

## **Appendix for Distributed Learning of Single Index Model with Differential Privacy and Massive Data**

Lingrui Wang<sup>1</sup>, Fode Zhang<sup>1</sup>, Hua Liang<sup>2</sup>

<sup>1</sup>*Southwestern University of Finance and Economics*

<sup>2</sup>*The George Washington University*

### **Supplementary Material**

This document succinctly compiles the supplementary materials accompanying the main text, including the complete proofs of all theorems and corollaries, the solving algorithm presented in Section S1, the differential privacy lemmas and guarantees in Section S2, and the full experimental results in Section S3, along with additional supplementary experiments.

Section S1 provides the detailed proofs for Section 2, where we derive convergence rates and estimation error bounds for the nonparametric components, establish asymptotic normality for the parametric component, and present the corresponding solving algorithm. Section S2 contains the proofs for Section 3, establishing convergence rates and error bounds for the differentially private procedures, together with the supporting lemmas required for the analysis of the DP algorithms. Finally, Section S3 presents additional experiments that further support the proposed methods

## S1 Proof of Section 2

### S1.1 Solving Algorithm for the Spline-Based Estimator

---

**Algorithm 1** Alternating Optimization for SIM

---

**Input:** Data  $\{(Y_i, \mathbf{X}_i)\}_{i \in G_j}$ , initial values  $(\boldsymbol{\tau}^{(0)}, \boldsymbol{\gamma}^{(0)})$  with  $\|\boldsymbol{\tau}^{(0)}\| < 1$ , tolerance  $\epsilon > 0$

**Output:**  $(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\gamma}})$

- 1: Set  $t = 0$
  - 2: **repeat**
  - 3:   Compute  $\mu_i^{(t)} = \boldsymbol{\alpha}(\boldsymbol{\tau}^{(t)})^\top \mathbf{X}_i$ , where  $\boldsymbol{\alpha}(\boldsymbol{\tau}) = (\sqrt{1 - \|\boldsymbol{\tau}\|^2}, \boldsymbol{\tau}^\top)^\top$
  - 4:   Update  $\boldsymbol{\gamma}^{(t+1)} = \left[ \frac{1}{n} \sum_{i \in G_j} \mathbf{B}(\mu_i^{(t)}) \mathbf{B}^\top(\mu_i^{(t)}) \right]^{-1} \left[ \frac{1}{n} \sum_{i \in G_j} Y_i \mathbf{B}(\mu_i^{(t)}) \right]$
  - 5:   Compute  $\nabla_{\boldsymbol{\tau}} l_n = \frac{1}{n} \sum_{i \in G_j} \left( Y_i - \boldsymbol{\gamma}^{(t+1)\top} \mathbf{B}(\mu_i^{(t)}) \right) \left( \boldsymbol{\gamma}^{(t+1)\top} \mathbf{B}'(\mu_i^{(t)}) \right) \mathbf{J}^\top(\boldsymbol{\tau}^{(t)}) \mathbf{X}_i$
  - 6:   Compute  $\nabla_{\boldsymbol{\tau}}^2 l_n = \mathbf{J}^\top(\boldsymbol{\tau}^{(t)}) \left[ \frac{1}{n} \sum_{i \in G_j} \left( \boldsymbol{\gamma}^{(t+1)\top} \mathbf{B}'(\mu_i^{(t)}) \right)^2 \mathbf{X}_i \mathbf{X}_i^\top \right] \mathbf{J}(\boldsymbol{\tau}^{(t)})$
  - 7:   Update  $\boldsymbol{\tau}^{(t+1)} = \boldsymbol{\tau}^{(t)} - (\nabla_{\boldsymbol{\tau}}^2 l_n)^{-1} \nabla_{\boldsymbol{\tau}} l_n$
  - 8:   If  $\|\boldsymbol{\tau}^{(t+1)}\| \geq 1$ , set  $\boldsymbol{\tau}^{(t+1)} = (1 - \eta) \boldsymbol{\tau}^{(t+1)} / \|\boldsymbol{\tau}^{(t+1)}\|$  for a small  $\eta > 0$
  - 9:   Set  $t = t + 1$
  - 10: **until**  $\|\boldsymbol{\tau}^{(t)} - \boldsymbol{\tau}^{(t-1)}\| < \epsilon$  and  $\|\boldsymbol{\gamma}^{(t)} - \boldsymbol{\gamma}^{(t-1)}\| < \epsilon$
  - 11: **Return**  $\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^{(t)}$  and  $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^{(t)}$
- 

Building on the algorithmic paradigms proposed by Zhang and Lian (2018); Wang and Cao (2018), we use an alternating optimization strategy to estimate the nonparametric and parametric components of the single-index model.

At iteration  $t$ , we first fix the current index parameter  $\boldsymbol{\tau}^{(t)}$  and compute

the projected covariates

$$\mu_i^{(t)} = \boldsymbol{\alpha}(\boldsymbol{\tau}^{(t)})^\top \mathbf{X}_i.$$

Given  $\boldsymbol{\tau}^{(t)}$ , the spline coefficients  $\boldsymbol{\gamma}$  are then updated by ordinary least squares. Since the loss is quadratic in  $\boldsymbol{\gamma}$  when  $\boldsymbol{\tau}$  is fixed, this step admits a closed-form solution.

Next, with  $\boldsymbol{\gamma}^{(t+1)}$  fixed, we update the index parameter  $\boldsymbol{\tau}$  using a quasi-Newton-type step based on the gradient and an approximate Hessian of  $l_n(\boldsymbol{\tau}, \boldsymbol{\gamma})$ . After each update, we apply a step to ensure that the constraint  $\|\boldsymbol{\tau}\| < 1$  is satisfied.

The algorithm iterates between these two steps until the successive changes in both  $\boldsymbol{\tau}$  and  $\boldsymbol{\gamma}$  are below a prescribed tolerance. This alternating scheme combines the closed-form update for the spline coefficients with an efficient low-dimensional update for the index parameter, leading to a practical estimation procedure for the SIM.

## S1.2 Technical lemmas for Section 2

We present several technical lemmas that will be employed to establish the principal results discussed in Section 2. Based on the result of (De Boor, 1978), for any function  $g \in \mathcal{H}, n \geq 1$ , there exists a function  $\tilde{g} \in \mathcal{S}_n$ , such that  $\|\tilde{g} - g\|_\infty \leq Ch^p$ , where  $C$  is some fixed positive constant. Thus, we

have  $\tilde{f}^{(j)} = \tilde{\gamma}^{(j)\top} \mathbf{B}(u)$ , such that  $\|\tilde{f}^{(j)} - f_0\|_\infty \leq Ch^p$ , and  $\tilde{f}^{(j)}(u)$  is the best least-squares projection of  $f_0(u)$  in the space  $S_n$ . This leads to

$$\langle \tilde{f}^{(j)}(u) - f_0(u), \mathbf{B}(u) \rangle = 0, j = 1, 2, \dots, s. \quad (\text{S1.1})$$

For convenience, we define

$$\tilde{\tau}^{(j)} = \arg \min_{\tau} \frac{1}{2n} \sum_{i \in G_j} (Y_i - \tilde{f}(\boldsymbol{\alpha}^\top(\tau) \mathbf{X}_i))^2,$$

$$\text{and let } \boldsymbol{\theta} = \begin{pmatrix} \tau \\ \gamma \end{pmatrix}, \tilde{\boldsymbol{\theta}}^{(j)} = \begin{pmatrix} \tilde{\tau}^{(j)} \\ \tilde{\gamma}^{(j)} \end{pmatrix}, \hat{\boldsymbol{\theta}}^{(j)} = \begin{pmatrix} \hat{\tau}^{(j)} \\ \hat{\gamma}^{(j)} \end{pmatrix}, \text{ let } U_{\tau,0i} = \boldsymbol{\alpha}^\top(\tau_0) \mathbf{X}_i.$$

**Lemma 1.** *Under assumption (A1)-(A5), we have*

$$\sqrt{n}(\tilde{\tau}^{(j)} - \tau_0) \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1} \boldsymbol{\Sigma} \mathbf{A}^{-1}),$$

and

$$\sqrt{n}(\tilde{\boldsymbol{\alpha}}^{(j)} - \boldsymbol{\alpha}_0) \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{J}(\tau_0) \mathbf{A}^{-1} \boldsymbol{\Sigma} \mathbf{A}^{-1} \mathbf{J}^\top(\tau_0)),$$

where

$$\begin{aligned} \boldsymbol{\Sigma} &= E \left[ \varepsilon^2 f'(U_{\tau,0i}) \mathbf{J}^\top(\tau_0) \mathbf{X} \mathbf{X}^\top \mathbf{J}(\tau_0) f'(U_{\tau,0i}) \right], \\ \mathbf{A} &= E \left[ \{f'_0(U_{\tau,0i})\}^2 \mathbf{J}^\top(\tau_0) \mathbf{X} \mathbf{X}^\top \mathbf{J}(\tau_0) \right]. \end{aligned}$$

*Proof.* Let  $\tilde{\mathbf{a}} = \sqrt{n}(\tilde{\tau}^{(j)} - \tau_0)$ , note that  $\tilde{\mathbf{a}}$  minimizes the function

$$\tilde{l}(\mathbf{a}) = \frac{1}{2} \left\{ \sum_{i \in G_j} \left( Y_i - \tilde{f}(U_{\tau,0i} + \frac{1}{\sqrt{n}} \mathbf{a}^\top \mathbf{J}^\top(\tilde{\tau}) \mathbf{X}_i) \right)^2 - \sum_i \left( Y_i - \tilde{f}(U_{\tau,0i}) \right)^2 \right\}.$$

Then we have

$$\begin{aligned}\tilde{l}(\mathbf{a}) &= - \sum_{i \in G_j} (Y_i - \tilde{f}(U_{\tau,0i})) (\tilde{f}(U_{\tau,0i} + \frac{1}{\sqrt{n}} \mathbf{a}^\top \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i) - \tilde{f}(U_{\tau,0i})) \\ &\quad + \frac{1}{2} \sum_i (\tilde{f}(U_{\tau,0i} + \frac{1}{\sqrt{n}} \mathbf{a}^\top \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i) - \tilde{f}(U_{\tau,0i}))^2.\end{aligned}$$

Using Taylor expansion for  $\tilde{f}$ , we have

$$\begin{aligned}\tilde{l}(\mathbf{a}) &= - \sum_{i \in G_j} \left( Y_i - \tilde{f}(U_{\tau,0i}) \right) \left[ \frac{1}{\sqrt{n}} \tilde{f}'(U_{\tau,0i}) \mathbf{a}^\top \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i + \frac{1}{2n} \tilde{f}''(\bar{U}_{\tau,0i}) \mathbf{a}^\top \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i \mathbf{X}_i^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{a} \right] \\ &\quad + \frac{1}{2} \sum_{i \in G_j} \left[ \frac{1}{\sqrt{n}} \tilde{f}'(U_{\tau,0i}) \mathbf{a}^\top \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i + \frac{1}{2n} \tilde{f}''(\bar{U}_{\tau,0i}) \mathbf{a}^\top \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i \mathbf{X}_i^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{a} \right]^2,\end{aligned}$$

where  $\bar{U}_{\tau,0i}$  lies between  $U_{\tau,0i}$  and  $U_{\tau,0i} + \frac{1}{\sqrt{n}} \mathbf{a}^\top \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i$ , which is independent of  $Y_i$ . Thus, we get

$$\tilde{l}(\mathbf{a}) = - \sum_{i \in G_j} (Y_i - \tilde{f}(U_{\tau,0i})) \frac{1}{\sqrt{n}} \tilde{f}'(U_{\tau,0i}) \mathbf{a}^\top \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i + \frac{1}{2} \mathbf{a}^\top \mathbf{A}_n \mathbf{a} (1 + o_{\mathbb{P}}(1)),$$

where

$$\mathbf{A}_n = \frac{1}{n} \sum_{i \in G_j} \left[ \left\{ \tilde{f}'(U_{\tau,0i}) \right\}^2 + \frac{1}{2} (Y_i - \tilde{f}(U_{\tau,0i})) \tilde{f}''(\bar{U}_{\tau,0i}) \right] \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i \mathbf{X}_i^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}).$$

From the structure of  $\mathbf{A}_n$  and Carroll et al. (1997), there exist large constants  $c_1$  and  $c_2$  such that

$$\begin{aligned}\left\| \mathbf{A}_n - \frac{1}{n} \sum_{i \in G_j} \left[ \{f'_0(U_{\tau,0i})\}^2 + \frac{1}{2} (Y_i - f_0(U_{\tau,0i})) \tilde{f}''(U_{\tau,0i}) \right] \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) \right\| \\ = o_{\mathbb{P}} \left\{ \frac{1}{n} \sum_{i \in G_j} (c_1 |Y_i| + c_2) \right\} = o_{\mathbb{P}}(1).\end{aligned}$$

The law of large numbers implies that

$$\begin{aligned} & \frac{1}{n} \sum_{i \in G_j} \left[ \{f_0'(U_{\tau,0i})\}^2 + \frac{1}{2}(Y_i - f_0(U_{\tau,0i}))\tilde{f}''(U_{\tau,0i}) \right] \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) \\ &= E \left[ \{f_0'(U_{\tau,0i})\}^2 \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X} \mathbf{X}^\top \mathbf{J}(\boldsymbol{\tau}_0) \right] + o_{\mathbb{P}}(1). \end{aligned}$$

Thus, we have

$$\mathbf{A}_n = E \left[ \{f_0'(U_{\tau,0i})\}^2 \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X} \mathbf{X}^\top \mathbf{J}(\boldsymbol{\tau}_0) \right] + o_{\mathbb{P}}(1) = \mathbf{A} + o_{\mathbb{P}}(1).$$

Note that

$$\begin{aligned} & \frac{1}{\sqrt{n}}(Y_i - \tilde{f}(U_{\tau,0i}))\tilde{f}'(U_{\tau,0i}) \\ &= \frac{1}{\sqrt{n}}(Y_i - f(U_{\tau,0i}))\tilde{f}'(U_{\tau,0i}) + \frac{1}{\sqrt{n}}(f(U_{\tau,0i}) - \tilde{f}(U_{\tau,0i}))\tilde{f}'(U_{\tau,0i}) \\ &= \frac{1}{\sqrt{n}}(Y_i - f(U_{\tau,0i}))f'(U_{\tau,0i}) + \frac{1}{\sqrt{n}}(Y_i - f(U_{\tau,0i}))(\tilde{f}'(U_{\tau,0i}) - f'(U_{\tau,0i})) \\ & \quad + \frac{1}{\sqrt{n}}(f(U_{\tau,0i}) - \tilde{f}(U_{\tau,0i}))(\tilde{f}'(U_{\tau,0i}) - f'(U_{\tau,0i})) \\ & \quad + \frac{1}{\sqrt{n}}(f(U_{\tau,0i}) - \tilde{f}(U_{\tau,0i}))f'(U_{\tau,0i}). \end{aligned}$$

Based on De Boor (1978),  $\|\tilde{f}' - f_0'\|_\infty = O(h^{p-1})$ , we have

$$\begin{aligned} & \sum_{i \in G_j} \frac{1}{\sqrt{n}}(Y_i - \tilde{f}(U_{\tau,0i}))\tilde{f}'(U_{\tau,0i})\mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i \\ &= \frac{1}{\sqrt{n}} \sum_{i \in G_j} \varepsilon_i f'(U_{\tau,0i})\mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i + o_{\mathbb{P}}(1). \end{aligned}$$

By the convexity lemma of Pollard (1991), we get

$$\tilde{\mathbf{a}} = \mathbf{A}^{-1} \frac{1}{\sqrt{n}} \sum_{i \in G_j} \varepsilon_i f'(U_{\tau,0i})\mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i + o_{\mathbb{P}}(1).$$

Let

$$\begin{aligned} & \text{Var} \{ \varepsilon f'(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X} \} \\ &= E \left[ \varepsilon^2 f'(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X} \mathbf{X}^\top \mathbf{J}(\boldsymbol{\tau}_0) f'(U_{\tau,0i}) \right] = \boldsymbol{\Sigma}. \end{aligned}$$

We get that

$$\sqrt{n}(\tilde{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1} \boldsymbol{\Sigma} \mathbf{A}^{-1}).$$

Applying the delta method, we can obtain the distribution of  $\sqrt{n}(\tilde{\boldsymbol{\alpha}}^{(j)} - \boldsymbol{\alpha}_0)$ ,

that is

$$\sqrt{n}(\tilde{\boldsymbol{\alpha}}^{(j)} - \boldsymbol{\alpha}_0) \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{J}(\boldsymbol{\tau}_0) \mathbf{A}^{-1} \boldsymbol{\Sigma} \mathbf{A}^{-1} \mathbf{J}^\top(\boldsymbol{\tau}_0)).$$

□

Let

$$\mathbf{D}_{\boldsymbol{\theta},i} = \begin{bmatrix} \boldsymbol{\gamma}^\top \mathbf{B}'(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}) \mathbf{X}_i) \mathbf{J}^\top(\boldsymbol{\tau}) & \mathbf{0} \\ \mathbf{0} & \mathbf{B}(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}) \mathbf{X}_i) \end{bmatrix},$$

and

$$\mathbf{W}_{n,\boldsymbol{\theta}} = \frac{1}{n} \sum_i \mathbf{D}_{\boldsymbol{\theta},i} \begin{pmatrix} \mathbf{X}_i \\ 1 \end{pmatrix} \begin{pmatrix} \mathbf{X}_i \\ 1 \end{pmatrix}^\top \mathbf{D}_{\boldsymbol{\theta},i}^\top, \quad \mathbf{W}_{\boldsymbol{\theta}} = E \left[ \mathbf{D}_{\boldsymbol{\theta},i} \begin{pmatrix} \mathbf{X}_i \\ 1 \end{pmatrix} \begin{pmatrix} \mathbf{X}_i \\ 1 \end{pmatrix}^\top \mathbf{D}_{\boldsymbol{\theta},i}^\top \right].$$

We have the following lemma.

**Lemma 2** (Lemma S.4 in the supplement of (Wang and Cao, 2018)). *There*

*exists a positive constant  $C$  such that*

$$\sup_{\boldsymbol{\theta}} \|\mathbf{W}_{\boldsymbol{\theta}}^{-1}\|_2 \leq C, \quad \text{and} \quad \sup_{\boldsymbol{\theta}} \|\mathbf{W}_{n,\boldsymbol{\theta}}^{-1}\|_2 \leq C, \text{ a.s.} \dots$$

**Lemma 3.** *Under assumptions (A1)-(A5), we have*

$$\left\| \hat{\boldsymbol{\theta}}^{(j)} - \tilde{\boldsymbol{\theta}}^{(j)} \right\| = O_{\mathbb{P}}(J_n^{1/2} n^{-1/2} + h^p).$$

*Proof.* Note that

$$\left. \frac{\partial l_n^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(j)}} - \left. \frac{\partial l_n^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}^{(j)}} = \left( \left. \frac{\partial^2 l_n^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right) (\hat{\boldsymbol{\theta}}^{(j)} - \tilde{\boldsymbol{\theta}}^{(j)}),$$

where  $\bar{\boldsymbol{\theta}}$  lies between  $\hat{\boldsymbol{\theta}}^{(j)}$  and  $\tilde{\boldsymbol{\theta}}^{(j)}$ . Since  $\hat{\boldsymbol{\theta}}^{(j)}$  minimizes the function  $l_n^{(j)}(\boldsymbol{\theta})$ ,

we have

$$\hat{\boldsymbol{\theta}}^{(j)} - \tilde{\boldsymbol{\theta}}^{(j)} = - \left( \left. \frac{\partial^2 l_n^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right)^{-1} \left. \frac{\partial l_n^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}^{(j)}}.$$

We have

$$\left. \frac{\partial l_n^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}^{(j)}} = -\frac{1}{n} \sum_{i \in G_j} \left( Y_i - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}) \mathbf{X}_i) \right) \begin{pmatrix} \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)} \mathbf{X}_i) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i \\ \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)} \mathbf{X}_i)) \end{pmatrix}. \quad (\text{S1.2})$$

Thus, the first term in (S1.2) can be expressed as follows

$$\begin{aligned} & \frac{1}{n} \sum_{i \in G_j} \left( Y_i - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i) \right) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)} \mathbf{X}_i) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i \\ &= \frac{1}{n} \sum_{i \in G_j} \left( Y_i - f_0(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i) \right) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)} \mathbf{X}_i) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i \\ & \quad + \frac{1}{n} \sum_{i \in G_j} \left( f_0(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i) - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i) \right) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)} \mathbf{X}_i) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i \\ & \quad + \frac{1}{n} \sum_{i \in G_j} \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)} \mathbf{X}_i) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) (\boldsymbol{\tau}_0 - \tilde{\boldsymbol{\tau}}^{(j)}) + O_{\mathbb{P}}(n^{-1/2}). \end{aligned} \quad (\text{S1.3})$$

The first term in (S1.3) can be expressed

$$\begin{aligned}
 & \frac{1}{n} \sum_{i \in G_j} \varepsilon_i \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)} \mathbf{X}_i) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i \\
 &= \frac{1}{n} \sum_{i \in G_j} \varepsilon_i f_0'(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0 \mathbf{X}_i) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i \\
 & \quad + \frac{1}{n} \sum_{i \in G_j} \varepsilon_i \left( \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)} \mathbf{X}_i) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) - f_0'(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i) \mathbf{J}^\top(\boldsymbol{\tau}_0) \right) \mathbf{X}_i \\
 &= \left\{ \frac{1}{n} \sum_{i \in G_j} \varepsilon_i f_0'(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i \right\} (1 + O_{\mathbb{P}}(1)).
 \end{aligned}$$

The Bersten's inequity leads to

$$E \left\| \frac{1}{n} \sum_{i \in G_j} \varepsilon_i f_0'(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i \right\|^2 \leq C n^{-1}.$$

Thus

$$\left\| \frac{1}{n} \sum_{i \in G_j} (Y_i - f_0(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i)) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)} \mathbf{X}_i) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i \right\| = O_{\mathbb{P}}(n^{-1/2}).$$

Note that  $\|\tilde{f}^{(j)} - f_0\|_\infty \leq C h^p$ , we have

$$\left\| \frac{1}{n} \sum_{i \in G_j} \left( f_0(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i) - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i) \right) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)} \mathbf{X}_i) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i \right\| = O_{\mathbb{P}}(h^p),$$

and by using Lemma 1, we get

$$\left\| \frac{1}{n} \sum_{i \in G_j} \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)} \mathbf{X}_i)) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) (\boldsymbol{\tau}_0 - \tilde{\boldsymbol{\tau}}^{(j)}) \right\| = O_{\mathbb{P}}(n^{-1/2}).$$

The second term in Eq. (S1.2) can be formulated as follows

$$\begin{aligned}
 & \frac{1}{n} \sum_{i \in G_j} \left( Y_i - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_i) \right) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_i) \\
 = & \frac{1}{n} \sum_{i \in G_j} \left( Y_i - f_0(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_i) \right) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_i) \\
 & + \frac{1}{n} \sum_{i \in G_j} \left( f_0(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_i) - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_i) \right) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_i) \\
 & + \frac{1}{n} \sum_{i \in G_j} f'_0(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_i) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}})\mathbf{X}_i) \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0)(\tilde{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) + O_{\mathbb{P}}(n^{-1/2}).
 \end{aligned}$$

Note that

$$E \left\| \frac{1}{n} \sum_{i \in G_j} \varepsilon_i \mathbf{B}(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_i) \right\|^2 \leq C J_n / n.$$

Thus we have

$$\begin{aligned}
 & \left\| \frac{1}{n} \sum_{i \in G_j} \varepsilon_i \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_i) \right\| \\
 = & \left\| \frac{1}{n} \sum_{i \in G_j} \varepsilon_i \mathbf{B}(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_i) \right\| + \left\| \frac{1}{n} \sum_{i \in G_j} \varepsilon_i \left( \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_i) - \mathbf{B}(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_i) \right) \right\| \\
 = & \left\| \frac{1}{n} \sum_{i \in G_j} \varepsilon_i \mathbf{B}(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_i) \right\| + \left\| \frac{1}{n} \sum_{i \in G_j} \varepsilon_i \mathbf{B}'(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_i) \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0)(\tilde{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \right\| + o(n^{-1/2}) \\
 = & O_{\mathbb{P}}(J_n^{1/2} n^{-1/2}).
 \end{aligned}$$

Since  $\left\| \tilde{f}^{(j)} - f_0 \right\|_{\infty} \leq C h^p$ , Lemma 1 and Eq. S1.1 lead to

$$\begin{aligned}
 & \left\| \frac{1}{n} \sum_{i \in G_j} \left( f_0(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_i) - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_i) \right) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_i) \right\| = 0, \\
 & \left\| \frac{1}{n} \sum_{i \in G_j} f'_0(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_i) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_i) \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0)(\tilde{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \right\| = O_{\mathbb{P}}(J_n^{1/2} n^{-1/2}).
 \end{aligned}$$

Thus, we have

$$\left\| \frac{\partial l_n^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}^{(j)}} \right\| = O_{\mathbb{P}}(J_n^{1/2} n^{-1/2} + h^p).$$

For the second order derivative, we have

$$\begin{aligned} V_{n,\boldsymbol{\theta}}^{(j)} &= \begin{pmatrix} \frac{\partial^2 l^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^\top} & \frac{\partial^2 l^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\gamma}^\top} \\ \frac{\partial^2 l^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} & \frac{\partial^2 l^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} \end{pmatrix} \\ &= \frac{1}{n} \sum_{i \in G_j} \mathbf{D}_{\boldsymbol{\theta},i} \begin{pmatrix} \mathbf{X}_i \\ 1 \end{pmatrix} \begin{pmatrix} \mathbf{X}_i \\ 1 \end{pmatrix}^\top \mathbf{D}_{\boldsymbol{\theta},i}^\top \\ &\quad - \frac{1}{n} \sum_{i \in G_j} (Y_i - f(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}) \mathbf{X}_i)) \\ &\quad \times \begin{pmatrix} f''(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}) \mathbf{X}_i) \mathbf{J}^\top(\boldsymbol{\tau}) \mathbf{X}_i \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}) & \mathbf{B}'(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}) \mathbf{X}_i) \mathbf{J}^\top(\boldsymbol{\tau}) \mathbf{X}_i \\ \mathbf{B}'(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}) \mathbf{X}_i) \mathbf{J}^\top(\boldsymbol{\tau}) \mathbf{X}_i & 0 \end{pmatrix} \\ &\quad - \frac{1}{n} \sum_{i \in G_j} (Y_i - f(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}) \mathbf{X}_i)) \begin{pmatrix} \mathbf{B}'(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}) \mathbf{X}_i) \left\{ \frac{\partial}{\partial \boldsymbol{\tau}} \mathbf{J} \right\}^\top \mathbf{X}_i & 0 \\ \mathbf{0} & 0 \end{pmatrix}. \end{aligned}$$

From Lemma 2, we have  $\sup_{\boldsymbol{\theta}} \left\| (V_{n,\boldsymbol{\theta}}^{(j)})^{-1} \right\|_2 = O(1)$ . This leads to

$$\left\| \hat{\boldsymbol{\theta}}^{(j)} - \tilde{\boldsymbol{\theta}}^{(j)} \right\| \leq \sup_{\boldsymbol{\theta}} \left\| (V_{n,\boldsymbol{\theta}}^{(j)})^{-1} \right\|_2 \left\| \frac{\partial l_n^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}^{(j)}} \right\| = O_{\mathbb{P}}(J_n^{1/2} n^{-1/2} + h^p).$$

□

**Lemma 4.** *Under assumptions (A1)-(A5), we have*

$$\left\| \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( Y_i - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij} \right\| = O_{\mathbb{P}}(N^{-1/2} + h^p).$$

*Proof.* It follows that

$$\begin{aligned}
 & \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( Y_i - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \right) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij} \\
 &= \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \varepsilon_{ij} \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij} \\
 & \quad + \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( f_0(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_{ij}) - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_{ij}) \right) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij} \\
 & \quad + \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left\{ \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \right\}^2 \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)})^\top \mathbf{X}_{ij} \mathbf{X}_{ij}^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}) (\tilde{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0).
 \end{aligned}$$

Utilizing an analogous procedure as outlined in Lemma 3, we have

$$\begin{aligned}
 & \left\| \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \varepsilon_{ij} \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij} \right\| = O_{\mathbb{P}}(N^{-1/2}), \\
 & \left\| \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( f_0(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_{ij}) - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_{ij}) \right) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij} \right\| = O_{\mathbb{P}}(h^p).
 \end{aligned}$$

Subsequently, we shall establish the bound for the third term in S1.4. By

Lemma 1, we have

$$\tilde{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0 = \mathbf{A}^{-1} \frac{1}{n} \sum_{i \in G_j} \varepsilon_i f'_0(\boldsymbol{\alpha}_0^\top \mathbf{X}_{ij}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_{ij} + o_{\mathbb{P}}(1),$$

where  $\mathbf{A} = E \left[ \{f'_0(U_{\tau,0i})\}^2 \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X} \mathbf{X}^\top \mathbf{J}(\boldsymbol{\tau}_0) \right]$ . Then, we get

$$\begin{aligned}
 & \sum_{i \in G_j} \left\{ \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \right\}^2 \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)})^\top \mathbf{X}_{ij} \mathbf{X}_{ij}^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}) (\tilde{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \\
 &= \sum_{i \in G_j} \varepsilon_i \left\{ \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \right\}^2 \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)})^\top \mathbf{X}_{ij} \mathbf{v}^{(j)},
 \end{aligned}$$

where  $\mathbf{v}^{(j)} = \sum_{i \in G_j} f'_0(\boldsymbol{\alpha}_0^\top \mathbf{X}_{ij}) \mathbf{X}_{ij}^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{A}^{-1} \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_{ij}$ . Thus, we have

$$\begin{aligned} & \frac{1}{N^2} E \left\{ \sum_{j=1}^s \sum_{i \in G_j} \left\{ \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\}^2 \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)})^\top \mathbf{X}_{ij} \mathbf{X}_{ij}^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}) (\tilde{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \right\}^2 \\ &= \frac{1}{N^2} E \left\{ \sum_{j=1}^s \sum_{i \in G_j} \varepsilon_i \left\{ \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\}^2 \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)})^\top \mathbf{X}_{ij} \mathbf{v}^{(j)} \right\}^2 \\ &\leq \frac{\sigma^2}{N^2} E \sum_{j=1}^s \sum_{i \in G_j} \left\{ \left\{ \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\}^2 \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)})^\top \mathbf{X}_{ij} \mathbf{v}^{(j)} \right\}^2. \end{aligned}$$

By Lemma 1, we know that

$$\frac{1}{n} \sum_{i \in G_j} \left\{ \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\}^2 \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)})^\top \mathbf{X}_{ij} \mathbf{X}_{ij}^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}) = A + o_{\mathbb{P}}(1).$$

Observing that

$$\begin{aligned} & \sum_{i \in G_j} \left\{ \left\{ \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\}^2 \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)})^\top \mathbf{X}_{ij} \mathbf{v}^{(j)} \right\}^2 \\ &= \sum_{i \in G_j} \left\{ \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\}^4 (\mathbf{v}^{(j)})^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)})^\top \mathbf{X}_{ij} \mathbf{X}_{ij}^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}) (\mathbf{v}^{(j)}) \\ &= n \left( \mathbf{v}^{(j)} \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right)^\top \left( \frac{1}{n} \sum_{i \in G_j} \left\{ \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\}^2 \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)})^\top \mathbf{X}_{ij} \mathbf{X}_{ij}^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}) \right) \\ &\quad \times \left( \mathbf{v}^{(j)} \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right) \\ &= \left( \frac{1}{\sqrt{n}} \sum_{i \in G_j} f_{ij} \mathbf{X}_{ij}^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_{ij} \right)^\top \mathbf{A}^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i \in G_j} f_{ij} \mathbf{X}_{ij}^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_{ij} \right), \end{aligned}$$

where  $f_{ij} = f'_0(\boldsymbol{\alpha}_0^\top \mathbf{X}_{ij}) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij})$ . The central limit theorem and

Slutsky's theorem give that

$$E \sum_{i \in G_j} \left\{ \left\{ \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\}^2 \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)})^\top \mathbf{X}_{ij} \mathbf{v}^{(j)} \right\}^2 = O_{\mathbb{P}}(1).$$

Therefore we have

$$\left\| \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left\{ \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \right\}^2 \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)})^\top \mathbf{X}_{ij} \mathbf{X}_{ij}^\top \mathbf{J}(\tilde{\boldsymbol{\tau}}^{(j)}) (\tilde{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \right\| = O_{\mathbb{P}}(N^{-1/2}).$$

This leads to

$$\left\| \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( Y_i - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \right) \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij} \right\| = O_{\mathbb{P}}(N^{-1/2} + h^p).$$

□

**Lemma 5.** *Based on the assumptions (A1)-(A5), we have*

$$\left\| \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( Y_i - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \right) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \right\| = O_{\mathbb{P}}(J_n^{1/2} N^{-1/2}).$$

*Proof.* It follows that

$$\begin{aligned} & \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( Y_i - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \right) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \\ &= \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \varepsilon_{ij} \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \\ & \quad + \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) - f_0(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \right) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \\ & \quad + \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( f_0(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}_{ij}) - f_0(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}) \right) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)})\mathbf{X}_{ij}). \end{aligned} \tag{S1.4}$$

Employing analogous procedure as applied in Lemma 3 and Lemma 4, we

have

$$\begin{aligned}
& \left\| \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \varepsilon_{ij} \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\| = O_{\mathbb{P}}(J_n^{1/2} N^{-1/2}), \\
& \left\| \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) - f_0(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\| = 0, \\
& \left\| \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( f_0(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0) \mathbf{X}_{ij}) - f_0(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\| = O_{\mathbb{P}}(J_n^{1/2} N^{-1/2}).
\end{aligned}$$

Thus, we have

$$\left\| \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( Y_i - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\| = O_{\mathbb{P}}(J_n^{1/2} N^{-1/2}).$$

□

### S1.3 Proof of Theorem 1

*Proof.* It follows that

$$\begin{aligned}
& \left\| \hat{f}^{(j)} - f_0 \right\|_2 \\
&= \left\| \hat{f}^{(j)} - \tilde{f}^{(j)} + \tilde{f}^{(j)} - f_0 \right\|_2 \\
&\leq \left\| \hat{f}^{(j)} - \tilde{f}^{(j)} \right\|_2 + \left\| \tilde{f}^{(j)} - f_0 \right\|_2.
\end{aligned}
\tag{S1.5}$$

The first term in Eq. S1.5 can be expressed as

$$\begin{aligned}
& \left\| \hat{f}^{(j)}(u) - \tilde{f}^{(j)}(u) \right\|_2^2 \\
&= \left\| \hat{\gamma}^{(j)\top} \mathbf{B}(u) - \tilde{\gamma}^{(j)\top} \mathbf{B}(u) \right\|_2^2 \\
&= (\hat{\gamma}^{(j)} - \tilde{\gamma}^{(j)})^\top E [\mathbf{B}(u) \mathbf{B}^\top(u)] (\hat{\gamma}^{(j)} - \tilde{\gamma}^{(j)}) \\
&\leq C \left\| \hat{\gamma}^{(j)} - \tilde{\gamma}^{(j)} \right\|^2 \leq C \left\| \hat{\boldsymbol{\theta}}^{(j)} - \tilde{\boldsymbol{\theta}}^{(j)} \right\|^2.
\end{aligned}$$

By Lemma 2, we have

$$\left\| \hat{f}^{(j)} - \tilde{f}^{(j)} \right\|_2 = O_{\mathbb{P}}(J_n^{1/2} n^{-1/2} + h^p).$$

Note that  $\left\| \tilde{f}^{(j)} - f_0 \right\|_2 \leq Ch^p$ , we have

$$\left\| \hat{f}^{(j)} - f_0 \right\|_2 = O_{\mathbb{P}}(J_n^{1/2} n^{-1/2} + h^p).$$

□

#### S1.4 Proof of Theorem 2

*Proof.* We use the notation  $\hat{U}_{\tau,i} = \boldsymbol{\alpha}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_i$ , and  $\hat{U}_\tau = \boldsymbol{\alpha}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \mathbf{X}$ . For any  $\mathbf{v} \in R^{d-1}$ , we define

$$\hat{m}(\mathbf{v}; \boldsymbol{\alpha}^\top(\hat{\boldsymbol{\tau}}^{(j)})) = \hat{f}^{(j)} \left( \hat{U}_\tau + \mathbf{v}^\top \mathbf{J}^\top(\hat{\boldsymbol{\tau}}) \tilde{\mathbf{X}}; \boldsymbol{\alpha}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \right),$$

and

$$\mathcal{M}_n = \{m(\mathbf{x}) = f(\boldsymbol{\alpha}^\top(\hat{\boldsymbol{\tau}})\mathbf{x}) : f \in \mathcal{H}\}.$$

Note that  $\hat{m}$  minimizes the function  $l(\hat{m}) = \frac{1}{2n} \sum_{i \in G_j} (Y_i - \hat{m}(\mathbf{X}_i))^2$  for all  $m \in \mathcal{M}_n$  when  $\mathbf{v} = \mathbf{0}$ , thus we have

$$\left. \frac{\partial l(\hat{m})}{\partial \mathbf{v}} \right|_{\mathbf{v}=\mathbf{0}} = \mathbf{0}.$$

Denote  $\hat{m}_i^{(j)} = \hat{f}^{(j)}(\hat{U}_{\tau,i}; \boldsymbol{\alpha}^\top(\hat{\boldsymbol{\tau}}^{(j)})) = \hat{\boldsymbol{\gamma}}^{(j)\top} \mathbf{B}(\hat{U}_{\tau,i})$ , thus get

$$\begin{aligned} \mathbf{0} &\equiv \left. \frac{\partial l(\hat{m})}{\partial \mathbf{v}} \right|_{\mathbf{v}} \\ &= -\frac{1}{n} \sum_{i \in G_j} \left( Y_i - \hat{m}_i^{(j)} \right) \hat{f}^{(j)'}(\hat{U}_{\tau,i}) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i \\ &= -\frac{1}{n} \sum_{i \in G_j} \left( Y_i - \hat{m}_i^{(j)} \right) f_0'(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \\ &\quad - \frac{1}{n} \sum_{i \in G_j} \left( Y_i - \hat{m}_i^{(j)} \right) f_0'(U_{\tau,0i}) (\mathbf{J}^\top(\boldsymbol{\tau}_0) - \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)})) \tilde{\mathbf{X}}_i \\ &\quad - \frac{1}{n} \sum_{i \in G_j} \left( Y_i - \hat{m}_i^{(j)} \right) \left( \hat{f}^{(j)'}(\hat{U}_{\tau,i}) - f_0'(U_{\tau,0i}) \right) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i. \end{aligned} \tag{S1.6}$$

Note that

$$\begin{aligned} \hat{f}^{(j)'}(\hat{U}_{\tau,i}) - f_0'(U_{\tau,0i}) &= \hat{f}^{(j)'}(\hat{U}_{\tau,i}) - \hat{f}^{(j)'}(U_{\tau,0i}) + \hat{f}^{(j)'}(U_{\tau,0i}) - f_0'(U_{\tau,0i}) \\ &= \hat{f}^{(j)''}(\hat{U}_{\tau,i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X}_i (\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) + \hat{f}^{(j)'}(U_{\tau,0i}) - f_0'(U_{\tau,0i}) + o_{\mathbb{P}}(n^{-1/2}). \end{aligned}$$

According to Lemma 7 of (Stone, 1986) and Theorem 2.1, we have

$$\left\| m_i - \hat{m}_i^{(j)} \right\|_{\infty} \leq C J_n^{1/2} \left\| m_i - \hat{m}_i^{(j)} \right\|_2 = O_{\mathbb{P}}(J_n n^{-1/2} + J_n^{1/2} h^p).$$

and by Lemma 3, we have

$$\begin{aligned} \left\| \hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0 \right\| &\leq \left\| \hat{\boldsymbol{\tau}}^{(j)} - \tilde{\boldsymbol{\tau}} \right\| + \left\| \tilde{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0 \right\| \\ &= O_{\mathbb{P}}(J_n^{1/2} n^{-1/2} + h^p). \end{aligned}$$

The definition of  $\tilde{\mathbf{X}}_i$  gives that

$$\begin{aligned} &\frac{1}{n} \sum_{i \in G_j} \left( Y_i - \hat{m}_i^{(j)} \right) \left( \hat{f}^{(j)'}(\hat{U}_{\tau,i}) - f'_0(U_{\tau,0i}) \right) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i \\ &= \frac{1}{n} \sum_{i \in G_j} \left( m_i - \hat{m}_i^{(j)} \right) \left( \hat{f}^{(j)'}(U_{\tau,0i}) - f'_0(U_{\tau,0i}) \right) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i \\ &\quad + \frac{1}{n} \sum_{i \in G_j} \left( m_i - \hat{m}_i^{(j)} \right) \hat{f}^{(j)''}(\bar{U}_{\tau,i}) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) (\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \\ &\quad + \frac{1}{n} \sum_{i \in G_j} \varepsilon_i \left( \hat{f}^{(j)'}(U_{\tau,0i}) - f'_0(U_{\tau,0i}) \right) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i \\ &\quad + \frac{1}{n} \sum_{i \in G_j} \varepsilon_i \hat{f}^{(j)''}(\bar{U}_{\tau,i}) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) (\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) + o_{\mathbb{P}}(n^{-1/2}) \\ &= o_{\mathbb{P}}(n^{-1/2}). \end{aligned}$$

We note that

$$\frac{1}{n} \sum_{i \in G_j} \left( Y_i - \hat{m}_i^{(j)} \right) f'_0(U_{\tau,0i}) (\mathbf{J}^\top(\boldsymbol{\tau}_0) - \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)})) \tilde{\mathbf{X}}_i = o_{\mathbb{P}}(n^{-1/2}).$$

This gives to

$$\begin{aligned} &\frac{1}{n} \sum_{i \in G_j} \left( Y_i - \hat{m}_i^{(j)} \right) \hat{f}^{(j)'}(\hat{U}_{\tau,i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \\ &= \frac{1}{n} \sum_{i \in G_j} \varepsilon_i f'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i + \frac{1}{n} \sum_{i \in G_j} \left( m_i - \hat{m}_i^{(j)} \right) f'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i. \end{aligned}$$

Note that

$$\begin{aligned}
m_i - \hat{m}_i^{(j)} &= f_0(U_{\tau,0i}) - \hat{f}^{(j)}(\hat{U}_{\tau,i}) \\
&= -f_0(U_{\tau,0i}) - f_0(\hat{U}_{\tau,i}) + f_0(\hat{U}_{\tau,i}) - \hat{f}^{(j)}(\hat{U}_{\tau,i}) \\
&= f'_0(U_{\tau,0i}) \tilde{\mathbf{X}}_i^\top \mathbf{J}(\boldsymbol{\tau}_0)(\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) + f'_0(U_{\tau,0i}) \boldsymbol{\Phi}(\mathbf{X}_i)^\top \mathbf{J}(\boldsymbol{\tau}_0)(\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \\
&\quad + f_0(\hat{U}_{\tau,i}) - \hat{f}^{(j)}(\hat{U}_{\tau,i}) + o_{\mathbb{P}}(n^{-1/2}).
\end{aligned}$$

Thus, the second term in Eq. S1.6 can be expressed

$$\begin{aligned}
&\frac{1}{n} \sum_{i \in G_j} (m_i - \hat{m}_i^{(j)}) f'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \\
&= -\frac{1}{n} \sum_{i \in G_j} \{f'_0\}^2(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^\top \mathbf{J}(\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \\
&\quad + \frac{1}{n} \sum_{i \in G_j} \{f'_0\}^2(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \boldsymbol{\Phi}(\mathbf{X}_i)^\top \mathbf{J}(\boldsymbol{\tau}_0)(\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \\
&\quad + \frac{1}{n} \sum_{i \in G_j} (f_0(\hat{U}_{\tau,i}) - \hat{f}^{(j)}(\hat{U}_{\tau,i})) f'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i.
\end{aligned}$$

The Lemma 3 and Theorem 2.1 leads to

$$\begin{aligned}
&\frac{1}{n} \sum_{i \in G_j} (f_0(\hat{U}_{\tau,i}) - \hat{f}^{(j)}(\hat{U}_{\tau,i})) f'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i = o_{\mathbb{P}}(n^{-1/2}), \\
&\frac{1}{n} \sum_{i \in G_j} \{f'_0\}^2(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \boldsymbol{\Phi}(\mathbf{X}_i)^\top \mathbf{J}(\boldsymbol{\tau}_0)(\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) = o_{\mathbb{P}}(n^{-1/2}).
\end{aligned}$$

Then we have

$$\begin{aligned}
\mathbf{0} &= \frac{1}{n} \sum_{i \in G_j} \varepsilon_i f'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \\
&\quad - \frac{1}{n} \sum_{i \in G_j} \{f'_0\}^2(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^\top \mathbf{J}(\boldsymbol{\tau}_0)(\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) + o_{\mathbb{P}}(n^{-1/2}).
\end{aligned}$$

The central limit theorem gives that

$$\sqrt{n}(\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1} \boldsymbol{\Sigma} \mathbf{A}^{-1}),$$

where

$$\begin{aligned} \mathbf{A} &= E \left[ \{f'_0\}^2 (U_{\tau,0}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}} \tilde{\mathbf{X}}^\top \mathbf{J}(\boldsymbol{\tau}_0) \right], \\ \boldsymbol{\Sigma} &= E \left[ \{\varepsilon f'_0(U_{\tau,0}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}\}^{\otimes 2} \right]. \end{aligned}$$

By using the delta method, we have the distribution of  $\sqrt{n}(\hat{\boldsymbol{\alpha}}^{(j)} - \boldsymbol{\alpha}_0)$  as follows

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}^{(j)} - \boldsymbol{\alpha}_0) \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{J}(\boldsymbol{\tau}_0) \mathbf{A}^{-1} \boldsymbol{\Sigma} \mathbf{A}^{-1} \mathbf{J}^\top(\boldsymbol{\tau}_0)).$$

□

### S1.5 Proof of Theorem 3

*Proof.* The triangle inequality shows that

$$\|\bar{f} - f_0\|_2 \leq \left\| \frac{1}{s} \sum_{j=1}^s \hat{f}^{(j)} - \frac{1}{s} \sum_{j=1}^s \tilde{f}^{(j)} \right\|_2 + \left\| \frac{1}{s} \sum_{j=1}^s \tilde{f}^{(j)} - f_0 \right\|_2. \quad (\text{S1.7})$$

The first term in Eq. S1.7 can be bounded as follows

$$\begin{aligned} & \left\| \frac{1}{s} \sum_{j=1}^s \hat{f}^{(j)} - \frac{1}{s} \sum_{j=1}^s \tilde{f}^{(j)} \right\|_2^2 \\ &= \frac{1}{s^2} \sum_{j=1}^s (\hat{\boldsymbol{\gamma}}^{(j)} - \tilde{\boldsymbol{\gamma}}^{(j)})^\top E [\mathbf{B}(u) \mathbf{B}^\top(u)] \sum_{j=1}^s (\hat{\boldsymbol{\gamma}}^{(j)} - \tilde{\boldsymbol{\gamma}}^{(j)}) \\ &\leq C \left\| \frac{1}{s} \sum_{j=1}^s (\hat{\boldsymbol{\gamma}}^{(j)} - \tilde{\boldsymbol{\gamma}}^{(j)}) \right\|^2 \leq C \left\| \frac{1}{s} \sum_{j=1}^s (\hat{\boldsymbol{\theta}}^{(j)} - \tilde{\boldsymbol{\theta}}) \right\|^2. \end{aligned}$$

In the following we focus on the term  $\left\| \frac{1}{s} \sum_{j=1}^s (\hat{\boldsymbol{\theta}}^{(j)} - \tilde{\boldsymbol{\theta}}^{(j)}) \right\|$ . The Lemma

3 gives that

$$\hat{\boldsymbol{\theta}}^{(j)} - \tilde{\boldsymbol{\theta}}^{(j)} = - \left( \frac{\partial^2 l^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \right)^{-1} \frac{\partial l^{(j)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}^{(j)}}.$$

Note that  $\left\| (V_{n,\theta}^{(j)})^{-1} \right\|_2 = O(1)$ , a.s., then the term  $\frac{1}{s} \sum_{j=1}^s (\hat{\boldsymbol{\theta}}^{(j)} - \tilde{\boldsymbol{\theta}}^{(j)})$  is equivalent to

$$\frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( Y_i - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right) \begin{pmatrix} \tilde{f}^{(j)'}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij} \\ \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \end{pmatrix}.$$

The Lemma 4 and 5 leads to

$$\begin{aligned} \left\| \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( Y_i - \tilde{f}^{(j)}(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij} \right) \mathbf{J}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij} \right\| &= O_{\mathbb{P}}(N^{-1/2} + h^p), \\ \left\| \frac{1}{N} \sum_{j=1}^s \sum_{i \in G_j} \left( Y_i - \tilde{f}^{(j)}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right) \mathbf{B}(\boldsymbol{\alpha}^\top(\tilde{\boldsymbol{\tau}}^{(j)}) \mathbf{X}_{ij}) \right\| &= O_{\mathbb{P}}(J_n^{1/2} N^{-1/2} + h^p). \end{aligned}$$

Thus, we have

$$\left\| \frac{1}{s} \sum_{j=1}^s (\hat{\boldsymbol{\theta}}^{(j)} - \tilde{\boldsymbol{\theta}}^{(j)}) \right\| = O_{\mathbb{P}}(J_n^{1/2} N^{-1/2} + h^p),$$

which implies that

$$\left\| \frac{1}{s} \sum_{j=1}^s \hat{f}^{(j)} - \frac{1}{s} \sum_{j=1}^s \tilde{f}^{(j)} \right\|_2 = O_{\mathbb{P}}(J_n^{-1/2} N^{-1/2} + h^p).$$

Now we give the bound of the second term in Eq. S1.7. Note that

$$\left\| \frac{1}{s} \sum_{j=1}^s \tilde{f}^{(j)} - f_0 \right\|_2 = \left\| \frac{1}{s} \sum_{j=1}^s (\tilde{f}^{(j)} - f_0) \right\|_2.$$

From  $\left\| \tilde{f}^{(j)} - f_0 \right\|_{\infty} \leq Ch^p$ , we have

$$\left\| \frac{1}{s} \sum_{j=1}^s \tilde{f}^{(j)} - f_0 \right\|_2 = O_{\mathbb{P}}(h^p).$$

Thus, we have

$$\left\| \bar{f} - f_0 \right\|_2 = O_{\mathbb{P}}(J_n^{1/2} N^{-1/2} + h^p).$$

□

## S1.6 Proof of Theorem 4

*Proof.* We first prove that  $E(\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) = o_{\mathbb{P}}(N^{-1/2})$ . Similar to the proof of Theorem 2.2, we will show that

$$\frac{1}{n} \sum_{i \in G_j} \left( Y_i - \hat{m}_i^{(j)} \right) \left( \hat{f}^{(j)'}(\hat{U}_{\tau,i}) - f_0'(U_{\tau,0i}) \right) \mathbf{J}^{\top}(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i = o_{\mathbb{P}}(N^{-1/2}).$$

From Theorem 2.1, we know that

$$\begin{aligned} \left\| \hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0 \right\| &= O_{\mathbb{P}}(J_n^{1/2} n^{-1/2} + h^p), \\ \left\| m_i - \hat{m}_i^j \right\|_{\infty} &= O_{\mathbb{P}}(J_n n^{-1/2} + J_n^{1/2} h^p). \end{aligned}$$

Thus, we have

$$\begin{aligned}
 & \frac{1}{n} \sum_{i \in G_j} \left( Y_i - \hat{m}_i^{(j)} \right) \left( \hat{f}^{(j)'}(U_{\tau,0i}) - f_0'(U_{\tau,0i}) \right) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i \\
 = & \frac{1}{n} \sum_{i \in G_j} \left( m_i - \hat{m}_i^{(j)} \right) \left( \hat{f}^{(j)'}(\hat{U}_{\tau,i}) - f_0'(\hat{U}_{\tau,i}) \right) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i \\
 & + \frac{1}{n} \sum_{i \in G_j} \left( m_i - \hat{m}_i^{(j)} \right) \hat{f}^{(j)''}(\bar{U}_{\tau,i}) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) (\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \\
 & + \frac{1}{n} \sum_{i \in G_j} \varepsilon_i \left( \hat{f}^{(j)'}(U_{\tau,0i}) - f_0'(U_{\tau,0i}) \right) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i \\
 & + \frac{1}{n} \sum_{i \in G_j} \varepsilon_i \hat{f}^{(j)''}(\bar{U}_{\tau,i}) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) (\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) + o_{\mathbb{P}}(J_n n^{-1} + J_n^{1/2} n^{-1/2} h^p) \\
 = & o_{\mathbb{P}}(J_n n^{-1} + J_n^{1/2} n^{-1/2} h^p).
 \end{aligned}$$

The assumption  $n \gg N^{5/8}$  leads to

$$\frac{1}{n} \sum_{i \in G_j} \left( Y_i - \hat{m}_i^{(j)} \right) \left( \hat{f}^{(j)'}(U_{\tau,0i}) - f_0'(U_{\tau,0i}) \right) \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)}) \tilde{\mathbf{X}}_i = o_{\mathbb{P}}(N^{-1/2}),$$

and

$$\frac{1}{n} \sum_{i \in G_j} \left( Y_i - \hat{m}_i^{(j)} \right) f_0'(U_{\tau,0i}) (\mathbf{J}^\top(\boldsymbol{\tau}_0) - \mathbf{J}^\top(\hat{\boldsymbol{\tau}}^{(j)})) \tilde{\mathbf{X}}_i = o_{\mathbb{P}}(N^{-1/2}).$$

Thus, we have

$$\begin{aligned}
 \mathbf{0} &= \frac{1}{n} \sum_{i \in G_j} \left( Y_i - \hat{m}_i^{(j)} \right) \hat{f}^{(j)'}(\hat{U}_{\tau,i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \\
 &= \frac{1}{n} \sum_{i \in G_j} \varepsilon_i f_0'(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i + \frac{1}{n} \sum_{i \in G_j} \left( m_i - \hat{m}_i^{(j)} \right) f_0'(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i + o_{\mathbb{P}}(N^{-1/2}).
 \end{aligned}$$

where the second term can be expressed

$$\begin{aligned}
 & \frac{1}{n} \sum_{i \in G_j} \left( m_i - \hat{m}_i^{(j)} \right) f'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \\
 = & \frac{1}{n} \sum_{i \in G_j} \{f'_0\}^2(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) (\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \\
 & + \frac{1}{n} \sum_{i \in G_j} \{f'_0\}^2(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \boldsymbol{\Phi}(\mathbf{X}_i)^\top \mathbf{J}(\boldsymbol{\tau}_0) (\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \\
 & + \frac{1}{n} \sum_{i \in G_j} \left( f_0(\hat{U}_{\tau,i}) - \hat{f}^{(j)}(\hat{U}_{\tau,i}) \right) f'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i.
 \end{aligned}$$

Note that

$$\begin{aligned}
 & \frac{1}{n} \sum_{i \in G_j} \{f'_0\}^2(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \boldsymbol{\Phi}(\mathbf{X}_i)^\top \mathbf{J}(\boldsymbol{\tau}_0) (\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) = o_{\mathbb{P}}(J_n^{1/2} n^{-1} + n^{-1/2} h^p) = o_{\mathbb{P}}(N^{-1/2}), \\
 & \frac{1}{n} \sum_{i \in G_j} \left( f_0(\hat{U}_{\tau,i}) - \hat{f}^{(j)}(\hat{U}_{\tau,i}) \right) f'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i = o_{\mathbb{P}}(J_n n^{-1} + J_n^{1/2} n^{-1/2} h^p) = o_{\mathbb{P}}(N^{-1/2}).
 \end{aligned}$$

Then, we have

$$\begin{aligned}
 \mathbf{0} &= \frac{1}{n} \sum_{i \in G_j} \varepsilon_i f'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \\
 &+ \frac{1}{n} \sum_{i \in G_j} \{f'_0\}^2(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) (\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) + o_{\mathbb{P}}(N^{-1/2}).
 \end{aligned}$$

Note that

$$0 < c \leq E \left\| \frac{1}{n} \sum_{i \in G_j} \{f'_0\}^2(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) \right\| \leq C.$$

This leads to

$$E(\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) = o_{\mathbb{P}}(N^{-1/2}).$$

Theorem 2.2 gives that

$$\sqrt{n}(\hat{\boldsymbol{\tau}}^{(j)} - \boldsymbol{\tau}_0) \rightarrow N(\mathbf{0}, \mathbf{A}^{-1} \boldsymbol{\Sigma} \mathbf{A}^{-1}),$$

where

$$\begin{aligned}\mathbf{A} &= E \left[ \{f'_0\}^2 (U_{\tau,0}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}} \tilde{\mathbf{X}}^\top \mathbf{J}(\boldsymbol{\tau}_0) \right], \\ \boldsymbol{\Sigma} &= E \left[ \{\varepsilon f'_0(U_{\tau,0}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \tilde{\mathbf{X}}\}^{\otimes 2} \right].\end{aligned}$$

Thus, we have

$$\sqrt{N}(\bar{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1} \boldsymbol{\Sigma} \mathbf{A}^{-1}).$$

From the delta method, we can obtain the distribution of  $\sqrt{N}(\bar{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$  as follows

$$\sqrt{N}(\bar{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{J}(\boldsymbol{\tau}_0) \mathbf{A}^{-1} \boldsymbol{\Sigma} \mathbf{A}^{-1} \mathbf{J}^\top(\boldsymbol{\tau}_0)).$$

□

## S2 Proof of Section 3

### S2.1 Supplementary Differential Privacy Results

In this appendix we collect the standard results on differential privacy that are used throughout the paper.

**Lemma 6** (Composition Theorem (Dwork et al., 2006)). *Let  $\mathcal{M}_0 : \mathbb{X}^n \rightarrow \mathbb{R}_0^d$  be  $(\varepsilon, \delta)$ -differentially private. For  $i = 1, \dots, k-1$ , let*

$$\mathcal{M}_i : \mathbb{X}^n \times \mathbb{R}_{i-1}^d \rightarrow \mathbb{R}_i^d$$

be  $(\varepsilon, \delta)$ -differentially private conditioned on the output of  $\mathcal{M}_{i-1}$ . Then the composed mechanism  $\mathcal{M} = \mathcal{M}_{k-1} \circ \dots \circ \mathcal{M}_0$  is  $(k\varepsilon, k\delta)$ -differentially private.

**Lemma 7** (Post-processing (Dwork et al., 2006)). *Let  $\mathcal{M} : \mathbb{X}^n \rightarrow \mathbb{R}^d$  be  $(\varepsilon, \delta)$ -differentially private. Then for any measurable  $g : \mathbb{R}^d \rightarrow \mathbb{R}^m$ ,  $g(\mathcal{M}(Z))$  is also  $(\varepsilon, \delta)$ -differentially private.*

**Lemma 8** (Moments Accountant and Rényi Divergence (Abadi et al., 2016; Mironov, 2017)). *Let  $\mathcal{M}$  be a randomized mechanism and  $Z, Z'$  neighboring datasets. For any  $\lambda > 0$ , define the  $\lambda$ -th moment of the privacy loss as*

$$\alpha_{\mathcal{M}}(\lambda) := \sup_{Z \sim Z'} \log \mathbb{E}_{o \sim \mathcal{M}(Z)} \left[ \left( \frac{\Pr[\mathcal{M}(Z)=o]}{\Pr[\mathcal{M}(Z')=o]} \right)^\lambda \right].$$

*If  $\mathcal{M}(Z) \sim P$  and  $\mathcal{M}(Z') \sim Q$  admit densities, then*

$$\alpha_{\mathcal{M}}(\lambda) = \lambda D_{\lambda+1}(P \| Q),$$

*where  $D_{\lambda+1}$  is the Rényi divergence of order  $\lambda + 1$ .*

**Lemma 9** (Parallel Composition (Dwork et al., 2006)). *Let  $\mathcal{M}_i : \mathbb{X}_i^{n_i} \rightarrow \mathbb{R}_i^d$ ,  $i = 1, \dots, k$ , be  $(\varepsilon, \delta)$ -differentially private mechanisms, where  $Z_i \in \mathbb{X}_i^{n_i}$  are mutually disjoint datasets. Then the joint mechanism  $\mathcal{M}(Z_1, \dots, Z_k) = (\mathcal{M}_1(Z_1), \dots, \mathcal{M}_k(Z_k))$  is also  $(\varepsilon, \delta)$ -differentially private.*

Together with the Gaussian mechanism (Definition 2.2 in the main text), these results provide the core privacy-accounting tools used in our algorithms.

## S2.2 Proof of Lemma 4

*Proof.* Define the joint parameter  $\boldsymbol{\theta} := \begin{pmatrix} \boldsymbol{\gamma}^{(j)} \\ \boldsymbol{\tau}^{(j)} \end{pmatrix}$  and set  $u_i := \boldsymbol{\alpha}(\boldsymbol{\tau}^{(j)})^\top \mathbf{X}_i$ .

The empirical loss is given by

$$L_n^{(j)}(\boldsymbol{\theta}) = \frac{1}{2n} \sum_{i=1}^n (\mathbf{B}^\top(u_i) \boldsymbol{\gamma}^{(j)} - Y_i)^2.$$

The gradient with respect to  $\boldsymbol{\theta}$  is

$$\nabla_{\boldsymbol{\theta}} L_n^{(j)}(\boldsymbol{\theta}) = \begin{bmatrix} \nabla_{\boldsymbol{\gamma}^{(j)}} L_n^{(j)} \\ \nabla_{\boldsymbol{\tau}^{(j)}} L_n^{(j)} \end{bmatrix} = \frac{1}{n} \sum_{i=1}^n \begin{bmatrix} (\boldsymbol{\gamma}^{\top(j)} \mathbf{B}(u_i) - Y_i) \mathbf{B}(u_i) \\ (\boldsymbol{\gamma}^{\top(j)} \mathbf{B}(u_i) - Y_i) \boldsymbol{\gamma}^{(j)\top} \mathbf{B}'(u_i) \mathbf{J}^\top(\boldsymbol{\tau}^{(j)}) \mathbf{X}_i \end{bmatrix}.$$

Using  $\|\mathbf{B}(u_i)\| \leq c_b$ ,  $\|\mathbf{B}'(u_i)\| \leq c'_b$ ,  $\|\boldsymbol{\gamma}^{(j)}\| \leq c_0$ ,  $\|\mathbf{J}(\boldsymbol{\tau}^{(j)})\| \leq C_\kappa$ ,  $\|\mathbf{X}_i\| \leq c_X$ , and  $|Y_i| \leq R$ , we obtain

$$|\mathbf{B}^\top(u_i) \boldsymbol{\gamma}^{(j)} - Y_i| \leq c_b c_0 + R, \quad |\boldsymbol{\gamma}^{(j)\top} \mathbf{B}'(u_i)| \leq c_0 c'_b.$$

Then the gradient norms satisfy

$$\|\nabla_{\boldsymbol{\gamma}^{(j)}} L_n^{(j)}(\boldsymbol{\theta})\| \leq (c_b c_0 + R) c_b, \quad \|\nabla_{\boldsymbol{\tau}^{(j)}} L_n^{(j)}(\boldsymbol{\theta})\| \leq (c_b c_0 + R) c_0 c'_b C_\kappa c_X.$$

Therefore,

$$\|\nabla_{\boldsymbol{\theta}} L_n^{(j)}(\boldsymbol{\theta})\| \leq (c_b c_0 + R) \sqrt{c_b^2 + (c_0 c'_b C_\kappa c_X)^2} =: L_\theta.$$

Hence,  $L_n^{(j)}(\boldsymbol{\theta})$  is  $L_\theta$ -Lipschitz.

The Hessian with respect to  $\boldsymbol{\theta}$  is block-structured:

$$\nabla_{\boldsymbol{\theta}}^2 L_n^{(j)}(\boldsymbol{\theta}) = \begin{bmatrix} H_{\gamma\gamma} & H_{\gamma\tau} \\ H_{\tau\gamma} & H_{\tau\tau} \end{bmatrix},$$

where

$$\begin{aligned}
 H_{\gamma\gamma} &= \frac{1}{n} \sum_{i=1}^n \mathbf{B}(u_i) \mathbf{B}^\top(u_i), \\
 H_{\gamma\tau} &= \frac{1}{n} \sum_{i=1}^n \left[ (\boldsymbol{\gamma}^{(j)\top} \mathbf{B}'(u_i)) \mathbf{J}^\top(\boldsymbol{\tau}^{(j)}) \mathbf{X}_i \mathbf{B}^\top(u_i) + (\mathbf{B}^\top(u_i) \boldsymbol{\gamma}^{(j)} - Y_i) \mathbf{J}^\top(\boldsymbol{\tau}^{(j)}) \mathbf{X}_i \mathbf{B}'^\top(u_i) \right], \\
 H_{\tau\tau} &= \frac{1}{n} \sum_{i=1}^n \left[ (\boldsymbol{\gamma}^{(j)\top} \mathbf{B}'(u_i))^2 \mathbf{J}^\top(\boldsymbol{\tau}^{(j)}) \mathbf{X}_i \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}^{(j)}) \right. \\
 &\quad + (\mathbf{B}^\top(u_i) \boldsymbol{\gamma}^{(j)} - Y_i) (\boldsymbol{\gamma}^{(j)\top} \mathbf{B}''(u_i)) \mathbf{J}^\top(\boldsymbol{\tau}^{(j)}) \mathbf{X}_i \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}^{(j)}) \\
 &\quad \left. + (\mathbf{B}^\top(u_i) \boldsymbol{\gamma}^{(j)} - Y_i) (\boldsymbol{\gamma}^{(j)\top} \mathbf{B}'(u_i)) \frac{\partial \mathbf{J}^\top(\boldsymbol{\tau}^{(j)})}{\partial \boldsymbol{\tau}^{(j)}} \mathbf{X}_i \right].
 \end{aligned}$$

By assumptions, we have

$$\begin{aligned}
 |\boldsymbol{\gamma}^{(j)\top} \mathbf{B}'(u_i)| &\leq c_0 c'_b, \quad |\boldsymbol{\gamma}^{(j)\top} \mathbf{B}''(u_i)| \leq c_0 c''_b, \\
 |\mathbf{B}^\top(u_i) \boldsymbol{\gamma}^{(j)} - Y_i| &\leq c_0 c_b + R, \quad \|\mathbf{J}(\boldsymbol{\tau}^{(j)})\|_2 \leq C_\kappa, \quad \left\| \frac{\partial \mathbf{J}^\top(\boldsymbol{\tau}^{(j)})}{\partial \boldsymbol{\tau}^{(j)}} \right\|_2 \leq C'_\kappa, \quad \|\mathbf{X}_i\| \leq c_X,
 \end{aligned}$$

Thus, we can bound these three terms,

$$\begin{aligned}
 \|H_{\gamma\gamma}\| &\leq c_b^2, \\
 \|H_{\gamma\tau}\| &\leq c_0 c'_b C_\kappa c_X c_b + (c_b c_0 + R) C_\kappa c_X c'_b, \\
 \|H_{\tau\tau}\| &\leq (c_0 c'_b)^2 C_\kappa^2 c_X^2 + (c_b c_0 + R) c_0 c'_b C_\kappa^2 c_X^2 + (c_b c_0 + R) c_0 c'_b C'_\kappa c_X^2.
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \|\nabla_{\boldsymbol{\theta}}^2 L_n^{(j)}(\boldsymbol{\theta})\|_2 &\leq \max\{c_b^2, (c_0 c'_b)^2 C_\kappa^2 c_X^2 + (c_b c_0 + R) c_0 c'_b C_\kappa^2 c_X^2 + (c_b c_0 + R) c_0 c'_b C'_\kappa c_X^2\} \\
 &\quad + c_0 c'_b C_\kappa c_X c_b + (c_b c_0 + R) C_\kappa c_X c'_b =: M_\theta.
 \end{aligned}$$

Hence,  $L_n^{(j)}(\boldsymbol{\theta})$  is  $L_\theta$ -Lipschitz and  $M_\theta$ -smooth. This completes the proof.  $\square$

### S2.3 proof of Theorem 5

*Proof.* In each iteration  $t$  of Algorithm 1, the update rule follows projected noisy gradient descent:

$$\begin{aligned}\boldsymbol{\gamma}_{t+1}^{(j)} &= \Pi_{c_\gamma} \left( \boldsymbol{\gamma}_t^{(j)} - \eta \left[ \nabla_\gamma L_n^{(j)}(\boldsymbol{\tau}_t^{(j)}, \boldsymbol{\gamma}_t^{(j)}) + \boldsymbol{\omega}_t^\gamma \right] \right), \quad \boldsymbol{\omega}_t^\gamma \sim \mathcal{N}(0, \sigma_\theta^2 I_{J_n + \rho + 1}), \\ \boldsymbol{\tau}_{t+1}^{(j)} &= \Pi_{c_\tau} \left( \boldsymbol{\tau}_t^{(j)} - \eta \left[ \nabla_\tau L_n^{(j)}(\boldsymbol{\tau}_t^{(j)}, \boldsymbol{\gamma}_t^{(j)}) + \boldsymbol{\omega}_t^\tau \right] \right), \quad \boldsymbol{\omega}_t^\tau \sim \mathcal{N}(0, \sigma_\theta^2 I_{d-1}),\end{aligned}$$

Define the constraint set  $\mathcal{C} = c_\gamma \times c_\tau$  and the gradient mapping with step size  $\eta$  as:

$$G_\eta(\boldsymbol{\theta}) := \frac{1}{\eta} (\boldsymbol{\theta} - \Pi_{\mathcal{C}}(\boldsymbol{\theta} - \eta \nabla_\theta L_n^{(j)}(\boldsymbol{\theta}))).$$

Thus, the update can be compactly written as

$$\boldsymbol{\theta}_{t+1}^{(j)} = \Pi_{\mathcal{C}} \left( \boldsymbol{\theta}_t^{(j)} - \eta [\nabla_\theta L_n^{(j)}(\boldsymbol{\theta}_t^{(j)}) + \boldsymbol{\omega}_t] \right),$$

where  $\boldsymbol{\omega}_t = \begin{pmatrix} \boldsymbol{\omega}_t^\gamma \\ \boldsymbol{\omega}_t^\tau \end{pmatrix} \sim \mathcal{N}(0, \sigma^2 I_p)$  and  $p = (J_n + \rho + 1) + (d - 1)$ . For clarity, we denote it by  $\tilde{\nabla}_{\boldsymbol{\theta}, t}$  (stacking the two blocks), rather than the gradient of any post-processed iterate.

To ensure privacy, we apply the Gaussian mechanism to the stacked vector:

$$\mathcal{M}_{\boldsymbol{\theta}, t} := \tilde{\nabla}_{\boldsymbol{\theta}, t} + \mathcal{N}(0, \sigma_\theta^2 I_p),$$

where  $p$  is the dimension of  $\boldsymbol{\theta}$ . The released noisy statistic is then post-processed (projection and parameter update), which preserves privacy.

Let  $Z$  and  $Z'$  be neighboring datasets differing in one sample. Define the output distributions:

$$P = \mathcal{N}\left(\tilde{\nabla}_{\boldsymbol{\theta},t}(Z), \sigma_{\theta}^2 I\right), \quad Q = \mathcal{N}\left(\tilde{\nabla}_{\boldsymbol{\theta},t}(Z'), \sigma_{\theta}^2 I\right).$$

By the moments accountant, the  $\lambda$ -th moment of the privacy loss is

$$\alpha_{\mathcal{M}_{\theta,t}}(\lambda) = \lambda \cdot D_{\lambda+1}(P \parallel Q),$$

and by the Gaussian Rényi divergence formula (Mironov, 2017),

$$D_{\lambda+1}(P \parallel Q) = \frac{\lambda+1}{2\sigma^2} \left\| \tilde{\nabla}_{\boldsymbol{\theta},t}(Z) - \tilde{\nabla}_{\boldsymbol{\theta},t}(Z') \right\|^2.$$

From Lemma 4, there exists a constant  $L_{\theta}$  such that

$$\left\| \tilde{\nabla}_{\boldsymbol{\theta},t}(Z) - \tilde{\nabla}_{\boldsymbol{\theta},t}(Z') \right\| \leq \frac{2L_{\theta}}{n}.$$

Substituting yields

$$\alpha_{\mathcal{M}_{\theta,t}}(\lambda) \leq \frac{2L_{\theta}^2 \lambda(\lambda+1)}{n^2 \sigma^2} \leq \frac{c L_{\theta}^2 \lambda^2}{n^2 \sigma_{\theta}^2},$$

for some absolute constant  $c > 0$ . By composition over  $T$  rounds,

$$\alpha_{\mathcal{M}_{\theta}}(\lambda) \leq c \cdot \frac{T L_{\theta}^2 \lambda^2}{n^2 \sigma_{\theta}^2}.$$

Hence, to ensure  $(\varepsilon, \delta)$ -differential privacy for the entire procedure, it suffices

to choose

$$\sigma_{\theta}^2 = C \cdot \frac{T L_{\theta}^2 \log(1/\delta)}{n^2 \varepsilon^2},$$

for a universal constant  $C > 0$ . Then there exists  $\lambda > 0$  such that

$$\alpha_{\mathcal{M}_\theta}(\lambda) \leq \frac{\lambda\varepsilon}{2}, \quad \exp\left(-\frac{\lambda\varepsilon}{2}\right) \leq \delta,$$

which implies that the *joint* mechanism  $\mathcal{M}_\theta$  is  $(\varepsilon, \delta)$ -DP. By post-processing, the released iterates and each block's update (including the  $\boldsymbol{\tau}$ - and  $\boldsymbol{\gamma}$ -components) are also  $(\varepsilon, \delta)$ -DP.

The above analysis establishes that the local mechanism executed on each machine satisfies  $(\varepsilon, \delta)$ -differential privacy. Since the local datasets are mutually disjoint, Lemma 9 implies that the collection of all local outputs is also  $(\varepsilon, \delta)$ -differentially private. Finally, the aggregated estimators are obtained by averaging the local outputs, which is a deterministic post-processing operation. Therefore, by Lemma 7, the aggregated estimators also satisfy  $(\varepsilon, \delta)$ -differential privacy.  $\square$

#### S2.4 Proof of Theorem 6

*Proof.* We analyze the convergence of the private estimator under the following high-probability events:

$$\mathcal{E}_1 := \{\Pi_R(Y_i) = Y_i, \forall i \in [n]\}.$$

The update rule follows projected noisy gradient descent:

$$\begin{aligned}\gamma_{t+1}^{(j)} &= \Pi_{c_\gamma} \left( \gamma_t^{(j)} - \eta \left[ \nabla_\gamma L_n^{(j)}(\tau_t^{(j)}, \gamma_t^{(j)}) + \omega_t^\gamma \right] \right), \quad \omega_t^\gamma \sim \mathcal{N}(0, \sigma_\theta^2 I_{J_n + \rho + 1}), \\ \tau_{t+1}^{(j)} &= \Pi_{c_\tau} \left( \tau_t^{(j)} - \eta \left[ \nabla_\tau L_n^{(j)}(\tau_t^{(j)}, \gamma_t^{(j)}) + \omega_t^\tau \right] \right), \quad \omega_t^\tau \sim \mathcal{N}(0, \sigma_\theta^2 I_{d-1}),\end{aligned}$$

where the step size is fixed at  $\eta = 1/M_\theta$ , and the loss is  $M_\theta$ -smooth in  $\theta = \begin{pmatrix} \gamma \\ \tau \end{pmatrix}$ .

Define the constraint set  $\mathcal{C} = c_\gamma \times c_\tau$  and the gradient mapping with step size  $\eta$  as:

$$G_\eta(\theta) := \frac{1}{\eta} (\theta - \Pi_{\mathcal{C}}(\theta - \eta \nabla_\theta L_n^{(j)}(\theta))).$$

Thus, the update can be compactly written as

$$\theta_{t+1}^{(j)} = \Pi_{\mathcal{C}} \left( \theta_t^{(j)} - \eta [\nabla_\theta L_n^{(j)}(\theta_t^{(j)}) + \omega_t] \right),$$

where  $\omega_t = \begin{pmatrix} \omega_t^\gamma \\ \omega_t^\tau \end{pmatrix}$ . Then, we have:

$$L_n^{(j)}(\theta_{t+1}^{(j)}) \leq L_n^{(j)}(\theta_t^{(j)}) - \frac{\eta}{2} \|G_\eta(\theta_t^{(j)})\|^2 + \frac{\eta}{2} \|\omega_t\|^2$$

Now, summing over all iterations  $t = 0$  to  $T - 1$  and dividing by  $T$ , we define the average gradient norm:

$$\bar{G}_T := \frac{1}{T} \sum_{t=0}^{T-1} \|G_\eta(\theta_t^{(j)})\|^2.$$

Then summing the inequality gives:

$$\begin{aligned}\bar{G}_T &\leq \frac{2M_\theta}{T} \sum_{t=0}^{T-1} \left( L_n^{(j)}(\boldsymbol{\theta}_t^{(j)}) - L_n^{(j)}(\boldsymbol{\theta}_{t+1}^{(j)}) \right) + \frac{2}{T} \sum_{t=0}^{T-1} \|\boldsymbol{\omega}_t\|^2 \\ &= \frac{2M_\theta}{T} \left( L_n^{(j)}(\boldsymbol{\theta}_0^{(j)}) - L_n^{(j)}(\boldsymbol{\theta}_T^{(j)}) \right) + \frac{2}{T} \sum_{t=0}^{T-1} \|\boldsymbol{\omega}_t\|^2.\end{aligned}$$

Now we control the noise term. By standard concentration inequalities, we have

$$\frac{1}{T} \sum_{t=0}^{T-1} \|\boldsymbol{\omega}_t\|^2 \lesssim \sigma_\theta^2(d + J_n + \rho),$$

with probability at least  $1 - \exp(-c_1\sqrt{T}(d + J_n + \rho))$ .

Finally, under the event  $\mathcal{E}_1$ , we have:

$$\bar{G}_T \lesssim \frac{L_\theta \sqrt{M_\theta \log(1/\delta)}}{n\varepsilon} + \frac{L_\theta(d + J_n + \rho) \sqrt{M_\theta \log(1/\delta)}}{n\varepsilon}.$$

with probability at least  $1 - \exp(-c_1\sqrt{T}(d + J_n + \rho))$ .

Similarly, we have  $\mathcal{E}_1$  occurs with probability at least  $1 - c_2 \exp(-c_3 n)$ .

Thus, we conclude that

$$\bar{G}_T \lesssim \frac{L_\theta \sqrt{M_\theta \log(1/\delta)}}{n\varepsilon} + \frac{L_\theta(d + J_n + \rho) \sqrt{M_\theta \log(1/\delta)}}{n\varepsilon}.$$

with probability at least  $1 - c_2 \exp(-c_3 n) - \exp(-c_1\sqrt{T}(d + J_n + \rho))$ .  $\square$

### S3 Supplementary Simulation Results

Unless otherwise specified, all supplementary simulations use cubic splines with degree  $\rho = 3$ , equally spaced knots, and spline basis dimensions  $J_n \in$

$\{n^{1/5}, n^{1/6}, n^{1/7}\}$ . We consider the following additional data-generating models:

$$\mathcal{M}_2 : Y = \sin(\boldsymbol{\alpha}^\top \mathbf{x}) + \varepsilon, \text{ where } \boldsymbol{\alpha} = (1/\sqrt{6}, 1/\sqrt{6}, 1/\sqrt{6}, -1/\sqrt{6}, 1/\sqrt{6}, 1/\sqrt{6})^\top, \\ \varepsilon \sim \mathcal{N}(0, 0.2^2), \text{ and } \mathbf{x} \sim U[-1, 1]^6.$$

$$\mathcal{M}_3 : Y = \exp\{\sin(\boldsymbol{\alpha}^\top \mathbf{x})\} + \varepsilon, \text{ where } \boldsymbol{\alpha} = (1/\sqrt{4}, 1/\sqrt{4}, 1/\sqrt{4}, 1/\sqrt{4})^\top, \\ \varepsilon \sim \mathcal{N}(0, 1), \text{ and } \mathbf{x} \sim U[-1, 1]^4.$$

### S3.1 Distributed estimation of SIM

This appendix presents additional DC simulation results for Models  $\mathcal{M}_2$  and  $\mathcal{M}_3$ , complementing the representative findings reported in the main text. The simulation setup is the same as in the main text: we consider  $N \in \{2^{11}, 2^{12}, 2^{13}, 2^{14}, 2^{15}\}$ ,  $s \in \{1, 4, 16, 64, 128\}$ , and  $n = N/s$ , under spline basis dimensions  $J_n \in \{n^{1/5}, n^{1/6}, n^{1/7}\}$ .

Figure 1 reports the average MSE of the aggregated nonparametric estimator  $\bar{f}$  under Models  $\mathcal{M}_2$  and  $\mathcal{M}_3$ . Tables 1 and 2 further report the nonparametric and index-estimation results for  $N = 16384$  and  $J_n = n^{1/7}$ . The benchmark methods first estimate the index parameter using SIR, MAVE, or IDC, respectively, and then estimate the nonparametric component using the same B-spline procedure with the estimated index fixed. For the index parameter, we report the mean squared error (MSE)

$\frac{1}{t} \sum_{j=1}^m \sum_{i=1}^t (\hat{\alpha}_{ji} - \alpha_j)^2$  and the absolute bias sum (ABS)  $\frac{1}{t} \sum_{j=1}^m \sum_{i=1}^t |\hat{\alpha}_{ji} - \alpha_j|$ , where  $m$  denotes the dimension of  $\boldsymbol{\alpha}$  and  $t$  denotes the number of Monte Carlo repetitions.

The results show that the proposed DC estimator performs comparably to the benchmark methods when  $s$  is small or moderate. Table 2 also shows that its index-estimation errors are generally comparable to those of the benchmark methods in these settings. As  $s$  increases and the local sample size decreases, IDC-Spline and MAVE-Spline can achieve smaller MSE and ABS values, particularly when the local sample size is extremely small. This improvement in index estimation is also reflected in Table 1, where the adapted benchmark methods and the two-stage refitting strategy generally yield smaller nonparametric MSEs, especially under  $\mathcal{M}_3$ . These gains require either additional communication or separate index-estimation steps, whereas the proposed estimator only requires one round of aggregation. Overall, the results illustrate the trade-off between statistical accuracy and communication cost.

Figures 2 and 3 report the empirical coverage probabilities of nominal 95% confidence regions for the aggregated parametric estimator under Models  $\mathcal{M}_2$  and  $\mathcal{M}_3$ , respectively. In both figures, the top row corresponds to marginal coverage and the bottom row to joint coverage, with columns

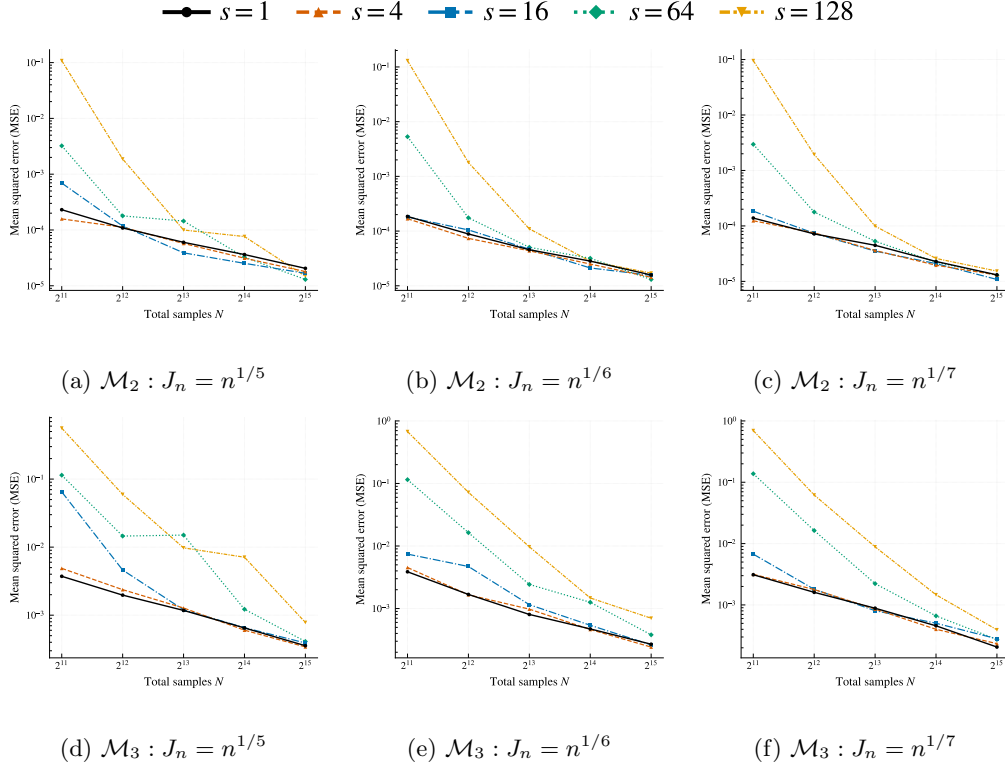


Figure 1: Average MSE of the aggregated estimator  $\bar{f}$  under Models  $\mathcal{M}_2$  (top row) and  $\mathcal{M}_3$  (bottom row). Columns correspond to different spline basis dimensions  $J_n$ , and curves correspond to different numbers of machines  $s$ .

representing different spline basis dimensions. The same qualitative pattern as in the main text is observed: when the local sample size is sufficiently large, both marginal and joint coverages remain close to the nominal level, whereas increasing  $s$  too aggressively leads to substantial undercoverage, especially for the joint confidence region. These supplementary results further support the theoretical scaling condition required for valid distributed

---

### S3. SUPPLEMENTARY SIMULATION RESULTS

---

Table 1: Nonparametric MSE comparison under Models  $\mathcal{M}_2$  and  $\mathcal{M}_3$  for  $N = 16384$  and  $J_n = n^{1/7}$ .

| Model           | s   | Proposed DC | SIR-Spline | IDC-Spline | MAVE-Spline | Two-stage refit |
|-----------------|-----|-------------|------------|------------|-------------|-----------------|
| $\mathcal{M}_2$ | 1   | 0.000023    | 0.000019   | 0.000019   | 0.000019    | 0.000019        |
|                 | 4   | 0.000020    | 0.000017   | 0.000017   | 0.000017    | 0.000017        |
|                 | 16  | 0.000021    | 0.000017   | 0.000017   | 0.000017    | 0.000017        |
|                 | 64  | 0.000022    | 0.000020   | 0.000020   | 0.000020    | 0.000020        |
|                 | 128 | 0.000026    | 0.000021   | 0.000021   | 0.000021    | 0.000021        |
| $\mathcal{M}_3$ | 1   | 0.000457    | 0.000455   | 0.000456   | 0.000456    | 0.000457        |
|                 | 4   | 0.000401    | 0.000394   | 0.000395   | 0.000394    | 0.000395        |
|                 | 16  | 0.000503    | 0.000476   | 0.000478   | 0.000477    | 0.000478        |
|                 | 64  | 0.000660    | 0.000477   | 0.000474   | 0.000474    | 0.000476        |
|                 | 128 | 0.001314    | 0.000666   | 0.000572   | 0.000572    | 0.000682        |

---

inference.

Tables 3–5 summarize the finite-sample performance of the aggregated parametric estimator  $\bar{\alpha}$  under Models  $\mathcal{M}_1$ – $\mathcal{M}_3$ . Across all three models, the same qualitative pattern is observed: for fixed  $s$ , both ABS and MSE decrease as  $N$  increases, indicating improved parametric estimation with larger total sample size. For fixed  $N$ , the estimation error remains stable when  $s$  is small or moderate, but increases substantially when  $s$  becomes too large, reflecting the deterioration caused by insufficient local sample size.

Table 2: Estimation errors of the index parameter under Models  $\mathcal{M}_1$ ,  $\mathcal{M}_2$ , and  $\mathcal{M}_3$  for  $N = 16384$  and  $J_n = n^{1/7}$ .

| Model           | s   |     | Proposed DC | SIR-Spline | IDC-Spline | MAVE-Spline |
|-----------------|-----|-----|-------------|------------|------------|-------------|
| $\mathcal{M}_1$ | 1   | ABS | 0.037806    | 0.044529   | 0.037809   | 0.042701    |
|                 |     | MSE | 0.000439    | 0.000628   | 0.000439   | 0.000582    |
|                 | 4   | ABS | 0.038455    | 0.047086   | 0.038459   | 0.044493    |
|                 |     | MSE | 0.000466    | 0.000690   | 0.000466   | 0.000602    |
|                 | 16  | ABS | 0.040028    | 0.045444   | 0.040059   | 0.044501    |
|                 |     | MSE | 0.000498    | 0.000656   | 0.000499   | 0.000628    |
|                 | 64  | ABS | 0.067533    | 0.086261   | 0.067692   | 0.076274    |
|                 |     | MSE | 0.001438    | 0.002377   | 0.001445   | 0.001825    |
|                 | 128 | ABS | 0.169194    | 0.208512   | 0.142781   | 0.147175    |
|                 |     | MSE | 0.009077    | 0.013378   | 0.006470   | 0.006923    |
| $\mathcal{M}_2$ | 1   | ABS | 0.013265    | 0.015083   | 0.013266   | 0.013775    |
|                 |     | MSE | 0.000047    | 0.000062   | 0.000047   | 0.000052    |
|                 | 4   | ABS | 0.012173    | 0.014002   | 0.012172   | 0.012525    |
|                 |     | MSE | 0.000040    | 0.000054   | 0.000040   | 0.000042    |
|                 | 16  | ABS | 0.012770    | 0.014874   | 0.012771   | 0.013000    |
|                 |     | MSE | 0.000043    | 0.000058   | 0.000043   | 0.000045    |
|                 | 64  | ABS | 0.014734    | 0.018389   | 0.014746   | 0.014980    |
|                 |     | MSE | 0.000056    | 0.000087   | 0.000057   | 0.000058    |
|                 | 128 | ABS | 0.019093    | 0.024882   | 0.019197   | 0.019760    |
|                 |     | MSE | 0.000098    | 0.000169   | 0.000100   | 0.000104    |
| $\mathcal{M}_3$ | 1   | ABS | 0.036590    | 0.041770   | 0.036590   | 0.038315    |
|                 |     | MSE | 0.000517    | 0.000663   | 0.000517   | 0.000562    |
|                 | 4   | ABS | 0.036436    | 0.040031   | 0.036449   | 0.037141    |
|                 |     | MSE | 0.000524    | 0.000635   | 0.000524   | 0.000532    |
|                 | 16  | ABS | 0.040988    | 0.045221   | 0.041014   | 0.042641    |
|                 |     | MSE | 0.000637    | 0.000786   | 0.000638   | 0.000687    |
|                 | 64  | ABS | 0.064765    | 0.076076   | 0.063120   | 0.059869    |
|                 |     | MSE | 0.001617    | 0.002185   | 0.001501   | 0.001403    |
|                 | 128 | ABS | 0.240886    | 0.233312   | 0.143380   | 0.143462    |
|                 |     | MSE | 0.019712    | 0.018258   | 0.007135   | 0.007343    |

### S3. SUPPLEMENTARY SIMULATION RESULTS

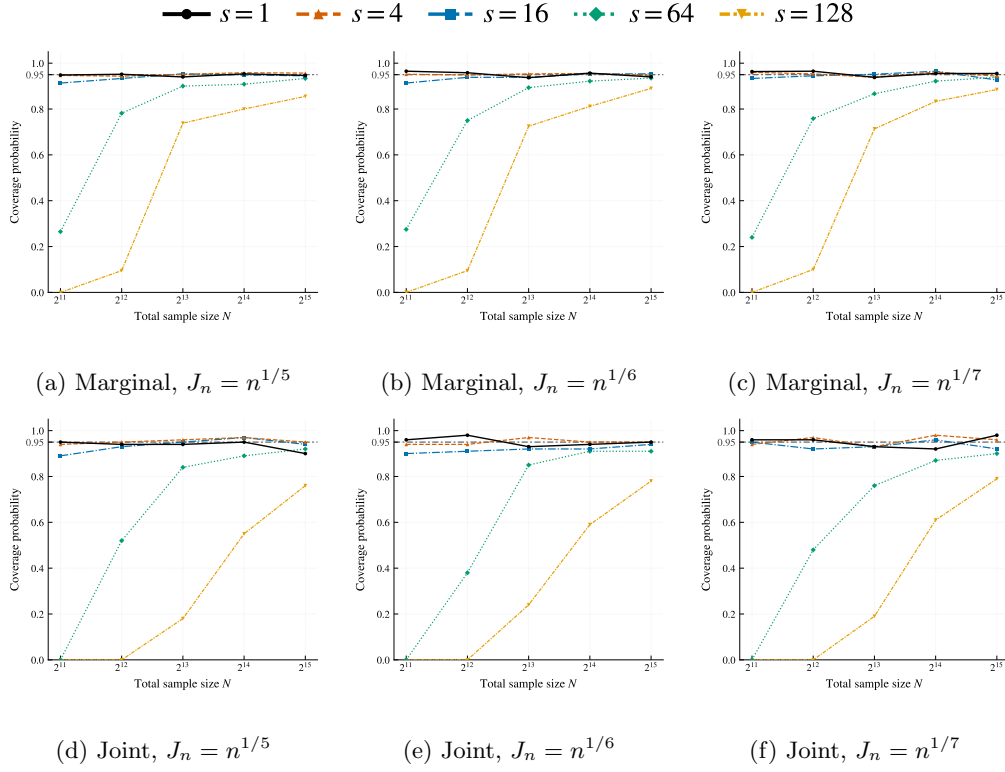


Figure 2: Empirical coverage probabilities of nominal 95% confidence regions for the aggregated parametric estimator under Model  $\mathcal{M}_2$ . The top row shows marginal coverage, and the bottom row shows joint coverage. Columns correspond to different spline basis dimensions  $J_n$ , and curves correspond to different numbers of machines  $s$ .

Overall, these results are consistent with the theoretical conclusion that the accuracy of the aggregated parametric estimator is primarily governed by the effective local sample size  $n = N/s$ .

Table 3: The ABS and MSE of parameter estimation for the model  $\mathcal{M}_1$  under different setting

| N           |           |     | 2048    | 4096    | 8192    | 16384   | 32768   |
|-------------|-----------|-----|---------|---------|---------|---------|---------|
| $J_n = 1/5$ | $s = 1$   | ABS | 0.10765 | 0.07286 | 0.05321 | 0.03874 | 0.02755 |
|             |           | MSE | 0.00359 | 0.00162 | 0.00091 | 0.00046 | 0.00024 |
|             | $s = 4$   | ABS | 0.10731 | 0.08034 | 0.05734 | 0.03740 | 0.02805 |
|             |           | MSE | 0.00354 | 0.00198 | 0.00100 | 0.00045 | 0.00025 |
|             | $s = 16$  | ABS | 0.18868 | 0.09906 | 0.06246 | 0.04078 | 0.02886 |
|             |           | MSE | 0.01124 | 0.00319 | 0.00122 | 0.00052 | 0.00026 |
|             | $s = 64$  | ABS | 0.62862 | 0.36120 | 0.15824 | 0.06948 | 0.03660 |
|             |           | MSE | 0.10453 | 0.03827 | 0.00798 | 0.00161 | 0.00042 |
|             | $s = 128$ | ABS | 0.84472 | 0.63827 | 0.35610 | 0.16133 | 0.06737 |
|             |           | MSE | 0.17547 | 0.10567 | 0.03610 | 0.00812 | 0.00147 |
| $J_n = 1/6$ | $s = 1$   | ABS | 0.10496 | 0.08060 | 0.05028 | 0.03914 | 0.02689 |
|             |           | MSE | 0.00355 | 0.00206 | 0.00081 | 0.00048 | 0.00022 |
|             | $s = 4$   | ABS | 0.11089 | 0.07584 | 0.05559 | 0.03801 | 0.02635 |
|             |           | MSE | 0.00398 | 0.00179 | 0.00098 | 0.00046 | 0.00022 |
|             | $s = 16$  | ABS | 0.18506 | 0.09946 | 0.05922 | 0.04058 | 0.02685 |
|             |           | MSE | 0.01081 | 0.00311 | 0.00111 | 0.00052 | 0.00023 |
|             | $s = 64$  | ABS | 0.63237 | 0.36081 | 0.14662 | 0.07016 | 0.03933 |
|             |           | MSE | 0.10577 | 0.03773 | 0.00693 | 0.00162 | 0.00048 |
|             | $s = 128$ | ABS | 0.84443 | 0.62519 | 0.35910 | 0.14420 | 0.06547 |
|             |           | MSE | 0.17522 | 0.10173 | 0.03665 | 0.00657 | 0.00141 |
| $J_n = 1/7$ | $s = 1$   | ABS | 0.11333 | 0.07051 | 0.05539 | 0.03781 | 0.02717 |
|             |           | MSE | 0.00399 | 0.00155 | 0.00092 | 0.00044 | 0.00023 |
|             | $s = 4$   | ABS | 0.11001 | 0.08151 | 0.05011 | 0.03845 | 0.02672 |
|             |           | MSE | 0.00382 | 0.00208 | 0.00080 | 0.00047 | 0.00022 |
|             | $s = 16$  | ABS | 0.17943 | 0.09032 | 0.06286 | 0.04005 | 0.02662 |
|             |           | MSE | 0.00996 | 0.00262 | 0.00129 | 0.00050 | 0.00022 |
|             | $s = 64$  | ABS | 0.62754 | 0.35401 | 0.14959 | 0.06802 | 0.03880 |
|             |           | MSE | 0.10384 | 0.03672 | 0.00713 | 0.00146 | 0.00046 |
|             | $s = 128$ | ABS | 0.85428 | 0.62973 | 0.36013 | 0.14806 | 0.06148 |
|             |           | MSE | 0.17899 | 0.10297 | 0.03676 | 0.00692 | 0.00124 |

---

### S3. SUPPLEMENTARY SIMULATION RESULTS

---

Table 4: The ABS and MSE of parameter estimation for the model  $\mathcal{M}_2$  under different settings

| N           |           |     | 2048    | 4096    | 8192    | 16384   | 32768   |
|-------------|-----------|-----|---------|---------|---------|---------|---------|
| $J_n = 1/5$ | $s = 1$   | ABS | 0.03687 | 0.02688 | 0.01882 | 0.01339 | 0.00960 |
|             |           | MSE | 0.00036 | 0.00019 | 0.00010 | 0.00005 | 0.00003 |
|             | $s = 4$   | ABS | 0.03771 | 0.02777 | 0.01924 | 0.01258 | 0.00937 |
|             |           | MSE | 0.00038 | 0.00020 | 0.00009 | 0.00004 | 0.00002 |
|             | $s = 16$  | ABS | 0.04174 | 0.02744 | 0.01876 | 0.01330 | 0.00934 |
|             |           | MSE | 0.00046 | 0.00020 | 0.00009 | 0.00005 | 0.00002 |
|             | $s = 64$  | ABS | 0.09535 | 0.04033 | 0.02239 | 0.01500 | 0.00963 |
|             |           | MSE | 0.00273 | 0.00044 | 0.00013 | 0.00006 | 0.00002 |
|             | $s = 128$ | ABS | 0.47472 | 0.08977 | 0.03408 | 0.02036 | 0.01241 |
|             |           | MSE | 0.05736 | 0.00249 | 0.00033 | 0.00011 | 0.00004 |
| $J_n = 1/6$ | $s = 1$   | ABS | 0.03509 | 0.02526 | 0.02007 | 0.01285 | 0.01005 |
|             |           | MSE | 0.00032 | 0.00017 | 0.00011 | 0.00004 | 0.00003 |
|             | $s = 4$   | ABS | 0.03782 | 0.02719 | 0.01845 | 0.01286 | 0.00937 |
|             |           | MSE | 0.00037 | 0.00020 | 0.00009 | 0.00004 | 0.00002 |
|             | $s = 16$  | ABS | 0.04222 | 0.02883 | 0.01933 | 0.01301 | 0.00937 |
|             |           | MSE | 0.00046 | 0.00021 | 0.00010 | 0.00005 | 0.00002 |
|             | $s = 64$  | ABS | 0.09818 | 0.04425 | 0.02299 | 0.01451 | 0.01037 |
|             |           | MSE | 0.00285 | 0.00053 | 0.00014 | 0.00006 | 0.00003 |
|             | $s = 128$ | ABS | 0.48577 | 0.08788 | 0.03323 | 0.01934 | 0.01139 |
|             |           | MSE | 0.05980 | 0.00236 | 0.00031 | 0.00010 | 0.00003 |
| $J_n = 1/7$ | $s = 1$   | ABS | 0.03547 | 0.02618 | 0.01952 | 0.01327 | 0.00910 |
|             |           | MSE | 0.00033 | 0.00018 | 0.00010 | 0.00005 | 0.00002 |
|             | $s = 4$   | ABS | 0.03645 | 0.02583 | 0.01990 | 0.01217 | 0.00955 |
|             |           | MSE | 0.00035 | 0.00018 | 0.00011 | 0.00004 | 0.00002 |
|             | $s = 16$  | ABS | 0.04079 | 0.02711 | 0.01879 | 0.01277 | 0.00997 |
|             |           | MSE | 0.00042 | 0.00020 | 0.00009 | 0.00004 | 0.00003 |
|             | $s = 64$  | ABS | 0.09553 | 0.04082 | 0.02318 | 0.01473 | 0.00995 |
|             |           | MSE | 0.00269 | 0.00045 | 0.00015 | 0.00006 | 0.00003 |
|             | $s = 128$ | ABS | 0.47949 | 0.08947 | 0.03603 | 0.01909 | 0.01119 |
|             |           | MSE | 0.05829 | 0.00245 | 0.00036 | 0.00010 | 0.00003 |

Table 5: The ABS and MSE of parameter estimation for the model  $\mathcal{M}_3$  under different settings

| N           |           |     | 2048    | 4096    | 8192    | 16384   | 32768   |
|-------------|-----------|-----|---------|---------|---------|---------|---------|
| $J_n = 1/5$ | $s = 1$   | ABS | 0.10576 | 0.07778 | 0.05240 | 0.03732 | 0.02544 |
|             |           | MSE | 0.00442 | 0.00230 | 0.00112 | 0.00058 | 0.00024 |
|             | $s = 4$   | ABS | 0.10932 | 0.07370 | 0.04828 | 0.03488 | 0.02541 |
|             |           | MSE | 0.00471 | 0.00210 | 0.00094 | 0.00047 | 0.00024 |
|             | $s = 16$  | ABS | 0.21041 | 0.10372 | 0.05961 | 0.03828 | 0.02683 |
|             |           | MSE | 0.01680 | 0.00423 | 0.00136 | 0.00059 | 0.00029 |
|             | $s = 64$  | ABS | 0.62078 | 0.37538 | 0.19179 | 0.07311 | 0.03453 |
|             |           | MSE | 0.11319 | 0.04506 | 0.01286 | 0.00201 | 0.00047 |
|             | $s = 128$ | ABS | 0.80747 | 0.63012 | 0.37068 | 0.18349 | 0.06932 |
|             |           | MSE | 0.18150 | 0.11447 | 0.04272 | 0.01139 | 0.00176 |
| $J_n = 1/6$ | $s = 1$   | ABS | 0.09242 | 0.07440 | 0.05592 | 0.03655 | 0.02305 |
|             |           | MSE | 0.00352 | 0.00221 | 0.00121 | 0.00054 | 0.00021 |
|             | $s = 4$   | ABS | 0.10905 | 0.07629 | 0.05220 | 0.03754 | 0.02366 |
|             |           | MSE | 0.00459 | 0.00220 | 0.00105 | 0.00053 | 0.00022 |
|             | $s = 16$  | ABS | 0.18496 | 0.10219 | 0.05678 | 0.04025 | 0.02638 |
|             |           | MSE | 0.01320 | 0.00415 | 0.00125 | 0.00061 | 0.00027 |
|             | $s = 64$  | ABS | 0.61758 | 0.37832 | 0.16298 | 0.07252 | 0.03663 |
|             |           | MSE | 0.11211 | 0.04598 | 0.00941 | 0.00198 | 0.00051 |
|             | $s = 128$ | ABS | 0.81503 | 0.62022 | 0.36868 | 0.15432 | 0.06906 |
|             |           | MSE | 0.18465 | 0.11116 | 0.04231 | 0.00819 | 0.00181 |
| $J_n = 1/7$ | $s = 1$   | ABS | 0.10644 | 0.07675 | 0.05452 | 0.03659 | 0.02679 |
|             |           | MSE | 0.00434 | 0.00236 | 0.00113 | 0.00052 | 0.00029 |
|             | $s = 4$   | ABS | 0.09776 | 0.06732 | 0.05686 | 0.03645 | 0.02371 |
|             |           | MSE | 0.00385 | 0.00180 | 0.00118 | 0.00052 | 0.00023 |
|             | $s = 16$  | ABS | 0.17330 | 0.09377 | 0.05714 | 0.04099 | 0.02486 |
|             |           | MSE | 0.01156 | 0.00344 | 0.00120 | 0.00064 | 0.00024 |
|             | $s = 64$  | ABS | 0.61532 | 0.36316 | 0.16221 | 0.06341 | 0.03634 |
|             |           | MSE | 0.11127 | 0.04264 | 0.00947 | 0.00151 | 0.00051 |
|             | $s = 128$ | ABS | 0.80955 | 0.63492 | 0.36366 | 0.15551 | 0.06090 |
|             |           | MSE | 0.18247 | 0.11605 | 0.04097 | 0.00832 | 0.00135 |

### S3. SUPPLEMENTARY SIMULATION RESULTS

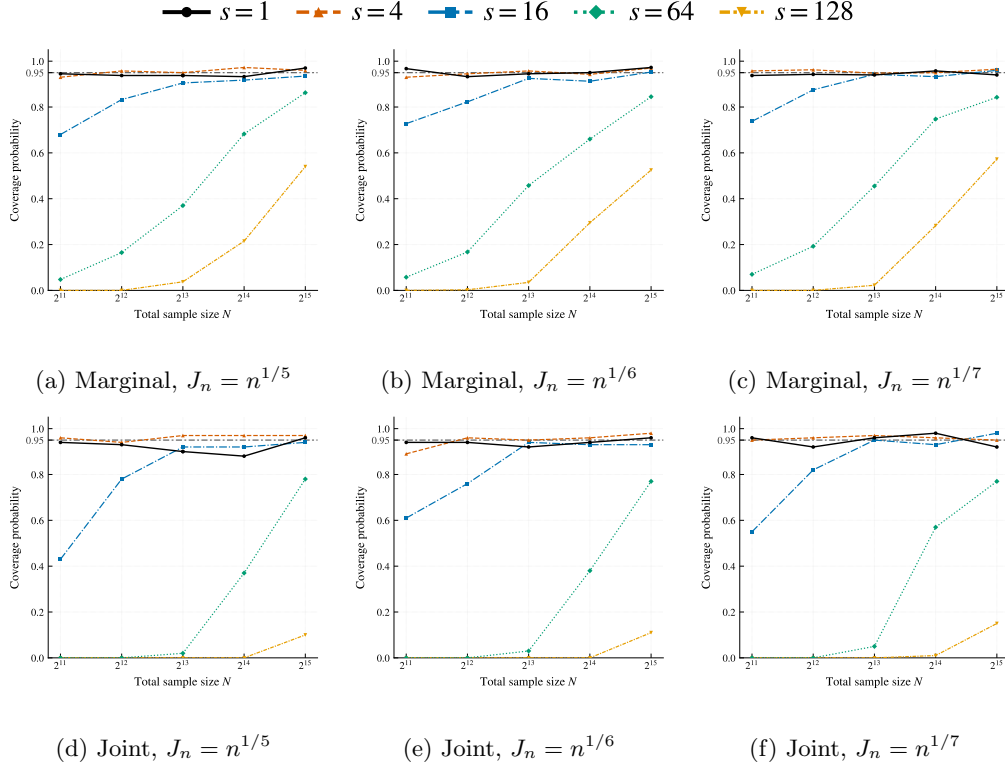


Figure 3: Empirical coverage probabilities of nominal 95% confidence regions for the aggregated parametric estimator under Model  $\mathcal{M}_3$ . The top row shows marginal coverage, and the bottom row shows joint coverage. Columns correspond to different spline basis dimensions  $J_n$ , and curves correspond to different numbers of machines  $s$ .

### S3.2 Distributed estimation with differential privacy

In addition to the single-machine differentially private results reported in the main text, we provide the corresponding supplementary results for Models  $\mathcal{M}_2$  and  $\mathcal{M}_3$ . The simulation setup is the same as in the main text: we consider  $n \in [5000, 50000]$ , use cubic splines with  $\rho = 3$  and equally spaced

knots, set  $J_n = n^{1/7}$ , and take  $(\varepsilon, \delta) \in \{(0.2, 1/n), (0.5, 1/n), (0.8, 1/n)\}$ , with 100 replications for each configuration. Figure 4 presents the corresponding estimation error curves, and Table 6 reports the average absolute bias (ABS) and mean squared error (MSE) of  $\hat{\alpha}$ . Across both models, the same qualitative pattern as in the main text is observed: for fixed privacy level, the estimation error decreases as  $n$  increases, while for fixed  $n$ , the error decreases as  $\varepsilon$  increases, reflecting the standard privacy–utility trade-off. Compared with the non-private benchmark, differential privacy leads to a noticeable loss in estimation accuracy, especially when  $n$  is small or the privacy level is strong. Overall, these supplementary results confirm that the private estimator becomes increasingly stable as the sample size grows and that the main conclusions are robust across different regression functions.

In addition to the private DC results reported in the main text, we provide the corresponding supplementary results for Models  $\mathcal{M}_2$  and  $\mathcal{M}_3$ . The simulation setup is the same as in the main text: we consider total sample sizes  $N \in [5000, 50000]$  and numbers of machines  $s \in \{1, 5, 25, 50\}$ . Under parallel composition, each machine uses the same privacy budget  $(\varepsilon, \delta) = (0.5, 1/N)$ , and each configuration is repeated 100 times. Figure 5 presents the corresponding estimation error curves, and Table 7 reports the

---

### S3. SUPPLEMENTARY SIMULATION RESULTS

---

Table 6: ABS and MSE of the single-machine estimator  $\alpha$  under Models  $\mathcal{M}_2$  and  $\mathcal{M}_3$  for privacy levels  $(0.2, 1/n)$ ,  $(0.5, 1/n)$ , and  $(0.8, 1/n)$ .

| $n$             |                     | 5000 | 10000   | 15000   | 20000   | 25000   | 30000   | 35000   | 40000   | 45000   | 50000   |
|-----------------|---------------------|------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| $\mathcal{M}_2$ | non                 | ABS  | 0.02626 | 0.01769 | 0.01416 | 0.01204 | 0.01089 | 0.01015 | 0.00925 | 0.00864 | 0.00776 |
|                 | private             | MSE  | 0.00018 | 0.00008 | 0.00005 | 0.00004 | 0.00003 | 0.00003 | 0.00002 | 0.00002 | 0.00002 |
|                 | $\varepsilon = 0.2$ | ABS  | 0.81164 | 0.86970 | 0.85037 | 0.77963 | 0.76738 | 0.74238 | 0.67485 | 0.65619 | 0.53290 |
|                 |                     | MSE  | 0.16775 | 0.19437 | 0.19114 | 0.15965 | 0.15120 | 0.14865 | 0.12643 | 0.11762 | 0.09997 |
|                 | $\varepsilon = 0.5$ | ABS  | 0.81199 | 0.63318 | 0.54048 | 0.52188 | 0.44025 | 0.34646 | 0.36930 | 0.34664 | 0.29861 |
|                 |                     | MSE  | 0.16807 | 0.10841 | 0.07936 | 0.07751 | 0.05435 | 0.03333 | 0.04054 | 0.03581 | 0.02663 |
|                 | $\varepsilon = 0.8$ | ABS  | 0.70435 | 0.50018 | 0.41081 | 0.35750 | 0.32334 | 0.29783 | 0.28106 | 0.26149 | 0.24914 |
|                 |                     | MSE  | 0.12739 | 0.06690 | 0.05111 | 0.03976 | 0.03268 | 0.02803 | 0.02553 | 0.02174 | 0.01966 |
|                 | non                 | ABS  | 0.06736 | 0.04642 | 0.03746 | 0.03161 | 0.02709 | 0.02377 | 0.02246 | 0.02138 | 0.01988 |
|                 | private             | MSE  | 0.00172 | 0.00079 | 0.00056 | 0.00039 | 0.00029 | 0.00022 | 0.00020 | 0.00018 | 0.00015 |
|                 | $\varepsilon = 0.2$ | ABS  | 0.87140 | 0.94878 | 0.85642 | 0.83030 | 0.78977 | 0.74331 | 0.72155 | 0.61485 | 0.54908 |
|                 |                     | MSE  | 0.24950 | 0.32727 | 0.26557 | 0.24354 | 0.22653 | 0.20153 | 0.19043 | 0.15537 | 0.12327 |
| $\mathcal{M}_3$ | $\varepsilon = 0.5$ | ABS  | 0.82336 | 0.68149 | 0.54001 | 0.55564 | 0.46956 | 0.31701 | 0.35000 | 0.27132 | 0.27760 |
|                 |                     | MSE  | 0.23069 | 0.17285 | 0.11799 | 0.12262 | 0.09417 | 0.05012 | 0.05623 | 0.03760 | 0.03853 |
|                 | $\varepsilon = 0.8$ | ABS  | 0.79188 | 0.49360 | 0.37085 | 0.35294 | 0.31604 | 0.27718 | 0.24773 | 0.19056 | 0.19581 |
|                 |                     | MSE  | 0.20732 | 0.10123 | 0.06626 | 0.06008 | 0.04766 | 0.04003 | 0.03568 | 0.01949 | 0.01848 |

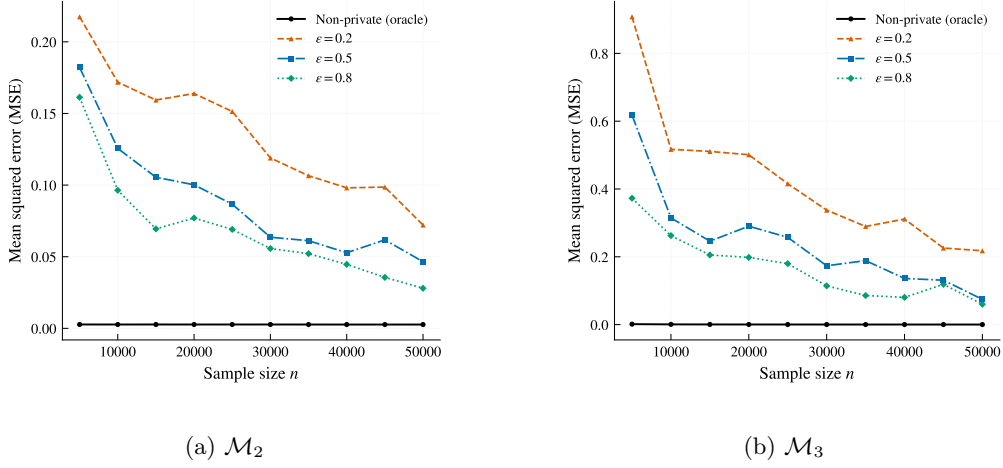


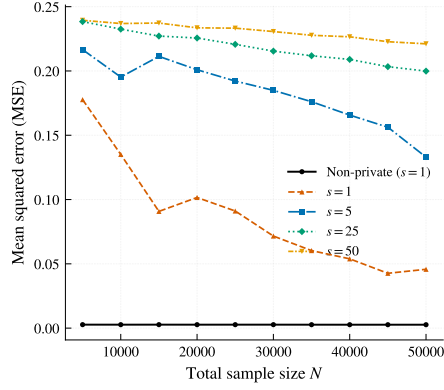
Figure 4: Estimation error of the single-machine differentially private estimator under Models  $\mathcal{M}_2$  and  $\mathcal{M}_3$  for different privacy levels.

average absolute bias (ABS) and mean squared error (MSE) of the aggregated estimator  $\bar{\alpha}_T$ . Across both models, the supplementary results exhibit patterns consistent with those observed in the main text. For fixed  $s$ , the estimation error decreases as  $N$  increases, while more distributed settings generally lead to larger estimation errors. Compared with the non-private benchmark, the private estimators exhibit larger ABS and MSE, indicating the accuracy loss introduced by differential privacy. This degradation becomes more evident when the number of machines is large. Overall, these results further confirm the impact of privacy protection on the accuracy of distributed estimation.

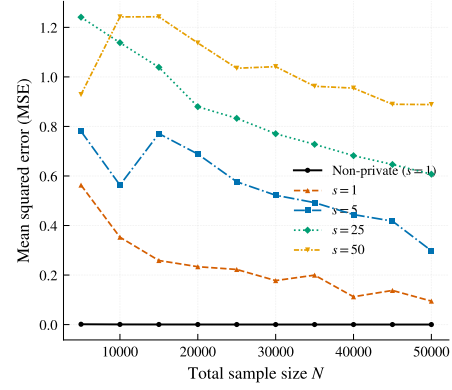
### S3. SUPPLEMENTARY SIMULATION RESULTS

Table 7: ABS and MSE of the aggregated estimator  $\bar{\alpha}_T$  under Models  $\mathcal{M}_2$  and  $\mathcal{M}_3$  in the private DC setting.

| N               |                     |          | 5000    | 10000   | 15000   | 20000   | 25000   | 30000   | 35000   | 40000   | 45000   | 50000   |         |         |
|-----------------|---------------------|----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| $\mathcal{M}_2$ | private             | $s = 1$  | ABS     | 0.02626 | 0.01769 | 0.01416 | 0.01204 | 0.01089 | 0.01015 | 0.00925 | 0.00864 | 0.00807 | 0.00776 |         |
|                 |                     |          | MSE     | 0.00018 | 0.00008 | 0.00005 | 0.00004 | 0.00003 | 0.00003 | 0.00002 | 0.00002 | 0.00002 | 0.00002 |         |
|                 |                     | $s = 5$  | ABS     | 0.02644 | 0.01774 | 0.01420 | 0.01205 | 0.01091 | 0.01014 | 0.00926 | 0.00867 | 0.00808 | 0.00778 |         |
|                 |                     |          | MSE     | 0.00018 | 0.00008 | 0.00005 | 0.00004 | 0.00003 | 0.00003 | 0.00002 | 0.00002 | 0.00002 | 0.00002 |         |
|                 |                     | $s = 25$ | ABS     | 0.02817 | 0.01834 | 0.01462 | 0.01231 | 0.01109 | 0.01029 | 0.00933 | 0.00876 | 0.00814 | 0.00781 |         |
|                 |                     |          | MSE     | 0.00020 | 0.00009 | 0.00005 | 0.00004 | 0.00003 | 0.00003 | 0.00002 | 0.00002 | 0.00002 | 0.00002 |         |
|                 | $s = 50$            | ABS      | 0.03280 | 0.01985 | 0.01544 | 0.01290 | 0.01163 | 0.01060 | 0.00958 | 0.00894 | 0.00830 | 0.00796 |         |         |
|                 |                     | MSE      | 0.00028 | 0.00010 | 0.00006 | 0.00004 | 0.00003 | 0.00003 | 0.00002 | 0.00002 | 0.00002 | 0.00002 |         |         |
|                 | $\varepsilon = 0.5$ | $s = 1$  | ABS     | 0.73465 | 0.64248 | 0.51191 | 0.49152 | 0.42635 | 0.42130 | 0.35135 | 0.32973 | 0.32350 | 0.32045 |         |
|                 |                     |          | MSE     | 0.13750 | 0.11062 | 0.07250 | 0.06942 | 0.05272 | 0.05115 | 0.03786 | 0.03178 | 0.03126 | 0.02946 |         |
|                 |                     | $s = 5$  | ABS     | 0.80663 | 0.76155 | 0.74437 | 0.70995 | 0.68726 | 0.67301 | 0.65137 | 0.59876 | 0.56606 | 0.54064 |         |
|                 |                     |          | MSE     | 0.16107 | 0.14624 | 0.13795 | 0.12647 | 0.11971 | 0.11594 | 0.10919 | 0.09513 | 0.08587 | 0.07857 |         |
|                 |                     | $s = 25$ | ABS     | 0.91892 | 0.83299 | 0.82267 | 0.78103 | 0.78450 | 0.78111 | 0.78093 | 0.76934 | 0.77023 | 0.75602 |         |
|                 |                     |          | MSE     | 0.19587 | 0.16490 | 0.16061 | 0.14666 | 0.14691 | 0.14601 | 0.14653 | 0.14197 | 0.14237 | 0.13791 |         |
|                 |                     | $s = 50$ | ABS     | 1.72488 | 0.90942 | 0.86997 | 0.84551 | 0.79412 | 0.80188 | 0.79755 | 0.80227 | 0.80285 | 0.80111 |         |
|                 |                     |          | MSE     | 0.57211 | 0.18965 | 0.17436 | 0.16598 | 0.14876 | 0.15138 | 0.15020 | 0.15141 | 0.15186 | 0.15111 |         |
|                 |                     | non      | $s = 1$ | ABS     | 0.06736 | 0.04642 | 0.03746 | 0.03161 | 0.02709 | 0.02377 | 0.02246 | 0.02138 | 0.02086 | 0.01988 |
|                 |                     |          |         | MSE     | 0.00172 | 0.00079 | 0.00056 | 0.00039 | 0.00029 | 0.00022 | 0.00020 | 0.00018 | 0.00017 | 0.00015 |
|                 |                     |          | $s = 5$ | ABS     | 0.06746 | 0.04636 | 0.03786 | 0.03162 | 0.02743 | 0.02396 | 0.02262 | 0.02150 | 0.02095 | 0.02008 |
|                 |                     |          |         | MSE     | 0.00173 | 0.00080 | 0.00057 | 0.00039 | 0.00029 | 0.00022 | 0.00020 | 0.00018 | 0.00017 | 0.00015 |
|                 | $s = 25$            |          | ABS     | 0.12996 | 0.06875 | 0.04943 | 0.03813 | 0.03200 | 0.02788 | 0.02539 | 0.02354 | 0.02259 | 0.02105 |         |
|                 |                     |          | MSE     | 0.00622 | 0.00175 | 0.00091 | 0.00057 | 0.00040 | 0.00029 | 0.00026 | 0.00022 | 0.00020 | 0.00017 |         |
|                 | $s = 50$            |          | ABS     | 0.23963 | 0.12384 | 0.08202 | 0.05912 | 0.04817 | 0.04017 | 0.03552 | 0.03125 | 0.02900 | 0.02628 |         |
|                 |                     |          | MSE     | 0.01943 | 0.00549 | 0.00243 | 0.00130 | 0.00087 | 0.00061 | 0.00048 | 0.00038 | 0.00033 | 0.00027 |         |
| $\mathcal{M}_3$ | $s = 1$             |          | ABS     | 0.78131 | 0.65242 | 0.58732 | 0.48255 | 0.50125 | 0.37858 | 0.37160 | 0.32381 | 0.33648 | 0.25685 |         |
|                 |                     |          | MSE     | 0.20160 | 0.15201 | 0.13051 | 0.09586 | 0.10459 | 0.06707 | 0.06494 | 0.05429 | 0.05476 | 0.03315 |         |
|                 | $s = 5$             |          | ABS     | 0.88394 | 0.82927 | 0.80991 | 0.79940 | 0.80119 | 0.79769 | 0.78560 | 0.78209 | 0.76756 | 0.68070 |         |
|                 |                     |          | MSE     | 0.23708 | 0.20872 | 0.19100 | 0.18522 | 0.18663 | 0.18426 | 0.17982 | 0.17866 | 0.17138 | 0.14586 |         |
|                 | $s = 25$            | ABS      | 0.95147 | 0.88516 | 0.85413 | 0.89411 | 0.87023 | 0.86519 | 0.85335 | 0.85705 | 0.83950 | 0.85100 |         |         |
|                 |                     | MSE      | 0.25334 | 0.21905 | 0.20397 | 0.22355 | 0.21280 | 0.20886 | 0.20372 | 0.20486 | 0.19697 | 0.20187 |         |         |
|                 | $s = 50$            | ABS      | 1.36242 | 0.97817 | 0.86273 | 0.89551 | 0.90935 | 0.86847 | 0.88419 | 0.86178 | 0.87887 | 0.85240 |         |         |
|                 |                     | MSE      | 0.48784 | 0.26199 | 0.20651 | 0.22101 | 0.22767 | 0.20830 | 0.21571 | 0.20502 | 0.21285 | 0.20121 |         |         |



(a)  $\mathcal{M}_2$



(b)  $\mathcal{M}_3$

Figure 5: Estimation error of the private aggregated estimator in the DC setting under Models  $\mathcal{M}_2$  and  $\mathcal{M}_3$ .

## References

- Abadi, M., A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308–318.
- Carroll, R. J., J. Fan, I. Gijbels, and M. P. Wand (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association* 92(438), 477–489.
- De Boor, C. (1978). *A practical guide to splines*, Volume 27. Springer New York.
- Dwork, C., F. McSherry, K. Nissim, and A. Smith (2006). Calibrating noise to sensitivity in private data analysis. In *In TCC 2006*, pp. 265–284. Springer.
- Mironov, I. (2017). Rényi differential privacy. In *2017 IEEE 30th computer security foundations*

## REFERENCES

---

- symposium (CSF)*, pp. 263–275. IEEE.
- Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7(2), 186–199.
- Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *The Annals of Statistics* 14(2), 590–606.
- Wang, L. and G. Cao (2018). Efficient estimation for generalized partially linear single-index models. *Bernoulli* 24(2), 1101–1127.
- Zhang, J. and H. Lian (2018). Partially linear additive models with unknown link functions. *Scandinavian Journal of Statistics* 45(2), 255–282.