Supplementary Material to "Distributed Logistic Regression for Massive Data with Rare Events"

Xuetong Li[1], Xuening Zhu[2], and Hansheng Wang[1]

[1]*Peking University and* [2]*Fudan University*

S1 Proof of Theorem 1

For each $\widehat{\theta}_{\mathrm{RMLE},k}$ $(1 \le k \le K)$, we apply the Taylor expansion and then obtain the following representation as

$$\dot{\mathcal{L}}_{\mathrm{R},k}\big(\widehat{\theta}_{\mathrm{RMLE},k}\big) = \dot{\mathcal{L}}_{\mathrm{R},k}\big(\theta^*\big) + \ddot{\mathcal{L}}_{\mathrm{R},k}\big(\theta^*\big)\Big(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*\Big) + \frac{1}{2}\Gamma_{\mathrm{R},k},$$

where $\Gamma_{\mathrm{R},k} = (\Gamma_{\mathrm{R},k,j}) \in \mathbb{R}^{p+1}$ with $1 \le j \le p+1$, $\Gamma_{\mathrm{R},k,j} = (\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)^\top \Delta_{\mathrm{R},k,j}(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)$, $\Delta_{\mathrm{R},k,j} = \big(\Delta^{(j_1 j_2)}_{\mathrm{R},k,j}\big) \in \mathbb{R}^{(p+1)\times(p+1)}$, $\Delta^{(j_1 j_2)}_{\mathrm{R},k,j} = \partial^2 \dot{\mathcal{L}}_{\mathrm{R},k,j}(\theta)/\partial\theta_{j_1}\partial\theta_{j_2}|_{\theta=\widetilde{\theta}_{\mathrm{R},k}}$ for $1 \le j_1, j_2 \le p+1$, $\dot{\mathcal{L}}_{\mathrm{R},k,j}(\theta)$ is the jth element of $\dot{\mathcal{L}}_{\mathrm{R},k}(\theta)$, and $\widetilde{\theta}_{\mathrm{R},k} = \eta_k \widehat{\theta}_{\mathrm{RMLE},k} + (1-\eta_k)\theta^*$ for some $0 \le \eta_k \le 1$. In the meanwhile, we define $\Gamma^*_{\mathrm{R},k} = (\Gamma^*_{\mathrm{R},k,j})$, $\Gamma^*_{\mathrm{R},k,j} = (\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)^\top \Delta^*_{\mathrm{R},k,j}(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)$ and $\Delta^*_{\mathrm{R},k,j} = \partial^2 \dot{\mathcal{L}}_{\mathrm{R},k,j}(\theta)/\partial\theta_{j_1}\partial\theta_{j_2}|_{\theta=\theta^*}$. We also define $\Gamma^{\Omega}_{\mathrm{R},k} = (\Gamma^{\Omega}_{\mathrm{R},k,j}) \in \mathbb{R}^{p+1}$ with $1 \le j \le p+1$, $\Gamma^{\Omega}_{\mathrm{R},k,j} = ne^{\alpha^*_N}(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)^\top \Omega_{\mathrm{R},j}(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)$ and $\Omega_{\mathrm{R},j} = E\{\Delta^*_{\mathrm{R},k,j}/(ne^{\alpha^*_N})\}$. More specifically, by definition we have

$\mathcal{L}_{\mathrm{R},k}(\theta) = \sum_{i=1}^{N} a_i^{(k)}\big\{Y_i Z_i^\top\theta - \log\big(1+e^{Z_i^\top\theta}\big)\big\}$. Thus, we have $\dot{\mathcal{L}}_{\mathrm{R},k}(\theta) = \sum_{i=1}^{N} a_i^{(k)}\big\{Y_i - p_i(\alpha_N,\beta)\big\}Z_i$, $\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta) = -\sum_{i=1}^{N} a_i^{(k)} p_i(\alpha_N,\beta)\big\{1-p_i(\alpha_N,\beta)\big\}Z_i Z_i^\top$, and $\Delta_{\mathrm{R},k,j} = -\sum_{i=1}^{N} a_i^{(k)} p_i(\alpha_N,\beta)\{1-p_i(\alpha_N,\beta)\}\{1-2p_i(\alpha_N,\beta)\}Z_{i,j}Z_iZ_i^\top$, respectively. Then, we obtain

$$\widehat{\theta}_{\mathrm{RMLE},k} - \theta^* \;=\; -\Big\{\ddot{\mathcal{L}}_{\mathrm{R},k}\big(\theta^*\big)\Big\}^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}\big(\theta^*\big) - \frac{1}{2}\Big\{\ddot{\mathcal{L}}_{\mathrm{R},k}\big(\theta^*\big)\Big\}^{-1}\Gamma_{\mathrm{R},k}.$$

For an arbitrary vector γ, define $\|\gamma\| = \sqrt{\gamma^\top\gamma}$. For an arbitrary square matrix $A=(a_{ij})$, define $\|A\| = \sqrt{\lambda_{\max}(A^\top A)}$, where $\lambda_{\max}(B)$ stands for the maximum eigenvalue of an arbitrary semi-positive definite matrix B. Define $\lambda_{\min}(B)$ as the minimum eigenvalue of an arbitrary semi-positive definite matrix B.

Subsequently, we should decompose the RMLE estimator into a total of five different terms as

$$\widehat{\theta}_{\mathrm{RMLE}} - \theta^* = Q_{\mathrm{R},1} + Q_{\mathrm{R},2}/2 + Q_{\mathrm{R},3}/2 + Q_{\mathrm{R},4}/2 + Q_{\mathrm{R},5}/2,$$

where $Q_{\mathrm{R},1} = K^{-1}\sum_{k=1}^{K}\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)$, $Q_{\mathrm{R},2} = K^{-1}\sum_{k=1}^{K}\Sigma^{*-1}(ne^{\alpha_N^*})^{-1}\Gamma^{\Omega}_{\mathrm{R},k} = B(\theta^*)/(ne^{\alpha_N^*})$, $Q_{\mathrm{R},3} = K^{-1}\sum_{k=1}^{K}(\widehat{\Sigma}_{\mathrm{R},k}^{-1} - \Sigma^{*-1})\{\Gamma^*_{\mathrm{R},k}/(ne^{\alpha_N^*})\}$, $Q_{\mathrm{R},4} = K^{-1}\sum_{k=1}^{K}\widehat{\Sigma}_{\mathrm{R},k}^{-1}\{(\Gamma_{\mathrm{R},k} - \Gamma^*_{\mathrm{R},k})/(ne^{\alpha_N^*})\}$, $Q_{\mathrm{R},5} = K^{-1}\sum_{k=1}^{K}\Sigma^{*-1}\{(\Gamma^*_{\mathrm{R},k} - \Gamma^{\Omega}_{\mathrm{R},k})(ne^{\alpha_N^*})^{-1}\}$, $\Sigma^* = E(e^{X_i^\top\beta^*}Z_iZ_i^\top)$, $\widehat{\Sigma}_{\mathrm{R},k} = -\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)/(ne^{\alpha_N^*})$ and $B(\theta^*) = K^{-1}\sum_{k=1}^{K}\Sigma^{*-1}\Gamma^{\Omega}_{\mathrm{R},k}$. Subsequently, we shall study the asymptotic property of the RMLE estimator in eight steps. In the 1st step, we show that $E(e^{t\|X_i\|}) = M_t < \infty$ for any $t \in \mathbb{R}$. In the 2nd step, we provide three uniform convergence results. We next verify in the 3rd step that $\max_k \|\widehat{\theta}_{\text{RMLE},k} - \theta^*\| = o_p(\sqrt{\log K}/\sqrt{ne^{\alpha_N^*}})$. We show in the 4th step that $K^{-1}\sum_{k=1}^K \|\widehat{\theta}_{\text{RMLE},k} - \theta^*\|^2 = O_p\{(ne^{\alpha_N^*})^{-1}\}$. The results obtained in Steps 1–4 are to be used in the subsequent technical proofs. In the 5th step, we prove that $\sqrt{Ne^{\alpha_N^*}}Q_{\text{R},1} \to_d N(0, \Sigma^{*-1})$ as $N \to \infty$. In the 6th step, we show that $Q_{\text{R},2}$ is a random bias term with $C_{\min} \le E\{\|B(\theta^*)\|\} \le C_{\max}$ for some fixed positive constants $C_{\min}$ and $C_{\max}$. In the 7th step, we demonstrate that $Q_{\text{R},3} = o_p\{(ne^{\alpha_N^*})^{-1}\}$. In the last step, we prove that $Q_{\text{R},4}$ and $Q_{\text{R},5}$ are of the order $o_p\{(ne^{\alpha_N^*})^{-1}\}$.

Step 1. By the theorem condition (C1), we have $P(\|X_i\| > M) \le 2\exp(-C_{\text{Tail}}M^2)$ for some positive constant C_{Tail}. We then have

$$E\left\{\exp\left(t\Big\|X_i\Big\|\right)\right\} = \int_0^\infty P\left\{\exp\left(t\Big\|X_i\Big\|\right) > s\right\} ds$$
$$= \int_0^\infty P\left(\Big\|X_i\Big\| > \frac{\log s}{t}\right) ds \le \int_0^\infty 2\exp\left(-C_{\text{Tail}}\frac{\log^2 s}{t^2}\right) ds.$$

By letting $z = \sqrt{C_{\text{Tail}}}\{\log s - t^2/(2C_{\text{Tail}})\}/t$, we then compute

$$E\left\{\exp\left(t\Big\|X_i\Big\|\right)\right\} \le \int_0^\infty 2\exp\left(-C_{\text{Tail}}\frac{\log^2 s}{t^2}\right) ds$$
$$= \frac{2t}{\sqrt{C_{\text{Tail}}}}\exp\left(\frac{t^2}{4C_{\text{Tail}}}\right)\int_{-\infty}^\infty e^{-z^2} dz = M_t < \infty,$$

where $M_t = 2\sqrt{\pi}t\exp\left\{t^2/(4C_{\text{Tail}})\right\}/\sqrt{C_{\text{Tail}}}$.

Step 2. We provide here three uniform convergence result, which is also to be used in the subsequent proofs. These three results are given by

$$\max_k \left\| \dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)/(ne^{\alpha_N^*}) \right\| = o_p\left(\sqrt{\frac{\log K}{ne^{\alpha_N^*}}}\right), \tag{S1.1}$$

$$\max_k \left\| \widehat{\Sigma}_{\mathrm{R},k}^{-1} - \Sigma^{*-1} \right\| = o_p\left(\sqrt{\frac{\log K}{ne^{\alpha_N^*}}}\right) + O(e^{\alpha_N^*}), \tag{S1.2}$$

$$\max_k \left\| \Delta_{\mathrm{R},k,j}^*/(ne^{\alpha_N^*}) - \Omega_{\mathrm{R},j} \right\| = o_p\left(\sqrt{\frac{\log K}{ne^{\alpha_N^*}}}\right). \tag{S1.3}$$

The technical details for proving (S1.1)–(S1.3) are very similar. Comparatively speaking, the proof of (S1.2) is technically more complicated. We thus provide the proof details for (S1.2) only. The proofs for (S1.1) and (S1.3) are omitted to save space. Specifically, by (A.9) in Jordan et al. (2018), we have

$$\begin{aligned}\max_k \left\| \widehat{\Sigma}_{\mathrm{R},k}^{-1} - \Sigma^{*-1} \right\| \le \left\| \Sigma^{*-1} \right\|^2 \max_k \left\| \widehat{\Sigma}_{\mathrm{R},k} - E(\widehat{\Sigma}_{\mathrm{R},k}) \right\| \\ + \left\| \Sigma^{*-1} \right\|^2 \left\| E(\widehat{\Sigma}_{\mathrm{R},k}) - \Sigma^* \right\|,\end{aligned}$$

where $E(\widehat{\Sigma}_{\mathrm{R},k}) = E\{e^{X_i^\top \beta^*}(1+e^{Z_i^\top \theta^*})^{-2} Z_i Z_i^\top\}$. We then study these two parts separately with details.

Step 2.1. We start with the first part $\|\Sigma^{*-1}\|^2 \max_k \|\widehat{\Sigma}_{\mathrm{R},k} - E(\widehat{\Sigma}_{\mathrm{R},k})\|$. Write $\widehat{\Sigma}_{\mathrm{R},k} - E(\widehat{\Sigma}_{\mathrm{R},k}) = (ne^{\alpha_N^*})^{-1} \sum_{i=1}^N U_{ki}$, where $U_{ki} = a_i^{(k)} p_i(\alpha_N^*, \beta^*)\{1 - p_i(\alpha_N^*, \beta^*)\} Z_i Z_i^\top - K^{-1} E[p_i(\alpha_N^*, \beta^*)\{1 - p_i(\alpha_N^*, \beta^*)\} Z_i Z_i^\top]$. Next, we define $U_{ki} = (U_{ki,j_1 j_2})$ with $1 \le j_1, j_2 \le p+1$ and $U_{ki,j_1 j_2}^M = U_{ki,j_1 j_2} I(\|Z_i\| \le M)$.

Consider an arbitrary small positive number ε, we have

$$P\left(\Big|\frac{1}{ne^{\alpha_N^*}}\sum_{i=1}^{N}U_{ki,j_1j_2}\Big|>\varepsilon\right)=P\left(\Big|\frac{1}{ne^{\alpha_N^*}}\sum_{i=1}^{N}U_{ki,j_1j_2}\Big|>\varepsilon,\mathcal{E}\right)$$
$$+P\left(\Big|\frac{1}{ne^{\alpha_N^*}}\sum_{i=1}^{N}U_{ki,j_1j_2}\Big|>\varepsilon,\overline{\mathcal{E}}\right),\tag{S1.4}$$

where $\mathcal{E}=\{\max_i\|Z_i\|\leq M\}$ is a random event and $\overline{\mathcal{E}}$ stands for its complement. For the first part in (S1.4), we have

$$P\left(\Big|\frac{1}{ne^{\alpha_N^*}}\sum_{i=1}^{N}U_{ki,j_1j_2}\Big|>\varepsilon,\mathcal{E}\right)\leq P\left(\Big|\frac{1}{ne^{\alpha_N^*}}\sum_{i=1}^{N}U^M_{ki,j_1j_2}\Big|>\varepsilon\right)$$
$$\leq P\left(\Big|\frac{1}{ne^{\alpha_N^*}}\sum_{i=1}^{N}\Big\{U^M_{ki,j_1j_2}-E\Big(U^M_{ki,j_1j_2}\Big)\Big\}\Big|+\Big|E\Big(Ke^{-\alpha_N^*}U^M_{ki,j_1j_2}\Big)\Big|>\varepsilon\right).$$

Next, note that $E(Ke^{-\alpha_N^*}U^M_{ki,j_1j_2})\to E(Ke^{-\alpha_N^*}U_{ki,j_1j_2})=0$ as $M\to\infty$. This suggests that for any $\varepsilon>0$, there should exists a sufficiently large but fixed positive constant $M_\varepsilon>0$ such that $|E(Ke^{-\alpha_N^*}U^M_{ki,j_1j_2})|\leq\varepsilon/2$ for any $M>M_\varepsilon$. Therefore, we have

$$P\left(\Big|\frac{1}{ne^{\alpha_N^*}}\sum_{i=1}^{N}U_{ki,j_1j_2}\Big|>\varepsilon,\mathcal{E}\right)$$
$$\leq P\left(\Big|\frac{1}{ne^{\alpha_N^*}}\sum_{i=1}^{N}\Big\{U^M_{ki,j_1j_2}-E\Big(U^M_{ki,j_1j_2}\Big)\Big\}\Big|>\frac{\varepsilon}{2}\right).$$

Moreover, we apply the Bernstein's inequality (Bernstein, 1926; Zhang and Chen, 2020) to $\{U^M_{ki,j_1j_2} - E(U^M_{ki,j_1j_2}) : 1 \le i \le N\}$ and obtain

$$P\left(\Big|\frac{1}{ne^{\alpha_N^*}}\sum_{i=1}^N\Big\{U^M_{ki,j_1j_2} - E\Big(U^M_{ki,j_1j_2}\Big)\Big\}\Big| > \frac{\varepsilon}{2}\right)$$

$$\le 2\exp\left(-\frac{n^2e^{2\alpha_N^*}\varepsilon^2/8}{\sum_{i=1}^N \mathrm{var}(U^M_{ki,j_1j_2}) + Mne^{\alpha_N^*}\varepsilon/6}\right). \tag{S1.5}$$

Further, we have $\mathrm{var}(U^M_{ki,j_1j_2}) \le E\{(U^M_{ki,j_1j_2})^2\} = E\{U^2_{ki,j_1j_2}I(\|Z_i\| \le M)\} \le E(U^2_{ki,j_1j_2}) = \mathrm{var}[a_i^{(k)}p_i(\alpha_N^*,\beta^*)\{1-p_i(\alpha_N^*,\beta^*)\}Z_{ij_1}Z_{ij_2}] \le E[a_i^{(k)}p_i^2(\alpha_N,\beta)\{1 - p_i(\alpha_N^*,\beta^*)\}^2Z^2_{ij_1}Z^2_{ij_2}] \le K^{-1}e^{\alpha_N^*}E(e^{X_i^\top\beta^*}Z^2_{ij_1}Z^2_{ij_2}) \le E(e^{t\|X_i\|})$. Here the last inequality holds for a sufficiently large but fixed constant t due to (S.1) in the supplementary material of Wang (2020). By Step 1, we have $E(e^{t\|X_i\|}) \le M_t$. Thus, we further bound the right hand side of (S1.5) by

$$P\left(\Big|\frac{1}{ne^{\alpha_N^*}}\sum_{i=1}^N\Big\{U^M_{ki,j_1j_2} - E\Big(U^M_{ki,j_1j_2}\Big)\Big\}\Big| > \frac{\varepsilon}{2}\right) \le 2\exp\left(-\frac{ne^{\alpha_N^*}\varepsilon^2/8}{M_t + M\varepsilon/6}\right). \tag{S1.6}$$

Next, by the theorem condition (C1), we have for the second part in (S1.4) that

$$P\left(\Big|(ne^{\alpha_N^*})^{-1}\sum_{i=1}^N U_{ki,j_1j_2}\Big| > \varepsilon, \overline{\mathcal{E}}\right) \le P\big(\overline{\mathcal{E}}\big) \le \sum_{i=1}^N P\left(\big\|Z_i\big\| > M\right)$$

$$\le 2\exp\Big(\log N - C_{\mathrm{Tail}}M^2\Big). \tag{S1.7}$$

Combining the two results (S1.6) and (S1.7), we obtain

$$P\left(\max_k \left\|\widehat{\Sigma}_{\mathrm{R},k} - E(\widehat{\Sigma}_{\mathrm{R},k})\right\| > \varepsilon\right) \le \sum_{k=1}^{K} P\left(\left\|\widehat{\Sigma}_{\mathrm{R},k} - E(\widehat{\Sigma}_{\mathrm{R},k})\right\| > \varepsilon\right)$$
$$\le 2\exp\left(\log K - \frac{ne^{\alpha_N^*}\varepsilon^2/8}{M_t + M\varepsilon/6}\right) + 2\exp\Big(\log K + \log N - C_{\mathrm{Tail}} M^2\Big). \tag{S1.8}$$

Next, we can define $\varepsilon = 4\sqrt{M_t \log K/(ne^{\alpha_N^*})}$ and $M = \max(M_\varepsilon, \sqrt{2\log N}/\sqrt{C_{\mathrm{Tail}}})$ for (S1.8), we have for the first part of (S1.8) that $2\exp\{-ne^{\alpha_N^*}\varepsilon^2(M_t + M\varepsilon/6)^{-1}/8\} \le 2\exp\big[-\log K\{2M_t(M_t{+}2\sqrt{2M_t \log N \log K}/\sqrt{C_{\mathrm{Tail}} ne^{\alpha_N^*}}/3)^{-1} - 1\}\big] \to 0$ as $N \to \infty$ due to the theorem condition (C3). For the second part of (S1.8), we have $2\exp\big(\log K + \log N - C_{\mathrm{Tail}} M^2\big) = 2\exp(\log K - \log N) = 2/n \to 0$ due to the theorem condition (C2). Therefore, the right hand side of (S1.8) shrinks towards 0 when N diverges to infinity. This suggests that $\max_k \|\widehat{\Sigma}_{\mathrm{R},k} - E(\widehat{\Sigma}_{\mathrm{R},k})\| = o_p(\sqrt{\log K}/\sqrt{ne^{\alpha_N^*}})$. This leads to $\|\Sigma^{*-1}\|^2 \max_k \|\widehat{\Sigma}_{\mathrm{R},k} - E(\widehat{\Sigma}_{\mathrm{R},k})\| = o_p(\sqrt{\log K}/\sqrt{ne^{\alpha_N^*}})$ for the first part in (S1.4).

Step 2.2. For the second part in (S1.4), we have $\|E(\widehat{\Sigma}_{\mathrm{R},k}) - \Sigma^*\| = \|E[e^{X_i^\top \beta^*}\{1-(1+e^{Z_i^\top \theta^*})^{-2}\}Z_i Z_i^\top]\| = e^{\alpha_N^*}\|E\{e^{2X_i^\top \beta^*}(2+e^{Z_i^\top \theta^*})(1+e^{Z_i^\top \theta^*})^{-2} Z_i Z_i^\top\}\| = O(e^{\alpha_N^*})$. Therefore, we finally have $\|\Sigma^{*-1}\|^2 \max_k \|E(\widehat{\Sigma}_{\mathrm{R},k}) - \Sigma^*\| = O(e^{\alpha_N^*})$. This completes the proof of (S1.2).

Step 3. We verify in this step that $\max_k \|\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*\| = o_p\{\sqrt{\log K}(n$

$e^{\alpha_N^*})^{-1/2}\}$. Since $\widehat{\theta}_{\mathrm{RMLE},k}$ is the maximizer of $\mathcal{L}_{\mathrm{R},k}(\theta)$ and $\mathcal{L}_{\mathrm{R},k}(\theta)$ is a strictly concave function for θ, we then have

$$P\left(\left\|\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\right\|>\varepsilon\right)\le P\left[\sup_{\|u\|\le 1}\left\{\mathcal{L}_{\mathrm{R},k}\big(\theta^*+u\varepsilon\big)-\mathcal{L}_{\mathrm{R},k}\big(\theta^*\big)\right\}\ge 0\right]$$
$$=P\left[\sup_{\|u\|\le 1}\left\{\dot{\mathcal{L}}_{\mathrm{R},k}\big(\theta^*\big)^\top u\varepsilon+u^\top\ddot{\mathcal{L}}_{\mathrm{R},k}\big(\theta^*\big)u\varepsilon^2+\widetilde{Q}_{\mathrm{R},k}\varepsilon^2\right\}\ge 0\right],$$

where $\widetilde{Q}_{\mathrm{R},k}=u^\top\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\widetilde{\theta}_{\mathrm{R},k})-\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\}u$ and $\widetilde{\theta}_{\mathrm{R},k}=\eta_k\widehat{\theta}_{\mathrm{US},k}+(1-\eta_k)\theta^*$ for some $0\le\eta_k\le 1$. Note that $\ddot{\mathcal{L}}_{\mathrm{R},k}(\widetilde{\theta}_{\mathrm{R},k})-\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)=-\sum_{i=1}^N a_i^{(k)}p_i(\bar{\theta}_{\mathrm{R},k})\{1-p_i(\bar{\theta}_{\mathrm{R},k})\}\{1-2p_i(\bar{\theta}_{\mathrm{R},k})\}Z_i^\top(\widetilde{\theta}_{\mathrm{R},k}-\theta^*)Z_iZ_i^\top$ with $\bar{\theta}_{\mathrm{R},k}=\bar{\eta}_k\widetilde{\theta}_{\mathrm{US},k}+(1-\bar{\eta}_k)\theta^*$ for some $0\le\bar{\eta}_k\le 1$. Consequently, we have $|\widetilde{Q}_{\mathrm{R},k}|\le\|\ddot{\mathcal{L}}_{\mathrm{R},k}(\widetilde{\theta}_{\mathrm{R},k})-\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\|\le\sum_{i=1}^N a_i^{(k)}\exp(Z_i^\top\bar{\theta}_{\mathrm{R},k})\|Z_i\|^3\|\widetilde{\theta}_{\mathrm{R},k}-\theta^*\|$. Since $\bar{\theta}_{\mathrm{R},k}=\bar{\eta}_k\widetilde{\theta}_{\mathrm{US},k}+(1-\bar{\eta}_k)\theta^*$, we have $\exp(Z_i^\top\bar{\theta}_{\mathrm{R},k})\le\exp(\alpha_N^*+\varepsilon+X_i^\top\beta^*+\|X_i\|\varepsilon)$. This suggest that $|\widetilde{Q}_{\mathrm{R},k}|\le\widetilde{\Delta}_{\mathrm{R},k}\varepsilon$, where $\widetilde{\Delta}_{\mathrm{R},k}=\sum_{i=1}^N a_i^{(k)}\exp(\alpha_N^*+\varepsilon+X_i^\top\beta^*+\|X_i\|\varepsilon)\|Z_i\|^3$. Then, let $\varepsilon=C_N/\sqrt{ne^{\alpha_N^*}}$ with $C_N>0$ and define $\lambda_{\mathcal{L}}=\lambda_{\min}\{-\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta)/(ne^{\alpha_N^*})\}>0$, we obtain

$$P\left(\left\|\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\right\|>\frac{C_N}{\sqrt{ne^{\alpha_N^*}}}\right)\le P\left\{\left\|\frac{\dot{\mathcal{L}}_{\mathrm{R},k}\big(\theta^*\big)}{\sqrt{ne^{\alpha_N^*}}}\right\|-\lambda_{\mathcal{L}}C_N+\frac{\widetilde{\Delta}_{\mathrm{R},k}C_N^2}{(ne^{\alpha_N^*})^{3/2}}>0\right\}$$
$$\le P\left(\left\|\frac{\dot{\mathcal{L}}_{\mathrm{R},k}\big(\theta^*\big)}{ne^{\alpha_N^*}}\right\|>\frac{\lambda_{\mathcal{L}}C_N}{2\sqrt{ne^{\alpha_N^*}}}\right)+P\left(\frac{\widetilde{\Delta}_{\mathrm{R},k}}{ne^{\alpha_N^*}}>\frac{\lambda_{\mathcal{L}}\sqrt{ne^{\alpha_N^*}}}{2C_N}\right). \qquad \text{(S1.9)}$$

We focus on these two part in (S1.9) separately. We first calculate the first part in (S1.9). By the similar arguments as in Step 2, we have

$$P\left(\left\|\frac{\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)}{ne^{\alpha_N^*}}\right\| > \frac{\lambda_{\mathcal{L}} C_N}{2\sqrt{ne^{\alpha_N^*}}}\right) \leq 2\exp\left(\log N - C_{\mathrm{Tail}} M^2\right)$$
$$+2\exp\left\{-\frac{\lambda_{\mathcal{L}}^2 C_N^2/32}{M_t + M\lambda_{\mathcal{L}} C_N/(12\sqrt{ne^{\alpha_N^*}})}\right\}. \tag{S1.10}$$

We next compute the second part in (S1.9). We also have

$$P\left(\frac{\widetilde{\Delta}_{\mathrm{R},k}}{ne^{\alpha_N^*}} > \frac{\lambda_{\mathcal{L}}\sqrt{ne^{\alpha_N^*}}}{2C_N}\right) \leq P\left\{E\left(\frac{\widetilde{\Delta}_{\mathrm{R},k}}{ne^{\alpha_N^*}}\right) > \frac{\lambda_{\mathcal{L}}\sqrt{ne^{\alpha_N^*}}}{4C_N}\right\}$$
$$P\left\{\frac{\widetilde{\Delta}_{\mathrm{R},k}}{ne^{\alpha_N^*}} - E\left(\frac{\widetilde{\Delta}_{\mathrm{R},k}}{ne^{\alpha_N^*}}\right) > \frac{\lambda_{\mathcal{L}}\sqrt{ne^{\alpha_N^*}}}{4C_N}\right\}. \tag{S1.11}$$

For the first part in (S1.11), we have

$$P\left\{\frac{\widetilde{\Delta}_{\mathrm{R},k}}{ne^{\alpha_N^*}} - E\left(\frac{\widetilde{\Delta}_{\mathrm{R},k}}{ne^{\alpha_N^*}}\right) > \frac{\lambda_{\mathcal{L}}\sqrt{ne^{\alpha_N^*}}}{4C_N}\right\} \leq 2\exp\left(\log N - C_{\mathrm{Tail}} M^2\right)$$
$$+2\exp\left\{-\frac{n^2 e^{2\alpha_N^*}\lambda_{\mathcal{L}}^2/(128C_N^2)}{M_t + M\lambda_{\mathcal{L}}\sqrt{ne^{\alpha_N^*}}/(24C_N)}\right\}. \tag{S1.12}$$

Combining (S1.10)–(S1.12), we finally have

$$P\left(\max_k \left\|\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*\right\| > \frac{C_N}{\sqrt{ne^{\alpha_N^*}}}\right) \leq \sum_{k=1}^{K} P\left(\left\|\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*\right\| > \frac{C_N}{\sqrt{ne^{\alpha_N^*}}}\right)$$
$$\leq 2\exp\left\{\log K - \frac{\lambda_{\mathcal{L}}^2 C_N^2/32}{M_t + M\lambda_{\mathcal{L}} C_N/(12\sqrt{ne^{\alpha_N^*}})}\right\} \tag{S1.13}$$
$$+2\exp\left\{\log K - \frac{n^2 e^{2\alpha_N^*}\lambda_{\mathcal{L}}^2/(128C_N^2)}{M_t + M\lambda_{\mathcal{L}}\sqrt{ne^{\alpha_N^*}}/(24C_N)}\right\} \tag{S1.14}$$
$$+KP\left\{E\left(\frac{\widetilde{\Delta}_{\mathrm{R},k}}{ne^{\alpha_N^*}}\right) > \frac{\lambda_{\mathcal{L}}\sqrt{ne^{\alpha_N^*}}}{4C_N}\right\} + 4\exp\left(\log K + \log N - C_{\mathrm{Tail}} M^2\right). \tag{S1.15}$$

Let $C_N = 8\sqrt{M_t \log K}/\lambda_{\mathcal{L}}$ and $M = \max(M_\varepsilon, \sqrt{2\log N/C_{\text{Tail}}})$. Then we have for the term in (S1.13) that $2\exp\{\log K - (M_t + M\lambda_{\mathcal{L}} C_N/12/\sqrt{ne^{\alpha_N^*}})^{-1} \lambda_{\mathcal{L}}^2 C_N^2/32\} \le 2\exp[-\log K\{2M_t(M_t + 2\sqrt{2M_t/C_{\text{Tail}}}\sqrt{\log N}\sqrt{\log K}/\sqrt{ne^{\alpha_N^*}} /3)\}] \to 0$ as $N \to \infty$ due to the theorem condition (C3). The term in (S1.14) can be computed as $2\exp[\log K - (128 C_N^2)^{-1} n^2 e^{2\alpha_N^*}\lambda_{\mathcal{L}}^2\{M_t + M\sqrt{ne^{\alpha_N^*}}\lambda_{\mathcal{L}}/(24C_N)\}^{-1}] \le \exp[\log K - (ne^{\alpha_N^*})^2\lambda_{\mathcal{L}}^4/64/\log K/M_t/128(M_t + \lambda_{\mathcal{L}}^2/192\sqrt{2/C_{\text{Tail}} M_t}\sqrt{ne^{\alpha_N^*}\log N/\log K})^{-1}] \to 0$ by the theorem condition (C3). For the first term in (S1.15), we have $E\{\widetilde{\Delta}_{\text{R},k}(ne^{\alpha_N^*})^{-1}\} = E\{\exp(\varepsilon + X_i^\top\beta^* + \|X_i\|\varepsilon)\|Z_i\|^3\} = E[\exp\{X_i^\top\beta^* + (1+\|X_i\|)C_N/\sqrt{ne^{\alpha_N^*}}\}\|Z_i\|^3] \to E\{\exp(X_i^\top\beta^*)\|Z_i\|^3\}$ since $C_N/\sqrt{ne^{\alpha_N^*}} \to 0$ as $N \to \infty$. Further, we have $E\{\exp(X_i^\top\beta^*)\|Z_i\|^3\} < E(e^{t\|X_i\|})$ with a sufficiently large but fixed constant t due to (S.1) in the supplementary material of Wang (2020) and $E(e^{t\|X_i\|}) \le M_t$ by Step 1. We then have $KP[E\{(ne^{\alpha_N^*})^{-1}\widetilde{\Delta}_{\text{R},k}\} > \lambda_{\mathcal{L}}\sqrt{ne^{\alpha_N^*}}/(4C_N)] \to 0$ as $N \to \infty$. For the second term in (S1.15), we have $4\exp\big(\log K + \log N - C_{\text{Tail}}M^2\big) = 4\exp(\log K - \log N) = 4/n \to 0$ due to the theorem condition (C2). This finishes the proof.

Step 4. We aim to show in this step that $K^{-1}\sum_{k=1}^K \|\widehat{\theta}_{\text{RMLE},k} - \theta^*\|^2 = K^{-1}\sum_{k=1}^K \|(ne^{\alpha_N^*})^{-1}E(\widehat{\Sigma}_{\text{R},k})^{-1}\dot{\mathcal{L}}_{\text{R},k}(\theta^*)\|^2\{1+o_p(1)\} = O_p\{(ne^{\alpha_N^*})^{-1}\}$. Here recall $E(\widehat{\Sigma}_{\text{R},k}) = E\{e^{X_i^\top\beta^*}(1+e^{Z_i^\top\theta^*})^{-2}Z_iZ_i^\top\}$. Note that $\widehat{\theta}_{\text{RMLE},k} - \theta^* = (ne^{\alpha_N^*})^{-1}E(\widehat{\Sigma}_{\text{R},k})^{-1}\dot{\mathcal{L}}_{\text{R},k}(\theta^*) + \{\widehat{\Sigma}_{\text{R},k}^{-1} - E(\widehat{\Sigma}_{\text{R},k})^{-1}\}\dot{\mathcal{L}}_{\text{R},k}(\theta^*)/(ne^{\alpha_N^*}) + \widehat{\Sigma}_{\text{R},k}^{-1}\Gamma_{\text{R},k}/$

$(ne^{\alpha_N^*})$ by Taylor expansion. Consequently, we have $K^{-1}\sum_{k=1}^K \|\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*\|^2 = K^{-1}\sum_{k=1}^K \|(ne^{\alpha_N^*})^{-1}E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\|^2 + K^{-1}\sum_{k=1}^K \|(ne^{\alpha_N^*})^{-1}\{\widehat{\Sigma}_{\mathrm{R},k}^{-1} - E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\|^2 + K^{-1}\sum_{k=1}^K \|(ne^{\alpha_N^*})^{-1}\widehat{\Sigma}_{\mathrm{R},k}^{-1}\Gamma_{\mathrm{R},k}\|^2 + 2K^{-1}\sum_{k=1}^K n^{-2} e^{-2\alpha_N^*}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^\top E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\{\widehat{\Sigma}_{\mathrm{R},k}^{-1} - E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*) + 2K^{-1}\sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^\top (ne^{\alpha_N^*})^{-2}E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\widehat{\Sigma}_{\mathrm{R},k}^{-1}\Gamma_{\mathrm{R},k} + 2K^{-1}\sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^\top\{\widehat{\Sigma}_{\mathrm{R},k}^{-1} - E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\}\widehat{\Sigma}_{\mathrm{R},k}^{-1}\Gamma_{\mathrm{R},k}(ne^{\alpha_N^*})^{-2}$. By the Cauchy-Schwarz inequality, it suffices to compute $K^{-1}\sum_{k=1}^K \|(ne^{\alpha_N^*})^{-1}\{\widehat{\Sigma}_{\mathrm{R},k}^{-1} - E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\|^2$ and $K^{-1}\sum_{k=1}^K \|(ne^{\alpha_N^*})^{-1}\widehat{\Sigma}_{\mathrm{R},k}^{-1}\Gamma_{\mathrm{R},k}\|^2$.

STEP 4.1. We start with $K^{-1}\sum_{k=1}^K \|(ne^{\alpha_N^*})^{-1}E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\|^2$. Since $K^{-1}\sum_{k=1}^K \|(ne^{\alpha_N^*})^{-1}E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\|^2 > 0$, We then calculate

$$
\begin{aligned}
&E\left\{K^{-1}\sum_{k=1}^K \left\|\left(ne^{\alpha_N^*}\right)^{-1}E\left(\widehat{\Sigma}_{\mathrm{R},k}\right)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}\left(\theta^*\right)\right\|^2\right\} \qquad \text{(S1.16)}\\
&= \left(ne^{\alpha_N^*}\right)^{-2}K^{-1}\sum_{k=1}^K \operatorname{tr}\left[E\left(\widehat{\Sigma}_{\mathrm{R},k}\right)^{-1}E\left\{\dot{\mathcal{L}}_{\mathrm{R},k}\left(\theta^*\right)\dot{\mathcal{L}}_{\mathrm{R},k}\left(\theta^*\right)^\top\right\}E\left(\widehat{\Sigma}_{\mathrm{R},k}\right)^{-1}\right].
\end{aligned}
$$

Recall $\dot{\mathcal{L}}_{\mathrm{R},k}(\theta) = \sum_{i=1}^N a_i^{(k)}\{Y_i - p_i(\alpha_N, \beta)\}Z_i$, we have

$$
\begin{aligned}
&\left(ne^{\alpha_N^*}\right)^{-1}E\left\{\dot{\mathcal{L}}_{\mathrm{R},k}\left(\theta^*\right)\dot{\mathcal{L}}_{\mathrm{R},k}\left(\theta^*\right)^\top\right\}\\
&= \left(ne^{\alpha_N^*}\right)^{-1}NE\left[a_i^{(k)}\left\{Y_i - p_i\left(\alpha_N^*, \beta^*\right)\right\}^2 Z_i Z_i^\top\right]\\
&= \left(Kne^{\alpha_N^*}\right)^{-1}NE\left\{\frac{e^{\alpha_N^* + X_i^\top\beta^*}}{\left(1 + e^{\alpha_N^* + X_i^\top\beta^*}\right)^2}Z_i Z_i^\top\right\} = E\left(\widehat{\Sigma}_{\mathrm{R},k}\right).
\end{aligned}
$$

Together with (S1), we can compute

$$
E\left\{K^{-1}\sum_{k=1}^K \left\|\left(ne^{\alpha_N^*}\right)^{-1}E\left(\widehat{\Sigma}_{\mathrm{R},k}\right)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}\left(\theta^*\right)\right\|^2\right\} = \left(ne^{\alpha_N^*}\right)^{-1}\operatorname{tr}\left\{E\left(\widehat{\Sigma}_{\mathrm{R},k}\right)^{-1}\right\}.
$$

It is noteworthy that $(1+e^{\alpha_N^*+X_i^\top\beta^*})^{-2}e^{X_i^\top\beta^*}Z_iZ_i^\top$ converges to $e^{X_i^\top\beta^*}Z_iZ_i^\top$ almost surely as $N\to\infty$ and $(1+e^{\alpha_N^*+X_i^\top\beta^*})^{-2}e^{X_i^\top\beta^*}\|Z_i\|^2\le e^{X_i^\top\beta^*}\|Z_i\|^2$. Further, we have $E(e^{X_i^\top\beta^*}\|Z_i\|^2)\le E\{\exp(t\|X_i\|)\}\le M_t$ with sufficiently large but fixed constant t by (S.1) in Wang (2020) and Step 1. Hence, we obtain $E\{K^{-1}\sum_{k=1}^K\|(ne^{\alpha_N^*})^{-1}E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\|^2\}=(ne^{\alpha_N^*})^{-1}\mathrm{tr}\{E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\}$ $=(ne^{\alpha_N^*})^{-1}\mathrm{tr}(\Sigma^{*-1})\{1+o(1)\}$ by the dominated convergence theorem.

STEP 4.2. Next, we study $K^{-1}\sum_{k=1}^K\|\{\widehat{\Sigma}_{\mathrm{R},k}^{-1}-E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)(ne^{\alpha_N^*})^{-1}\|^2$. We can compute that $K^{-1}\sum_{k=1}^K\|\{\widehat{\Sigma}_{\mathrm{R},k}^{-1}-E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)(ne^{\alpha_N^*})^{-1}\|^2\le\max_k\|(ne^{\alpha_N^*})^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\|^2\max_k\|\widehat{\Sigma}_{\mathrm{R},k}^{-1}-E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\|^2=o_p\{(ne^{\alpha_N^*})^{-2}\log^2 K\}=o_p\{(ne^{\alpha_N^*})^{-1}\}$ by (S1.1)–(S1.2) and theorem condition (C3). Lastly, we study $K^{-1}\sum_{k=1}^K\|(ne^{\alpha_N^*})^{-1}\widehat{\Sigma}_{\mathrm{R},k}^{-1}\Gamma_{\mathrm{R},k}\|^2$. We calculate $K^{-1}\sum_{k=1}^K\|(ne^{\alpha_N^*})^{-1}\widehat{\Sigma}_{\mathrm{R},k}^{-1}\Gamma_{\mathrm{R},k}\|^2\le\max_k\|\widehat{\Sigma}_{\mathrm{R},k}^{-1}\|^2\max_k\|(ne^{\alpha_N^*})^{-1}\Gamma_{\mathrm{R},k}\|^2$. As we have $\max_k\|\widehat{\Sigma}_{\mathrm{R},k}^{-1}\|^2\le\max_k\|\widehat{\Sigma}_{\mathrm{R},k}^{-1}-\Sigma^{*-1}\|^2+\|\Sigma^{*-1}\|^2=O_p(1)$ by (S1.2), we shall focus on $\max_k\|\Gamma_{\mathrm{R},k}/(ne^{\alpha_N^*})\|^2$. Note that $\max_k\|(ne^{\alpha_N^*})^{-1}\Gamma_{\mathrm{R},k}\|^2\le\max_k\|(ne^{\alpha_N^*})^{-1}\Gamma_{\mathrm{R},k}^\Omega\|^2+\max_k\|(\Gamma_{\mathrm{R},k}-\Gamma_{\mathrm{R},k}^\Omega)/(ne^{\alpha_N^*})\|^2$. One can verify that $\max_k\|(ne^{\alpha_N^*})^{-1}\Gamma_{\mathrm{R},k}^\Omega\|^2\le(p+1)\max_j\max_k\{\Gamma_{\mathrm{R},k,j}^\Omega/(ne^{\alpha_N^*})\}^2$. Further, we have $\max_k|\Gamma_{\mathrm{R},k,j}^\Omega(ne^{\alpha_N^*})^{-1}|\le\max_k\|\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\|^2\|\Omega_{\mathrm{R},j}\|$. It suffices to compute $\max_k\|\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\|^2$ and $\|\Omega_{\mathrm{R},j}\|$ separately. By Step 3, we have $\max_k\|\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\|^2=o_p\{\log K/(ne^{\alpha_N^*})\}$. By Step 1, we know $\|\Omega_{\mathrm{R},j}\|\le\|E(e^{X_i^\top\beta^*}Z_{i,j}Z_iZ_i^\top)\|\le E(e^{X_i^\top\beta^*}\|Z_i\|^3)\le M_t$ with a sufficiently large but fixed constant t. We then have $\max_k |\Gamma^{\Omega}_{\mathrm{R},k,j}(ne^{\alpha_N^*})^{-1}| = o_p\{\log K/(ne^{\alpha_N^*})\}$. This suggests that $\max_k \|\Gamma^{\Omega}_{\mathrm{R},k}/(ne^{\alpha_N^*})\|^2 = o_p\{\log^2 K/(ne^{\alpha_N^*})^2\} = o_p\{1/(ne^{\alpha_N^*})\}$ by the theorem condition (C3). By similar arguments, we have $\max_k \|(ne^{\alpha_N^*})^{-1}(\Gamma_{\mathrm{R},k} - \Gamma^{\Omega}_{\mathrm{R},k})\|^2 = o_p\{1/(ne^{\alpha_N^*})\}$. Combining the above results, we obtain $K^{-1}\sum_{k=1}^K \|\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*\|^2 = K^{-1}\sum_{k=1}^K \|(ne^{\alpha_N^*})^{-1}E(\widehat{\Sigma}_{\mathrm{R},k})^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\|^2\{1+o_p(1)\} = O_p\{(ne^{\alpha_N^*})^{-1}\}$.

Step 5. We prove here $\sqrt{Ne^{\alpha_N^*}}Q_{\mathrm{R},1} \to_d N(0, \Sigma^{*-1})$ as $N \to \infty$. We first compute the expectation of $Q_{\mathrm{R},1}$. Recall $Q_{\mathrm{R},1} = K^{-1}\sum_{k=1}^K \ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1} \dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)$. It is obvious that $E(Q_{\mathrm{R},1}) = E\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\}$. Note that the conditional expectation $E(\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)|\mathbb{X}, \mathcal{A}) = 0$, where $\mathcal{A}$ is the σ–field of $\{a_i^{(k)} : 1 \le i \le N, 1 \le k \le K\}$ and $\mathbb{X}$ is the σ–field of $\{X_i : 1 \le i \le N\}$. As a result, we know that $E\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)|\mathbb{X}, \mathcal{A}\} = \ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}E(\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*) |\mathbb{X}, \mathcal{A}) = 0$. This leads to $E(Q_{\mathrm{R},1}) = 0$. We then compute the variance of $Q_{\mathrm{R},1}$. We can calculate $\mathrm{var}(Q_{\mathrm{R},1}) = K^{-1}\mathrm{var}\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\}$. We then focus on $\mathrm{var}\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\}$. Note that $E\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)|\mathbb{X}\} = E\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)|\mathbb{X}, \mathcal{A}\} = 0$. We then compute $\mathrm{var}\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\} = E[\mathrm{var}\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)|\mathbb{X}\}] = E[\mathrm{var}\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)|\mathbb{X}, \mathcal{A}\}|\mathbb{X}]$. Then, we calculate that $\mathrm{var}(Q_{\mathrm{R},1}) = K^{-1}E[\mathrm{var}\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)|\mathbb{X}, \mathcal{A}\}] = K^{-1}E [E\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^\top\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}|\mathbb{X}, \mathcal{A}\}]$. One can verify $E\{\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*) \dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^\top|\mathbb{X}, \mathcal{A}\} = \ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)^{-1}$. Consequently, we compute that $\mathrm{var}(Q_{\mathrm{R},1}) =$

$K^{-1}E\{\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\}^{-1} = (Kne^{\alpha_N^*})^{-1}E\{(ne^{\alpha_N^*})^{-1}\ddot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\}^{-1} = (Ne^{\alpha_N^*})^{-1}\Sigma^{*-1}$ $\{1+o(1)\}$. Then by the Central Limit Theorem (Schenker and Welsh, 1988), we have $\sqrt{Ne^{\alpha_N^*}}Q_{\mathrm{R},1} \to_d N(0,\Sigma^{*-1})$ as $N\to\infty$.

Step 6. We show here that $Q_{\mathrm{R},2}$ is a random bias term being of the order $K/(Ne^{\alpha_N^*})$ in the sense that $C_{\min} \le E\{\|B(\theta^*)\|\} \le C_{\max}$ for some fixed positive constants $C_{\min}$ and $C_{\max}$.

Step 6.1. We start with the upper bound of $E\{\|B(\theta^*)\|\}$. Recall that $Q_{\mathrm{R},2} = (ne^{\alpha_N^*})^{-1}B(\theta^*) = (ne^{\alpha_N^*})^{-1}\Sigma^{*-1}K^{-1}\sum_{k=1}^K \Gamma_{\mathrm{R},k}^{\Omega}$. We then have $\|B(\theta^*)\| \le \lambda_{\max}(\Sigma^{*-1})\|K^{-1}\sum_{k=1}^K \Gamma_{\mathrm{R},k}^{\Omega}\|$. We further compute $\Omega_{\mathrm{R},j} = E\{(ne^{\alpha_N^*})^{-1}\Delta_{\mathrm{R},k,j}^*\} = -E\{e^{X_i^\top\beta^*}(1-e^{Z_i^\top\theta^*})(1+e^{Z_i^\top\theta^*})^{-3}Z_{i,j}Z_iZ_i^\top\}$. Further, we have $\|\Omega_{\mathrm{R},j}\| \le \|E(e^{X_i^\top\beta^*}Z_{i,j}Z_iZ_i^\top)\| \le E(e^{X_i^\top\beta^*}\|Z_i\|^3) \le E\{\exp(t\|X_i\|)\}$ $\le M_t$ with sufficiently large but fixed constant t by (S.1) in Wang (2020) and Step 1. Therefore, we can calculate

$$\begin{aligned}\Big\|K^{-1}\sum_{k=1}^K \Gamma_{\mathrm{R},k}^{\Omega}\Big\| &= ne^{\alpha_N^*}\left[K^{-1}\sum_{k=1}^K\sum_{j=1}^{p+1}\left\{\left(\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\right)^\top\Omega_{\mathrm{R},j}\left(\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\right)\right\}^2\right]^{-1/2}\\ &\le \sqrt{p+1}ne^{\alpha_N^*}\max_j\left|K^{-1}\sum_{k=1}^K\left(\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\right)^\top\Omega_{\mathrm{R},j}\left(\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\right)\right|\\ &\le \sqrt{p+1}M_t\left(ne^{\alpha_N^*}K^{-1}\sum_{k=1}^K\left\|\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\right\|^2\right).\end{aligned}$$

By Step 4, we have that $K^{-1}\sum_{k=1}^K\|\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\|^2 = K^{-1}\sum_{k=1}^K\|\Sigma^{*-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)$ $\|^2(ne^{\alpha_N^*})^{-1}\{1+o_p(1)\}$ and $E\{K^{-1}\sum_{k=1}^K\|(ne^{\alpha_N^*})^{-1}\Sigma^{*-1}\dot{\mathcal{L}}_{\mathrm{R},k}(\theta^*)\|^2\} = (ne^{\alpha_N^*})^{-1}$ $\mathrm{tr}(\Sigma^{*-1})\{1+o(1)\}$. Then we define $C_{\max} = 2\sqrt{p+1}M_t\mathrm{tr}(\Sigma^{*-1})$. This leads to $E\{\|B(\theta^*)\|\} \leq C_{\max}$.

STEP 6.2. We next study the lower bound of $E\{\|B(\theta^*)\|\}$. One can verify that $E\{\|B(\theta^*)\|\} \geq \lambda_{\max}^{-1}(\Sigma^*)E(\|K^{-1}\sum_{k=1}^K \Gamma_{\mathrm{R},k}^{\Omega}\|)$ and $E(\|K^{-1}\sum_{k=1}^K \Gamma_{\mathrm{R},k}^{\Omega}\|) = E[\sum_{j=1}^{p+1}\{K^{-1}\sum_{k=1}^K ne^{\alpha_N^*}(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)^\top \Omega_{\mathrm{R},j}(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)\}^2]^{-1/2} \geq \min_j E\{|K^{-1}\sum_{k=1}^K ne^{\alpha_N^*}(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)^\top \Omega_{\mathrm{R},j}(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)|\}\sqrt{p+1}$. Here $ne^{\alpha_N^*}(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)^\top \Omega_{\mathrm{R},j}(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*) \to_d \sum_i \kappa_{ji}\chi_i^2(1)$, where κ_{ji} is the ith eigenvalue of $\Sigma^{*1/2}\Omega_{\mathrm{R},j}\Sigma^{*1/2}$ and $\chi_i^2(1)$ stands for the ith Chi-square distribution with one degree of freedom. Define $C_{\min} = \sqrt{p+1}\lambda_{\max}^{-1}(\Sigma^*)\min_j E\{|K^{-1}\sum_{k=1}^K ne^{\alpha_N^*}(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)^\top \Omega_{\mathrm{R},j}(\widehat{\theta}_{\mathrm{RMLE},k} - \theta^*)|\}/2$. This leads to $E\{\|B(\theta^*)\|\} \geq C_{\min}$. This accomplishes the proof.

STEP 7. In this step, we demonstrate that $Q_{\mathrm{R},3} = o_p\{(ne^{\alpha_N^*})^{-1}\}$. Recall $Q_{\mathrm{R},3} = K^{-1}\sum_{k=1}^K(\widehat{\Sigma}_{\mathrm{R},k}^{-1} - \Sigma^{*-1})\{\Gamma_{\mathrm{R},k}^*/(ne^{\alpha_N^*})\}$. We then have

$$\left\|Q_{\mathrm{R},3}\right\| \leq \frac{1}{K}\sum_{k=1}^K \left\|\widehat{\Sigma}_{\mathrm{R},k}^{-1} - \Sigma^{*-1}\right\| \left\|\frac{\Gamma_{\mathrm{R},k}^*}{ne^{\alpha_N^*}}\right\| \leq Q_{3a}Q_{3b},$$

where $Q_{\mathrm{R},3a} = \max_k \|\widehat{\Sigma}_{\mathrm{R},k}^{-1} - \Sigma^{*-1}\|$ and $Q_{\mathrm{R},3b} = K^{-1}\sum_{k=1}^K \|\Gamma_{\mathrm{R},k}^*/(ne^{\alpha_N^*})\|$. By applying (S1.2), we have verified $Q_{\mathrm{R},3a} = o_p(\sqrt{\log K}/\sqrt{ne^{\alpha_N^*}}) + O(e^{\alpha_N^*}) = o_p(1)$ by the theorem condition (C3) and (2.2). We then focus on $Q_{\mathrm{R},3b}$. We have

$$Q_{\mathrm{R},3b} \leq \sum_{k=1}^K \left\|\left(\Gamma_{\mathrm{R},k}^* - \Gamma_{\mathrm{R},k}^{\Omega}\right)/\left(ne^{\alpha_N^*}\right)\right\|/K + \sum_{k=1}^K \left\|\Gamma_{\mathrm{R},k}^{\Omega}/\left(ne^{\alpha_N^*}\right)\right\|/K. \tag{S1.17}$$

For the first part of (S1.17), we have $K^{-1}\sum_{k=1}^{K}\|(\Gamma^*_{\mathrm{R},k}-\Gamma^{\Omega}_{\mathrm{R},k})/(ne^{\alpha^*_N})\| \leq \sqrt{p+1}\max_j K^{-1}\sum_{k=1}^{K}|\Gamma^*_{\mathrm{R},k,j}-\Gamma^{\Omega}_{\mathrm{R},k,j}|/(ne^{\alpha^*_N})$. Consequently, it suffices to compute $K^{-1}\sum_{k=1}^{K}|\Gamma^*_{\mathrm{R},k,j}-\Gamma^{\Omega}_{\mathrm{R},k,j}|/(ne^{\alpha^*_N})$. Note that $K^{-1}\sum_{k=1}^{K}|\Gamma^*_{\mathrm{R},k,j} - \Gamma^{\Omega}_{\mathrm{R},k,j}|/(ne^{\alpha^*_N}) \leq \max_k\|\Delta^*_{\mathrm{R},k,j}(ne^{\alpha^*_N})^{-1}-\Omega_{\mathrm{R},j}\|K^{-1}\sum_{k=1}^{K}\|\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\|^2 = o_p(\sqrt{\log K}/\sqrt{ne^{\alpha^*_N}})O_p\{(ne^{\alpha^*_N})^{-1}\} = o_p\{(ne^{\alpha^*_N})^{-1}\}$ by (S1.3), the theorem condition (C3) and Step 4. For the second part of (S1.17), we have $K^{-1}\sum_{k=1}^{K}\|\Gamma^{\Omega}_{\mathrm{R},k}/(ne^{\alpha^*_N})\| \leq \|\Omega_{\mathrm{R},j}\|K^{-1}\sum_{k=1}^{K}\|\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\|^2 \leq O_p\{(ne^{\alpha^*_N})^{-1}\}M_t = O_p\{(ne^{\alpha^*_N})^{-1}\}$ by Step 4. Here $\|\Omega_{\mathrm{R},j}\| \leq \|E(e^{X_i^\top\beta^*}Z_{i,j}Z_iZ_i^\top)\| \leq E(e^{X_i^\top\beta^*}\|Z_i\|^3) \leq E\{\exp(t\|X_i\|)\} \leq M_t$ with sufficiently large but fixed constant t by (S.1) in Wang (2020) and Step 1. Therefore, we have $Q_{3b} = O_p\{(ne^{\alpha^*_N})^{-1}\}$. This finally leads to $Q_{\mathrm{R},3} = o_p\{(ne^{\alpha^*_N})^{-1}\}$.

Step 8. Lastly, we show in this step that $Q_{\mathrm{R},4} = o_p\big\{(ne^{\alpha^*_N})^{-1}\big\}$. Recall $Q_{\mathrm{R},4} = K^{-1}\sum_{k=1}^{K}\widehat{\Sigma}^{-1}_{\mathrm{R},k}\{(\Gamma_{\mathrm{R},k}-\Gamma^*_{\mathrm{R},k})/(ne^{\alpha^*_N})\}$. We then have $\|Q_{\mathrm{R},4}\| \leq Q_{4a}Q_{4b}$, where $Q_{\mathrm{R},4a} = \max_k\|\widehat{\Sigma}^{-1}_{\mathrm{R},k}\|$ and $Q_{\mathrm{R},4b} = K^{-1}\sum_{k=1}^{K}\|(\Gamma_{\mathrm{R},k}-\Gamma^*_{\mathrm{R},k})(ne^{\alpha^*_N})^{-1}\|$. Note that $Q_{\mathrm{R},4a} = \max_k\|\widehat{\Sigma}^{-1}_{\mathrm{R},k}\| \leq \max_k\|\widehat{\Sigma}^{-1}_{\mathrm{R},k}-\Sigma^{*-1}\| + \|\Sigma^{*-1}\| = O_p(1)$ by (S1.2). We then focus on $Q_{\mathrm{R},4b}$. Since we have $Q_{\mathrm{R},4b} \leq \sqrt{p+1}\max_j K^{-1}\sum_{k=1}^{K}|\Gamma_{\mathrm{R},k,j}-\Gamma^*_{\mathrm{R},k,j}|/(ne^{\alpha^*_N})$, it suffices to compute $\sum_{k=1}^{K} K^{-1}|\Gamma_{\mathrm{R},k,j}-\Gamma^*_{\mathrm{R},k,j}|/(ne^{\alpha^*_N})$. Note that $K^{-1}\sum_{k=1}^{K}|\Gamma_{\mathrm{R},k,j}-\Gamma^*_{\mathrm{R},k,j}|/(ne^{\alpha^*_N}) \leq \max_k(ne^{\alpha^*_N})^{-1}\|\Delta_{\mathrm{R},k,j}-\Delta^*_{\mathrm{R},k,j}\|K^{-1}\sum_{k=1}^{K}\|\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\|^2$. Further, we have $\max_k(ne^{\alpha^*_N})^{-1}\|\Delta_{\mathrm{R},k,j}-\Delta^*_{\mathrm{R},k,j}\| \leq \max_k(ne^{\alpha^*_N})^{-1}\|\Delta_{\mathrm{R},k,j}-E(\Delta_{\mathrm{R},k,j})\|+\|(ne^{\alpha^*_N})^{-1}$

$E(\Delta_{\mathrm{R},k,j})-\Omega_{\mathrm{R},j}\|+\max_k \|(ne^{\alpha_N^*})^{-1}\Delta^*_{\mathrm{R},k,j}-\Omega_{\mathrm{R},j}\| = o_p(1)$ by (S1.3) and the definition of $\Delta_{\mathrm{R},k,j}$. In the meanwhile, we have $K^{-1}\sum_{k=1}^K \|\widehat{\theta}_{\mathrm{RMLE},k}-\theta^*\|^2 = O_p\big\{(ne^{\alpha_N^*})^{-1}\big\}$ by Step 4. This leads to $Q_{\mathrm{R},4} = o_p\big\{(ne^{\alpha_N^*})^{-1}\big\}$. We next prove that $Q_{\mathrm{R},5} = o_p\big\{(ne^{\alpha_N^*})^{-1}\big\}$. Recall that $Q_{\mathrm{R},5} = (ne^{\alpha_N^*})^{-1}\Sigma^{*-1}K^{-1}\sum_{k=1}^K (\Gamma^*_{\mathrm{R},k}-\Gamma^{\Omega}_{\mathrm{R},k})$. We have verified that $K^{-1}\sum_{k=1}^K \|(\Gamma^*_{\mathrm{R},k}-\Gamma^{\Omega}_{\mathrm{R},k})/(ne^{\alpha_N^*})\| = o_p\{(ne^{\alpha_N^*})^{-1}\}$ in Step 7. Then we have $\|Q_{\mathrm{R},5}\| \le \|\Sigma^{*-1}\|K^{-1}\sum_{k=1}^K \|(\Gamma^*_{\mathrm{R},k}-\Gamma^{\Omega}_{\mathrm{R},k})/(ne^{\alpha_N^*})\| = o_p\{(ne^{\alpha_N^*})^{-1}\}$. This completes the entire theorem proof.

S2 Proof of Theorem 2

Define $\widetilde{\theta}^* = \theta^* + \flat$, where $\flat = \big(\log K, 0, \cdots, 0\big) \in \mathbb{R}^{p+1}$. For each $\widehat{\theta}_{\mathrm{US},k}(1 \le k \le K)$, we apply the Taylor expansion and have

$$\dot{\mathcal{L}}_{\mathrm{US},k}\big(\widehat{\theta}_{\mathrm{US},k}\big) = \dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}_{\mathrm{US},k}\big) + \ddot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}_{\mathrm{US},k}\big)\Big(\widehat{\theta}_{\mathrm{US},k}-\widetilde{\theta}^*\Big),$$

where $\Gamma_{\mathrm{US},k} = (\Gamma_{\mathrm{US},k,j}) \in \mathbb{R}^{p+1}$, $\Gamma_{\mathrm{US},k,j} = (\widehat{\theta}_{\mathrm{US},k}-\widetilde{\theta}^*)^\top \Delta_{\mathrm{US},k,j}(\widehat{\theta}_{\mathrm{US},k}-\widetilde{\theta}^*)$, $\Delta_{\mathrm{US},k,j} = \big(\Delta^{(j_1 j_2)}_{\mathrm{US},k,j}\big) \in \mathbb{R}^{(p+1)\times(p+1)}$ for $1 \le j \le p+1$, $\Delta^{(j_1 j_2)}_{\mathrm{US},k,j} = \partial^2 \dot{\mathcal{L}}_{\mathrm{US},k,j}(\theta)/\partial\theta_{j_1}\partial\theta_{j_2}|_{\theta=\widetilde{\theta}_{\mathrm{US},k}}$ for $1 \le j_1, j_2 \le p+1$, $\dot{\mathcal{L}}_{\mathrm{US},k,j}(\theta)$ is the jth element of $\dot{\mathcal{L}}_{\mathrm{US},k}(\theta)$, and $\widetilde{\theta}_{\mathrm{US},k} = \eta_k\widehat{\theta}_{\mathrm{US},k}+(1-\eta_k)\widetilde{\theta}^*$ for some $0 \le \eta_k \le 1$. In the meanwhile, we define $\Gamma^*_{\mathrm{US},k} = (\Gamma^*_{\mathrm{US},k,j})$, $\Gamma^*_{\mathrm{US},k,j} = (\widehat{\theta}_{\mathrm{US},k}-\theta^*)^\top\Delta^*_{\mathrm{US},k,j}(\widehat{\theta}_{\mathrm{US},k}-\theta^*)$ and $\Delta^*_{\mathrm{US},k,j} = \partial^2\dot{\mathcal{L}}_{\mathrm{US},k,j}(\theta)/\partial\theta_{j_1}\partial\theta_{j_2}|_{\theta=\theta^*}$. We also define $\Gamma^{\Omega}_{\mathrm{US},k} = (\Gamma^{\Omega}_{\mathrm{US},k,j}) \in \mathbb{R}^{p+1}$ with $1 \le j \le p+1$, $\Gamma^{\Omega}_{\mathrm{US},k,j} = ne^{\alpha_N^*}(\widehat{\theta}_{\mathrm{US},k}-\theta^*)^\top\Omega_{\mathrm{US},j}(\widehat{\theta}_{\mathrm{US},k}-\theta^*)$

and $\Omega_{\mathrm{US},j} = E\{\Delta^*_{\mathrm{US},k,j}/(ne^{\alpha^*_N})\}$. More specifically, by definition we have $\mathcal{L}_{\mathrm{US},k}(\theta) = \sum_{i=1}^N \big\{Y_i Z_i^\top \theta - \big(a_i^{(k)} + Y_i - a_i^{(k)} Y_i\big) \log\big(1 + e^{Z_i^\top \theta}\big)\big\}$. Thus, we have $\dot{\mathcal{L}}_{\mathrm{US},k}(\theta) = \sum_{i=1}^N \big\{Y_i - (a_i^{(k)} + Y_i - a_i^{(k)} Y_i) p_i(\alpha_N, \beta)\big\} Z_i$, $\ddot{\mathcal{L}}_{\mathrm{US},k}(\theta) = -\sum_{i=1}^N \big(a_i^{(k)} + Y_i - a_i^{(k)} Y_i\big) p_i(\alpha_N, \beta)\big\{1 - p_i(\alpha_N, \beta)\big\} Z_i Z_i^\top$. and $\Delta_{\mathrm{US},k,j} = -\sum_{i=1}^N (a_i^{(k)} + Y_i - a_i^{(k)} Y_i) p_i(\alpha_N, \beta)\{1 - p_i(\alpha_N, \beta)\}\{1 - 2p_i(\alpha_N, \beta)\} Z_{i,j} Z_i Z_i^\top$, respectively. Then, we have

$$\widehat{\theta}_{\mathrm{US},k} - \widetilde{\theta}^* = -\Big\{\ddot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big)\Big\}^{-1} \dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big) - \frac{1}{2}\Big\{\ddot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big)\Big\}^{-1} \Gamma_{\mathrm{US},k}.$$

Subsequently, we should decompose the US estimator into a total of six different terms as follows

$$\widehat{\theta}_{\mathrm{US}} - \theta^* = Q_{\mathrm{US},1} + Q_{\mathrm{US},2} + Q_{\mathrm{US},3}/2 + Q_{\mathrm{US},4}/2 + Q_{\mathrm{US},5}/2 + Q_{\mathrm{US},6}/2,$$

where $Q_{\mathrm{US},1} = K^{-1} \Sigma_2^{*-1} \sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{US},k}(\widetilde{\theta}^*)/(Ne^{\alpha^*_N})$, $Q_{\mathrm{US},2} = K^{-1} \sum_{k=1}^K (\widehat{\Sigma}^{-1}_{\mathrm{US},k} - \Sigma_2^{*-1})(Ne^{\alpha^*_N})^{-1} \dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big)$, $Q_{\mathrm{US},3} = K^{-1} \sum_{k=1}^K \Sigma_2^{*-1}\{\Gamma^\Omega_{\mathrm{US},k}/(Ne^{\alpha^*_N})\}$, $Q_{\mathrm{US},4} = K^{-1} \sum_{k=1}^K (\widehat{\Sigma}^{-1}_{\mathrm{US},k} - \Sigma_2^{*-1})\{\Gamma^*_{\mathrm{US},k}/(Ne^{\alpha^*_N})\}$, $Q_{\mathrm{US},5} = K^{-1} \sum_{k=1}^K \widehat{\Sigma}^{-1}_{\mathrm{US},k}\{(\Gamma_{\mathrm{US},k} - \Gamma^*_{\mathrm{US},k})/(Ne^{\alpha^*_N})\}$, $Q_{\mathrm{US},6} = K^{-1} \sum_{k=1}^K \Sigma_2^{*-1}\{(\Gamma^*_{\mathrm{US},k} - \Gamma^\Omega_{\mathrm{US},k})/(Ne^{\alpha^*_N})\}$, $\widehat{\Sigma}_{\mathrm{US},k} = -\ddot{\mathcal{L}}_{\mathrm{US},k}\big(\theta^*\big)/(Ne^{\alpha^*_N})$ and $\Sigma_2^* = E\big\{e^{X_i^\top \beta^*}(1 + \gamma e^{X_i^\top \beta^*})^{-1} Z_i Z_i^\top\big\}$ with $\gamma = \lim_{N\to\infty} Ke^{\alpha^*_N}$. We would investigate the asymptotic property of the US estimator in the following two steps. In the 1st step, we show that $\sqrt{Ne^{\alpha^*_N}} Q_{\mathrm{US},1} \to_d N\big(0, \Sigma_2^{*-1}\Sigma_1^*\Sigma_2^{*-1}\big)$ as $N \to \infty$, where $\Sigma_1^* = E\big\{\big(1+\gamma e^{X_i^\top \beta^*}\big)^{-2} e^{X_i^\top \beta^*} Z_i Z_i^\top\big\}$. In the 2nd step, we demonstrate that $Q_{\mathrm{US},2} = o_p\{(Ne^{\alpha^*_N})^{-1/2}\}$. The technical details for computing $Q_{\mathrm{US},3}$–$Q_{\mathrm{US},6}$ are very similar with those of $Q_{\mathrm{R},3}$–$Q_{\mathrm{R},6}$. We thus provide the proof details for $Q_{\mathrm{US},1}$ and $Q_{\mathrm{US},2}$ only. The proofs for $Q_{\mathrm{US},3}$–$Q_{\mathrm{US},6}$ are omitted to save space.

STEP 1. In this step, we verify $\sqrt{Ne^{\alpha_N^*}}Q_{\mathrm{US},1} \to_d N\big(0, \Sigma_2^{*-1}\Sigma_1^*\Sigma_2^{*-1}\big)$ as $N \to \infty$. To this end, we calculate the mean and covariance of $(Ne^{\alpha_N^*})^{-1/2}K^{-1}\sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big)$, respectively. We start with $E\big\{\big(Ne^{\alpha_N^*}\big)^{-1/2}K^{-1}\sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big)\big\}$. Recall $\dot{\mathcal{L}}_{\mathrm{US},k}(\theta) = \sum_{i=1}^N \big\{Y_i - (a_i^{(k)} + Y_i - a_i^{(k)}Y_i)p_i(\alpha_N, \beta)\big\}Z_i$ and $\sum_{k=1}^K a_i^{(k)} = 1$. We then have $(Ne^{\alpha_N^*})^{-1/2}K^{-1}\sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big) = (Ne^{\alpha_N^*})^{-1/2}\sum_{i=1}^N \big\{Y_i - (K^{-1} + Y_i - K^{-1}Y_i)p_i(\alpha_N, \beta)\big\}Z_i$. Furthermore, since we have

$$
\begin{aligned}
&E\Big[\Big\{Y_i - \Big(K^{-1} + Y_i - K^{-1}Y_i\Big)p_i(\alpha_N^* + \log K, \beta^*)\Big\}\Big] \\
&= E\Big[E\Big\{Y_i\Big(1 - p_i\big(\alpha_N^* + \log K, \beta^*\big)\Big) - K^{-1}\Big(1 - Y_i\Big)p_i\big(\alpha_N^* + \log K, \beta^*\big)\big|Z_i\Big\}\Big] \\
&= E\Big[p_i\big(\alpha_N^*, \beta^*\big)\Big\{1 - p_i\big(\alpha_N^* + \log K, \beta^*\big)\Big\}\Big] \\
&\quad - E\Big[K^{-1}\Big\{1 - p_i\big(\alpha_N^*, \beta^*\big)\Big\}p_i\big(\alpha_N^* + \log K, \beta^*\big)\Big] \\
&= E\Bigg[\Big\{1 - p_i\big(\alpha_N^* + \log K, \beta^*\big)\Big\}\Big\{1 - p_i\big(\alpha_N^*, \beta\big)\Big\} \\
&\quad \left\{\frac{p_i(\alpha_N^*, \beta^*)}{1 - p_i(\alpha_N^*, \beta^*)} - \frac{p_i(\alpha_N^* + \log K, \beta^*)/K}{1 - p_i(\alpha_N^* + \log K, \beta^*)}\right\}\Bigg] = 0.
\end{aligned}
$$

The last equality is because $p_i(\alpha_N^*, \beta^*)/\{1 - p_i(\alpha_N^*, \beta^*)\} = e^{Z_i^\top\theta^*}$ and $p_i(\alpha_N^* + \log K, \beta^*)\{1 - p_i(\alpha_N^* + \log K, \beta^*)\}^{-1} = e^{Z_i^\top\theta^* + \log K}$. Hence, we calculate $E\{(Ne^{\alpha_N^*})^{-1/2}K^{-1}\sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{US},k}(\widetilde{\theta}^*)g\} = 0$.

Next, we compute the covariance of $(Ne^{\alpha_N^*})^{-1/2}K^{-1}\sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big)$ as

$$\mathrm{cov}\Big\{\big(Ne^{\alpha_N^*}\big)^{-1/2}K^{-1}\sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big)\Big\}$$

$$= e^{-\alpha_N^*}\mathrm{cov}\Big[\Big\{Y_i\big(1-p_i\big(\alpha_N^*+\log K,\beta^*\big)\big)-K^{-1}\big(1-Y_i\big)p_i\big(\alpha_N^*+\log K,\beta^*\big)\Big\}Z_i\Big]$$

$$= e^{-\alpha_N^*}E\Big[\Big\{Y_i\big(1-p_i\big(\alpha_N+\log K,\beta^*\big)\big)^2$$

$$+K^{-2}\big(1-Y_i\big)p_i^2\big(\alpha_N^*+\log K,\beta^*\big)\Big\}Z_iZ_i^\top\Big]$$

$$= e^{-\alpha_N^*}E\Big[p_i\big(\alpha_N^*,\beta^*\big)\big\{1-p_i\big(\alpha_N^*+\log K,\beta^*\big)\big\}^2 Z_iZ_i^\top$$

$$+K^{-2}\big\{1-p_i\big(\alpha_N^*,\beta^*\big)\big\}p_i^2\big(\alpha_N^*+\log K,\beta^*\big)Z_iZ_i^\top\Big]$$

$$= e^{-\alpha_N^*}E\Big[\Big\{p_i\big(\alpha_N^*,\beta^*\big)\big\{1-p_i\big(\alpha_N^*+\log K,\beta^*\big)\big\}^2\big\{1-p_i\big(\alpha_N^*,\beta^*\big)\big\}^{-1}Z_iZ_i^\top\Big]$$

$$= E\left\{\frac{e^{X_i^\top\beta^*}}{\big(1+e^{X_i^\top\beta^*+\alpha_N^*+\log K}\big)^2}Z_iZ_i^\top\right\}.$$

One can verify that $\big(1+e^{X_i^\top\beta^*+\alpha_N^*+\log K}\big)^{-2}e^{X_i^\top\beta^*}Z_iZ_i^\top$ converges to $\big(1+\gamma e^{X_i^\top\beta^*}\big)^{-2}e^{X_i^\top\beta^*}Z_iZ_i^\top$ almost surely as $N\to\infty$, and $\big(1+e^{X_i^\top\beta^*+\alpha_N^*+\log K}\big)^{-2}$ $e^{X_i^\top\beta^*}\|Z_i\|^2 \le e^{X_i^\top\beta^*}\|Z_i\|^2$. Here $E\big(e^{X_i^\top\beta^*}\|Z_i\|^2\big)\le M_t$ with sufficiently large but fixed constant t by (S.1) in Wang (2020) and Step 1 in S1. Then by the dominated convergence theorem (Royden and Fitzpatrick, 1988), we have

$$\mathrm{cov}\left\{\frac{1}{K\sqrt{Ne^{\alpha_N^*}}}\sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big)\right\}$$

$$= E\left\{\frac{e^{X_i^\top\beta^*}}{\big(1+e^{X_i^\top\beta^*+\alpha_N^*+\log K}\big)^2}Z_iZ_i^\top\right\} = \Sigma_1^*\Big\{1+o(1)\Big\},$$

where $\Sigma_1^* = E\{e^{X_i^\top\beta^*}(1+\gamma e^{X_i^\top\beta^*})^{-2}Z_iZ_i^\top\}$.

Lastly, we check the Lindeberg-Feller condition (Van der Vaart, 2000). Recall $\dot{\mathcal{L}}_{\mathrm{US},k}(\theta) = \sum_{i=1}^N \big\{Y_i - (a_i^{(k)} + Y_i - a_i^{(k)} Y_i) p_i(\alpha_N, \beta)\big\} Z_i$ and $\sum_k a_i^{(k)} = 1$. We then have $K^{-1}\sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big) = \sum_{i=1}^N \big\{Y_i - (K^{-1}\sum_{k=1}^K a_i^{(k)} + Y_i - K^{-1}\sum_{k=1}^K a_i^{(k)} Y_i) p_i(\alpha_N, \beta)\big\} Z_i = \sum_{i=1}^N \big\{Y_i - \big(K^{-1} + Y_i - K^{-1} Y_i\big) p_i\big(\alpha_N^* + \log K, \beta^*\big)\big\} Z_i$. Then for any $\varepsilon > 0$, we have

$$\sum_{i=1}^N E\Big[\Big\|\big\{Y_i - \big(K^{-1} + Y_i - K^{-1}Y_i\big)p_i\big(\widetilde{\theta}^*\big)\big\}Z_i\Big\|^2$$
$$I\Big(\big\|\big\{Y_i - \big(K^{-1} + Y_i - K^{-1}Y_i\big)p_i\big(\widetilde{\theta}^*\big)\big\}Z_i\big\| > \sqrt{Ne^{\alpha_N^*}}\varepsilon\Big)\Big]$$
$$= NE\Big[p_i\big(\theta^*\big)\Big\|\big\{1 - p_i\big(\widetilde{\theta}^*\big)\big\}Z_i\Big\|^2 I\Big(\big\|\big\{1 - p_i\big(\widetilde{\theta}^*\big)\big\}Z_i\big\| > \sqrt{Ne^{\alpha_N^*}}\varepsilon\Big)\Big]$$
$$+NE\Big[\big\{1 - p_i\big(\theta^*\big)\big\}\Big\|K^{-1}p_i\big(\widetilde{\theta}^*\big)Z_i\Big\|^2 I\Big(\big\|K^{-1}p_i\big(\widetilde{\theta}^*\big)Z_i\big\| > \sqrt{Ne^{\alpha_N^*}}\varepsilon\Big)\Big]$$
$$\le NE\Big\{p_i\big(\theta^*\big)\|Z_i\|^2 I\Big(\|Z_i\| > \sqrt{Ne^{\alpha_N^*}}\varepsilon\Big)\Big\}$$
$$+Ne^{2\alpha_N^*}E\Big[\Big\|e^{X_i^\top\beta^*}Z_i\Big\|^2 I\Big(\big\|Z_i\big\| > \sqrt{Ne^{\alpha_N^*}}\varepsilon\Big)\Big]$$
$$\le Ne^{\alpha_N^*}E\Big\{e^{X_i^\top\beta^*}\big\|Z_i\big\|^2 I\Big(\big\|Z_i\big\| > \sqrt{Ne^{\alpha_N^*}}\varepsilon\Big)\Big\}$$
$$+Ne^{2\alpha_N^*}E\Big\{e^{2X_i^\top\beta^*}\big\|Z_i\big\|^2 I\Big(\big\|Z_i\big\| > \sqrt{Ne^{\alpha_N^*}}\varepsilon\Big)\Big\} = o\Big(Ne^{\alpha_N^*}\Big).$$

The last step is because of the dominated convergence theorem (Royden and Fitzpatrick, 1988). Therefore, by applying the Lindeberg Feller Central Limit Theorem (Van der Vaart, 2000), we accomplish the proof of Step 1.

Step 2. Recall $Q_{\mathrm{US},2} = K^{-1}\sum_{k=1}^K(\widehat{\Sigma}_{\mathrm{US},k}^{-1} - \Sigma_2^{*-1})\dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big)/(Ne^{\alpha_N^*})$. We prove here $Q_{\mathrm{US},2} = o_p\{(Ne^{\alpha_N^*})^{-1/2}\}$. We have $\|Q_{\mathrm{US},2}\| \le Q_{2a}Q_{2b}$, where $Q_{2a} = \max_k \|\widehat{\Sigma}_{\mathrm{US},k}^{-1} - \Sigma_2^{*-1}\|$ and $Q_{2b} = K^{-1}\sum_{k=1}^K \|\dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big)/(Ne^{\alpha_N^*})\|$. By the similar arguments used by (S1.2), we have $Q_{2a} = o_p(1)$. We then focus on proving $Q_{2b} = O_p\{(Ne^{\alpha_N^*})^{-1/2}\}$. By Jensen's inequality, we have $\{K^{-1} \sum_{k=1}^K \|(Ne^{\alpha_N^*})^{-1}\dot{\mathcal{L}}_{\text{US},k}(\widetilde{\theta}^*)\|\}^2 \leq K^{-1}\sum_{k=1}^K \|(Ne^{\alpha_N^*})^{-1}\dot{\mathcal{L}}_{\text{US},k}(\widetilde{\theta}^*)\|^2$. Moreover, it is obvious that $K^{-1}\sum_{k=1}^K \|(Ne^{\alpha_N^*})^{-1}\dot{\mathcal{L}}_{\text{US},k}(\widetilde{\theta}^*)\|^2 \geq 0$. Then, it suffices to calculate $E\{K^{-1}\sum_{k=1}^K \|(Ne^{\alpha_N^*})^{-1}\dot{\mathcal{L}}_{\text{US},k}(\widetilde{\theta}^*)\|^2\}$. We have

$$E\left\{K^{-1}\sum_{k=1}^K \left\|\left(Ne^{\alpha_N^*}\right)^{-1}\dot{\mathcal{L}}_{\text{US},k}\left(\widetilde{\theta}^*\right)\right\|^2\right\}$$
$$= K^{-1}\sum_{k=1}^K \text{tr}\left[E\left\{\left(Ne^{\alpha_N^*}\right)^{-2}\dot{\mathcal{L}}_{\text{US},k}\left(\widetilde{\theta}^*\right)\dot{\mathcal{L}}_{\text{US},k}\left(\widetilde{\theta}^*\right)^\top\right\}\right]. \tag{S2.18}$$

Furthermore, we have

$$E\left\{\left(Ne^{\alpha_N^*}\right)^{-2}\dot{\mathcal{L}}_{\text{US},k}\left(\widetilde{\theta}^*\right)\dot{\mathcal{L}}_{\text{US},k}\left(\widetilde{\theta}^*\right)^\top\right\} = \text{cov}\left\{\left(Ne^{\alpha_N^*}\right)^{-1}\dot{\mathcal{L}}_{\text{US},k}\left(\widetilde{\theta}^*\right)\right\}$$
$$= N^{-1}e^{-2\alpha_N^*}\text{cov}\Big[\Big\{Y_i\Big(1 - p_i\big(\alpha_N^* + \log K, \beta^*\big)\Big)$$
$$-a_i^{(k)}\big(1 - Y_i\big)p_i\big(\alpha_N^* + \log K, \beta^*\big)\Big\}Z_i\Big]$$
$$= N^{-1}e^{-2\alpha_N^*}E\Big[\Big\{Y_i\Big(1 - p_i\big(\alpha_N^* + \log K, \beta^*\big)\Big)^2$$
$$+a_i^{(k)}\Big(1 - Y_i\Big)p_i^2\big(\alpha_N^* + \log K, \beta^*\big)\Big\}Z_iZ_i^\top\Big]$$

$$= N^{-1}e^{-2\alpha_N^*}E\Big[\Big\{p_i\big(\alpha_N^*, \beta^*\big)\Big(1 - p_i\big(\alpha_N^* + \log K, \beta^*\big)\Big)^2$$
$$+K^{-1}\Big(1 - p_i\big(\alpha_N^*, \beta^*\big)\Big)p_i^2\big(\alpha_N^* + \log K, \beta^*\big)\Big\}Z_iZ_i^\top\Big]$$
$$= \left(Ne^{\alpha_N^*}\right)^{-1}E\left\{\frac{e^{X_i^\top\beta^*}}{(1 + e^{X_i^\top\beta^* + \alpha_N^*})(1 + e^{X_i^\top\beta^* + \alpha_N^* + \log K})}Z_iZ_i^\top\right\}. \tag{S2.19}$$

Combining (S2.18) and (S2.19), we have

$$E\left\{K^{-1}\sum_{k=1}^{K}\left\|\left(Ne^{\alpha_N^*}\right)^{-1}\dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big)\right\|^2\right\}$$
$$=\left(Ne^{\alpha_N^*}\right)^{-1}E\left\{\frac{e^{X_i^\top\beta^*}}{(1+e^{X_i^\top\beta^*+\alpha_N^*})(1+e^{X_i^\top\beta^*+\alpha_N^*+\log K})}\|Z_i\|^2\right\}.$$

Note that $(1+e^{X_i^\top\beta^*+\alpha_N^*})^{-1}(1+e^{X_i^\top\beta^*+\alpha_N^*+\log K})^{-1}e^{X_i^\top\beta^*}\|Z_i\|^2$ converges to $(1+\gamma e^{X_i^\top\beta^*})^{-1}e^{X_i^\top\beta^*}\|Z_i\|^2$ almost surely as $N\to\infty$, and $(1+e^{X_i^\top\beta^*+\alpha_N^*+\log K})^{-1}$ $(1+e^{X_i^\top\beta^*+\alpha_N^*})^{-1}e^{X_i^\top\beta^*}\|Z_i\|^2 \le e^{X_i^\top\beta^*}\|Z_i\|^2$. Here $E\big(e^{X_i^\top\beta^*}\|Z_i\|^2\big)\le M_t$ with sufficiently large but fixed constant t by (S.1) in Wang (2020) and Step 1 in S1. Thus, from the dominated convergence theorem (Royden and Fitzpatrick, 1988), the last equality holds. We then obtain $K^{-1}\sum_{k=1}^K\|(Ne^{\alpha_N^*})^{-1}$ $\dot{\mathcal{L}}_{\mathrm{US},k}\big(\widetilde{\theta}^*\big)\|^2=O_p\{(Ne^{\alpha_N^*})^{-1}\}$. This suggests that $Q_{2b}=O_p\{(Ne^{\alpha_N^*})^{-1/2}\}$. This finally leads to $Q_{\mathrm{US},2}=o_p\{(Ne^{\alpha_N^*})^{-1/2}\}$.

In addition, to demonstrate that $\widehat{\theta}_{\mathrm{US}}$ is less efficient, we need to prove $\Sigma_2^{*-1}\Sigma_1\Sigma_2^{*-1}-\Sigma^{*-1}$ is semi-positive definite. Denote $g=\sqrt{e^{X_i^\top\beta^*}}\Sigma^{*-1}Z_i-(1+\gamma e^{X_i^\top\beta^*})^{-1}\sqrt{e^{X_i^\top\beta^*}}\Sigma_2^{*-1}Z_i$. Subsequently, we have $E\left(gg^\top\right)=\Sigma^{*-1}E(e^{X_i^\top\beta^*}$ $Z_iZ_i^\top)\Sigma^{*-1}+\Sigma_2^{*-1}E\big\{(1+\gamma e^{X_i^\top\beta^*})^{-2}e^{X_i^\top\beta^*}Z_iZ_i^\top\big\}\Sigma_2^{*-1}-2\Sigma_2^{*-1}E\big\{e^{X_i^\top\beta^*}(1+$ $\gamma e^{X_i^\top\beta^*})^{-1}Z_iZ_i^\top\big\}\Sigma^{*-1}=\Sigma^{*-1}\Sigma^*\Sigma^{*-1}+\Sigma_2^{*-1}\Sigma_1^*\Sigma_2^{*-1}-2\Sigma_2^{*-1}\Sigma_2^*\Sigma^{*-1}=\Sigma_2^{*-1}\Sigma_1$ $\Sigma_2^{*-1}-\Sigma^{*-1}$. Since $E\left(gg^\top\right)$ is semi-positive definite, we then verify that $\Sigma_2^{*-1}\Sigma_1\Sigma_2^{*-1}-\Sigma^{*-1}$ is semi-positive definite.

S3 Proof of Theorem 3

For each $\widehat{\theta}_{\text{IPW},k}$ $(1 \le k \le K)$, we apply the Taylor expansion and have

$$\dot{\mathcal{L}}_{\text{IPW},k}\big(\widehat{\theta}_{\text{IPW},k}\big) = \dot{\mathcal{L}}_{\text{IPW},k}\big(\theta^*\big) + \ddot{\mathcal{L}}_{\text{IPW},k}\big(\theta^*\big)\Big(\widehat{\theta}_{\text{IPW},k} - \theta^*\Big) + \frac{1}{2}\Gamma_{\text{IPW},k},$$

where $\Gamma_{\text{IPW},k} = \big(\Gamma_{\text{IPW},k,j}\big) \in \mathbb{R}^{p+1}$, $\Gamma_{\text{IPW},k,j} = (\widehat{\theta}_{\text{IPW},k} - \theta^*)^\top \Delta_{\text{IPW},k,j}(\widehat{\theta}_{\text{IPW},k} - \theta^*)$ $\Delta_{\text{IPW},k,j} = \big(\Delta^{(j_1 j_2)}_{\text{IPW},k,j}\big) \in \mathbb{R}^{(p+1)\times(p+1)}$ for $1 \le j \le p+1$, $\Delta^{(j_1 j_2)}_{\text{IPW},k,j} = \partial^2 \dot{\mathcal{L}}_{\text{IPW},k,j}(\theta)/\partial\theta_{j_1}\partial\theta_{j_2}|_{\theta=\widetilde{\theta}_k}$ for $1 \le j_1, j_2 \le p+1$, $\dot{\mathcal{L}}_{\text{IPW},k,j}(\theta)$ is the j th element of $\dot{\mathcal{L}}_{\text{IPW},k}(\theta)$, and $\widetilde{\theta}_k = \eta_k\widehat{\theta}_{\text{IPW},k} + (1-\eta_k)\theta^*$ for some $0 \le \eta_k \le 1$. In addition, we define $\Gamma^*_{\text{IPW},k} = (\Gamma^*_{\text{IPW},k,j})$, $\Gamma^*_{\text{IPW},k,j} = (\widehat{\theta}_{\text{IPW},k} - \theta^*)^\top \Delta^*_{\text{IPW},k,j}(\widehat{\theta}_{\text{IPW},k} - \theta^*)$ and $\Delta^*_{\text{IPW},k,j} = \partial^2 \dot{\mathcal{L}}_{\text{IPW},k,j}(\theta)/\partial\theta_{j_1}\partial\theta_{j_2}|_{\theta=\theta^*}$. We also define $\Gamma^{\Omega}_{\text{IPW},k} = (\Gamma^{\Omega}_{\text{IPW},k,j}) \in \mathbb{R}^{p+1}$ with $1 \le j \le p+1$, $\Gamma^{\Omega}_{\text{IPW},k,j} = ne^{\alpha^*_N}(\widehat{\theta}_{\text{IPW},k} - \theta^*)^\top \Omega_{\text{IPW},j}(\widehat{\theta}_{\text{IPW},k} - \theta^*)$ and $\Omega_{\text{IPW},j} = E\{\Delta^*_{\text{IPW},k,j}/(ne^{\alpha^*_N})\}$. More specifically, by definition we have $\mathcal{L}_{\text{IPW},k}(\theta) = \sum_{i=1}^N \big\{Y_i Z_i^\top\theta - \big(a_i^{(k)}K - a_i^{(k)}Y_iK + Y_i\big)\log(1+e^{Z_i^\top\theta})\big\}$, Thus, we have $\dot{\mathcal{L}}_{\text{IPW},k}(\theta) = \sum_{i=1}^N \big\{Y_i - \big(a_i^{(k)}K + Y_i - a_i^{(k)}Y_iK\big)p_i(\alpha_N,\beta)\big\}Z_i$, $\ddot{\mathcal{L}}_{\text{IPW},k}(\theta) = -\sum_{i=1}^N (a_i^{(k)}K + Y_i - a_i^{(k)}Y_iK)p_i(\alpha_N,\beta)$ $\big\{1 - p_i(\alpha_N,\beta)\big\}Z_iZ_i^\top$, and $\Delta_{\text{IPW},k,j} = -\sum_{i=1}^N \big(a_i^{(k)}K + Y_i - a_i^{(k)}Y_iK\big)p_i(\alpha_N,\beta)$ $\{1 - p_i(\alpha_N,\beta)\}\{1 - 2p_i(\alpha_N,\beta)\}Z_{i,j}Z_iZ_i^\top$, respectively. Then, we have

$$\widehat{\theta}_{\text{IPW},k} - \theta^* = -\Big\{\ddot{\mathcal{L}}_{\text{IPW},k}\big(\theta^*\big)\Big\}^{-1}\dot{\mathcal{L}}_{\text{IPW},k}\big(\theta^*\big) - \frac{1}{2}\Big\{\ddot{\mathcal{L}}_{\text{IPW},k}\big(\theta^*\big)\Big\}^{-1}\Gamma_{\text{IPW},k}.$$

Subsequently, we should decompose the IPW estimator into a total of six different terms as follows

$$\widehat{\theta}_{\mathrm{IPW}} - \theta^* = Q_{\mathrm{IPW},1} + Q_{\mathrm{IPW},2} + Q_{\mathrm{IPW},3}/2 + Q_{\mathrm{IPW},4}/2 + Q_{\mathrm{IPW},5}/2 + Q_{\mathrm{IPW},6}/2,$$

where $Q_{\mathrm{IPW},1} = K^{-1}\Sigma^{*-1}\sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{IPW},k}(\theta^*)/(Ne^{\alpha_N^*})$, $Q_{\mathrm{IPW},2} = (Ne^{\alpha_N^*})^{-1}K^{-1}$ $\sum_{k=1}^K(\widehat{\Sigma}_{\mathrm{IPW},k}^{-1} - \Sigma^{*-1})\dot{\mathcal{L}}_{\mathrm{IPW},k}\big(\theta^*\big)$, $Q_{\mathrm{IPW},3} = K^{-1}\sum_{k=1}^K \Sigma^{*-1}\{\Gamma_{\mathrm{IPW},k}^{\Omega}/(Ne^{\alpha_N^*})\}$, $Q_{\mathrm{IPW},4} = K^{-1}\sum_{k=1}^K(\widehat{\Sigma}_{\mathrm{IPW},k}^{-1} - \Sigma^{*-1})\{\Gamma_{\mathrm{IPW},k}^*/(Ne^{\alpha_N^*})\}$, $Q_{\mathrm{IPW},5} = (Ne^{\alpha_N^*})^{-1}$ $K^{-1}\sum_{k=1}^K \widehat{\Sigma}_{\mathrm{IPW},k}^{-1}(\Gamma_{\mathrm{IPW},k} - \Gamma_{\mathrm{IPW},k}^*)$, $Q_{\mathrm{IPW},6} = K^{-1}\sum_{k=1}^K \Sigma^{*-1}\{(\Gamma_{\mathrm{IPW},k}^* - \Gamma_{\mathrm{IPW},k}^{\Omega})/(Ne^{\alpha_N^*})\}$ and $\widehat{\Sigma}_{\mathrm{IPW},k} = -(Ne^{\alpha_N^*})^{-1}\ddot{\mathcal{L}}_{\mathrm{IPW},k}\big(\theta^*\big)$. We would investigate the asymptotic property of the US estimator in the following two steps. In the 1st step, we show that $\sqrt{Ne^{\alpha_N^*}}Q_{\mathrm{IPW},1} \to_d N\big(0, \Sigma^{*-1}\big)$ as $N \to \infty$. In the 2nd step, we demonstrate that $Q_{\mathrm{IPW},2} = o_p\{(Ne^{\alpha_N^*})^{-1/2}\}$. The technical details for computing $Q_{\mathrm{IPW},3}$–$Q_{\mathrm{IPW},6}$ are very similar with those of $Q_{\mathrm{R},3}$–$Q_{\mathrm{R},6}$. We thus provide the proof details for $Q_{\mathrm{IPW},1}$ and $Q_{\mathrm{IPW},2}$ only. The proofs for $Q_{\mathrm{IPW},3}$–$Q_{\mathrm{IPW},6}$ are omitted to save space.

STEP 1. Note that $\sum_{k=1}^K a_i^{(k)} = 1$ for $1 \le i \le N$, we have

$$\sum_{k=1}^K \dot{\mathcal{L}}_{\mathrm{IPW},k}\big(\theta^*\big) = \sum_{i=1}^N \Big\{KY_i - \Big(K + KY_i - KY_i\Big)p_i\big(\alpha_N, \beta\big)\Big\}Z_i = K\dot{\mathcal{L}}\big(\theta^*\big).$$

Consequently, we have $Q_{\mathrm{IPW},1} = (Ne^{\alpha_N^*})^{-1}\Sigma^{*-1}\dot{\mathcal{L}}\big(\theta^*\big)$. According to Theorem 1 in Wang (2020), we know that $\big(Ne^{\alpha_N^*}\big)^{-1/2}\dot{\mathcal{L}}\big(\theta^*\big) \to_d N\big(0, \Sigma^*\big)$ when $N \to \infty$, where $\Sigma^* = E\big(e^{X_i^\top\beta^*}Z_iZ_i^\top\big)$. This suggests that $\big(Ne^{\alpha_N^*}\big)^{-1/2}\Sigma^{*-1}$ $\dot{\mathcal{L}}\big(\theta^*\big) \to_d N\big(0, \Sigma^{*-1}\big)$ when $N \to \infty$.

STEP 2. We prove in this step $Q_{\mathrm{IPW},2} = o_p\{(Ne^{\alpha_N^*})^{-1/2}\}$. Recall $Q_{\mathrm{IPW},2} = K^{-1}\sum_{k=1}^K(\widehat{\Sigma}_{\mathrm{IPW},k}^{-1} - \Sigma^{*-1})\dot{\mathcal{L}}_{\mathrm{IPW},k}(\theta^*)/(Ne^{\alpha_N^*})$. Consequently, we have $\|Q_{\mathrm{IPW},2}\| \le \widetilde{Q}_{2a}\widetilde{Q}_{2b}$, where $\widetilde{Q}_{2a} = \max_k \|\widehat{\Sigma}_{\mathrm{IPW},k}^{-1} - \Sigma^{*-1}\|$ and $\widetilde{Q}_{2b} = K^{-1}\sum_{k=1}^K \|\dot{\mathcal{L}}_{\mathrm{IPW},k}(\theta^*)/(Ne^{\alpha_N^*})\|$. By the similar arguments used by (S1.2), we have $\widetilde{Q}_{2a} = o_p(1)$. We then focus on proving $\widetilde{Q}_{2b} = O_p\{(Ne^{\alpha_N^*})^{-1/2}\}$. We have $\{K^{-1}\sum_{k=1}^K \|(Ne^{\alpha_N^*})^{-1}\dot{\mathcal{L}}_{\mathrm{IPW},k}(\theta^*)\|\}^2 \le (Ne^{\alpha_N^*})^{-1}\sum_{k=1}^K \|\dot{\mathcal{L}}_{\mathrm{IPW},k}(\theta^*)\|^2$ K^{-1} by Jensen's inequality. Moreover, we know that $K^{-1}\sum_{k=1}^K \|(Ne^{\alpha_N^*})^{-1}$ $\dot{\mathcal{L}}_{\mathrm{IPW},k}(\theta^*)\|^2 \ge 0$. Then, it is sufficient to compute $E\{K^{-1}\sum_{k=1}^K \|(Ne^{\alpha_N^*})^{-1}$ $\dot{\mathcal{L}}_{\mathrm{IPW},k}(\theta^*)\|^2\}$. We have

$$
\begin{aligned}
&E\left\{K^{-1}\sum_{k=1}^K \left\|\left(Ne^{\alpha_N^*}\right)^{-1}\dot{\mathcal{L}}_{\mathrm{IPW},k}\left(\theta^*\right)\right\|^2\right\} \\
&= K^{-1}\sum_{k=1}^K \mathrm{tr}\left[E\left\{\left(Ne^{\alpha_N^*}\right)^{-2}\dot{\mathcal{L}}_{\mathrm{IPW},k}\left(\theta^*\right)\dot{\mathcal{L}}_{\mathrm{IPW},k}\left(\theta^*\right)^\top\right\}\right]. \qquad \text{(S3.20)}
\end{aligned}
$$

Furthermore, we have

$$
\begin{aligned}
&E\left\{\left(Ne^{\alpha_N^*}\right)^{-1}\dot{\mathcal{L}}_{\mathrm{IPW},k}\left(\theta^*\right)\dot{\mathcal{L}}_{\mathrm{IPW},k}\left(\theta^*\right)^\top\right\} = \mathrm{cov}\left\{\left(Ne^{\alpha_N^*}\right)^{-1/2}\dot{\mathcal{L}}_{\mathrm{IPW},k}\left(\theta^*\right)\right\} \\
&= \left(Ne^{\alpha_N^*}\right)^{-1}N\mathrm{cov}\left[\left\{Y_i\left(1-p_i\left(\alpha_N^*,\beta^*\right)\right) - a_i^{(k)}\left(1-Y_i\right)p_i\left(\alpha_N^*,\beta^*\right)K\right\}Z_i\right] \\
&= e^{-\alpha_N^*}E\left[\left\{Y_i\left(1-p_i\left(\alpha_N^*,\beta^*\right)\right)^2 + a_i^{(k)}\left(1-Y_i\right)p_i^2\left(\alpha_N^*,\beta^*\right)K^2\right\}Z_iZ_i^\top\right] \\
&= e^{-\alpha_N^*}E\left[p_i\left(\alpha_N^*,\beta^*\right)\left\{1-p_i\left(\alpha_N^*,\beta^*\right)\right\}\left\{1-p_i\left(\alpha_N^*,\beta^*\right)+Kp_i\left(\alpha_N^*,\beta^*\right)\right\}Z_iZ_i^\top\right] \\
&= E\left\{\frac{1+e^{X_i^\top\beta^*+\alpha_N^*+\log K}}{\left(1+e^{X_i^\top\beta^*+\alpha_N^*}\right)^3}e^{X_i^\top\beta^*}Z_iZ_i^\top\right\}. \qquad \text{(S3.21)}
\end{aligned}
$$

Combining (S3.20) and (S3.21), we have

$$
\begin{aligned}
&E\left\{K^{-1}\sum_{k=1}^{K}\left\|\left(Ne^{\alpha_N^*}\right)^{-1}\dot{\mathcal{L}}_{\mathrm{IPW},k}\left(\theta^*\right)\right\|^2\right\}\\
&=\left(Ne^{\alpha_N^*}\right)^{-1}\mathrm{tr}\left[E\left\{\frac{1+e^{X_i^\top\beta^*+\alpha_N^*+\log K}}{\left(1+e^{X_i^\top\beta^*+\alpha_N^*}\right)^3}e^{X_i^\top\beta^*}Z_iZ_i^\top\right\}\right]\\
&=\left(Ne^{\alpha_N^*}\right)^{-1}E\left\{\left(1+\gamma e^{X_i^\top\beta^*}\right)e^{X_i^\top\beta^*}\left\|Z_i\right\|^2\right\}\left\{1+o(1)\right\}.
\end{aligned}
$$

One can verify that $(1+e^{X_i^\top\beta^*+\alpha_N^*+\log K})(1+e^{X_i^\top\beta^*+\alpha_N^*})^{-3}e^{X_i^\top\beta^*}\|Z_i\|^2$ converges to $(1+\gamma e^{X_i^\top\beta^*})e^{X_i^\top\beta^*}\|Z_i\|^2$ almost surely as $N\to\infty$, and $(1+e^{X_i^\top\beta^*+\alpha_N^*+\log K})(1+e^{X_i^\top\beta^*+\alpha_N^*})^{-3}e^{X_i^\top\beta^*}\|Z_i\|^2\le\big(1+\gamma e^{X_i^\top\beta^*}\big)e^{X_i^\top\beta^*}\|Z_i\|^2$ with $E\big\{\big(1+\gamma e^{X_i^\top\beta^*}\big)e^{X_i^\top\beta^*}\|Z_i\|^2\big\}<\infty$. Here $E\big(e^{X_i^\top\beta^*}\|Z_i\|^2\big)\le M_t$ with sufficiently large but fixed constant t by (S.1) in Wang (2020) and Step 1 in S1. Thus, from the dominated convergence theorem (Royden and Fitzpatrick, 1988), the last equality holds. We then obtain $(Ne^{\alpha_N^*})^{-1}\sum_{k=1}^{K}\|\dot{\mathcal{L}}_{\mathrm{IPW},k}\big(\theta^*\big)\|^2 K^{-1}=O_p\{(Ne^{\alpha_N^*})^{-1}\}$. This suggests that $\widetilde{Q}_{2b}=O_p\{(Ne^{\alpha_N^*})^{-1/2}\}$. This finally leads to $Q_{\mathrm{IPW},2}=o_p\{(Ne^{\alpha_N^*})^{-1/2}\}$.

Bibliography

Bernstein, S. (1926), *Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle: professées à la Aorbonne*, vol. 30, Gauthier-Villars.

Jordan, M. I., Lee, J. D., and Yang, Y. (2018), "Communication-efficient distributed statistical inference," *Journal of the American Statistical Association*.

Royden, H. L. and Fitzpatrick, P. (1988), *Real analysis*, vol. 32, Macmillan New York.

Schenker, N. and Welsh, A. H. (1988), "Asymptotic results for multiple imputation," *The Annals of Statistics*, 16, 1550–1566.

Van der Vaart, A. W. (2000), *Asymptotic statistics*, vol. 3, Cambridge university press.

Wang, H. (2020), "Logistic Regression for Massive Data with Rare Events," in *International Conference on Machine Learning*, PMLR, pp. 9829–9836.

Zhang, H. and Chen, S. X. (2020), "Concentration inequalities for statistical inference," *arXiv preprint arXiv:2011.02258*.