Supplementary of "Residual-based Alternative Partial Least Squares for Generalized Functional Linear Models"

Yue Wang[1], Xiao Wang[2], Joseph G. Ibrahim[3], and Hongtu Zhu[3]

[1]*Department of Biostatistics and Informatics, Colorado School of Public Health*

[2]*Department of Statistics, Purdue University*

[3]*Department of Biostatistics, University of North Carolina at Chapel Hill*

Theory for the PFLM

We first consider the theory for the PFLM $y_i = \mathbf{z}_i^\intercal \boldsymbol{\alpha} + \int x_i(s)b(s)ds + \epsilon_i$. The results established in this section will be used as lemmas for proving the theorems in the main paper.

We first recall the assumptions made in the main paper.

(A1) $\|b^*\|+\mathbb{E}(\|x\|^4) < \infty$ and $\lambda_{\min}(\mathbb{E}(\mathbf{z}^{\otimes 2})) > 0$, where $\lambda_{\min}(M)$ denotes the smallest eigenvalue of any symmetric matrix M.

(A2) $\|\mathcal{C}\| < 1$ and $p = O(n^{1/2})$ as $n \to \infty$.

The following result establishes the asymptotic properties of $\widehat{b}_p(\cdot)$ under the PFLM.

Theorem S 1. Suppose Assumptions (A1) and (A2) hold. As $n \to \infty$, if $n^{-1/2}\lambda_p^{-2} = o(1)$, then the RAPLS estimate $\widehat{b}_p(\cdot)$ satisfies

$$\|\widehat{b}_p - b^*\| = O_p(n^{-1/2}\lambda_p^{-2}). \tag{1}$$

To prove Theorem S1, we first introduce the following lemma that characterizes the convergence rate of the empirical kernel $\widehat{\mathcal{C}}(s,t)$.

Lemma 1. *Suppose Assumption (A2) holds. Then, we have*

$$\widehat{\mathcal{C}}(s,t) = \mathcal{C}(s,t) + n^{-1/2}A(s,t),$$

where $\|A\|_F = O_p(1)$ *as* $n \to \infty$.

Proof. First, we write

$$\begin{aligned}
&\widehat{\mathcal{C}}(s,t) - \mathcal{C}(s,t) = n^{-1}\mathbf{X}(s)^\intercal\mathbf{X}(t) - \mathbb{E}\{x(s)x(t)\} \\
&+ \left[n^{-1}\mathbf{X}(s)^\intercal\mathbf{Z} - \mathbb{E}\{\mathbf{z}^\intercal x(s)\}\right](n^{-1}\mathbf{Z}^\intercal\mathbf{Z})^{-1}n^{-1}\mathbf{Z}^\intercal\mathbf{X}(t) + \\
&+ n^{-1}\mathbf{X}(s)^\intercal\mathbf{Z}\left[(n^{-1}\mathbf{Z}^\intercal\mathbf{Z})^{-1} - \{\mathbb{E}(\mathbf{z}^{\otimes 2})\}^{-1}\right]n^{-1}\mathbf{Z}^\intercal\mathbf{X}(t) + \\
&+ n^{-1}\mathbf{X}(s)^\intercal\mathbf{Z}(n^{-1}\mathbf{Z}^\intercal\mathbf{Z})^{-1}[n^{-1}\mathbf{Z}^\intercal\mathbf{X}(t) - \mathbb{E}\{\mathbf{z}x(t)\}] \\
&= I_1(s,t) + I_2(s,t) + I_3(s,t) + I_4(s,t).
\end{aligned}$$

By Assumption (A2) and the central limit theorem (CLT), we get $I_1(s,t) = O_p(n^{-1/2})$ for any fixed s and t. Since q (the dimension of $\mathbf{z}$) is fixed, the CLT yields that $\|n^{-1}\mathbf{X}(s)^\intercal\mathbf{Z} - \mathbb{E}\{\mathbf{z}^\intercal x(s)\}\| = O_p(n^{-1/2})$ for any fixed $s \in \mathcal{S}$.

This also indicates that $\|n^{-1}\mathbf{X}(s)^\intercal\mathbf{Z}\|\leq \|\mathbb{E}[\mathbf{z}^\intercal x(s)]\|+O_p(n^{-1/2}) = O_p(1)$ for any fixed s. Similarly, the CLT yields that

$$\|n^{-1}\mathbf{Z}^\intercal\mathbf{Z} - \mathbb{E}(\mathbf{z}^{\otimes 2})\|_2= O_p(n^{-1/2}),$$

where $\mathbf{z}^{\otimes 2} = \mathbf{z}\mathbf{z}^\intercal$. Then, using the matrix variant of Taylor series, we write

$$(n^{-1}\mathbf{Z}^\intercal\mathbf{Z})^{-1} = \left\{\mathbb{E}(\mathbf{z}^{\otimes 2}) + n^{-1}\mathbf{Z}^\intercal\mathbf{Z} - \mathbb{E}(\mathbf{z}^{\otimes 2})\right\}^{-1}$$
$$= \left[\mathbf{I}_q + \{\mathbb{E}(\mathbf{z}^{\otimes 2})\}^{-1}\{n^{-1}\mathbf{Z}^\intercal\mathbf{Z} - \mathbb{E}(\mathbf{z}^{\otimes 2})\}\right]^{-1}\{\mathbb{E}(\mathbf{z}^{\otimes 2})\}^{-1}$$

Since $\|n^{-1}\mathbf{Z}^\intercal\mathbf{Z} - \mathbb{E}(\mathbf{z}^{\otimes 2})\|_2= O_p(n^{-1/2})$ and $\lambda_{\min}(\mathbb{E}(\mathbf{z}^{\otimes 2})) > 0$, one can write

$$(n^{-1}\mathbf{Z}^\intercal\mathbf{Z})^{-1} = \sum_{k=0}^{\infty}(-1)^k\left[\{\mathbb{E}(\mathbf{z}^{\otimes 2})\}^{-1}\{n^{-1}\mathbf{Z}^\intercal\mathbf{Z} - \mathbb{E}(\mathbf{z}^{\otimes 2})\}\right]^k\{\mathbb{E}(\mathbf{z}^{\otimes 2})\}^{-1}.$$

Thus, as $n \to \infty$

$$\|(n^{-1}\mathbf{Z}^\intercal\mathbf{Z})^{-1} - \{\mathbb{E}(\mathbf{z}^{\otimes 2})\}^{-1}\|_2\leq \sum_{k=1}^{\infty}\|\{\mathbb{E}(\mathbf{z}^{\otimes 2})\}^{-1}\{n^{-1}\mathbf{Z}^\intercal\mathbf{Z} - \mathbb{E}(\mathbf{z}^{\otimes 2})\}\|_2^k$$
$$\leq \frac{\|\{\mathbb{E}(\mathbf{z}^{\otimes 2})\}^{-1}\{n^{-1}\mathbf{Z}^\intercal\mathbf{Z} - \mathbb{E}(\mathbf{z}^{\otimes 2})\}\|_2}{1 - \|\{\mathbb{E}(\mathbf{z}^{\otimes 2})\}^{-1}\{n^{-1}\mathbf{Z}^\intercal\mathbf{Z} - \mathbb{E}(\mathbf{z}^{\otimes 2})\}\|_2} = O_p(n^{-1/2}),$$

indicating $\|\{n^{-1}\mathbf{Z}^\intercal\mathbf{Z}\}^{-1}\|_2= O_p(1)$.

Combining all the assertions above, one can easily see that for any fixed s and t, we have $I_j(s,t) = O_p(n^{-1/2})$ for $j = 1,\ldots,4$. Finally, since $\mathcal{S}$ is a bounded set, we have $\|I_j\|_F= O_p(n^{-1/2})$ for $j = 1,\ldots,4$. as $n \to \infty$. This completes the proof. □

For any square integrable function f on $\mathcal{S}$, we define $A(f) : f \to \int_{\mathcal{S}} A(s,t)f(t)dt$, where $A(s,t)$ is defined in Lemma 1. Also, denote $\zeta(t) =$

$n^{-1/2}\mathbf{X}^\intercal(t)M_Z\boldsymbol{\epsilon}$, where $M_Z = I_n - \mathbf{Z}(\mathbf{Z}^\intercal\mathbf{Z})^{-1}\mathbf{Z}^\intercal$. The following result characterizes the distance between the population and empirical RAPLS basis functions.

Lemma 2. *Suppose Assumptions (A1) and (A2) hold. For $j \geq 1$,*

$$\widehat{\mathcal{C}}^j(b)(t) - \mathcal{C}^j(b^*)(t) = n^{-1/2}\xi_j(t) + n^{-1}\eta_j(t), \tag{2}$$

where

$$\xi_{j+1}(t) = \sum_{k=0}^{j} \mathcal{C}^k \left(A\left(\mathcal{C}^{j-k}(b^*)\right)\right)(t) + \mathcal{C}^j(\zeta)(t) \text{ for } j \geq 0, \tag{3}$$

$\eta_1(t) = 0$, *and*

$$\eta_{j+2}(t) = \sum_{k=0}^{j} (\mathcal{C} + n^{-1/2}A)^k \left(A(\xi_{j+1-k})\right)(t) \text{ for } j \geq 0. \tag{4}$$

Proof. Using Lemma 1, it is easy to see that $\widehat{\mathcal{C}}(b)(t) - \mathcal{C}(b^*)(t) = n^{-1/2}\xi_1(t) + n^{-1}\eta_1(t)$ with $\xi_1(t) = A(b^*)(t) + \zeta(t)$ and $\eta_1(t) = 0$. To show (3) and (4) for arbitrary $j \geq 0$, we first write

$$\widehat{\mathcal{C}}^{j+2}(b)(t) = \int_{\mathcal{S}} \{\mathcal{C}(s,t) + n^{-1/2}A(s,t)\}\{\mathcal{C}^{j+1}(b^*)(s) + n^{-1/2}\xi_{j+1}(s) + n^{-1}\eta_{j+1}(s)\}ds$$

$$= \mathcal{C}^{j+2}(b^*)(t) + n^{-1/2}\{A(\mathcal{C}^{j+1}(b^*))(t) + \mathcal{C}(\xi_{j+1})(t)\} + n^{-1}\{A(\xi_{j+1})(t) + (\mathcal{C} + n^{-1/2}A)(\eta_{j+1})(t)\}.$$

This indicates that

$$\xi_{j+2}(t) = A(\mathcal{C}^{j+1}(b^*))(t) + \mathcal{C}(\xi_{j+1})(t), \; \eta_{j+2}(t) = A(\xi_{j+1})(t) + (\mathcal{C} + n^{-1/2}A)(\eta_{j+1})(t).$$

Now we suppose (3) holds for $j+1$. We verify it for $j+2$. Using the equation above, we have

$$\begin{aligned}\xi_{j+2}(t) &= A(\mathcal{C}^{j+1}(b^*))(t) + \sum_{k=0}^{j} \mathcal{C}^{k+1}\left(A\left(\mathcal{C}^{j-k}(b^*)\right)\right)(t) + \mathcal{C}^{j+1}(\zeta)(t) \\ &= \sum_{k=0}^{j+1} \mathcal{C}^{k}\left(A\left(\mathcal{C}^{j+1-k}(b^*)\right)\right)(t) + \mathcal{C}^{j+1}(\zeta)(t);\end{aligned}$$

this proves (3). To show (4), we first notice that $\eta_2(t) = A(\xi_1)(t)$ satisfies (4). Now suppose (4) holds for $j+1$. Thus,

$$\begin{aligned}\eta_{j+2}(t) &= A(\xi_{j+1})(t) + (\mathcal{C} + n^{-1/2}A)\left(\eta_{j+1}\right)(t) \\ &= \sum_{k=0}^{j} (\mathcal{C} + n^{-1/2}A)^k \left(A(\xi_{j+1-k})\right)(t);\end{aligned}$$

this proves (4), which completes the proof.

□

Recall that $\widehat{h}_{jk} = \int_{\mathcal{S}} \widehat{\mathcal{C}}^{j+1}(b)(s)\widehat{\mathcal{C}}^{k}(b)(s)ds$ and $\widehat{\beta}_j = \int_{\mathcal{S}} \widehat{\mathcal{C}}(b)(s)\widehat{\mathcal{C}}^{j+1}(b)(s)ds$. The next result characterizes the convergence rate of $\widehat{h}_{jk}$ and $\widehat{\beta}_j$.

Lemma 3. *Suppose Assumptions (A2) and (A3) hold. We have*

$$\begin{aligned}\widehat{h}_{jk} &= h^*_{jk} + n^{-1/2}\int_{\mathcal{S}} \{\mathcal{C}^k(b^*)(s)\xi_{j+1}(s) + \mathcal{C}^{j+1}(b^*)(s)\xi_k(s)\}ds \\ &\quad + O_p\left(n^{-1}j^2k^2\|\mathcal{C}\|^{j+k-1}\right) \qquad (5)\end{aligned}$$

and

$$\widehat{\beta}_j = \beta_j^* + n^{-1/2} \int_{\mathcal{S}} \{\xi_1(s)\mathcal{C}^k(b^*)(s) + \xi_j(s)\mathcal{C}(b^*)(s)\} ds + O_p\left(n^{-1} j^2 \|\mathcal{C}\|^j\right), \tag{6}$$

where $\xi_j(\cdot)$ *is defined in Lemma 2.*

Proof. We first write

$$\begin{aligned}
\widehat{h}_{jk} = h_{jk}^* &+ n^{-1/2} \int_{\mathcal{S}} \left\{\xi_{j+1}(s)\mathcal{C}^k(b^*)(s) + \xi_k(s)\mathcal{C}^j(b^*)(s)\right\} ds \\
&+ n^{-1} \int_{\mathcal{S}} \left\{\xi_{j+1}(s)\xi_k(s) + \mathcal{C}^k(b^*)(s)\eta_{j+1}(s) + \mathcal{C}^{j+1}(b^*)(s)\eta_k(s)\right\} ds \\
&+ n^{-3/2} \int_{\mathcal{S}} \{\xi_{j+1}(s)\eta_k(s) + \xi_k(s)\eta_{j+1}(s)\} ds + n^{-2} \int_{\mathcal{S}} \eta_{j+1}(s)\eta_k(s) ds \\
&= h_{jk}^* + n^{-1/2} \int_{\mathcal{S}} \left\{\xi_{j+1}(s)\mathcal{C}^k(b^*)(s) + \xi_k(s)\mathcal{C}^j(b^*)(s)\right\} ds + n^{-1}R_{1,jk} + n^{-3/2}R_{2,jk} + n^{-2}R_{3,jk}.
\end{aligned}$$

It follows from (3) in Lemma 2 that for all $j \geq 0$, with probability approaching 1, as $n \to \infty$,

$$\|\xi_{j+1}\| \leq C_\xi \|\mathcal{C}\|^j (j + 1 + \|\zeta\|) \tag{7}$$

for some uniform constant C_ξ for all j, where $\zeta(\cdot) = n^{-1/2}\mathbf{X}^\intercal(\cdot)M_Z\boldsymbol{\epsilon}$ is defined in previous discussions. Here, we also use $\|A\|_F = O_p(1)$ and $\|b^*\| <$

∞. For $j \geq 1$, it follows from (4) that

$$\begin{aligned}
\|\eta_{j+1}\| &\leq \|A\| \sum_{k=0}^{j-1} \|\mathcal{C} + n^{-1/2} A\|^k \|\xi_{j-k}\| \\
&\leq C_\xi \|A\| \sum_{k=0}^{j-1} \|\mathcal{C} + n^{-1/2} A\|^k \|\mathcal{C}\|^{j-k} (j - k + \|\zeta\|) \\
&\leq C_\xi \|\mathcal{C}\|^j \sum_{k=0}^{j} (1 + n^{-1/2} \|A\| \|\mathcal{C}\|^{-1})^k (j - k + \|\zeta\|). \qquad (8)
\end{aligned}$$

Since $\|\mathcal{C}\| < 1$, $\|A\| = O_p(1)$, and $k \leq p = O(n^{1/2})$, then as $n \to \infty$, we have $(1 + n^{-1/2} \|A\| \|\mathcal{C}\|)^k = O_p(1)$. This indicates that

$$\|\eta_{j+1}\| \leq C_\xi \|\mathcal{C}\|^j j (j + \|\zeta\|)$$

uniformly for all j.

Next, we bound $\widehat{h}_{jk} - h^*_{jk}$. Note that as $n \to \infty$, with probability approaching 1, we have

$$\begin{aligned}
|R_{1,jk}| &\leq \|\xi_{j+1}\| \|\xi_k\| + \|\mathcal{C}\|^k \|\eta_{j+1}\| \|b^*\| + \|\mathcal{C}\|^{j+1} \|\eta_k\| \|b^*\| \\
&\leq C_\xi \|\mathcal{C}\|^{j+k-1} \{(j + \|\zeta\|)(k + \|\zeta\|) + j(j + \|\zeta\|) + k(k + \|\zeta\|)\};
\end{aligned}$$

here, we use $\|\mathcal{C}\| < 1$. Similarly, as $n \to \infty$, with probability approaching 1,

$$\begin{aligned}
|R_{2,jk}| &\leq \|\xi_{j+1}\| \|\eta_k\| + \|\xi_k\| \|\eta_{j+1}\| \\
&\leq C_\xi \|\mathcal{C}\|^{j+k-1} (j + k)(j + \|\zeta\|)(k + \|\zeta\|)
\end{aligned}$$

and $|R_{3,jk}| \leq \|\eta_{j+1}\| \|\eta_k\| \leq C_\xi \|\mathcal{C}\|^{j+k-1} jk(j + \|\zeta\|)(k + \|\zeta\|)$ uniformly for

$j, k \leq p$. Therefore, we have

$$\left|\widehat{h}_{jk} - h^*_{jk} + n^{-1/2}\int_{\mathcal{S}}\{\xi_{j+1}(s)\mathcal{C}^k(b^*)(s) + \xi_k(s)\mathcal{C}^j(b^*)(s)\}ds\right|$$
$$= O_p\left(n^{-1}jk(j+\|\zeta\|)(k+\|\zeta\|)\|\mathcal{C}\|^{j+k-1}\right); \tag{9}$$

Similarly, for $\beta^*_j = \int_{\mathcal{S}}\mathcal{C}(b^*)(s)\mathcal{C}^j(b^*)(s)ds = h^*_{0,j}$, we get

$$\left|\widehat{\beta}_j - \beta^*_j - n^{-1/2}\int_{\mathcal{S}}\{\xi_1(s)\mathcal{C}^j(b^*)(s) + \xi_j(s)\mathcal{C}(b^*)(s)\}ds\right| = O_p\left\{n^{-1}j(j+\|\zeta\|)(1+\|\zeta\|)\|\mathcal{C}\|^j\right\}. \tag{10}$$

For the PFLM, using the CLT, we get $\|\zeta\| = O_p(1)$ where $\zeta(t) = n^{-1/2}\mathbf{X}^\intercal(t)M_Z\boldsymbol{\epsilon}$, which completes the proof. □

The next result provides asymptotic results for $\widehat{\boldsymbol{\gamma}}$, where $\widehat{\boldsymbol{\gamma}} = \widehat{\mathbf{H}}^{-1}\widehat{\boldsymbol{\beta}}$. For short-hand notations, denote $\Delta_{1jk} = n^{-1/2}\int_{\mathcal{S}}\left\{\xi_{j+1}(s)\mathcal{C}^k(b^*)(s) + \xi_k(s)\mathcal{C}^j(b^*)(s)\right\}ds$, $\Delta_{2,jk} = \widehat{h}_{jk} - h^*_{jk} - \Delta_{1jk}$, and recall $\lambda_p = \lambda_{\min}(\mathbf{H}^*)$. Also, denote $\Delta_1 = (\Delta_{1,jk})_{jk}, \Delta_2 = (\Delta_{2,jk})_{jk}$, and $\boldsymbol{\delta} = (\Delta_{1,01}, \cdots, \Delta_{1,0p})^\intercal$.

Lemma 4. *Suppose Assumptions (A1) and (A2) hold. If $n^{-1/2}\lambda_p^{-2} = o(1)$ as $n \to \infty$, then we have*

$$\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*\| = O_p\left(n^{-1/2}\lambda_p^{-2}(1 + n^{-1/2}\lambda_p^{-1})\right). \tag{11}$$

Proof. Since $\Delta = \Delta_1 + \Delta_2$, we know that $\|\Delta\| \leq \|\Delta_1\| + \|\Delta_2\|$. Following the proof of Lemma 3, we get

$$|\Delta_{1jk}|= O_p\left(n^{-1/2}\|\mathcal{C}\|^{j+k}(j+k+\|\zeta\|)\right),\ |\Delta_{2,jk}|= O_p\left(n^{-1}jk(j+\|\zeta\|)(k+\|\zeta\|)\|\mathcal{C}\|^{j+k}\right) \tag{12}$$

uniformly for $j, k \leq p = O(n^{1/2})$. Since $\|\mathcal{C}\|< 1$ and $p = O(n^{1/2})$, using Theorem 3.3 in Rudin et al. (1976), we have $\|\Delta_1\|_F= O_p\{n^{-1/2}(1+\|\zeta\|)\}$ and $\|\Delta_2\|_F= O_p\{n^{-1}(1+\|\zeta\|^2)\}$. This indicates that $||\Delta||= O_p\{n^{-1/2}(1+\|\zeta\|)+n^{-1}\|\zeta\|^2\}$. Under the PFLM, since $\|\zeta\|= O_p(1)$, we have $\|\Delta\|= O_p(n^{-1/2})$. Since $n^{-1/2}\lambda_p^{-2} = o(1)$, we have $||\Delta||/\lambda_p < 1$ with probability approaching 1 as $n \to \infty$. Then, using the matrix variant of the Taylor series, we get

$$\widehat{\mathbf{H}}^{-1} = (\mathbf{I}_p + \mathbf{H}^{*-1}\Delta)^{-1}\mathbf{H}^{*-1} = \left\{\mathbf{I}_p - \mathbf{H}^{*-1}\Delta + O_p(||\Delta||^2/\lambda_p^2)\right\}\mathbf{H}^{*-1}, \tag{13}$$

where with potential misuse of notations, $O_p(||\Delta||^2/\lambda_p^2)$ denotes a matrix with its Frobenious norm bounded by $O_p(||\Delta||^2/\lambda_p^2)$. This further indicates

$$\widehat{\mathbf{H}}^{-1} - \mathbf{H}^{*-1} = -\mathbf{H}^{*-1}\Delta\mathbf{H}^{*-1} + \mathbf{R}_{\mathbf{H}},$$

where $\mathbf{R}_{\mathbf{H}} = O_p(\lambda_p^{-3}\|\Delta\|^2) = O_p\left\{\lambda_p^{-3}n^{-1}\left(1+\|\zeta\|+n^{-1/2}\|\zeta\|^2\right)^2\right\}$.

Similar arguments yield that $\|\boldsymbol{\delta}\|= O_p\{n^{-1/2}(1+\|\zeta\|)\}$ and $\|\boldsymbol{\theta}\|= O_p\{n^{-1}(1+\|\zeta\|^2)\}$, where $\boldsymbol{\theta} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* - \boldsymbol{\delta}$. Also, note that

$$\|\boldsymbol{\beta}^*\|\leq \sum_{j=1}^{p}\|\mathcal{C}\|^{j+1}\|b^*\|^2= O(1).$$

and $\|\boldsymbol{\gamma}^*\| \leq \lambda_p^{-1}\|\boldsymbol{\beta}^*\| \leq \lambda_p^{-1}\sum_{j=1}^p \|\mathcal{C}\|^{j+1}\|b^*\|^2 = O(\lambda_p^{-1})$. Therefore, combining all the assertions above, we get

$$\begin{aligned}
\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*\| \leq & |\mathbf{H}^{*-1}(\boldsymbol{\delta}+\boldsymbol{\theta})\| + \|\mathbf{H}^{*-1}\Delta\mathbf{H}^{*-1}(\boldsymbol{\delta}+\boldsymbol{\theta})\| + \|\mathbf{H}^{*-1}\Delta\boldsymbol{\gamma}^*\| \\
& + \|\mathbf{R_H}(\boldsymbol{\delta}+\boldsymbol{\theta})\| + \|\mathbf{R_H}\boldsymbol{\beta}^*\| = O_p\left\{n^{-1/2}\lambda_p^{-1}\left(1+\|\zeta\|+n^{-1/2}\|\zeta\|^2\right)\right\} \\
& + O_p\left\{n^{-1}\lambda_p^{-2}\left(1+\|\zeta\|+n^{-1/2}\|\zeta\|^2\right)^2\right\} \\
& + O_p\left\{n^{-1/2}\lambda_p^{-2}\left(1+\|\zeta\|+n^{-1/2}\|\zeta\|^2\right)\right\} \\
& + O_p\left\{n^{-3/2}\lambda_p^{-3}\left(1+\|\zeta\|+n^{-1/2}\|\zeta\|^2\right)^3\right\} \\
& + O_p\left\{n^{-1}\lambda_p^{-3}\left(1+\|\zeta\|+n^{-1/2}\|\zeta\|^2\right)\right\}.
\end{aligned} \tag{14}$$

For the PFLM, taking $\|\zeta\| = O_p(1)$ completes the proof. □

We now prove Theorem S1 based on previous lemmas.

Proof. First, we write

$$\begin{aligned}
\widehat{b}_p(s) &= \sum_{j=1}^p \left\{\mathcal{C}^j(b^*)(s) + n^{-1/2}\xi_j(s) + n^{-1}\eta_j(s)\right\}(\gamma_j + \widehat{\gamma}_j - \gamma_j^*) \\
&= b_p^*(s) + n^{-1/2}\sum_{j=1}^p \xi_j(s)\widehat{\gamma}_j + n^{-1}\sum_{j=1}^p \eta_j(s)\widehat{\gamma}_j + \sum_{j=1}^p \mathcal{C}^j(b^*)(s)(\widehat{\gamma}_j - \gamma_j^*)
\end{aligned}$$

Using Lemmas 2-4 and $\|\mathcal{C}\| < 1$, we have

$$\begin{aligned}
& \|n^{-1/2}\sum_{j=1}^p \xi_j(s)\widehat{\gamma}_j\| \leq O(n^{-1/2}) \times (\|\boldsymbol{\gamma}^*\| + \|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*\|)\sum_{j=1}^p j\|\mathcal{C}\|^{j-1} \\
&= O_p\left(n^{-1/2}\lambda_p^{-1}\right) + O_p\left(n^{-1}\lambda_p^{-2}(1+n^{-1/2}\lambda_p^{-1})\right).
\end{aligned} \tag{15}$$

Since $\lambda_p^{-2}n^{-1/2} = o(1)$, the RHS of (15) can be simplified as $O_p(n^{-1/2}\lambda_p^{-1})$.

Similarly, we get

$$\|n^{-1}\sum_{j=1}^{p}\eta_j(s)\widehat{\gamma}_j\|= O_p\left(n^{-1}\lambda_p^{-1}\right),$$

$$\|\sum_{j=1}^{p}\mathcal{C}^j(b^*)(s)(\widehat{\gamma}_j-\gamma_j^*)\|= O_p(n^{-1/2}\lambda_p^{-2}).$$

Therefore, we have

$$\|\widehat{b}_p - b_p^*\|= O_p(n^{-1/2}\lambda_p^{-2}). \tag{16}$$

□

Proof for Theorem 1 in the main paper

In this section, we prove Theorem 1 in the main paper based on Lemmas 1-4.

We first rewrite the pseudo response $\tilde{y}_i^{(m)}$ as

$$\tilde{y}_i^{(m)} = \eta_i^* + \eta_i^{(m)} - \eta_i^* + \{w_i^{(m)}\}^{-1}r_i^{(m)}.$$

Denoting $\delta_i^{(m)} = \eta_i^{(m)} - \eta_i^* + \{w_i^{(m)}\}^{-1}r_i^{(m)}$, we get

$$\tilde{y}_i^{(m)} = \eta_i^* + \delta_i^{(m)}.$$

Thus, this model has the same form as the PFLM studied in Section 1 except that $\mathbb{E}[\delta_i^{(m)} \mid \mathbf{z}_i, x_i(\cdot)]$ is not necessarily 0.

Let $\boldsymbol{\delta}^{(m)} = (\delta_1^{(m)}, \ldots, \delta_n^{(m)})^\intercal$ and $\varepsilon^{(m)}(t) = n^{-1/2}\mathbf{X}^\intercal(t)M_Z\boldsymbol{\delta}^{(m)}$. It can be seen from Lemmas 1-4 that the bound of $\|b^{(m+1)} - b_p^*\|$ depends on $\|\varepsilon^{(m)}\|$.

In particular, if $\|\varepsilon^{(m)}\|= O_p(1)$, which is the case under the PFLM, then $\|b^{(m+1)} - b^*_p\|= O_p(n^{-1/2}\lambda_p^{-2})$.

We first consider deterministic initial values $\alpha_0^{(0)}, \boldsymbol{\alpha}^{(0)}$, and $b^{(0)}(\cdot)$ for the iterative RAPLS algorithms. The next lemma links $\|\varepsilon^{(0)}\|$ with $\{\mathbb{E}\left(\eta^{(0)} - \eta^*\right)^4\}^{1/2}$.

Lemma 5. *Suppose Assumptions (A1) and (A2) hold. Let $\eta^{(0)}$ distribute like every $\eta_i^{(0)}$. Then, we have* $\|\varepsilon^{(0)}\|\leq O_p(\sqrt{n})\left\{\mathbb{E}\left(\eta^{(0)} - \eta^*\right)^4\right\}^{1/2}$.

Proof. Let $\delta^{(0)}$ distribute like every $\delta_i^{(0)}$. Since $\varepsilon^{(0)}(t) = n^{-1/2}\mathbf{X}^{\intercal}(t)\mathbf{Z}(\mathbf{Z}^{\intercal}\mathbf{Z})^{-1}\mathbf{Z}^{\intercal}\boldsymbol{\delta}^{(0)}$, the CLT yields that, as $n\to\infty$,

$$\|\varepsilon^{(0)}\|\leq O_p(\sqrt{n}) \times \lambda_{\min}^{-1}(\mathbb{E}[\mathbf{z}^{\otimes 2}])(\mathbb{E}\|x\|^2)^{1/2}(\mathbb{E}\|\mathbf{z}\|^2)^{1/2}\|\mathbb{E}\{\delta^{(0)}\mathbf{z}\}\|. \tag{17}$$

Thus, under Assumptions (A1) and (A2), we get

$$\|\varepsilon^{(0)}\|= O_p(\sqrt{n}) \times \|\mathbb{E}\{\delta^{(0)}\mathbf{z}\}\|.$$

Recall that $\delta^{(0)} = \eta^{(0)} - \eta^* + \{w^{(0)}\}^{-1}r^{(0)}$. Letting $r^* = r(y, \eta^*) = T(y) - \dot{A}(\eta^*)$, we have $\mathbb{E}[r^*|\mathbf{z}, x(\cdot)] = 0$. Thus, using the Taylor expansion, we have

$$\begin{aligned}\mathbb{E}\{\delta^{(0)}\mathbf{z}\} &= \mathbb{E}\left\{(\eta^{(0)} - \eta^*)\mathbf{z}\right\} + \mathbb{E}\left\{(w^{(0)})^{-1}(r^{(0)} - r^*)\mathbf{z}\right\}\\ &= \mathbb{E}\left[\left\{1 - (w^{(0)})^{-1}\widetilde{w}^{(0)}\right\}(\eta^{(0)} - \eta_i^*)\mathbf{z}\right]\\ &= \mathbb{E}\left\{(w^{(0)})^{-1}\dot{w}(\check{\eta}^{(0)})(\widetilde{\eta}^{(0)} - \eta^{(0)})(\eta^{(0)} - \eta^*)\mathbf{z}\right\};\end{aligned} \tag{18}$$

here, $\widetilde{w}^{(0)} = w(\widetilde{\eta}^{(0)})$ with $\widetilde{\eta}^{(0)}$ located on the line segment between $\eta^{(0)}$ and η^*; $\check{\eta}^{(0)}$ locates on the line segment between $\widetilde{\eta}^{(0)}$ and $\eta^{(0)}$; $\dot{w}(\cdot)$ is the derivative of $w(\cdot)$. Note that $w(\cdot) = \ddot{A}(\cdot)$ is a smooth positive function, and is bounded away from 0 on any bounded set; also, $\dot{w}(\cdot)$ is bounded on any bounded set.

Therefore, using the Cauchy-Schwarz inequality, we get

$$\|\mathbb{E}[\delta^{(0)}\mathbf{z}]\| \leq O(1) \times \left\{\mathbb{E}[(\eta^{(0)} - \eta^*)^4]\right\}^{1/2}$$

which completes the proof. □

Since $\eta^{(0)} = \alpha_0^{(0)} + \mathbf{z}^\intercal \boldsymbol{\alpha}^{(0)} + \int_{\mathcal{S}} x(s) b^{(0)}(s) ds$, we also get

$$|\eta^{(0)} - \eta^*| \leq |\alpha_0^{(0)} - \alpha_0^*| + \|\mathbf{z}\| \|\boldsymbol{\alpha}^{(0)} - \boldsymbol{\alpha}^*\| + \|x\| \|b^{(0)} - b^*\|. \tag{19}$$

We next prove Proposition 1 in the main text.

Proof. Since $\|b^{(0)} - b^*\| = O(1)$, $\|\boldsymbol{\alpha}^{(0)} - \boldsymbol{\alpha}^*\| = O(1)$, and $\|\alpha_0^{(0)} - \alpha_0^*\| = O(1)$, we have $\mathbb{E}(\eta^{(0)} - \eta^*)^4 = O(1)$ as $n \to \infty$. Thus, $\|\varepsilon^{(0)}\| = O_p(\sqrt{n})$, indicating that $1 + \|\varepsilon^{(0)}\| + n^{-1/2} \|\varepsilon^{(0)}\|^2 \asymp \|\varepsilon^{(0)}\|$. Thus, one can derive from (14) that

$$\|\boldsymbol{\gamma}^{(1)} - \boldsymbol{\gamma}^*\| = O_p(1) \times \left[\lambda_p^{-2} \left\{\mathbb{E}(\eta^{(0)} - \eta^*)^4\right\}^{1/2} + \sum_{k=1}^{3} \left(\lambda_p^{-1} \left\{\mathbb{E}(\eta^{(0)} - \eta^*)^4\right\}^{1/2}\right)^k\right]$$

Following the same derivations as those in (15) and (16), we can get

$$\begin{aligned} \|b^{(1)} - b_p^*\| &= O_p(1) \\ &\times \left[\lambda_p^{-2} \left\{\mathbb{E}(\eta^{(0)} - \eta^*)^4\right\}^{1/2} + \sum_{k=1}^{3} \left(\lambda_p^{-1} \left\{\mathbb{E}(\eta^{(0)} - \eta^*)^4\right\}^{1/2}\right)^k\right]. \end{aligned} \tag{20}$$

Therefore, since $\mathbb{E}(\eta^{(0)} - \eta^*)^4 = O(1)$, we get $\|b^{(1)} - b_p^*\| = O_p(\lambda_p^{-3})$. □

Next, we consider data-driven initial values $\alpha_{0,n}^{(0)}, \boldsymbol{\alpha}_n^{(0)}$, and $b_n^{(0)}$. The next result shows the consistency of the first RAPLS iterate $b^{(1)}(\cdot)$.

Proof. Recall from Assumption (A3) that the data driven initial values satisfy $|\alpha_{0,n}^{(0)} - \alpha^*| + \|\boldsymbol{\alpha}_n^{(0)} - \boldsymbol{\alpha}^*\| + \|b_n^{(0)} - b^*\| = O(\tau_n)$. Then, using (19), we get $\left\{\mathbb{E}(\eta^{(0)} - \eta^*)^4\right\}^{1/2} = O(\tau_n^2)$ as $n \to \infty$. Then, since $\lambda_p^{-2}\tau_n = O(1)$, we know $\lambda_p^{-1}\tau_n^2 = o(1)$. Therefore, we get from (20) that

$$\|b^{(1)} - b^*\| = O_p\left(\lambda_p^{-2}\tau_n^2\right). \tag{21}$$

□

We next extend the results to any arbitrary iterate $b^{(m)}(\cdot)$.

Proof. To show this, we first study the asymptotic property of the simple plug-in estimates $\boldsymbol{\alpha}^{(1)}$ and $\alpha_0^{(1)}$. For ease of notation, let $\widetilde{\mathbf{z}}_i = (1, \mathbf{z}_i^\intercal)^\intercal$, $\widetilde{\boldsymbol{\alpha}}^* = (\alpha_0^*, \boldsymbol{\alpha}^{*\intercal})^\intercal$ and $\widetilde{\boldsymbol{\alpha}}^{(1)} = (\alpha_0^{(1)}, \boldsymbol{\alpha}^{(1)\intercal})^\intercal$. Note that $n^{-1}\sum_{i=1}^n r(y_i, \widetilde{\mathbf{z}}_i^\intercal\widetilde{\boldsymbol{\alpha}}^{(1)} + \int_{\mathcal{S}} x_i(s)b^{(1)}(s)ds)\widetilde{\mathbf{z}}_i = \mathbf{0}$ and $\mathbb{E}[r(y_i, \widetilde{\mathbf{z}}_i^\intercal\widetilde{\boldsymbol{\alpha}}^* + \int_{\mathcal{S}} x_i(s)b^*(s)ds)\widetilde{\mathbf{z}}_i] = \mathbf{0}$.

Since $\lambda_p^{-2}\tau_n = O(1)$ and $\tau_n = o(1)$ as $n \to \infty$, we get $\lambda_p^{-2}\tau_n^2 = o(1)$, leading to $\|b^{(1)} - b^*\| = o_p(1)$. Hence, using the law of large numbers, we have $n^{-1}\sum_{i=1}^n r(y_i, \alpha_0 + \mathbf{z}_i^\intercal\boldsymbol{\alpha} + \int_{\mathcal{S}} x_i(s)b^{(1)}(s)ds)\mathbf{z}_i$ converges to $\mathbb{E}[r(y_i, \alpha_0 + \mathbf{z}_i^\intercal\boldsymbol{\alpha} + \int_{\mathcal{S}} x_i(s)b^*(s)ds)\mathbf{z}_i]$ uniformly for all α_0 and $\boldsymbol{\alpha}$ as $n \to \infty$. Thus, by the theory of maximum likelihood estimation (MLE), we know that $\alpha_0^{(1)}$ and $\boldsymbol{\alpha}^{(1)}$ are consistent estimators of α_0^* and $\boldsymbol{\alpha}^*$, respectively. We use the

Taylor expansion to get the convergence rate of $\widetilde{\boldsymbol{\alpha}}_0^{(1)}$. Specifically,

$$\begin{aligned}\mathbf{0} &= n^{-1}\sum_{i=1}^{n} r(y_i, \widetilde{\mathbf{z}}_i^\intercal \widetilde{\boldsymbol{\alpha}}^{(1)} + \int_{\mathcal{S}} x_i(s) b^{(1)}(s) ds)\widetilde{\mathbf{z}}_i \\ &= n^{-1}\sum_{i=1}^{n} r(y_i, \widetilde{\mathbf{z}}_i^\intercal \widetilde{\boldsymbol{\alpha}}^{*} + \int_{\mathcal{S}} x_i(s) b^{*}(s) ds)\widetilde{\mathbf{z}}_i \\ &+ n^{-1}\sum_{i=1}^{n} \ddot{A}(\eta_i^\dagger)\left[\widetilde{\mathbf{z}}_i^\intercal \left(\widetilde{\boldsymbol{\alpha}}^{*} - \widetilde{\boldsymbol{\alpha}}^{(1)}\right) + \int_{\mathcal{S}} x_i(s)\left\{b^*(s) - b^{(1)}(s)\right\} ds\right]\widetilde{\mathbf{z}}_i,\end{aligned}$$

where $\eta_i^\dagger$ is on the line segment between $\widetilde{\mathbf{z}}_i^\intercal \widetilde{\boldsymbol{\alpha}}^{(1)} + \int_{\mathcal{S}} x_i(s) b^{(1)}(s) ds$ and $\widetilde{\mathbf{z}}_i^\intercal \widetilde{\boldsymbol{\alpha}}^{*} + \int_{\mathcal{S}} x_i(s) b^{*}(s) ds$. This leads to

$$\begin{aligned}\widetilde{\boldsymbol{\alpha}}^{(1)} - \widetilde{\boldsymbol{\alpha}}^{*} = &\left\{n^{-1}\sum_{i=1}^{n} \ddot{A}(\eta_i^\dagger)\widetilde{\mathbf{z}}_i\widetilde{\mathbf{z}}_i^\intercal\right\}^{-1} \\ &\times \left[n^{-1}\sum_{i=1}^{n} r(y_i, \widetilde{\mathbf{z}}_i^\intercal \widetilde{\boldsymbol{\alpha}}^{*} + \int_{\mathcal{S}} x_i(s) b^{*}(s) ds)\widetilde{\mathbf{z}}_i \right. \\ &\left. + n^{-1}\sum_{i=1}^{n} \ddot{A}(\eta_i^\dagger)\int_{\mathcal{S}} x_i(s)\left\{b^*(s) - b^{(1)}(s)\right\} ds\widetilde{\mathbf{z}}_i\right]\end{aligned}$$

Since $\widetilde{\boldsymbol{\alpha}}^{(1)}$ and $b^{(1)}$ are consistent, we know $\left|\eta_i^\dagger - \eta_i^*\right| = o_p(1)$. Hence, by the law of large numbers and some matrix algebra, we get

$$\left\|\left\{n^{-1}\sum_{i=1}^{n} \ddot{A}(\eta_i^\dagger)\widetilde{\mathbf{z}}_i\widetilde{\mathbf{z}}_i^\intercal\right\}^{-1} - \left[\mathbb{E}\left\{\ddot{A}(\eta_i)\widetilde{\mathbf{z}}_i\widetilde{\mathbf{z}}_i^\intercal\right\}\right]^{-1}\right\| = o_p(1).$$

Since $\mathbb{E}\left[r(y_i, \widetilde{\mathbf{z}}_i^\intercal \widetilde{\boldsymbol{\alpha}}^{*} + \int_{\mathcal{S}} x_i(s) b^{*}(s) ds)\widetilde{\mathbf{z}}_i\right] = \mathbf{0}$, the CLT yields

$$n^{-1}\sum_{i=1}^{n} r(y_i, \widetilde{\mathbf{z}}_i^\intercal \widetilde{\boldsymbol{\alpha}}^{*} + \int_{\mathcal{S}} x_i(s) b^{*}(s) ds)\widetilde{\mathbf{z}}_i = O_p(n^{-1/2}).$$

Since $\left\|b^{(1)} - b^*\right\| = O_p(\lambda_p^{-2}\tau_n^2)$, by Assumptions (A1)–(A3), we get

$$\left\|n^{-1}\sum_{i=1}^{n} \ddot{A}(\eta_i^\dagger)\int_{\mathcal{S}} x_i(s)\left\{b^*(s) - b^{(1)}(s)\right\} ds\widetilde{\mathbf{z}}_i\right\| = O_p(\tau_n^2\lambda_p^{-2}).$$

This leads to $\|\widetilde{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha}^*\| = O_p(\tau_n^2 \lambda_p^{-2})$. Furthermore, since $\tau_n \lambda_p^{-2} = O(1)$, we know

$$\|\widetilde{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha}^*\| + \|b^{(1)} - b^*\| = O_p(\tau_n),$$

which satisfies Assumption (A3). Therefore, the same arguments can be used for proving $\|b^{(m)} - b^*\| = O_p(\tau_n^2 \lambda_p^{-2})$ and $\|\widetilde{\boldsymbol{\alpha}}^{(m)} - \boldsymbol{\alpha}^*\| = O_p(\tau_n^2 \lambda_p^{-2})$ for $m \geq 2$. □

Proof of Theorem 2 in the main paper

We first prove a supporting lemma regarding the estimation of $\theta_k^*(\cdot)$.

Recall that $\mathbf{z}_i = \int_{\mathcal{S}} x_i(s) \boldsymbol{\theta}^*(s) ds + \boldsymbol{\zeta}_i$ for $i = 1, \ldots, n$, where $\boldsymbol{\theta}^*(\cdot) = \big(\theta_1^*(\cdot), \ldots, \theta_q^*(\cdot)\big)^\intercal$ with $\theta_k^*(\cdot) = K_w^{-1}(\mathbb{E}[w_i^* \zeta_{ik} x_i(\cdot)])$, $\mathbb{E}[\ddot{A}(\eta_i^*) \boldsymbol{\zeta}_i x_i(\cdot)] = \mathbf{0}$ and $\mathbb{E}[\ddot{A}(\eta_i^*) \boldsymbol{\zeta}_i \boldsymbol{\zeta}_i^\intercal] = \widetilde{\Sigma}_\zeta$. Also, define $\theta_{k,s_n}^*(s) = \sum_{j=1}^{s_n} \theta_{kj}^* \pi_j(s)$ as the approximation of $\theta_k^*(\cdot)$ with the first s_n basis functions, where $\theta_{kj}^* = \int \theta_k^*(s) \pi_j(s) ds$.

The following lemma shows that $\widehat{\theta}_k(s)$, constructed from the three-step procedure from the main paper, are consistent estimators with a convergence rate no slower than $n^{-1/4}$.

Lemma 6. *Suppose Assumptions (A1)-(A3) hold. If* $\|\theta_k^*(s) - \theta_{k,s_n}^*(s)\|^2 = O(s_n^{1-2b}), s_n \approx n^a,$ *and* $1/\{2(2b-1)\} < a \leq 1/4,$ *and* $\tau_n^2 \lambda_p^{-2} = o(n^{-1/4}),$ *then* $\|\widehat{\theta}_k(s) - \theta_k^*(s)\| = o_p(n^{-1/4})$.

Proof. We write

$$\left\|\widehat{\theta}_k(s) - \theta^*_k(s)\right\| = \left\|\sum_{j=1}^{s_n}(\widehat{\theta}_{kj} - \theta^*_{kj})\pi_j(s)\right\| + \left\|\sum_{j=s_n+1}^{\infty}\theta^*_{kj}\pi_j(s)\right\|$$
$$\leq \left\|\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}^*_k\right\| + \left(\sum_{j=s_n+1}^{\infty}{\theta^*_{kj}}^2\right)^{1/2}, \tag{22}$$

where $\boldsymbol{\theta}^*_k = (\theta^*_{k1}, \ldots, \theta^*_{k,s_n})^\intercal$ and $\widehat{\boldsymbol{\theta}_k} = (\widehat{\theta}_{k1}, \ldots, \widehat{\theta}_{k,s_n})^\intercal$. Note that

$$\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}^*_k = (U_{s_n}^\intercal \widehat{A} U_{s_n})^{-1} U_{s_n}^\intercal \widehat{A}\left(\sum_{j=s_n+1}^{\infty} U_{(\cdot,j)}\theta^*_{kj} + W_k\right),$$

where $U_{(\cdot,j)} = (U_{1j}, \ldots, U_{nj})^\intercal$ and $W_k = (\zeta_{1k}, \ldots, \zeta_{nk})^\intercal$. Thus, we know

$$\left\|\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}^*_k\right\| \leq \left\|\left(\frac{U_{s_n}^\intercal \widehat{A} U_{s_n}}{n}\right)^{-1}\right\| \sqrt{\sum_{l=1}^{s_n}\left\{\sum_{j=s_n+1}^{\infty} U_{(\cdot,l)}^\intercal \widehat{A} U_{(\cdot,j)}/n\right\}^2} \sqrt{\sum_{j=s_n+1}^{\infty}{\theta^*_{kj}}^2}$$
$$+ \left\|\left(\frac{U_{s_n}^\intercal \widehat{A} U_{s_n}}{n}\right)^{-1}\right\| \sqrt{\sum_{l=1}^{s_n}\left\{U_{(\cdot,l)}^\intercal \widehat{A} W_k/n\right\}^2}$$

Note that

$$\left\|\left(U_{s_n}^\intercal \widehat{A} U_{s_n}/n\right)^{-1} - \left(\mathbb{E}\left[\ddot{A}(\eta^*_1)U_{(1,\cdot)}U_{(1,\cdot)}^\intercal\right]\right)^{-1}\right\|$$
$$\leq \left\|\left(U_{s_n}^\intercal \widehat{A} U_{s_n}/n\right)^{-1} - \left(U_{s_n}^\intercal A U_{s_n}/n\right)^{-1}\right\| + \left\|\left(U_{s_n}^\intercal A U_{s_n}/n\right)^{-1} - \left(\mathbb{E}\left[\ddot{A}(\eta^*_1)U_{(1,\cdot)}U_{(1,\cdot)}^\intercal\right]\right)^{-1}\right\|$$
$$= J_1 + J_2$$

To bound J_1 and J_2, we will rely on the matrix Taylor expansion: for any matrix M and some small addition term Δ_M and get $\|(M+\Delta_M)^{-1} - M^{-1}\| \leq O(1) \times \|\Delta_M\|$. This leads to

$$\|J_1\| = O_p\left(\tau_n^2 \lambda_p^{-2}\sqrt{s_n}\right).$$

For J_2, we combine the matrix Taylor expansion with the central limit theorem and get $\|J_2\|= O_p(\sqrt{s_n/n})$. Noting that for $l, k = 1, \ldots, s_n$, we get

$$\frac{U_{(\cdot,l)}^{\intercal}(\widehat{A}-A)U_{(\cdot,k)}}{n} = O_p(\lambda_p^{-2}\tau_n^2)$$

Then, the same matrix expansion leads to

$$\|J_1\|= O_p\left(\tau_n^2\lambda_p^{-2}\sqrt{s_n}\right)$$

Since $s_n \asymp n^a$ with $a \leq 1/4$, $\tau_n^2\lambda_p^{-2} = o(n^{-1/4})$, we get

$$\left\|\left(U_{s_n}^{\intercal}\widehat{A}U_{s_n}/n\right)^{-1} - \left(\mathbb{E}\left[\ddot{A}(\eta_1^*)U_{(1,\cdot)}U_{(1,\cdot)}^{\intercal}\right]\right)^{-1}\right\| = o_p(1).$$

Similarly, we get

$$\left\|U_{(\cdot,l)}^{\intercal}\widehat{A}\sum_{j=s_n+1}^{\infty}U_{(\cdot,j)}/n - \mathbb{E}\left[\ddot{A}(\eta_1^*)U_{1l}\sum_{j=s_n+1}^{\infty}U_{1j}\right]\right\| = O_p\left(\tau_n^2\lambda_p^{-2}\right) = o_p(n^{-1/4}).$$

Also, since $\|x_1(s)\|^2= \sum_{j=1}^{\infty}U_j^2$ and $\mathbb{E}\|x\|^4< \infty$, we get

$$\begin{aligned}
&\sum_{l=1}^{s_n}\left\{\sum_{j=s_n+1}^{\infty}U_{(\cdot,l)}^{\intercal}\widehat{A}U_{(\cdot,j)}/n\right\}^2 = \sum_{l=1}^{s_n}\left(\mathbb{E}\left[\ddot{A}(\eta_1^*)U_{1l}\sum_{j=s_n+1}^{\infty}U_{1j}\right] + o_p(n^{-1/4})\right)^2 \\
&\leq o_p(s_n n^{-1/4}) + \mathbb{E}\left[\ddot{A}^2(\eta_1^*)\sum_{l=1}^{s_n}U_{1l}^2\left(\sum_{j=s_n+1}^{\infty}U_{1j}\right)^2\right] \\
&\leq o_p(s_n n^{-1/4}) + O_p\left\{\mathbb{E}\left(\sum_{j=s_n+1}^{\infty}U_{1j}\right)^4\right\} = O_p(1).
\end{aligned}$$

Similarly, since $\tau_n^2\lambda_p^{-2} = o(n^{-1/4})$, we get

$$\begin{aligned}
&\left|U_{(\cdot,l)}^{\intercal}\widehat{A}W/n\right| \leq \left|U_{(\cdot,l)}^{\intercal}AW/n\right| + \left|U_{(\cdot,l)}^{\intercal}(\widehat{A}-A)W/n\right| \\
&= O_p(n^{-1/2}) + O_p(\tau_n^2\lambda_p^{-2}) = o_p(n^{-1/4}).
\end{aligned}$$

Hence, we can get

$$\sum_{l=1}^{s_n} \left\{ U_{(\cdot,l)}^{\intercal} \widehat{A} W_k / n \right\}^2 = o_p \left(s_n n^{-1/2} \right).$$

Since

$$\sum_{j=s_n+1}^{\infty} {\theta_{kj}^*}^2 = O(s_n^{1-2b}).$$

and $s_n = n^a$, we get

$$\left\| \widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k^* \right\| = O_p \left(n^{a(1-2b)/2} \right) + o_p \left(n^{a-1/2} \right),$$

which is $o_p(n^{-1/4})$ because $s_n \asymp n^a$ with $1/\{2(2b-1)\} < a \leq 1/4$. Combining all the assertions above concludes the proof. □

For short-hand notation, let $\boldsymbol{\varpi}^*(s) = b^*(s) + \boldsymbol{\alpha}_p^{\intercal} \boldsymbol{\theta}^*(s)$ and $\widehat{\boldsymbol{\varpi}}_p(s) = \widehat{b}_p(s) + \widehat{\boldsymbol{\alpha}}_p^{\intercal} \widehat{\boldsymbol{\theta}}(s)$. Define

$$S_n(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{\varpi}(s)) = n^{-1} \sum_{i=1}^{n} \left\{ T(y_i) - \dot{A} \left(\int_{\mathcal{S}} x_i(s) \boldsymbol{\varpi}(s) ds + \alpha_0 + \boldsymbol{\zeta}_i^{\intercal} \boldsymbol{\alpha} \right) \right\} \boldsymbol{\zeta}_i$$

and

$$S(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{\varpi}(s)) = \mathbb{E} \left[\left\{ T(y_i) - \dot{A} \left(\int_{\mathcal{S}} x_i(s) \boldsymbol{\varpi}(s) ds + \alpha_0 + \boldsymbol{\zeta}_i^{\intercal} \boldsymbol{\alpha} \right) \right\} \boldsymbol{\zeta}_i \right]$$

We conclude this section by proving Theorem 3 in the main paper.

Proof. We first show that $\widehat{\boldsymbol{\alpha}}_p^{\text{cal}}$ is a consistent estimator of $\boldsymbol{\alpha}^*$. Note that $\widehat{\boldsymbol{\alpha}}_p^{\text{cal}}$ is the solution to $S_n(\widehat{\alpha}_{0,p}, \boldsymbol{\alpha}, \widehat{\boldsymbol{\varpi}}_p(s)) = \mathbf{0}$, and $\boldsymbol{\alpha}^*$ is the solution to $S(\alpha_0^*, \boldsymbol{\alpha}, \boldsymbol{\varpi}^*(s)) = \mathbf{0}$. Using the maximum likelihood theory, to establish the consistency of $\widehat{\boldsymbol{\alpha}}_p^{\mathrm{cal}}$, we only need to show that $S_n(\widehat{\alpha}_{0,p}, \boldsymbol{\alpha}, \widehat{\boldsymbol{\varpi}}_p(s)) \rightarrow S(\alpha_0^*, \boldsymbol{\alpha}, \boldsymbol{\varpi}^*(s))$ in probability uniformly for all $\boldsymbol{\alpha}$. More specifically, we write

$$
\begin{aligned}
& S_n(\widehat{\alpha}_{0,p}, \boldsymbol{\alpha}, \widehat{\boldsymbol{\varpi}}_p(s)) - S(\alpha_0^*, \boldsymbol{\alpha}, \boldsymbol{\varpi}^*(s)) \\
&= n^{-1} \sum_{i=1}^{n} \left\{ \dot{A}\left(\int_{\mathcal{S}} x_i(s)\boldsymbol{\varpi}^*(s)ds + \alpha_0^* + \boldsymbol{\zeta}_i^{\intercal}\boldsymbol{\alpha} \right) - \dot{A}\left(\int_{\mathcal{S}} x_i(s)\widehat{\boldsymbol{\varpi}}_p(s)ds + \widehat{\alpha}_{0,p} + \widehat{\boldsymbol{\zeta}}_i^{\intercal}\boldsymbol{\alpha} \right) \right\} \widehat{\boldsymbol{\zeta}}_i \\
&+ n^{-1} \sum_{i=1}^{n} \left\{ T(y_i) - \dot{A}\left(\int_{\mathcal{S}} x_i(s)\boldsymbol{\varpi}^*(s)ds + \alpha_0^* + \boldsymbol{\zeta}_i^{\intercal}\boldsymbol{\alpha} \right) \right\} \left\{ \widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i \right\} \\
&+ n^{-1} \sum_{i=1}^{n} \left\{ T(y_i) - \dot{A}\left(\int_{\mathcal{S}} x_i(s)\boldsymbol{\varpi}^*(s)ds + \alpha_0^* + \boldsymbol{\zeta}_i^{\intercal}\boldsymbol{\alpha} \right) \right\} \boldsymbol{\zeta}_i - S(\alpha_0^*, \boldsymbol{\alpha}, \boldsymbol{\varpi}^*(s)) \\
&= I_1 + I_2 + I_3 \qquad (23)
\end{aligned}
$$

For some $\widetilde{\eta}_{i,1}$ between $\int_{\mathcal{S}} x_i(s)\boldsymbol{\varpi}^*(s)ds + \alpha_0^* + \boldsymbol{\zeta}_i^{\intercal}\boldsymbol{\alpha}$ and $\int_{\mathcal{S}} x_i(s)\widehat{\boldsymbol{\varpi}}_p(s)ds + \widehat{\alpha}_{0,p} + \widehat{\boldsymbol{\zeta}}_i^{\intercal}\boldsymbol{\alpha}$, we get

$$
I_1 = n^{-1} \sum_{i=1}^{n} \ddot{A}(\widetilde{\eta}_{i,1}) \left\{ \int_{\mathcal{S}} x_i(s) \left(\boldsymbol{\varpi}^*(s) - \widehat{\boldsymbol{\varpi}}_p(s) \right) ds + (\alpha_0^* - \widehat{\alpha}_{0,p}) + (\boldsymbol{\zeta}_i - \widehat{\boldsymbol{\zeta}}_i)^{\intercal}\boldsymbol{\alpha} \right\} \widehat{\boldsymbol{\zeta}}_i
$$

Based on Lemma 6, we get

$$
\left\| \boldsymbol{\varpi}^*(s) - \widehat{\boldsymbol{\varpi}}_p(s) \right\| = O_p\left(\tau_n^2 \lambda_p^{-2} \right) + o_p(n^{-1/4}).
$$

Letting $\Delta_{\boldsymbol{\zeta}_i} = \widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i$, we get

$$
\|\Delta_{\boldsymbol{\zeta}_i}\| = \left\| \int_{\mathcal{S}} x_i(s) \left(\widehat{\boldsymbol{\theta}}(s) - \boldsymbol{\theta}^*(s) \right) ds \right\| = o_p\left(n^{-1/4} \right).
$$

Also, as $n, p \rightarrow \infty$, since $\alpha_0^* - \widehat{\alpha}_{0,p} = o_p(1)$, we know $\widetilde{\eta}_{i,1} - \left(\int_{\mathcal{S}} x_i(s)\boldsymbol{\varpi}^*(s)ds + \alpha_0^* + \boldsymbol{\zeta}_i^{\intercal}\boldsymbol{\alpha} \right) =$

$o_p(1)$, leading to $\ddot{A}(\widetilde{\eta}_{i,1}) = O_p(1)$. Therefore, we get

$$\|I_1\| \leq n^{-1}\sum_{i=1}^{n}\ddot{A}(\widetilde{\eta}_{i,1})\left\{\|x_i\|\|\boldsymbol{\varpi}^*(s)-\widehat{\boldsymbol{\varpi}}_p(s)\|+|\alpha_0^*-\widehat{\alpha}_{0,p}|+\|\boldsymbol{\zeta}_i-\widehat{\boldsymbol{\zeta}}_i\|\|\boldsymbol{\alpha}\|\right\}$$
$$\times\left(\|\boldsymbol{\zeta}_i\|+\|\boldsymbol{\zeta}_i-\widehat{\boldsymbol{\zeta}}_i\|\right) = o_p(1).$$

Similarly, one can show that $\|I_2\| = o_p(1)$. The law of large numbers guarantees that $\|I_3\| = o_p(1)$. Therefore, $\|S_n(\widehat{\alpha}_{0,p},\boldsymbol{\alpha},\widehat{\boldsymbol{\varpi}}_p(s)) - S(\alpha_0^*,\boldsymbol{\alpha},\boldsymbol{\varpi}^*(s))\| = o_p(1)$, which guarantees the consistency of $\widehat{\boldsymbol{\alpha}}_p^{\mathrm{cal}}$.

To establish the asymptotic normality, we use a similar idea. Note that

$$\begin{aligned}
\mathbf{0} &= S_n(\widehat{\alpha}_{0,p},\widehat{\boldsymbol{\alpha}}_p^{\mathrm{cal}},\widehat{\boldsymbol{\varpi}}_p(s)) \\
&= n^{-1}\sum_{i=1}^{n}\left\{\dot{A}\left(\int_{\mathcal{S}}x_i(s)\widehat{\boldsymbol{\varpi}}_p(s)ds+\widehat{\alpha}_{0,p}+\widehat{\boldsymbol{\zeta}}_i^{\intercal}\widehat{\boldsymbol{\alpha}}_p^{\mathrm{cal}}\right)-\dot{A}\left(\int_{\mathcal{S}}x_i(s)\widehat{\boldsymbol{\varpi}}_p(s)ds+\widehat{\alpha}_{0,p}+\widehat{\boldsymbol{\zeta}}_i^{\intercal}\boldsymbol{\alpha}^*\right)\right\}\widehat{\boldsymbol{\zeta}}_i \\
&+ n^{-1}\sum_{i=1}^{n}\left\{\dot{A}\left(\int_{\mathcal{S}}x_i(s)\widehat{\boldsymbol{\varpi}}_p(s)ds+\widehat{\alpha}_{0,p}+\widehat{\boldsymbol{\zeta}}_i^{\intercal}\boldsymbol{\alpha}^*\right)-\dot{A}\left(\int_{\mathcal{S}}x_i(s)\boldsymbol{\varpi}^*(s)ds+\alpha_0^*+\boldsymbol{\zeta}_i^{\intercal}\boldsymbol{\alpha}^*\right)\right\}\widehat{\boldsymbol{\zeta}}_i \\
&+ n^{-1}\sum_{i=1}^{n}\left\{T(y_i)-\dot{A}\left(\int_{\mathcal{S}}x_i(s)\boldsymbol{\varpi}^*(s)ds+\alpha_0^*+\boldsymbol{\zeta}_i^{\intercal}\boldsymbol{\alpha}^*\right)\right\}\left\{\widehat{\boldsymbol{\zeta}}_i-\boldsymbol{\zeta}_i\right\} \\
&+ n^{-1}\sum_{i=1}^{n}\left\{T(y_i)-\dot{A}\left(\int_{\mathcal{S}}x_i(s)\boldsymbol{\varpi}^*(s)ds+\alpha_0^*+\boldsymbol{\zeta}_i^{\intercal}\boldsymbol{\alpha}^*\right)\right\}\boldsymbol{\zeta}_i \\
&= I_4+I_5+I_6+I_7 \qquad (24)
\end{aligned}$$

For some $\widetilde{\eta}_{i,2}$ between $\int_{\mathcal{S}}x_i(s)\widehat{\boldsymbol{\varpi}}_p(s)ds+\widehat{\alpha}_{0,p}+\widehat{\boldsymbol{\zeta}}_i^{\intercal}\widehat{\boldsymbol{\alpha}}_p^{\mathrm{cal}}$ and $\int_{\mathcal{S}}x_i(s)\widehat{\boldsymbol{\varpi}}_p(s)ds+\widehat{\alpha}_{0,p}+\widehat{\boldsymbol{\zeta}}_i^{\intercal}\boldsymbol{\alpha}^*$, we get

$$I_4 = n^{-1}\sum_{i=1}^{n}\ddot{A}(\widetilde{\eta}_{i,2})\widehat{\boldsymbol{\zeta}}_i\widehat{\boldsymbol{\zeta}}_i^{\intercal}\left(\widehat{\boldsymbol{\alpha}}_p^{\mathrm{cal}}-\boldsymbol{\alpha}^*\right). \qquad (25)$$

Similar to I_1, we can show that $\widetilde{\eta}_{i,2}-\eta_i^* = o_p(1)$, where $\eta_i^* = \int_{\mathcal{S}}x_i(s)\boldsymbol{\varpi}^*(s)ds+$

$\alpha_0^* + \boldsymbol{\zeta}_i^\intercal \boldsymbol{\alpha}^*$. This leads to $\ddot{A}(\widetilde{\eta}_{i,2}) - \ddot{A}(\eta_i^*) = o_p(1)$ due to the smoothness of $A(\cdot)$. Then, we write

$$n^{-1}\sum_{i=1}^{n} \ddot{A}(\widetilde{\eta}_{i,2})\widehat{\boldsymbol{\zeta}}_i\widehat{\boldsymbol{\zeta}}_i^\intercal = n^{-1}\sum_{i=1}^{n}\Big(\ddot{A}(\eta_i^*) + o_p(1)\Big)(\boldsymbol{\zeta}_i + \Delta_{\boldsymbol{\zeta}_i})(\boldsymbol{\zeta}_i + \Delta_{\boldsymbol{\zeta}_i})^\intercal$$

this leads to

$$\left\| n^{-1}\sum_{i=1}^{n} \ddot{A}(\widetilde{\eta}_{i,2})\widehat{\boldsymbol{\zeta}}_i\widehat{\boldsymbol{\zeta}}_i^\intercal - n^{-1}\sum_{i=1}^{n} \ddot{A}(\eta_i^*)\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^\intercal \right\| = o_p(1)$$

Then, we know with probability approaching 1, $n^{-1}\sum_{i=1}^{n} \ddot{A}(\widetilde{\eta}_{i,2})\widehat{\boldsymbol{\zeta}}_i\widehat{\boldsymbol{\zeta}}_i^\intercal$ is invertible due to the low-dimensionality of $\boldsymbol{\zeta}_i$. Again, using the same matrix expansion technique, we can show that

$$\|\Delta\| = \left\| \left(n^{-1}\sum_{i=1}^{n} \ddot{A}(\widetilde{\eta}_{i,2})\widehat{\boldsymbol{\zeta}}_i\widehat{\boldsymbol{\zeta}}_i^\intercal\right)^{-1} - \left(n^{-1}\sum_{i=1}^{n} \ddot{A}(\eta_i^*)\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^\intercal\right)^{-1} \right\| = o_p(1). \tag{26}$$

Similarly, for some $\widetilde{\eta}_{i,3} = \eta_i + \Delta_{\widetilde{\eta}_{i,3}}$ and $\Delta_{A_{i,3}} = \ddot{A}(\widetilde{\eta}_{i,3}) - \ddot{A}(\eta_i^*)$, we get

$$\begin{aligned}
I_5 &= n^{-1} \sum_{i=1}^{n} \ddot{A}(\widetilde{\eta}_{i,3}) \left(\int_{\mathcal{S}} x_i(s) \left(\widehat{\boldsymbol{\varpi}}_p(s) - \boldsymbol{\varpi}^*(s) \right) ds + \widehat{\alpha}_{0,p} - \alpha_0^* \right) \left(\boldsymbol{\zeta}_i + \Delta_{\boldsymbol{\zeta}_i} \right) \\
&= n^{-1} \sum_{i=1}^{n} \left(\ddot{A}(\eta_i^*) + \Delta_{A_{i,3}} \right) \left(\int_{\mathcal{S}} x_i(s) \left(\widehat{\boldsymbol{\varpi}}_p(s) - \boldsymbol{\varpi}^*(s) \right) ds + \widehat{\alpha}_{0,p} - \alpha_0^* \right) \left(\boldsymbol{\zeta}_i + \Delta_{\boldsymbol{\zeta}_i} \right) \\
&= n^{-1} \sum_{i=1}^{n} \ddot{A}(\eta_i^*) \left(\int_{\mathcal{S}} x_i(s) \left(\widehat{\boldsymbol{\varpi}}_p(s) - \boldsymbol{\varpi}^*(s) \right) ds + \widehat{\alpha}_{0,p} - \alpha_0^* \right) \boldsymbol{\zeta}_i \\
&+ n^{-1} \sum_{i=1}^{n} \ddot{A}(\eta_i^*) \left(\int_{\mathcal{S}} x_i(s) \left(\widehat{\boldsymbol{\varpi}}_p(s) - \boldsymbol{\varpi}^*(s) \right) ds + \widehat{\alpha}_{0,p} - \alpha_0^* \right) \Delta_{\boldsymbol{\zeta}_i} \\
&+ n^{-1} \sum_{i=1}^{n} \Delta_{A_{i,3}} \left(\int_{\mathcal{S}} x_i(s) \left(\widehat{\boldsymbol{\varpi}}_p(s) - \boldsymbol{\varpi}^*(s) \right) ds + \widehat{\alpha}_{0,p} - \alpha_0^* \right) \boldsymbol{\zeta}_i \\
&+ n^{-1} \sum_{i=1}^{n} \Delta_{A_{i,3}} \left(\int_{\mathcal{S}} x_i(s) \left(\widehat{\boldsymbol{\varpi}}_p(s) - \boldsymbol{\varpi}^*(s) \right) ds + \widehat{\alpha}_{0,p} - \alpha_0^* \right) \Delta_{\boldsymbol{\zeta}_i} \\
&= I_{51} + I_{52} + I_{53} + I_{54}
\end{aligned}$$

Here, due to the smoothness of $A(\cdot)$, we get

$$\Delta_{A_{i,3}} = O_p(1) \times \left(\int_{\mathcal{S}} x_i(s) \left\{ \widehat{\boldsymbol{\varpi}}_p(s) - \boldsymbol{\varpi}^*(s) \right\} ds + (\widehat{\alpha}_{0,p} - \alpha_0^*) \right) = \quad O_p\left(\tau_n^2 \lambda_p^{-2}\right) + o_p(n^{-1/4}).$$

Here, we use the fact that $\widehat{\alpha}_{0,p} - \alpha_0^* = O_p(\tau_n^2 \lambda_p^{-2})$. Hence, we get $\|I_{53}\| = \{O_p\left(\tau_n^2 \lambda_p^{-2}\right) + o_p(n^{-1/4})\}^2$ and $\|I_{54}\| = \{O_p\left(\tau_n^2 \lambda_p^{-2}\right) + o_p(n^{-1/4})\}^2 \times o_p(n^{-1/4})$. Since $\tau_n^2 \lambda_p^{-2} = o(n^{-1/4})$, we get $\|I_{53}\| = o_p(n^{-1/2})$ and $\|I_{54}\| = o_p(n^{-1/2})$. Similarly, we show $\|I_{52}\| = o_p(n^{-1/2})$. For I_{51}, we write

$$\begin{aligned}
I_{51} = &\int_{\mathcal{S}} n^{-1} \sum_{i=1}^{n} x_i(s) \ddot{A}(\eta_i^*) \boldsymbol{\zeta}_i (\widehat{\boldsymbol{\varpi}}_p(s) - \boldsymbol{\varpi}^*(s)) ds \\
&+ \left(n^{-1} \sum_{i=1}^{n} \ddot{A}(\eta_i^*) \boldsymbol{\zeta}_i \right) (\widehat{\alpha}_{0,p} - \alpha_0) \\
= &I_{511} + I_{512}.
\end{aligned}$$

Then, the CLT yields

$$\|I_{511}\| \leq \left(\sum_{k=1}^{q} \left\| n^{-1} \sum_{i=1}^{n} x_i(s) \ddot{A}(\eta_i^*) \zeta_{ik} \right\|^2 \right)^{1/2} \|\widehat{\boldsymbol{\varpi}}_p(s) - \xi_p(s)\|$$
$$= O_p\left(n^{-1/2}\tau_n^2 \lambda_p^{-2}\right) + o_p(n^{-1/2}) = o_p(n^{-1/2});$$

here, we use the fact that $\mathbb{E}[x_i(s)\ddot{A}(\eta_i^*)\zeta_{ik}] = 0$.

Similarly,

$$\|I_{512}\| \leq \left(\sum_{k=1}^{q} \left\| n^{-1} \sum_{i=1}^{n} \ddot{A}(\eta_i^*) \zeta_{ik} \right\|^2 \right)^{1/2} \|\widehat{\alpha}_{0,p} - \alpha_0^*\| = O_p\left(n^{-1/2}\lambda_p^{-2}\tau_n^2\right) = o_p(n^{-1/2}).$$

Thus, we get $\|I_5\| = o_p(n^{-1/2})$. For I_6, note that $\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i = \int_{\mathcal{S}} x_i(s) \left(\boldsymbol{\theta}^*(s) - \widehat{\boldsymbol{\theta}}(s)\right) ds$.

Thus,

$$I_6 = \int_{\mathcal{S}} \left(n^{-1} \sum_{i=1}^{n} \left\{ T(y_i) - \dot{A}\left(\int_{\mathcal{S}} x_i(s)\boldsymbol{\varpi}^*(s)ds + \alpha_0^* + \boldsymbol{\zeta}_i^\intercal \boldsymbol{\alpha}^* \right) \right\} x_i(s) \right) \left(\boldsymbol{\theta}^*(s) - \widehat{\boldsymbol{\theta}}(s)\right) ds$$

Since $\mathbb{E}[T(y_i) \mid x_i(\cdot), \boldsymbol{\zeta}_i] = \dot{A}\left(\int_{\mathcal{S}} x_i(s)\boldsymbol{\varpi}^*(s)ds + \alpha_0^* + \boldsymbol{\zeta}_i^\intercal \boldsymbol{\alpha}^*\right)$, thus, we know

$$\mathbb{E}\left[\left\{ T(y_i) - \dot{A}\left(\int_{\mathcal{S}} x_i(s)\boldsymbol{\varpi}^*(s)ds + \alpha_0^* + \boldsymbol{\zeta}_i^\intercal \boldsymbol{\alpha}^* \right) \right\} x_i(s) \right] = 0$$

for all s. Thus, similar to I_{511} and I_{512}, by the CLT, we get

$$\|I_6\| \leq \left\| n^{-1} \sum_{i=1}^{n} \left\{ T(y_i) - \dot{A}\left(\int_{\mathcal{S}} x_i(s)\boldsymbol{\varpi}^*(s)ds + \alpha_0^* + \boldsymbol{\zeta}_i^\intercal \boldsymbol{\alpha}^* \right) \right\} x_i(s) \right\| \times \sum_{k=1}^{q} \|\theta_k^*(s) - \widehat{\theta}_k(s)\|$$
$$= O_p\left(n^{-1/2}\right) \times o_p(n^{-1/4}) = o_p(n^{-1/2}).$$

For I_7, by the CLT, we get

$$\sqrt{n} I_7 \xrightarrow{d} N\left(\mathbf{0}, \mathbb{E}\left[\left\{ T(y_i) - \dot{A}\left(\eta_i^*\right) \right\}^2 \boldsymbol{\zeta}_i \boldsymbol{\zeta}_i^\intercal \right] \right)$$

Some algebra yields that

$$\begin{aligned}
&\mathbb{E}\left[\left\{T(y_i) - \dot{A}\left(\eta_i^*\right)\right\}^2 \boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^\intercal\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left\{T(y_i) - \dot{A}\left(\eta_i^*\right)\right\}^2 \mid x_i(\cdot), \boldsymbol{\zeta}_i\right] \boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^\intercal\right] \\
&= \mathbb{E}\left[\ddot{A}(\eta_i^*)\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^\intercal\right] = \widetilde{\Sigma}_\zeta.
\end{aligned}$$

Combining all the assertions above, we get

$$\sqrt{n}\left(\widehat{\boldsymbol{\alpha}}_p^{\mathrm{cal}} - \boldsymbol{\alpha}^*\right) = -\left\{\left(n^{-1}\sum_{i=1}^n \ddot{A}(\eta_i^*)\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^\intercal\right\}^{-1} + \Delta\right) I_7 + o_p(1),$$

where Δ is defined in (26) with $\|\Delta\| = o_p(1)$. Finally, since $\left(n^{-1}\sum_{i=1}^n \ddot{A}(\eta_i^*)\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^\intercal\right)^{-1} \to \widetilde{\Sigma}_\zeta^{-1}$ in probability, we know

$$\sqrt{n}\left(\widehat{\boldsymbol{\alpha}}_p^{\mathrm{cal}} - \boldsymbol{\alpha}\right) \xrightarrow{d} N\left(\mathbf{0}, \widetilde{\Sigma}_\zeta^{-1}\right);$$

this completes the proof.

□

Supporting Information of the Real Data Analysis

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year publicprivate partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org."

PET image preprocessing. The PET images used in the analysis underwent four main preprocessing steps, which made more uniform PET data available and provide consistent starting points and simplify sequence ADNI analyses. In the first step, separate frames were extracted from the original raw image file for registration purposes. Six five-minute frames (ADNI1) were acquired 30 to 60 minutes post-injection. Each extracted frame was co-registered to the first extracted frame of the raw image file. All co-registered frames were recombined into a co-registered dynamic image set. These image sets have the same image size (for example, $128 \times 128 \times 63$) and voxel dimensions (for example, $2.0 \times 2.0 \times 2.0$ mm) and remain in the same spatial orientation as the original PET image data. In the second step, a single 30 min PET image was created by averaging the 6 five-minute frames of the co-registered dynamic image set from step 1. In the third step, each subject's co-registered averaged image from their baseline PET scan was then reoriented into a standard $160 \times 160 \times 96$ voxel image grid, having 1.5 mm cubic voxels. The individual frames from each following PET scan (6-month scan, 12-month scan, etc.) were co-registered to this baseline reference image. In the fourth step, the final PET image was the result of smoothing of the image from step 3. Each image set was filtered with a scanner-specific filter function (can be a non-isotropic filter) to produce images of a uniform isotropic resolution of 8 mm FWHM, the approximate resolution of the lowest resolution scanners used in ADNI. Image sets from higher resolution scanners have been smoothed more than image sets from lower resolution scanners. The specific filter functions were determined from the Hoffman phantom PET scans that were acquired during the certification process.

Demographic information. Among the 302 individuals, 195 participants were male, and 107 were female; 283 were right-handed, and 19 were left-handed; For the marital status, 239 were married, 35 were windowed, 22 were divorced, and 6 were never married. The individuals had an average of 15.41 years of education with a standard deviation 3.01 years. The minimum education length was 4 years and the maximum education length was 20 years. The average age was 75.2 years with a standard deviation of 7.3 years. The youngest person was 55 years old, and the oldest person was 89 years old.

Bibliography

Rudin, W. et al. (1976). *Principles of mathematical analysis*, Volume 3. McGraw-hill New York.