat{K} - K$. Then we obtain

$$\begin{aligned}[\tfrac{1}{n}\textstyle\sum_{l=1}^{n} Y_l^*(\hat{\xi}_{lj} - \xi_{lj})]^2 &\le 2(\textstyle\sum_{k\neq j} \frac{\vec{\xi}_k}{\hat{\lambda}_j - \lambda_k}\int \Delta\hat{\phi}_j\phi_k)^2 + 2(\vec{\xi}_j\int(\hat{\phi}_j - \phi_j)\phi_j)^2 \\ &\le 2[\textstyle\sum_{k\neq j} \frac{\vec{\xi}_k^2}{(\hat{\lambda}_j - \lambda_k)^2}][\sum_{k=1}^{\infty}(\int \Delta\hat{\phi}_j\phi_k)^2] + 2\vec{\xi}_j^2(\int(\hat{\phi}_j - \phi_j)\phi_j)^2,\end{aligned}$$

where $\vec{\xi}_j = \frac{1}{n}\sum_{l=1}^{n} Y_l^*\xi_{lj}$. Lemma 1 of Cardot et al. (2007) yields that

$$|\lambda_j - \lambda_k| \ge \lambda_j - \lambda_{j+1} \ge \lambda_m - \lambda_{m+1} \ge \lambda_m/(m+1) \ge \lambda_m/(2m)$$

uniformly for $1 \le j \le m$. From (5.2) of Hall and Horowitz (2007) we have $\sup_{j\ge1}|\hat{\lambda}_j - \lambda_j| \le |||\Delta||| = O_p(n^{-1/2})$ and

$$(\textstyle\int(\hat{\phi}_j - \phi_j)\phi_j)^2 \le \|\hat{\phi}_j - \phi_j\|^2 \le C\frac{|||\Delta|||^2}{(\lambda_j - \lambda_{j+1})^2} \le C|||\Delta|||^2\lambda_j^{-2}j^2, \tag{A.5}$$

where $|||\Delta||| = (\int_{\mathcal{T}}\int_{\mathcal{T}} \Delta^2(s,t)dsdt)^{1/2}$. Using Parseval's identity, we obtain

$$\sum_{k=1}^{\infty}(\int \Delta\hat{\phi}_j\phi_k)^2 = \int(\int \Delta\hat{\phi}_j)^2 \le |||\Delta|||^2 = O_p(n^{-1}).$$

Assumption 5' implies that $|\hat{\lambda}_j - \lambda_j| = o_p(\lambda_m/m)$. Consequently, $\sum_{k\neq j} \frac{\vec{\xi}_k^2}{(\hat{\lambda}_j - \lambda_k)^2} = \sum_{k\neq j} \frac{\vec{\xi}_k^2}{(\lambda_j - \lambda_k)^2}[1 + o_p(1)]$, where $o_p(1)$ holds uniformly for $1 \le j \le m$. Using Lemma 2 of Cardot et al. (2007) and the fact that $(\lambda_j-\lambda_k)^2\geq(\lambda_k-\lambda_{k+1})^2$, we deduce that

$$\sum_{k\neq j}\frac{1}{(\lambda_j-\lambda_k)^2}E(\vec{\xi}_k^2)$$

$$\leq C\sum_{k\neq j}\frac{1}{(\lambda_j-\lambda_k)^2}[n^{-1}\lambda_k+a_k^{*2}\lambda_k^2]$$

$$\leq C[\frac{1}{n(\lambda_j-\lambda_{j+1})}\sum_{k\neq j}\frac{\lambda_k}{|\lambda_j-\lambda_k|}+\sum_{k=1}^{j-1}\frac{\lambda_k^2a_k^{*2}}{(\lambda_k-\lambda_{k+1})^2}+\sum_{k=j+1}^{2j}\frac{j^2a_k^{*2}}{(k-j)^2}+\sum_{k=2j+1}^{\infty}\frac{\lambda_k^2a_k^{*2}}{(\lambda_j-\lambda_{2j})^2}]$$

$$\leq C(n^{-1}\lambda_j^{-1}j^2\log j+1).$$

where $a_k^*=a_k+\sum_{r=1}^q v_{rk}^*\alpha_{0r}$. Assumption 2 yields that

$$\sum_{j=1}^m\lambda_j^{-2}j^2\log j\leq m^{-2}\lambda_m^{-2}\sum_{j=1}^m j^4\log j\leq\lambda_m^{-2}m^3\log m$$

and $\sum_{j=1}^m\lambda_j^{-1}\leq\lambda_m^{-1}m$. Therefore,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\check{Y}_{i21}^2 &\leq(\sum_{j=1}^m\frac{1}{\lambda_j}[\frac{1}{n}\sum_{l=1}^n Y_l^*(\hat{\xi}_{lj}-\xi_{lj})]^2)(\sum_{j=1}^m\frac{1}{n\lambda_j}\sum_{i=1}^n\xi_{ij}^2)\\ &=O_p(n^{-2}\lambda_m^{-2}m^4\log m+n^{-1}\lambda_m^{-1}m^2).\end{aligned}\tag{A.6}$$

Decomposing $\frac{1}{n}\sum_{l=1}^n Y_l^*\hat{\xi}_{lj}=\vec{\xi}_j+\frac{1}{n}\sum_{l=1}^n Y_l^*(\hat{\xi}_{lj}-\xi_{lj})$ and using (A.6), we obtain

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\check{Y}_{i22}^2 &\leq C\sum_{j=1}^m\frac{(\hat{\lambda}_j-\lambda_j)^2}{\lambda_j^3}(\frac{1}{n}\sum_{l=1}^n Y_l^*\hat{\xi}_{lj})^2[1+o_p(1)](\sum_{j=1}^m\frac{1}{n\lambda_j}\sum_{i=1}^n\xi_{ij}^2)\\ &=O_p(n^{-1}\lambda_m^{-1}m+n^{-3}\lambda_m^{-4}m^4\log m+n^{-2}\lambda_m^{-3}m^2).\end{aligned}\tag{A.7}$$

By (A.10) of Tang (2015), it holds that

$$\|\hat{\phi}_j-\phi_j\|^2=O_p(n^{-1}j^2\log j)\tag{A.8}$$

uniformly for $1\leq j\leq m$. Using (A.7) and (A.8), we obtain

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\check{Y}_{i23}^2 &\leq(\sum_{j=1}^m\frac{1}{\hat{\lambda}^2}(\frac{1}{n}\sum_{l=1}^n Y_l^*\hat{\xi}_{lj})^2)(\frac{1}{n}\sum_{i=1}^n\|X_i\|^2)(\sum_{j=1}^m\|\hat{\phi}_j-\phi_j\|^2)\\ &=O_p((n^{-1}m^3+n^{-3}\lambda_m^{-3}m^6\log m+n^{-2}\lambda_m^{-2}m^4)\log m).\end{aligned}\tag{A.9}$$

Then by (A.3), (A.6), (A.7), (A.9) and Assumption 5', we conclude that

$$\frac{1}{n}\sum_{i=1}^n\check{Y}_{i2}^2=O_p(n^{-2}\lambda_m^{-2}m^4\log m+n^{-1}\lambda_m^{-1}m^2)=o_p(h_0^2).\tag{A.10}$$

Define $\xi_j^* = \frac{1}{n}\sum_{l=1}^n \lambda_j^{-1/2}\xi_{lj}Y_l^*$. Since $E[\max_{1\le j\le m}(\xi_j^* - E(\xi_j^*))^2] \le \frac{1}{n}\sum_{j=1}^m \lambda_j^{-1}E(\xi_j Y^*)^2 \le Cn^{-1}m$, we then have $\max_{1\le j\le m}|\xi_j^* - E(\xi_j^*)| = O_p(n^{-1/2}m^{1/2})$. Hence, we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \check{Y}_{i1}^2 &= \frac{1}{n}\sum_{i=1}^n Y_i^{*2} - 2\sum_{j=1}^m \xi_j^{*2} + \sum_{j=1}^m \frac{\xi_j^{*2}}{n\lambda_j}(\textstyle\sum_{i=1}^n \xi_{ij}^2) + \sum_{j\ne j'}\xi_j^*\xi_{j'}^*\bar{\xi}_{jj'}\\ &= \sum_{j=1}^\infty (a_j + \textstyle\sum_{r=1}^q v_{rj}^*\alpha_{0r})^2\lambda_j + E(\check{V}^T\boldsymbol{\alpha}_0 + g(Z^T\boldsymbol{\beta}_0))^2\\ &\quad -2\sum_{j=1}^m (a_j + \textstyle\sum_{r=1}^q v_{rj}^*\alpha_{0r})^2\lambda_j + \sum_{j=1}^m (a_j + \sum_{r=1}^q v_{rj}^*\alpha_{0r})^2\lambda_j + o_p(1)\\ &= E(\check{V}^T\boldsymbol{\alpha}_0 + g(Z^T\boldsymbol{\beta}_0))^2 + o_p(1),\end{aligned} \tag{A.11}$$

where $\bar{\xi}_{jj'} = \frac{1}{n(\lambda_j\lambda_{j'})^{1/2}}\sum_{i=1}^n \xi_{ij}\xi_{ij'}$. Combining (A.2), (A.10), (A.11) and (3.1), we conclude that

$$\frac{1}{n}\check{\boldsymbol{Y}}^T\check{\boldsymbol{Y}} = \boldsymbol{\alpha}_0^T E(\check{V}\check{V}^T)\boldsymbol{\alpha}_0 + 2\boldsymbol{b}_0^T E[\boldsymbol{B}_{\boldsymbol{\beta}_0}(Z^T\boldsymbol{\beta}_0)\check{V}^T]\boldsymbol{\alpha}_0 + \boldsymbol{b}_0^T\Gamma(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0)\boldsymbol{b}_0 + o_p(1). \tag{A.12}$$

Similar to the proof of (A.12), we obtain that

$$\frac{1}{n}\tilde{\boldsymbol{W}}^T\tilde{\boldsymbol{W}} = E(\check{V}\check{V}^T) + o_p(1), \quad \frac{1}{n}\tilde{\boldsymbol{Y}}^T\tilde{\boldsymbol{W}} = \boldsymbol{\alpha}_0 E(\check{V}^T\check{V}) + \boldsymbol{b}_0^T E[\boldsymbol{B}_{\boldsymbol{\beta}_0}(Z^T\boldsymbol{\beta}_0)\check{V}] + o_p(1).$$

Now Lemma A.1 follows from (A.12) and the preceding expression.

Lemma A.2. Under Assumptions 1, 4 and 5', it holds that

$$\sup_{\boldsymbol{\beta}\in\Theta_{\rho_0}}\max_{1\le j\le m}\max_{1\le k\le K_{\boldsymbol{\beta}}}\lambda_j^{-\frac{1}{2}}|\frac{1}{n}\sum_{i=1}^n \xi_{ij}B_{k\boldsymbol{\beta}}^{(r)}(Z_i^T\boldsymbol{\beta})| = o_p(n^{-\frac{1}{2}}h_0^{\frac{1}{4}-r}\log n),$$

$$\sup_{\boldsymbol{\beta}\in\Theta_{\rho_0}}\max_{k,k'}|\frac{1}{n}\sum_{i=1}^n B_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})B_{k'\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta}) - E[B_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})B_{k'\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})]| = o_p(n^{-\frac{1}{2}}h_0^{\frac{1}{2}}\log n),$$

$$\sup_{\boldsymbol{\beta}\in\Theta_{\rho_0}}\max_{k,k'}|\frac{1}{n}\sum_{i=1}^n B'_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})B'_{k'\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta}) - E[B'_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})B'_{k'\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})]| = o_p(n^{-\frac{1}{2}}h_0^{-\frac{3}{2}}\log n),$$

and

$$\sup_{\boldsymbol{\beta}\in\Theta_{\rho_0}}\max_{k,k'}|\frac{1}{n}\sum_{i=1}^n B_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})B''_{k'\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta}) - E[B_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})B''_{k'\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})]| = o_p(n^{-\frac{1}{2}}h_0^{-\frac{3}{2}}\log n)$$

for $r = 0, 1, 2$.

Proof. We give only the proof for the first step with $r = 2$, as the first step with $r = 0, 1$ and the other steps follow from similar arguments. Define $\eta_{jki}(Z_i^T\boldsymbol{\beta}) = \lambda_j^{-1/2}\xi_{ij}B''_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})$. Applying Assumptions 1 and Lemma E.1 of Kato (2012), we have $\max_{1\leq j\leq m, 1\leq i\leq n}|\lambda_j^{-1/2}\xi_{ij}| = O_p((mn)^{1/4})$. Hence, by Assumption 5', for any $\varepsilon > 0$ and $\epsilon > 0$, there exists a positive constant $\tilde{C}_1$ such that

$$P\{\max_{1\leq j\leq m, 1\leq i\leq n}|\lambda_j^{-1/2}\xi_{ij}| \geq \tilde{C}_1 n^{1/2}h_0^{1/4}(\log n)^{-1}\} < \epsilon/4. \tag{A.13}$$

Using Assumptions 1 and the fact that $|B''_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})| \leq Ch_0^{-2}$, we obtain

$$\begin{aligned}
&|E[\lambda_j^{-\frac{1}{2}}\xi_{ij}B''_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})I_{\{|\lambda_j^{-\frac{1}{2}}\xi_{ij}|\geq\tilde{C}_1 n^{\frac{1}{2}}h_0^{\frac{1}{4}}(\log n)^{-1}\}}]| \\
&\leq Cn^{-\frac{3}{2}}h_0^{-\frac{11}{4}}(\log n)^3 E[\lambda_j^{-\frac{1}{2}}\xi_{ij}]^4 < \varepsilon n^{-\frac{1}{2}}h_0^{-\frac{7}{4}}\log n/2.
\end{aligned}$$

Denote

$$\begin{aligned}
\tilde{\eta}_{jki}(Z_i^T\boldsymbol{\beta}) \quad &= \lambda_j^{-\frac{1}{2}}\xi_{ij}B''_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})I_{\{|\lambda_j^{-\frac{1}{2}}\xi_{ij}|<\tilde{C}_1 n^{\frac{1}{2}}h_0^{\frac{1}{4}}(\log n)^{-1}\}} \\
&-E[\lambda_j^{-\frac{1}{2}}\xi_{ij}B''_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})I_{\{|\lambda_j^{-\frac{1}{2}}\xi_{ij}|<\tilde{C}_1 n^{\frac{1}{2}}h_0^{\frac{1}{4}}(\log n)^{-1}\}}].
\end{aligned}$$

Then we have

$$\begin{aligned}
&P\{\sup_{\boldsymbol{\beta}\in\Theta_{\rho_0}}\max_{j,k}|\tfrac{1}{n}\textstyle\sum_{i=1}^n \eta_{jki}(Z_i^T\boldsymbol{\beta})| \geq \varepsilon n^{-\frac{1}{2}}h_0^{-\frac{7}{4}}\log n\} \\
&\leq P\{\max_{j,i}|\lambda_j^{-\frac{1}{2}}\xi_{ij}| \geq \tilde{C}_1 n^{\frac{1}{2}}h_0^{\frac{1}{4}}(\log n)^{-1}\} \\
&+P\{\sup_{\boldsymbol{\beta}\in\Theta_{\rho_0}}\max_{j,k}|\tfrac{1}{n}\textstyle\sum_{i=1}^n \tilde{\eta}_{jki}(Z_i^T\boldsymbol{\beta})| \geq \varepsilon n^{-\frac{1}{2}}h_0^{-\frac{7}{4}}\log n/2\}.
\end{aligned} \tag{A.14}$$

Using the fact that $|B''_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})| \leq Ch_0^{-2}$, again we obtain

$$|\tilde{\eta}_{jki}(Z_i^T\boldsymbol{\beta})| \leq Cn^{\frac{1}{2}}h_0^{-\frac{7}{4}}(\log n)^{-1}. \tag{A.15}$$

From Assumption 1, it follows that

$$\sum_{i=1}^n E(\tilde{\eta}_{jki}^2(Z_i^T\boldsymbol{\beta})) \leq Cn\lambda_j^{-1}(E[B''^4_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})]E(\xi_j^4))^{1/2} \leq Cnh_0^{-7/2}. \tag{A.16}$$

For $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{1d})^T \in \Theta_{\rho_0}$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \ldots, \beta_{2d})^T \in \Theta_{\rho_0}$, define $|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1| = \max_{1 \le r \le d-1} |\beta_{2r} - \beta_{1r}|$. Since $\sum_{j=1}^m \frac{1}{n} \sum_{i=1}^n \lambda_j^{-\frac{1}{2}} |\xi_{ij}| = O_p(m)$, then there exists a positive $\tilde{C}_2$ such that

$$P\{\sum_{j=1}^m \frac{1}{n} \sum_{i=1}^n \lambda_j^{-\frac{1}{2}} |\xi_{ij}| \ge \tilde{C}_2 m\} < \epsilon/4. \tag{A.17}$$

From (2.6), for all $\boldsymbol{\beta} \in \Theta_{\rho_0}$, the total of different $B_{k\boldsymbol{\beta}}(u)$ is not more than $(s+1)k_n$. Let Θ_{ρ_0} be divided into N disjoint parts $\Theta_{\rho_0 1}, \cdots, \Theta_{\rho_0 N}$ such that for any $\boldsymbol{\beta} \in \Theta_{\rho_0 l}, 1 \le l \le N$ and any $1 \le j \le m, 1 \le k \le (s+1)k_n$, when $\sum_{j=1}^m \frac{1}{n} \sum_{i=1}^n \lambda_j^{-\frac{1}{2}} |\xi_{ij}| < \tilde{C}_2 m$,

$$\begin{aligned}
&\sup_{\boldsymbol{\beta} \in \Theta_{\rho_0 l}} |\tfrac{1}{n} \textstyle\sum_{i=1}^n \tilde{\eta}_{jki}(Z_i^T \boldsymbol{\beta}) - \tfrac{1}{n} \sum_{i=1}^n \tilde{\eta}_{jki}(Z_i^T \boldsymbol{\beta}_l)| \\
&\le \sup_{\boldsymbol{\beta} \in \Theta_{\rho_0 l}} \lambda_j^{-\frac{1}{2}} \Big(\tfrac{1}{n} \textstyle\sum_{i=1}^n |\xi_{ij}| |B''_{k\boldsymbol{\beta}}(Z_i^T \boldsymbol{\beta}) - B''_{k\boldsymbol{\beta}}(Z_i^T \boldsymbol{\beta}_l)| + E(|\xi_{ij}| |B''_{k\boldsymbol{\beta}}(Z_i^T \boldsymbol{\beta}) - B''_{k\boldsymbol{\beta}}(Z_i^T \boldsymbol{\beta}_l)|) \Big) \\
&\le \sup_{\boldsymbol{\beta} \in \Theta_{\rho_0 l}} C h_0^{-3} \textstyle\sum_{j=1}^m \Big(\tfrac{1}{n} \sum_{i=1}^n \lambda_j^{-\frac{1}{2}} |\xi_{ij}| + E(\lambda_j^{-\frac{1}{2}} |\xi_{ij}|) \Big) |\boldsymbol{\beta} - \boldsymbol{\beta}_l| \\
&\le C m h_0^{-3} |\boldsymbol{\beta} - \boldsymbol{\beta}_l| < \varepsilon n^{-\frac{1}{2}} h_0^{-\frac{7}{4}} \log n / 4.
\end{aligned}$$

This can be done with $N = C(mn^{1/2}/(\varepsilon h_0^{5/4} \log n))^{d-1}$. Using Bernstein inequality and (A.15), (A.16) and Assumption 5', for sufficiently large n, it follows that

$$\begin{aligned}
&P\Big(\sup_{\boldsymbol{\beta} \in \Theta_{\rho_0}} \max_{j,k} |\tfrac{1}{n} \textstyle\sum_{i=1}^n \tilde{\eta}_{jki}(Z_i^T \boldsymbol{\beta})| \ge \varepsilon n^{-\frac{1}{2}} h_0^{-\frac{7}{4}} \log n / 2, \sum_{j=1}^m \tfrac{1}{n} \sum_{i=1}^n \lambda_j^{-\frac{1}{2}} |\xi_{ij}| < \tilde{C}_2 m \Big) \\
&\le P\Big(\cup_{l=1}^N \{ \max_{j,k} |\tfrac{1}{n} \textstyle\sum_{i=1}^n \tilde{\eta}_{jki}(Z_i^T \boldsymbol{\beta}_l)| \ge \varepsilon n^{-\frac{1}{2}} h_0^{-\frac{7}{4}} \log n / 4 \} \Big) \\
&\le C m k_n N \exp\Big\{ - \frac{\varepsilon^2 n h_0^{-\frac{7}{2}} (\log n)^2}{32 C n h_0^{-7/2} + 4 C n^{\frac{1}{2}} h_0^{-\frac{7}{4}} (\log n)^{-1} \varepsilon n^{\frac{1}{2}} h_0^{-\frac{7}{4}} \log n} \Big\} < \epsilon/2.
\end{aligned}$$

Now Lemma A.2 follows from (A.13), (A.14), (A.17) and the preceding inequality.

Lemma A.3. Assume that Assumptions 1, 2, 4 and 5' hold. Then it holds that

$$\frac{1}{n} \tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}) \tilde{\boldsymbol{B}}(\boldsymbol{\beta}) = \Gamma(\boldsymbol{\beta}, \boldsymbol{\beta}) + o_p(h_0^2),$$

where $o_p(h_0^2)$ holds uniformly for $1 \le k, k' \le K_{\boldsymbol{\beta}}$ and $\boldsymbol{\beta} \in \Theta_{\rho_0}$.

Proof. Define

$$\tilde{B}_{k\boldsymbol{\beta}1}(Z_i^T \boldsymbol{\beta}) = B_{k\boldsymbol{\beta}}(Z_i^T \boldsymbol{\beta}) - \frac{1}{n} \sum_{l=1}^n B_{k\boldsymbol{\beta}}(Z_l^T \boldsymbol{\beta}) \check{\xi}_{il}, \quad \tilde{B}_{k\boldsymbol{\beta}2}(Z_i^T \boldsymbol{\beta}) = \frac{1}{n} \sum_{l=1}^n B_{k\boldsymbol{\beta}}(Z_l^T \boldsymbol{\beta}) (\tilde{\xi}_{il} - \check{\xi}_{il}).$$

We decompose the (k,k')th element of $\frac{1}{n}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta})\tilde{\boldsymbol{B}}(\boldsymbol{\beta})$ as

$$\frac{1}{n}\sum_{i=1}^n \tilde{B}_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})\tilde{B}_{k'\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \Big(\tilde{B}_{k\boldsymbol{\beta}1}(Z_i^T\boldsymbol{\beta})\tilde{B}_{k'\boldsymbol{\beta}1}(Z_i^T\boldsymbol{\beta}) - \tilde{B}_{k\boldsymbol{\beta}1}(Z_i^T\boldsymbol{\beta})\tilde{B}_{k'\boldsymbol{\beta}2}(Z_i^T\boldsymbol{\beta})$$
$$-\tilde{B}_{k\boldsymbol{\beta}2}(Z_i^T\boldsymbol{\beta})\tilde{B}_{k'\boldsymbol{\beta}1}(Z_i^T\boldsymbol{\beta}) + \tilde{B}_{k\boldsymbol{\beta}2}(Z_i^T\boldsymbol{\beta})\tilde{B}_{k'\boldsymbol{\beta}2}(Z_i^T\boldsymbol{\beta})\Big).$$

Applying the Cauchy-Schwarz inequality, Lemma A.2, (A.8) and Assumptions 2 and 5', we obtain

$$\sup_{\boldsymbol{\beta}\in\Theta_{\rho_0}} \max_k \frac{1}{n}\sum_{i=1}^n \Big(\sum_{j=1}^m \frac{1}{\lambda_j}[\frac{1}{n}\sum_{l=1}^n B_{k\boldsymbol{\beta}}(Z_l^T\boldsymbol{\beta})(\hat{\xi}_{lj}-\xi_{lj})]\xi_{ij}\Big)^2$$
$$\le \sup_{\boldsymbol{\beta}\in\Theta_{\rho_0}} \max_k \Big(\sum_{j=1}^m \frac{1}{\lambda_j}[\frac{1}{n}\sum_{l=1}^n B_{k\boldsymbol{\beta}}(Z_l^T\boldsymbol{\beta})(\hat{\xi}_{lj}-\xi_{lj})]^2\Big)\Big(\sum_{j=1}^m \frac{1}{n\lambda_j}\sum_{i=1}^n \xi_{ij}^2\Big)$$
$$\le (\sup_{\boldsymbol{\beta}\in\Theta_{\rho_0}} \max_k \frac{1}{n}\sum_{l=1}^n B^2_{k\boldsymbol{\beta}}(Z_l^T\boldsymbol{\beta}))(\frac{1}{n}\sum_{l=1}^n \|X_l\|^2)(\sum_{j=1}^m \frac{\|\hat{\phi}_j-\phi_j\|^2}{\lambda_j})(\sum_{j=1}^m \frac{1}{n\lambda_j}\sum_{i=1}^n \xi_{ij}^2)$$
$$= O_p(n^{-1}\lambda_m^{-1}m^4 h_0 \log m) = o_p(h_0^3). \tag{A.18}$$

Similar to the proof of (A.7), (A.9) and using Lemma A.2, we then deduce that

$$\sup_{\boldsymbol{\beta}\in\Theta_{\rho_0}} \max_k \frac{1}{n}\sum_{i=1}^n \Big(\sum_{j=1}^m (\frac{\hat{\xi}_{ij}}{\hat{\lambda}_j} - \frac{\xi_{ij}}{\lambda_j})(\frac{1}{n}\sum_{l=1}^n B_{k\boldsymbol{\beta}}(Z_l^T\boldsymbol{\beta})\hat{\xi}_{lj})\Big)^2$$
$$= o_p(n^{-2}\lambda_m^{-2}m^2 h_0^{1/2}(\log n)^2) + o_p(n^{-2}\lambda_m^{-1}m^3 h_0^{1/2}(\log n)^2) + O_p(n^{-2}\lambda_m^{-3}m^4 h_0\log m)$$
$$+O_p(n^{-2}\lambda_m^{-2}m^6 h_0(\log m)^2) = o_p(h_0^3). \tag{A.19}$$

Using Lemma A.2 and Assumption 5', we conclude that

$$\frac{1}{n}\sum_{i=1}^n \tilde{B}_{k\boldsymbol{\beta}1}(Z_i^T\boldsymbol{\beta})\tilde{B}_{k'\boldsymbol{\beta}1}(Z_i^T\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n B_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})B_{k'\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta}) - 2\sum_{j=1}^m \rho_{kj}\rho_{k'j}$$
$$+\sum_{j=1}^m \rho_{kj}\rho_{k'j}(\frac{1}{n\lambda_j}\sum_{i=1}^n \xi_{ij}^2) + \sum_{j\neq j'} \rho_{kj}\rho_{k'j'}\bar{\xi}_{jj'}$$
$$= E[B_{k\boldsymbol{\beta}}(Z^T\boldsymbol{\beta})B_{k'\boldsymbol{\beta}}(Z^T\boldsymbol{\beta})] + o_p(h_0^2),$$

where $\rho_{kj} = \frac{1}{n\lambda_j^{1/2}}\sum_{l=1}^n \xi_{lj}B_{k\boldsymbol{\beta}}(Z_l^T\boldsymbol{\beta})$. Now Lemma A.3 follows from (A.18), (A.19) and the preceding equation.

Proof of Theorem 3.1. By arguments similar to those used in the proof of Lemmas A.1

and A.3, it follows that

$$\frac{1}{n}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta})(\check{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha}) = \Pi(\boldsymbol{\alpha},\boldsymbol{\beta}) + o_p(h_0). \tag{A.20}$$

Using Lemma A.3, (A.20) and arguments similar to those used in the proof of Lemma 1 of Tang (2013), we then deduce that

$$\frac{1}{n}(\check{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha})^T\tilde{\boldsymbol{B}}(\boldsymbol{\beta})(\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta})\tilde{\boldsymbol{B}}(\boldsymbol{\beta}))^{-1}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta})(\check{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha}) = \Pi^T(\boldsymbol{\alpha},\boldsymbol{\beta})\Gamma^{-1}(\boldsymbol{\beta},\boldsymbol{\beta})\Pi(\boldsymbol{\alpha},\boldsymbol{\beta})+o_p(1). \tag{A.21}$$

Therefore, Lemma A.1 and (A.21) imply that

$$\begin{aligned}\tfrac{1}{n}(\check{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha})^T\boldsymbol{P}(\boldsymbol{\beta})(\check{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha}) &= \rho(\boldsymbol{\alpha}) - \Pi^T(\boldsymbol{\alpha},\boldsymbol{\beta})\Gamma^{-1}(\boldsymbol{\beta},\boldsymbol{\beta})\Pi(\boldsymbol{\alpha},\boldsymbol{\beta}) + o_p(1) \\ &=: \tilde{G}(\boldsymbol{\alpha},\boldsymbol{\beta}) + o_p(1),\end{aligned} \tag{A.22}$$

where $o_p(1)$ holds uniformly for $\boldsymbol{\beta}\in\Theta_{\rho_0}$ and $\boldsymbol{\alpha}$ is in any bounded neighborhood of $\boldsymbol{\alpha}_0$. Similar to the proof of Lemmas A.1 and A.3, it holds that $\frac{1}{n}\tilde{\boldsymbol{\varepsilon}}^T\tilde{\boldsymbol{\varepsilon}} = \sigma^2 + o_p(1)$, $\frac{1}{n}(\check{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha})^T\tilde{\boldsymbol{\varepsilon}} = o_p(h_0)$ and $\frac{1}{n}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta})\tilde{\boldsymbol{\varepsilon}} = o_p(h_0)$. Similar to the proof of (A.21) and (A.22), we further have $\frac{1}{n}(\check{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha})^T\boldsymbol{P}(\boldsymbol{\beta})\tilde{\boldsymbol{\varepsilon}} = o_p(1)$ and $\frac{1}{n}\tilde{\boldsymbol{\varepsilon}}^T\boldsymbol{P}(\boldsymbol{\beta})\tilde{\boldsymbol{\varepsilon}} = \sigma^2 + o_p(1)$. Therefore, from (A.1), (A.22) and (3.2), it follows that

$$\tilde{G}_n(\boldsymbol{\alpha},\boldsymbol{\beta}) = G(\boldsymbol{\alpha},\boldsymbol{\beta}) + o_p(1), \tag{A.23}$$

where $o_p(1)$ holds uniformly for $\boldsymbol{\beta}\in\Theta_{\rho_0}$ and $\boldsymbol{\alpha}$ is in any bounded neighborhood of $\boldsymbol{\alpha}_0$. By the fact that $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$ is the minimizer of $\tilde{G}_n(\boldsymbol{\alpha},\boldsymbol{\beta})$ and using (A.23), we have

$$\tilde{G}_n(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) \le \tilde{G}_n(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0) = G(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0) + o_p(1). \tag{A.24}$$

By (A.1) and (A.22), we have that $\tilde{G}(\boldsymbol{\alpha},\boldsymbol{\beta}) \ge 0$ and $G(\boldsymbol{\alpha},\boldsymbol{\beta}) \ge \sigma^2$. From (3.2), one obtains $G(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0) = \sigma^2 + o_p(1)$. Applying (A.23) and (A.24), we obtain that $\sigma^2 \le G(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) = \tilde{G}_n(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})+o_p(1) \le G(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)+o_p(1)$. Therefore, $|G(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})-G(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)| = o_p(1)$; that is, $|G^*(\hat{\boldsymbol{\theta}}_{-d})-G^*(\boldsymbol{\theta}_{0,-d})| = o_p(1)$. Since $G^*(\boldsymbol{\theta}_{-d})$ is locally convex at $\boldsymbol{\theta}_{0,-d}$, it follows that $\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0 = o_p(1)$ and $\hat{\boldsymbol{\beta}}_{-d}-\boldsymbol{\beta}_{0,-d} = o_p(1)$. This completes the proof of (3.3).

From (A.11), Assumption 5 and the fact that $\lambda_j \le C/(j\log j)$, we have

$$\sum_{j=m+1}^{\infty}(a_j+\sum_{r=1}^{q}v_{rj}^{*}\alpha_{0r})^2\lambda_j \le Cm^{-2\gamma}=o(h_0^2). \tag{A.25}$$

Applying Assumption 5 and (A.25), we can easily prove that $\frac{1}{n}(\check{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha})^T(\check{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha})=\rho(\boldsymbol{\alpha})+o_p(h_0^2)$ in Lemma A.1, $\frac{1}{n}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta})\tilde{\boldsymbol{B}}(\boldsymbol{\beta})=\Gamma(\boldsymbol{\beta},\boldsymbol{\beta})+o_p(h_0^4)$ in Lemma A.3 and $\frac{1}{n}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta})(\check{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha})=\Pi(\boldsymbol{\alpha},\boldsymbol{\beta})+o_p(h_0^3)$. Consequently, it follows that $\tilde{G}_n(\boldsymbol{\alpha},\boldsymbol{\beta})=G(\boldsymbol{\alpha},\boldsymbol{\beta})+o_p(h_0^2)$ and $|G(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})-G(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)|=o_p(h_0^2)$. Now (3.4) follows from Assumption 7. This completes the proof of Theorem 3.1.

Lemma A.4. Under Assumptions 1-6, it holds that

$$\ddot{G}_n(\boldsymbol{\theta}_{-d},\tilde{\boldsymbol{b}}(\boldsymbol{\theta}_{-d}))=2\Omega(\boldsymbol{\theta}_{-d})+o_p(1),$$

where $o_p(1)$ holds uniformly for $\boldsymbol{\beta}\in\Theta_{\rho_0}$, $\boldsymbol{\alpha}$ is in any bounded neighborhood of $\boldsymbol{\alpha}_0$ and $\Omega(\boldsymbol{\beta}_{-d})=(\pi_{kr})_{(q+d-1)\times(q+d-1)}$ with

$$\pi_{kr}=E(\check{V}_k\check{V}_r)-E[\boldsymbol{B}(Z^T\boldsymbol{\beta})\check{V}_k]^T\Gamma^{-1}(\boldsymbol{\beta},\boldsymbol{\beta})E[\boldsymbol{B}(Z^T\boldsymbol{\beta})\check{V}_r],\quad k,r=1,\dots,q, \tag{A.26}$$

$$\pi_{k(q+r)}=E[\dot{\boldsymbol{B}}_r(Z^T\boldsymbol{\beta})\check{V}_k]^T\bar{\boldsymbol{b}}(\boldsymbol{\alpha},\boldsymbol{\beta})+E[\boldsymbol{B}(Z^T\boldsymbol{\beta})\check{V}_k]^T\check{\boldsymbol{b}}_r(\boldsymbol{\alpha},\boldsymbol{\beta}), \tag{A.27}$$

for $k,=1,\dots,q; r=1,\dots,d-1$, and

$$\begin{aligned}\pi_{(q+k)(q+r)} &= [\bar{\boldsymbol{b}}^T(\boldsymbol{\alpha},\boldsymbol{\beta})R_{rk}(\boldsymbol{\beta},\boldsymbol{\beta})+\check{\boldsymbol{b}}_r^T(\boldsymbol{\alpha},\boldsymbol{\beta})H_k(\boldsymbol{\beta},\boldsymbol{\beta})]\bar{\boldsymbol{b}}(\boldsymbol{\alpha},\boldsymbol{\beta})-[\ddot{\Pi}_{kr}^T(\boldsymbol{\alpha},\boldsymbol{\beta})\\ &\quad -\bar{\boldsymbol{b}}^T(\boldsymbol{\alpha},\boldsymbol{\beta})M_{kr}(\boldsymbol{\beta},\boldsymbol{\beta})]\bar{\boldsymbol{b}}(\boldsymbol{\alpha},\boldsymbol{\beta})+[\dot{\Pi}_k^T(\boldsymbol{\alpha},\boldsymbol{\beta})-\bar{\boldsymbol{b}}^T(\boldsymbol{\alpha},\boldsymbol{\beta})H_k(\boldsymbol{\beta},\boldsymbol{\beta})]\check{\boldsymbol{b}}_r(\boldsymbol{\alpha},\boldsymbol{\beta}),\end{aligned} \tag{A.28}$$

for $k,r=1,\dots,d-1$, $\bar{\boldsymbol{b}}(\boldsymbol{\alpha},\boldsymbol{\beta})=\Gamma^{-1}(\boldsymbol{\beta},\boldsymbol{\beta})\Pi(\boldsymbol{\alpha},\boldsymbol{\beta})$, $\check{\boldsymbol{b}}_r(\boldsymbol{\alpha},\boldsymbol{\beta})=-\Gamma^{-1}(\boldsymbol{\beta},\boldsymbol{\beta})(H_r^T(\boldsymbol{\beta},\boldsymbol{\beta})+H_r(\boldsymbol{\beta},\boldsymbol{\beta}))\bar{\boldsymbol{b}}(\boldsymbol{\alpha},\boldsymbol{\beta})+\Gamma^{-1}(\boldsymbol{\beta},\boldsymbol{\beta})\dot{\Pi}_r(\boldsymbol{\alpha},\boldsymbol{\beta})$, $\dot{\Pi}_r(\boldsymbol{\alpha},\boldsymbol{\beta})=\frac{\partial\Pi(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\beta_r}$ and $\ddot{\Pi}_{kr}(\boldsymbol{\alpha},\boldsymbol{\beta})=\frac{\partial^2\Pi(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\beta_r\beta_k}$, $M_{kr}(\boldsymbol{\beta},\boldsymbol{\beta}')$ is a $K_n\times K_n$ matrix whose (l,l')th element is $E[B_l(Z^T\boldsymbol{\beta})\ddot{B}_{l'kr}(Z^T\boldsymbol{\beta}')]$ and $\ddot{B}_{lkr}(Z^T\boldsymbol{\beta})=\frac{\partial^2 B_l(Z^T\boldsymbol{\beta})}{\partial\beta_k\partial\beta_r}$.

Proof. Let $\tilde{\pi}_{kr}$ be the (k,r)th element of $\ddot{G}_n(\boldsymbol{\theta}_{-d},\tilde{\boldsymbol{b}}(\boldsymbol{\theta}_{-d}))$. From (3.6) and (3.7), we have that

$$\tilde{\pi}_{kr}=\frac{2}{n}[\tilde{\mathbf{W}}_k^T\tilde{\mathbf{W}}_r-\tilde{\mathbf{W}}_k^T\tilde{\boldsymbol{B}}(\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}\tilde{\boldsymbol{B}}^T\tilde{\mathbf{W}}_r],\quad k,r=1,\dots,q, \tag{A.29}$$

$$\tilde{\pi}_{k(q+r)} = \frac{2}{n}[\tilde{\mathbf{W}}_k^T \dot{\tilde{\boldsymbol{B}}}_r \tilde{\boldsymbol{b}} + \tilde{\mathbf{W}}_k^T \tilde{\boldsymbol{B}} \dot{\boldsymbol{b}}_r], \quad k, = 1, \ldots, q; r = 1, \ldots, d-1, \tag{A.30}$$

$$\tilde{\pi}_{(q+k)(q+r)} = \frac{2}{n}(\dot{\tilde{\boldsymbol{B}}}_r \tilde{\boldsymbol{b}} + \tilde{\boldsymbol{B}} \dot{\boldsymbol{b}}_r)^T \dot{\tilde{\boldsymbol{B}}}_k \tilde{\boldsymbol{b}} - \frac{2}{n}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha} - \tilde{\boldsymbol{B}}\tilde{\boldsymbol{b}})^T (\ddot{\tilde{\boldsymbol{B}}}_{kr} \tilde{\boldsymbol{b}} + \dot{\tilde{\boldsymbol{B}}}_k \dot{\boldsymbol{b}}_r), \tag{A.31}$$

for $k, r = 1, \ldots, d-1$, where $\tilde{\mathbf{W}}_k = (\tilde{W}_{1k}, \ldots, \tilde{W}_{nk})^T$ for $k = 1, \ldots, q$, $\tilde{\boldsymbol{B}} = \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_{-d})$, $\dot{\tilde{\boldsymbol{B}}}_r = \dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_{-d})$ and $\tilde{\boldsymbol{b}} = \tilde{\boldsymbol{b}}(\boldsymbol{\alpha}, \boldsymbol{\beta}_{-d})$, with for simplicity of notation, $\dot{\boldsymbol{b}}_r = \dot{\boldsymbol{b}}_r(\boldsymbol{\alpha}, \boldsymbol{\beta}_{-d}) = \frac{\partial \tilde{\boldsymbol{b}}(\boldsymbol{\alpha}, \boldsymbol{\beta}_{-d})}{\partial \beta_r}$ and $\ddot{\tilde{\boldsymbol{B}}}_{kr} = \ddot{\tilde{\boldsymbol{B}}}_{kr}(\boldsymbol{\beta}_{-d}) = \frac{\partial^2 \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_{-d})}{\partial \beta_k \partial \beta_r}$. Since $(\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}})^{-1} \tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}} = I$, we then have

$$\frac{\partial (\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}})^{-1}}{\partial \beta_r} \tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}} + (\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}})^{-1} \Big(\frac{\partial \tilde{\boldsymbol{B}}^T}{\partial \beta_r} \tilde{\boldsymbol{B}} + \tilde{\boldsymbol{B}}^T \frac{\partial \tilde{\boldsymbol{B}}}{\partial \beta_r} \Big) = 0.$$

Hence,

$$\frac{\partial (\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}})^{-1}}{\partial \beta_r} = -(\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}})^{-1} (\dot{\tilde{\boldsymbol{B}}}_r^T \tilde{\boldsymbol{B}} + \tilde{\boldsymbol{B}}^T \dot{\tilde{\boldsymbol{B}}}_r)(\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}})^{-1}.$$

Note that $\tilde{\boldsymbol{b}} = (\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}})^{-1} \tilde{\boldsymbol{B}}^T (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha})$. We further have

$$\dot{\boldsymbol{b}}_r = -(\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}})^{-1} [(\dot{\tilde{\boldsymbol{B}}}_r^T \tilde{\boldsymbol{B}} + \tilde{\boldsymbol{B}}^T \dot{\tilde{\boldsymbol{B}}}_r)(\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}})^{-1} \tilde{\boldsymbol{B}}^T (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}) - \dot{\tilde{\boldsymbol{B}}}_r^T (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha})]. \tag{A.32}$$

Similar to the proof of Lemmas A.2 and A.3, we obtain that

$$\frac{1}{n}(\dot{\tilde{\boldsymbol{B}}}_r^T \tilde{\boldsymbol{B}} + \tilde{\boldsymbol{B}}^T \dot{\tilde{\boldsymbol{B}}}_r) = H_r^T(\boldsymbol{\beta}, \boldsymbol{\beta}) + H_r(\boldsymbol{\beta}, \boldsymbol{\beta}) + o_p(h_0^3). \tag{A.33}$$

Furthermore, under Assumption 5, Lemma A.3 yields that $\frac{1}{n} \tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}} = \Gamma(\boldsymbol{\beta}, \boldsymbol{\beta}) + o_p(h_0^4)$. Similar to the proof of Lemma 1 of Tang (2013), we obtain that $|(\frac{K_n}{n} \tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}})^{-1} - (K_n \Gamma(\boldsymbol{\beta}, \boldsymbol{\beta}))^{-1}|_\infty = o_p(h_0^3)$. By Lemma 9 of Huang et al. (2004), we also have that $\|(\frac{K_n}{n} \tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}})^{-1}\|_\infty \le C$ and $\|(K_n \Gamma(\boldsymbol{\beta}, \boldsymbol{\beta}))^{-1}\|_\infty \le C$. Using (A.33), we have $\|H_r^T(\boldsymbol{\beta}, \boldsymbol{\beta}) + H_r(\boldsymbol{\beta}, \boldsymbol{\beta})\|_\infty = O(1)$ and

$$\|\frac{1}{n}(\dot{\tilde{\boldsymbol{B}}}_r^T \tilde{\boldsymbol{B}} + \tilde{\boldsymbol{B}}^T \dot{\tilde{\boldsymbol{B}}}_r)\|_\infty = \|H_r^T(\boldsymbol{\beta}, \boldsymbol{\beta}) + H_r(\boldsymbol{\beta}, \boldsymbol{\beta})\|_\infty + o_p(h_0^2) = O_p(1). \tag{A.34}$$

Similar to the proof of (A.20), we obtain $\frac{1}{n} \tilde{\boldsymbol{B}}^T (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}) = \Pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) + o_p(h_0^3)$. Observe that $\|\Pi(\boldsymbol{\alpha}, \boldsymbol{\beta})\|_\infty = O(1)$ and hence $\|\frac{1}{n} \tilde{\boldsymbol{B}}^T (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha})\|_\infty = O_p(1)$. Let $\vec{\boldsymbol{B}}_r = \frac{1}{n}(\dot{\tilde{\boldsymbol{B}}}_r^T \tilde{\boldsymbol{B}} + \tilde{\boldsymbol{B}}^T \dot{\tilde{\boldsymbol{B}}}_r)$,

$\vec{H}_r(\boldsymbol{\beta},\boldsymbol{\beta}) = H_r^T(\boldsymbol{\beta},\boldsymbol{\beta}) + H_r(\boldsymbol{\beta},\boldsymbol{\beta})$ and $\vec{\boldsymbol{Y}} = \frac{1}{n}\tilde{\boldsymbol{B}}^T(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha})$. Then

$$|(\tfrac{K_n}{n}\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}\vec{\boldsymbol{B}}_r(\tfrac{K_n}{n}\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}\vec{\boldsymbol{Y}} - (K_n\Gamma(\boldsymbol{\beta},\boldsymbol{\beta}))^{-1}\vec{\boldsymbol{B}}_r(\tfrac{K_n}{n}\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}\vec{\boldsymbol{Y}}$$

$$\le |(\tfrac{K_n}{n}\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1} - (K_n\Gamma(\boldsymbol{\beta},\boldsymbol{\beta}))^{-1}|_\infty\|\vec{\boldsymbol{B}}_r\|_\infty\|(\tfrac{K_n}{n}\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}\|_\infty\|\vec{\boldsymbol{Y}}\|_\infty$$

$$= o_p(h_0^3)O_p(1)O_p(1)O_p(1) = o_p(h_0^3)$$

and

$$|(K_n\Gamma(\boldsymbol{\beta},\boldsymbol{\beta}))^{-1}\vec{\boldsymbol{B}}_r(\tfrac{K_n}{n}\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}\vec{\boldsymbol{Y}} - (K_n\Gamma(\boldsymbol{\beta},\boldsymbol{\beta}))^{-1}\vec{H}_r(\boldsymbol{\beta},\boldsymbol{\beta})(\tfrac{K_n}{n}\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}\vec{\boldsymbol{Y}}|_\infty$$

$$\le \|(K_n\Gamma(\boldsymbol{\beta},\boldsymbol{\beta}))^{-1}\|_\infty|\vec{\boldsymbol{B}}_r - \vec{H}_r(\boldsymbol{\beta},\boldsymbol{\beta})|_\infty\|(\tfrac{K_n}{n}\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}\|_\infty\|\vec{\boldsymbol{Y}}\|_\infty$$

$$= O(1)o_p(h_0^3)O_p(1)O_p(1) = o_p(h_0^3).$$

Furthermore, it holds that

$$\begin{aligned}&|(\tfrac{K_n}{n}\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}\vec{\boldsymbol{B}}_r(\tfrac{K_n}{n}\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}\vec{\boldsymbol{Y}}\\ &-(K_n\Gamma(\boldsymbol{\beta},\boldsymbol{\beta}))^{-1}\vec{H}_r(\boldsymbol{\beta},\boldsymbol{\beta})(K_n\Gamma(\boldsymbol{\beta},\boldsymbol{\beta}))^{-1}\Pi(\boldsymbol{\alpha},\boldsymbol{\beta})|_\infty = o_p(h_0^3)\end{aligned} \tag{A.35}$$

Under Assumption 5, similar to the proof of (A.20), we deduce that

$$\frac{1}{n}\dot{\tilde{\boldsymbol{B}}}_r^T(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}) = \dot{\Pi}_r(\boldsymbol{\alpha},\boldsymbol{\beta}) + o_p(h_0^2). \tag{A.36}$$

Similar to the proof of (A.35), we further deduce that

$$|(\frac{K_n}{n}\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}(\frac{1}{n}\dot{\tilde{\boldsymbol{B}}}_r^T(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha})) - (K_n\Gamma(\boldsymbol{\beta},\boldsymbol{\beta}))^{-1}\dot{\Pi}_r(\boldsymbol{\alpha},\boldsymbol{\beta})|_\infty = o_p(h_0^2). \tag{A.37}$$

Combining (A.32), (A.35) and (A.37), we then have

$$|\dot{\boldsymbol{b}}_r - \check{\boldsymbol{b}}_r(\boldsymbol{\alpha},\boldsymbol{\beta})|_\infty = o_p(h_0). \tag{A.38}$$

By arguments similar to those used in the proof of (A.35), we further have that

$$\frac{1}{n}\dot{\boldsymbol{b}}_r^T\boldsymbol{B}^T\dot{\tilde{\boldsymbol{B}}}_k\tilde{\boldsymbol{b}} = \check{\boldsymbol{b}}_r^T(\boldsymbol{\alpha},\boldsymbol{\beta})H_k(\boldsymbol{\beta},\boldsymbol{\beta})\bar{\boldsymbol{b}}(\boldsymbol{\alpha},\boldsymbol{\beta}) + o_p(1). \tag{A.39}$$

Similar to the proof of (A.39), we obtain that

$$\frac{1}{n}\tilde{\boldsymbol{b}}^T\dot{\tilde{\boldsymbol{B}}}_r^T\dot{\tilde{\boldsymbol{B}}}_k\tilde{\boldsymbol{b}} = \bar{\boldsymbol{b}}^T(\boldsymbol{\alpha},\boldsymbol{\beta})R_{rk}(\boldsymbol{\beta},\boldsymbol{\beta})\bar{\boldsymbol{b}}(\boldsymbol{\alpha},\boldsymbol{\beta}) + o_p(1), \tag{A.40}$$

$$\frac{1}{n}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha} - \tilde{\boldsymbol{B}}\tilde{\boldsymbol{b}})^T(\ddot{\tilde{\boldsymbol{B}}}_{kr}\tilde{\boldsymbol{b}} + \dot{\tilde{\boldsymbol{B}}}_k\dot{\boldsymbol{b}}_r) = [\ddot{\Pi}_{kr}^T(\boldsymbol{\alpha},\boldsymbol{\beta}) - \bar{\boldsymbol{b}}^T(\boldsymbol{\alpha},\boldsymbol{\beta})M_{kr}(\boldsymbol{\beta},\boldsymbol{\beta})]\bar{\boldsymbol{b}}(\boldsymbol{\alpha},\boldsymbol{\beta})$$

$$+[\dot{\Pi}_k^T(\boldsymbol{\alpha},\boldsymbol{\beta}) - \bar{\boldsymbol{b}}^T(\boldsymbol{\alpha},\boldsymbol{\beta})H_k(\boldsymbol{\beta},\boldsymbol{\beta})]\check{\boldsymbol{b}}_r(\boldsymbol{\alpha},\boldsymbol{\beta}) + o_p(1).$$

Now (A.28) follows from (A.31), (A.39), (A.40) and the preceding expression. Using the fact that $\frac{1}{n}\sum_{i=1}^n(W_{ik} - \frac{1}{n}\sum_{l=1}^n W_{lk}\check{\xi}_{il})(W_{ir} - \frac{1}{n}\sum_{l=1}^n W_{lr}\check{\xi}_{il}) = E(\check{V}_k\check{V}_r) + o_p(1)$, (A.26) and (A.27) can be proved in a similar fashion. This completes the proof of Lemma A.4.

Lemma A.5. Under Assumptions 1 to 3 and 5, it holds that

$$\sum_{j=1}^m \lambda_j[a_j - \frac{1}{\hat{\lambda}_j}(\frac{1}{n}\sum_{l=1}^n \zeta_l\hat{\xi}_{lj})]^2 = O_p(n^{-1}\lambda_m^{-1}m),$$

where $\zeta_l = \sum_{q=1}^\infty a_q\xi_{lq}$.

Proof Set $S_1 = \sum_{j=1}^m \lambda_j[a_j - \frac{1}{\lambda_j}(\frac{1}{n}\sum_{l=1}^n \zeta_l\xi_{lj})]^2$, $S_2 = \sum_{j=1}^m \frac{1}{\lambda_j}[\frac{1}{n}\sum_{l=1}^n \zeta_l(\hat{\xi}_{lj} - \xi_{lj})]^2$ and $S_3 = \sum_{j=1}^m \lambda_j(\frac{1}{\hat{\lambda}_j} - \frac{1}{\lambda_j})^2(\frac{1}{n}\sum_{l=1}^n \zeta_l\hat{\xi}_{lj})^2$. Note that

$$\sum_{j=1}^m \lambda_j[a_j - \frac{1}{\hat{\lambda}_j}(\frac{1}{n}\sum_{l=1}^n \zeta_l\hat{\xi}_{lj})]^2 \le 3(S_1 + S_2 + S_3). \tag{A.41}$$

Since $E[a_j - \frac{1}{\lambda_j}(\frac{1}{n}\sum_{l=1}^n \zeta_l\xi_{lj})] = 0$, then from Assumptions 1to 3, we obtain

$$E(S_1) = \sum_{j=1}^m \frac{1}{\lambda_j}Var(\frac{1}{n}\sum_{l=1}^n \zeta_l\xi_{lj}) \le \sum_{j=1}^m \frac{1}{n^2\lambda_j}\sum_{l=1}^n E(\zeta_l^2\xi_{lj}^2) \le Cm/n. \tag{A.42}$$

Similar to the proof of (A.6), (A.7) and using Assumption 5, we deduce that

$$S_2 = O_p(n^{-2}\lambda_m^{-2}m^3\log m + n^{-1}\lambda_m^{-1}m) = O_p(n^{-1}\lambda_m^{-1}m) \tag{A.43}$$

and

$$\begin{aligned} S_3 &\le C\sum_{j=1}^m \frac{(\hat{\lambda}_j - \lambda_j)^2}{\lambda_j^3}\Big(\bar{\zeta}_j^2 + [\frac{1}{n}\sum_{l=1}^n \zeta_l(\hat{\xi}_{lj} - \xi_{lj})]^2\Big)[1 + o_p(1)] \\ &= O_p(n^{-1}\lambda_m^{-1} + n^{-3}\lambda_m^{-4}m^3\log m + n^{-2}\lambda_m^{-3}m) = O_p(n^{-1}\lambda_m^{-1}). \end{aligned} \tag{A.44}$$

Now Lemma A.5 follows from combining (A.41) to (A.44).

Lemma A.6. Denote

$$\dot{g}_{0r}(Z_i) = \frac{\partial g_0(Z_i^T\boldsymbol{\beta})}{\partial \beta_r}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = \sum_{k=1}^{K_n} b_{0k}B_k'(Z_i^T\boldsymbol{\beta}_0)\Big(Z_{ir} - \frac{\beta_{0r}Z_{id}}{\sqrt{1 - (\beta_{01}^2 + \ldots + \beta_{0(d-1)}^2)}}\Big)$$

for $r = 1, \ldots, d-1$ and $A_{ri} = \dot{g}_{0r}(Z_i) - \frac{1}{n}\sum_{l=1}^n \dot{g}_{0r}(Z_l)\tilde{\xi}_{il}$. Under Assumptions 1, 2, 4 and 5, it holds that

$$\sum_{j=1}^m \lambda_j^{-1}(\sum_{i=1}^n \xi_{ij}A_{ri})^2 = O_p(nm + \lambda_m^{-2}m^4\log m).$$

Proof Let $A_{ri}^* = \dot{g}_{0r}(Z_i) - \sum_{j'=1}^m \frac{1}{\lambda_{j'}}(\frac{1}{n}\sum_{l=1}^n \dot{g}_{0r}(Z_l)\xi_{lj'})\xi_{ij'}$. Observe that

$$\begin{aligned}(\textstyle\sum_{i=1}^n \xi_{ij}A_{ri})^2 \quad &\leq 4\Big(\textstyle\sum_{i=1}^n \xi_{ij}A_{ri}^*\Big)^2 \\ &+4\Big(\textstyle\sum_{i=1}^n \xi_{ij}\sum_{j'=1}^m \frac{1}{\lambda_{j'}}[\frac{1}{n}\sum_{l=1}^n \dot{g}_{0r}(Z_l)(\hat{\xi}_{lj'} - \xi_{lj'})]\xi_{ij'}\Big)^2 \\ &+4\Big(\textstyle\sum_{i=1}^n \xi_{ij}\sum_{j'=1}^m (\frac{1}{\hat{\lambda}_{j'}} - \frac{1}{\lambda_{j'}})[\frac{1}{n}\sum_{l=1}^n \dot{g}_{0r}(Z_l)\hat{\xi}_{lj'}]\xi_{ij'}\Big)^2 \qquad (A.45) \\ &+4\Big(\textstyle\sum_{i=1}^n \xi_{ij}\sum_{j'=1}^m \frac{1}{\hat{\lambda}_{j'}}[\frac{1}{n}\sum_{l=1}^n \dot{g}_{0r}(Z_l)\hat{\xi}_{lj'}](\hat{\xi}_{ij'} - \xi_{ij'})\Big)^2 \\ &=: 4(T_{j1} + T_{j2} + T_{j3} + T_{j4}).\end{aligned}$$

By direct computations and using Assumption 1, we obtain

$$\begin{aligned}E(\xi_{ij}^2 {A_{ri}^*}^2) \quad &\leq 2E(\xi_{ij}^2 \dot{g}_{0r}^2(Z_i)) + 2E[\xi_{ij}^2(\textstyle\sum_{j'=1}^m \frac{1}{\lambda_{j'}}(\frac{1}{n}\sum_{l=1}^n \dot{g}_{0r}(Z_l)\xi_{lj'})\xi_{ij'})^2] \\ &\leq C(\lambda_j + m\lambda_j/n^2 + (n-1)m\lambda_j/n^2 + m^2\lambda_j/n^2) \leq C\lambda_j\end{aligned}$$

and

$$|\sum_{i_1 \neq i_2} E(\xi_{i_1 j}\xi_{i_2 j}A_{ri_1}^* A_{ri_2}^*)| \leq C[(n-1)(n+2)\lambda_j/n + (n-1)m\lambda_j/n] \leq Cn\lambda_j.$$

Hence, it follows that

$$E(T_{j1}) = \sum_{i=1}^n E(\xi_{ij}^2 {A_{ri}^*}^2) + \sum_{i_1 \neq i_2} E(\xi_{i_1 j}\xi_{i_2 j}A_{ri_1}^* A_{ri_2}^*) \leq Cn\lambda_j. \qquad (A.46)$$

Similar to the proof of (A.6) and using Assumption 1, we have

$$\sum_{j'=1}^m \frac{1}{\lambda_{j'}}[\frac{1}{n}\sum_{l=1}^n \dot{g}_{0r}(Z_l)(\hat{\xi}_{lj'} - \xi_{lj'})]^2 = O_p(n^{-2}\lambda_m^{-2}m^3\log m).$$

Since $\sum_{j'=1}^m \frac{1}{\lambda_{j'}} E(\sum_{i=1}^n \xi_{ij}\xi_{ij'})^2 \le Cn^2\lambda_j$, then

$$\begin{aligned}\sum_{j=1}^m \lambda_j^{-1} T_{j2} &\le \Big(\sum_{j'=1}^m \frac{1}{\lambda_{j'}} [\frac{1}{n}\sum_{l=1}^n \dot g_{0r}(Z_l)(\hat\xi_{lj'} - \xi_{lj'})]^2 \Big) \\ &\times \Big(\sum_{j=1}^m \lambda_j^{-1} \sum_{j'=1}^m \frac{1}{\lambda_{j'}} (\sum_{i=1}^n \xi_{ij}\xi_{ij'})^2 \Big) \\ &= O_p(n^{-2}\lambda_m^{-2} m^3 \log m) O_p(n^2 m) = O_p(\lambda_m^{-2} m^4 \log m).\end{aligned} \tag{A.47}$$

Similar to the proof (A.7) and using Assumption 5, we deduce that

$$\begin{aligned}\sum_{j=1}^m \lambda_j^{-1} T_{j3} &\le \Big(\sum_{j'=1}^m \lambda_{j'} (\frac{1}{\hat\lambda_{j'}} - \frac{1}{\lambda_{j'}})^2 [\frac{1}{n}\sum_{l=1}^n \dot g_{0r}(Z_l)\hat\xi_{lj'}]^2 \Big) \\ &\times \Big(\sum_{j=1}^m \lambda_j^{-1} \sum_{j'=1}^m \frac{1}{\lambda_{j'}} (\sum_{i=1}^n \xi_{ij}\xi_{ij'})^2 \Big) \\ &= O_p(\lambda_m^{-2} m^2 + n^{-1}\lambda_m^{-4} m^4 \log m) = O_p(\lambda_m^{-2} m^2 \log m).\end{aligned} \tag{A.48}$$

and

$$\begin{aligned}\sum_{j=1}^m \lambda_j^{-1} T_{j4} &\le \Big(\sum_{j'=1}^m \frac{1}{\lambda_{j'}^2} [\frac{1}{n}\sum_{l=1}^n \dot g_{0r}(Z_l)\hat\xi_{lj'}]^2 \Big) [1 + o_p(1)] \\ &\times \Big(\sum_{j=1}^m \frac{1}{\lambda_j} \sum_{i=1}^n \xi_{ij}^2 \Big) \Big(\sum_{j'=1}^m \sum_{i=1}^n (\hat\xi_{ij'} - \xi_{ij'})^2 \Big) \\ &= O_p(n^{-1}\lambda_m^{-1} m^5 \log m + n^{-2}\lambda_m^{-3} m^7 (\log m)^2) = o_p(\lambda_m^{-2} m^4 \log m).\end{aligned} \tag{A.49}$$

Now Lemma A.6 follows from combining (A.45)-(A.49) and Assumption 5.

Lemma A.7. Under the Assumptions 1-3 and 5, it holds that

$$n^{-1/2} \Big| \sum_{j=1}^m \frac{1}{\hat\lambda_j} (\frac{1}{n}\sum_{l=1}^n \zeta_l \hat\xi_{lj}) \sum_{i=1}^n (\hat\xi_{ij} - \xi_{ij}) A_{ri} \Big| = o_p(1).$$

Proof Let $\check\zeta_j = \frac{1}{n}\sum_{l=1}^n \zeta_l \hat\xi_{lj}$. Applying the Cauchy-Schwarz inequality, we obtain

$$\Big(\sum_{j=1}^m \frac{1}{\hat\lambda_j} \check\zeta_j \sum_{i=1}^n (\hat\xi_{ij} - \xi_{ij}) A_{ri} \Big)^2 \le \Big(\sum_{j=1}^m \frac{1}{\hat\lambda_j^2} \check\zeta_j^2 \Big) \Big(\sum_{j=1}^m (\sum_{i=1}^n (\hat\xi_{ij} - \xi_{ij}) A_{ri})^2 \Big).$$

Using (A.4), (A.5), Assumption 5, Parseval's identity and some arguments similar to those used to prove Lemma A.6, we deduce that

$$\sum_{j=1}^{m}(\sum_{i=1}^{n}(\hat{\xi}_{ij}-\xi_{ij})A_{ri})^2$$
$$\le 2\sum_{j=1}^{m}[(\sum_{k\neq j}(\hat{\lambda}_j-\lambda_k)^{-1}\int\Delta\hat{\phi}_j\phi_k\sum_{i=1}^{n}\xi_{ik}A_{ri})^2+(\sum_{i=1}^{n}\xi_{ij}A_{ri})^2(\int(\hat{\phi}_j-\phi_j)\phi_j)^2]$$
$$\le C|||\Delta|||^2\sum_{j=1}^{m}[\sum_{k\neq j}(\hat{\lambda}_j-\lambda_k)^{-2}(\sum_{i=1}^{n}\xi_{ik}A_{ri})^2+\lambda_j^{-2}j^2(\sum_{i=1}^{n}\xi_{ij}A_{ri})^2]$$
$$=O_p(\lambda_m^{-1}m^3\log m+n^{-1}\lambda_m^{-3}m^6\log m)=o_p(n).$$

Similar to the proof of (A.7) and using Assumption 5 , we obtain that

$$\sum_{j=1}^{m}\frac{1}{\hat{\lambda}_j^2}\zeta_j^2=O_p(n^{-1}\lambda_m^{-1}m+1+n^{-2}\lambda_m^{-3}m^3\log m+n^{-1}\lambda_m^{-2}m)=o_p(1).$$

This completes the proof of Lemma A.7.

Lemma A.8. Set $\tilde{\zeta}_i=\zeta_i-\frac{1}{n}\sum_{l=1}^{n}\zeta_l\tilde{\xi}_{il}$. Under Assumptions 1-4 and 5, it holds that

$$n^{-1/2}|\sum_{i=1}^{n}\tilde{\zeta}_iA_{ri}|=o_p(1).$$

Proof Observe that

$$\begin{aligned}\sum_{i=1}^{n}\tilde{\zeta}_iA_{ri} &= \sum_{j=1}^{m}[a_j-\tfrac{1}{\hat{\lambda}_j}(\tfrac{1}{n}\sum_{l=1}^{n}\zeta_l\hat{\xi}_{lj})]\sum_{i=1}^{n}\xi_{ij}A_{ri}\\ &\quad-\sum_{j=1}^{m}\tfrac{1}{\hat{\lambda}_j}(\tfrac{1}{n}\sum_{l=1}^{n}\zeta_l\hat{\xi}_{lj})\sum_{i=1}^{n}(\hat{\xi}_{ij}-\xi_{ij})A_{ri}+\sum_{j=m+1}^{\infty}a_j\sum_{i=1}^{n}\xi_{ij}A_{ri}.\end{aligned}\quad\text{(A.50)}$$

Lemmas A.5, A.6 and Assumption 5 imply that

$$n^{-\frac{1}{2}}|\sum_{j=1}^{m}[a_j-\tfrac{1}{\hat{\lambda}_j}(\tfrac{1}{n}\sum_{l=1}^{n}\zeta_l\hat{\xi}_{lj})]\sum_{i=1}^{n}\xi_{ij}A_{ri}|$$
$$\le n^{-\frac{1}{2}}\Big(\sum_{j=1}^{m}\lambda_j[a_j-\tfrac{1}{\hat{\lambda}_j}(\tfrac{1}{n}\sum_{l=1}^{n}\zeta_l\hat{\xi}_{lj})]^2\Big)^{\frac{1}{2}}\Big(\sum_{j=1}^{m}\lambda_j^{-1}(\sum_{i=1}^{n}\xi_{ij}A_{ri})^2\Big)^{\frac{1}{2}}\quad\text{(A.51)}$$
$$=O_p(n^{-1/2}\lambda_m^{-1/2}m+n^{-1}\lambda_m^{-3/2}m^{5/2}(\log m)^{1/2})=o_p(1).$$

By arguments similar to those used in the proof of Lemma A.6 and using Lemma 6.1 of Cardot et al. (2007), we obtain that

$$
\begin{aligned}
(\sum_{j=m+1}^{\infty} a_j \sum_{i=1}^{n} \xi_{ij} A_{ri})^2 &\leq (\sum_{j=m+1}^{\infty} a_j^2)(\sum_{j=m+1}^{\infty}(\sum_{i=1}^{n} \xi_{ij} A_{ri})^2) \\
&= O_p(nm^{-2\gamma+1} + \lambda_m^{-2} m^{-2\gamma+4} \log m) \sum_{j=m+1}^{\infty} \lambda_j = o_p(n). \qquad (A.52)
\end{aligned}
$$

Now Lemma A.8 follows from combining (A.50)-(A.52) and Lemma A.7.

Lemma A.9. Suppose that Assumptions 1-5 hold. Then

$$n^{-1/2}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0 - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)^T \dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\boldsymbol{b}_0 = n^{-1/2} \sum_{i=1}^{n} \dot{g}_{0r}(Z_i)\varepsilon_i + o_p(1).$$

Proof Using arguments similar to those used to prove Lemmas A.6 and A.7, we deduce that $\sum_{i=1}^n A_{ri}^2 = O_p(n)$, $n^{-1/2}\sum_{i=1}^n \varepsilon_i(\frac{1}{n}\sum_{l=1}^n \dot{g}_{0r}(Z_l)\tilde{\xi}_{il}) = o_p(1)$ and

$$n^{-1/2} \sum_{i=1}^{n} (\frac{1}{n} \sum_{l=1}^{n} \varepsilon_l \tilde{\xi}_{il}) A_{ri} = o_p(1), \quad n^{-1/2} \sum_{i=1}^{n} (\frac{1}{n} \sum_{l=1}^{n} R(Z_l^T \boldsymbol{\beta}_0) \tilde{\xi}_{il}) A_{ri} = o_p(1).$$

Hence

$$n^{-1/2} \tilde{\boldsymbol{\varepsilon}}^T \dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\boldsymbol{b}_0 = n^{-1/2} \sum_{i=1}^{n} \dot{g}_{0r}(Z_i)\varepsilon_i + o_p(1). \qquad (A.53)$$

Using (3.1) and the assumption that $nh^{2p} \to 0$, it follows that $(\sum_{i=1}^n R(Z_i^T\boldsymbol{\beta}_0)A_{ri})^2 \leq (\sum_{i=1}^n R^2(Z_i^T\boldsymbol{\beta}_0))(\sum_{i=1}^n A_{ri}^2) = o_p(n)$. Consequently, we have

$$n^{-1/2} \tilde{\boldsymbol{R}}^T \dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\boldsymbol{b}_0 = o_p(1), \qquad (A.54)$$

where $\tilde{\boldsymbol{R}} = (\tilde{R}(Z_1^T\boldsymbol{\beta}_0), \ldots, \tilde{R}(Z_n^T\boldsymbol{\beta}_0))^T$ and $\tilde{R}(Z_i^T\boldsymbol{\beta}_0) = R(Z_i^T\boldsymbol{\beta}_0) - \frac{1}{n}\sum_{l=1}^n R(Z_l^T\boldsymbol{\beta}_0)\tilde{\xi}_{il}$. Now Lemma A.9 follows from Lemma A.8, (A.53) and (A.54).

Lemma A.10. Under the assumptions of Theorem 2, it holds that

$$n^{-\frac{1}{2}}(\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0)^T \tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0) \dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\boldsymbol{b}_0 = n^{-\frac{1}{2}} \boldsymbol{\varepsilon}^T \boldsymbol{B}(\boldsymbol{\beta}_0) \Gamma^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0) H_r(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0)\boldsymbol{b}_0 + o_p(1),$$

where $\boldsymbol{B}(\boldsymbol{\beta}_0) = (\boldsymbol{B}(Z_1^T\boldsymbol{\beta}_0), \ldots, \boldsymbol{B}(Z_n^T\boldsymbol{\beta}_0))^T$.

Proof Note that $\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)-\boldsymbol{b}_0=(\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0))^{-1}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)(\tilde{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0-\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)$. By (A.33), we obtain

$$\begin{aligned}|\tfrac{1}{n}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\boldsymbol{b}_0|_\infty &= |H_r(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0)\boldsymbol{b}_0|_\infty+o_p(h_0^2)\\ &\le \max\nolimits_{1\le k\le K_n} E[B_k(Z^T\boldsymbol{\beta}_0)|\dot{g}_{0r}(Z)|]+o_p(h_0^2)=O_p(h_0)\end{aligned}$$

Similar to the proof of Lemma A.9, we have $\|n^{-\frac{1}{2}}(\tilde{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0-\boldsymbol{\varepsilon}-\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)^T\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\|_\infty=o_p(1)$ and $\|n^{-\frac{1}{2}}\boldsymbol{\varepsilon}^T\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\|_\infty=O_p(K_n^{1/2})$. Hence

$$\begin{aligned}&n^{-\frac{1}{2}}|(\tilde{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0-\boldsymbol{\varepsilon}-\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)^T\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)(\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0))^{-1}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\boldsymbol{b}_0|\\ &\le K_n\|n^{-\frac{1}{2}}|(\tilde{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0-\boldsymbol{\varepsilon}-\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)^T\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\|_\infty\|(\tfrac{K_n}{n}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0))^{-1}\|_\infty\\ &\times|\tfrac{1}{n}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\boldsymbol{b}_0|_\infty=K_no_p(1)O_p(1)O_p(h_0)=o_p(1).\end{aligned}$$

Using arguments similar to those used in the proof of (A.35), we can deduce that

$$|(\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0))^{-1}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\boldsymbol{b}_0-\Gamma^{-1}(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0)H_r(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0)\boldsymbol{b}_0|_\infty=o_p(h_0).$$

Hence

$$\begin{aligned}&|n^{-\frac{1}{2}}\boldsymbol{\varepsilon}^T\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)[(\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0))^{-1}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\boldsymbol{b}_0-\Gamma^{-1}(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0)H_r(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0)\boldsymbol{b}_0]|\\ &\le\|n^{-\frac{1}{2}}\boldsymbol{\varepsilon}^T\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\|_\infty|(\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0))^{-1}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\boldsymbol{b}_0-\Gamma^{-1}(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0)H_r(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0)\boldsymbol{b}_0|_\infty\\ &=O_p(K_n^{1/2})o_p(h_0)=o_p(1).\end{aligned}$$

Using arguments similar to those used to prove Lemmas A.6 and A.7, we deduce that

$$\|n^{-\frac{1}{2}}\boldsymbol{\varepsilon}^T(\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)-\boldsymbol{B}(\boldsymbol{\beta}_0))\|_\infty=n^{-\frac{1}{2}}\sum_{k=1}^{K_n}|\sum_{i=1}^n\varepsilon_i[\frac{1}{n}\sum_{l=1}^nB_k(Z_l^T\boldsymbol{\beta}_0)\tilde{\xi}_{il}]|=o_p(1).$$

Therefore,

$$\begin{aligned}&|n^{-\frac{1}{2}}\boldsymbol{\varepsilon}^T(\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)-\boldsymbol{B}(\boldsymbol{\beta}_0))\Gamma^{-1}(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0)H_r(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0)\boldsymbol{b}_0]|\\ &\le K_n\|n^{-\frac{1}{2}}\boldsymbol{\varepsilon}^T(\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)-\boldsymbol{B}(\boldsymbol{\beta}_0))\|_\infty\|(K_n\Gamma(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0))^{-1}\|_\infty|H_r(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0)\boldsymbol{b}_0|_\infty\\ &=K_no_p(1)O_p(1)O_p(h_0)=o_p(1).\end{aligned}$$

This completes the proof of Lemma A.10.

Proof of Theorem 3.2. From Lemma A.4 and Assumption 7, we have

$$\ddot{G}_n(\boldsymbol{\beta}^*_{-d}, \tilde{\boldsymbol{b}}(\boldsymbol{\beta}^*_{-d})) = 2\Omega(\boldsymbol{\beta}^*_{-d}) + o_p(1) = 2\Omega(\boldsymbol{\beta}_{0,-d}) + o_p(1) = 2\Omega_0 + o_p(1). \tag{A.55}$$

Note that $(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0 - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0))^T \dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ can be written as

$$\begin{aligned}
&(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0 - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0))^T \dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\\
&= (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0 - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)^T \dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\boldsymbol{b}_0 - (\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0)^T \tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\boldsymbol{b}_0\\
&+(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0 - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)^T \dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)(\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0)\\
&-(\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0)^T \tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)(\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0).
\end{aligned} \tag{A.56}$$

Similar to the proof of Lemma A.9, we have $\|n^{-\frac{1}{2}}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0 - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)^T \dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\|_\infty = O_p(K_n^{3/2})$ and $|n^{-\frac{1}{2}}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0 - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)^T \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)|_\infty = O_p(h_0^{1/2})$. Hence

$$\begin{aligned}
|\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0|_\infty &\leq \tfrac{K_n}{n}\|(\tfrac{K_n}{n}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0))^{-1}\|_\infty |\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0 - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)|_\infty\\
&= O_p(n^{-\frac{1}{2}}h_0^{-\frac{1}{2}}).
\end{aligned}$$

Further, we deduce that

$$\begin{aligned}
&n^{-\frac{1}{2}}|(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0 - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)^T \dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)(\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0)|\\
&\leq \|n^{-\frac{1}{2}}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}_0 - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)^T \dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\|_\infty |\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0|_\infty\\
&= O_p(n^{-1/2}h_0^{-2}) = o_p(1).
\end{aligned} \tag{A.57}$$

Applying (A.34), we have

$$\begin{aligned}
&n^{-\frac{1}{2}}|(\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0)^T \tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)(\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0)|\\
&\leq n^{\frac{1}{2}}K_n\|\tfrac{1}{n}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}_0)\|_\infty |\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0|_\infty^2 = O_p(n^{-1/2}h_0^{-2}) = o_p(1).
\end{aligned} \tag{A.58}$$

Now $(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{B}}(\hat{\boldsymbol{\beta}}_{-d})\hat{\boldsymbol{b}})^T \tilde{\boldsymbol{W}}_k$ can be written as

$$(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{B}}(\hat{\boldsymbol{\beta}}_{-d})\hat{\boldsymbol{b}})^T \tilde{\boldsymbol{W}}_k = (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)^T \tilde{\boldsymbol{W}}_k - (\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0)^T \tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\tilde{\boldsymbol{W}}_k$$

for $k = 1, \ldots, q$. Similar to the proof of Lemma A.9, we deduce that

$$n^{-\frac{1}{2}}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_0)\boldsymbol{b}_0)^T\tilde{\boldsymbol{W}}_k = n^{-\frac{1}{2}}\boldsymbol{\varepsilon}^T\tilde{\boldsymbol{W}}_k + o_p(1).$$

We decompose $\boldsymbol{\varepsilon}^T\tilde{\boldsymbol{W}}_k$ into three terms as

$$\begin{aligned}\boldsymbol{\varepsilon}^T\tilde{\boldsymbol{W}}_k \quad &= \textstyle\sum_{i=1}^n \varepsilon_i\Big(W_{ik} - \sum_{j=1}^m \frac{E(W_{lk}\xi_j)}{\lambda_j}\xi_{ij}\Big) - \sum_{i=1}^n \varepsilon_i \sum_{j=1}^m \frac{\xi_{ij}}{\lambda_j}\Big(\frac{1}{n}\sum_{l=1}^n W_{lk}\xi_{lj} - E(W_{lk}\xi_j)\Big)\\ &\quad - \textstyle\sum_{i=1}^n \varepsilon_i \frac{1}{n}\sum_{l=1}^n W_{lk}(\tilde{\xi}_{il} - \check{\xi}_{il}).\end{aligned}$$

Similar to the proof of Lemma A.8, we have $\sum_{i=1}^n \varepsilon_i \frac{1}{n}\sum_{l=1}^n W_{lk}(\tilde{\xi}_{il} - \check{\xi}_{il}) = o_p(n)$. Since

$$\sum_{i=1}^n \varepsilon_i\Big(W_{ik} - \sum_{j=1}^m \frac{E(W_{lk}\xi_j)}{\lambda_j}\xi_{ij}\Big) = \sum_{i=1}^n \varepsilon_i \check{V}_{ik} + \sum_{i=1}^n \varepsilon_i \sum_{j=m+1}^{\infty} v_{kj}^*\xi_{ij},$$

$\sum_{i=1}^n \varepsilon_i \sum_{j=1}^m \frac{\xi_{ij}}{\lambda_j}\Big(\frac{1}{n}\sum_{l=1}^n W_{lk}\xi_{lj} - E(W_{lk}\xi_j)\Big) = o_p(n)$ and $\sum_{i=1}^n \varepsilon_i \sum_{j=m+1}^{\infty} v_{kj}^*\xi_{ij} = o_p(n)$, it follows that $n^{-\frac{1}{2}}\boldsymbol{\varepsilon}^T\tilde{\boldsymbol{W}}_k = n^{-\frac{1}{2}}\boldsymbol{\varepsilon}^T\check{\boldsymbol{V}}_k + o_p(1)$, where $\check{\boldsymbol{V}}_k = (\check{V}_{1k}, \ldots, \check{V}_{nk})^T$. Similar to the proof of Lemma A.10, we have

$$n^{-\frac{1}{2}}(\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \boldsymbol{b}_0)^T\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_0)\tilde{\boldsymbol{W}}_k = n^{-\frac{1}{2}}\boldsymbol{\varepsilon}^T\boldsymbol{B}(\boldsymbol{\beta}_0)\Gamma^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0)E(\boldsymbol{B}(Z^T\boldsymbol{\beta}_0)W_k) + o_p(1).$$

Hence

$$n^{-\frac{1}{2}}(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{B}}(\hat{\boldsymbol{\beta}}_{-d})\hat{\boldsymbol{b}})^T\tilde{\boldsymbol{W}}_k = n^{-\frac{1}{2}}\boldsymbol{\varepsilon}^T(\check{\boldsymbol{V}}_k - \boldsymbol{B}(\boldsymbol{\beta}_0)\Gamma^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0)E(\boldsymbol{B}(Z^T\boldsymbol{\beta}_0)W_k)) + o_p(1). \tag{A.59}$$

Now (3.10) follows from (A.55)-(A.59), Lemmas A.9 and A.10, and the Central Limit Theorem. This completes the proof of Theorem 3.2.

Lemma A.11. Under the assumptions of Theorem 3.3, it holds that

$$\|\tilde{\boldsymbol{b}}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \boldsymbol{b}_0\|^2 = O_p(n^{-1}K_n^2).$$

Proof. From Assumption 6 and Lemma A.3, all the eigenvalues of $(\frac{K_n}{n}\tilde{\boldsymbol{B}}^T(\hat{\boldsymbol{\beta}})\tilde{\boldsymbol{B}}(\hat{\boldsymbol{\beta}}))^{-1}$ are bounded away from zero and infinity, except possibly on an event whose probability tends to zero. We then have

$$\|\tilde{\boldsymbol{b}}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})-\boldsymbol{b}_0\|^2 \le CK_n^2\|\tilde{\boldsymbol{B}}^T(\hat{\boldsymbol{\beta}})(\tilde{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{B}}(\hat{\boldsymbol{\beta}})\boldsymbol{b}_0)\|^2/n^2, \tag{A.60}$$

where $\|a\|=(a_1^2+\ldots+a_k^2)^{1/2}$ for a vector $a=(a_1,\ldots,a_k)^T$. Let $\boldsymbol{F}(\boldsymbol{\alpha},\boldsymbol{\beta}_{-d})=\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}_{-d})(\tilde{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha}-\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_{-d})\boldsymbol{b}_0)$. By a Taylor expansion, we have that

$$\boldsymbol{F}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{-d})=\boldsymbol{F}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_{0,-d})-\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}^\star_{-d})\tilde{\boldsymbol{W}}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)+\frac{\partial \boldsymbol{F}}{\partial\boldsymbol{\beta}_{-d}}\Big|_{\boldsymbol{\beta}_{-d}=\boldsymbol{\beta}^\star_{-d}}(\hat{\boldsymbol{\beta}}_{-d}-\boldsymbol{\beta}_{0,-d}), \tag{A.61}$$

where $(\boldsymbol{\alpha}^{\star T},\boldsymbol{\beta}^{\star}_{-d}{}^T)^T$ is between $(\hat{\boldsymbol{\alpha}}^T,\hat{\boldsymbol{\beta}}_{-d}^T)^T$ and $(\boldsymbol{\alpha}_0^T,\boldsymbol{\beta}_{0,-d}^T)^T$, and

$$\frac{\partial \boldsymbol{F}}{\partial\beta_r}\Big|_{\boldsymbol{\beta}_{-d}=\boldsymbol{\beta}^\star_{-d}}=\dot{\tilde{\boldsymbol{B}}}_r^T(\boldsymbol{\beta}^\star_{-d})(\tilde{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha}^\star-\tilde{\boldsymbol{B}}(\boldsymbol{\beta}^\star_{-d})\boldsymbol{b}_0)-\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}^\star_{-d})\dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}^\star_{-d})\boldsymbol{b}_0.$$

Similar to the proof of (A.33) and (A.36), we obtain that

$$\|\frac{1}{n}\dot{\tilde{\boldsymbol{B}}}_r^T(\boldsymbol{\beta}^\star_{-d})(\tilde{\boldsymbol{Y}}-\tilde{\boldsymbol{W}}\boldsymbol{\alpha}^\star-\tilde{\boldsymbol{B}}(\boldsymbol{\beta}^\star_{-d})\boldsymbol{b}_0)\|^2=\|E[\dot{\boldsymbol{B}}_r(Z^T\boldsymbol{\beta}^\star_{-d})\check{V}^T](\boldsymbol{\alpha}^\star-\boldsymbol{\alpha}_0)\|^2+o_p(1)=o_p(1)$$

and $\|\frac{1}{n}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}^\star)\dot{\tilde{\boldsymbol{B}}}_r(\boldsymbol{\beta}^\star)\boldsymbol{b}_0\|^2=O_p(1)$. From Theorem 3.2, it holds that $\|\hat{\boldsymbol{\beta}}_{-d}-\boldsymbol{\beta}_{0,-d}\|^2=O_p(n^{-1})$. Hence

$$\Big\|\frac{\partial \boldsymbol{F}}{\partial\boldsymbol{\beta}_{-d}}\Big|_{\boldsymbol{\beta}_{-d}=\boldsymbol{\beta}^\star_{-d}}(\hat{\boldsymbol{\beta}}_{-d}-\boldsymbol{\beta}_{0,d})\Big\|^2\le\sum_{r=1}^{d-1}\Big\|\frac{\partial \boldsymbol{F}}{\partial\beta_r}\Big|_{\boldsymbol{\beta}_{-d}=\boldsymbol{\beta}^\star_{-d}}\Big\|^2\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2=o_p(n). \tag{A.62}$$

It is easy to prove that $\|\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta}^\star_{-d})\tilde{\boldsymbol{W}}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\|^2=o_p(n)$. By arguments similar to those used to prove Lemma A.9, we can prove that $\|\boldsymbol{F}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_{0,-d})\|^2=O_p(n)$. Now Lemma A.11 follows from (A.60)-(A.62). This completes the proof of Lemma A.11.

Lemma A.12. Define $\check{a}_j=\frac{1}{\lambda_j}E[(Y-W^T\boldsymbol{\alpha}_0-g(Z^T\boldsymbol{\beta}_0))\xi_j]$. Under the assumptions of Theorem 3.3, it holds that

$$\sum_{j=1}^{\tilde{m}}(\hat{a}_j-\check{a}_j)^2=O_p(n^{-1}\tilde{m}\lambda_{\tilde{m}}^{-1}+n^{-2}\tilde{m}\lambda_{\tilde{m}}^{-2}\sum_{j=1}^{\tilde{m}}a_j^2\lambda_j^{-2}j^3).$$

Proof. Note that $E[(Y-W^T\boldsymbol{\alpha}_0-g(Z^T\boldsymbol{\beta}_0))\xi_j]=a_j\lambda_j$. Define $I_1=\frac{1}{n}\sum_{i=1}^n[Y_i-W_i^T\boldsymbol{\alpha}_0-g(Z_i^T\boldsymbol{\beta}_0)]\xi_{ij}-a_j\lambda_j$, $I_2=\frac{1}{n}\sum_{i=1}^n[Y_i-W_i^T\boldsymbol{\alpha}_0-g(Z_i^T\boldsymbol{\beta}_0)](\hat{\xi}_{ij}-\xi_{ij})$ and $I_3=\frac{1}{n}\sum_{i=1}^n[W_i^T(\hat{\boldsymbol{\alpha}}-$

$\boldsymbol{\alpha}_0) + (\hat{g}(Z_i^T\hat{\boldsymbol{\beta}}) - g(Z_i^T\boldsymbol{\beta}_0))]\hat{\xi}_{ij}$. Then we have

$$\sum_{j=1}^{\tilde{m}}(\hat{a}_j - \check{a}_j)^2 \leq 3\sum_{j=1}^{\tilde{m}}\lambda_j^{-2}(I_1^2 + I_2^2 + I_3^2)[1 + o_p(1)], \tag{A.63}$$

where $o_p(1)$ holds uniformly for $j = 1, \ldots, \tilde{m}$. Since $E(I_1) = 0$ and $E(I_1^2) \leq \frac{1}{n}[\sum_{k=1}^{\infty} a_k^2 E(\xi_k^2\xi_j^2) + \sigma^2\lambda_j] \leq C\lambda_j/n$, we obtain that

$$\sum_{j=1}^{\tilde{m}}\lambda_j^{-2}I_1^2 = O_p(n^{-1}\sum_{j=1}^{\tilde{m}}\lambda_j^{-1}) = O_p(n^{-1}\tilde{m}\lambda_{\tilde{m}}^{-1}). \tag{A.64}$$

Let $M(t) = E[(Y_i - W_i^T\boldsymbol{\alpha}_0 - g(Z_i^T\boldsymbol{\beta}_0))X_i(t)] = \sum_{k=1}^{\infty} a_k\lambda_k\phi_k(t)$. Then

$$\begin{aligned} I_2^2 &\leq 2\int_{\mathcal{T}}\Big(\tfrac{1}{n}\textstyle\sum_{i=1}^n[Y_i - W_i^T\boldsymbol{\alpha}_0 - g(Z_i^T\boldsymbol{\beta}_0)]X_i(t) - M(t)\Big)^2 dt\|\hat{\phi}_j - \phi_j\|^2 \\ &+2\Big(\int_{\mathcal{T}} M(t)(\hat{\phi}_j(t) - \phi_j(t))dt\Big)^2. \end{aligned}$$

Applying Assumption 1, it holds that

$$\begin{aligned} &E\Big(\int_{\mathcal{T}}\Big(\tfrac{1}{n}\textstyle\sum_{i=1}^n[Y_i - W_i^T\boldsymbol{\alpha}_0 - g(Z_i^T\boldsymbol{\beta}_0)]X_i(t) - M(t)\Big)^2 dt\Big) \\ &\leq \tfrac{1}{n}\int_{\mathcal{T}} E([Y_i - W_i^T\boldsymbol{\alpha}_0 - g(Z_i^T\boldsymbol{\beta}_0)]^2X_i^2(t))dt = O(n^{-1}). \end{aligned}$$

From (A.8), we obtain $\sum_{j=1}^{\tilde{m}}\lambda_j^{-2}\|\hat{\phi}_j - \phi_j\|^2 = O_p(n^{-1}\tilde{m}^3\lambda_{\tilde{m}}^{-2}\log\tilde{m})$. By arguments similar to those used in the proof of (5.15) of Hall and Horowitz (2007), it follows that

$$\sum_{j=1}^{\tilde{m}}\lambda_j^{-2}\Big(\int_{\mathcal{T}} M(t)(\hat{\phi}_j(t) - \phi_j(t))dt\Big)^2 = O_p\Big(\frac{\tilde{m}}{n\lambda_{\tilde{m}}} + \frac{\tilde{m}}{n^2\lambda_{\tilde{m}}^2}\sum_{j=1}^{\tilde{m}} a_j^2\lambda_j^{-2}j^3 + \frac{\tilde{m}^3\log\tilde{m}}{n^2\lambda_{\tilde{m}}^2}\Big).$$

Hence, using the assumption that $n^{-1}\tilde{m}^2\lambda_{\tilde{m}}^{-1}\log\tilde{m} \to 0$, we obtain

$$\sum_{j=1}^{\tilde{m}}\lambda_j^{-2}I_2^2 = O_p(n^{-1}\tilde{m}\lambda_{\tilde{m}}^{-1} + n^{-2}\tilde{m}\lambda_{\tilde{m}}^{-2}\sum_{j=1}^{\tilde{m}} a_j^2\lambda_j^{-2}j^3). \tag{A.65}$$

Define $I_{31} = \frac{1}{n}\sum_{i=1}^n[\hat{g}(Z_i^T\hat{\boldsymbol{\beta}}) - g_0(Z_i^T\hat{\boldsymbol{\beta}})]\hat{\xi}_{ij}$, $I_{32} = \frac{1}{n}\sum_{i=1}^n[W_i^T(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + (g_0(Z_i^T\hat{\boldsymbol{\beta}}) - g(Z_i^T\boldsymbol{\beta}_0))]\hat{\xi}_{ij}$ and $L_j = (l_{jkk'})_{K_n\times K_n}$ with $l_{jkk'} = (\frac{1}{n}\sum_{i=1}^n B_k(Z_i^T\hat{\boldsymbol{\beta}})\hat{\xi}_{ij})(\frac{1}{n}\sum_{i=1}^n B_{k'}(Z_i^T\hat{\boldsymbol{\beta}})\hat{\xi}_{ij})$. We write $\frac{1}{n}\sum_{i=1}^n B_k(Z_i^T\hat{\boldsymbol{\beta}})\hat{\xi}_{ij} = \frac{1}{n}\sum_{i=1}^n[B_k(Z_i^T\boldsymbol{\beta}_0)\xi_{ij} + (B_k(Z_i^T\hat{\boldsymbol{\beta}}) - B_k(Z_i^T\boldsymbol{\beta}_0))\xi_{ij} + B_k(Z_i^T\hat{\boldsymbol{\beta}})(\hat{\xi}_{ij} - \xi_{ij})]$. Then we have

$$|L_j|_\infty = \max_{k,k'}|l_{jkk'}| \leq \sum_{k=1}^{K_n}\Big(\frac{1}{n}\sum_{i=1}^n B_k(Z_i^T\boldsymbol{\beta}_0)\xi_{ij}\Big)^2 + \frac{C}{n}\sum_{i=1}^n[h_0^{-2}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2\xi_{ij}^2 + (\hat{\xi}_{ij} - \xi_{ij})^2].$$

Simple calculations yield $\sum_{k=1}^{K_n} E\Big(\frac{1}{n}\sum_{i=1}^{n} B_k(Z_i^T\boldsymbol{\beta}_0)\xi_{ij}\Big)^2 \le Cn^{-1}\lambda_j$. Applying Lemma A.11, we obtain that $\|\tilde{\boldsymbol{b}}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - \boldsymbol{b}_0\|_\infty^2 \le K_n\|\tilde{\boldsymbol{b}}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - \boldsymbol{b}_0\|^2 = O_p(n^{-1}K_n^3)$. Hence, under the assumptions of Theorem 3.3, it holds that

$$\begin{aligned}\sum_{j=1}^{\tilde{m}} \lambda_j^{-2} I_{31}^2 &\le \sum_{j=1}^{\tilde{m}} \lambda_j^{-2}\|\tilde{\boldsymbol{b}}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - \boldsymbol{b}_0\|_\infty^2 \cdot |L_j|_\infty \\ &= O_p(n^{-2}\tilde{m}\lambda_{\tilde{m}}^{-1}h_0^{-3} + n^{-2}\tilde{m}\lambda_{\tilde{m}}^{-1}h_0^{-5} + n^{-2}\tilde{m}^3\lambda_{\tilde{m}}^{-2}h_0^{-3}\log\tilde{m}) \\ &= O_p(n^{-1}\tilde{m}\lambda_{\tilde{m}}^{-1}).\end{aligned} \tag{A.66}$$

Using a Taylor expansion, Theorem 3.2, and the assumption that $nh_0^{2p} \to 0$, we deduce that

$$\begin{aligned}\sum_{j=1}^{\tilde{m}} \lambda_j^{-2} I_{32}^2 &\le \Big(\sum_{j=1}^{\tilde{m}} \frac{1}{n\lambda_j^2}\sum_{i=1}^{n}\hat{\xi}_{ij}^2\Big)\Big(\frac{1}{n}\sum_{i=1}^{n}[W_i^T(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + (g_0(Z_i^T\hat{\boldsymbol{\beta}}) - g(Z_i^T\boldsymbol{\beta}_0))]^2\Big) \\ &= O_p(\tilde{m}\lambda_{\tilde{m}}^{-1} + n^{-1}\tilde{m}^3\lambda_{\tilde{m}}^{-2}\log\tilde{m})O_p(n^{-1} + h_0^{2p}) = O_p(n^{-1}\tilde{m}\lambda_{\tilde{m}}^{-1}).\end{aligned} \tag{A.67}$$

Now Lemma A.12 follows from combining (A.63)-(A.67).

Proof of Theorem 3.3. Note that

$$\int_{\mathcal{T}}[\hat{a}(t) - a(t)]^2 dt \le C\Big(\sum_{j=1}^{\tilde{m}}(\hat{a}_j - \breve{a}_j)^2 + \sum_{j=1}^{\tilde{m}}(\breve{a}_j - a_j)^2 + \tilde{m}\sum_{j=1}^{\tilde{m}} a_j^2\|\hat{\phi}_j - \phi_j\|^2 + \sum_{j=\tilde{m}+1}^{\infty} a_j^2\Big) \tag{A.68}$$

and

$$\sum_{j=1}^{\tilde{m}}(\breve{a}_j - a_j)^2 = \sum_{j=1}^{\tilde{m}} \frac{(\hat{\lambda}_j - \lambda_j)^2}{\lambda_j^2} a_j^2[1 + o_p(1)] = O_p(n^{-1}\lambda_{\tilde{m}}^{-1}\sum_{j=1}^{\tilde{m}} a_j^2\lambda_j^{-1}). \tag{A.69}$$

Assumption 3 implies that $\tilde{m}\sum_{j=1}^{\tilde{m}} a_j^2\|\hat{\phi}_j - \phi_j\|^2 = O_p(\tilde{m}n^{-1}\sum_{j=1}^{\tilde{m}} a_j^2 j^2\log j) = o_p(\tilde{m}/n)$ and $\sum_{j=\tilde{m}+1}^{\infty} a_j^2 = O(\tilde{m}^{-2\gamma+1})$. Now (3.11) follows from Lemma A.12, (A.68) and (A.69). This completes the proof of Theorem 3.3.

Proof of Theorem 3.4. From Assumption 6 and Lemma A.3, all the eigenvalues of $(\frac{K_n^*}{n}\boldsymbol{B}^{*T}(\hat{\boldsymbol{\beta}})\boldsymbol{B}^*(\hat{\boldsymbol{\beta}}))^{-1}$ are bounded away from zero and infinity, except possibly on an event whose probability tends to zero. Similar to (3.1), there exists a spline function $g^*(u) = \sum_{k=1}^{K_n^*} b_{0k}^* B_k^*(u)$ such that

$$\sup_{u\in[U_{\boldsymbol{\beta}_0}, U^{\boldsymbol{\beta}_0}]} |g(u) - g^*(u)| \le Ch^p. \tag{A.70}$$

Let $\boldsymbol{b}_0^* = (b_{01}^*, \ldots, b_{0K_n^*}^*)^T$. Using the properties of B-splines (de Boor 1978), we obtain

$$\int_{U_{\boldsymbol{\beta}_0}}^{U^{\boldsymbol{\beta}_0}} (\hat{g}(u) - g(u))^2 du \quad \leq C(\|\boldsymbol{b}^*(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \boldsymbol{b}_0^*\|^2 / K_n^* + h^{2p}). \tag{A.71}$$

Using arguments similar to those used to prove Lemma A.11 and using the fact that

$$\sum_{k=1}^{K_n^*} (\sum_{i=1}^{n} B_k^*(Z_i \boldsymbol{\beta}_0) R^*(Z_i \boldsymbol{\beta}_0))^2 = O_p(n^2 h^{2p+1}),$$

where $R^*(u) = g(u) - \sum_{k=1}^{K_n^*} b_{0k}^* B_k^*(u)$, one can prove that

$$\|\boldsymbol{b}^*(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \boldsymbol{b}_0^*\|^2 = O_p(n^{-1} K_n^{*2}) + O_p(h^{2p-1}). \tag{A.72}$$

Now (3.13) follows from (A.71) and the fact that $h = O(K_n^{*-1})$. This completes the proof of Theorem 3.4.

Proof of Theorem 3.5. Observe that

$$\text{MSPE} \leq 3\{\|\hat{a} - a\|_K^2 + (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)^T E(WW^T)(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + E([\hat{g}(Z_{n+1}^T \hat{\boldsymbol{\beta}}) - g(Z_{n+1}^T \boldsymbol{\beta}_0)]^2 | \mathcal{S})\}, \tag{A.73}$$

where $\|\hat{a} - a\|_K^2 = \int_{\mathcal{T}} \int_{\mathcal{T}} K(s,t)[\hat{a}(s) - a(s)][\hat{a}(t) - a(t)] ds dt$. Under the assumptions of Theorem 3.5, using arguments similar to those used in the proof of Theorem 2 of Tang (2015), we deduce that

$$\|\hat{a} - a\|_K^2 = O_p(n^{-(\delta+2\gamma-1)/(\delta+2\gamma)}). \tag{A.74}$$

Write

$$\hat{g}(Z_{n+1}^T \hat{\boldsymbol{\beta}}) - g(Z_{n+1}^T \boldsymbol{\beta}_0) = \hat{g}(Z_{n+1}^T \hat{\boldsymbol{\beta}}) - g^*(Z_{n+1}^T \hat{\boldsymbol{\beta}}) + g^*(Z_{n+1}^T \hat{\boldsymbol{\beta}}) - g(Z_{n+1}^T \boldsymbol{\beta}_0).$$

Using a Taylor expansion, Theorems 3.2 and 3.4, (A.71), and the property of B-spline function, we obtain

$$
\begin{aligned}
E([\hat{g}(Z_{n+1}^T\hat{\boldsymbol{\beta}}) - g^*(Z_{n+1}^T\hat{\boldsymbol{\beta}})]^2|\mathcal{S}) &\leq 2E([\hat{g}(Z_{n+1}^T\boldsymbol{\beta}_0) - g^*(Z_{n+1}^T\boldsymbol{\beta}_0)]^2|\mathcal{S}) \\
&+Ch^{-2}(\textstyle\sum_{k=1}^{K_n^*}|\hat{b}_k - b_{0k}^*|)^2(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)^T E(ZZ^T)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \\
&= O_p(n^{-\frac{2p}{2p+1}}) + O_p(n^{-2}h^{-5} + n^{-1}h^{2p-4}) = O_p(n^{-\frac{2p}{2p+1}}).
\end{aligned}
$$

Using a Taylor expansion, Theorem 3.2 and (A.70), we also obtain

$$
\begin{aligned}
E([g^*(Z_{n+1}^T\hat{\boldsymbol{\beta}}) - g(Z_{n+1}^T\boldsymbol{\beta}_0]^2|\mathcal{S}) &\leq 2E([g^*(Z_{n+1}^T\boldsymbol{\beta}_0) - g(Z_{n+1}^T\boldsymbol{\beta}_0)]^2|\mathcal{S}) \\
&+C(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)^T E(ZZ^T)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = O_p(h^{2p}).
\end{aligned}
$$

Hence, $E([\hat{g}(Z_{n+1}^T\hat{\boldsymbol{\beta}}) - g(Z_{n+1}^T\boldsymbol{\beta}_0)]^2|\mathcal{S}) = O_p(n^{-2p/(2p+1)})$. Now (3.15) follows from (A.73), (A.74) and Theorem 3.2. This completes the proof of Theorem 3.5.