SEMIPARAMETRIC ANALYSIS FOR PAIRED COMPARISONS WITH COVARIATES

Haoyue Song[1], Lianqiang Qu[1], Ting Yan[1], Yuguo Chen[2]

[1] *Central China Normal University*

[2] *University of Illinois Urbana-Champaign*

Supplementary Material

This Supplementary Material contains additional simulation results, and the proofs for Lemma 1, Corollary 1, sufficient conditions for Condition 4, and Theorems 1–3. The proofs for Lemma 1 and Corollary 1 are given in Sections S1. The proof for sufficient conditions for Condition 4 in Section S2. The proofs for Theorems 1–3 are given in Sections S3–S5, respectively. Section S6 contains additional simulation results under Cases (i), (ii) and (iii) in Section 4.1. Section S7 presents comparison with the covariate-adjusted Bradley–Terry model in Yan (2025). Section S8 presents simulation results for evaluating robust of the choices of bandwidths and kernel functions, and reports simulation results under sparse paired comparisons. Section S9 presents simulation results for a comparison with the win counting method. Section S10 shows Table 2 in the main text with 95% confidence intervals. Section S11 reproduces Hoeffding's inequality in Hoeffding (1963) and the error bounds for nonparametric density estimators in Andrews (1995).

S1 Proofs of Lemma 1 and Corollary 1

S1.1 Proof of Lemma 1

Proof. For convenience, let

$$\widetilde{Y}_{ijt} = \theta_i^* - \theta_j^* + \boldsymbol{Z}_{ijt}^\top \boldsymbol{\eta}^* - \varepsilon_{ijt}.$$

According to the definition of a_{ijt} in (2.1), we have

$$a_{ijt} = \mathbb{I}(\widetilde{Y}_{ijt} + \gamma_0^* X_{ijt,0} > 0).$$

It follows from Conditions 1 and 2 that

$$\begin{aligned}
\mathbb{E}(Y_{ijt}|\boldsymbol{Z}_{ijt}, \varepsilon_{ijt}) &= \int_{-B_U}^{B_U} \frac{\mathbb{I}(\widetilde{Y}_{ijt} + \gamma_0^* x > 0) - \mathbb{I}(\gamma_0^* x > 0)}{f(x|\boldsymbol{Z}_{ijt})} f(x|\boldsymbol{Z}_{ijt}) dx \\
&= \int_{-B_U}^{B_U} \mathbb{I}(\widetilde{Y}_{ijt} + \gamma_0^* x > 0) - \mathbb{I}(\gamma_0^* x > 0) dx,
\end{aligned}$$

where $B_U > 0$ by Condition 2. To compute the above integration, there are two cases for γ_0^*: (Case I) $\gamma_0^* > 0$ and (Case II) $\gamma_0^* < 0$.

Consider Case I first. Observe that for any $y > 0$, $[\mathbb{I}(y + \gamma_0^* x > 0) - \mathbb{I}(\gamma_0^* x > 0)]$ is equal to 1 if $x \in (-y/\gamma_0^*, 0]$ and 0 otherwise. Therefore, if $\widetilde{Y}_{ijt} > 0$ and $\gamma_0^* > 0$, then

$$\mathbb{E}(Y_{ijt}|\boldsymbol{Z}_{ijt}, \varepsilon_{ijt}) = \int_{-\widetilde{Y}_{ijt}/\gamma_0^*}^{0} dx = \widetilde{Y}_{ijt}/\gamma_0^*. \tag{S1.1}$$

Similarly, for any $y \leq 0$, we have $[\mathbb{I}(y + \gamma_0^* x > 0) - \mathbb{I}(\gamma_0^* x > 0)]$ is equal to -1 if $x \in (0, -y/\gamma_0^*]$ and 0 otherwise. Therefore, in the case of $\widetilde{Y}_{ijt} \leq 0$ and $\gamma_0^* > 0$, we have

$$\mathbb{E}(Y_{ijt}|\boldsymbol{Z}_{ijt}, \varepsilon_{ijt}) = -\int_{0}^{-\widetilde{Y}_{ijt}/\gamma_0^*} dx = \widetilde{Y}_{ijt}/\gamma_0^*. \tag{S1.2}$$

In view of (S1.1) and (S1.2), by Condition 3, we have

$$\mathbb{E}(Y_{ijt}|\boldsymbol{Z}_{ijt}) = \mathbb{E}\big\{\mathbb{E}(Y_{ijt}|\boldsymbol{Z}_{ijt}, \epsilon_{ijt})|\boldsymbol{Z}_{ijt}\big\} = (\theta_i^* - \theta_j^* + \boldsymbol{Z}_{ijt}^\top \boldsymbol{\eta}^*)/\gamma_0^*.$$

Consider Case II. Observe that for any $y > 0$, $[\mathbb{I}(y + \gamma_0^* x > 0) - \mathbb{I}(\gamma_0^* x >$

0)] is equal to 1 if $x \in [0, -y/\gamma_0^*)$ and 0 otherwise. Therefore, if $\widetilde{Y}_{ijt} > 0$ and $\gamma_0^* < 0$, then

$$\mathbb{E}(Y_{ijt}|\boldsymbol{Z}_{ijt}, \varepsilon_{ijt}) = \int_0^{-\widetilde{Y}_{ijt}/\gamma_0^*} dx = -\widetilde{Y}_{ijt}/\gamma_0^*. \tag{S1.3}$$

Similarly, for any $y \leq 0$, we have $[\mathbb{I}(y + \gamma_0^* x > 0) - \mathbb{I}(\gamma_0^* x > 0)]$ is equal to -1 if $x \in [-y/\gamma_0^*, 0)$ and 0 otherwise. Therefore, in the case of $\widetilde{Y}_{ijt} \leq 0$ and $\gamma_0^* < 0$, we have

$$\mathbb{E}(Y_{ijt}|\boldsymbol{Z}_{ijt}, \varepsilon_{ijt}) = -\int_{-\widetilde{Y}_{ijt}/\gamma_0^*}^{0} dx = \widetilde{Y}_{ijt}/\gamma_0^*. \tag{S1.4}$$

In view of (S1.3) and (S1.4), by Condition 3, we have

$$\mathbb{E}(Y_{ijt}|\boldsymbol{Z}_{ijt}) = \mathbb{E}\big\{\mathbb{E}(Y_{ijt}|\boldsymbol{Z}_{ijt}, \varepsilon_{ijt})|\boldsymbol{Z}_{ijt}\big\} = (\theta_i^* - \theta_j^* + \boldsymbol{Z}_{ijt}^\top \boldsymbol{\eta}^*)/\gamma_0^*.$$

This completes the proof. □

S1.2 Proof of Corollary 1

Proof. Recall that $\boldsymbol{e}_i$ denotes the standard base in $\mathbb{R}^n$ with the ith entry 1 and others 0, and in (2.2), (2.3), (2.4) and (2.6), we define

$$\boldsymbol{U}_{0j} = -\boldsymbol{e}_j \in \mathbb{R}^n, \quad j = 1, \dots, n, \qquad \boldsymbol{U}_{ij} = \boldsymbol{e}_i - \boldsymbol{e}_j \in \mathbb{R}^n, \quad 1 \leq i < j \leq n,$$

$$\boldsymbol{U} = (\boldsymbol{U}_{01}, \dots, \boldsymbol{U}_{0n}, \boldsymbol{U}_{12}, \dots, \boldsymbol{U}_{1n}, \dots, \boldsymbol{U}_{(n-1)n})^\top \in \mathbb{R}^{N\times n},$$

$$\boldsymbol{V} = \boldsymbol{U}^\top \boldsymbol{U} \in \mathbb{R}^{n\times n}, \qquad \boldsymbol{D} = \boldsymbol{I}_{N\times N} - \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top \in \mathbb{R}^{N\times N},$$

$$\overline{\boldsymbol{Z}} = \frac{1}{T}\sum_{t\in[T]} \boldsymbol{Z}_t \in \mathbb{R}^{N\times p},$$

$$\boldsymbol{Z}_t = (\boldsymbol{Z}_{01t}, \boldsymbol{Z}_{02t}, \dots, \boldsymbol{Z}_{0nt}, \boldsymbol{Z}_{12t}, \dots, \boldsymbol{Z}_{1nt}, \dots, \boldsymbol{Z}_{(n-1)nt})^\top,$$

where $\boldsymbol{Z}_{ijt} = (X_{ijt,1}, \ldots, X_{ijt,p})^\top \in \mathbb{R}^p$. Under the restricted conditions $\theta_0^* = 0$ and $\gamma_0^* = 1$, by Lemma (1), we have

$$\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{Z}) = \overline{\boldsymbol{Z}}^\top \boldsymbol{D}(\boldsymbol{U}\boldsymbol{\theta}^* + \overline{\boldsymbol{Z}}\boldsymbol{\eta}^*) = \overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}}\boldsymbol{\eta}^*, \tag{S1.5}$$

where the last equality is due to

$$\boldsymbol{D}\boldsymbol{U} = (\boldsymbol{I}_{N\times N} - \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top)\boldsymbol{U} = \boldsymbol{0}, \tag{S1.6}$$

which follows from the definition of $\boldsymbol{D}$ in (2.6). Taking the expectations of both sides of (S1.5), by Condition 4, we have

$$\boldsymbol{\eta}^* = \mathbb{E}(\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\mathbb{E}(\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Y}}).$$

This shows that $\boldsymbol{\eta}^*$ exists uniquely and is identifiable.

Now, we consider the identifiability of $\boldsymbol{\theta}$. Let $\{\widetilde{\theta}_i\}_{i=0}^n$ be another parameter set satisfying

$$\widetilde{\theta}_0 = 0, \qquad \widetilde{\theta}_i - \widetilde{\theta}_j = \theta_i^* - \theta_j^*.$$

Because $\widetilde{\theta}_0 = 0$ and $\theta_0^* = 0$, we have

$$\widetilde{\theta}_i = \theta_i^*, \quad i = 1, \ldots, n.$$

Therefore, θ_i^* is identifiable. This completes the proof. □

S2 Sufficient conditions for Condition 4

If $\boldsymbol{Z}_{ijt}$'s for $i \neq j \in [n]$ and $t \in [T]$ are independently generated from a p-dimensional non-degenerated multivariate symmetric distribution, then Condition 4 holds.

Proof. Recall

$$\boldsymbol{U} = (\boldsymbol{U}_{01}, \dots, \boldsymbol{U}_{0n}, \boldsymbol{U}_{12}, \dots, \boldsymbol{U}_{1n}, \dots, \boldsymbol{U}_{(n-1)n})^\top \in \mathbb{R}^{N \times n}$$

and $\boldsymbol{U}_{ij} = \boldsymbol{e}_i - \boldsymbol{e}_j \in \mathbb{R}^n$, where $\boldsymbol{e}_i \in \mathbb{R}^n$ is the standard basis vector with the ith element 1 and others 0. $\boldsymbol{e}_0$ is an n-dimensional column vector with all elements 0. By noticing $\overline{\boldsymbol{Z}}_{ij} = -\overline{\boldsymbol{Z}}_{ji}$, a direct calculation gives

$$\boldsymbol{U}^\top \overline{\boldsymbol{Z}} = \begin{pmatrix} \sum_{j \neq 1} \overline{\boldsymbol{Z}}_{1j}^\top \\ \sum_{j \neq 2} \overline{\boldsymbol{Z}}_{2j}^\top \\ \vdots \\ \sum_{j \neq n} \overline{\boldsymbol{Z}}_{nj}^\top \end{pmatrix} \tag{S2.7}$$

and

$$\boldsymbol{1}_n^\top \boldsymbol{U}^\top \overline{\boldsymbol{Z}} = -\sum_{j \neq 0} \overline{\boldsymbol{Z}}_{0j}. \tag{S2.8}$$

According the definition of $\boldsymbol{D} = \boldsymbol{I}_{N \times N} - \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top \in \mathbb{R}^{N \times N}$ and $\boldsymbol{V}^{-1} = \frac{1}{n+1}(\boldsymbol{1}_n \boldsymbol{1}_n^\top + \boldsymbol{I}_{n \times n})$, we have

$$\overline{\boldsymbol{Z}}^\top \boldsymbol{D} \overline{\boldsymbol{Z}} = \sum_{i<j} \overline{\boldsymbol{Z}}_{ij} \overline{\boldsymbol{Z}}_{ij}^\top - \frac{1}{n+1}(\overline{\boldsymbol{Z}}^\top \boldsymbol{U} \boldsymbol{1}_n \boldsymbol{1}_n^\top \boldsymbol{U}^\top \overline{\boldsymbol{Z}} + \overline{\boldsymbol{Z}}^\top \boldsymbol{U} \boldsymbol{U}^\top \overline{\boldsymbol{Z}}). \tag{S2.9}$$

Combining (S2.7), (S2.8) and (S2.9), it yields

$$\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}} = \sum_{i<j} \overline{\boldsymbol{Z}}_{ij}\overline{\boldsymbol{Z}}_{ij}^\top - \frac{1}{n+1}\sum_{i=0}^{n}\left(\sum_{j,k=0,j,k\neq i}^{n} \overline{\boldsymbol{Z}}_{ij}\overline{\boldsymbol{Z}}_{ik}^\top\right).$$

If $\boldsymbol{Z}_{ijt}$'s for $i \neq j \in [n]$ and $t \in [T]$ are independently generated from a p-dimensional non-degenerated multivariate symmetric distribution, whose general random vector is denoted by $\boldsymbol{Z}$, then, by the large sample theory, $N^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}}$ converges to $\mathrm{Cov}(\boldsymbol{Z})$ almost surely, which implies Condition 4. □

S3 Proof for Theorem 1

Before proving Theorem 1, we present some preliminary lemmas. Let $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$ be the ith columns of $\boldsymbol{U}$ and $\boldsymbol{V}$, respectively. Define

$$\begin{aligned} F_i(\boldsymbol{\theta},\boldsymbol{\eta}) &= \boldsymbol{U}_i^\top(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}) - \boldsymbol{V}_i^\top\boldsymbol{\theta} \\ &= \sum_{j=0,j\neq i}^{n} [\overline{Y}_{ij} - (\theta_i - \theta_j + \overline{\boldsymbol{Z}}_{ij}^\top\boldsymbol{\eta})], \quad i = 1,\ldots,n. \end{aligned} \tag{S3.10}$$

$$\begin{aligned} \boldsymbol{F}(\boldsymbol{\theta},\boldsymbol{\eta}) &= \boldsymbol{U}^\top(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}) - \boldsymbol{V}\boldsymbol{\theta} \\ &= (F_1(\boldsymbol{\theta},\boldsymbol{\eta}),\ldots,F_n(\boldsymbol{\theta},\boldsymbol{\eta}))^\top : \mathbb{R}^{n+p} \mapsto \mathbb{R}^n. \end{aligned} \tag{S3.11}$$

Observe that

$$\sum_{i=1}^{n} F_i(\boldsymbol{\theta},\boldsymbol{\eta}) = \sum_{j=1}^{n}[\overline{Y}_{0j} - (\theta_0 - \theta_j + \overline{\boldsymbol{Z}}_{0j}^\top\boldsymbol{\eta})] := F_0(\boldsymbol{\theta},\boldsymbol{\eta}). \tag{S3.12}$$

For a given $\boldsymbol{\eta}$, we denote

$$\boldsymbol{F}_{\boldsymbol{\eta}}(\boldsymbol{\theta}) = \left(F_{1,\boldsymbol{\eta}}(\boldsymbol{\theta}), \ldots, F_{n,\boldsymbol{\eta}}(\boldsymbol{\theta})\right)^{\top}.$$

Define $\boldsymbol{\theta}^*_{\boldsymbol{\eta}}$ as the solution to $\mathbb{E}(\boldsymbol{F}_{\boldsymbol{\eta}}(\boldsymbol{\theta}))$ for a given $\boldsymbol{\eta}$. By (S3.11), it satisfies

$$\boldsymbol{\theta}^*_{\boldsymbol{\eta}} = (\theta^*_{1,\boldsymbol{\eta}}, \ldots, \theta^*_{n,\boldsymbol{\eta}})^{\top} = \boldsymbol{V}^{-1}\boldsymbol{U}^{\top}\mathbb{E}(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}), \tag{S3.13}$$

where, for $i = 1, \ldots, n$,

$$\theta^*_{i,\boldsymbol{\eta}} = \frac{1}{n+1}\sum_{j=0, j\neq i}^{n} \mathbb{E}(\overline{Y}_{ij} - \overline{\boldsymbol{Z}}_{ij}^{\top}\boldsymbol{\eta}) + \frac{1}{n+1}\sum_{j=1}^{n}\mathbb{E}(\overline{Y}_{j0} - \overline{\boldsymbol{Z}}_{j0}^{\top}\boldsymbol{\eta}), \tag{S3.14}$$

and $\theta^*_{0,\boldsymbol{\eta}} = 0$.

Lemma S3.1. *Let Θ be a neighbourhood containing the true parameter η^*. Under Conditions 3–6, we have*

$$\sup_{\boldsymbol{\eta}\in\Theta}\max_{i\in[n]_0} |F_{i,\boldsymbol{\eta}}(\boldsymbol{\theta})| = O_p\left(\frac{\sqrt{n\log n}}{\varpi}\right), \tag{S3.15}$$

where $F_{i,\boldsymbol{\eta}}(\boldsymbol{\theta})$, $i = 0, \ldots, n$, are defined in (S3.10) *and* (S3.12).

Proof. Let $\{\boldsymbol{\eta}_k \in \mathbb{R}^p, k \in [J_n]\}$ be the equal grid points on Θ and $\boldsymbol{\eta}_k^0$ be an interior point of $[\boldsymbol{\eta}_k, \boldsymbol{\eta}_{k+1}) = [\eta_{k,1}, \eta_{k+1,1}) \times \cdots \times [\eta_{k,p}, \eta_{k+1,p})$. We choose a sufficiently large J_n such that, for all $k \in [J_n]$ and any $\boldsymbol{\eta}_k^1$ and $\boldsymbol{\eta}_k^2 \in [\boldsymbol{\eta}_k, \boldsymbol{\eta}_{k+1})$,

$$\|\boldsymbol{\eta}_k^1 - \boldsymbol{\eta}_k^2\|_2 < C, \tag{S3.16}$$

where C is a constant. For any $\delta > 0$, because the following two events are equivalent,

$$\big\{\sup_{\boldsymbol{\eta}\in\Theta}\max_{i\in[n]}|F_{i,\boldsymbol{\eta}}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}})|>\delta\big\} \text{ and } \big\{\bigcup_{k\in[J_n]}\sup_{\boldsymbol{\eta}\in[\boldsymbol{\eta}_k,\boldsymbol{\eta}_{k+1})}\max_{i\in[n]}|F_{i,\boldsymbol{\eta}}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}})|>\delta\big\},$$

applying the union bound to the second event above, we have

$$\begin{aligned}
&\sum_{k\in[J_n]}P\left(\sup_{\boldsymbol{\eta}\in[\boldsymbol{\eta}_k,\boldsymbol{\eta}_{k+1})}\max_{i\in[n]_0}|F_{i,\boldsymbol{\eta}}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}})|>\delta\right)\\
&\le\sum_{k\in[J_n]}P\left(\sup_{\boldsymbol{\eta}\in[\boldsymbol{\eta}_k,\boldsymbol{\eta}_{k+1})}\max_{i\in[n]_0}|F_{i,\boldsymbol{\eta}}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}})-F_{i,\boldsymbol{\eta}^0_k}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}^0_k})|+\max_{i\in[n]_0}|F_{i,\boldsymbol{\eta}^0_k}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}^0_k})|>\delta\right)\\
&\le\underbrace{\sum_{k\in[J_n]}P\left(\sup_{\boldsymbol{\eta}\in[\boldsymbol{\eta}_k,\boldsymbol{\eta}_{k+1})}\max_{i\in[n]_0}|F_{i,\boldsymbol{\eta}}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}})-F_{i,\boldsymbol{\eta}^0_k}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}^0_k})|>\frac{\delta}{2}\right)}_{(I)}\\
&\quad+\underbrace{\sum_{k\in[J_n]}P\left(\max_{i\in[n]_0}|F_{i,\boldsymbol{\eta}^0_k}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}^0_k})|>\frac{\delta}{2}\right)}_{(II)}. \qquad \text{(S3.17)}
\end{aligned}$$

Consider (I) in (S3.17). According to the definitions of $\boldsymbol{V}_i$, $\boldsymbol{V}$, and $\boldsymbol{U}$, we have

$$\boldsymbol{V}_i^\top\boldsymbol{V}^{-1}\boldsymbol{U}^\top=\boldsymbol{U}_i^\top.$$

For $i=0,1,\ldots,n$, the above equation, together with (S3.11), (S3.13) and Condition 5, gives

$$\begin{aligned}
|F_{i,\boldsymbol{\eta}}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}})-F_{i,\boldsymbol{\eta}^0_k}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}^0_k})|&=|[\boldsymbol{U}_i^\top(\overline{\boldsymbol{Y}}-\overline{\boldsymbol{Z}}\boldsymbol{\eta})-\boldsymbol{V}_i^\top\boldsymbol{\theta}^*_{\boldsymbol{\eta}}]-[\boldsymbol{U}_i^\top(\overline{\boldsymbol{Y}}-\overline{\boldsymbol{Z}}\boldsymbol{\eta}^0_k)-\boldsymbol{V}_i^\top\boldsymbol{\theta}^*_{\boldsymbol{\eta}^0_k}]|\\
&=|\boldsymbol{U}_i^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}-\boldsymbol{\eta}^0_k)-\boldsymbol{V}_i^\top\boldsymbol{V}^{-1}\boldsymbol{U}^\top\mathbb{E}(\overline{\boldsymbol{Z}})(\boldsymbol{\eta}-\boldsymbol{\eta}^0_k)|\\
&=|\boldsymbol{U}_i^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}-\boldsymbol{\eta}^0_k)-\boldsymbol{U}_i^\top\mathbb{E}(\overline{\boldsymbol{Z}})(\boldsymbol{\eta}-\boldsymbol{\eta}^0_k)|\\
&\le|\boldsymbol{U}_i^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}-\boldsymbol{\eta}^0_k)|+|\boldsymbol{U}_i^\top\mathbb{E}(\overline{\boldsymbol{Z}})(\boldsymbol{\eta}-\boldsymbol{\eta}^0_k)|\\
&\le 2\kappa n\|\boldsymbol{\eta}-\boldsymbol{\eta}^0_k\|_2<2\kappa nC,
\end{aligned}$$

where the last inequality is due to (S3.16). By setting $C = \delta/4\kappa n$, we have

$$(I) = 0. \tag{S3.18}$$

We now consider (II) in (S3.17). Recall the definition of $\overline{\boldsymbol{Y}}$:

$$\overline{\boldsymbol{Y}} = (\overline{Y}_{01}, \ldots, \overline{Y}_{0n}, \overline{Y}_{12}, \ldots, \overline{Y}_{1n}, \ldots, \overline{Y}_{(n-1)n})^\top, \quad \overline{Y}_{ij} = \frac{1}{T}\sum_{t\in[T]} Y_{ijt}, \tag{S3.19}$$

where

$$Y_{ijt} = \frac{a_{ijt} - \mathbb{I}(X_{ijt,0} > 0)}{f(X_{ijt,0}|\boldsymbol{Z}_{ijt})}, \text{ for } 0 \le i < j \le n \text{ and } t \in [T].$$

Condition 6 implies

$$\overline{Y}_{ij} \in [-1/\varpi, 1/\varpi].$$

By the definition of $F_i(\boldsymbol{\theta}, \boldsymbol{\eta})$ in (S3.10), we have

$$F_i(\boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{j=0, j\neq i}^{n} [\overline{Y}_{ij} - (\theta_i - \theta_j + \overline{\boldsymbol{Z}}_{ij}^\top \boldsymbol{\eta})], \quad i = 0, 1, \ldots, n.$$

By the definition of $\boldsymbol{U}$ in (2.3), its column vector satisfies

$$\boldsymbol{U}_i^\top \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U} = \boldsymbol{U}_i^\top, \qquad i = 1, \ldots, n.$$

Define $\boldsymbol{U}_0 = (\mathbf{1}_n^\top, \mathbf{0}_{n^2}^\top)^\top \in \mathbb{R}^N$, which satisfies

$$\boldsymbol{U}_0^\top \boldsymbol{U} = -\mathbf{1}_n^\top, \quad \boldsymbol{U}_0^\top \boldsymbol{U}\boldsymbol{V}^{-1} = -\mathbf{1}_n^\top, \quad \boldsymbol{U}_0^\top \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U} = \boldsymbol{U}_0^\top.$$

According to the expression of $\boldsymbol{\theta}_{\boldsymbol{\eta}}^*$ in (S3.13), conditional on $\{\overline{\boldsymbol{Z}}_{ij}\}_{i<j}$, we

have

$$
\begin{aligned}
|\boldsymbol{F}_{i,\boldsymbol{\eta}_k^0}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}_k^0})| &= |\boldsymbol{U}_i^\top(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}_k^0) - \boldsymbol{U}_i^\top \boldsymbol{U}\boldsymbol{\theta}^*_{\boldsymbol{\eta}_k^0}| \\
&= |\boldsymbol{U}_i^\top(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}_k^0) - \boldsymbol{U}_i^\top \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top \mathbb{E}(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}_k^0)| \\
&= |\boldsymbol{U}_i^\top(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}_k^0) - \boldsymbol{U}_i^\top \mathbb{E}(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}_k^0)| \\
&= |\boldsymbol{U}_i^\top[\overline{\boldsymbol{Y}} - \mathbb{E}(\boldsymbol{Y})]|, \quad i = 0, 1, \ldots, n.
\end{aligned}
$$

Because $\boldsymbol{U}_i^\top \overline{\boldsymbol{Y}} = \sum_{j=0, j\neq i}^n \overline{Y}_{ij}$ is a sum of n independent random variables, by Hoeffding's inequality, we have

$$
\begin{aligned}
& \mathbb{P}\left(\max_{i\in[n]_0} |F_{i,\boldsymbol{\eta}_k^0}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}_k^0})| \geq \frac{\delta}{2} \Big| \{\overline{\boldsymbol{Z}}_{ij}\}_{i<j}\right) \\
\leq\ & \sum_{i=0}^n \mathbb{P}\left(|\boldsymbol{U}_i^\top[\overline{\boldsymbol{Y}} - \mathbb{E}(\boldsymbol{Y})]| \geq \frac{\delta}{2} \Big| \{\overline{\boldsymbol{Z}}_{ij}\}_{i<j}\right) \\
\leq\ & 2(n+1)\exp\left(-\frac{\delta^2}{8n/\varpi^2}\right) \\
\leq\ & 2(n+1)\exp\left(-\frac{\delta^2}{8(n+1)/\varpi^2}\right).
\end{aligned}
$$

By setting $\delta = c\sqrt{(n+1)\log(n+1)}/\varpi$, we have

$$
\begin{aligned}
(II) &= \sum_{k\in[J_n]} \mathbb{P}\left(\max_{i\in[n]_0} |F_{i,\boldsymbol{\eta}_k^0}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}_k^0})| > \frac{c\sqrt{(n+1)\log(n+1)}}{2\varpi} \Big| \{\overline{\boldsymbol{Z}}_{ij}\}_{i<j}\right) \\
&\leq \sum_{k\in[J_n]} 2(n+1)\exp\left(-\frac{c^2(n+1)\log(n+1)/\varpi^2}{8(n+1)/\varpi^2}\right) \\
&= \sum_{k\in[J_n]} 2(n+1)\exp\left(-\frac{c^2\log(n+1)}{8}\right) \\
&= \sum_{k\in[J_n]} 2(n+1)^{1-\frac{c^2}{8}} \\
&= O(n^p)\, 2(n+1)^{1-\frac{c^2}{8}}
\end{aligned}
$$

$$\leq \quad O\left((n+1)^{p+1-\frac{c^2}{8}}\right) \tag{S3.20}$$

for $c > \sqrt{8(p+1)}$, where the second to last equality is due to $J_n = O(n^p)$. Combining (S3.17), (S3.18) and (S3.20) yields (S3.15). This completes the proof. □

Lemma S3.2. *Let Θ be a neighbourhood containing the true parameter $\boldsymbol{\eta}^*$. Under Conditions 4–7, we have*

$$\sup_{\boldsymbol{\eta}\in\Theta} \|\widehat{\boldsymbol{\theta}}_{\boldsymbol{\eta}} - \boldsymbol{\theta}^*_{\boldsymbol{\eta}}\|_\infty = O_p\left(\frac{1}{\varpi^2}\left(\sqrt{\frac{\log n}{n}} + \frac{\sqrt{\log n}}{nh^{p+1}} + h^r\right)\right).$$

Proof. Note that $\breve{\boldsymbol{Y}}$ is defined as

$$\breve{\boldsymbol{Y}} = (\breve{Y}_{01}, \ldots, \breve{Y}_{0n}, \breve{Y}_{12}, \ldots, \breve{Y}_{1n}, \ldots, \breve{Y}_{(n-1)n})^\top, \quad \breve{Y}_{ij} = \frac{1}{T}\sum_{t\in[T]} \widehat{Y}_{ijt}, \tag{S3.21}$$

where

$$\widehat{Y}_{ijt} = \frac{a_{ijt} - \mathbb{I}(X_{ijt,0} > 0)}{\hat{f}(X_{ijt,0}|\boldsymbol{Z}_{ijt})}, \text{ for } 0 \leq i < j \leq n \text{ and } t \in [T].$$

For each given $\boldsymbol{\eta} \in \Theta$, recall

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\theta}} \|\breve{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{\theta} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}\|_2^2. \tag{S3.22}$$

A direct calculation leads to

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\eta}} = \boldsymbol{V}^{-1}\boldsymbol{U}^\top(\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}). \tag{S3.23}$$

By the definition of $\boldsymbol{\theta}^*_{\boldsymbol{\eta}}$ in (S3.13), it follows that

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\eta}} - \boldsymbol{\theta}^*_{\boldsymbol{\eta}} \quad = \quad \boldsymbol{V}^{-1}[\boldsymbol{U}^\top(\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}) - \boldsymbol{V}\boldsymbol{\theta}^*_{\boldsymbol{\eta}}]$$

$$
\begin{aligned}
&= \boldsymbol{V}^{-1}[\boldsymbol{U}^\top\overline{\boldsymbol{Y}} - \boldsymbol{U}^\top\overline{\boldsymbol{Z}}\boldsymbol{\eta} - \boldsymbol{V}\boldsymbol{\theta}^*_{\boldsymbol{\eta}} + \boldsymbol{U}^\top\breve{\boldsymbol{Y}} - \boldsymbol{U}^\top\overline{\boldsymbol{Y}}] \\
&= \underbrace{\boldsymbol{V}^{-1}\boldsymbol{F}_{\boldsymbol{\eta}}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}})}_{(I)} + \underbrace{\boldsymbol{V}^{-1}\boldsymbol{U}^\top(\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Y}})}_{(II)}. \qquad \text{(S3.24)}
\end{aligned}
$$

We first consider (I) in (S3.24). By Condition 4, (S3.12) and Lemma S3.1, we have

$$
\begin{aligned}
\|\boldsymbol{V}^{-1}\boldsymbol{F}_{\boldsymbol{\eta}}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}})\|_\infty &= \frac{1}{n+1}\max_{i\in[n]}\left|\sum_{j\in[n]} F_{j,\boldsymbol{\eta}}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}}) + F_{i,\boldsymbol{\eta}}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}})\right| \\
&= \frac{1}{n+1}\max_{i\in[n]}\left|F_{0,\boldsymbol{\eta}}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}}) + F_{i,\boldsymbol{\eta}}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}})\right| \\
&\le \frac{2}{n+1}\max_{i\in[n]_0}|F_{i,\boldsymbol{\eta}}(\boldsymbol{\theta})| \\
&= O_p\left(\sqrt{\frac{\log n}{\varpi^2 n}}\right). \qquad \text{(S3.25)}
\end{aligned}
$$

We then consider (II) in (S3.24). According to the definitions of $\breve{Y}_{ij}$ in (S3.21) and $\overline{Y}_{ij}$ in (S3.19), we have

$$
\begin{aligned}
&\|\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Y}}\|_\infty = \max_{0\le i<j\le n}|\breve{Y}_{ij} - \overline{Y}_{ij}| \\
= &\max_{0\le i<j\le n}\left|\frac{1}{T}\sum_{t=1}^T(\breve{Y}_{ijt} - \overline{Y}_{ijt})\right| \le \max_{0\le i<j\le n}|\breve{Y}_{ijt} - \overline{Y}_{ijt}| \\
= &\max_{0\le i<j\le n}\left|\frac{a_{ijt} - \mathbb{I}(X_{ijt,0}\ge 0)}{\hat{f}(X_{ijt,0} = x_{ijt,0}|\boldsymbol{Z}_{ijt} = \boldsymbol{z}_{ijt})} - \frac{a_{ijt} - \mathbb{I}(X_{ijt,0}\ge 0)}{f(X_{ijt,0} = x_{ijt,0}|\boldsymbol{Z}_{ijt} = \boldsymbol{z}_{ijt})}\right| \\
= &\max_{0\le i<j\le n}\left|(a_{ijt} - \mathbb{I}(X_{ijt,0}\ge 0))\left(\frac{1}{\hat{f}(x_{ijt,0}|\boldsymbol{z}_{ijt})} - \frac{1}{f(x_{ijt,0}|\boldsymbol{z}_{ijt})}\right)\right| \\
\le &\max_{0\le i<j\le n}\left|\frac{1}{\hat{f}(x_{ijt,0}|\boldsymbol{z}_{ijt})} - \frac{1}{f(x_{ijt,0}|\boldsymbol{z}_{ijt})}\right| \\
= &\max_{0\le i<j\le n}\left|\frac{\hat{f}(\boldsymbol{z}_{ijt})}{\hat{f}(x_{ijt,0}, \boldsymbol{z}_{ijt})} - \frac{f(\boldsymbol{z}_{ijt})}{f(x_{ijt,0}, \boldsymbol{z}_{ijt})}\right|
\end{aligned}
$$

$$
\begin{aligned}
&= \max_{0\le i<j\le n}\left|\frac{\hat f(\boldsymbol{z}_{ijt})}{\hat f(x_{ijt,0},\boldsymbol{z}_{ijt})}-\frac{f(\boldsymbol{z}_{ijt})}{\hat f(x_{ijt,0},\boldsymbol{z}_{ijt})}+\frac{f(\boldsymbol{z}_{ijt})}{\hat f(x_{ijt,0},\boldsymbol{z}_{ijt})}-\frac{f(\boldsymbol{z}_{ijt})}{f(x_{ijt,0},\boldsymbol{z}_{ijt})}\right| \\
&= \max_{0\le i<j\le n}\left|\frac{\hat f(\boldsymbol{z}_{ijt})-f(\boldsymbol{z}_{ijt})}{\hat f(x_{ijt,0},\boldsymbol{z}_{ijt})}+\frac{f(x_{ijt,0},\boldsymbol{z}_{ijt})-\hat f(x_{ijt,0},\boldsymbol{z}_{ijt})}{\hat f(x_{ijt,0},\boldsymbol{z}_{ijt})f(x_{ijt,0}|\boldsymbol{z}_{ijt})}\right| \\
&\le \max_{0\le i<j\le n}\frac{|\hat f(\boldsymbol{z}_{ijt})-f(\boldsymbol{z}_{ijt})|}{\hat f(x_{ijt,0},\boldsymbol{z}_{ijt})}+\max_{0\le i<j\le n}\frac{|f(x_{ijt,0},\boldsymbol{z}_{ijt})-\hat f(x_{ijt,0},\boldsymbol{z}_{ijt})|}{\hat f(x_{ijt,0},\boldsymbol{z}_{ijt})f(x_{ijt,0}|\boldsymbol{z}_{ijt})}. \quad \text{(S3.26)}
\end{aligned}
$$

Here, $\hat f(\boldsymbol{z}_{ijt})$ and $\hat f(x_{ijt,0},\boldsymbol{z}_{ijt})$ denote the kernel estimators for $f(\boldsymbol{z}_{ijt})$ and $f(x_{ijt,0},\boldsymbol{z}_{ijt})$, respectively, which are defined as follows:

$$
\hat f(\boldsymbol{z}_{ijt}) = \frac{1}{NT}\sum_{0\le s\ne k\le n}\sum_{l\in[T]}\mathcal{K}_{z,h}(\boldsymbol{z}_{ijt}-\boldsymbol{Z}_{skl}), \quad \text{(S3.27)}
$$

$$
\hat f(x_{ijt,0},\boldsymbol{z}_{ijt}) = \frac{1}{NT}\sum_{0\le s\ne k\le n}\sum_{l\in[T]}\mathcal{K}_{xz,h}(x_{ijt,0}-X_{skl,0},\boldsymbol{z}_{ijt}-\boldsymbol{Z}_{skl}). \quad \text{(S3.28)}
$$

By Theorem 1(a) of Andrews (1995, page 9) (reproduced in Lemma S11.2) for nonparametric density estimators, we have

$$
|\hat f(\boldsymbol{z}_{ijt})-f(\boldsymbol{z}_{ijt})| = O_p\left(\frac{1}{nh^p}+h^r\right) \le O_p\left(\frac{\sqrt{\log n}}{nh^p}+h^r\right),
$$

$$
|\hat f(x_{ijt,0},\boldsymbol{z}_{ijt})-f(x_{ijt,0},\boldsymbol{z}_{ijt})| = O_p\left(\frac{1}{nh^{p+1}}+h^r\right) \le O_p\left(\frac{\sqrt{\log n}}{nh^{p+1}}+h^r\right).
$$

By the strong law of large number, we have

$$
\hat f(x_{ijt,0},\boldsymbol{z}_{ijt}) \to f(x_{ijt,0},\boldsymbol{z}_{ijt}), \quad \hat f(\boldsymbol{z}_{ijt}) \to f(\boldsymbol{z}_{ijt}), \text{ as } n\to\infty, \quad \text{(S3.29)}
$$

almost surely. Since $f(x_{ijt,0},\boldsymbol{z}_{ijt}) > \varpi$ and $f(\boldsymbol{z}_{ijt}) > \varpi$ in Condition 6, we have

$$\hat{f}(x_{ijt,0}, \boldsymbol{z}_{ijt}) > \varpi, \quad \hat{f}(\boldsymbol{z}_{ijt}) > \varpi \tag{S3.30}$$

almost surely. Again, since $f(x_{ijt,0}|\boldsymbol{z}_{ijt}) > \varpi$ in Condition 6, it follows that (S3.26) satisfies

$$\begin{aligned} \|\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Y}}\|_\infty &= \max_{0\le i<j\le n} \frac{|\hat{f}(\boldsymbol{z}_{ijt}) - f(\boldsymbol{z}_{ijt})|}{\hat{f}(x_{ijt,0}, \boldsymbol{z}_{ijt})} + \max_{0\le i<j\le n} \frac{\left|f(x_{ijt,0}, \boldsymbol{z}_{ijt}) - \hat{f}(x_{ijt,0}, \boldsymbol{z}_{ijt})\right|}{\hat{f}(x_{ijt,0}, \boldsymbol{z}_{ijt}) f(x_{ijt,0}|\boldsymbol{z}_{ijt})} \\ &\le \frac{1}{\varpi} O_p\left(\frac{\sqrt{\log n}}{nh^p} + h^r\right) + \frac{1}{\varpi^2} O_p\left(\frac{\sqrt{\log n}}{nh^{p+1}} + h^r\right) \\ &\le O_p\left(\frac{1}{\varpi^2}\left(\frac{\sqrt{\log n}}{nh^{p+1}} + h^r\right)\right). \end{aligned} \tag{S3.31}$$

Therefore, we have

$$\begin{aligned} \|\boldsymbol{V}^{-1}\boldsymbol{U}^\top(\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Y}})\|_\infty &= \max_{i\in[n]} |\boldsymbol{V}_i^{-1}\boldsymbol{U}^\top(\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Y}})| \\ &= \frac{1}{n+1} \max_{i\in[n]} |\sum_{i\in[n]} \sum_{j=0, j\neq i}^{n} (\breve{Y}_{ij} - \overline{Y}_{ij})| \\ &= \frac{1}{n+1} \max_{i\in[n]} |\sum_{j\in[n]/i} (\breve{Y}_{ij} - \overline{Y}_{ij})| \\ &\le \frac{n-1}{n+1} \|\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Y}}\|_\infty \\ &= O_p\left(\frac{1}{\varpi^2}\left(\frac{\sqrt{\log n}}{nh^{p+1}} + h^r\right)\right). \end{aligned} \tag{S3.32}$$

Combining (S3.24), (S3.25) and (S3.32), we have

$$\sup_{\boldsymbol{\eta}\in\Theta} \|\widehat{\boldsymbol{\theta}}_{\boldsymbol{\eta}} - \boldsymbol{\theta}^*_{\boldsymbol{\eta}}\|_\infty = O_p\left(\frac{1}{\varpi^2}\left(\sqrt{\frac{\log n}{n}} + \frac{\sqrt{\log n}}{nh^{p+1}} + h^r\right)\right).$$

This completes the proofs. □

We are now ready to prove Theorem 1.

Proof of Theorem 1. Define

$$\begin{aligned}\widehat{\boldsymbol{Q}}(\boldsymbol{\eta}) &= \overline{\boldsymbol{Z}}^\top(\check{\boldsymbol{Y}} - \boldsymbol{U}\widehat{\boldsymbol{\theta}}_{\boldsymbol{\eta}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}),\\ \widetilde{\boldsymbol{Q}}(\boldsymbol{\eta}) &= \overline{\boldsymbol{Z}}^\top(\check{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{\theta}^*_{\boldsymbol{\eta}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}),\\ \boldsymbol{Q}(\boldsymbol{\eta}) &= \overline{\boldsymbol{Z}}^\top(\overline{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{\theta}^*_{\boldsymbol{\eta}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}).\end{aligned} \tag{S3.33}$$

Then, we have

$$\begin{aligned}\|\widehat{\boldsymbol{Q}}(\boldsymbol{\eta}) - \boldsymbol{Q}(\boldsymbol{\eta})\|_2 &\le \|\widehat{\boldsymbol{Q}}(\boldsymbol{\eta}) - \widetilde{\boldsymbol{Q}}(\boldsymbol{\eta})\|_2 + \|\widetilde{\boldsymbol{Q}}(\boldsymbol{\eta}) - \boldsymbol{Q}(\boldsymbol{\eta})\|_2\\ &= \underbrace{\|\overline{\boldsymbol{Z}}^\top\boldsymbol{U}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\eta}} - \boldsymbol{\theta}^*_{\boldsymbol{\eta}})\|_2}_{(I)} + \underbrace{\|\overline{\boldsymbol{Z}}^\top(\check{\boldsymbol{Y}} - \overline{\boldsymbol{Y}})\|_2}_{(II)}.\end{aligned} \tag{S3.34}$$

We first consider (I) in (S3.34). Note that

$$\sum_{i\in[n]_0}\sum_{j=0,j\neq i}^{n}\overline{Z}_{ij,k} = 0,$$

due to $\overline{Z}_{ij,k} + \overline{Z}_{ji,k} = 0$. By Lemma S3.2, we have

$$\begin{aligned}(I) &= \left[\sum_{k\in[p]}\left(\sum_{i\in[n]}\sum_{j=0,j\neq i}^{n}\overline{Z}_{ij,k}(\widehat{\theta}_{\boldsymbol{\eta},i} - \theta^*_{\boldsymbol{\eta},i})\right)^2\right]^{1/2}\\ &\le \left[\sum_{k\in[p]}\left(\sum_{i\in[n]}\sum_{j=0,j\neq i}^{n}\overline{Z}_{ij,k}\right)^2\right]^{1/2}\|\widehat{\boldsymbol{\theta}}_{\boldsymbol{\eta}} - \boldsymbol{\theta}^*_{\boldsymbol{\eta}}\|_\infty\\ &= \left[\sum_{k\in[p]}\left(\sum_{j\in[n]}\overline{Z}_{j0,k}\right)^2\right]^{1/2}\|\widehat{\boldsymbol{\theta}}_{\boldsymbol{\eta}} - \boldsymbol{\theta}^*_{\boldsymbol{\eta}}\|_\infty\\ &\le \sqrt{p}\kappa n \times O_p\left(\frac{1}{\varpi^2}\left(\sqrt{\frac{\log n}{n}} + \frac{\sqrt{\log n}}{nh^{p+1}} + h^r\right)\right)\end{aligned}$$

$$= O_p\left(\frac{\kappa}{\varpi^2}\left(\sqrt{n\log n} + \frac{\sqrt{\log n}}{h^{p+1}} + nh^r\right)\right). \tag{S3.35}$$

We then consider (II) in (S3.34). Under Conditions 5–7, by (S3.31), we have

$$\begin{aligned}
(II) &= \sqrt{\sum_{k\in[p]}\left(\sum_{0\le i<j\le n} \overline{Z}_{ij,k}(\breve{Y}_{ij} - \overline{Y}_{ij})\right)^2} \\
&\le \sqrt{\sum_{k\in[p]}\left(\sum_{0\le i<j\le n} \kappa\|\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Y}}\|_\infty\right)^2} \\
&\le \kappa\sqrt{p}N \times O_p\left(\frac{1}{\varpi^2}\left(\frac{\sqrt{\log n}}{nh^{p+1}} + h^r\right)\right) \\
&= O_p\left(\frac{\kappa}{\varpi^2}\left(\frac{n\sqrt{\log n}}{h^{p+1}} + n^2h^r\right)\right). \qquad \text{(S3.36)}
\end{aligned}$$

Combining (S3.34), (S3.35) and (S3.36) yields

$$\|\widehat{\boldsymbol{Q}}(\boldsymbol{\eta}) - \boldsymbol{Q}(\boldsymbol{\eta})\|_2 = O_p\left(\frac{\kappa}{\varpi^2}\left(\sqrt{n\log n} + \frac{n\sqrt{\log n}}{h^{p+1}} + n^2h^r\right)\right). \tag{S3.37}$$

We now bound $\boldsymbol{Q}(\boldsymbol{\eta}) - \mathbb{E}(\boldsymbol{Q}(\boldsymbol{\eta}))$. Note that, in (S3.13), we have

$$\boldsymbol{\theta}^*_{\boldsymbol{\eta}} = (\theta^*_{1,\boldsymbol{\eta}}, \ldots, \theta^*_{n,\boldsymbol{\eta}})^\top = \boldsymbol{V}^{-1}\boldsymbol{U}^\top\mathbb{E}(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}).$$

According to the definition of $\boldsymbol{Q}(\boldsymbol{\eta})$ in (S3.33) and the expression of $\mathbb{E}(\overline{Y}|\boldsymbol{Z})$ in Lemma 1, conditional on $\{\boldsymbol{Z}_{ij}\}_{i<j}$, we have

$$\begin{aligned}
\boldsymbol{Q}(\boldsymbol{\eta}) &= \overline{\boldsymbol{Z}}^\top(\overline{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{\theta}^*_{\boldsymbol{\eta}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}) \\
&= \overline{\boldsymbol{Z}}^\top[\overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}}) + \mathbb{E}(\overline{\boldsymbol{Y}}) - \boldsymbol{U}\boldsymbol{\theta}^*_{\boldsymbol{\eta}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}] \\
&= \overline{\boldsymbol{Z}}^\top[\overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}})] + \overline{\boldsymbol{Z}}^\top[\mathbb{E}(\overline{\boldsymbol{Y}}) - \boldsymbol{U}\boldsymbol{\theta}^*_{\boldsymbol{\eta}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}] \\
&= \overline{\boldsymbol{Z}}^\top[\overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}})] + \overline{\boldsymbol{Z}}^\top(\boldsymbol{U}\boldsymbol{\theta}^* + \overline{\boldsymbol{Z}}\boldsymbol{\eta}^* - \boldsymbol{U}\boldsymbol{\theta}^*_{\boldsymbol{\eta}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta})
\end{aligned}$$

$$
\begin{aligned}
&= \overline{\boldsymbol{Z}}^\top[\overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}})] + \overline{\boldsymbol{Z}}^\top \boldsymbol{U}(\boldsymbol{\theta}^* - \boldsymbol{\theta}^*_{\boldsymbol{\eta}}) + \overline{\boldsymbol{Z}}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \boldsymbol{\eta}) \\
&= \overline{\boldsymbol{Z}}^\top[\overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}})] + \overline{\boldsymbol{Z}}^\top \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top[\mathbb{E}(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}^*) - \mathbb{E}(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta})] + \overline{\boldsymbol{Z}}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \boldsymbol{\eta}) \\
&= \overline{\boldsymbol{Z}}^\top[\overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}})] - \overline{\boldsymbol{Z}}^\top \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta} - \boldsymbol{\eta}^*) + \overline{\boldsymbol{Z}}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \boldsymbol{\eta}) \\
&= \overline{\boldsymbol{Z}}^\top[\overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}})] + \overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \boldsymbol{\eta}). \qquad \text{(S3.38)}
\end{aligned}
$$

Because $\boldsymbol{D}\boldsymbol{U} = \boldsymbol{0}$ in (S1.6), conditional on $\{\boldsymbol{Z}_{ij}\}_{i<j}$, we have

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{Q}(\boldsymbol{\eta}^*)) &= \mathbb{E}[\overline{\boldsymbol{Z}}^\top(\overline{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{\theta}^*_{\boldsymbol{\eta}^*} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}^*)] \\
&= \mathbb{E}[\overline{\boldsymbol{Z}}^\top(\overline{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top\mathbb{E}(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}^*) - \overline{\boldsymbol{Z}}\boldsymbol{\eta}^*)] \\
&= \overline{\boldsymbol{Z}}^\top\mathbb{E}(\overline{\boldsymbol{Y}}) - \overline{\boldsymbol{Z}}^\top\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top\mathbb{E}(\overline{\boldsymbol{Y}}) + \overline{\boldsymbol{Z}}^\top\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top\overline{\boldsymbol{Z}}\boldsymbol{\eta}^* - \overline{\boldsymbol{Z}}^\top\overline{\boldsymbol{Z}}\boldsymbol{\eta}^* \\
&= \overline{\boldsymbol{Z}}^\top\boldsymbol{D}\mathbb{E}(\overline{\boldsymbol{Y}}) - \overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}}\boldsymbol{\eta}^* \\
&= \overline{\boldsymbol{Z}}^\top\boldsymbol{D}(\boldsymbol{U}\boldsymbol{\theta}^* + \overline{\boldsymbol{Z}}\boldsymbol{\eta}^*) - \overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}}\boldsymbol{\eta}^* \\
&= \overline{\boldsymbol{Z}}^\top\boldsymbol{D}\boldsymbol{U}\boldsymbol{\theta}^* + \overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}}\boldsymbol{\eta}^* - \overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}}\boldsymbol{\eta}^* = \boldsymbol{0}. \qquad \text{(S3.39)}
\end{aligned}
$$

Conditional on $\{\boldsymbol{Z}_{ij}\}_{i<j}$, it follows from (S3.38) that

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta})] &= \mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta}) - \boldsymbol{Q}(\boldsymbol{\eta}^*)] \\
&= \mathbb{E}[\overline{\boldsymbol{Z}}^\top(\overline{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{\theta}^*_{\boldsymbol{\eta}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}) - \overline{\boldsymbol{Z}}^\top(\overline{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{\theta}^*_{\boldsymbol{\eta}^*} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}^*)] \\
&= \mathbb{E}[\overline{\boldsymbol{Z}}^\top\boldsymbol{U}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}^*} - \boldsymbol{\theta}^*_{\boldsymbol{\eta}}) - \overline{\boldsymbol{Z}}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \boldsymbol{\eta})] \\
&= \mathbb{E}[\overline{\boldsymbol{Z}}^\top\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top(\mathbb{E}(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}^*) - \mathbb{E}(\overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta})) - \overline{\boldsymbol{Z}}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \boldsymbol{\eta})] \\
&= \mathbb{E}[\overline{\boldsymbol{Z}}^\top\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta} - \boldsymbol{\eta}^*) - \overline{\boldsymbol{Z}}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \boldsymbol{\eta})] \\
&= \mathbb{E}[\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \boldsymbol{\eta})] \\
&= \overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \boldsymbol{\eta}). \qquad \text{(S3.40)}
\end{aligned}
$$

Combining with (S3.38) and (S3.40), conditional on $\{\boldsymbol{Z}_{ij}\}_{i<j}$, yields

$$\begin{aligned}\boldsymbol{Q}(\boldsymbol{\eta})-\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta})] &= \overline{\boldsymbol{Z}}^\top[\overline{\boldsymbol{Y}}-\mathbb{E}(\overline{\boldsymbol{Y}})]+\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^*-\boldsymbol{\eta})-\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^*-\boldsymbol{\eta}) \\ &= \overline{\boldsymbol{Z}}^\top[\overline{\boldsymbol{Y}}-\mathbb{E}(\overline{\boldsymbol{Y}})]. \end{aligned} \tag{S3.41}$$

By Conditions 5 and 6, we have

$$|\overline{\boldsymbol{Z}}_{ij,k}\overline{Y}_{ij}| \le \frac{c_0\kappa}{\varpi}.$$

Since $\overline{\boldsymbol{Z}}_k^\top\overline{\boldsymbol{Y}}=\sum_{0\le i<j\le n}\overline{Z}_{ij,k}\overline{Y}_{ij}$ is the sum of N independently bounded random variables, by Hoeffding's inequality, we have

$$\mathbb{P}\left(|\overline{\boldsymbol{Z}}_k(\overline{\boldsymbol{Y}}-\mathbb{E}\overline{\boldsymbol{Y}})|\ge\varepsilon\Big|\{\overline{\boldsymbol{Z}}_{ij}\}_{i<j}\right)\le 2\exp\left(-\frac{\varepsilon^2}{2c_0^2\kappa^2N/\varpi^2}\right).$$

It follows from the union bound that (S3.41) satisfies

$$\begin{aligned}&\mathbb{P}\left(\|\boldsymbol{Q}(\boldsymbol{\eta})-\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta})]\|_\infty\ge\varepsilon\Big|\{\overline{\boldsymbol{Z}}_{ij}\}_{i<j}\right)\\ =\;&\mathbb{P}\left(\|\overline{\boldsymbol{Z}}^\top(\overline{\boldsymbol{Y}}-\mathbb{E}\overline{\boldsymbol{Y}})\|_\infty\ge\varepsilon\Big|\{\overline{\boldsymbol{Z}}_{ij}\}_{i<j}\right)\\ \le\;&\sum_{k=1}^p\mathbb{P}\left(|\overline{\boldsymbol{Z}}_k^\top(\overline{\boldsymbol{Y}}-\mathbb{E}\overline{\boldsymbol{Y}})|\ge\varepsilon\Big|\{\overline{\boldsymbol{Z}}_{ij}\}_{i<j}\right)\\ \le\;&2p\exp\left(-\frac{\varepsilon^2}{2c_0^2\kappa^2N/\varpi^2}\right).\end{aligned}$$

Conditional on $\{\overline{Z}_{ij,k}\}_{i<j}$, taking $\varepsilon=\sqrt{2c_0^2\kappa^2N\log(2pn^{c_2})}/\varpi=O(\kappa n\sqrt{\log n}/\varpi)$ for a constant $c_2>0$ yields

$$\|\boldsymbol{Q}(\boldsymbol{\eta})-\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta})]\|_\infty=O\left(\frac{\kappa n\sqrt{\log n}}{\varpi}\right), \tag{S3.42}$$

with probability at least $1-n^{-c_2}$. It follows that

$$\|\boldsymbol{Q}(\boldsymbol{\eta})-\mathbb{E}(\boldsymbol{Q}(\boldsymbol{\eta}))\|_2 \leq p\|\boldsymbol{Q}(\boldsymbol{\eta})-\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta})]\|_\infty = O\left(\frac{\kappa n\sqrt{\log n}}{\varpi}\right)$$

with probability at least $1-n^{-c_2}$. This, together with (S3.37), gives

$$\begin{aligned} & \|\widehat{\boldsymbol{Q}}(\boldsymbol{\eta})-\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta})]\|_2 \\ \leq\ & \|\widehat{\boldsymbol{Q}}(\boldsymbol{\eta})-\boldsymbol{Q}(\boldsymbol{\eta})\|_2+\|\boldsymbol{Q}(\boldsymbol{\eta})-\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta})]\|_2 \\ =\ & O_p\left(\frac{\kappa}{\varpi^2}\left(n\sqrt{\log n}+\frac{n\sqrt{\log n}}{h^{p+1}}+n^2h^r\right)\right). \end{aligned} \tag{S3.43}$$

We are now ready to bound the error of $\widehat{\boldsymbol{\eta}}$. Let ϑ_0 be any positive constant such that $\|\boldsymbol{\eta}-\boldsymbol{\eta}^*\|_2 \leq \vartheta_0$. For any unit vector $\boldsymbol{w}\in\mathbb{R}^p$ satisfying $\|\boldsymbol{w}\|_2=1$, the derivative of $-\boldsymbol{w}^\top\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta}+\boldsymbol{w}\vartheta)]$ with respect to ϑ is

$$-\frac{\partial \boldsymbol{w}^\top\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta}+\boldsymbol{w}\vartheta)]}{\partial\vartheta} = \boldsymbol{w}^\top\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}}\boldsymbol{w} > 0$$

by Condition 4. This shows that the function $-\boldsymbol{w}^\top\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta}+\boldsymbol{w}\vartheta)]$ increases with ϑ. For any $\vartheta>\vartheta_0>0$, the Cauchy-Schwarz inequality gives

$$\begin{aligned} & \|\boldsymbol{w}\|_2\|\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta}^*+\boldsymbol{w}\vartheta)]-\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta}^*)]\|_2 \\ \geq & |\boldsymbol{w}^\top\{\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta}^*+\boldsymbol{w}\vartheta)]-\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta}^*)]\}| \\ = & \boldsymbol{w}^\top\overline{\boldsymbol{Z}}\boldsymbol{D}\overline{\boldsymbol{Z}}^\top\boldsymbol{w}\vartheta \\ > & \lambda N\vartheta_0 = O(\lambda n^2\vartheta_0), \end{aligned}$$

where the last inequality is due to Condition 4 and $N=(n+1)n/2$. Therefore,

$$\inf_{\|\boldsymbol{\eta}-\boldsymbol{\eta}^*\|_2>\vartheta_0}\|\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta})]-\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta}^*)]\|_2 > O(\lambda n^2\vartheta_0). \tag{S3.44}$$

Assume that

$$\widehat{\boldsymbol{\eta}} \in B(\boldsymbol{\eta}^*; \|\boldsymbol{\eta} - \boldsymbol{\eta}^*\|_2 > \vartheta_0).$$

Because $\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta}^*)] = \mathbf{0}$ in (S3.39), by (S3.44), we have

$$\inf_{\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*\|_2 > \vartheta_0} \|\mathbb{E}[\boldsymbol{Q}(\widehat{\boldsymbol{\eta}})]\|_2 > O(\lambda n^2 \vartheta_0). \tag{S3.45}$$

Because $\widehat{\boldsymbol{Q}}(\widehat{\boldsymbol{\eta}}) = \mathbf{0}$, where $\widehat{\boldsymbol{Q}}(\widehat{\boldsymbol{\eta}})$ is defined in (S3.33), by (S3.43), we have

$$\begin{aligned} \|\mathbb{E}[\boldsymbol{Q}(\widehat{\boldsymbol{\eta}})]\|_2 &= \|\widehat{\boldsymbol{Q}}(\widehat{\boldsymbol{\eta}}) - \mathbb{E}[\boldsymbol{Q}(\widehat{\boldsymbol{\eta}})]\|_2 \\ &= O_p\left(\frac{\kappa}{\varpi^2}\left(n\sqrt{\log n} + \frac{n\sqrt{\log n}}{h^{p+1}} + n^2 h^r\right)\right). \end{aligned} \tag{S3.46}$$

Note that κ, λ and ϑ_0 are constants. When $n \to \infty$, we have

$$\frac{h^{p+1}}{\sqrt{\log n}/n} \to \infty,$$

which leads to

$$\frac{\frac{\kappa}{\varpi^2}\left(n\sqrt{\log n} + \frac{n\sqrt{\log n}}{h^{p+1}} + n^2 h^r\right)}{\lambda n^2 \vartheta_0} = \frac{\kappa}{\varpi^2 \lambda \vartheta_0}\left(\frac{\sqrt{\log n}}{n} + \frac{\sqrt{\log n}}{n h^{p+1}} + h^r\right) \to 0, \text{ as } n \to \infty.$$

This shows that

$$\|\mathbb{E}[\boldsymbol{Q}(\widehat{\boldsymbol{\eta}})]\|_2 \ll \lambda n^2 \vartheta_0,$$

which contradicts (S3.45). Therefore, we have

$$\widehat{\boldsymbol{\eta}} \in B(\boldsymbol{\eta}^*; \|\boldsymbol{\eta} - \boldsymbol{\eta}^*\|_2 \le \vartheta_0).$$

According to $\mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta}^*)] = \mathbf{0}$ in (S3.39) and the definition of $\boldsymbol{Q}(\boldsymbol{\eta})$ in

(S3.33), we have

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{Q}(\widehat{\boldsymbol{\eta}})] =& \mathbb{E}[\boldsymbol{Q}(\widehat{\boldsymbol{\eta}}) - \boldsymbol{Q}(\boldsymbol{\eta}^*)] \\
=& \mathbb{E}[\overline{\boldsymbol{Z}}^\top(\overline{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{\theta}^*_{\widehat{\boldsymbol{\eta}}} - \overline{\boldsymbol{Z}}\widehat{\boldsymbol{\eta}}) - \overline{\boldsymbol{Z}}^\top(\overline{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{\theta}^*_{\boldsymbol{\eta}^*} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}^*)] \\
=& \mathbb{E}[\overline{\boldsymbol{Z}}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \widehat{\boldsymbol{\eta}}) - \overline{\boldsymbol{Z}}^\top\boldsymbol{U}(\boldsymbol{\theta}^*_{\boldsymbol{\eta}^*} - \boldsymbol{\theta}^*_{\widehat{\boldsymbol{\eta}}})] \\
=& \mathbb{E}[\overline{\boldsymbol{Z}}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \widehat{\boldsymbol{\eta}}) - \overline{\boldsymbol{Z}}^\top\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top\mathbb{E}(\overline{\boldsymbol{Z}})(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*)] \\
=& \mathbb{E}[\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}}](\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*),
\end{aligned}
$$

where the second to last equality is due to the definition of $\boldsymbol{\theta}^*$ in Corollary 1, and the last equality is due to the definition of $\boldsymbol{D}$ in (2.6). Thus we have

$$
\begin{aligned}
\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*\|_2 &= \|\mathbb{E}(\overline{\boldsymbol{Z}}\boldsymbol{D}\overline{\boldsymbol{Z}}^\top)^{-1}\{\mathbb{E}[\boldsymbol{Q}(\widehat{\boldsymbol{\eta}})] - \mathbb{E}[\boldsymbol{Q}(\boldsymbol{\eta}^*)]\}\|_2 \\
&\leq \underbrace{\|\mathbb{E}(\overline{\boldsymbol{Z}}\boldsymbol{D}\overline{\boldsymbol{Z}}^\top)^{-1}\|_2}_{(I)} \underbrace{\|\mathbb{E}[\boldsymbol{Q}(\widehat{\boldsymbol{\eta}}) - \boldsymbol{Q}(\boldsymbol{\eta}^*)]\|_2}_{(II)}. \qquad \text{(S3.47)}
\end{aligned}
$$

We first consider (I) in (S3.47). By Condition 4, we have

$$
(I) \leq O(1/\lambda N) = O(1/\lambda n^2). \qquad \text{(S3.48)}
$$

Furthermore, the bound of (II) is shown in (S3.46). Combining (S3.47), (S3.48) and (S3.46) yields

$$
\begin{aligned}
\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*\|_\infty &\leq \|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*\|_2 \\
&\leq O\left(\frac{1}{n^2\lambda}\right) O_p\left(\frac{\kappa}{\varpi^2}\left(n\sqrt{\log n} + \frac{n\sqrt{\log n}}{h^{p+1}} + n^2 h^r\right)\right) \\
&= O_p\left(\frac{\kappa}{\varpi^2\lambda}\left(\frac{\sqrt{\log n}}{n} + \frac{\sqrt{\log n}}{nh^{p+1}} + h^r\right)\right). \qquad \text{(S3.49)}
\end{aligned}
$$

Now, we bound the error of $\widehat{\boldsymbol{\theta}}$ in the second part of Theorem 1. We divide it into two parts:

$$\|\widehat{\boldsymbol{\theta}}_{\widehat{\boldsymbol{\eta}}} - \boldsymbol{\theta}^*_{\boldsymbol{\eta}^*}\|_\infty = \underbrace{\|\widehat{\boldsymbol{\theta}}_{\widehat{\boldsymbol{\eta}}} - \widehat{\boldsymbol{\theta}}_{\boldsymbol{\eta}^*}\|_\infty}_{(I)'} + \underbrace{\|\widehat{\boldsymbol{\theta}}_{\boldsymbol{\eta}^*} - \boldsymbol{\theta}^*_{\boldsymbol{\eta}^*}\|_\infty}_{(II)'}. \tag{S3.50}$$

We first consider $(I)'$ in (S3.50). In view of (S3.23) and (S3.49), we have

$$\begin{aligned}
(I)' &= \|\boldsymbol{V}^{-1}\boldsymbol{U}^\top(\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\widehat{\boldsymbol{\eta}}) - \boldsymbol{V}^{-1}\boldsymbol{U}^\top(\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}^*)\|_\infty \\
&= \|\boldsymbol{V}^{-1}\boldsymbol{U}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \widehat{\boldsymbol{\eta}})\|_\infty \\
&\le O(p\kappa)\|\boldsymbol{\eta}^* - \widehat{\boldsymbol{\eta}}\|_\infty \\
&= O(p\kappa)\,O_p\left(\frac{\kappa}{\varpi^2\lambda}\left(\frac{\sqrt{\log n}}{n} + \frac{\sqrt{\log n}}{nh^{p+1}} + h^r\right)\right) \\
&= O\left(\frac{\kappa^2}{\varpi^2\lambda}\left(\frac{\sqrt{\log n}}{n} + \frac{\sqrt{\log n}}{nh^{p+1}} + h^r\right)\right).
\end{aligned} \tag{S3.51}$$

The bound of $(II)'$ is shown in Lemma S3.2, where

$$(II)' = O_p\left(\frac{1}{\varpi^2}\left(\sqrt{\frac{\log n}{n}} + \frac{\sqrt{\log n}}{nh^{p+1}} + h^r\right)\right). \tag{S3.52}$$

Combining (S3.50), (S3.51), and (S3.52) yields

$$\|\widehat{\boldsymbol{\theta}}_{\widehat{\boldsymbol{\eta}}} - \boldsymbol{\theta}^*_{\boldsymbol{\eta}^*}\|_\infty = O_p\left(\frac{1}{\varpi^2}\left(1 + \frac{\kappa^2}{\lambda}\right)\left(\sqrt{\frac{\log n}{n}} + \frac{\sqrt{\log n}}{nh^{p+1}} + h^r\right)\right).$$

This completes the proof.

□

S4 Proof for Theorem 2

Recall that we define $\overline{\boldsymbol{\tau}} = \overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{X}_0, \boldsymbol{Z})$ in (3.2). Before proving Theorem 2, we present a lemma.

Lemma S4.1. *Under Conditions 1–7, we have*

$$\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = (\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{\tau}} + O_p\left(\frac{\kappa}{\lambda\varpi}\left(\frac{\sqrt{\log n}}{n^2 h^{p+1}} + h^r\right)\right).$$

Proof. Recall in Lemma 1, we proved that

$$\mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{Z}) = \boldsymbol{U}\boldsymbol{\theta}^* + \overline{\boldsymbol{Z}}\boldsymbol{\eta}^*.$$

Because $\boldsymbol{D}\boldsymbol{U} = \mathbf{0}$ in (S1.6), base on the expression of $\widehat{\boldsymbol{\eta}}$, we have

$$\begin{aligned}
\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* &= (\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\breve{\boldsymbol{Y}} - \boldsymbol{\eta}^* \\
&= (\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\breve{\boldsymbol{Y}} - (\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}}\boldsymbol{\eta}^* - (\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\boldsymbol{U}\boldsymbol{\theta}^* \\
&= (\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}[\breve{\boldsymbol{Y}} - (\boldsymbol{U}\boldsymbol{\theta}^* + \overline{\boldsymbol{Z}}\boldsymbol{\eta}^*)] \\
&= (\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}[\breve{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{Z})] \\
&= (\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}[\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Y}} + \overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{Z})] \\
&= (\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}[\overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{Z})] + \underbrace{(\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}(\breve{\boldsymbol{Y}} - \overline{\boldsymbol{Y}})}_{(I)}. \quad \text{(S4.53)}
\end{aligned}$$

We consider (I) in (S4.53). We first define

$$\boldsymbol{A} = (\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D} = (\boldsymbol{A}_{01}, \ldots, \boldsymbol{A}_{(n-1)n}) \in \mathbb{R}^{p\times N}, \qquad \text{(S4.54)}$$

where $\boldsymbol{A}_{ij}$ is a p-dimensional column vector. Under Condition 4, we have

$$\begin{aligned}
\max_{0\le i<j\le n} \|\boldsymbol{A}_{ij}\|_2 &\le \max_{0\le i<j\le j} \|(\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\|_2 \|\overline{\boldsymbol{Z}}^\top \boldsymbol{D}_{ij}\|_2 \\
&\le \max_{0\le i<j\le j} \|(\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\|_2 \sqrt{\sum_{k\in[p]}\left(\sum_{s=0}^{n-1}\sum_{l=1,l<s}^{n} \overline{Z}_{slk} D_{ij,sl}\right)^2} \\
&\le \|(\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\|_2 \sqrt{p}\kappa \|\boldsymbol{D}\|_\infty
\end{aligned}$$

$$\leq \quad O\left(\frac{1}{n^2\lambda}\right)\sqrt{p}\kappa O(1) = O\left(\frac{\kappa}{n^2\lambda}\right), \tag{S4.55}$$

where the last inequality is due to

$$\|\boldsymbol{D}\|_\infty = \|\boldsymbol{I}_{N\times N} - \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top\|_\infty = \frac{3(n-1)}{n+1} = O(1). \tag{S4.56}$$

Note that $\boldsymbol{D}$ is fixed. We have

$$\mathbb{E}(\boldsymbol{A}|\boldsymbol{Z}) = \boldsymbol{A}. \tag{S4.57}$$

According to the definitions of $\breve{\boldsymbol{Y}}$ in (S3.21) and $\overline{Y}_{ij}$ in (S3.19), we have

$$\begin{aligned}
(I) &= \sum_{0\leq i<j\leq n} \boldsymbol{A}_{ij}\left(\breve{Y}_{ij} - \overline{Y}_{ij}\right) \\
&= \frac{1}{T}\sum_{0\leq i<j\leq n}\sum_{t\in[T]} \boldsymbol{A}_{ij}\left(\widehat{Y}_{ijt} - Y_{ijt}\right) \\
&= \frac{1}{T}\sum_{0\leq i<j\leq n}\sum_{t\in[T]} \boldsymbol{A}_{ij}[a_{ijt} - \mathbb{I}(X_{ijt,0} > 0)]\left(\frac{1}{\hat{f}(x_{ijt,0}|\boldsymbol{z}_{ijt})} - \frac{1}{f(x_{ijt,0}|\boldsymbol{z}_{ijt})}\right) \\
&= \frac{1}{T}\sum_{0\leq i<j\leq n}\sum_{t\in[T]} \boldsymbol{A}_{ij}Y_{ijt}\left(\frac{f(x_{ijt,0}|\boldsymbol{z}_{ijt})}{\hat{f}(x_{ijt,0}|\boldsymbol{z}_{ijt})} - 1\right) \\
&= \boldsymbol{A}[\mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{Z}) - \mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{X}_0, \boldsymbol{Z})] + O_p\left(\frac{\kappa}{n^2\lambda\varpi}\left(\frac{\sqrt{\log n}}{h^{p+1}} + n^2h^r\right)\right) \\
&= \boldsymbol{A}[\mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{Z}) - \mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{X}_0, \boldsymbol{Z})] + O_p\left(\frac{\kappa}{\lambda\varpi}\left(\frac{\sqrt{\log n}}{n^2h^{p+1}} + h^r\right)\right),
\end{aligned} \tag{S4.58}$$

where the second to last inequality is due to Theorem 1(b) of Andrews (1995, page 9) (reproduced in Lemma S11.2) for nonparametric density estimators. Combining the above equation with (S4.53), we have

$$\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = \boldsymbol{A}[\overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{Z})] + \boldsymbol{A}[\mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{Z}) - \mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{X}_0, \boldsymbol{Z})] + O_p\left(\frac{\kappa}{\lambda\varpi}\left(\frac{\sqrt{\log n}}{n^2h^{p+1}} + h^r\right)\right)$$

$$= \boldsymbol{A}\underbrace{[\overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{X}_0, \boldsymbol{Z})]}_{\overline{\boldsymbol{\tau}}} + O_p\left(\frac{\kappa}{\lambda\varpi}\left(\frac{\sqrt{\log n}}{n^2 h^{p+1}} + h^r\right)\right).$$

This completes the proof. □

We are now ready to prove Theorem 2.

Proof of Theorem 2. By Lemma S4.1, if

$$\frac{\kappa}{\lambda\varpi}\left(\frac{\sqrt{\log n}}{nh^{p+1}} + nh^r\right) = o(1),$$

then

$$n(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) = n(\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{\tau}} := \sum_{0\le i<j\le n} n\boldsymbol{A}_{ij}\overline{\tau}_{ij}.$$

To show

$$\sum_{0\le i<j\le n} n\boldsymbol{A}_{ij}\overline{\tau}_{ij} \rightsquigarrow N_p\Big(\mathbf{0}, n^2\mathbb{E}[(\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}]\mathbb{E}(\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\boldsymbol{\Sigma}_\tau \boldsymbol{D}\overline{\boldsymbol{Z}})\mathbb{E}[(\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}]\Big), \tag{S4.59}$$

it is sufficient to verify the Linderberg condition

$$\sum_{0\le i<j\le n} \mathbb{E}\Big[\|n\boldsymbol{A}_{ij}\overline{\tau}_{ij}\|_2^2 \mathbb{I}(\|n\boldsymbol{A}_{ij}\overline{\tau}_{ij}\|_2 > \varepsilon)\Big] \rightsquigarrow 0, \tag{S4.60}$$

where $\rightsquigarrow$ denotes "convergence in distribution."

Because

$$\begin{aligned}
\boldsymbol{D}\boldsymbol{D}^\top &= (\boldsymbol{I}_{N\times N} - \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top)(\boldsymbol{I}_{N\times N} - \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top)^\top \\
&= \boldsymbol{I}_{N\times N} - \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top - \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top + \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top \\
&= \boldsymbol{I}_{N\times N} - \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top - \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top + \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{V}^{-1}\boldsymbol{U}^\top \\
&= \boldsymbol{I}_{N\times N} - \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top = \boldsymbol{D},
\end{aligned}$$

the matrix $\boldsymbol{A}$ satisfies

$$\boldsymbol{A}\boldsymbol{A}^\top = (\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\boldsymbol{D}\overline{\boldsymbol{Z}}(\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1} = (\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}. \quad \text{(S4.61)}$$

Under Condition 4, it follows that

$$\sum_{0\le i<j\le n} \|n\boldsymbol{A}_{ij}\|_2^2 = \text{trace}(n^2\boldsymbol{A}\boldsymbol{A}^\top) \le pn^2\lambda_{\max}(\overline{\boldsymbol{Z}}^\top \boldsymbol{D}\overline{\boldsymbol{Z}})^{-1} = \frac{pn^2}{N\lambda} = O\left(1\right). \quad \text{(S4.62)}$$

Note that

$$\mathbb{E}(\overline{\tau}_{ij}^2) = \mathbb{E}^2(\overline{\tau}_{ij}) + Var(\overline{\tau}_{ij}) = \sigma_{\tau,ij}^2 < \infty.$$

Therefore, to verify (S4.60), it is sufficient to prove

$$\max_{i,j} \|n\boldsymbol{A}_{ij}\|_2 \to 0; \quad \text{(S4.63)}$$

see Van der Vaart (2000, page 21).

We now show (S4.63). Under Conditions 4, 5, and (S4.55), as $n\to\infty$, we have

$$\max_{0\le i<j\le n} \|n\boldsymbol{A}_{ij}\|_2 \le \max_{0\le i<j\le j} n\|\boldsymbol{A}_{ij}\|_2 \le nO\left(\frac{\kappa}{n^2\lambda}\right) = O\left(\frac{\kappa}{n\lambda}\right) \to 0.$$

This verifies (S4.63).

By (S4.61), we have

$$\begin{aligned}
\sum_{0\le i<j\le n} Cov(n\boldsymbol{A}_{ij}\overline{\tau}_{ij}) &= \sum_{0\le i<j\le n} \mathbb{E}[Cov(n\boldsymbol{A}_{ij}\overline{\tau}_{ij})|\{\overline{\boldsymbol{Z}}_{ij}\}_{i<j}] \\
&= n^2\mathbb{E}\left[\sum_{0\le i<j\le n} Cov(\boldsymbol{A}_{ij}\overline{\tau}_{ij})|\{\overline{\boldsymbol{Z}}_{ij}\}_{i<j}\right] \\
&= n^2\mathbb{E}\left[\sum_{0\le i<j\le n} \sigma_{\tau,ij}^2\boldsymbol{A}_{ij}\boldsymbol{A}_{ij}^\top|\{\overline{\boldsymbol{Z}}_{ij}\}_{i<j}\right]
\end{aligned}$$

$$= n^2\mathbb{E}(\boldsymbol{A}\boldsymbol{\Sigma}_\tau\boldsymbol{A}^\top|\{\overline{\boldsymbol{Z}}_{ij}\}_{i<j})$$
$$= n^2\mathbb{E}[(\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}]\mathbb{E}(\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\boldsymbol{\Sigma}_\tau\boldsymbol{D}\overline{\boldsymbol{Z}})\mathbb{E}[(\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}].$$

It follows from (S4.59) that we have

$$\sum_{0\le i<j\le n} n\boldsymbol{A}_{ij}\overline{\tau}_{ij} \rightsquigarrow N_p\left(\boldsymbol{0}, n^2\mathbb{E}[(\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}]\mathbb{E}(\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\boldsymbol{\Sigma}_\tau\boldsymbol{D}\overline{\boldsymbol{Z}})\mathbb{E}[(\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}]\right),$$

where $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a p-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Note that

$$n(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) = \sum_{0\le i<j\le n} n\boldsymbol{A}_{ij}\overline{\tau}_{ij}.$$

It follows that

$$n(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \rightsquigarrow N_p\left(\boldsymbol{0}, n^2\mathbb{E}[(\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}]\mathbb{E}(\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\boldsymbol{\Sigma}_\tau\boldsymbol{D}\overline{\boldsymbol{Z}})\mathbb{E}[(\overline{\boldsymbol{Z}}^\top\boldsymbol{D}\overline{\boldsymbol{Z}})^{-1}]\right).$$

This completes the proof of Theorem 2. □

S5 Proof for Theorem 3

Recall $\overline{\boldsymbol{\xi}} = \overline{\boldsymbol{Y}} - \mathbb{E}(\overline{\boldsymbol{Y}}|\boldsymbol{Z}) = \overline{\boldsymbol{Y}} - \overline{\boldsymbol{Z}}\boldsymbol{\eta}^* - \boldsymbol{U}\boldsymbol{\theta}^*$ in (3.3). For convenience, we write $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\widehat{\boldsymbol{\eta}}}$ and $\boldsymbol{\theta}^* = \boldsymbol{\theta}^*_{\boldsymbol{\eta}^*}$.

Lemma S5.1. *Under Conditions 1–7, we have*

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = \boldsymbol{V}^{-1}\boldsymbol{U}^\top\overline{\boldsymbol{\xi}} + \boldsymbol{R},$$

where

$$\boldsymbol{R} := \boldsymbol{V}^{-1}\boldsymbol{U}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^* - \widehat{\boldsymbol{\eta}}) + \boldsymbol{V}^{-1}\boldsymbol{U}^\top(\check{\boldsymbol{Y}} - \overline{\boldsymbol{Y}}) \qquad \text{(S5.64)}$$

satisfying

$$\|\boldsymbol{R}\|_\infty = O_p\left(\frac{1}{\varpi^2}\left(1+\frac{\kappa^2}{\lambda}\right)\left(\frac{\sqrt{\log n}}{n}+\frac{\sqrt{\log n}}{nh^{p+1}}+h^r\right)\right).$$

Proof. Base on the expression of $\widehat{\boldsymbol{\theta}}$, we have

$$\begin{aligned}\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^* &= \boldsymbol{V}^{-1}\boldsymbol{U}^\top(\breve{\boldsymbol{Y}}-\overline{\boldsymbol{Z}}\widehat{\boldsymbol{\eta}}-\boldsymbol{U}\boldsymbol{\theta}^*+\overline{\boldsymbol{Y}}-\overline{\boldsymbol{Y}}+\overline{\boldsymbol{Z}}\boldsymbol{\eta}^*-\overline{\boldsymbol{Z}}\boldsymbol{\eta}^*)\\ &= \boldsymbol{V}^{-1}\boldsymbol{U}^\top(\underbrace{\overline{\boldsymbol{Y}}-\overline{\boldsymbol{Z}}\boldsymbol{\eta}^*-\boldsymbol{U}\boldsymbol{\theta}^*}_{\overline{\boldsymbol{\xi}}})+\underbrace{\boldsymbol{V}^{-1}\boldsymbol{U}^\top\overline{\boldsymbol{Z}}(\boldsymbol{\eta}^*-\widehat{\boldsymbol{\eta}})}_{(I)}+\underbrace{\boldsymbol{V}^{-1}\boldsymbol{U}^\top(\breve{\boldsymbol{Y}}-\overline{\boldsymbol{Y}})}_{(II)}. \quad \text{(S5.65)}\end{aligned}$$

By (S3.51), we have

$$(I) = O\left(\frac{\kappa^2}{\varpi^2\lambda}\left(\frac{\sqrt{\log n}}{n}+\frac{\sqrt{\log n}}{nh^{p+1}}+h^r\right)\right). \quad \text{(S5.66)}$$

By (S3.32), we have

$$(II) = O_p\left(\frac{1}{\varpi^2}\left(\frac{\sqrt{\log n}}{nh^{p+1}}+h^r\right)\right). \quad \text{(S5.67)}$$

Combining (S5.65), (S5.66) and (S5.67) yields

$$\begin{aligned}\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^* &= \boldsymbol{V}^{-1}\boldsymbol{U}^\top\overline{\boldsymbol{\xi}}+O\left(\frac{\kappa^2}{\varpi^2\lambda}\left(\frac{\sqrt{\log n}}{n}+\frac{\sqrt{\log n}}{nh^{p+1}}+h^r\right)\right)+O_p\left(\frac{1}{\varpi^2}\left(\frac{\sqrt{\log n}}{nh^{p+1}}+h^r\right)\right)\\ &= \boldsymbol{V}^{-1}\boldsymbol{U}^\top\overline{\boldsymbol{\xi}}+O_p\left(\frac{1}{\varpi^2}\left(1+\frac{\kappa^2}{\lambda}\right)\left(\frac{\sqrt{\log n}}{n}+\frac{\sqrt{\log n}}{nh^{p+1}}+h^r\right)\right).\end{aligned}$$

This completes the proof. □

We are now ready to prove Theorem 3.

Proof of Theorem 3. By Lemma S5.1, if (3.7) holds, then

$$\sqrt{n}\boldsymbol{c}^\top(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) = \sqrt{n}\boldsymbol{c}^\top\boldsymbol{V}^{-1}\boldsymbol{U}^\top\overline{\boldsymbol{\xi}}+\sqrt{n}\boldsymbol{c}^\top\boldsymbol{R},$$

where $\boldsymbol{R} \in \mathbb{R}^n$ is defined in (S5.64) and satisfies

$$\begin{aligned} \|\sqrt{n}\boldsymbol{c}^\top \boldsymbol{R}\|_\infty &= \sqrt{n}\sum_{i=1}^n |c_i| O_p\left(\frac{1}{\varpi^2}\left(1+\frac{\kappa^2}{\lambda}\right)\left(\frac{\sqrt{\log n}}{n}+\frac{\sqrt{\log n}}{nh^{p+1}}+h^r\right)\right) \\ &= O_p\left(\frac{1}{\varpi^2}\left(1+\frac{\kappa^2}{\lambda}\right)\left(\sqrt{\frac{\log n}{n}}+\sqrt{\frac{\log n}{nh^{2p+2}}}+\sqrt{n}h^r\right)\right). \quad \text{(S5.68)} \end{aligned}$$

It is sufficient to show

$$\sqrt{n}\boldsymbol{c}^\top \boldsymbol{V}^{-1}\boldsymbol{U}^\top \overline{\boldsymbol{\xi}} \rightsquigarrow N(0, n\boldsymbol{C}^\top \boldsymbol{\Sigma}_\xi \boldsymbol{C}). \quad \text{(S5.69)}$$

We first show (S5.69). It is sufficient to verify the Lindeberg condition

$$\sum_{0\le i<j\le n} \mathbb{E}\Big[|\sqrt{n}C_{ij}\overline{\xi}_{ij}|^2 \mathbb{I}(|\sqrt{n}C_{ij}\overline{\xi}_{ij}| > \varepsilon)\Big] \to 0 \quad \text{(S5.70)}$$

for any given $\varepsilon > 0$. Write

$$\sqrt{n}\boldsymbol{c}^\top \boldsymbol{V}^{-1}\boldsymbol{U}^\top \overline{\boldsymbol{\xi}} = \sum_{0\le i<j\le n} \sqrt{n}C_{ij}\overline{\xi}_{ij},$$

where we define

$$\boldsymbol{C} := (\boldsymbol{c}^\top \boldsymbol{V}^{-1}\boldsymbol{U}^\top)^\top = (C_{01}, \ldots, C_{(n-1)n})^\top \in \mathbb{R}^N.$$

By the explicit expression of $\boldsymbol{V}^{-1}$ in (2.5) and $\sum_{i=1}^n c_i^2 \le \sum_{i=1}^n |c_i| < \infty$, we have

$$\begin{aligned} \boldsymbol{C}^\top \boldsymbol{C} &= \boldsymbol{c}^\top \boldsymbol{V}^{-1}\boldsymbol{U}^\top \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{c} = \boldsymbol{c}^\top \boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{V}^{-1}\boldsymbol{c} = \boldsymbol{c}^\top \boldsymbol{V}^{-1}\boldsymbol{c} \\ &= \frac{1}{n+1}\left(\sum_{i=1}^n |c_i|\right)^2 + \frac{1}{n+1}\sum_{i=1}^n c_i^2 \le \frac{2}{n+1}\left(\sum_{i=1}^n |c_i|\right)^2. \quad \text{(S5.71)} \end{aligned}$$

Then, we have

$$\sum_{0\le i<j\le n} |\sqrt{n}C_{ij}|^2 = n\boldsymbol{C}^\top \boldsymbol{C} \le \frac{2n}{n+1}\left(\sum_{i=1}^n |c_i|\right)^2 < \infty. \tag{S5.72}$$

By the explicit expression of $\boldsymbol{V}^{-1}$ in (2.5), we have

$$\begin{aligned}\max_{0\le i<j\le j} |\sqrt{n}C_{ij}| &= \sqrt{n}\max\left\{\frac{1}{n}\left(\max_{i\in[n]} c_i + \sum_{i=1}^n |c_i|\right), \frac{1}{n}\max_{i\in[n]}(c_i - c_j)\right\} \\ &= O\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n |c_i|\right) \to 0 \text{ as } n\to\infty.\end{aligned}$$

This, together with (S5.72), verifies (S5.70); see Van der Vaart (2000, page 21).

By (S5.71), we have

$$\sum_{0\le i<j\le n} \mathrm{Var}(\sqrt{n}C_{ij}\bar{\xi}_{ij}) = n\sum_{0\le i<j\le n} \sigma^2_{\xi,ij}C^2_{ij} = n\boldsymbol{C}^\top \boldsymbol{\Sigma}_\xi \boldsymbol{C}.$$

By the Lindeberg-Feller Central Limit Theorem (Lindeberg, 1922; Feller, 1935), we have

$$\sqrt{n}\boldsymbol{c}^\top(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = \sum_{0\le i<j\le n} \sqrt{n}C_{ij}\bar{\xi}_{ij} \rightsquigarrow N(0, n\boldsymbol{C}^\top \boldsymbol{\Sigma}_\xi \boldsymbol{C}).$$

This shows (S5.69). This completes the proof. □

S6 Additional simulation results under Cases (i)–(iii)

In this section, we report additional simulation results under Cases (i)–(iii): (i) the latent random error is generated from the standard normal distribution $N(0,1)$; (ii) the latent random error is drawn from the logistic distribution $\varepsilon_{ijt} \sim$ Logistic$(0, \sqrt{3}/\pi)$; (iii) $\varepsilon_{ijt} \sim$ Mix-Norm, where Mix-Norm means that a random variable is drawn from $N(-0.3, 0.91)$ with probability 0.75 and $N(0.9, 0.19)$ with probability 0.25. All other settings remain the same as those in Section 4.1 of the main paper. The results for Cases (ii) and (iii) are presented in Tables 3 and 4. To assess the normality of the parameter estimators, we draw QQ plots (quantile-quantile plots) of the kernel-based least squares estimators. We report the QQ plots of η_1 and θ_1 under these three cases in Figures 1–6.

Table 3: Simulation results when the noise $\varepsilon_{ijt} \sim$ Logistic$(0, \sqrt{3}/\pi)$.

		$T=1$			$T=3$		
n		Bias	SD	CP	Bias	SD	CP
50	θ_1	0.0166	0.381	0.957	0.0090	0.211	0.957
	θ_{12}	-0.0035	0.371	0.958	0.0018	0.216	0.954
	θ_{25}	0.0098	0.360	0.957	-0.0040	0.216	0.961
	θ_{37}	-0.0123	0.376	0.957	-0.0021	0.219	0.956
	θ_{50}	-0.0071	0.371	0.958	0.0012	0.210	0.961
	η_1	-0.0041	0.049	0.948	-0.0025	0.027	0.947
	η_2	0.0026	0.048	0.948	0.0014	0.028	0.946
100	θ_1	-0.0092	0.282	0.954	-0.0028	0.170	0.954
	θ_{25}	-0.0146	0.274	0.956	0.0025	0.173	0.962
	θ_{50}	-0.0099	0.281	0.960	-0.0041	0.168	0.952
	θ_{75}	-0.0213	0.292	0.961	-0.0035	0.191	0.971
	θ_{100}	-0.0160	0.272	0.956	-0.0186	0.168	0.959
	η_1	-0.0023	0.024	0.949	-0.0024	0.015	0.955
	η_2	0.0030	0.025	0.940	0.0008	0.015	0.953

Table 4: Simulation results when the noise $\varepsilon_{ijt} \sim$ Mix-Norm.

		$T=1$			$T=3$		
n		Bias	SD	CP	Bias	SD	CP
50	θ_1	−0.0179	0.370	0.942	0.0060	0.212	0.950
	θ_{12}	0.0034	0.374	0.951	−0.0024	0.203	0.949
	θ_{25}	0.0038	0.359	0.950	0.0015	0.206	0.948
	θ_{37}	−0.0079	0.344	0.956	0.0142	0.207	0.955
	θ_{50}	−0.0044	0.349	0.947	0.0053	0.211	0.944
	η_1	−0.0048	0.046	0.949	−0.0057	0.027	0.946
	η_2	0.0064	0.046	0.944	0.0046	0.027	0.946
100	θ_1	0.0115	0.262	0.955	−0.0027	0.159	0.955
	θ_{25}	−0.0037	0.264	0.952	0.0074	0.155	0.951
	θ_{50}	0.0065	0.263	0.952	−0.0014	0.152	0.944
	θ_{75}	−0.0027	0.273	0.958	0.0034	0.152	0.950
	θ_{100}	0.0061	0.250	0.952	−0.0018	0.151	0.952
	η_1	−0.0044	0.024	0.938	−0.0041	0.014	0.945
	η_2	0.0050	0.024	0.944	0.0041	0.014	0.938

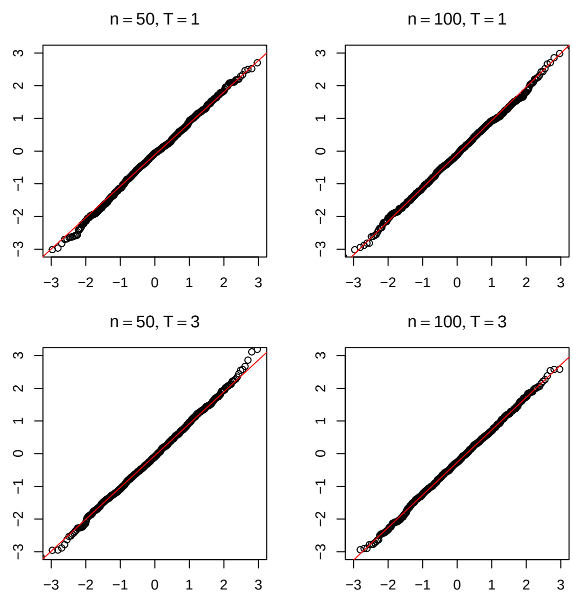

Figure 1: The QQ plots of η_1 when $\varepsilon_{ijt} \sim N(0,1)$.

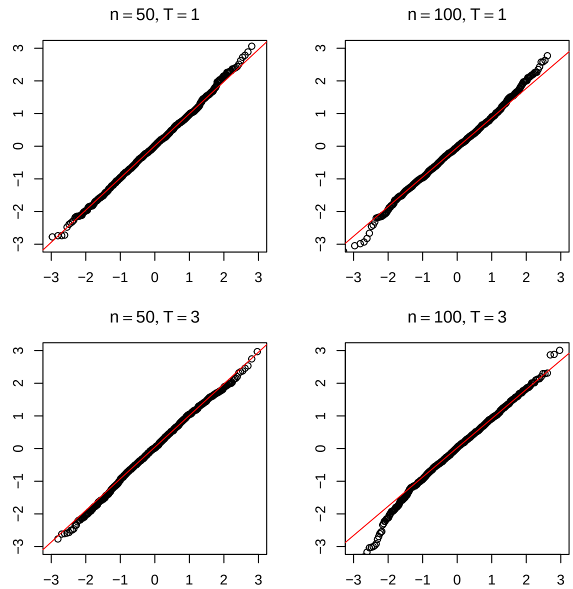

Figure 2: The QQ plots of θ_1 when $\varepsilon_{ijt} \sim N(0,1)$.

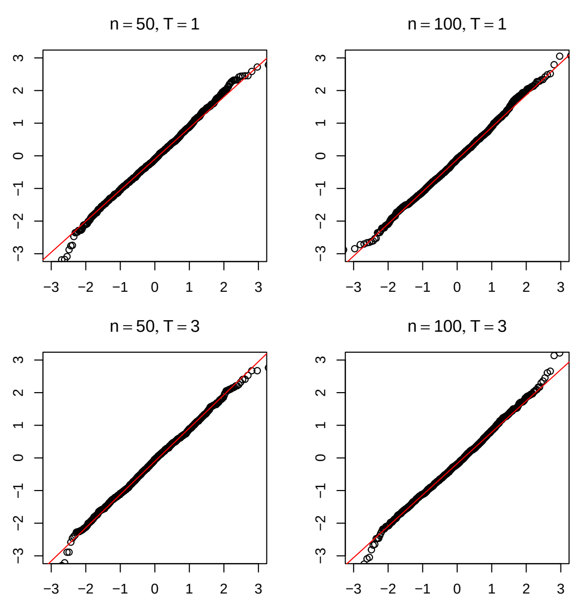

Figure 3: The QQ plots of η_1 when $\varepsilon_{ijt} \sim \text{Logistic}(0, \sqrt{3}/\pi)$.

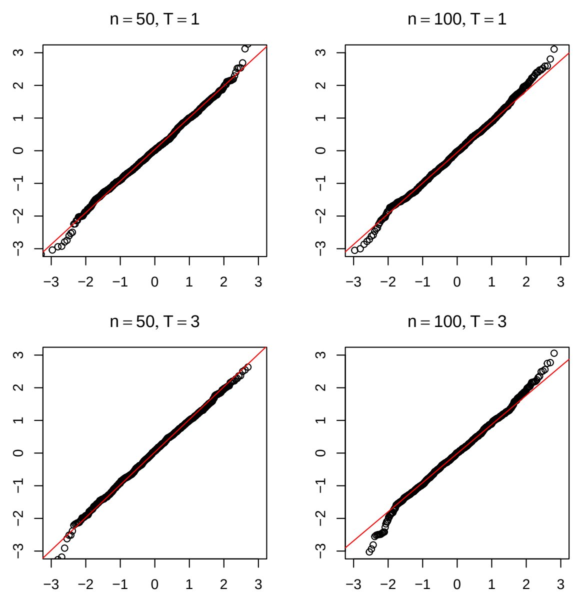

Figure 4: The QQ plots of θ_1 when $\varepsilon_{ijt} \sim \text{Logistic}(0, \sqrt{3}/\pi)$.

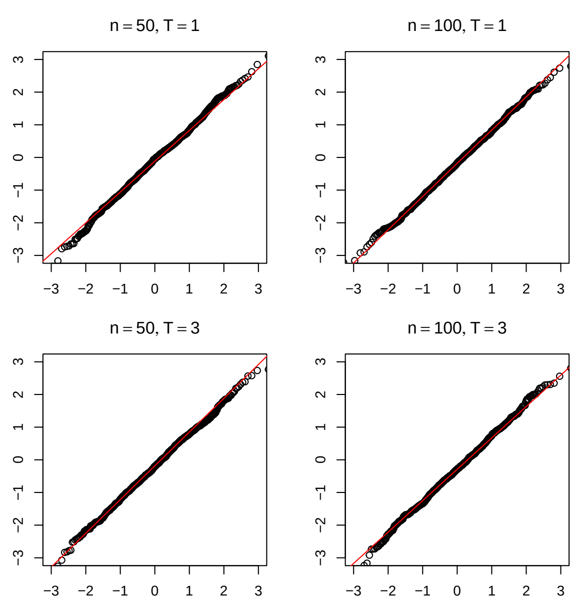

Figure 5: The QQ plots of η_1 when $\varepsilon_{ijt} \sim$ Mix-Norm.

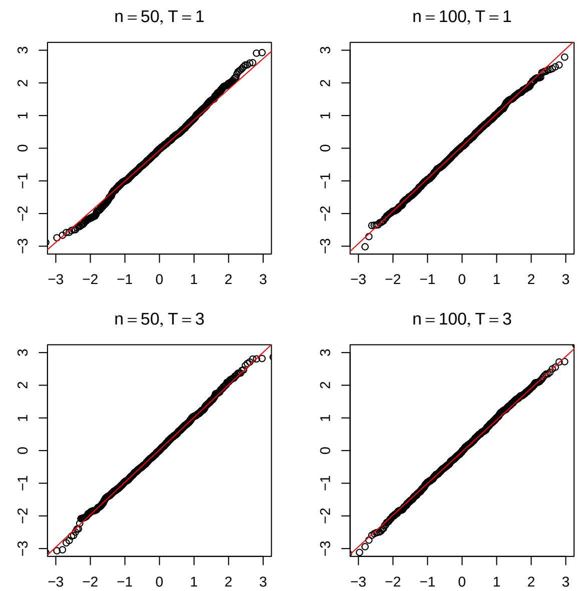

Figure 6: The QQ plots of θ_1 when $\varepsilon_{ijt} \sim$ Mix-Norm.

S7 Comparison with the covariate-adjusted Bradley–Terry model

We conduct simulation studies for a comprehensive comparison between the semiparametric model with the parametric model (i.e., the covariate-adjusted Bradley–Terry model (Yan, 2025)). Specifically, we consider the covariate-adjusted Bradley–Terry model as follows:

$$\mathbb{P}(A_{ijt} = 1 | X_{ijt,0}, \boldsymbol{Z}_{ijt}) = \frac{\exp(\theta_i - \theta_j + X_{ijt,0} + \boldsymbol{Z}_{ijt}^\top \boldsymbol{\eta})}{1 + \exp(\theta_i - \theta_j + X_{ijt,0} + \boldsymbol{Z}_{ijt}^\top \boldsymbol{\eta})}, \tag{S7.73}$$

where θ_0 is set to be 0 and the coefficient of $X_{ijt,0}$ is fixed at 1 for a fair comparison. Note that Fan et al. (2024) assumed that the covariates of each item enter into the Bradley–Terry model with an additive form, where item

i beats item j with probability $e^{\mu_{ij}}/(1+e^{\mu_{ij}})$ with $\mu_{ij} = \theta_i - \theta_i + X_i^\top \gamma - X_j^\top \gamma$. Here, X_i denotes the covariate of item i. It can be viewed as a special case of the model in Yan (2025), where the pairwise covariates associated with each comparison are used. We use the maximum likelihood to estimate the unknown parameters when fitting the above model to the simulated data.

When model (S7.73) is correctly specified, the noise in model (2.1) follows a standard logistic distribution, denoted as Logistic$(0,1)$. We consider four scenarios for the noise ϵ_{ijt}: (1) $\epsilon \sim N(0,1)$, (2) $\epsilon \sim$ Logistic$(0, \sqrt{3}/\pi)$, (3) $\epsilon \sim$ Mix-Norm and (4) $\epsilon \sim$ Logistic$(0,1)$, where the density of a logistic distribution is given in (4.2) and Mix-Norm denotes a mixture of normal distributions that is the same as case (iii) in the simulation. We set the merit parameters θ_i^* as $\theta_i^* = 0.2i \log n/n$ and the covariate parameter $\boldsymbol{\eta}^* = (-0.5, 0.5)^\top$. Each simulation was repeatedly 1000 times. We compare the biases of the estimator obtained from the semiparametric model with that of the MLE in the covariate-adjusted Bradley–Terry model. The results are presented in Table 5.

Table 5: The comparison results of the bias of the estimators obtained by our method and the MLE method with $n = 100$ and $T = 1$.

	$N(0,1)$		Logistic$(0, \sqrt{3}/\pi)$		Mix-Norm		Logistic$(0,1)$	
	Our method	MLE	Our method	MLE	Our method	MLE	Our method	MLE
θ_1	−0.0025	0.0009	−0.0093	−0.0024	0.0115	0.0083	−0.0054	−0.0090
θ_{25}	0.0025	0.0880	−0.0099	0.0972	0.0065	0.1145	−0.0448	−0.0123
θ_{50}	−0.0081	0.1672	−0.0160	0.2059	0.0061	0.2159	−0.0699	0.0122
η_1	−0.0025	0.0009	−0.0023	−0.0005	−0.0044	−0.0006	−0.0259	0.0000
η_2	0.0037	0.0001	0.0030	0.0014	0.0050	0.0001	0.0281	0.0014

From Table 5, we observe that when the covariate-adjusted Bradley–Terry model is correctly specified (i.e., the noise is distributed according to a standard logistic distribution), in most cases, the MLE has smaller biases than the estimator obtained from the semiparametric model. This is expected since parametric models use more information and are more efficient than semiparametric models when they are correctly specified. When the covariate-adjusted Bradley–Terry model is not correctly specified (i.e., the noise is not generated from the standard logistic distribution), the semiparametric estimator has smaller biases than the MLE. For instance, when $\epsilon \sim N(0,1)$ or $\epsilon \sim \text{Logistic}(0, \sqrt{3/\pi})$, the biases of the semiparametric estimator are close to zero, whereas the bias of the MLE for parameter θ_{50} in the misspecified model (S7.73) exceeds 0.16. Additionally, we note that the MLE for the covariate parameters remains unbiased, even when the noise distribution is misspecified. This phenomenon may arise because n^2 samples are used to estimate the fixed-dimensional covariate parameters, thereby effectively attenuating the influence of the misspecified noise distribution on the MLE of the covariate parameters.

S8 Evaluating robust of the choices of bandwidths and kernel functions, sparse paired comparisons

We have conducted simulation studies to evaluate the sensitivity of the proposed method to the bandwidth selection and the choice of kernel function. First, we evaluate the performance of the proposed method under different bandwidths. Specifically, we set $h = c(NT)^{-1/7}$ with $c = (2.4, 2.6, 2.8, 3.0)$. We fix $T = 1$ and $n = 50$. Other settings are the same as outlined in Section 4.1. The results presented in Table 6 indicate. We observe that the bias and the standard deviation of the estimators are comparable under different bandwidths, and the coverage frequencies are all close to the 95% target level.

For the selection of kernel functions, we consider comparing the Gaussian kernel $\mathcal{K}(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$ and the uniform kernel $\mathcal{K}(x) = \frac{1}{2}\mathbb{I}(|x| \leq 1)$

with the quartic kernel function in (4.1). The bandwidth h is determined according to the proposed procedure. The results are presented in Table 7. The results indicate that the performance of the proposed method remains comparable under different kernel functions.

Table 6: Comparison results under different bandwidths, with $n = 50$ and $T = 1$

		$N(0,1)$			Logistic$(0, \sqrt{3}/\pi)$			Mix-Norm		
		Bias	SD	CP	Bias	SD	CP	Bias	SD	CP
$c = 2.4$	θ_1	0.0044	0.343	0.944	0.0146	0.368	0.957	−0.0181	0.362	0.942
	θ_{12}	−0.0072	0.345	0.947	−0.0093	0.358	0.952	−0.0025	0.365	0.952
	θ_{25}	−0.0221	0.343	0.950	−0.0029	0.347	0.952	−0.0078	0.350	0.947
	θ_{37}	−0.0232	0.353	0.949	−0.0302	0.364	0.953	−0.0258	0.337	0.955
	θ_{50}	−0.0401	0.359	0.957	−0.0308	0.357	0.955	−0.0277	0.340	0.946
	η_1	0.0223	0.045	0.921	0.0230	0.046	0.918	0.0228	0.044	0.915
	η_2	−0.0248	0.044	0.916	−0.0244	0.046	0.909	−0.0214	0.044	0.923
$c = 2.6$	θ_1	0.0048	0.346	0.943	0.0157	0.375	0.958	−0.0180	0.366	0.942
	θ_{12}	−0.0044	0.350	0.945	−0.0061	0.365	0.956	0.0007	0.370	0.951
	θ_{25}	−0.0164	0.347	0.950	0.0040	0.354	0.954	−0.0017	0.355	0.949
	θ_{37}	−0.0144	0.358	0.951	−0.0206	0.371	0.957	−0.0163	0.341	0.956
	θ_{50}	−0.0282	0.363	0.964	−0.0181	0.365	0.958	−0.0153	0.345	0.947
	η_1	0.0078	0.046	0.948	0.0088	0.048	0.943	0.0083	0.045	0.935
	η_2	−0.0104	0.046	0.951	−0.0102	0.047	0.936	−0.0068	0.045	0.944
$c = 2.8$	θ_1	0.0054	0.348	0.939	0.0166	0.381	0.957	−0.0179	0.370	0.942
	θ_{12}	−0.0019	0.353	0.948	−0.0035	0.371	0.958	0.0034	0.374	0.951
	θ_{25}	−0.0113	0.350	0.948	0.0098	0.360	0.957	0.0038	0.359	0.950
	θ_{37}	−0.0066	0.361	0.951	−0.0123	0.376	0.957	−0.0079	0.344	0.956
	θ_{50}	−0.0178	0.367	0.964	−0.0071	0.371	0.958	−0.0044	0.349	0.947
	η_1	−0.0054	0.047	0.947	−0.0041	0.049	0.948	−0.0048	0.046	0.949
	η_2	0.0026	0.047	0.949	0.0026	0.048	0.948	0.0064	0.046	0.944
$c = 3.0$	θ_1	0.0059	0.351	0.942	0.0173	0.387	0.958	−0.0179	0.373	0.942
	θ_{12}	0.0004	0.356	0.951	−0.0012	0.376	0.960	0.0058	0.377	0.953
	θ_{25}	−0.0066	0.353	0.949	0.0150	0.364	0.958	0.0088	0.362	0.952
	θ_{37}	0.0005	0.364	0.949	−0.0049	0.381	0.962	−0.0003	0.346	0.955
	θ_{50}	−0.0084	0.369	0.963	0.0026	0.376	0.960	0.0056	0.352	0.945
	η_1	−0.0175	0.048	0.933	−0.0160	0.050	0.938	−0.0170	0.047	0.937
	η_2	0.0146	0.048	0.939	0.0146	0.049	0.939	0.0187	0.047	0.927

We also conducted simulations to evaluate the performance of the proposed estimator under sparse comparison graphs. Following the setting of Han et al. (2024), we simulate sparsity by modeling $n_{ij} \sim \text{Bernoulli}(T, p_{ij,n})$, where $p_{ij,n}$ is generated from the uniform distribution $U(p_n, q_n)$ with $p_n = 1/\sqrt{n}$ and $q_n = p_n \log n$. All other setting are the same as before. Tables 8, 9 and 10 report the simulation results under Cases (i), (ii) and (iii). From these three tables, we can see that the biases of the estimators are very small and their simulated coverage frequencies are close to the target level.

Table 7: Comparison results for three kernel functions when $n = 50$ and $T = 1$.

		Gaussian Kernel			Quartic Kernel			Uniform Kernel		
ε_{ijt}		Bias	SD	CP	Bias	SD	CP	Bias	SD	CP
$N(0,1)$	θ_1	0.0054	0.343	0.942	0.0054	0.348	0.939	0.0077	0.376	0.954
	θ_{12}	−0.0040	0.348	0.948	−0.0019	0.353	0.948	0.0009	0.381	0.950
	θ_{25}	−0.0153	0.344	0.950	−0.0113	0.350	0.948	−0.0089	0.384	0.961
	θ_{37}	−0.0127	0.355	0.950	−0.0066	0.361	0.951	−0.0048	0.388	0.955
	θ_{50}	−0.0262	0.360	0.963	−0.0178	0.367	0.964	−0.0113	0.395	0.959
	η_1	0.0019	0.046	0.952	−0.0054	0.047	0.947	−0.0017	0.052	0.945
	η_2	−0.0046	0.045	0.958	0.0026	0.047	0.949	−0.0017	0.052	0.948
Logistic$(0, \sqrt{3}/\pi)$	θ_1	0.0155	0.370	0.957	0.0166	0.381	0.957	0.0189	0.419	0.961
	θ_{12}	−0.0064	0.359	0.953	−0.0035	0.371	0.958	0.0021	0.404	0.962
	θ_{25}	0.0045	0.349	0.953	0.0098	0.360	0.957	0.0140	0.394	0.965
	θ_{37}	−0.0188	0.365	0.954	−0.0123	0.376	0.957	−0.0063	0.433	0.965
	θ_{50}	−0.0161	0.358	0.956	−0.0071	0.371	0.958	−0.0009	0.405	0.960
	η_1	0.0030	0.047	0.947	−0.0041	0.049	0.948	0.0003	0.054	0.943
	η_2	−0.0045	0.046	0.947	0.0026	0.048	0.948	−0.0014	0.054	0.946
Mix-Norm	θ_1	−0.0180	0.364	0.941	−0.0179	0.370	0.942	−0.0158	0.397	0.945
	θ_{12}	0.0014	0.367	0.953	0.0034	0.374	0.951	0.0044	0.394	0.956
	θ_{25}	0.0000	0.353	0.947	0.0038	0.359	0.950	0.0046	0.382	0.954
	θ_{37}	−0.0134	0.338	0.956	−0.0079	0.344	0.956	−0.0063	0.374	0.959
	θ_{50}	−0.0122	0.343	0.944	−0.0044	0.349	0.947	0.0011	0.372	0.952
	η_1	0.0021	0.045	0.940	−0.0048	0.046	0.949	0.0008	0.050	0.945
	η_2	−0.0007	0.045	0.948	0.0064	0.046	0.944	0.0016	0.051	0.948

In addition, the average standard deviation of the estimator decreases as n or T increases, as expected.

Table 8: Simulation results when $\varepsilon_{ijt} \sim N(0,1)$, $p_n = 1/\sqrt{n}$ and $q_n = p_n \log n$.

		$T = 1$			$T = 3$		
n		Bias	SD	CP	Bias	SD	CP
50	θ_1	−0.0062	0.609	0.944	−0.0053	0.353	0.947
	θ_{12}	−0.0204	0.632	0.951	−0.0121	0.342	0.950
	θ_{25}	−0.0188	0.608	0.957	−0.0173	0.343	0.946
	θ_{37}	−0.0208	0.645	0.956	−0.0036	0.351	0.952
	θ_{50}	−0.0405	0.600	0.942	−0.0199	0.367	0.957
	η_1	−0.0059	0.080	0.949	−0.0063	0.046	0.941
	η_2	0.0046	0.083	0.948	0.0060	0.047	0.955
100	θ_1	0.0012	0.513	0.946	0.0054	0.297	0.962
	θ_{25}	0.0066	0.492	0.959	−0.0181	0.296	0.951
	θ_{50}	0.0086	0.482	0.949	−0.0077	0.287	0.962
	θ_{75}	−0.0024	0.498	0.949	−0.0216	0.280	0.953
	θ_{100}	−0.0076	0.493	0.950	−0.0126	0.291	0.955
	η_1	−0.0048	0.046	0.949	−0.0030	0.027	0.955
	η_2	0.0055	0.045	0.945	0.0050	0.028	0.954

Table 9: Simulation results when $\varepsilon_{ijt} \sim \text{Logistic}(0, \sqrt{3}/\pi)$, $p_n = 1/\sqrt{n}$ and $q_n = p_n \log n$.

n		$T=1$ Bias	SD	CP	$T=3$ Bias	SD	CP
50	θ_1	-0.0058	0.641	0.948	-0.0082	0.354	0.954
	θ_{12}	0.0025	0.619	0.948	-0.0203	0.360	0.953
	θ_{25}	0.0010	0.612	0.946	-0.0264	0.347	0.945
	θ_{37}	0.0149	0.608	0.946	-0.0284	0.356	0.941
	θ_{50}	-0.0031	0.615	0.947	-0.0327	0.349	0.951
	η_1	-0.0031	0.085	0.960	-0.0041	0.047	0.944
	η_2	0.0008	0.085	0.951	0.0006	0.047	0.961
100	θ_1	0.0014	0.541	0.963	0.0147	0.294	0.955
	θ_{25}	-0.0032	0.536	0.954	0.0108	0.300	0.952
	θ_{50}	0.0151	0.507	0.954	0.0185	0.296	0.959
	θ_{75}	-0.0013	0.515	0.950	0.0122	0.294	0.959
	θ_{100}	-0.0112	0.538	0.961	-0.0015	0.308	0.957
	η_1	-0.0041	0.047	0.951	-0.0027	0.028	0.953
	η_2	0.0037	0.045	0.953	0.0036	0.025	0.950

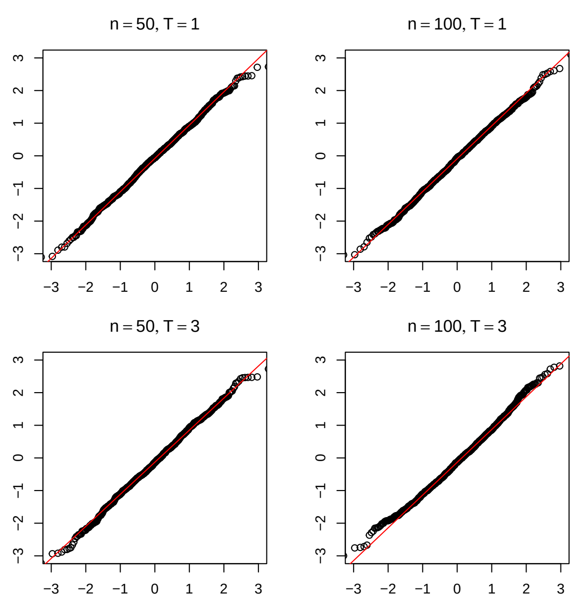

Figure 7: The QQ plots of η_1 when $\varepsilon_{ijt} \sim N(0,1)$, $p_n = 1/\sqrt{n}$ and $q_n = p_n \log n$.

Table 10: Simulation results when $\varepsilon_{ijt} \sim$ Mix-Norm, $p_n = 1/\sqrt{n}$ and $q_n = p_n \log n$.

		$T=1$			$T=3$		
n		Bias	SD	CP	Bias	SD	CP
50	θ_1	0.0186	0.610	0.952	0.0046	0.350	0.948
	θ_{12}	0.0238	0.612	0.948	−0.0056	0.350	0.940
	θ_{25}	0.0108	0.613	0.958	−0.0026	0.360	0.947
	θ_{37}	−0.0095	0.626	0.949	−0.0077	0.344	0.955
	θ_{50}	−0.0189	0.598	0.951	−0.0013	0.350	0.955
	η_1	−0.0077	0.083	0.956	−0.0038	0.046	0.951
	η_2	0.0093	0.087	0.946	0.0042	0.046	0.950
100	θ_1	0.0168	0.491	0.949	0.0177	0.293	0.952
	θ_{25}	0.0027	0.499	0.953	0.0040	0.294	0.951
	θ_{50}	−0.0132	0.491	0.949	−0.0033	0.291	0.942
	θ_{75}	−0.0016	0.489	0.954	0.0010	0.278	0.949
	θ_{100}	0.0068	0.483	0.944	0.0019	0.285	0.946
	η_1	−0.0079	0.044	0.949	−0.0047	0.025	0.946
	η_2	0.0060	0.046	0.942	0.0063	0.027	0.950

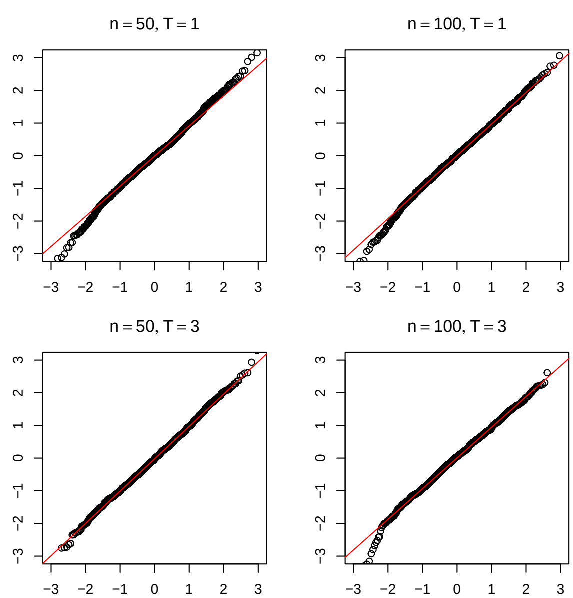

Figure 8: The QQ plots of θ_1 when $\varepsilon_{ijt} \sim N(0,1)$, $p_n = 1/\sqrt{n}$ and $q_n = p_n \log n$.

Table 11: The comparison results of the bias of the estimators obtained by our method and Yan (2025)'s method with $T = 1$.

		$N(0,1)$		Logistic$(0, \sqrt{3}/\pi)$		Mix-Norm		Logistic$(0,1)$	
n		Ours	Yan	Ours	Yan	Ours	Yan	Ours	Yan
50	θ_1	0.0054	0.0313	0.0166	0.0325	−0.0179	−0.0017	0.0135	0.0193
	θ_{12}	−0.0019	0.0900	−0.0035	0.0898	0.0034	0.0972	−0.0307	0.0070
	θ_{25}	−0.0113	0.1703	0.0098	0.1999	0.0038	0.1903	−0.0661	0.0193
	θ_{37}	−0.0066	0.2546	−0.0123	0.2752	−0.0079	0.2802	−0.1063	0.0245
	θ_{50}	−0.0178	0.3341	−0.0071	0.3659	−0.0044	0.3817	−0.1271	0.0286
	η_1	−0.0054	−0.0004	−0.0041	−0.0014	−0.0048	−0.0015	−0.0299	0.0001
	η_2	0.0026	−0.0013	0.0026	−0.0011	0.0064	0.0014	0.0272	−0.0021
100	θ_1	−0.0025	0.0009	−0.0093	−0.0024	0.0115	0.0083	−0.0054	−0.0090
	θ_{25}	0.0025	0.0880	−0.0099	0.0972	0.0065	0.1145	−0.0448	−0.0123
	θ_{50}	−0.0081	0.1672	−0.0160	0.2059	0.0061	0.2159	−0.0699	0.0122
	θ_{75}	−0.0140	0.2549	−0.0259	0.2885	0.0097	0.3148	−0.1171	−0.0050
	θ_{100}	−0.0177	0.3586	−0.0308	0.3864	0.0075	0.4216	−0.1387	0.0032
	η_1	−0.0025	0.0009	−0.0023	−0.0005	−0.0044	−0.0006	−0.0259	0.0000
	η_2	0.0037	0.0001	0.0030	0.0014	0.0050	0.0001	0.0281	0.0014

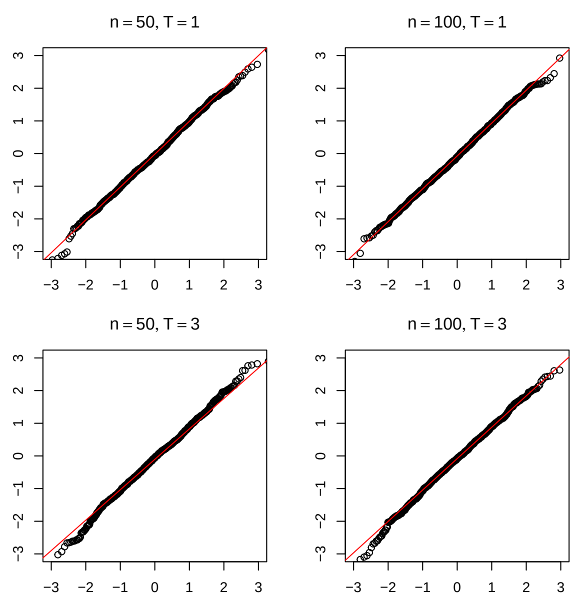

Figure 9: The QQ plots of η_1 when $\varepsilon_{ijt} \sim$ Logistic$(0, \sqrt{3}/\pi)$, $p_n = 1/\sqrt{n}$ and $q_n = p_n \log n$.

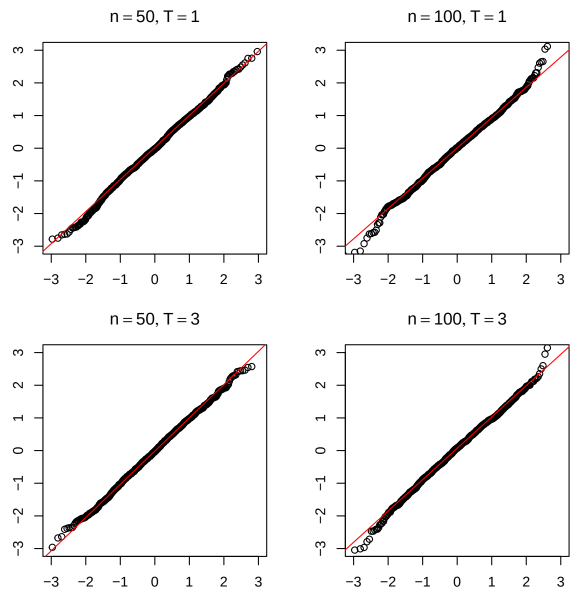

Figure 10: The QQ plots of θ_1 when $\varepsilon_{ijt} \sim$ Logistic$(0, \sqrt{3}/\pi)$, $p_n = 1/\sqrt{n}$ and $q_n = p_n \log n$.

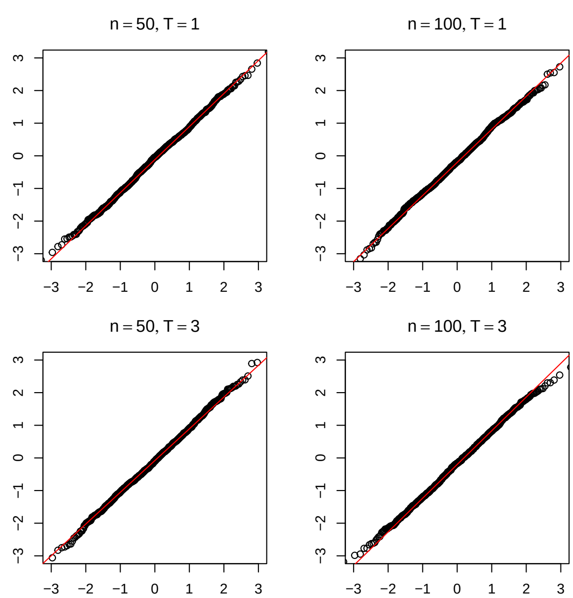

Figure 11: The QQ plots of η_1 when $\varepsilon_{ijt} \sim$ Mix-Norm, $p_n = 1/\sqrt{n}$ and $q_n = p_n \log n$.

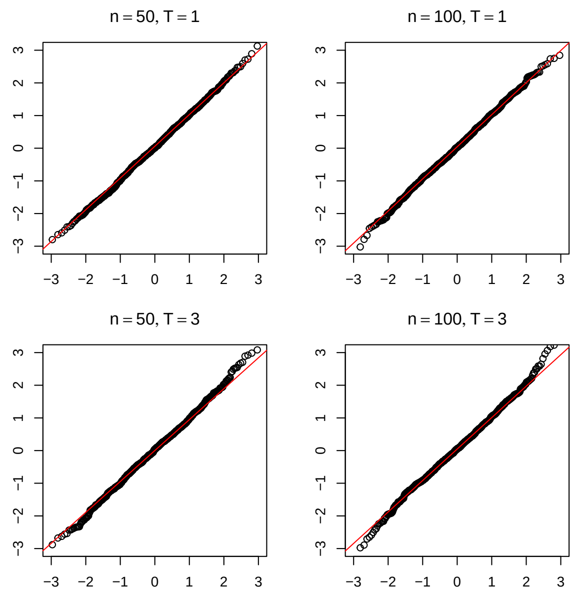

Figure 12: The QQ plots of θ_1 when $\varepsilon_{ijt} \sim$ Mix-Norm, $p_n = 1/\sqrt{n}$ and $q_n = p_n \log n$.

S9 Comparison with the win counting method

We use the recovery of the top-K items to compare the finite-sample performance between our method and the win counting method. Specifically, we define the following error rate as, i.e., Hamming distance, the comparison rule:

$$H_K(\widehat{R}, R^*) = \frac{1}{2K}\left(\sum_{i=1}^{n} \mathbb{I}(\widehat{R}_i > K, R_i^* \le K) + \sum_{i=1}^{n} \mathbb{I}(\widehat{R}_i \le K, R_i^* > K)\right),$$

where R_i^* is the true rank of item i and $\widehat{R}_i$ is the rank estimators based on the estimators of the merit parameters or the win counts. Here, $\mathbb{I}(\cdot)$ is an indicator function. This definition measures the proportion of mistakes for the top-K ranking problem.

In the balanced setting, the number of comparisons between any two items i and j is fixed at T. We set $K = 5$ and chose three values for T: $T = 1, 3, 5$. We set $n = 50$. The merit parameters were $\theta_i^* = 0.4i/(n-K)$ for $i = 0, 1, \ldots, n-K$ and $\theta_{n-K+i}^* = 1 - 0.1(K-i)$ for $i = 1, \ldots, K$. Here, the gap value between the top-5 items and the other items is $[1-0.1(5-1)] - 0.4 = 0.2$. All other settings are the same as those presented in the main paper. The results are presented in Table 12. From this table, we can see that the error rate under the semiparametric paired comparison model is comparable to the win counting method and the latter is a little smaller as expected. Furthermore, we compare these two methods under unbalanced settings. Specifically, the number of comparisons between any two items i and j follows a Bernoulli distribution Bernoulli$(T, p_{n,ij})$, where the probability $p_{ij,n}$ follows a uniform distribution $U(p_n, q_n)$ with $p_n = 1/\sqrt{n}$ and $q_n = \log n/\sqrt{n}$. All other settings are the same as before. The simulation results are also reported in Table 12. From this table, we can see that the ranking estimator based on the win counting method has a lager error rate than the ranking estimator obtained by our method.

Table 12: Hamming distance results for the win-count-based method and our method with $n = 50$.

		$N(0,1)$		Logistic$(0, \sqrt{3}/\pi)$		Mix-Norm	
		Win counts	Our method	Win counts	Our method	Win counts	Our method
Balanced	$T = 1$	0.3352	0.3556	0.3188	0.3496	0.3372	0.3572
	$T = 3$	0.1384	0.1568	0.1350	0.1646	0.1394	0.1590
	$T = 5$	0.0806	0.0924	0.0760	0.1004	0.0790	0.0886
Unbalanced	$T = 1$	0.6444	0.5858	0.6340	0.5846	0.6318	0.5924
	$T = 3$	0.4292	0.3588	0.4130	0.3586	0.4222	0.3550
	$T = 5$	0.3280	0.2528	0.3164	0.2554	0.3188	0.2464

S10 Table 2 with confidence intervals

Table 13 extends Table 2 with the 95% confidence intervals for merit parameters.

Table 13: The estimates of θ_i in 2018-19 NBA regular season.

Eastern Conference					Western Conference				
Team	a_i	$\widehat{\theta}_i$	$\widehat{R}_i$	CI	Team	a_i	$\widehat{\theta}_i$	$\widehat{R}_i$	CI
Milwaukee Bucks	60	1.981	1	[1.694,2.267]	Golden State Warriors	57	1.874	3	[1.583,2.165]
Toronto Raptors	58	1.978	2	[1.687,2.269]	Denver Nuggets	54	1.669	5	[1.378,1.960]
Philadelphia 76ers	51	1.646	6	[1.355,1.937]	Houston Rockets	53	1.680	4	[1.389,1.971]
Boston Celtics	49	1.608	7	[1.317,1.899]	Portland Trail Blazers	53	1.365	11	[1.076,1.653]
Indiana Pacers	48	1.579	8	[1.288,1.870]	Utah Jazz	50	1.394	10	[1.105,1.682]
Orlando Magic	42	0.879	21	[0.592,1.166]	Oklahoma City Thunder	49	1.491	9	[1.205,1.778]
Brooklyn Nets	42	1.044	16	[0.758,1.331]	San Antonio Spurs	48	1.323	12	[1.032,1.614]
Detroit Pistons	41	1.241	14	[0.954,1.527]	LA Clippers	48	1.243	13	[0.955,1.532]
Miami Heat	39	1.062	15	[0.771,1.353]	Sacramento Kings	39	0.880	20	[0.593,1.166]
Charlotte Hornets	39	0.905	19	[0.614,1.196]	Los Angeles Lakers	37	0.970	17	[0.679,1.261]
Washington Wizards	32	0.795	23	[0.509,1.082]	Minnesota Timberwolves	36	0.950	18	[0.659,1.241]
Atlanta Hawks	29	0.448	26	[0.161,0.734]	New Orleans Pelicans	33	0.824	22	[0.533,1.115]
Chicago Bulls	22	0.200	27	[−0.091,0.491]	Dallas Mavericks	33	0.696	25	[0.404,0.987]
Cleveland Cavaliers	19	0.111	28	[−0.176,0.397]	Memphis Grizzlies	33	0.713	24	[0.425,1.001]
New York Knicks	17	0	30	–	Phoenix Suns	19	0.043	29	[−0.243,0.330]

S11 Error bounds for independent random variables and nonparameteric density estimators

We reproduce Hoeffding's inequality in Hoeffding (1963) and error bounds for nonparameteric density estimators in Andrews (1995) here.

Lemma S11.1 (Hoeffding's inequality). *Let $X_1, X_2, \ldots, X_n$ be n independent random variables satisfying $X_i \in [a_i, b_i]$ for some constants a_i and b_i. Let $\mathcal{X} = \sum_{i=1}^n X_i$. For any $\varepsilon \geq 0$,*

$$\mathbb{P}\left(|\mathcal{X} - \mathbb{E}(\mathcal{X})| \geq \varepsilon\right) \leq 2\exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Lemma S11.2 (Andrews (1995)). *Under Conditions 1, 6 and 7, we consider a triangular array of random variables $\{(Y_t, \boldsymbol{X}_t) : 1 \leq t \leq \mathcal{T}\}$, where*

$Y_t \in \mathbb{R}$ and $\boldsymbol{X}_t \in \mathbb{R}^k$. Define the kernel estimator for the density function $f(\boldsymbol{x})$ of $\boldsymbol{X}_t$ as:

$$\hat{f}(\boldsymbol{x}) = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \mathcal{K}_{x,h}(\boldsymbol{x} - \boldsymbol{X}_t).$$

Here

$$\mathcal{K}_{x,h}(\boldsymbol{x} - \boldsymbol{X}_t) = \frac{1}{h^k} \mathcal{K}_x\left(\frac{\boldsymbol{x} - \boldsymbol{X}_t}{h}\right),$$

where h denotes the bandwidth parameter and $\mathcal{K}_x(\cdot)$ is a kernel function. Define $g(\boldsymbol{x}) = \mathbb{E}(Y_t|\boldsymbol{X}_t = \boldsymbol{x})$, and its kernel estimator is

$$\hat{g}(\boldsymbol{x}) = \frac{\frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} Y_t \mathcal{K}_{x,h}(\boldsymbol{x} - \boldsymbol{X}_t)}{\hat{f}(\boldsymbol{x})}.$$

Define D^λ as the λth order derivative for a function, where $\lambda \geq 0$. Let $f_t(\boldsymbol{x})$ be the density function of $\boldsymbol{x}$, and assume it is continuously differentiable to integral order ω in $\boldsymbol{x}$. Denote $\eta \in (0, \infty]$ as the near-epoch dependent number.

Then we have

$$\begin{aligned} &\sup_{\boldsymbol{x} \in R^k} \left| D^\lambda \hat{f}(\boldsymbol{x}) - \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} D^\lambda f_t(\boldsymbol{x}) \right| \\ = \;& O_p\left(\mathcal{T}^{-\frac{\eta}{2\eta+1}} h^{-k-|\lambda|-\frac{1}{2\eta+1}}\right) + O_p\left(h^{\omega-|\lambda|}\right) \end{aligned}$$

and

$$\begin{aligned} &\sup_{\boldsymbol{x}: \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} D^\lambda f_t(\boldsymbol{x}) \geq d_{\mathcal{T}}} \left| D^\lambda \hat{g}(\boldsymbol{x}) - D^\lambda g(\boldsymbol{x}) \right| \\ = \;& O_p\left(\mathcal{T}^{-\frac{\eta}{2\eta+1}} h^{-k-|\lambda|-\frac{1}{2\eta+1}} d_{\mathcal{T}}^{-2-|\lambda|}\right) + O_p\left(h^{\omega-|\lambda|} d_{\mathcal{T}}^{-2-|\lambda|}\right). \end{aligned}$$

In this paper, we choose $\lambda = 0$, $\mathcal{T} = NT$, $\eta = \infty$, $\omega = r$, and $d_{\mathcal{T}} = \varpi$. The value of k depends on the dimensionality of the random variable in

$\mathcal{K}(\cdot)$, taking the value p or $p+1$.

References

Andrews, D. W. K. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory*, 11(3):560–586.

Fan, J., Hou, J., and Yu, M. (2024). Uncertainty quantification of MLE for entity ranking with covariates. *Journal of Machine Learning Research*, 25(358):1–83.

Feller, W. (1935). Uber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 40:521–559.

Han, R., Tang, W., and Xu, Y. (2024). Statistical inference for pairwise comparison models. *arXiv preprint arXiv:2401.08463*.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.

Lindeberg, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15(1):211–225.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, volume 3. Cambridge University Press.

Yan, T. (2025). Inference in a generalized Bradley-Terry model for paired comparisons with covariates and a growing number of subjects. *arXiv:2507.22472*.